

Organization of intracellular reactions with rationally designed scaffolding systems



THESE 2009

| | |
|----------------------------|---|
| Title of the project | Organization of intracellular reactions with rationally designed scaffolding systems |
| PhD student | Camille Delebecque |
| University-Doctoral school | Frontières du Vivant (FdV), Faculté de Médecine, Paris-Descartes |
| Supervisors | -Pamela Silver, System Biology Department, Harvard Medical School, 200 Longwood avenue, Boston MA 02115, USA - Ariel Lindner, Evolutionary Systems Biology, Faculté de Médecine Paris-Descartes, 24 rue du faubourg Saint-Jacques, 75014 Paris, France |
| Laboratories of welcome | - System Biology Department, Harvard Medical School, Boston - Centre de Recherches Interdisciplinaires, (CRI), Faculté de Médecine , Université Paris-Descartes |
| Responsibles for thesis | - Pamela Silver, Boston - Ariel Lindner, Paris |
| Duration | Trois ans (fin 2009 à fin 2012). Thèse soutenue le 15 Novembre 2012 à Paris |

Summary

Hydrogen has been considered a potential fuel for the future since it is carbon-free and oxidized to water. In this project we intend to address the difficulty of biologically producing hydrogen using bacteria as micro and fully automatized factories. Using tools of molecular biology, synthetic biology and genetic engineering we aim at creating within the bacterium cell a localized anaerobic microenvironment suitable for the work of highly oxygen sensitive hydrogenases.

Alongside with the innovative synthetic biology approach, this system has lots of promises concerning hydrogen yields and as well as economically viable industrial applications.

Results

In cells bio-enzymatic pathways are often spatially organized into complexes, into organelles or onto protein scaffolds. Spatial organization limits diffusion and helps channels substrates between enzymatic cores, limiting competing reactions, insulating and increasing yields of sequential metabolic reactions. In this PhD thesis work, we engineered new tools to control the precise spatial organization of enzymes and increase the titer of specific pathways.

We design and engineer "artificial organelles" made of assembling RNA nanostructures. These scaffolds are made out of assembling non-coding RNA molecules we specifically design to polymerize into multi-dimensional

nanostructures inside bacterial cells. These structures have docking sites to target enzymes onto them and control their respective distance and stoichiometry.

We demonstrate the validity of our approach by optimizing and improving the production of biohydrogen and designing a protocol to simplify and standardize the use of RNA scaffold. Moreover, we develop a new synthetic biology “chassis” by developing strategies to engineer *Anabaena* PCC7120 and control the spatial localization of metabolic pathway at the cellular level. By targeting specific enzymes into oxygen-depleting heterocysts, metabolic engineers can now implement oxygen-sensitive pathways into oxygen evolving cyanobacteria. This PhD work opens the door to an array of new applications spanning synthetic biology, structural biology to nanotechnology.

Livrables

[Manuscrit de la thèse](#)

[Page de couverture de la thèse](#)

Publications

Camille J Delebecque et al. (2011) Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science*, 333, 470

Camille J Delebecque, Pamela A Silver and Ariel B Lindner (2012) Designing and using RNA scaffolds to assemble proteins in vivo. *Nature Protocols*, 7, 10, 1797-1807

Contacts

Pamela A. Silver

Harvard Medical School, Boston

pamela_silver@hms.harvard.edu

Ariel Lindner

Faculté de Médecine Paris-Descartes

ariel.lindner@inserm.fr

UNIVERSITÉ PARIS-DESCARTES

Ecole Doctorale - Frontières du Vivant

DOCTORAT

Discipline: Sciences de la vie et de la Santé

Spécialité: Biologie Synthétique

Camille Delebecque

Organization of intracellular reactions
with rationally designed scaffolding systems.

Directeur: Ariel Lindner & Pamela Silver

Soutenue le 15 Novembre 2012

Jury:

Dr. Tom Ellis

Pr. Eva Pebay-Peloubay

Pr. Miroslav Radman

Dr. Vincent Schachter

Dr. Gilles Truan

Acknowledgments

I remember a time when as a kid I was trying to calculate when I would be done with my studies, already with the specific goal in mind to become a doctor. I remember this feeling of trying to apprehend what looked like a n eternity, trying to imagine what the world would look like then, and what I would be looking like. Surely, we would have recovered from a disastrous Kyoto meeting, surely Cousteau's concerned words about the environment would have been listened to, surely we would have reached and pasted the oil peak production, surely by now we should have green cars. The world is changing and one must step out from being just a spectator and actively gear this change towards the right direction.

Many people I met along the way were great source of inspiration and I need to thank them for helping me, in some ways or another, getting where I am now. On the welcoming shoulders of these giants, I have been trying to understand where and what they are looking at while bringing my own twisted view to the equation. My first words need to go to Ariel Lindner and Pamela Silver for welcoming me in their laboratories, supporting and guiding me along the way. The craziness of this PhD project with its melting pot of scientific disciplines could have not been possible anywhere else in the world. Many many thanks. Then, there is François Taddei and his twisted views on education that brought me to this unique community that is the Center for Research and Interdisciplinary in Paris. People motivated to make things happen and change the world, no matter the obstacles. Before, at AgroParisTech and during my preparatory classes at Descartes I was given the chance to nurture my budding interest for research and was given great food of thoughts by Francis Denise and Bruno Latour. Thank you all.

Of course, sailing across a scientific ocean is not a personal endeavor, and I need to thank all of my friends in the Silver Lab and Taddei lab that contributed to make a very special favorable and enjoyable working environment. I need to specifically mention here two post-docs I worked closely with and that were a great source of inspiration and help. Thank you Faisal Aldaye and Daniel Ducat for being part of this adventure, I had a great time working together.

Then, there is family and friends. Thank you Mom, Dad and Lou for pushing me to cross oceans and become an eager and curious Camille. I need to thank my girlfriend Jamie for being a great support and for not being jealous of my bacteria I was sometimes spending much more time with. I need to thank my great friends Pierre Thireau, Jean-Philippe Krieger, Virgile Baudry, Océane Albert, Clémentine Roux, Denis Losdat, Daphney Dagneau, Bérengère Bouguet and many many others for being a source of inspiration, for the great times we have spent together and the many others to come.

World, bring it on!

Camille, September 2012

You must be the change you wish to see in the world.

- Mahatma Gandhi

Abstract

In cells bio-enzymatic pathways are often spatially organized into complexes, into organelles or onto protein scaffolds. Spatial organization limits diffusion and helps channels substrates between enzymatic cores, limiting competing reactions, insulating and increasing yields of sequential metabolic reactions. In this PhD thesis work, we engineered new tools to control the precise spatial organization of enzymes and increase the titer of specific pathways.

We design and engineer “artificial organelles” made of assembling RNA nanostructures. These scaffolds are made out of assembling non-coding RNA molecules we specifically design to polymerize into multi-dimensional nanostructures inside bacterial cells. These structures have docking sites to target enzymes onto them and control their respective distance and stoichiometry. We demonstrate the validity of our approach by optimizing and improving the production of biohydrogen and designing a protocol to simplify and standardize the use of RNA scaffold. Moreover, we develop a new synthetic biology “chassis” by developing strategies to engineer *Anabaena PCC7120* and control the spatial localization of metabolic pathway at the cellular level. By targeting specific enzymes into oxygen-depleting heterocysts, metabolic engineers can now implement oxygen-sensitive pathways into oxygen evolving cyanobacteria. This PhD work opens the door to an array of new applications spanning synthetic biology, structural biology to nanotechnology.

Résumé

Au sein des cellules, les voies enzymatiques sont souvent organisées spatialement sous forme de complexes, sur des structures protéiques ou dans des micro-compartiments. Cette organisation spatiale aide au déroulement optimal des réactions enzymatiques en limitant les pertes d'intermédiaires métaboliques, en isolant les voies de signalisations et en augmentant le rendement des réactions enzymatiques. Dans ce travail de thèse nous avons étudié la possibilité de créer des outils permettant de contrôler et optimiser de novo l'organisation spatiale de voies métaboliques *in vivo*.

Nous avons dessiné et assemblé des structures d'ARN non codants utilisées comme support pour organiser le métabolisme bactérien. Ces ARNs s'assemblent spontanément *in vivo* en des structures à une ou deux dimensions avec des sites distincts d'attachement protéique. Nous démontrons l'utilité de cette approche via l'optimisation d'une voie enzymatique de synthèse de biohydrogène et démocratisons l'utilisation de ces structures d'ARN en développant un protocole simplifié. Nous étendons cette étude à d'autres stratégies d'organisation, notamment via l'ingénierie des cellules spécialisées dans la fixation de l'azote atmosphérique de la cyanobactérie *Anabaena PCC7120*, les hétérocystes. Ce travail de thèse ouvre de nouvelles portes à la biologie de synthèse à la biologie structurale et aux nanotechnologies.

Contents

ACKNOWLEDGMENTS

ABSTRACT / RÉSUMÉ

Table of Figures **14**

PREAMBLE **18**

INTRODUCTION **29**

I - Intra-cellular spatial organization & Synthetic scaffolding Strategies
33

The need for Intracellular Organization.....33

Challenges faced by metabolism

Concentration gradients and substrate channeling

Nature's Solutions36

Enzyme complexes

Micro-compartments

Cellular differentiation

Emerging Synthetic Scaffold approach.....41

Using and engineering organelles and microcompartments

Using and engineering protein scaffolds

Using and engineering nucleic acids scaffolds

II - From the RNA world hypothesis to RNA nanotechnologies **47**

RNA World Hypothesis47

The primordial RNA world hypothesis

The modern RNA world

| | |
|---|-----------|
| RNA Nanotechnology..... | 51 |
| Structural properties and modularity of RNA nanostructures | |
| Strategies for programmable RNA self-assembly | |
| Nanostructures used as scaffolds | |
| III - Doing Synthetic Biology with RNA | 59 |
| Tools for engineering RNA devices..... | 59 |
| Harvesting Nature's components - the world of RNA Aptamers | |
| Evolving new RNA components - SELEX | |
| Computational approach | |
| RNA Control Devices in synthetic biology | 64 |
| Riboregulator and Riboswitch-based devices | |
| Applications in synthetic biology | |
| IV - The challenges of biologically producing hydrogen | 68 |
| Bio-hydrogen, an attractive biofuel | 68 |
| Biofuel economy | |
| Hydrogen advantages | |
| Strategies for biologically producing hydrogen..... | 71 |
| Light-driven hydrogen production | |
| Fermentation-based hydrogen production | |
| Designing a synthetic electron transfer pathway | 74 |
| Spatial organization of metabolic electron transfers | |
| Synthetic bio-hydrogen pathways | |
| RESULTS | 78 |
| V - Organization of intracellular pathways with rationally designed RNA Assemblies | 79 |
| Introduction..... | 80 |
| Engineering RNA scaffolds | 81 |
| In vivo and in vitro characterization of RNA scaffolds | 83 |

| | |
|---|-----------|
| Organizing proteins onto RNA scaffolds | 86 |
| VI - Standardizing the construction and use of RNA scaffolds | 91 |
| Introduction..... | 92 |
| Natural and engineered ncRNA | |
| Advantages and applications of synthetic RNA scaffolds | |
| Experimental Design | 95 |
| RNA scaffold design; RNAdesigner (Steps 1-8). | |
| RNA scaffold design; choosing aptamers to tether proteins to the RNA scaffolds (Steps 1-2). | |
| RNA scaffold design; choosing a folding scheme (Steps 3-8). | |
| RNA scaffold optimization (Steps 9-11). | |
| Cloning the designed RNA scaffold into an expression system (Steps 12-24). | |
| Expression systems (Steps 25-27). | |
| DNA capture probes for scaffold expression analysis by in vivo pull-down (Step 27Biii). | |
| Primer design for quantitative RT-PCR (Step 27Ciii). | |
| Stability considerations. | |
| Targeting proteins onto the RNA scaffold (Steps 28-30). | |
| Controls. | |
| Materials..... | 107 |
| Reagents | |
| Equipment | |
| Bioinformatics RNA Design tools: | |
| Protocol | 111 |
| RNA scaffold design - Timing 2 h | |
| RNA scaffold optimization - Timing 1 h | |
| Cloning the designed RNA scaffold into an expression system TIMING 12-15 d | |
| Induction of RNA scaffold expression - Timing 1-2 d | |
| Expression analysis | |
| Targeting Proteins onto the RNA Scaffold TIMING 2-3d | |
| Timing | |
| Troubleshooting | |

Anticipated Results

| | |
|---|------------|
| VII - METHODS | 128 |
| Introduction..... | 128 |
| In silico characterization of RNA structures..... | 129 |
| RNA folding | |
| Determination of RNA Structures | |
| Challenges in designing RNA Scaffolds | 131 |
| Discrete vs. polymerizing scaffolds | |
| Strategies for scaffold polymerization | |
| Probing the stability of RNA scaffolds | 138 |
| Considerations about RNA half life in E.coli | |
| Developing a qPCR strategy for short and highly complex ncRNAs | |
| Purifying RNA scaffolds | 140 |
| In vitro transcription | |
| Developing a pull-down protocol for in vivo purification | |
| Imaging the Scaffolds | 142 |
| Using fluorophores | |
| Using Electron Microscopy | |
| VIII - DISCUSSION | 148 |
| Small synthetic ncRNA as tools to study protein-protein interactions..... | 149 |
| Building a new in vivo scaffolding reporter system | |
| Controlling protein relative distance, orientation and complex sizes | |
| E. coli DsrA ncRNA - a naturally assembling RNA? | |
| In vivo Assembly..... | 161 |
| Dark matter RNA..... | 166 |
| Saving the world with hydrogen? | 168 |
| APPENDIX | 172 |
| Appendix I - Engineering Anabaena's Heterocysts | 173 |

| | |
|--|------------|
| Heterocysts - an attractive natural compartmentalization strategy to engineer. | 173 |
| Cyanobacterial heterocysts, a striking example of prokaryotic cell differentiation | |
| Gene expression and morphogenesis during heterocyst development. | |
| Challenges in engineering <i>Anabaena</i> PCC 7120. | 178 |
| <i>Anabaena</i> cultures | |
| Developing a working transformation protocol | |
| Controlling the expression of heterologous proteins in time, space and intensity | 181 |
| Developing a promoter library | |
| FACS characterization | |
| Optical Microscopy characterization of an endogenous promoter library | |
| Appendix II - Science Paper | 191 |
| Appendix III - Nature Protocol Paper | 196 |
| Supporting Material | 208 |
| Plasmid and strains | 208 |
| RNA scaffolds: aptamer domains | 209 |
| RNA scaffolds: synthesis | 209 |
| RNA scaffolds: mutations to prevent self-assembly | 210 |
| RNA scaffolds: mutations to prevent protein binding | 210 |
| Atomic force microscopy | 211 |
| RNA scaffolds: inhibitory strands | 212 |
| PP7 / MS2 proteins | 212 |
| Protein production levels | 212 |
| Gel-shift assays | 213 |
| Hydrogen constructs | 215 |
| Hydrogen production assays | 215 |
| Growth curves | 217 |
| Strains, plasmids, sequences | 219 |
| References | 224 |

Table of Figures

Figure 1: “*What I cannot create, I do not understand*”.

Figure 2: The popular iGEM blackboard.

Figure 3: Schematic of the cellulosome architecture.

Figure 4: Physical organization of the carboxysome.

Figure 5: Heterocysts in nitrogen starved *Anabaena PCC7120*.

Figure 6: Synthetic protein-based scaffold.

Figure 7: Examples of RNA tertiary structures.

Figure 8: A-form and B-form RNA and DNA helices.

Figure 9: pRNA-based assemblies.

Figure 10: Diversity of RNA and DNA nanostructures.

Figure 11: MS2 and PP7 protein crystal structure.

Figure 12: Systematic Evolution of Ligands by Exponential Enrichment (SELEX).

Figure 13: Diversity of applications of RNA-based controllers in synthetic biology.

Figure 14: Ethanol and Biodiesel Production from 2000 to 2010.

Figure 15: Hydrogen producing fermentation pathways.

Figure 16: Chloroplast thylakoid membrane and associated electron transfer through the photosynthetic machinery.

Figure 17: Design of RNA modules to organize proteins.

Figure 18: Characterization of RNA assemblies.

Figure 19: Fluorescence protein complementation *in vivo*.

Figure 20: Scaffolding hydrogen production.

Figure 21: General workflow for design, induction, expression and experimental testing of the RNA scaffold.

Figure 22: Designing and optimizing RNA scaffolds using RNAdesigner.

Figure 23: Compatible aptamer pair.

Figure 24: Examples of RNA scaffold folding scheme designs.

Figure 25: Modular cloning approach to the design of RNA Scaffolds.

Figure 26: Example of RNAdesigner input and output sequences and structures.

Figure 27: Troubleshooting Table.

Figure 28: Scheme illustrating the approach used to design the RNA modules.

Figure 29: Schematic representation of the steps involved in designing the self-assembling RNA module d₁.

Figure 30: Secondary structures of the RNA building blocks d₀, d₁, d₂', and d₂'.

Figure 31: Primary sequences for the discrete, one- and two-dimensional modules.

Figure 32: qRT-PCR of RNA modules.

Figure 33: Initial experiments of RNA scaffold assisted fluorescence complementation.

Figure 34: Gold elemental mapping of RNA scaffold *in vivo*

Figure 35: Simplified bacterial luciferase biochemical reaction.

Figure 36: Bioluminescence assay of scaffolded vs. unscaffolded luciferase pathway.

Figure 37: Discrete RNA scaffold variant library.

Figure 38: Computed secondary structure of a large discrete scaffold.

Figure 39: DsrA RNA polymers

Figure 40: Table of DsrA design variants secondary structures

Figure 41: Evolving assembling RNA. Schematic of a tailored SELEX process.

Figure 42: Evolving Assembling RNAs. Selection of *in vivo* polymerizing RNAs.

Figure 43: *Anabaena PCC7120* and heterocyst development.

Figure 44: Heterocyst differentiation and pattern formation.

Figure 45: *Anabaena* growth assays.

Figure 46: Table of all the characterized *Anabaena* promoters.

Figure 47: FACS Characterization of wild-type fragmented *Anabaena PCC 7120*.

Figure 48: Spatiotemporal characterization of the *Anabaena* promoter library via fluorescence imaging.

Figure 49: Protein production levels.

Figure 50: Protein binding assay onto RNA scaffold via reverse gel shift assay.

Figure 51: GC chromatographs of hydrogen production.

Figure 52: Growth curves of *E. coli* cells expressing RNA scaffolds

Figure 53: Comparison of RNA production levels in different BL21 cells.

Figure 54: Plasmid, sequences and strain tables.

“Man can go Nature one better”

Luther Burbank (1923)

“Evolution has to become an experimental science, which must first be controlled and studied then conducted and finally shaped to the use of man.”

Hugo de Vries, 1904



PREAMBLE

“So...what is it you’re actually studying again?”

Synthetic biology that is. The engineering of living organisms at the molecular level to have them execute and perform novel functions (Endy 2005). This short definition might satisfy most scientists, but is probably far from satisfying the broader public stuck by the apparent conflict in between the two terms “synthetic” and “biology”.

Historical context is of a great help to soften the apparent conceptual clash behind the name of this new field and explain why one can actually think about engineering biology. The non-linearity in scientific progress (Kuhn 1996), unroots many of the modern

scientific fields from their own weird and tortuous historical heritage - this is especially true of synthetic biology, most people dating its birth about a decade ago.

Yet, it is probably this ignored heritage that best conceptualizes and explains this new scientific field - and helps come with a satisfactory answer during those awkward moments when friends or family ask you “So...what is it you’re actually studying again...?”

In the case of synthetic biology, this unspoken mix of sociology, enthusiasm and scientific advancement that made what the field is today, is still understudied and probably worth a thesis on its own. I will only touch upon it here inspired by the work of Drew Endy (Endy 2005), Luis Campos (Campos 2009) and others, to give a broader context to my PhD work.

Engineering the living

A vast majority of people put synthetic biology’s birth around a decade ago when a bunch of engineers started to look closer at biology. Fewer would maybe go as far as the 70s with the discoveries of restriction enzymes by Daniel Nathans, Werner Arber and Hamilton O. Smith (Roberts 2005). But the idea of engineering biology, the central concept behind synthetic biology, is actually a recurring theme throughout all of the 19th and 20th century.

While the invention of agriculture around 10000 years ago is arguably the first attempt at controlling the living (Allaby et al. 2008), our understanding of the living was profoundly shaken in the 1800s with the rise of Mendelian experimental plant breeding and the development of darwinistic evolutionary theories (Moose & Mumm 2008).

Darwin compiled a huge amount of data on biological diversity and adaptation during his trip on the HMS Beagle. However, the strength of his arguments in his famous book The origin of Species lies mainly on artificial rather than natural evolution through the observation countless breeding examples he reports (Darwin 1859). He later published a book dedicated to the topic detailing the power of artificial selection through artificial evolution (Darwin 1868). This conceptual switch from life fixed and frozen in time to life evolving through natural selection really shook the scientific world.

The concomitant rise of experimental plant breeding is an interesting point to which to trace synthetic biology. It also makes for a nice connection with my former background as an agronomical engineer. A few names are to remember. Luther Burbank was a very successful plant breeder accredited with having created more than 800 new plant varieties throughout his career. Although criticized by his peers for his non-scientific approach, Burbank has to be credited for his progressive conception of life. His citations could easily be today's synthetic biologists' mottos:

"Plant breeding to be successful must be conducted like architecture."

(Burbank 1909)

"Man can go Nature one better." (Burbank 1923)

Hugo de Vries, is another scientist to highlight for his visionary ideas about engineering the living.

"Burbank crosses species, I seek to create new ones"

“Evolution has to become an experimental science, which must first be controlled and studied then conducted and finally shaped to the use of man” (Cold spring harbor laboratory inauguration, 1904)

de Vries is remembered as one of the first geneticists, well known for rediscovering Mendel’s work on the law of heredity in the 1890s and his mutation theory. He could arguably be also considered as one of the founding grandfathers of synthetic biology.

“What I cannot create, I do not understand”

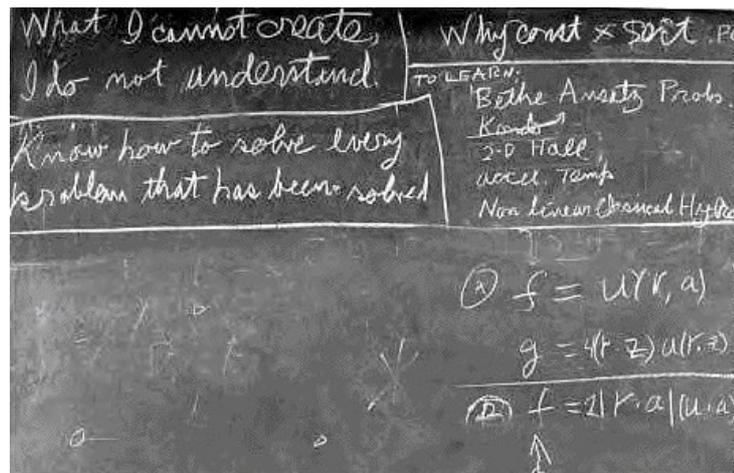


Figure 1: “What I cannot create, I do not understand” - Feynman’s famous scrawl on a Caltech chalkboard in 1988 - Courtesy of the California Institute of Technology.

This famous quote by Richard Feynman is widely used in the Synthetic biology community to convey the idea that descriptive biology has reached some limits (Drubin et al. 2007). This idea applied to the field of biology actually precedes Feynman and was

introduced by a French biologist named Stéphane Leduc in the 1910s. Leduc is also credited with coining the term “synthetic biology”.

Jusqu'à present la biologie n'a eu recours qu'à l'observation et à l'analyse. L'unique utilisation de l'observation et de l'analyse, l'exclusion de la méthode synthétique, est une des causes qui retardent le progrès de la biologie. [La méthode synthétique] semble devoir être la plus féconde, la plus apte à nous révéler les mécanismes physiques des phénomènes de la vie dont l'étude n'est même pas ébauchée.¹

La biologie synthétique représente une méthode nouvelle, légitime, scientifique; la synthèse appliquée à la biologie et une méthode féconde, inspiratrice de recherches; le programme consistant à chercher à reproduire, en dehors des êtres vivants, chacun des phénomènes de la vie suggère immédiatement un nombre infini d'expériences, c'est une direction pour l'activité.²

Leduc was a professor of medicine in Nantes studying and trying to recreate forms “reminiscent of life” using osmotic forces, chemical gradients and diffusion. Leduc’s very distinct approach to biology, postulating that one needs to be able to recreate bio-

¹ Up until now, biology has only been using observation and analysis. The sole use of observation and analysis, excluding the synthetic method, is one of the causes slowing down the progress of biology. [Synthesis] seems to be the most efficient and capable at revealing the physical mechanisms of life, whose understanding has barely started.

² Synthetic biology is a new method, legitimate and scientific; Synthesis applied to biology is a fruitful and inspiring research method; Trying to reproduce, outside of the living, the individual phenomena comprising life will foster an infinite number of experiments, it is a worthwhile research area.

logical phenomena to understand them, is now one of the core principles of modern synthetic biology.

(Partially) understanding the molecular details of biology

Many recent discoveries deeply influenced the birth of synthetic biology and many of the engineering tools now used across synthetic biology labs were developed in the 70s and 80s. They result from a new fundamental understanding of many of biology's core processes.

In 1953, Watson and Crick discovered the molecular structure of DNA which contains the genetic instructions used in the development and functioning of all known living organisms. This opened the door to a new era in biology.

An early milestone in the history of synthetic biology is the discovery of transcriptional feedback and mathematical logic in gene regulation by Monod and Jacob (*i.e.* Lac operon in 1961 (Jacob & MONOD 1961)). This first important grasp at understanding genetic regulation is at the foundation of a central concept in Synthetic biology: **abstraction**. A high level description of biological systems is often useful when engineering biology and is enabled through the mathematical description of genetic networks.

A few years later, the first gene coding for a yeast tRNA was fully synthesized by Har-gobind Khorana and coworkers (Khorana 1979). Restriction enzymes were then discovered and awarded the Nobel price in 1978 to Daniel Nathans, Werner Arber and

Hamilton Smith (Smith & Wilcox 1970; Danna & Nathans 1971). In an editorial comment published in *Gene*, the Polish scientist Waclaw Szybalski wrote:

*“The work on restriction nucleases not only permits us easily to construct recombinant DNA molecules and to analyze individual genes, but also has led us into the **new era of synthetic biology** where not only existing genes are described and analyzed but also new gene arrangements can be constructed and evaluated”*

Indeed, in 1977 the first micro-organism was genetically engineered to produce the hormone somatostatin by Herbert Boyer (Keiichi 1977). This led to the creation of the first modern biotech company, Genentech and the new era of recombinant DNA (I. S. Johnson 1983). Finally, in 1983, PCR was developed by Kary Mullis completing most of the toolbox of the modern synthetic biologist (Boyle & Silver 2009).

And yet, we will have to wait another two decades for the rise of modern synthetic biology.

Engineering + Biology

The emergence of contemporary Synthetic Biology is closely linked to trends in the software industry as well as the social context surrounding it. At the turn of the millennium and after the dot-com bubble crash, the open-source software movement gained momentum. In parallel a number of engineers, including Tom Knight, Ron Weiss, Drew Endy and others were working on amorphous computing, a field applica-

ble to many natural systems including molecular biology and gene networks. The first discussions about “open source biology” and new approaches to engineering biology between Drew Endy, Bob Carlson and Roger Brent take place at the Molecular Sciences Institute in Berkeley in 1999 and were certainly fueled by this whole context (Campos 2009).

Around the same time, another field was gaining momentum, Systems Biology. Although evolution works by random tinkering, it is appearing clearer and clearer that within biological complexity, generalizable principles do emerge. Systems Biology is attempting to discern these patterns and formulates general laws applying to biological networks. The mathematical framework developing with Systems Biology to describe the inherent simplicity of biological systems greatly influenced the growth of Synthetic Biology (Alon 2006).

Yet, Synthetic Biology would only gain momentum with the inaugural “Synthetic Biology 1.0” conference. Charismatic researchers such as Pamela Silver, George Church, James J Collins, Craig Venter, Jay Keasling, Uri Alon amongst others, promoted the growth of the field with ground-breaking works. Interestingly, the student iGEM competition (international Genetically Engineered Machines Competition) also had a major role in democratizing synthetic biology which from then on would also be strongly driven by students. To put things back in context, this is how I got first involved with the field and how synthetic biology arrived in France. The first French iGEM team was hosted by the Center for Interdisciplinary Research in Paris in 2007 as well as the first synthetic biology club in France.

To overcome these challenges, Tom Knight and colleagues introduced the idea of **standardization** in 2004 through the use of “Biobrick Parts” - basic DNA sequences clearly defined, and characterized with a description of standard conditions to use a given Biobrick in combination with others. The Biobrick Parts are made to be easily sharable, bypassing the need for long and painful Material Transfer Agreement, and easily used or reused through a standardized cloning strategy. To support the development of Biobricks, a Registry of Standard Biological Parts was created at MIT and is widely used by iGEM teams.

Along with the idea of **standardization**, the concept of **decoupling** and **abstraction** are at the root of modern synthetic biology:

***Decoupling** is the idea that it is useful to separate a complicated problem into many simpler problems that can be worked on independently, such that the resulting work can eventually be combined to produce a functioning whole. (Endy 2005)*

*Biological engineering **abstraction** hierarchies must (1) allow individuals to work at any one level of complexity without regard for the details that define other levels, yet (2) allow for the principled exchange of limited information across levels. (Endy 2005)*

To some extent, the concept of **standardization** is today already being challenged by the fast evolving cloning technologies such as the Gibson (Gibson et al. 2009) or Golden Gate assembly methods, and more than ever by the decreasing cost of DNA synthesis (Carlson 2009; Ellis et al. 2011). As I started my PhD the synthesis cost was

still around \$0.50/ base pair (bp). In 2012, the prices are around \$0.20/bp with a 4 time faster turnabout (4 working days). Projects that would have required months if not years of cloning can now be done in a few weeks, also helped by the development of great software to assist the scientist such as J5 (Hillson et al. 2012). Synthetic Biology is more than adopting these changes, it is greatly fostering them.

Synthetic Biology is certainly still in its adolescence. However, its ability to embrace sociological, scientific and technological trends promises a bright future with hopefully many innovations getting out of the labs for the greater good.

INTRODUCTION

The Hitchhiker's guide to my thesis

“It is a mistake to think you can solve any problem with just potatoes” (Adams 1979). Indeed. And with this in mind, the first few Chapters aim at understanding the deep and rich scientific context surrounding this thesis work. It is the result of an interdisciplinary approach to research and bridges the gaps between previously unrelated fields.

In the first Chapter, we explore the basis of structural complexity in the living and how cells cope with various metabolic challenges: competing intracellular reactions, toxic intermediates, bottlenecks and futile enzymatic reactions cycles. Nature evolved a number of solutions to insulate metabolic pathways by using intracellular spatial organization. Synthetic biology is taking inspiration from these natural strategies to increase fidelity and titers of synthetic circuits. However, the different approaches implemented before our work were lacking access to higher order architectures and modularity.

An emerging and quickly expanding field specifically addresses the spatial organization of biomolecules, the RNA nanotechnology field. We spend a few paragraphs at the beginning of Chapter II introducing how our understanding of the RNA molecule evolved, from an information only carrier to an incredibly versatile molecule with enzymatic and structural properties rivaling those of proteins which gave ground to the “RNA world hypothesis”. “RNA nanotechnology” harvests the structural modularity of RNA to make RNA-based nanostructures *in vitro*. Its explorations have been at the basis of the idea of assemble non-coding RNAs to make scaffolds *in vivo*.

RNA's properties have also been put to use in Synthetic Biology to control transcription and translation. It is a rapidly growing area of interest and a number of computational tools and experimental protocols have been developed to work with and engineer RNA. We used and were inspired by most of them in our work. We also put a specific emphasis here on aptamers as intermediates enabling interactions with the cellular machinery, binding proteins or small molecules. Overall, this third Chapter brings a deeper understanding of how RNA circuits work *in vivo*, and how they are implemented.

With the toolbox figured out, we then move onto meaningful challenges to solve. Bio-hydrogen is an ongoing area of research. Hydrogen can be produced biologically in a more environmentally friendly way than current production processes such as steam reformation or electrophoresis. Hydrogen is also a much more energetically dense molecule than any other biofuels. Our labs have a history of closely looking at potential ways to produce biohydrogen. Electron transfer pathways had been shown to improve reaction titers which made this biological pathway a great candidate to test our scaffolding platform in a meaningful way.

All of this comes together in Chapter V and Chapter VI where we present the main results of my PhD work. As highlighted earlier, RNA synthetic biology was showing a lot of promises and so did protein-based scaffolding systems to enhance yields and fidelity of sequential metabolic reactions. The missing link came from an understanding of RNA nanotechnology and the potential of the RNA molecule as a structural element *in vivo*. Inspired by this rich scientific heritage, we developed, characterized and tested RNA-based *in vivo* scaffolding systems in our *Science* paper presented here in Chapter V. Chapter IV presents our effort to popularize, standardize and make as seamless as possible the use of RNA as an *in vivo* scaffolding platform. We highlight the vast array of potential applications from synthetic biology to structural biology and develop a simplified and modular protocol.

Chapter VII focuses on methods, with a specific emphasis on going deeper into novel technics we had to develop to work with RNA at the crossroad between biology and nanotechnologies, from electron microscopy to non denaturing RNA pull-down assays. We also further discuss the design of RNA scaffolds and synthesis in order to clarify, if necessary, these points already addressed in the papers.

We conclude the main body of the thesis with a discussion taking us from exploring the different doors opened by this work to the potential of biohydrogen as a biofuel while putting our results in context. In a first part we present our continuing effort towards using RNA scaffolds as tools to study protein interactions *in vivo*. We also introduce our work on a small *E. coli* RNA thought to naturally polymerize *in vivo*. We then explore some of the doors opened by our work and propose follow-up studies, from exploring “dark matter RNA” to directed evolution schemes to evolve *in vivo* polymerizing RNAs. Finally, we touch upon an ongoing project in the Appendix I with a new take on spatially organizing metabolism. We develop a brand new approach to spatial organization by sequestering oxygen sensitive hydrogen evolving pathways in heterocysts, differentiated nitrogen-fixing anoxic cells of *Anabaena PCC7120*. This opens new trails for synthetic biology hitchhikers, as synthetic oxygen sensitive metabolic pathways have been incompatible with oxygen evolving photosynthetic chassis.

A towel, it says, is about the most massively useful thing an interstellar hitchhiker can have. (Adams 1979)



I - Intra-cellular spatial organization & Synthetic scaffolding

Strategies

1. The need for Intracellular Organization

1. Challenges faced by metabolism

Spatial organization of metabolic pathways helps overcome a number of specific challenges faced by metabolism. First, intermediate metabolites can often participate in

competing intracellular reactions. Malonyl-CoA is a classical example and closely looked at due to underlying industrial interests. It is an important precursor for a wide variety of biomolecules including biodiesel, but its intracellular availability is limited as it is closely associated with cell growth. It is used in the biosynthesis of polyketides and flavones but also consumed during fatty acid or phospholipid synthesis (Xu et al. 2011). Another example is ferredoxin, a very common electron carrier that can reduce other molecules through non-specific electrostatic interactions (Agapakis et al. 2010). We further investigate the scaffolding and insulation of a specific ferredoxin-hydrogenase pair in Chapter V.

Growth inhibition from specific toxic metabolic intermediates can also be very problematic. An interesting example is the biodegradation of halogenated compounds. The initial oxidative or hydrolytic activation enzymatic reactions produce intermediates that are highly toxic and reactive that very few micro-organisms can deal with (van Hylckama Vlieg et al. 2000). Another interesting example is the one of mimosine, a free amino-acid that can be found in the Central American crop *Leucaena leucocephala*. Successful utilization of leucaena as a ruminant forage depends on colonization of the rumen by bacteria that degrade dihydroxypyridines a highly toxic intermediates in the metabolism of mimosine (Allison et al. 1990). Therefore this crop does not export very well.

Bottlenecks are fairly common in sequential metabolic reactions in which some transformation steps can be slower than others. For bio-engineering purposes, many of these bottlenecks have been identified and characterized, as in the case of the co-

enzyme Q10 (Cluis et al. 2011), the melvanoate (Pitera et al. 2007) and the L-tyrosine (Lütke-Eversloh & Stephanopoulos 2008) pathways in *E. coli*.

Finally, intermediates are sometimes susceptible to extra cycles of synthesis or degradation. It is the case of the beta oxidation of fatty acids (Marchesini 2003). In the bacterial luciferase pathway, reduced flavin mononucleotide can undergo auto-oxidation leading to decreased bioluminescence (Becvar et al. 1975). We further elaborate on this specific pathway in Appendix I and develop it as a reporter of *in vivo* scaffolding.

2. Concentration gradients and substrate channeling

Lewis Wolpert first had the insight that gradients might provide the information necessary for structuring developing embryos (Wolpert 1969). Changing scale and moving forward in time, it was shown that diffusion gradients are also very important at the intracellular level in bacteria. Examples are plentiful and include the MinCD protein gradient for cell division in *E. coli* (Endres 2011) or phosphorylated CtrA in *C. crescentus* (Endres 2011). But concentration gradients are also widely used at an even smaller scale where substrate channeling improves the rate and fidelity of multi-enzymatic reactions (Miles et al. 1999).

Substrate channeling is the process of transfer of intermediates between interacting enzyme in sequential biosynthetic pathways (Miles et al. 1999). Substrate channeling reduces loss of intermediates by diffusion, prevents competing pathways from using the intermediates and protects labile intermediates from

solvent (Miles et al. 1999). The term was first coined for direct contact in between interacting enzymes, though nature uses its principles in a wide variety of bio-compartments that limit intermediate diffusion.

2. Nature's Solutions

1. Enzyme complexes

Multi-enzyme complexes are one of the most common strategies seen in nature to deal with these metabolic challenges. Very large enzymatic complexes are frequent in Eukaryotes (Narayanaswamy et al. 2009; Noree et al. 2010). The purinosome is an interesting example. Purine synthesis requires ten different chemical reactions performed by six different enzymes. The enzymes cluster in the cytoplasm to form a large complex, the "purinosome" whose assembly is dynamically regulated by changes in purine levels (An et al. 2008). Another interesting example is the one of the MAP-kinase signaling modules. The MAP-kinase modules form multi-enzymatic complexes helped by scaffolding proteins that facilitate the pathway activation in response to specific stimuli and insulate it from irrelevant stimuli (Whitmarsh & Davis 1998).

In prokaryotes, enzymatic complexes are also very common. The pyruvate dehydrogenase complex of gram negative bacteria improves this reaction by more than 50 times through substrate channeling between the three enzymes forming this complex (de Kok et al. 1998). In salmonella, enzymes catalyzing the last

two steps of tryptophan synthesis form a complex to channel the intermediate indole metabolite between the two active sites (Miles 2001). The cellulosome is another interesting multi-enzyme complex of anaerobic bacteria. Insolubility and heterogeneity of cellulose makes it a poorly bio-available substrate. To meet this challenge, anaerobic bacteria organize the eleven necessary enzymes onto a protein scaffold to make a complex called the cellulosome, thus ensuring correct enzymatic ratio and specific spatial alignment (Nordon et al. 2008; Schwarz 2001).

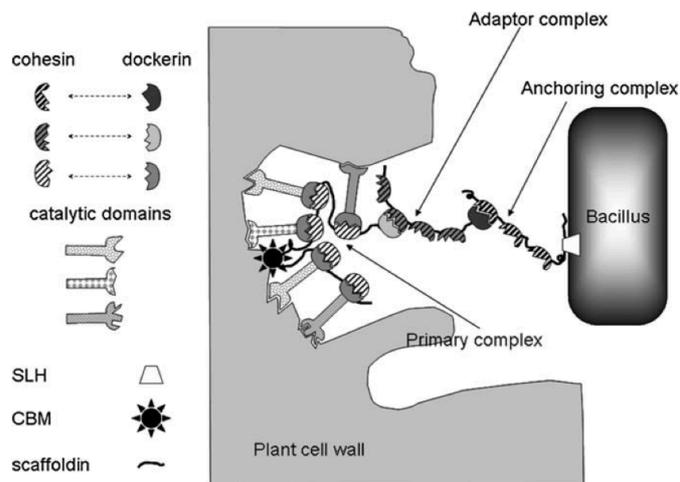


Figure 3: Schematic of the cellulosome architecture showing the colocalization of catalytic domains for efficient plant cell wall degradation. The backbone of the complex is made of flexible scaffoldin molecules. Primary scaffoldins include cohesin modules as well as carbohydrate-binding modules (CBM). Adaptor scaffoldins contain cohesin and dockerin domains. The anchorin scaffoldin

include an S-layer homology domain (SLH). The complex is flexible to conform to plant geometries. (Adapted from Nordon et al., 2011).

Another interesting example, and one of the best studied for its evolutionary flexibility, is the one of the polyketide synthase complexes. Polyketides are a large and diverse group of molecules with interesting pharmaceutical properties (Walsh 2004). The polyketide synthase complex acts as an assembly line to lengthen the growing polyketide chain by passing on the molecule from one domain to another (Tran et al. 2010). The order and domain composition of the enzymatic polyketide synthase complex determines the output molecule (Khosla et al. 2009).

2. Micro-compartments

Eukaryotes have a wide variety of membrane bound organelles to insulate various enzymatic reactions. The peroxisome is an interesting example. It has a central role in the generation but also the scavenging of hydrogen peroxide, thus protecting the cytoplasm from reactive oxygen species (Schrader & Fahimi 2004). Peroxisomes are particularly important in methylotrophic yeast species which perform the toxic first step of methanol metabolism in peroxisomes, namely its oxidation to formaldehyde with concomitant production of hydrogen peroxide (OZIMEK et al. 2005).

Bacteria also have micro-compartments in the form of polyhedral organelles to optimize metabolic processes. Carboxysomes were the first ones to be identified in cyanobacteria (Gantt & Conti 1969). These four nanometer protein shells are filled with carbonic anhydrase RuBisCo which catalyzes the CO₂ fixation step of the Calvin cycle. By creating localized CO₂ gradients, they favour carbon fixation over photorespiration (Savage et al. 2010). *Salmonella enterica* has an evolutionary related microcompartment involved with the degradation of 1,2-Propanediol, the Pdu microcompartment. It protects cytoplasm from propionaldehyde, a toxic intermediate of this metabolic pathway (Yeates et al. 2010; Bobik 2006).

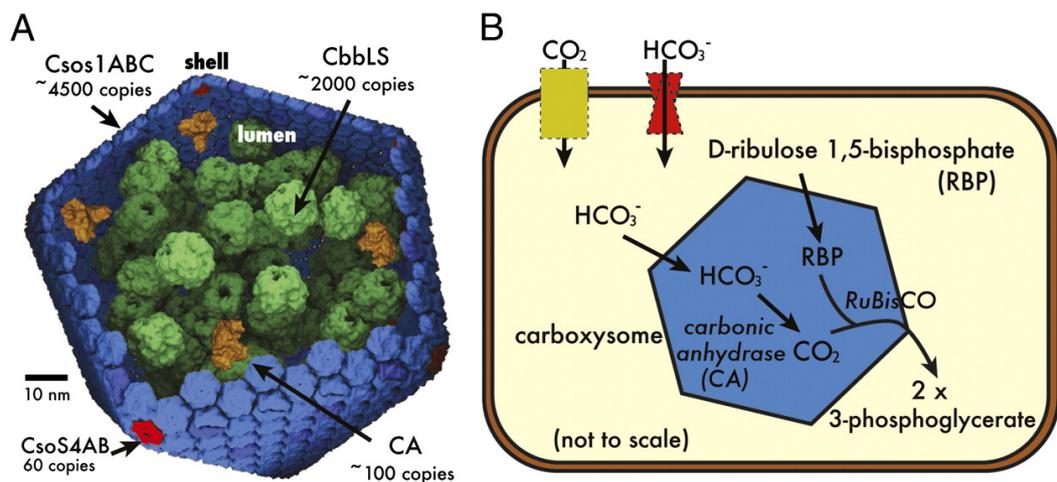


Figure 4: Physical organization of the carboxysome. (A) Partial model of the carboxysome showing RuBisCO (green), carbon anhydrase (orange) and shell

proteins (blue and red). (B) Schematic of the carbon concentration mechanism. (Adapted from Bonacci et al. 2012)

3. Cellular differentiation

A few multicellular prokaryotes found another solution to accommodate incompatible metabolic reactions. Cyanobacteria have evolved multiple specialized cell types, including nitrogen-fixing heterocysts and spore-like akinetes (Kumar et al. 2010).

The development of heterocysts in the filamentous cyanobacteria *Anabaena PCC 7120*, is certainly one of the most striking example of cell differentiation in prokaryote (Kumar et al. 2010). Oxygenic photosynthesis and nitrogen fixation are two essential but incompatible processes in cyanobacteria metabolism because nitrogenase is inactivated by oxygen. Most cyanobacteria separate the two processes in time using their circadian clock. *Anabaena* however fixes nitrogen in specialized differentiated anoxic cells called heterocysts. The strategy here is to compartmentalize metabolism in different cell-types, heterocysts provide fixed nitrogen to vegetative cells which in turn provide sugars (Golden & Yoon 2003) (Kumar et al. 2010).

We further investigate heterocysts in Appendix II developing molecular biology tools to engineer then and use *Anabaena* new chassis for synthetic biology.

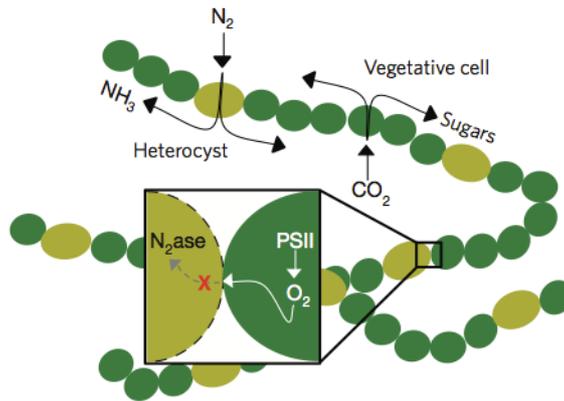


Figure 5: Heterocysts (pale green) in nitrogen starved *Anabaena PCC7120* (adapted from Agapakis et al, 2012)

3. Emerging Synthetic Scaffold approach

The first attempts at creating de novo cellular organization in order to improve specific metabolic fluxes was made through the use of enzyme fusions (Conrado et al. 2008). However, a number of specific disadvantages exist with enzyme fusions. Notably, this strategy is not amenable to pathways containing more than two enzymes and it does not provide the option to balance stoichiometry (Lee et al. 2011). We will thus only cover the recent developments in post-transcriptional assembly methods which are directly relevant to our work.

1. Using and engineering organelles and microcompartments

It is possible to specifically address enzymes in specific compartments in eukaryotes by using short specific targeting sequences (Siddiqui et al. 2012) and

in yeast about 75% of the proteome is specifically localized at the subcellular level (Huh et al. 2003). These targeting sequences are small N-terminal or C-terminal peptide used transiently by the cellular machinery to address proteins to specific localizations (Heijne 1990).

Using this strategy, a plant isoprenoid-producing pathway was isolated from competing reaction by targeting the specific enzymes to the mitochondria in engineered yeast. The idea here was to both isolate the pathway, but also to take advantage of the presence of a pool of farnesyl diphosphate in this organelle, an important intermediate of this pathway Farhi:2011gd}. Other examples include the sequestration in yeast vacuole of a pathway producing methyl halides, a chemical used as agricultural fumigants and a precursor of many chemicals and fuels. This strategy both sequester halogenated intermediates and enable the pathway to use a native pool of substrate in the vacuole, leading to overall higher yields of methyl halides production (Bayer et al. 2009)

A recent paper takes this strategy to the next level. Chloroplasts, alongside mitochondria and hydrogenosomes all evolved from ancient prokaryotes that became endosymbionts (Howe et al. 2008). In an effort to implement photosynthesis in animals, researchers engineered cyanobacteria to enter, maintain and divide in the cytoplasm of non-photosynthetic cells (Agapakis et al. 2011). It is a first interesting step which could lead to engineering a true endosymbiosis by implementing mutual metabolic exchanges.

In bacteria, work is in progress to target specific enzymes into carboxysomes. Carboxysomes have been expressed in *E. coli* and shown to be functional *in vitro* but much work remains to fully understand targeting into these micro-compartments (Bonacci et al. 2012).

2. Using and engineering protein scaffolds

One of the first uses of a protein scaffold in synthetic biology is by Wendell Lim and colleagues and their use of the Ste5 MAP-Kinase scaffold. By tethering a unique set of kinases onto this protein scaffold, they created a synthetic MAP-kinase pathway with artificial input and output properties (Park et al. 2003) (Bashor et al. 2008).

Dueber et al were the first to apply protein scaffolding technologies to increase yields of specific enzymatic reactions. They used three eukaryotic interaction domains (GBD, SH3, and PDZ) to recruit three mevalonate biosynthetic enzymes (atoB, HMGS, and HMGR) C-terminally tagged with peptide ligands specific for these interaction domains. This specific pathway produces a toxic intermediate and suffers from a bottleneck reducing the specific turnover of the whole pathway. Scaffolding resulted in an increase in yield of 77 folds (Dueber et al. 2009).

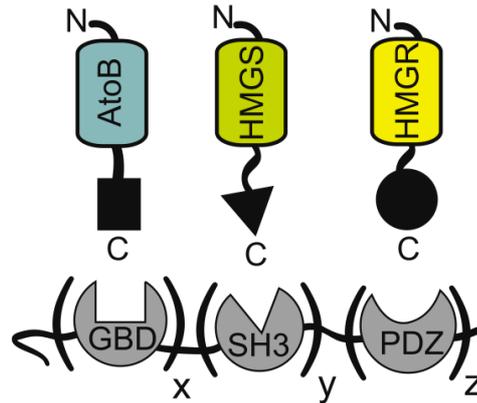


Figure 6: Synthetic protein-based scaffold. Three modular protein-protein interaction domains are fused and recruit three mevalonate biosynthetic enzymes (atoB, HMGS, and HMGR) C-terminally tagged with peptide ligands specific for these interaction domains. (Adapted from Lee et al., 2011)

This scaffolding strategy using the same interaction domains has since been extended to other pathways with variable titer improvements (Lee et al. 2011). For the pathway producing D-glucaric acid, the second transformation step is catalyzed by an enzyme, myo-inositol oxygenase, whose activity is strongly influenced by the concentration of the myo-inositol substrate. Scaffolding proved to be successful in raising its local concentration and five fold yield improvements were observed (Moon et al. 2010). The scaffold was tested in yeast on a pathway producing resveratrol, an antioxidant molecule interesting for its effects on longevity in a number of animal models (Valenzano et al. 2006). The enzymes (4-coumarate:CoA ligase (4CL1) and stilbene synthase (STS)) of resveratrol biosynthesis pathway were scaffolded and a five fold yield improvement was observed (Wang & Yu 2012).

Finally, and very relevant to our work, Agapakis and colleagues tested this scaffolding platform on a redox pathway producing hydrogen. This specific pathway is discussed

in greater details in Chapters IV. It suffers from competing reactions and requires a direct contact between a reduced ferredoxin and a [Fe-Fe] Hydrogenase. The pathway benefited from scaffolding and a four-fold yield improvement was observed (Agapakis et al. 2010).

3. Using and engineering nucleic acids scaffolds

Many fewer attempts at using nucleic acids as scaffolds *in vivo* have been made and our work pioneers this technology. Specific challenges include the bioavailability of DNA *in vivo* and the instability of RNA. Nucleic acid nanotechnologies have been investigating scaffolding *in vitro* as illustrated in Chapter III, but remained to be applied *in vivo*. Specific challenges included the need for an isothermal assembly mechanism, enzyme-scaffold interaction, stability and recombinant expression of large amount of nucleic acids. All of these challenge are addressed in details in Chapters V and VI. Here, I would like to highlight a recent paper, which although does not use nucleic acid assemblies per say, still pushes the field forward with interesting ideas and innovations.

The use of plasmid DNA as a stable and modular scaffold was recently shown in *E. coli* (Conrado et al. 2012). Researchers used DNA-binding zinc fingers to target proteins onto plasmids. A number of pathways were tested for the effect of scaffolding including, resveratrol, 1,2-propanediol and mevalonate pathways and showed increase in yields. This approach is very innovative but suffers from a number of drawbacks including limited maximal number of plasmids per cells, metabolic burden of maintain-

ing the plasmids, and no access to higher-order 2D or 3D architectures. As a very interesting and compelling example of how iGEM is driving synthetic biology it is worth noting that this work was done by the 2010 Slovenian iGEM team (Conrado et al. 2012).

“One can contemplate an RNA world, containing only RNA molecules, that serve to catalyze the synthesis of themselves”

Walter Gilbert (1986)

“In the beginning the Universe was created. This has made a lot of people very angry and has been widely regarded as a bad move.”

Douglas Adams
The Hitchhiker's Guide to the Galaxy



II - From the RNA world hypothesis to RNA nanotechnologies

1. RNA World Hypothesis

1. The primordial RNA world hypothesis

The concept of RNA as a primordial molecule arose in the late 1960s. Crick (1968), Orgel (1968) and Woese (1967) postulated that primordial systems could have consisted of RNA and Proteins. RNA would serve as information support and protein molecules

would provide enzymatic activities (Eigen et al. 1981). At the time, RNA was only thought to be involved in translation and to come in only three “flavours”: tRNA, rRNA and mRNA.

The discovery of RNA catalytic properties in the 1980s (Guerrier-Takada et al. 1983; Kruger et al. 1982) provided a much stronger basis for the plausibility of an RNA-base primordial world. In 1986 the term “RNA world” is finally coined by Walter Gilbert. In a commentary following the discovery of RNA enzymatic properties in *E. coli* by Cech (Zaug & Cech 1986), Gilbert postulated that RNA could not only support of heredity but also activity in primordial systems (Gilbert 1986):

“One can contemplate an RNA world, containing only RNA molecules, that serve to catalyze the synthesis of themselves”(Gilbert 1986)

The primordial RNA world of Gilbert is a world where RNA serves as both genotype and phenotype. In the early 90s, the ability to evolve and select a broad range of RNA with catalytic properties via a new method (SELEX - see Chapter III) further fueled the enthusiasm for the RNA world theory and made it possible to conceive a ribo-organism carrying out complex metabolism (Tuerk & Gold 1990; Ellington & Szostak 1990) (Benner et al. 1989). It is still today a much debated and relevant hypothesis. In the latest views on the RNA world, it is thought that various self-replicating molecular systems probably preceded RNA, but proteins large enough to self-fold and have useful activities came about only after RNA was available. The RNA world probably evolved into a world of ribonucleoproteins, before giving rise to the DNA and RNA world of today. DNA is thought to have taken over the role of data storage due to its

increased stability. On the other side of the spectrum, proteins, through a greater variety of monomers (amino acids), replaced RNA's role in specialized biocatalysis (Cech 2012; Eddy 2001).

2. The modern RNA world

The modern RNA world is not hypothetical. It is the one of the modern biological systems in which RNA comes in a wide range of flavours, with many probably still to discover. In this context, it is useful to use Cech's RNA classification when talking about the wide range of biological activities of RNA and ask what RNA can do by itself versus with proteins or DNA (Cech 2012).

By itself, RNA can first be the sole support of genetic information. RNA viruses have a single or double stranded RNA molecule as genetic material and rely on host cells to replicate (Hiscox 2007). By itself, RNA can also bind with high affinity small molecules switching from one configuration to another. These RNA are called riboswitches and they may tightly control genetic regulation (commonly by initiating/terminating transcription or cleaving transcripts) in a wide range of micro-organisms spanning from gram-negative bacteria to plants (Winkler & Breaker 2003; Breaker 2012). We will further develop on the subject in Chapter III.

RNA often works in conjugation with proteins, as in the case of ribonucleoproteins. The ribosome is an interesting example, as RNA takes on the central roles of the molecule, including decoding mRNA templates, mRNA start-site selection and codon

anti-codon interaction (Moore & Steitz 2011). The eukaryotic spliceosome and the telomerases also work in intimate relation with RNA molecules at their core (Blackburn & K. Collins 2011; Will & Luhrmann 2011).

RNA also can intimately work with DNA to regulate gene expression. In eukaryotes several classes of non-coding RNAs (ncRNA) take on this role through a diversity of mechanisms involved in RNA interference (Volpe & Martienssen 2011). In bacteria, small RNA (sRNA) can act via a variety of principles. They can be true anti-sense RNAs, synthesized from the strand complementary to the mRNA they regulate, they can also act by pairing but have limited complementarity with their targets, and they can regulate proteins by binding to and affecting protein activity (Gottesman & Storz 2011). DsrA is one of these interesting sRNA. It is involved in the RpoS pathway in association with the Hfq protein and has been shown to polymerize *in vitro* (Majdalani et al. 1998; Cayrol et al. 2009). We further detail and engineer DsrA in the discussion Chapter VIII.

Novel synthetic RNAs with engineered function are also a very interesting characteristic of the modern RNA world. As our understanding grows about the structural and functional modularity of RNA as a molecule, new RNA uses have emerged in a variety of scientific fields. In Chapter III we will review the use of engineered RNA molecules in synthetic biology. In the following paragraph, we introduce the young and growing RNA nanotechnology field which uses RNA properties to build functional architectures *in vitro* with nanometer precision.

2. RNA Nanotechnology

1. Structural properties and modularity of RNA nanostructures

Nanotechnology is the manipulation of matter at the molecular and atomic scale. Since biological molecules such as RNA, DNA and proteins have defined structures at the nanoscale, they may be useful building blocks for bottom-up fabrication of nano-devices. About 30 years ago, Ned Seeman pioneered this concept using DNA as a building block and its intrinsic Watson-Crick base pairing rule to control assembly (Seeman 1982). The field is now known as DNA nanotechnology and significantly grew and matured since its inception. Numerous two and three-dimensional DNA architectures have since been rationally constructed and self assembled as shown in Figure 10 (Aldaye et al. 2008; Lin et al. 2009).

The uniqueness of the RNA molecule, which as seen earlier opens the door to a dazzling array of functionality in biological systems, tends to separate the emerging field of RNA nanotechnology from DNA nanotechnology. RNA is composed of four nucleotides: adenine, uridine, cytosine, and guanine. In addition to Watson-Crick base pairing, the molecule also possesses non-canonical base pairing which promotes intra-molecular interactions and higher order folding. RNA-based architectures commonly harness the resulting secondary or

tertiary structure motifs while still borrowing from the DNA nanotechnology tool-box.

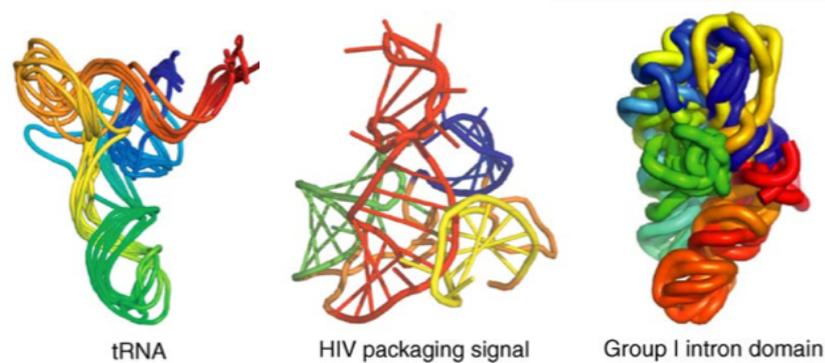


Figure 7: Examples of RNA tertiary structures (adapted from Weeks, 2010)
(Weeks 2010)

Tertiary structure allows for the formation of loop-receptor interactions, central to many biological processes (see Chapter III). It is necessary to point out that the RNA/RNA double helix is very stable. It is actually the most stable among the three helices: RNA/RNA, RNA/ DNA and DNA/DNA. Hydrogen bonding in the RNA molecule is very similar to the DNA molecule. However, the 2'-OH in the RNA ribose favors an A-type helix as opposed to B-type for DNA. The small difference in base stacking in between the two configurations is enough to make a significant difference in helix stability. Consequently, once formed RNA architectures are actually more stable than DNA architectures (Guo 2010; Leontis et al. 2006).

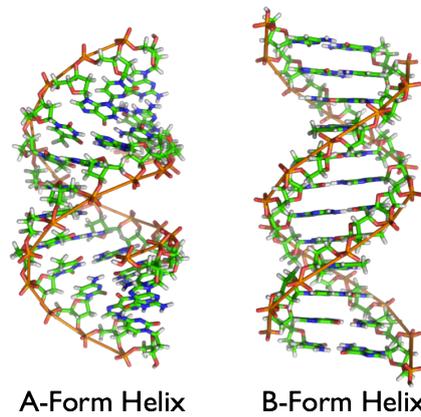


Figure 8: A-form and B-form RNA and DNA helices. The A-form is much more compact and the small difference in base stacking makes it more stable.

Currently, RNAs up to 80 bases are commercially available with a turn-around of a few business days. Most interestingly, high amount of single-stranded RNA can be transcribed *in vitro* but also *in vivo* by the cell's machinery (see Chapter V and VI).

2. Strategies for programmable RNA self assembly

To engineer complex RNA architectures, the use of predictable and addressable self-assembling building blocks is necessary. RNA nanotechnology borrows from both the DNA nanotechnology's and nature's toolboxes to achieve stable two and three dimensional structures.

There are two basic ways for RNA modules to self-assemble: in a templated or non-templated fashion (Guo 2010). Templated assembly requires the intervention of other components to assist in the polymerization of RNAs. This is the case in many of the natural systems we know where RNA modules interact

with cellular components: the ribosome, phi29 pRNA for example (Xiao et al. 2005). Non-templated assembly includes most of the RNA-RNA or RNA-DNA interactions found *in vivo*, as well as DNA nanotechnology's assembly strategies.

The array of RNA architectures built is quickly growing as the field is expanding. There are different tactics for constructing these architectures (Guo 2010). A first idea is to borrow from nature, and one of the earliest artificial RNA structure made was based on pRNA assemblies (Shu et al. 2004). We build upon this strategy in Chapter VIII by engineering DsrA RNAs.

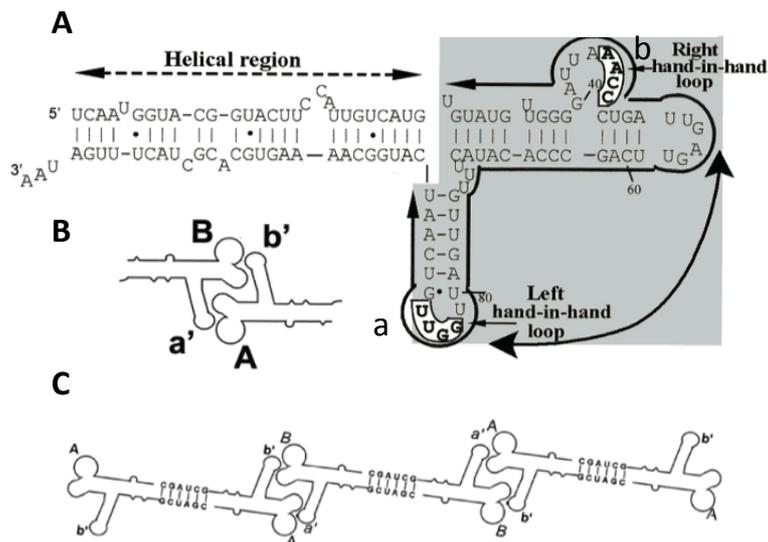


Figure 9: pRNA-based assemblies. (A) Sequence and structure of wild type phi29 pRNA. It is possible to modify the highlighted sequence (a) and (b) to form pRNA dimers (B) by making them complementary (A-a' and B-b'). Based on this principle, it is possible to synthesize RNA molecules made of two engi-

neered pRNA that are going to self assemble and polymerize (C). (Adapted from Guo et al., 2004)

A second strategy is to borrow from DNA nanotechnology and thus use non-templated assembly. Here assembly does not rely on hydrogen-bound interactions between self-folded molecules, but rather on Watson-Crick base pairing between the RNA strands. An example is the synthesis of cubic RNA scaffolds where directly after *in vitro* transcription, RNA strands were programmed to polymerize to form the desired architecture (Afonin et al. 2010) (Famulok & Ackermann 2010). Our *in vivo* assembly strategy demonstrated in Chapter V is based on this idea. A third strategy is called RNA tectonics and relies on designing artificial specific modules with defined reach and stack that self assemble using hairpin interactions into higher order structures such as RNA filaments or jigsaw puzzles (Jaeger & Chworos 2006).

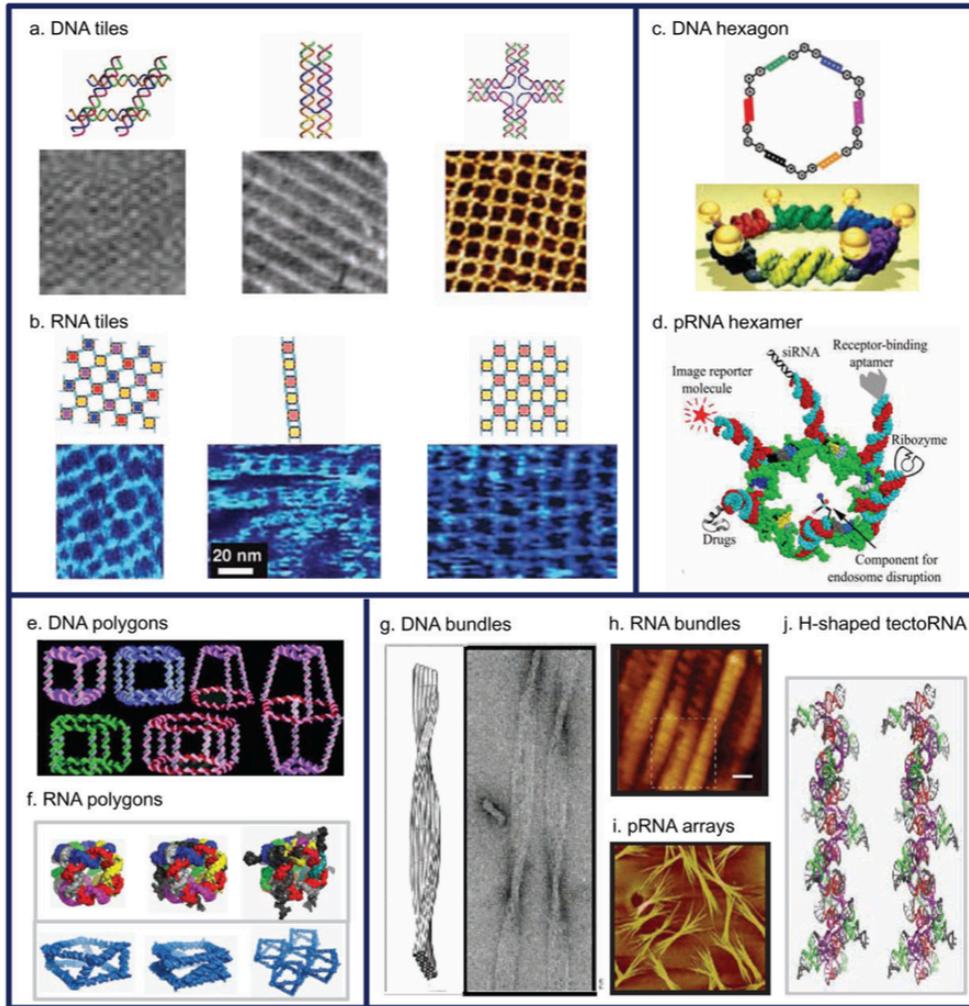


Figure 10: Diversity of RNA and DNA nanostructures. Images of DNA tiles (a)_{1,2} and RNA tiles via tectosquares (b)₁₃; illustration of hexameric DNA gold nanoparticle (c)_{1,2}, pRNA hexameric ring (d)_{50,87}; DNA 3D polygons (e)₂; and RNA cubic scaffolds (f) _{30,77}; images of DNA bundles (g)₇₀, RNA bundles (h)₂₃, pRNA arrays (i)₁₂; and 3D model of H-shaped tectoRNA (j)₂₁. All images are taken by AFM except (g) and (j), as well as the first two images of (a), which are TEM images. All images were adapted from the individual references with permission.

RNA is the candidate of choice for *in vivo* synthetic architectures due to its intrinsic single strandedness and the fact that it can easily be expressed in high concentration intracellularly using synthetic biology tools. However, simply designing a platform is not our end goal, as to be useful for synthetic biology architectures need to be useable as a scaffold to template metabolic pathways.

3. Nanostructures used as scaffolds

Nanostructures are very good templates for the organization of small molecules with nanometer precision. Scaffolding has been an active field of experimentation in DNA nanotechnology as nanoparticles, proteins and peptides were successfully organized onto DNA architectures (Williams et al. 2007) (Kuzyk et al. 2009; Tan et al. 2011).

Fewer ventured into scaffolding full metabolic pathways and studying yield improvements (Teller & Willner 2010). The first demonstration came from the scaffolding of the bacterial luciferase redox pathway on a double-stranded DNA scaffold with an observed three fold increase in activity (Niemeyer et al. 2002). In Appendix I, we further study and develop the pathway as a reporter system for *in vivo* scaffolding applications. Two dimensional systems were also successfully used, and example include the assembly of a glucose oxidase (GOx) and horseradish peroxidase (HRP) on 2D hexagonal DNA strips where a 10-fold enhancement was observed (Wilner et al. 2009).

Finally, recent publications have focused on understanding the effect of relative protein distance on overall yields. Examples include varying the distance in be-

tween the BMR reductase domain and the BMP porphyrin domain of the cytochrome P₄₅₀ BM₃ complex (Erkelenz et al. 2011) or organizing a GOx/HRP cascade on DNA origami tiles with precisely controlled spatial positions (Fu et al. 2012). Relative distance in between the enzymatic catalytic centers do indeed regulate the efficiency of the enzymatic reactions. These recent developments remained *in vitro* examples but, together with our work, they demonstrate that nucleic acid nanotechnology can bring a lot to metabolic engineering and structural biology.



III - Doing Synthetic Biology with RNA

1. Tools for engineering RNA devices

1. Harvesting Nature's components – the world of RNA Aptamers

Our understanding of RNA modularity has increased ever since discovering that RNA is much more than just an information carrier in the early 1980s.

(Guerrier-Takada et al. 1983; Kruger et al. 1982). The Szostak lab is credited with coining the term “Aptamer”, from the latin “Aptus” meaning “to fit” (Tsuji et al. 2009; Bunka & Stockley 2006). RNA’s ability to fold into complex three-dimensional shapes rivaling those of proteins enables it to bind specific target molecules with high affinity.

Naturally occurring RNA aptamers can be found in riboswitches - specific RNA sequences involved in regulating transcription or translation (Winkler & Breaker 2003). These riboswitches can fold into two mutually exclusive structures based on ligand binding or unbinding onto the aptamer, thus dictating the activity of the RNA module (Winkler & Breaker 2003).

The most studied natural aptamer is found in the MS2 virus genomic RNA, an icosahedral positive-sense single-stranded RNA virus that infects *E. coli*. The MS2 aptamer controls the temporal translation of the MS2 RNA. Upon binding to the MS2 coat protein, a conformational switch of the mRNA template enables the translation of the replicase gene (Peabody 1993). The MS2 aptamer has been a model for RNA-protein recognition since the early 90s when Uhlenbeck and colleagues defined the 19 bases of the minimal RNA consensus binding site (Stockley et al. 1995; Witherell et al. 1991). This translational operator has been characterized biochemically, by NMR and X-ray structures of the MS2 protein-RNA complex are now available (Parrott et al. 2000).

Such a well characterized aptamer has been an obvious choice to use as a tool for molecular biology, specifically to follow the spatio-temporal distribution of

mRNA. mRNA molecules can be tagged on their 3' end with large repeats of the MS2 aptamer, typically 96, allowing fluorescent proteins fused to the MS2 protein to colocalize and reveal the spatial localization of the mRNA molecules (Golding & Cox 2009; Golding et al. 2005). Interestingly, the coat protein-viral replicase RNA is apparently conserved but the specifics of the coat proteins and RNA structures they recognize have diverged during evolution. In these regards, the PP7 bacteriophage is very interesting as the divergence went far enough so that the PP7 RNA aptamer is unrecognizable to other phages coat proteins (Lim & Peabody 2002). The complex in between the aptamer and the PP7 coat protein was recently crystalized (Chao et al. 2007), and it was chosen in our work for its tight binding and orthogonality to MS2.

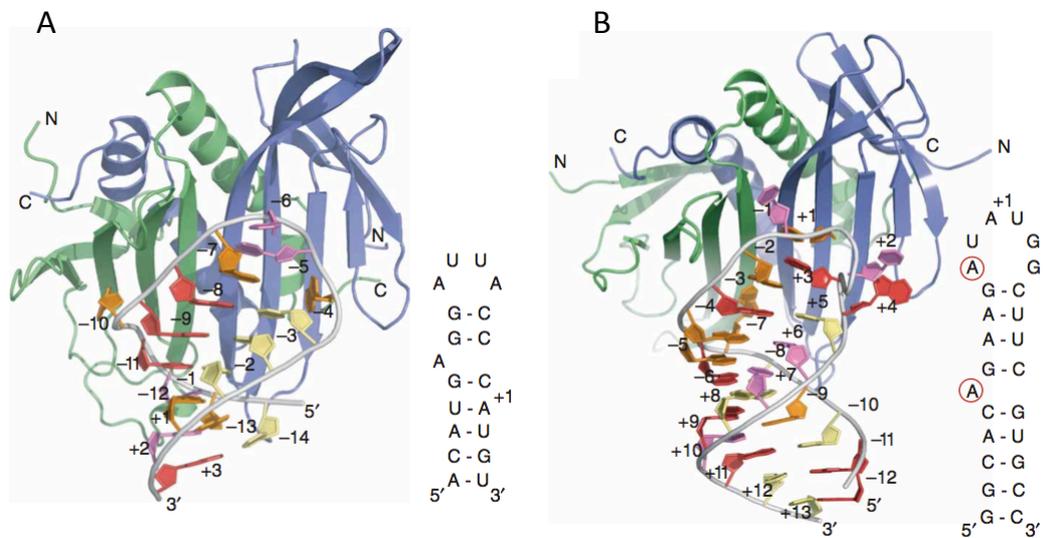


Figure 11: MS2 and PP7 protein crystal structure. Crystal structure of the MS2 coat protein - RNA complex (A) and the PP7 coat protein - RNA complex (B). The adenine residues position and orientation of the PP7 aptamer

(circled in red) differ considerably from those of the MS2 coat protein providing the orthogonality. (adapted from Jeffrey et al., 2008)

2. Evolving new RNA components – SELEX

Aptamers can be selected artificially to bind a wide array of molecule by iterative round of *in vitro* selection. The process called SELEX (Systematic Evolution of Ligands by Exponential Enrichment) was pioneered in 1990 by the Gold and Szostak labs (Tuerk & Gold 1990; Ellington & Szostak 1990) and significantly contributed to popularizing the RNA world concept (Cech 2012).

During SELEX, randomized pool of RNA (or ssDNA) molecules are incubated with the molecular target of interest. During each iterative round, binding species are separated from non-binders and amplified to generate a new pool. This cycle is repeated until a starting pool of generally upwards 10^{15} RNA is reduced to only a few sequences with the highest specificity to the desired target (Tuerk & Gold 1990; Ellington & Szostak 1990). SELEX opens the door to developing orthogonal RNA aptamer libraries, specific to the enzymes to be scaffolded.

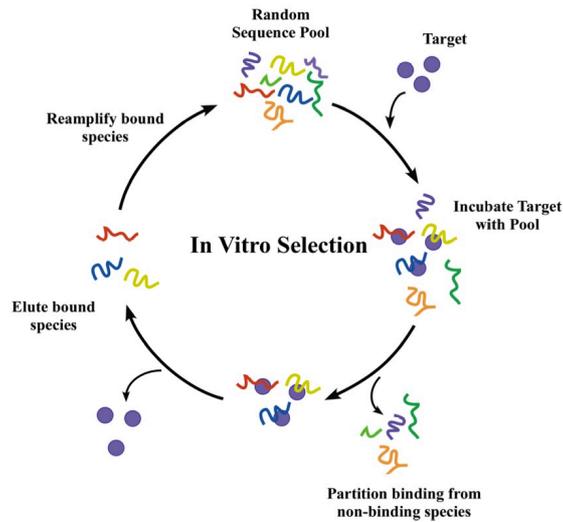


Figure 12: Systematic Evolution of Ligands by Exponential Enrichment (SELEX). (adapted from Chee Yan et al, 2008)

3. Computational approach

Many computational tools are available to assist researchers in designing and implementing functional RNA-based devices. To control and fine tune translation, a ribosome binding site (RBS) design software was recently released (Salis et al. 2009). The difference in free energy in between unbound 30S sub-unit and mRNA versus the formed complex was used to determine translation initiation rates of RBSs. This method was then used to fine tune and control the expression of a fluorescent reporter.

There are many RNA folding software packages available that rely on a free energy minimization algorithm to predict RNA structure. Mfold is currently the reference and has been cited over 3000 times (Zuker 2003). It has been used for example to design RNA-based devices to control gene expression in a

temperature dependent manner (Chowdhury et al. 2006). We extensively used RNA designer that has the specificity to let the user define a secondary structure and computes a primary sequence respecting the folding scheme (Andronescu et al. 2003). Nupack is an interesting new addition as it let users study interaction strength and specificity between several RNA strands (Zadeh et al. 2010).

2. RNA Control Devices in synthetic biology

We extensively define synthetic biology in the introduction. Many higher order functions such as oscillators or logic circuits do not have constraints on the specific building blocks used, but traditionally rely on native biological regulators such as transcription factors (Alon 2003; Drubin et al. 2007). Increasingly however, RNA is emerging as an interesting alternative for tunable components to be used in genetic circuits to reprogram cellular behaviour (Benenson 2012). RNA synthetic biology is following the footsteps of RNA nanotechnology as it discovers the modularity of RNA-based devices. We will cover here the recent advancement of RNA synthetic biology in prokaryotes.

1. Riboregulator and Riboswitch-based devices

RNA sensors can be engineered to detect temperature changes based on temperature-sensitive secondary structure rearrangement. For example, in many gram-negative bacteria an RNA hairpin is present in the 5' UTR modulat-

ing RBS accessibility and thus tuning gene expression (Chowdhury et al. 2006). Based on the same principle, RNA thermometers can be engineered (Neupert & Bock 2009) to generate temperature-sensitive RNA control device for synthetic biology (Kortmann et al. 2011).

RNA itself by Watson-Crick base pairing can act as a input for RNA-based control devices. Engineered riboregulators have been elegantly used to control biological networks. In one of the earliest examples, a small sequence complementary to the RBS was designed upstream from the RBS. Upon transcription, a stem-loop forms, effectively blocking the RBS and transcription. By expressing a small RNA complementary to the blocking sequence, effective translation is possible (Isaacs et al. 2004). More complex RNA-responsive designs have since then been successfully used in *E. coli* (Callura et al. 2010; Lucks et al. 2011).

Riboswitches have demonstrated to be great synthetic biology tools to respond to small molecule or protein inputs. The vast majority of designs in prokaryotes have been made to control translation. For example, a theophylline aptamer was linked to the RBS through a linker sequence capable of structural rearrangement modulating RBS accessibility (Desai & Gallivan 2004). The theophylline aptamer can also be linked to a hammerhead ribozyme which sequesters a RBS. Upon ligand-induced cleavage, the RBS is released and translation can happen (Ogawa & Maeda 2008).

2. Applications in synthetic biology

RNA-based devices in synthetic biology have great potential (Khalil & J. J. Collins 2010) with many applications in agriculture, bio-manufacturing and health (Isaacs et al. 2006; Saito & Inoue 2009; Benenson 2012). An interesting example of an environmental application was the use of an atrazine aptamer controlling cell motility (Sinha et al. 2010). An atrazine responsive RBS-based device controls the expression of the *cheZ* gene based on atrazine availability in the environment. The engineered *E. coli* strain also bear an atrazine-degrading pathway thus effectively degrading the herbicide. In genetic network geared towards bio-manufacturing, like our project, many opportunities remain. Recently, an RNA control device was used to detect metabolite accumulation in yeast where the xanthine responsive ribozyme was linked to a fluorescent reporter gene. Upon accumulation of the metabolite, GFP expression level increases (Win & Smolke 2008).

Many more applications are geared towards health and medicine, mostly in the field of eukaryotic synthetic biology (Khalil & J. J. Collins 2010). Here, the interest of RNA-based devices lie in the fact that RNA-only systems are compatible with size-limitations of most delivery solutions (viral vectors for example), and do not need additional transgenic proteins for their actuation (for a comprehensive review see (Liang et al. 2011)).

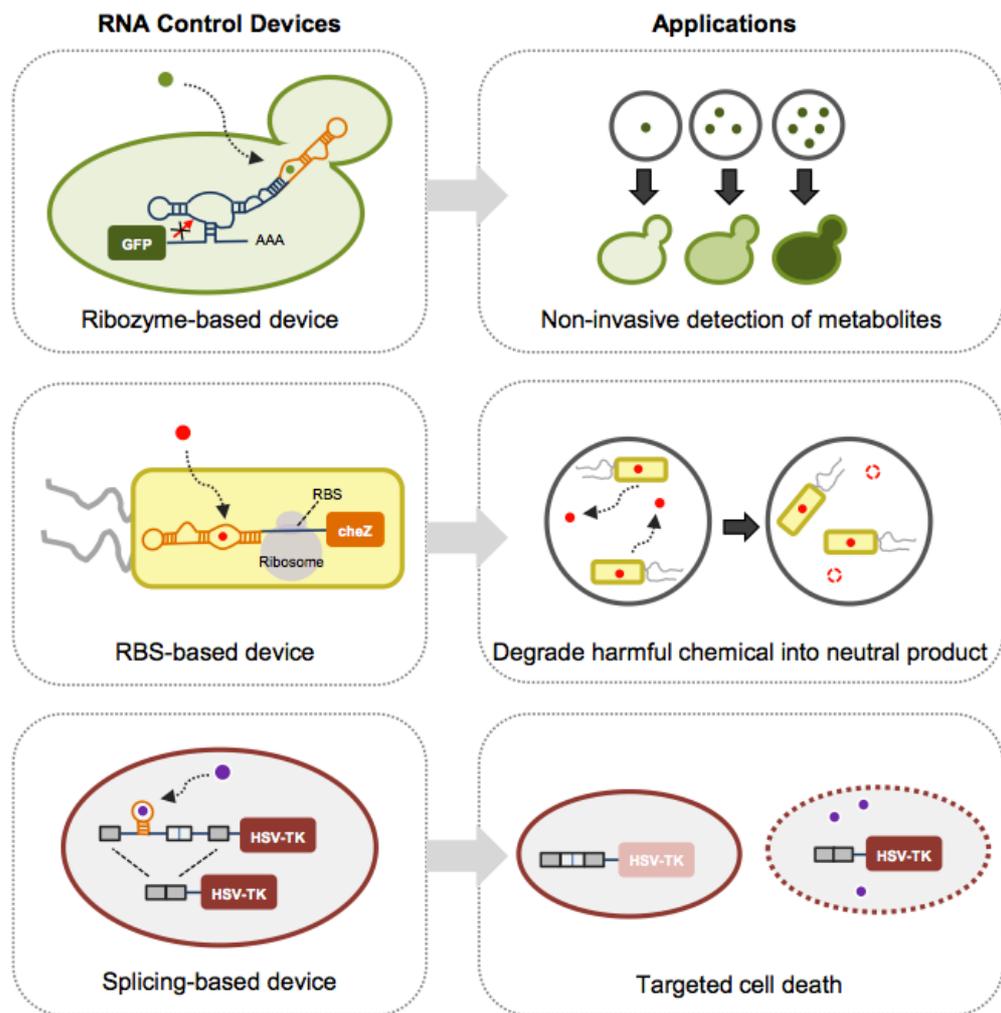


Figure 13: Diversity of applications of RNA-based controllers in synthetic biology (adapted from Liang et al, 2012).



IV - The challenges of biologically producing hydrogen

1. Bio-hydrogen, an attractive biofuel

1. Biofuel economy

Replacing fossil fuels with biofuels - fuels whose energy is derived from biological carbon fixation - could reduce many of the problems associated with fossil fuel production and use. These include the emission of greenhouse gas pollutant emissions, the exhaustion of non-renewable fossil resources and the dependence on unstable foreign suppliers.

The first generation of biofuels - made from the fermentation of carbohydrates usually derived from sugar or starch crops such as sugarcane, beet or corn - is receiving a lot of interest from politicians, scientists and the general public; however, they also have many drawbacks. These include increased food costs as they compete with human and animal nutrition, and other problems associated with intensive farming practices such as air and ground water pollution, or land and water resource requirement. Moreover, first generation biofuels tend to require heavy subsidies from governments and other type of market interventions to be economically competitive with fossil fuels. Yet, according to the WorldWatch Institute, global first generation biofuel production has reached an astonishing 105 billion liters in 2010 up by 17% from 2009 in response to rising oil prices. US and Brazil are the world leaders in ethanol and Europe is the largest producer of Biodiesel. The first generation of biofuel provided 2.7% of the world's transport fuel in 2010 (Ren21 Report, 2011).

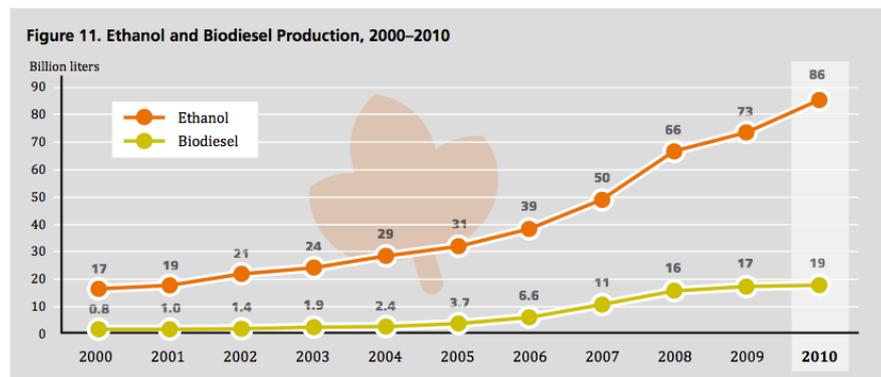


Figure 14: Ethanol and Biodiesel Production from 2000 to 2010. (from Renewables 2011 Global Status Report).

There is a urgent need for advanced biofuels - biofuels produced from sustainable feedstock or derived from photosynthetically fixing carbon dioxide - to reduce the pressure on world food supply and biodiversity associated with the first generation. Although not produced commercially at large scale yet, the diversity of players in the advanced biofuel industry has significantly increased. Both the aviation industry and the traditional oil industry have shown interest making advanced biofuels a reality (Ademe Report 2010; Ren21 Report 2011).

2. Hydrogen advantages

Over the past few years, hydrogen has emerged as a promising advanced biofuel as pressure mounts to replace fossil fuels. Its primary advantage is that hydrogen is energetically denser than all other combustible fuels (hydrogen 141.9 kJ/g vs. ethanol 19.7 kJ/g or diesel 45.8 kJ/g). Hydrogen also combusts into pure water and the energy efficiency of a fuel cell is generally between 40-60% (up to 85% if waste heat is captured for use) - far more efficient than combustion engines limited by the Carnot cycle below 40% and practically averaging around 18-20%.

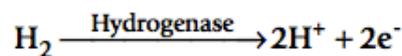
Currently, hydrogen is produced commercially primarily by coal gasification, steam reformation of natural gas, and water electrolysis. These methods are not only too costly for hydrogen to compete with gasoline as a fuel, but they are also energy intensive and therefore not environmentally friendly (Mormilan & Veziroglu 2002).

Hydrogen production thus presents an attractive alternative to other fuels, especially biological hydrogen which can be produced using renewable biomass or sunlight as its primary energy source (Waks & Silver 2009; Savage et al. 2008).

2. Strategies for biologically producing hydrogen

1. Light-driven hydrogen production

This strategy is the primary focus of one of my thesis projects and is presented here in Appendix II. This approach uses the power of photosynthesis to split water. This process, photobiolysis, is attractive as it only requires a source of sunlight, water and carbon dioxide (Dutta et al. 2005). Light is captured via the bacterial photosynthetic system and part of the reducing power generated is re-routed towards a ferredoxin. Ferredoxin then catalyses hydrogen evolution via a hydrogenase and the reduction of free protons:



Despite several decades of research, a number of serious drawbacks still need to be addressed, engineering related and biology related (Hallenbeck et al. 2012). First, the mixture of gases evolved, mainly oxygen and hydrogen, is explosive and would require safety precautions to be implemented if produced at high concentration. The second engineering related challenge has to do with

the engineering of photobioreactors. It would take a breakthrough in material sciences to build low-cost photobioreactors that are both transparent and hydrogen impermeable (Dasgupta et al. 2010).

The biggest challenge is actually biology-related and has to do once more with the mixture of gases evolved. Unfortunately, the most efficient hydrogenases, the [Fe-Fe] hydrogenases, are highly oxygen sensitive and irreversibly inactivated by short term exposure to low oxygen concentrations. Much research is being done to address this fundamental obstacle, albeit without much success so far. This includes protein engineering (Leroux et al. 2010), protein directed evolution (Stapleton & Swartz 2010; Bingham et al. 2012), and using alternative hydrogenases with lower activities (Melis et al. 1999). In Appendix II, we take on a new approach consisting of isolating the hydrogen evolving pathway from the oxygen evolving one by controlling their respective intra-cellular spatial localization in *Anabaena PCC7120*.

2. Fermentation-based hydrogen production

With the high hydrogenase oxygen sensitivity and the engineering challenges associated with light driven biohydrogen production, all eyes are on alternative ways to biologically evolve hydrogen. Fermentation-based hydrogen production is an attractive solution as anaerobic conditions preserve hydrogenase integrity and fermentation is fairly well mastered at the industrial scale. This is the strategy used in association with scaffolding in the central part of my thesis, presented here in Chapter V.

There are basically two ways to make hydrogen through fermentative processes. Pyruvate is either reduced to acetyl-CoA and formate through the Pyruvate-Formate Lyase pathway (PFL) which yields a maximum of 2 moles of hydrogen per mole of glucose, or less commonly to acetyl-CoA and a reduced ferredoxin through the Pyruvate-Ferredoxin Oxido-Reductase pathway (PFOR). The former can yield a maximum of 4 mol of hydrogen per mole of glucose.

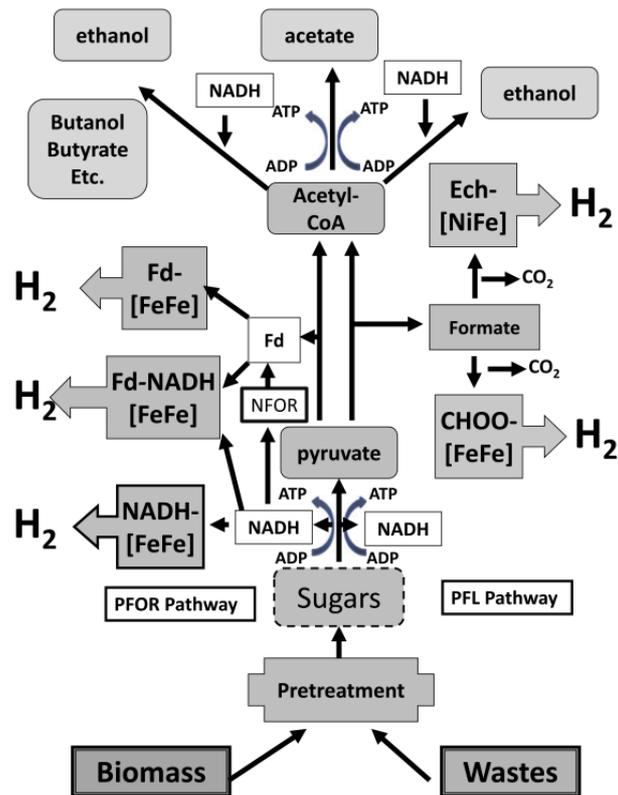


Figure 15: Hydrogen producing fermentation pathways (from Hallenbeck et al., 2012)

Other challenges arise with fermentation-based hydrogen production. Only one third of the substrate, be it wastes, sugar or biomass, can theoretically be converted to hydrogen. The remaining two thirds are used to maintain the organism redox balance and ATP pool, hence the interest in optimizing pathway flux, minimizing electron transfer loss and protein-protein interactions (Hal- lenbeck et al. 2012).

3. Designing a synthetic electron transfer pathway

1. Spatial organization of metabolic electron transfers

Electron transfers are central to many core biological processes, from reducing inorganic chemicals into biologically active biomolecules to harvesting solar energy and breaking down organic compounds (Agapakis & Silver 2010).

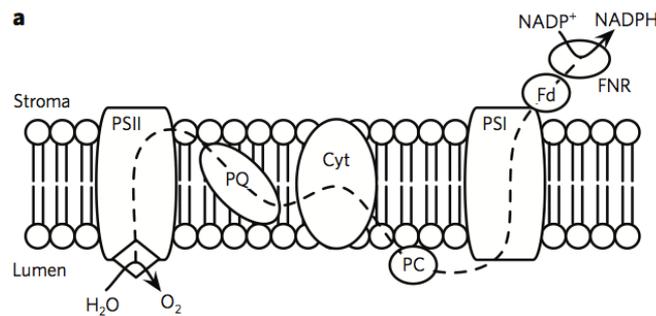


Figure 16: Chloroplast thylakoid membrane and associated electron transfer through the photosynthetic machinery. (From Agapakis, Boyle and Silver, 2012)

Interestingly, physical interactions between electron transfer components is essential for efficient electron transfer. Electrons have to quantum-mechanically transfer in between the enzymes' metal clusters and this process requires both correct orientation and proximity between interacting enzymes, optimally 14Å (Page et al. 1999). In many cases, and especially in thylakoids, cytochromes, quinones and other redox enzymes are membrane-bound to reduce the search space to finding the correct partner to only two dimensions (Dekker & Boekema 2005; Kirchhoff et al. 2008).

But this also extends to cytoplasmic enzymes. Connecting large clusters of redox enzymes has been shown to improve electron transfer both *in vitro* and *in vivo*. Fusion between the mammalian cytochrome P450 to yeast (Shiota et al. 2000) or bacterial (Gilardi et al. 2001) reductases or fusions between ferredoxins and [Fe-Fe]-hydrogenase enzymes have been shown to improve pathway flux (Aliverti & Zanetti 1997). In a ground-breaking and inspiring work in the Silver lab, ferredoxins and [Fe-Fe]-hydrogenase were scaffolded to small synthetic protein scaffolds resulting in a four fold increase in hydrogen yields (Agapakis et al. 2010). Internal electron transfer rate appears to correlate with interface recognition in between interacting enzymes (Aigrain et al. 2011) and electron channeling efficiency to the hydrogenase can be fine tuned by switching ferredoxin types (Agapakis et al. 2010).

2. Synthetic bio-hydrogen pathways

E. coli only has the PFL pathway and most research efforts have been focused on improving the energy flux through this pathway, and yields close to the theoretical maximum have been reached (Hallenbeck & Ghosh 2009). An interesting option would be to use *Clostridium sp* which has the PFOR pathway, but the molecular biology tools available to work with this microorganism are still currently limited. We took the synthetic biology approach and used *E. coli* as a chassis to express an improved PFOR pathway. This pathway was optimized by Christina Agapakis and Danny Ducat in the Silver lab (Agapakis et al. 2010).

The PFOR pathway uses ferredoxin-dependent [FeFe]-hydrogenases. [FeFe]-hydrogenases thermodynamically favor hydrogen production, have fairly high hydrogen production activity and a relatively simple maturation pathway, making them excellent candidates for recombinant expression in *E. coli*. Heterologous expression of [FeFe]-hydrogenases in *E. coli* is sufficient for small measurable hydrogen production. However, co-expression of PFOR, ferredoxin, and [FeFe]-hydrogenase further enhances the production of hydrogen by coupling the breakdown of glucose with the establishment of a reduced pool of ferredoxin.

Ferredoxin-dependent [FeFe]-hydrogenases are a valuable tool for synthetic biology as most redox enzymes in *E. coli* have a potential of -320mv and interact with NADH. [FeFe]-hydrogenases reducing potential of about -420mv opens

the door to using many plants or anaerobic bacteria derived redox enzymes. Hence, the synthetic pathway was further tuned by fine picking the best ferredoxin to reduce [FeFe]-hydrogenase in *E. coli*. Typically, plants have a wide variety of fairly conserved ferredoxins which evolved accordingly to the different requirements of the many redox pathways. For the synthetic electron transfer pathway in Chapter V, we used a spinach leaf ferredoxin (Agapakis et al. 2010).



RESULTS



V - Organization of intracellular pathways with rationally designed RNA Assemblies

This Chapter is at the core of my PhD work and was published in Science (see Appendix II). At the beginning of my PhD, it was clear that scaffolding in synthetic biology had a lot of potential and was useful to increase titer and fidelity of sequential metabolic reactions. However protein scaffolding was clearly missing the characteristic modularity and engineer-ability of other synthetic biology tools. Moreover, protein scaffolding was limited to “discrete” scaffolds, scaffolds where only a handful of enzymes could at most be brought together - far away from a dreamt “synthetic organelle”.

Around the same time, two other fields were quickly gaining momentum. RNA synthetic biology was quickly expanding, with notably many ground-breaking contributions from the Smolke and Collins labs and RNA nanotechnology labs such as the Guo and Jaegger labs were pushing the boundaries of possible RNA based *in vitro* assemblies inspired by the DNA nanotechnology field.

Our contribution results from this fruitful melting pot exposed earlier. Not only did we increase bio-hydrogen production, this paper also demonstrates for the first time the use of RNA scaffolds in Synthetic Biology and that RNA based assemblies can be done *in vivo*. The edited paper is also provided in Appendix II for easier read

1. Introduction

In cells, multi-enzymatic pathways are often physically and spatially organized onto scaffolds, clusters or into micro-compartments (Burack & Shaw 2010). Spatial organization helps substrates flow between interacting proteins, limits cross-talk between signaling pathways, and increases yields of sequential metabolic reactions (Burack & Shaw 2010; Savage et al. 2010). The ability to spatially organize protein complexes and biological pathways presents a strategy to engineer cells (Dueber et al. 2009; Park et al. 2003).

The spatial organization of biomolecules has been the focus of DNA nanotechnology (Rothemund 2006; Seeman 2010; Lin et al. 2009; Winfree et al. 1998). This approach utilizes DNA's base-pairing to generate one-, two- and three-dimensional assemblies *in vitro*. However, DNA structures have largely remained limited to *in vitro* applica-

tions (Lin et al. 2008). RNA provides a compatible material for *in vivo* nucleic acid based construction (Guo 2010). It can be produced via the transcription machinery, and forms stable interactions. RNA has been used to build higher-order assemblies *in vitro* (Jaeger & Chworos 2006; Cayrol et al. 2009) and can potentially be used *in vivo* to engineer the intracellular environment.

Here, we engineered a new class of synthetic RNA modules that assemble *in vivo* into functional discrete, one-, and two-dimensional scaffolds. These materials were used to control the spatial organization of bound proteins.

2. Engineering RNA scaffolds

We developed an approach for the *in vivo* isothermal assembly of extended RNA scaffolds by constructing sequence-symmetric RNA building blocks (Figure 17, C and D) inspired by 2D DNA analogs (Liu et al. 2005; Liu et al. 2006). These RNA strands possess dimerization domains (DDs) and polymerization domains (PDs). To prevent the formation of ill-defined networks, it was necessary to disfavor the collapse of the palindromic regions (Yin et al. 2008) and control assembly order by insuring tile formation before polymerization. We achieved this by designing PDs that fold intramolecularly into kinetically protected hairpin structures (Figure 17D, step i). The stem of these hairpins is an overlapping shared domain with the DD that discourages collapse (Figure 17D, red segments), allowing the DD to activate the PD upon self-binding (Figure 17D, step iii). We further destabilized the collapsed state by incorporating wobble pairs and mismatches (Figures 29 to 31, Chapter VII).

The 1D RNA assembly D_1 was derived from a single RNA d_1 with PP7 and MS2 binding domains (Figure 17E). d_1 assembled into d_1-1 (step i), which self-assembled into d_1-2 (step iii). The torsion in d_1-2 induced folding into an RNA nanotube capable of growing into the 1D scaffold D_1 (step iv). The 2D RNA assembly D_2 was formed from d_2' and d_2'' , each carrying a distinct PP7 and MS2 aptamer (Figure 17F). The dormant tile d_2' spontaneously generated the pro-tile d_2-1 (step i), which interacted with d_2'' to generate tile d_2-2 (step ii). d_2-2 then self-assembled into the 2D RNA scaffold D_2 with PP7 and MS2 binding domains (step iv).

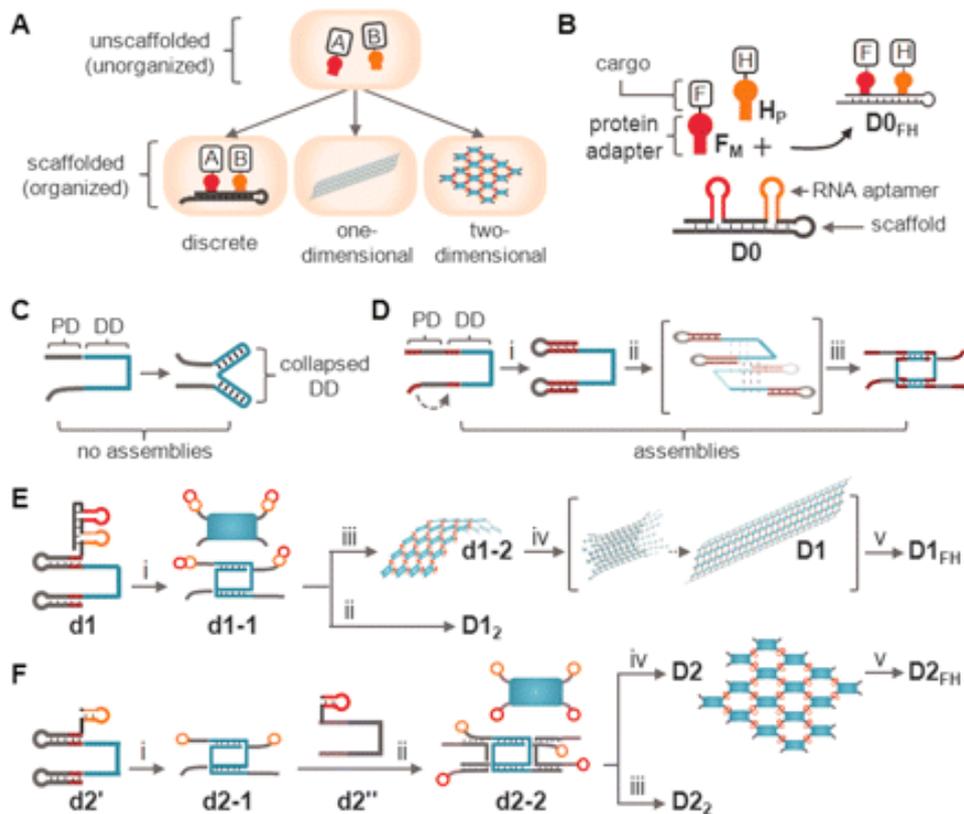


Figure 17: Design of RNA modules to organize proteins. (A) Proteins A and B scaffolded onto discrete, 1D, and 2D RNA assemblies. (B) D_0 is a RNA strand that folds

into a duplex with PP7 and MS2 sites. Ferredoxin/MS2 (FM) and hydrogenase/PP7 (HP) bind Do to generate DoFH. (C and D) RNA with DDs and PDs initiates the formation of extended assemblies. Capping the palindromic sequences in DDs with PDs prevents its collapse (i) and allows for self-assembly (ii) into functioning tiles (iii). (E) D₁ is constructed from a RNA strand d₁ bearing PP7 and MS2, and it assembles into tile d₁₋₁ (i). d₁₋₁ assembles into a ribbon D₁₂ (ii) or into a nanotube d₁₋₂ (iii) that grows into D₁ (iv). D₁ organizes FM and HP into D₁FH (v). (F) D₂ is constructed from d_{2'} and d_{2''} bearing PP7 and MS2, respectively. d_{2'} assembles into the pro-tile d₂₋₁ (i) and interacts with d_{2''} to generate d₂₋₂ (ii). d₂₋₂ self-assembles into a nanotube D₂₂ (iii) or the 2D D₂ (iv). D₂ organizes FM and HP into D₂FH (v).

3. *In vivo* and *in vitro* characterization of RNA scaffolds

We used atomic force microscopy (AFM) to characterize *in vitro* transcribed RNA modules d₁ and d_{2'}/d_{2''}. d₁ formed 1D RNA fibers (D₁), whereas d_{2'}/d_{2''} assembled into 2D extended RNA fibers (D₂) (Figure 18A). The width of D₁ (~5 nm, a few tiles wide) is smaller than that of its DNA analog (Liu et al. 2006) and might also correspond to 1D ribbons (D₁₂) constructed from a continuous line of single tiles (Figure 17E, step ii). Given that D₂ preferentially grows in a single direction when compared with its DNA analog (Liu et al. 2005), it might also correspond to RNA nanotubes (D₂₂) that are relatively wider than D₁ (Figure 17D, step iii). To confirm the validity of our assemblies, we used analogs of d₁ and d_{2'} with a poly-T stretch in place of the DD

incapable of assembling; drT and dz'T did not generate extended assemblies (Figure 18B)

For *in vivo* characterization, we developed a DNA-based precipitation (DP) method to purify our RNA assemblies from cells. Streptavidin-coated magnetic beads with a biotinylated DNA capture probe (DPC) were added to cell lysates. The capture domain of DPC binds the T7 terminator in our RNA molecules (Figure 18C, step i). The RNA assemblies were released upon addition of DPR that bound the release domain of DPC (Figure 18C, step ii). We were able to capture and release RNA (Figure 18D).

In vivo synthesized D₁ and D₂ revealed extended 1D and 2D assemblies (Figure 18E). Cross-sectional height analysis showed D₁ to have two populations of distinct height (3 and 6 nm), which is characteristic of open versus closed nanotubes. *In vivo* D₂ assembled into 2D structures that are smaller and somewhat different than their *in vitro* counterparts, suggesting that the assembly process in cells is of lower fidelity. To confirm that the assemblies formed *in vivo*, we engineered a set of inhibitory strands (ISs) that bound the trigger domains of dr' and dz'. The inhibition by these strands was confirmed *in vitro* (Figure 18F). The purification of D₁/D₂ in the presence of excess ISs did not eliminate the observed 1D and 2D assemblies (Figure 18G), confirming the formation of D₁ and D₂ in cells pre-lysis.

Transmission electron microscopy (TEM) analysis of whole bacterial cells expressing D₁ or D₂ confirmed their assembly in cells. The RNA assemblies were tagged with gold-binding metallothionein-PP7 fusion proteins (PAu) that form clusters (Figure 18H) (Diestra et al. 2009). Cells coexpressing PAu and D₁ formed thin filaments with

lengths of 200 to 300 nm, whereas cells coexpressing D2 formed compact sphere-like structures ~ 100 nm in diameter. D₀, D₁, or D₂ does not affect cell growth (Figure 52, Appendix IV). Cells carrying the D₁ and D₂ scaffolds had higher RNA levels relative to cells expressing mutated poly-T RNA analogs (Figure 18I), consistent with the formation of degradation resistant assemblies. Thus, d₁ and d₂'/d₂" assembled *in vivo* into D₁ and D₂.

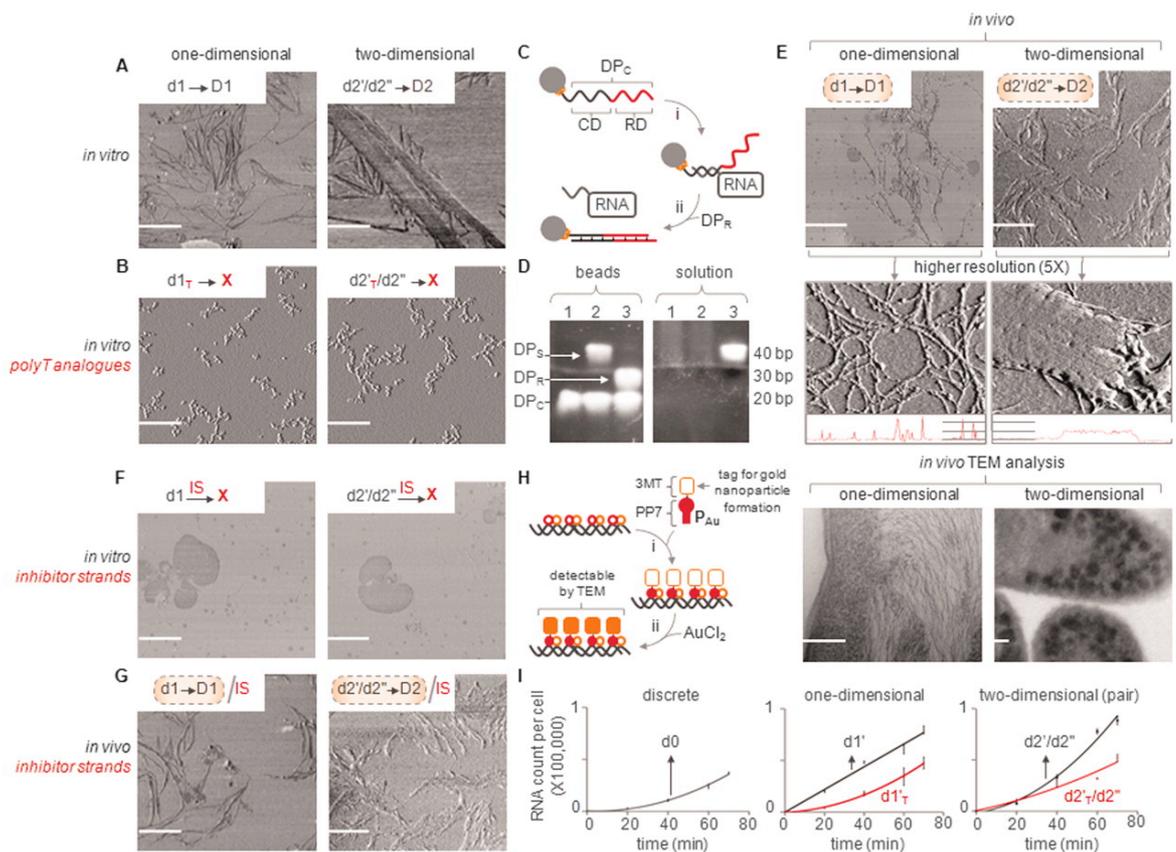


Figure 18: Characterization of RNA assemblies. (A) *In vitro* transcribed d₁ and d₂'/d₂" assemble into D₁ and D₂ (AFM; phase images; scale bars, 0.25 μ m). (B) *In vitro*

transcribed mutated RNA d_1T and $d_2'T/d_2''$ do not assemble. (C) DNA-based precipitation of *in vivo* RNA assemblies uses DPC (i) and a release probe (DPR) for recovery (ii). (D) Capture and release of substrate DPS (left gel, beads; right gel, solution). Lane 1, conjugation of DPC to streptavidin-coated magnetic beads; lane 2, capture of DPS; lane 3, release of DPS using DPR. (E) AFM analysis of purified assemblies. (F) ISs bind the DDs of d_1 and d_2' to prevent their assembly into D_1 or D_2 (circular structures are drying artifacts). (G) When used during the purification of d_1 and d_2'/d_2'' , D_1 and D_2 assemblies are still observed. (H) TEM analysis revealed the formation of 1D assemblies for D_1 and 2D aggregates for D_2 (scale bars, 100 nm). (I) Quantitative real-time fluorescence polymerase chain reaction analysis of *in vivo* RNA production levels. Error bars indicate SEM.

4. Organizing proteins onto RNA scaffolds

RNA can be used to spatially organize proteins in cells. We used fluorescence complementation to detect protein assembly on our RNA scaffolds (Valencia-Burton et al. 2007). Green fluorescent protein (GFP) split into two halves (FA and FB) fused to the PP7 or MS2 aptamer binding proteins was used (Figure 19A). Cells expressing FA and FB alone (Figure 19B) or D_0 , D_1 , or D_2 without the split GFPs displayed little fluorescence. However, the coexpression of D_0 , D_1 , or D_2 with the split GFPs showed increased fluorescence (Figure 19C). Thus, our RNA scaffolds served as docking sites to promote protein-protein interactions in cells.

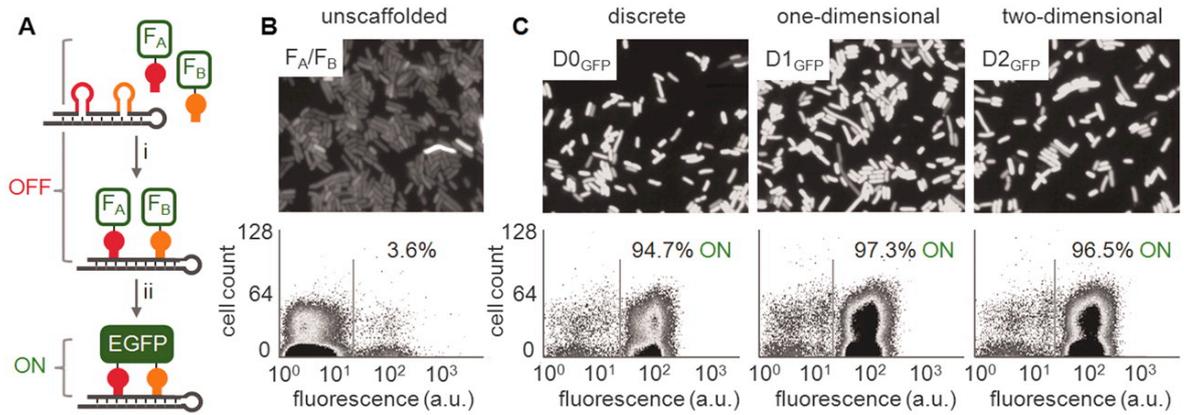


Figure 19: Fluorescence protein complementation *in vivo*. (A) GFP split into two halves, each of which is fused to PP7 or MS2 (F_A and F_B). F_A and F_B bind their respective aptamers (i) and reconstruct functional fluorescent GFP (ii). EGFP, enhanced green fluorescent protein. (B) Fluorescence microscopy imaging of cells expressing F_A and F_B revealed little to no fluorescence (scale bars, 10 μm). a.u., arbitrary units. (C) Cells coexpressing F_A and F_B with D_0 , D_1 , or D_2 reveal an increase in fluorescence, indicating that D_0 , D_1 , and D_2 scaffold PP7 and MS2 protein chimeras. Gray lines in flow cytometry plots separate OFF and ON cells.

Biological hydrogen production has both fundamental and practical implications. Co-expression of [FeFe]-hydrogenase and ferredoxin catalyzes the reduction of protons to hydrogen through electron transfer (Agapakis et al. 2010). We used this system to assess the potential of our RNA scaffolds to constrain flux through spatial organization. We fused the hydrogenase to a single copy of PP7 (HP) and ferredoxin to a dimer of MS2 (FM), and we conducted electrophoretic gel-shift analysis of the binding of FM and HP to D_0 (Figure 20A). Addition of HP to D_0 resulted in a single product termed

DoH. The addition of FP to Do resulted in the formation of DoF. The addition of HP and FM to Do resulted in a single product assigned to the protein-RNA assembly DoFH. HP and FM assembled onto Do in cells to form DoFH (Figure 20B).

To determine whether our RNA scaffolds increased hydrogen biosynthesis, we used gas chromatography to analyze cells expressing the hydrogen-producing pathway, along with the different RNA assemblies (Figure 51, Appendix IV). The relative levels of FM and HP expression in Do, D1, and D2 cells were comparable (Figure 49, Appendix IV). Do, D1, and D2 assembled FM and HP into DoFH, D1FH, and D2FH (Figure 51, Appendix IV). Do resulted in a 4.0 ± 1.3 -fold increase in hydrogen production compared with unscaffolded HP and FM (Figure 20C). Hydrogen output with the extended assemblies D1 and D2 resulted in a 11 ± 2.8 - and 48 ± 1.5 -fold increase in hydrogen production (Figure 20C). When normalized against the amount of RNA in cells (Figure 18I and Figure 32, Chapter VII), Do, D1, and D2 resulted in a 4.0-, 6.2-, and 24-fold increase. The increase with D2 is consistent with its assembly *in vivo* into “organelle-like” structures effective at concentrating proteins and their products (Figure 18H). Mutating the PP7 and MS2 binding sites prevented protein scaffolding (Figure 20, D to G). Thus, RNA can be used to organize enzymatic pathways *in vivo* to increase output as a function of architecture.

We controlled the spatial organization of proteins in cells using RNA molecules that are sequence-programmed to isothermally assemble into predefined discrete, 1D, and 2D structures *in vivo*. These assemblies scaffolded proteins and were used to organize a hydrogen-producing biosynthetic pathway. Hydrogen production was optimized as a function of scaffold architecture. Unlike protein-based approaches (Dueber et al.

2009; Park et al. 2003; Agapakis et al. 2010), RNA-based scaffolds allow for the formation of complex multidimensional architectures with nanometer precision. *In vivo* RNA assemblies can thus be used to engineer biological pathways through spatial constraints (Isaacs et al. 2006; Win & Smolke 2008).

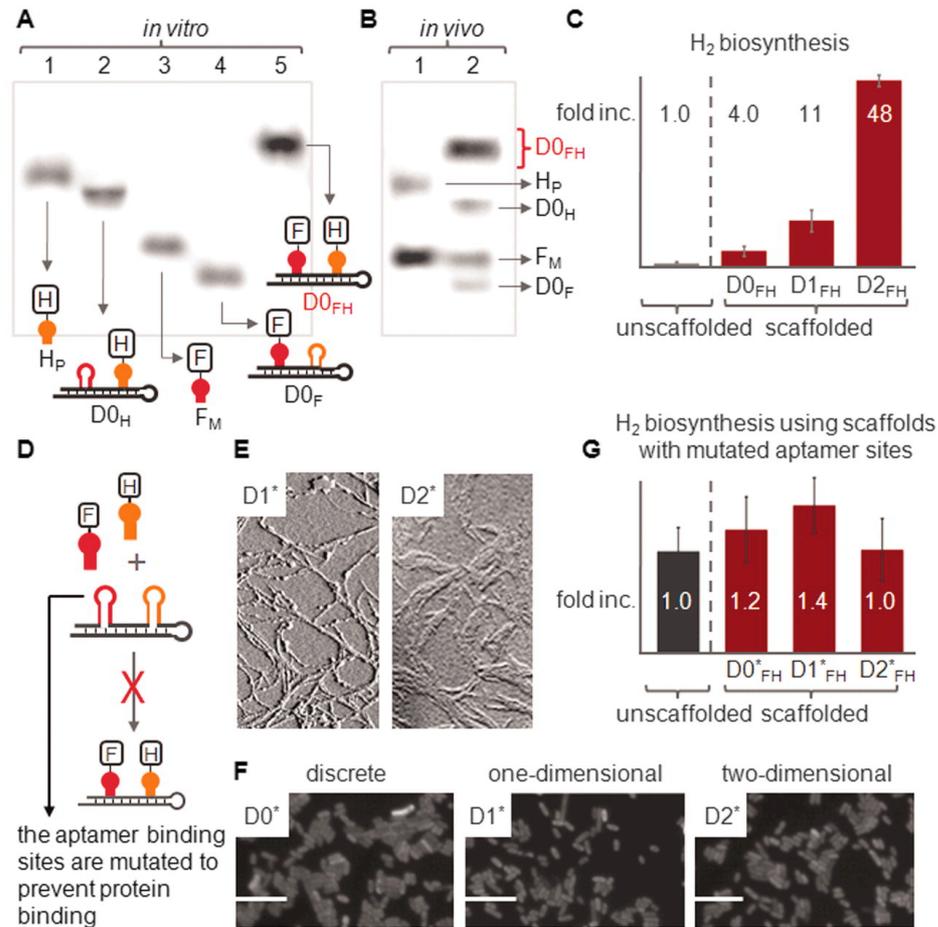


Figure 20: Scaffolding hydrogen production. (A) *In vitro* gel shift of HP (lane 1) binds Do to form DoH (lane 2). FM (lane 3) binds Do to form DoF (lane 4). HP and FM

bind Do to form DoFH (lane 5). (B) *In vivo* gel shift of HP and FM (lane 1) and HP and FM in the presence of Do (lane 2). (C) Hydrogen biosynthesis as a function of scaffold, normalized to unscaffolded cells expressing HP and FM. (D) Mutating aptamer binding sites (E) do not affect self-assembly, (F) but do prevent protein binding (scale bars, 10 μm) and (G) hydrogen production. Error bars indicate SEM. Dashed lines in (C) and (G) denote separation between scaffolded and unscaffolded proteins.



VI - Standardizing the construction and use of RNA scaffolds

The following Nature Protocol paper is an attempt at streamlining the making of RNA scaffolds. We took the core principles of Synthetic Biology as mottos to write this contribution: abstraction, standardization and modularity. A synthetic biologist does not need to be an expert at RNA nanotechnology to design and make useful RNA scaffolds. We designed a cloning strategy around a set of well defined vectors and restriction sites to enable the rapid testing and easy sharing of RNA scaffold designs. Last but not least, we highlight the modularity of RNA scaffolding and potential de-

sign variants. Hopefully, it also highlights the array of potential applications of RNA scaffolding and enables an easier access to this technology - be it for the molecular biologist, the synthetic biologist or the chemist. The edited paper is also provided in Appendix III for easier read.

1. Introduction

1. Natural and engineered ncRNA

Natural ncRNA molecules derive their diverse range of behaviors from their unique ability to fold into complex tertiary structures with recognition and even catalytic properties (Eddy 2001). These properties in turn inspired the RNA hypothesis for the origins of life and much effort has been devoted to evolving novel RNA functions by SELEX (Tuerk & Gold 1990; Ellington & Szostak 1990). Recently, RNA molecules were rationally designed to perform specific tasks both *in vitro* and *in vivo* (Isaacs et al. 2006; Win & Smolke 2008; Callura et al. 2010), scaffolding being a compelling new application (Delebecque et al. 2011). As our understanding about the causal relationship between primary sequence, secondary structure and function grows, RNA is now viewed as a modular molecule with extensive engineering potential.

2. Advantages and applications of synthetic RNA scaffolds

Scaffolding is widely used in nature. For example, multi-enzyme pathways are often physically and spatially organized onto clusters through protein domain interactions (Conrado et al. 2008), microcompartments (Savage et al. 2010; Burack & Shaw 2010),

or natural RNA scaffolds (Zappulla 2004; Cayrol et al. 2009; Shevtsov & Dunder 2011). Spatial organization helps direct substrate flow between interacting enzymes, limiting cross-talk and increasing the yields of sequential metabolic reactions (Adam & Delbruck 1968; Dueber et al. 2009; Lee et al. 2011; Park et al. 2003; Agapakis et al. 2010; Moon et al. 2010; Conrado et al. 2012).

Previously, synthetic protein scaffolds were developed to direct flux in synthetic metabolic pathways (e.g., improving titers of mevalonate in *E. coli* (Dueber et al. 2009)). These scaffolds were built by fusing three eukaryotic protein-protein interaction domains (PDZ, SH3 and GBD) and co-expressing proteins to be scaffolded as fusions with their cognate binding domains. This strategy allows for the localization and stoichiometric control of a limited number of proteins and was successfully applied to improve yields of hydrogen and glucaric acid synthesis in *E. coli* (Agapakis et al. 2010; Moon et al. 2010). Plasmid DNA was also recently used for scaffolding and improving the titer of resveratrol, 1,2 propanediol and melvanoate (Conrado et al. 2012).

The attraction of RNA scaffolds is their ability to be rationally programmed using the rules of base-pairing. This offers access to larger scaffolds, in which hundreds of proteins are gathered to work together, and confers the ability to control not only stoichiometry but also the distance and orientation between interacting proteins. With hundreds of different, orthogonal, characterized aptamer domains (Bunka & Stockley 2006) and thus an expansive range of different binding domains, RNA scaffolds may bring together large complex pathways.

RNA scaffolds are applicable in many areas in which the spatial organization of biomolecules is desirable. In the synthetic biology realm, RNA scaffolds bring an added level of control by offering a tunable platform to control the spatial organization of proteins (Delebecque et al. 2011). However, current tested designs are limited to the assembly of two distinct aptamers, but can be in principle enlarged either by using the discrete system (see Experimental Design) or by mixing different aptamers in given ratios. Scaffold and assembly size can also be further controlled by adding non-polymerizing 'ends' or by using the discrete system as described in the Experimental Design. Furthermore, RNA scaffold libraries will also be expanded by applying future advances in the RNA nanotechnology field (Shu et al. 2004; Guo 2010; Afonin et al. 2010) or through inspiration from naturally existing structures (e.g., DsrA RNA (Cayrol et al. 2009)) as well as by designing directed evolution selections for functional ncRNA-mediated assemblies.

2. Experimental Design

The general workflow for the design, induction and analysis of RNA scaffold expression is illustrated in Figure 21 below.

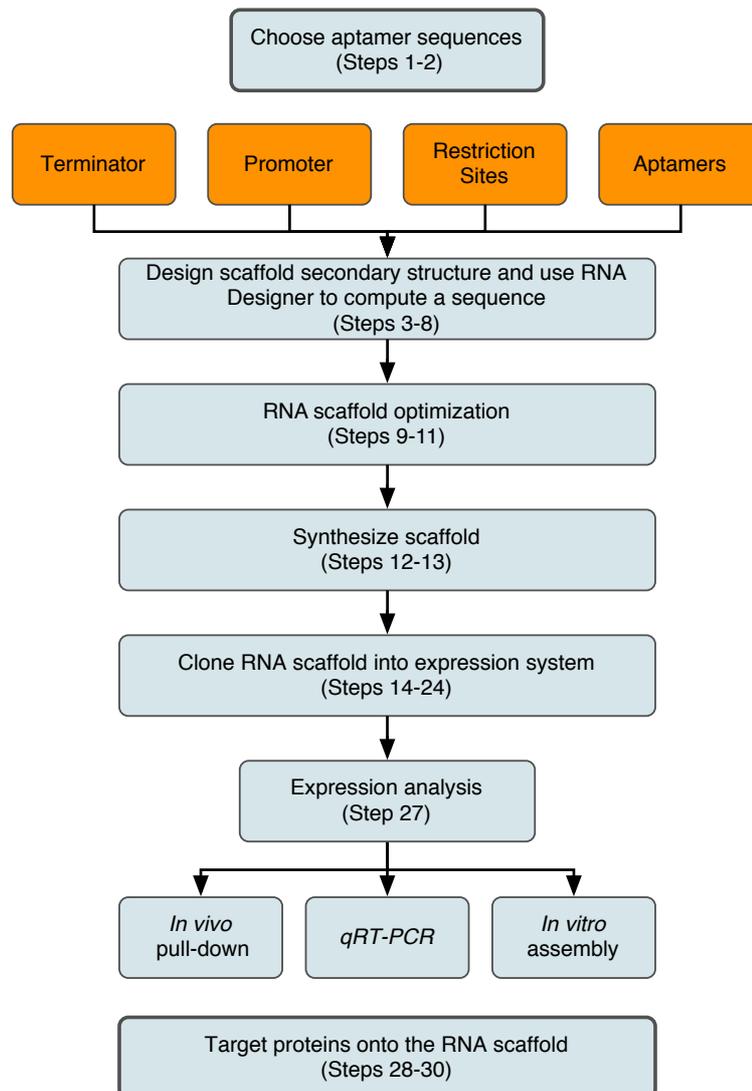


Figure 21: General workflow for design, induction, expression and experimental testing of the RNA scaffold. Orange boxes represent sequence constraints to be input into RNAdesigner.

1. RNA scaffold design; RNA designer (Steps 1–8).

RNA scaffold design and optimization uses RNA folding software that relies on a free energy minimization algorithm to predict sequences that are amenable to *in vivo* scaffolding. RNA designer (Andronescu et al. 2003), for example, can be used to design the primary sequence of an RNA molecule that folds into a desired secondary structure (Figure 22).

This software computes an RNA sequence that folds into the specified secondary structure given a number of optional sequence constraints. In the present context, the sequence constraints to be specified consist of the sequence of aptamers, terminator and restriction enzyme sites that will be used. The secondary structure specified depends on the RNA folding scheme to be used (see below). The simulation should be run at physiological temperature (e.g., 37°C for *E. coli*) and with a target GC percentage matching the genomic content of the organism in which the RNA will be expressed (e.g., about 51% for *E. coli*).

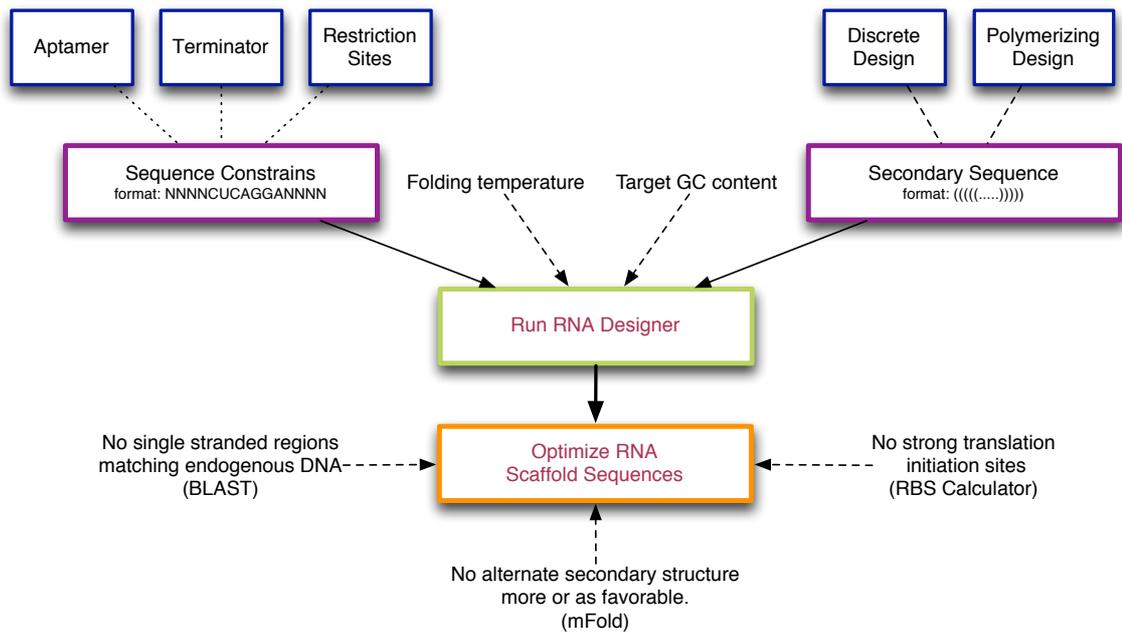


Figure 22: Designing and optimizing RNA scaffolds using RNAdesigner. The design process starts with choosing both the sequence constraints and the secondary structure folding scheme. Parameters to input are linked through dotted lines: for the sequence constraints, aptamer, terminator and restriction enzyme site sequences need to be input; while the secondary structure should be input in dot-bracket format and will vary according to whether a discrete or polymerizing folding scheme is chosen. Run RNAdesigner taking into account folding temperature and a target GC content. Optimizing the RNA scaffold sequence output is a critical part of the design process and important parameters to consider are annotated with dotted lines. Bioinformatic programs to use to check these parameters are shown in brackets.

2. RNA scaffold design; choosing aptamers to tether proteins to the RNA scaffolds (Steps 1–2).

To choose the aptamers for the scaffold to be designed, a number of parameters should be considered: (i) the binding affinity between the aptamer and its binding (adaptor) protein should be as high as possible (e.g. nanomolar range); (ii) the binding should also be as specific as possible, leading to mutually orthogonal aptamers and (iii) the binding protein sequence should be optimized for bacterial expression in terms of codon usage and stability.

One such set of mutually orthogonal and extensively studied aptamers are those from MS2 and PP7 bacteriophages (Chao et al. 2007; Convery et al. 2004) with dissociation constants of about 82 nM (F6 aptamer) and 1 nM respectively (Lim et al. 2001; Parrott et al. 2000) (Figure 23). This set of two aptamers will enable repetitive scaffolding of two different proteins from a chosen pathway (e.g. [Fe-Fe] Hydrogenase and Ferredoxin).

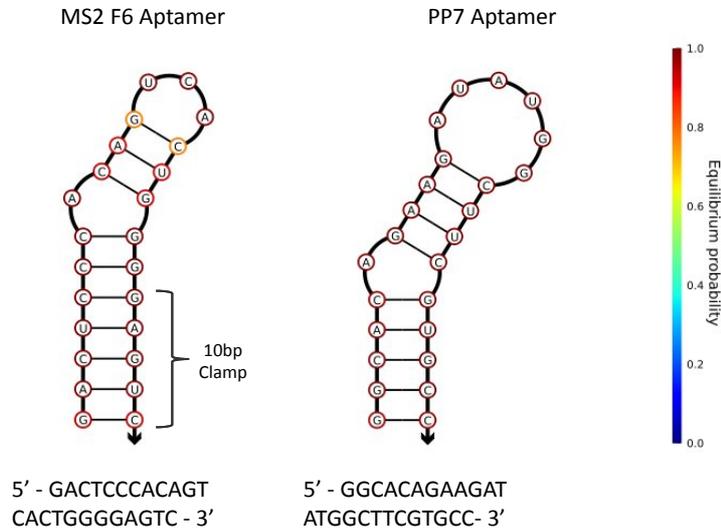


Figure 23: Compatible aptamer pair. Primary sequence and secondary structure of the MS2 F6 aptamer and PP7 aptamers. Both aptamers are orthogonal to each other and are usable for RNA scaffolds. A 10bp clamp (as indicated on figure) is added to the stem of the 14bp MS2 F6 aptamer to adjust its length to that of the PP7 aptamer. Visualization with NuPack, the color coded scale depicts the base-pairing probabilities.

3. RNA scaffold design; choosing a folding scheme (Steps 3–8).

RNA scaffolds can be expressed *in vivo* as discrete (Figure 24, A and B) or polymerizing molecules (Figure 24C). The latter requires a more complex approach to both design and characterization but allows for more complex architectures and RNA-RNA

interactions to be studied *in vivo* (see Step 3B). Discrete scaffolds are easier to engineer and can also be used as tags for mRNA expression studies (Golding & Cox 2009; Valencia-Burton et al. 2007).

For metabolic engineering purposes, the scaffolds should be modular in length and number of docking sites, as well as easy to characterize. For this application, we suggest using an initial “discrete” scaffold. This ncRNA contains multiple copies of the chosen aptamer flanked by spacers that define the relative distance and orientation between the folded aptamers, as successfully used in previous work aiming at mRNA tagging (Golding & Cox 2009; Valencia-Burton et al. 2007) (Figure 24A). This design can be serially cloned to reach a desired scaffold length (typically 96-mer or more, (Golding et al. 2005)).

To achieve more complex scaffold geometries, scaffolds can instead be made of polymerizing RNA molecules. This relies on the molecular cross-assembly of RNA strands based on principles from the toolbox of RNA and DNA nanotechnology, including symmetry and kinetic considerations, and Watson-Crick as well as non-canonical interactions (Guo 2010; Aldaye et al. 2008; Seeman 2007). Polymerizing scaffolds are made of short RNA strands cross-polymerizing to create extended structures. Much exploration remains to be done to better understand and further expand the library of assembling RNAs and assembly schemes. Here, we give recommendations on how to utilize both symmetry and kinetic assembly, which are key to our published assembly scheme.

Our polymerizing scaffold design strategy relies on two principles. Firstly, the use of palindromic sequences minimizes the number of different interacting strands necessary to form the extended structures. Using sequence symmetry, it is possible to design nanotubes or two-dimensional sheets with only one or two different polymerizing strands, respectively (Guo 2010; Liu et al. 2006; Liu et al. 2005). Secondly, kinetic assembly pathways (Yin et al. 2008) enable the assembly to occur isothermally. In our design, RNA molecules assemble in a two-step process in which all non-interacting regions are locked into metastable assembly-intermediate hairpins. They only unfold upon cross-interaction through their “dimerization region” leading to the polymerization of the RNA into extended structures (Delebecque et al. 2011).

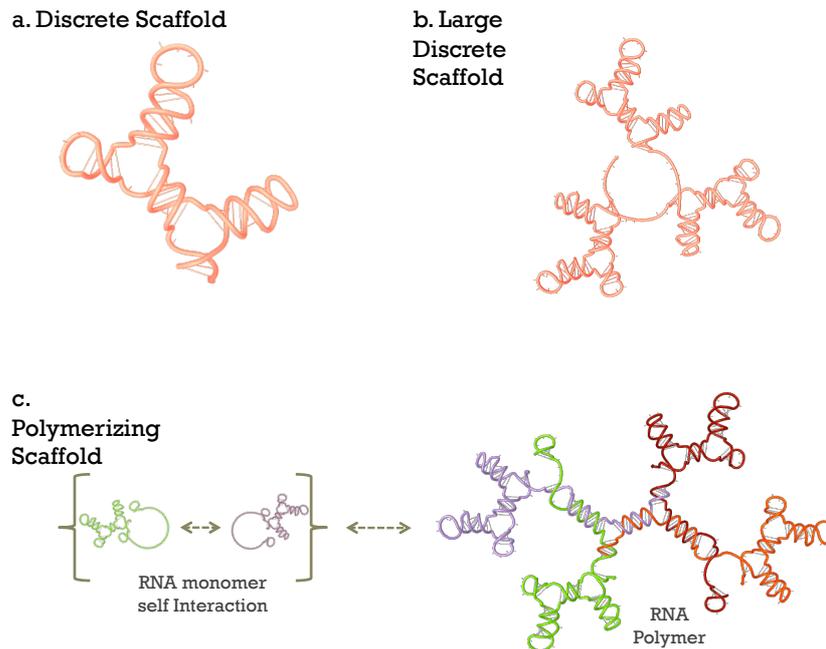


Figure 24: Examples of RNA scaffold folding scheme designs. (a) Computed secondary structure of a small discrete scaffold design comprised of two aptamer protein

binding sites. (b) Secondary structure of an extended discrete scaffold expressed as a repeat of the small discrete scaffold. (c) Proposed secondary structure for a polymerizing scaffold. Small RNA modules interact through single stranded domains and polymerize into the multimeric scaffold. Visualization with NuPack.

4. RNA scaffold optimization (Steps 9–11).

Often, a given sequence output from RNA designer folds into a number of alternative secondary structures. To further optimize the structure, these should be analyzed using another RNA secondary structure program such as mFold (Zuker 2003) or Nupack (Zadeh et al. 2010). By screening the top RNA designer outputs, one should pick the sequence that optimizes the thermodynamic gap between the desired folding and the next most favorable structure. Finally, the candidate sequences should be screened against any ribosome binding sites using an RBS calculator (Salis et al. 2009) (<https://salis.psu.edu/software/>) in order to avoid translation and against complete match of free single stranded regions with any endogenous mRNA using the BLAST server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The workflow for RNA scaffold design and optimization is illustrated in Figure 21.

5. Cloning the designed RNA scaffold into an expression system (Steps 12–24).

We have taken a modular cloning approach in which the key elements—promoters, assembly tags, aptamer protein binding sites and terminators—are separated by unique restriction sites to allow for tuning of the fundamental properties of the scaffold through quick and precise changes (Figure 25A). This cloning strategy allows users to

switch between the two different scaffold architectures: (i) the concatemerization of the aptamer module to create large discrete scaffolds and (ii) assembly by polymerization of discrete ncRNA molecules (Figure 25B). Once the construct is verified, it is cloned into commercial plasmids (e.g., pETDuet-1, EMD Chemicals) for *in vivo* expression.

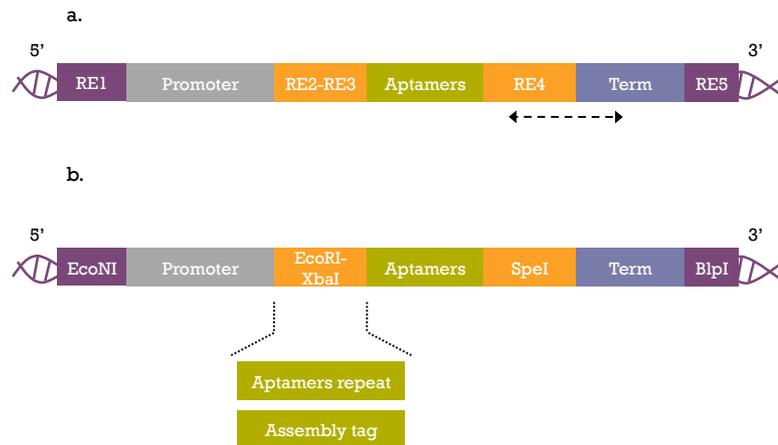


Figure 25: Modular cloning approach to the design of RNA Scaffolds. (a) Separation of promoter (grey box), aptamer (green box) and terminator (blue box) sequences by five unique restriction sites (RE; Purple and yellow boxes). The dotted arrow indicates the suggested region to hybridize probes for the pull-down assay (see Step 27Biii) (b) Suggested restriction enzymes. RE₁=EcoNI and RE₅=BlnI enable cloning into the pETDuet-1 Vector (EMD Chemicals). RE₂=EcoRI, RE₃=XbaI and RE₄=SpeI enable repeated cloning of the aptamer domain to create large discrete scaffolds or the cloning of an assembly tag (see Step 3Bi) to create polymerizing RNA molecules.

6. Expression systems (Steps 25–27).

For initial characterization of a candidate RNA scaffold, we recommend the use of a T7 expression system. This well-characterized expression system can be used for preliminary *in vitro* expression studies using commercially available IVT systems (e.g., MEGAscript T7 Kit from Invitrogen), and can be tuned for a wide range of expression levels, yielding very high expression levels upon full induction. We thus suggest cloning the RNA scaffold expression cassette into a pET based vector (e.g., PetDuet-1, EMD Chemicals) and transforming it into BL21 DE3 *E. coli* cells. Scaffold concentration is tunable *in vivo* by using different amounts of the inducing molecule (e.g., Isopropyl β -D-1-thiogalactopyranoside (IPTG) or arabinose). A number of compatible T7-based plasmids have been developed and can be used for the simultaneous expression of up to six different proteins in addition to the RNA scaffold (e.g., EMD Chemicals Duet Vectors system). It should be noted that the T7 system is the only expression system tested in ref. 1, and other systems may also be useful. *in vivo*

7. DNA capture probes for scaffold expression analysis by *in vivo* pull-down (Step 27Biii).

The DNA capture probe has two functional regions with well-defined melting temperatures (T_m). The binding domain is designed to interact with a ~10 bases long constant region that is single stranded and non-assembled (e.g. the region between RE4 and the start of the terminator, see Figure 25). The annealing temperature of this region should be ~20°C. “The release region of the DNA capture probe should be designed to be about 10 bases. These extra 10 bases should be designed so that they do

not form hairpins with the binding domain and bring the T_m of the 20 bases full DNA capture probe to -35°C . The probe should then be ordered as a 5' biotinylated oligo.

8. Primer design for quantitative RT-PCR (Step 27Ciii).

Quantitative RT-PCR of small RNAs with complex secondary structures can be complicated. One strategy used by commercial kits is to add a poly-A tail to the RNA scaffold during retro-transcription

(Benes & Castoldi 2010). This provides a specific site to anneal a reverse primer for the quantitative PCR. A specific forward primer is equally important and we recommend designing three of them with a dedicated software program using a target T_m of 60°C (e.g., primer-BLAST), hybridizing between the RE4 and the end of the terminator stem and assessing their performance by incorporating a melting curve analysis step at the end of the PCR program (refer to instrument manual for specific programming). Choose a primer giving a clear single peak in the melting curve graph.

9. Stability considerations.

Different strategies can be considered to enhance the stability of the RNA scaffold. RNA turnover is a natural component of gene expression. Half-lives of most bacterial RNAs range from 40 seconds to 60 minutes (Selinger 2003). Steady-state transcript concentrations are a result of degradation and synthesis rate (Richards et al. 2008). Strategies to enhance RNA scaffold level include the use of highly efficient expression systems (e.g., T7 based expression system) but also the implementation of design considerations to minimize decay.

In *E. coli*, 5' and 3'-end accessibility is of particular importance in initiating the decay process. For example, RNase E requires 5' single stranded RNA of at least four nucleotides in length for efficient binding, while RNase II and PNPase are unable to bind substrates with fewer than 6-10 unpaired bases at the 3' end (Richards et al. 2008). Therefore, minimizing single stranded regions and locking both ends of the RNA scaffold with hairpins is an important part of the design strategy (Molinaro & Tinoco 1995) (Figure 24). Additionally, using RNaseE knock-out strains for expression (e.g., BL21 DE3 Star, Invitrogen) may also help further stabilize the RNA transcripts (Richards et al. 2008).

10. Targeting proteins onto the RNA scaffold (Steps 28–30).

The chosen aptamer-binding proteins are fused to the proteins to be scaffolded. For this step, it is advised to explore the space of possible linker length and fusion orientation so as to both optimize scaffolding and protein interactions. It is also useful to examine three-dimensional structures of the protein domains to be used to determine whether an N- or C-terminus can be used as a fusion junction without interfering with the function of the protein.

11. Controls.

Impairing RNA polymerization (Step 27Aiii): Short DNA oligos are designed to match and hybridize to the dimerization region of the polymerizing scaffold to prevent RNA polymerization. RT-qPCR control (Step 27Cv): GapA or MreB *E. coli* housekeeping genes are used as internal references to assess the relative RNAs given their scaffold concentration following induction. Impairing Protein binding (Step 30): RNA scaffold

folds with aptamers mutated to poly-T and A sequences are designed to prevent protein binding.

3. Materials

1. Reagents

All solutions and buffers should be kept sterile and according to the manufacturer's recommendation.

Chemicals and solvents:

- 1 Kb plus ladder (Invitrogen, USA, cat. no. 10787-018)
- 6% TBE-Urea gels (Novex gels; Invitrogen, USA, cat. no. EC68652BOX)
- Acetic Acid (Fisher Scientific, USA, cat. no. A35-500)
- Ampicillin (Sigma-Aldrich, USA, cat. no. A9518)
- Bacterial peptone (Fisher Scientific, USA, cat. no. BP1420-2)
- Bacterial Protein Extraction Reagent (Pierce, USA, cat. no. 78243)
- ddH₂O, Sterile
- Dithiothreitol (Invitrogen, USA, cat. no. 15508-013)
- Dynabeads M-280 Streptavidin (Invitrogen, USA, cat. no. 112-05D)
- EDTA (ethylenediaminetetraacetic acid; Invitrogen, USA, cat. no. AM9261)
- Ethidium Bromide (1% (wt/vol) solution; Sigma-Aldrich, USA, cat. no. E-8751) !CAUTION Ethidium bromide is mutagenic, always use nitrile gloves.

- Glycerol (Sigma-Aldrich, USA, cat. no. G5516-100ML)
- IPTG (Anatrace, USA, cat. no. I1002)
- Lysozyme (Sigma-Aldrich, USA, cat. no. L6876-5G)
- Potassium Chloride (Sigma-Aldrich, USA, cat. no. P5405)
- SeaKem LE Agarose (Lonza Cologne GmbH, Germany, cat. no. 50002)
- Sodium Chloride (Fisher Scientific, USA, cat. no. S671-500)
- Tris (tris(hydroxymethyl)aminomethane; Fisher Scientific, USA, cat. no. BP152-5)
- Yeast extract (VWR, USA, cat. no. 97063-370)
- Luria Broth (LB) (Sigma-Aldrich, USA, cat. no. L2542-500ML)

Bacterial strains and vectors:

- E. coli* cloning strain (Turbo Competent *E. coli*; NEB, USA, cat. no. C2984H)
- E. coli* expression strain (One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* Invitrogen, Germany, cat. no C6010-03)
- pETDuet-1 T7 expression vector (EMD Chemicals, USA, cat. no. 71146-3)
- pACYCDuet-1 T7 expression vector (EMD Chemicals, USA, cat. no. 71147-3)
- pCOLADuet-1 T7 expression vector (EMD Chemicals, USA, cat. no. 71406-3)
- pCDFDuet-1 T7 expression vector (EMD Chemicals, USA, cat. no. 71340-3)

Kits:

- MEGAscript T7 Kit (Invitrogen, USA, cat. no. AM1334M)
- NCode VILO miRNA cDNA Synthesis Kit (Invitrogen, USA, cat. no. A11193-050)

This kit allows for the detection and quantification of small RNAs by adding a poly-A tail prior to the retro-transcription step.

- QIAprep Spin Miniprep Kit (Qiagen, USA, cat. no. 27104)
- QIAquick Gel Extraction Kit (Qiagen, USA, cat. no. 28704)
- SYBR Green Supermix (Invitrogen, USA, cat. no. 4309155)
- Total RNA Purification 96-Well Kit (Norgen, Canada, cat. no. 24300)

This column-based purification kit allows for the purification of total RNA without cut-off in size.

Enzymes:

- Restriction endonucleases: EcoNI, BspI (Fermentas, Germany, cat. nos. FD1304 and FD0094, respectively).
- T4 DNA ligase (Promega, Germany, cat. no. M1801)

Buffers:

- Assembly buffer: 150 mM KCl, 0.1 mM dithiothreitol, 0.1 mM EDTA, 10 mM Tris, in ddH₂O, made fresh and adjusted to pH 7.4.

PCR primers (Integrated DNA Technologies, Inc., USA)

All PCR primers should be designed using appropriate software (e.g., primer-BLAST), in order to avoid self or hetero-dimers as well as complementary to genomic DNA in order to avoid non-specific amplifications. Primers should be diluted to 100 μ M and kept at -20°C until use.

- DNA oligos (Integrated DNA Technologies, Inc., USA)
- Biotinylated DNA capture probes (Integrated DNA Technologies, Inc., USA)

2. Equipment

- Asylum MFP-3D Atomic Force Microscope (Asylum, USA)
- Cuvettes (Semi-Micro Cuvettes; BrandTech Scientific, USA, cat. no. 2711010)
- Eppendorf centrifuge (Eppendorf AG, Germany, cat. no. 5702R)
- Eppendorf Mastercycler ep realplex (Eppendorf AG, Germany, cat. no. 6302 000.601)
- Etched silicon cantilever (Olympius, USA, cat. no. OMCL-AC160TS)
- Highest Grade V1 Mica, 25 x 76mm (Tedpella, USA, cat. no. 56)
- Incubation facilities for bacterial culture
- Liquid nitrogen
- Spectrophotometer (Ultrospec 3100 pro; GE Healthcare, Germany, .cat. no. 80-2112-32)

3. Bioinformatics RNA Design tools:

- RNAdesigner (<http://www.rnasoft.ca/cgi-bin/RNAsoft/RNAdesigner/rnadesign.pl>)
- Mfold (<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>)
- Nupack (<http://www.nupack.org/>)
- RBS Calculator (<https://salis.psu.edu/software/>)

4. Protocol

1. RNA scaffold design – Timing 2 h

Start by choosing an appropriate set of aptamers (e.g., MS2 and PP7, see Experimental Design for further details and Figure 23).

If the aptamer stem loops are of different lengths, add complementary bases at both sides of the shorter stem's root in order to equilibrate their length (see example in Figure 24 - use RNAdesigner).

Choose an appropriate folding scheme (see Experimental Design for further details) and design the secondary structure of the scaffold. Use option A for designing discrete scaffolds made of single RNA strands and option B for designing polymerizing scaffolds made of cross-assembling RNA strands.

A- Discrete scaffold design.

- i) Design the scaffold RNA strand to be relatively short, up to 200 – 300 bases so that it can be synthesized.
- ii) Design the scaffold so that aptamer binding sites are protruding and spaced according to your choice (e.g. 11bp per one RNA helix, see example in Figure 26A).
- iii) Design a complementary 3' region of 20-30 bases that folds back on itself thus minimizing single stranded regions and locking the whole structure into a duplex (see example in Figure 26A, discrete scaffold bases 80 to 106).

B – Polymerizing scaffold design.

i) Start by designing the scaffold domain responsible for polymerization, the “assembly tag” (see Figure 25). The assembly tag is divided into two functional regions, both palindromic. The dimerization region is involved in the first step of the assembly process and it should remain available until assembly. To prevent it from collapsing due to its symmetric sequence, design the polymerization region so that it partly folds back onto the dimerization region (see example in Figure 26B). This both stabilizes the dimerization region and makes it available for the first step of the assembly process. Upon cross-interaction, the polymerization domain becomes available by strand displacement and tiles can then assemble.

NB: Stabilizing the dimerization region using the polymerization region might require the use of wobble base pairing and some shuffling. See Figure 26 for a detailed example.

first probe with RNA designer to fold the polymerization domain onto the dimerization domain, thus creating the metastable intermediate. We also ask it to start making palindromes by asking for hairpins (in yellow are extra sequences to enable RNA designer to find a solution). Some tweaking might be necessary to keep the metastable intermediate folding once the palindromes are manually completed. This involves using G-Us (Wobble pairs) or A-Ts base pairs instead of G-Cs (“shuffling”) to weaken some regions until the correct folding is achieved in mFold.

4) Using the dot-bracket format (Figure 22), input the secondary structure of the desired scaffold according to its folding scheme (from Step 3) into RNA designer.

5) Select and input the required primary sequence elements into the constraint field of RNAdesigner according to Figure 22. Specify the sequence constraints according to Figure 22, which should include the RNA aptamer sequences (from Steps 1-2) and the rho-independent terminator. “N” denotes positions where any bases can be used.

NB: If you are using the T7 promoter, the three last bases of the promoter are transcribed and should be considered in the sequence constraints.

6) Match target GC content to the organism into which the scaffold is to be expressed (a good database: <http://bionumbers.hms.harvard.edu>) and adjust the simulation temperature to temperature of organism growth.

7) Ensure that secondary and primary structure inputs match and are of equal length.

8) Run RNA Designer.

2. RNA scaffold optimization – Timing 1 h

- 9) Check the predicted structure of the top RNA designer results and their theoretical stabilities (free energies) using a tool such as the Mfold web server. Evaluate the probability of alternative secondary-structure formation by comparing the free energy of predicted alternative structures with that of the desired structure (see Experimental Design).
- 10) Select the most stable primary sequence with the desired secondary structure, which optimizes the thermodynamic gap between the desired folding and the next most favorable structure.
- 11) Confirm that the selected primary sequence is lacking any strong ribosome binding sites using, for example, the RBS calculator web server.

3. Cloning the designed RNA scaffold into an expression system TIMING 12–15 d

- 12) Add the features enabling expression and cloning to the selected primary sequence from Step 11: cloning site (RE1) and full promoter on the 5' end, and the chosen cloning site (RE5) after the terminator on the 3' end.
- 13) Order the full construct as a synthetic gene cloned into a high-copy cloning vector.

- 14) Transform competent bacterial cells (e.g., heat-shock competent *E. coli* Turbo cells from NEB) with about 0.1 ng of the high copy plasmid containing the scaffold using standard protocols⁴⁴.
- 15) Spread the transformed bacteria on LB plates (100–150 µL bacterial suspension per plate) containing the appropriate antibiotic for selection (e.g., 100 mg/ml ampicillin for high-copy number *E. coli* plasmids of the pUC or pBluescript series) and incubate them overnight at 37°C.
- 16) Inoculate a colony into an individual aliquot of 5 mL LB medium supplemented with the correct antibiotic (e.g., 100 µg/mL ampicillin), incubate the culture overnight at 37°C.
- 17) Isolate the plasmid using a plasmid isolation kit (e.g., Qiagen Miniprep kit), following the manufacturer's instructions.
- 18) Digest the plasmid using a suitable combination of restriction enzymes flanking the cassette (e.g., Fermentas FastDigest EcoNI and BspI restriction enzymes – see Figure 25B), following the manufacturer's instructions.
- 19) Separate the restriction enzyme-digested samples by electrophoresis in 1.5–2% agarose gels, using standard protocols⁴⁴. Verify the fragment size by using a suitable DNA ladder.
- 20) Purify the cassette-containing DNA using an agarose gel purification kit, following the manufacturer's instructions.

- 21) Digest the chosen expression vector (e.g., EMD Chemicals pETDuet-1 vector) with restriction enzymes corresponding to the 5' and 3' overhangs of the cassette (e.g., Fermentas FastDigest EcoNI and BlnI restriction enzymes), following the manufacturer's instructions.
- 22) Ligate the purified cassette from Step 20 into the digested expression vector from Step 21, using a T₄ DNA ligase (e.g., Promega - T₄ DNA ligase) using standard protocols (Tolia & Joshua-Tor 2006).
- 23) Transform into competent bacterial cells (e.g., Invitrogen - heat shock competent *E. coli* BL21 DE2-star for the suggested T7-based expression system), using standard protocols.
- 24) Identify correct clones by running purified and digested plasmids in a gel electrophoresis to confirm insert size corresponding to your scaffold using standard protocols. This should be further confirmed by sequencing. Correct clones can be kept as glycerol stock at -80°C.

4. Induction of RNA scaffold expression – Timing 1–2 d

- 25) Inoculate one colony of the scaffold-bearing strain from Step 23 into liquid growth medium (e.g., Luria-Bertani Broth (LB) or a minimal growth medium for *E. coli*) with appropriate antibiotic (e.g., 100 mg/ml ampicillin) in a 5 ml culture tube.
- 26) After overnight culture at 37°C, dilute the culture with fresh LB with appropriate antibiotic to 1/20 and let grow until an optical density (OD) of approximately 0.3 at

600nm is reached (about 2 hours, see ref. 44) and induce scaffold production with the appropriate molecule (e.g., with 0.1 to 1mM IPTG).

NB: Induction conditions are critical. Cells should be incubated at the temperature at which the RNA scaffold was designed (i.e. 37°C or 30°C). Induce the cells when they reach the beginning of exponential phase (i.e. OD 0.2-0.3).

5. Expression analysis

27) Proceed with expression analysis using one or more of options A-C, depending on whether the RNA scaffold is designed to polymerize. Option A is suitable for studying RNA polymerization *in vitro* and relies on the choice of a T7, T3 or SP6 promoter in the design. Option B is suitable to purify *in vivo* produced RNA samples or RNA-protein complexes. It is a modified pull-down assay where cross-linking is not necessary and biological samples are not denatured. Option C enables the precise quantification of *in vivo* RNA scaffolds.

A- *In vitro* assembly TIMING 2-3 d

i) Set up an *in vitro* transcription reaction using an appropriate kit (e.g., MEGAscript T7 Kit, Applied Biosystems). Use 500ng of linearized purified plasmid (e.g., digested with Fermentas FastDigest PstI and purified with QIAquick PCR purification Kit, according to manufacturers' protocols) containing the RNA scaffold expression cassette from Step 24. Perform *in vitro* transcription overnight at 37°C according to manufacturer's instructions.

NB: Circular plasmid templates will generate extremely long heterogeneous RNA transcripts because RNA polymerases are processive and rho-independent terminators may not provide efficient termination *in vitro*. Thus, it is preferable to digest the plasmid with an appropriate restriction enzyme downstream of the RNA scaffold cassette. It is worthwhile to also gel-purify the linearized plasmid.

ii) Purify the *in vitro* transcribed RNA scaffold using the Norgen RNA purification kit according to manufacturer's instruction. The purified *in vitro* transcribed RNA scaffold may be stored at -20°C for 2-3 days. It is recommended that samples be placed at -80°C for long-term storage.

iii) Deposit ~ 25 ng of the RNA onto freshly cleaved mica (2.0 cm^2), allow to dry for 20 minutes at the appropriate assembly temperature (e.g., 30°C or 37°C). As an appropriate negative control, allow for assembly with a 10 fold molar excess of DNA oligos designed to impair RNA interactions. These oligos should be fully complementary to the interacting regions of the RNA molecules (e.g., dimerization region in Figure 26)

iv) Visualize the presence of RNA assemblies (e.g. RNA polymers of length 10-100nm) using an Atomic Force Microscope (e.g., Asylum MFP-3D). Perform acquisition in air, at room temperature, using an etched silicon cantilever with a resonance frequency of ~ 300 kHz, a spring constant of $\sim 42\text{N/m}$ and a tip radius of ~ 10 nm.

B- *In vivo* pull-down TIMING 2-3 d

- i) Take 500 mL of induced cells from Step 26 and centrifuge at 5,000g for 5 minutes at room temperature.
- ii) Remove the supernatant and place the cells on ice. Perform cell lysis by adding 200 μ L of ice cold Bacterial Protein Extraction Reagent and gently pipette up and down until the cell suspension is homogeneous. Leave on ice for 5 minutes.
- iii) Add 1nM of a DNA-biotinylated capture probe designed to specifically hybridize with the scaffold (see Experimental Design) and allow for interaction with the RNA scaffold at room temperature for 5 minutes.
- iv) Add 1mg of Streptavidin coated Dynabeads M-280 and allow for interaction with the sample for 5 minutes on ice.
- v) Place the Lysate on a magnet and rinse the beads twice with ice-cold assembly buffer made fresh
- vi) Resuspend the beads in 10mL of assembly buffer complemented with 10nM of a DNA probe fully complementary to the capture probe. Incubate at 37°C for 5 minutes with intermittent gentle hand shaking.
- vii) Run on a 6% TBE-urea gel and visualize the RNA scaffold with ethidium bromide staining, according to the manufacturer's instructions. The purified scaffold should be visible as a clear single band.

C- *In vivo* expression analysis – qRT-PCR TIMING 2-3 d

- i) Purify RNA from 1 mL of induced cells from Step 26 and from 1 mL of non-induced cells from Step using the Norgen Kit, according to the manufacturer's instructions.
- ii) Prepare cDNA from 0.5 mg of purified total RNA using NCode™ VILO™ according to manufacturer's instructions. cDNA may be kept in the provided buffer at -20°C.
- iii) To a 4.5 mL aliquot of 10 fold-diluted cDNA in ddH₂O, add 3 mL of a 2 nM solution of a specific forward primer, the provided poly-A annealing reverse primer according to the manufacturer's instructions and 7.5 mL of SYBR Green Supermix (Ambion)
- iv) Set up the PCR reaction using SYBR Green Supermix according to the manufacturer's instructions and perform the quantitative PCR on an appropriate thermocycler (e.g., Eppendorf Mastercycler ep realplex) with the following program: 50°C for 2 minutes then 95°C for 5 minutes to activate the enzyme. The PCR cycles are then as follows: 15 seconds at 95°C, 30 seconds at 60°C and 30 seconds at 72°C, repeated 45 times.
- v) Calculate the relative concentration of the RNA scaffold by using a stable mRNA internal reference (e.g., GapA or MreB for *E. coli*): assess total RNA samples from Step 27Ci for the concentration of the internal reference RNA. Then calculate the relative concentration of each RNA sample (induced vs. non induced from Step 27Ci), relative to the mRNA internal reference using the qPCR software Biogazelle Plus according to the manufacturer's instructions.

NB: Good quality, stable housekeeping genes are essential for high quality RNA quantification by qRT-PCR⁴⁶. We recommend designing primer pairs for at least three candidate housekeeping genes and assessing their stability in the assay conditions using dedicated software (e.g., GeNorm).

6. Targeting Proteins onto the RNA Scaffold TIMING 2–3d

28) Make fusion proteins between the chosen aptamer proteins and the proteins to be scaffolded according to standard protocols (e.g., using the biobrick cloning system⁴⁶).

29) Clone the genes into compatible expression plasmids (e.g., Duet vector EMD Chemicals) with appropriate restriction enzymes according to standard protocols.

30) Co-transform the RNA scaffold plasmid from Step 23 alongside the protein coding plasmid from Step 29 into competent bacterial cells (e.g., heat-shock competent *E. coli* BL21 DE2-star). Appropriate negative controls here are RNA scaffolds with mutated aptamers designed to prevent protein binding (e.g., modifying PP7 or MS2 stem loops to a stretch of T and A bases). Plate the transformed bacteria, pick colonies and identify correct clones according to standard protocols. Correct clones can be kept as glycerol stocks at -80°C and aliquots used for further experiments.

7. Timing

- Steps 1-8, RNA scaffold design: 2h
- Steps 9-11, RNA scaffold optimization: 1h.
- Steps 12-24, Cloning the designed RNA scaffold into an expression system: 12-15d
(Step 13 takes most of this time, typically 5-10 business days)
- Steps 25-26, Induction of RNA scaffold expression: 1-2d
- Step 27, Expression analysis: 2-3d (depending on which option A-C is chosen)
- Steps 28-30, Targeting proteins onto the RNA scaffold: 2-3d (typically Steps 28 and 29 can be done while waiting for the scaffold synthesis in Step 13)

8. Troubleshooting

| Step | Problem | Possible Reason | Solution |
|------|--|---|--|
| 8 | RNA Designer cannot find a sequence folding into the desired secondary structure | Sequence constraints are too restrictive. | Extend the double stranded region of the desired secondary structure and add more complementary bases using RNA designer to promote its formation. |

| Step | Problem | Possible Reason | Solution |
|------|---|---|---|
| 8 | RNA Designer does not allow the design of symmetric sequence ((((0)))) | Function not supported by the software. | As exemplified in Figure 26B, trick RNA designer by designing a fake hairpin: add a few unpaired bases in between the symmetrical sequences and remove them after the results. ((((...)))) Alternatively you can use the Nupack Design package. |
| 10 | Prediction of several alternative structures | The program computes not only the most favorable structure, but also less favorable structures with relatively lower free energy. | Compare the different free energies. If necessary, pick another RNA designer result. |
| 13 | Gene encoding RNA scaffold cannot be synthesized | Long sequence with complicated secondary structures and repeats | Divide the sequence into two genes. In the case of extended discrete scaffolds a solution is to synthesize the discrete RNA module with biobrick restriction sites and serially clone it until desired scaffold size is reached (Figure 25B). |

| Step | Problem | Possible Reason | Solution |
|--------|---------------------------|---|---|
| 20 | Low DNA Recovery | Fuzzy DNA band on the agarose gel. | Increase the agarose concentration up to 3% |
| 24 | No correct clone | Low ligation efficiency. | To ensure maximum ligation efficiency, dephosphorylate the vector and adjust the molar ratio between vector and insert to approximately 1:3. |
| 27Bvii | No RNA scaffold recovered | Low Pull-down efficiency RNase contamination | Screen for a number of DNA capture probes Make sure to use standard practice when working with RNA. Use gloves, filter tips, and a decontaminated workspace. |
| 27Cv | Poor qPCR readings | Non-specific primers | Check the primer melting curve. Design and try a new pair of primers. |

Figure 27: Troubleshooting Table.

9. Anticipated Results

The described protocol results in a number of RNA scaffolds cloned and expressed *in vivo*, and capable of binding target proteins with prominent affinity and selectivity. Both polymerizing and discrete scaffolds can be designed, though as discussed above, characterization is slightly more complex for polymerizing scaffolds as assembly efficiency should be assessed isothermally *in vitro* first (Step 27A). *In vitro* assembling candidates can be then be evaluated for *in vivo* assembly and scaffolding efficiency by purifying them from cells (Step 27B). Polymerizing RNA scaffolds can be expected to reach tens of nanometers in size and gather hundreds of proteins. Discrete scaffolds can also gather hundreds of proteins depending on design choice (i.e. the number of aptamers). Aptamer occupancy was estimated to be of at least 70% for a repeat of 96 MS2 aptamers by Cox et al (Golding et al. 2005). It is expected to be equal or higher here due to the complementary 3' region stabilizing the aptamers (see Step 3Aiii). Finally, depending on the expression system, fully induced cells can be expected to produce tens of thousands of RNA molecules without excessive metabolic burden (Delebecque et al. 2012).



VII - METHODS

1. Introduction

Many of the challenges and associated tricks involved in designing, using and characterizing RNA scaffolds are covered in detail in Chapter V and especially in Chapter VI. We provide here further details and highlight some of the very specific methods we had to develop to work at the boundary in between Synthetic Biology and RNA Nanotechnology.

2. *In silico* characterization of RNA structures

1. RNA folding

RNA molecules *in vivo* adopt specific secondary structures essential to their many biological functions. The secondary structure is the set of base pairs formed when a RNA strand folds on itself with each base involved in, at most, one Watson-Crick pair.

Interestingly, RNA is a remarkably dynamic molecule *in vivo*, more so than *in vitro*. *In vivo*, RNA molecules have the ability to assemble into very precise structures and undergo transitions from one defined structure to another on a biological time scale (Uhlenbeck 1995; Woodson 2000). We make good use of this structural flexibility to implement a kinetic assembly pathway (Yin et al. 2008) in our polymerizing scaffold assembly to enable the assembly to occur isothermally as detailed earlier.

Two main mechanisms seem to control efficient RNA folding *in vivo*. First, a number of proteins with RNA chaperone activity have been characterized *in vitro* (Coetzee et al. 1994; SEMRAD 2004) and might facilitate thermodynamic equilibration by lowering free energy barriers between folds (Schroeder et al. 2002). Second, the sequence and timing with which regions of nascent RNA become available during transcription dictate folding pathways. RNA secondary structure folding happens on a microsecond time scale (Gralla & Crothers 1973; Pörschke 1974), much faster than *E. coli* RNA polymerase transcription rate, which is about 80 nucleotides per second (Vogel & Jensen 1994).

This is a very important and often disregarded point and a major difference between isothermal *in vivo* RNA folding and temperature controlled *in vitro* folding. The sequentiality of RNA transcription favors correct folding of engineered RNA structures and limits misfolds of RNAs with large extended repeats such as discrete RNA scaffolds.

2. Determination of RNA Structures

Determining RNA structure is easier than the typical protein folding problem. First, it takes 20 amino acids with very different structures to build a protein with a large number of different possible interactions (hydrophilic, hydrophobic, polar, etc...). Then, the existence of secondary structural elements (β -sheets, α -helices, β -turns, coils, etc...) is highly contextual and may not form *in vitro*. Overall, tertiary structures are dependent on secondary elements and vice-versa, thus the energetic contributions of each element are not separable which makes it very difficult to predict three-dimensional structures of a protein from its sequence (Hartl & Hayer-Hartl 2009).

Determining RNA structures is much simpler (Jr & Bustamante 1999). Their fundamental building blocks are very much alike - only four nucleotides each made of a base, a ribose and a phosphate. Electrostatic interactions are well understood theoretically and experimentally and there are only four basic secondary structure elements in RNA (helices, bulges, junctions and loops). Helices are all A-form Watson-Crick base pairing helices and the other structural elements are non-Watson Crick regions always terminated by helices. Finally, secondary elements are more stable than

tertiary interactions, they exist and are stable by themselves which makes it relatively easy for software to compute structures (Zuker 2003).

RNA folding software algorithms are based on standard free energy models that provide a measure of the thermodynamic stability of secondary structures, as a function of sequence, temperature, and salt concentration (Jr & Bustamante 1999). We used a number of different tools as described earlier, including RNA designer (Andronescu et al. 2003), NuPack (Zadeh et al. 2010), and mFold (Zuker 2003). Interestingly, all of these software were developed for *in vitro* work and disregard the sequentiality of transcription and of *in vivo* RNA folding. Thus, when interpreting folding results, results with similar free energy can be sorted and some disregarded based on this non sequential folding criterium.

3. Challenges in designing RNA Scaffolds

1. Discrete vs. polymerizing scaffolds

We designed and used two kinds of scaffolds, discrete (Figure 24A and B in Chapter VI) and polymerizing (Fig 24C in Chapter VI).

Polymerizing scaffolds require a more complex approach to both design and characterization but allow for more complex architecture and RNA-RNA interactions to be studied *in vivo*. Discrete scaffolds on the other hand are easier to engineer and can also be used in mRNA expression studies. Discrete scaffolds are interesting molecular tools for metabolic engineering as number of docking sites, inter-enzymatic distance and complex size can be precisely controlled.

2. Strategies for scaffold polymerization

We provide here complementary information about scaffold design. The sequence of the RNA modules were designed using RNA Designer (Andronescu et al. 2003), BLAST (Altschul et al. 1990), RBS calculator (Salis et al. 2009), and mFold (Zuker 2003) (Figure 28). The approach involves the initial use of RNA Designer to generate a list of RNA sequences that could potentially generate the desired secondary motifs with the sequence constraints provided (i.e. aptamer sequence, symmetry, and folding). 'dist' is a parameter provided by RNA Designer, and refers to how close we are to the desired correctly-folded structure (i.e. our target). This list of sequence outputs was then screened against *E. coli*'s genome using BLAST, and narrowed to eliminate any sequences of high similarity. We then used the RBS calculator to further narrow the list of hits to include sequences with the lowest probability for ribosomal binding. If needed, the selected output sequences were then verified using mFold and this process was iterated a number of times until the sequence output provided the RNA module of desired structure/function with the highest probability to fold correctly during sequential translation.

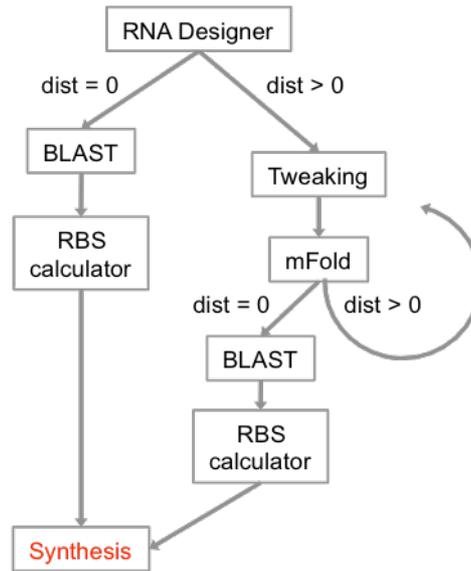


Figure 28: Scheme illustrating the approach used to design the RNA modules.

Discrete (D_0), one-dimensional (D_1) and two-dimensional (D_2) RNA assemblies were constructed *in vivo*. D_0 was constructed from a single RNA strand d_0 , designed to fold into a duplex with protruding PP7 and MS2 binding domains. D_1 was constructed from a single RNA strand d_1 , designed to initially self-assemble into a tile (d_{1-1}) capable of polymerizing into an extended assembly (d_{1-2}), which spontaneously folds into the one-dimensional RNA nanotube D_1 , with protruding PP7 and MS2 binding domains. d_1 possess a trigger (τ_8 bases) and polymerization (τ_{10} bases) domains that are palindromic. Standard free energy calculations reveal that d_1 would preferentially collapse into an inert motif incapable of assembling into tiles or extended assemblies. Although heating and cooling could achieve access to the thermodynamically most stable extended assemblies, doing so *in vivo* is not feasible. To overcome this problem,

we modified the RNA module d_1 so that we would be able to program the kinetics of the assembly pathway, and to generate extended assemblies isothermally. We prevented d_1 's trigger domain from collapsing, by locking it into an inert motif. This involved using part of the polymerization domain to fold back on itself, and cap part of the trigger domain. By doing so, we favored the nucleation and formation of d_{1-1} and its extension into d_{1-2} and assembly to D_1 .

This process is exemplified using d_1 . In the following scheme we present scenarios in which a d_1 analogue is constructed that is not capable of capping itself, as well as a d_1 analogue that is capable of capping itself. In the following scheme (Figure 29), the convention of RNA designer, is such that “()” represents complementary bases while “...” denotes non-interacting bases. “...” denotes regions included to “trick” RNA designer into designing palindromes (referred to as ‘extra sequences’ in Figure 29). The red sequences correspond to polymerization domains. The blue sequences correspond to the trigger domains.

D_2 was constructed from two RNA strands d_2' and d_2'' . d_2' possesses the PP7 binding domain, while d_2'' possess the MS2 binding domain. The assembly of D_2 was similarly programmed using locked versus open d_2' and d_2'' confirmations. d_2' self-assembles into the pro-tile d_{2-1} , which interacts with d_2'' to generate the fully functioning tile d_{2-2} . d_{2-2} then self-assembles into the extended RNA scaffold D_2 . The minimized structures of d_0 , d_1 , d_2' , and d_2'' is illustrated in Figure 30, and a schematic of each assembled module is illustrated in Figure 31.

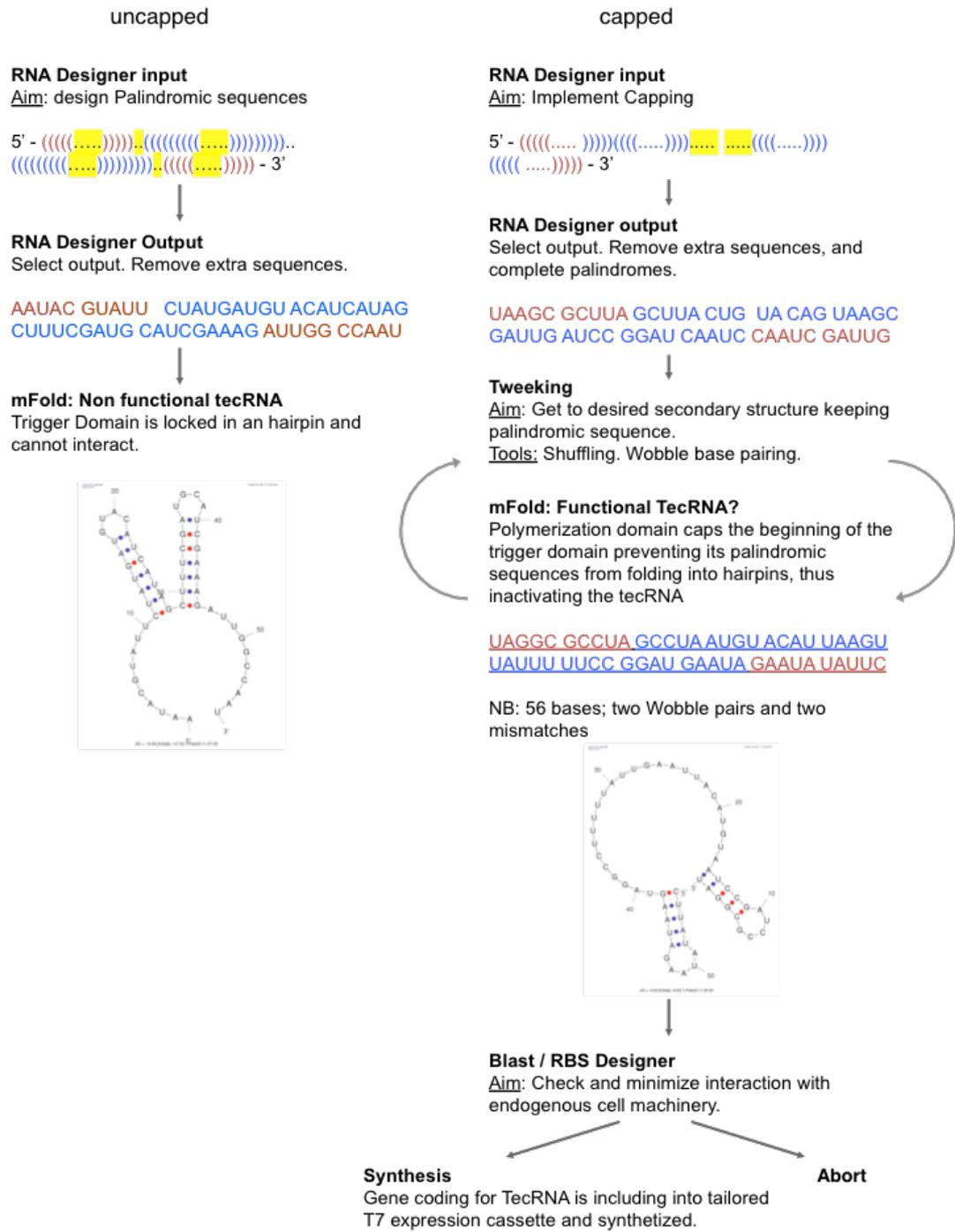


Figure 29: Schematic representation of the steps involved in designing the self-assembling RNA module dr.

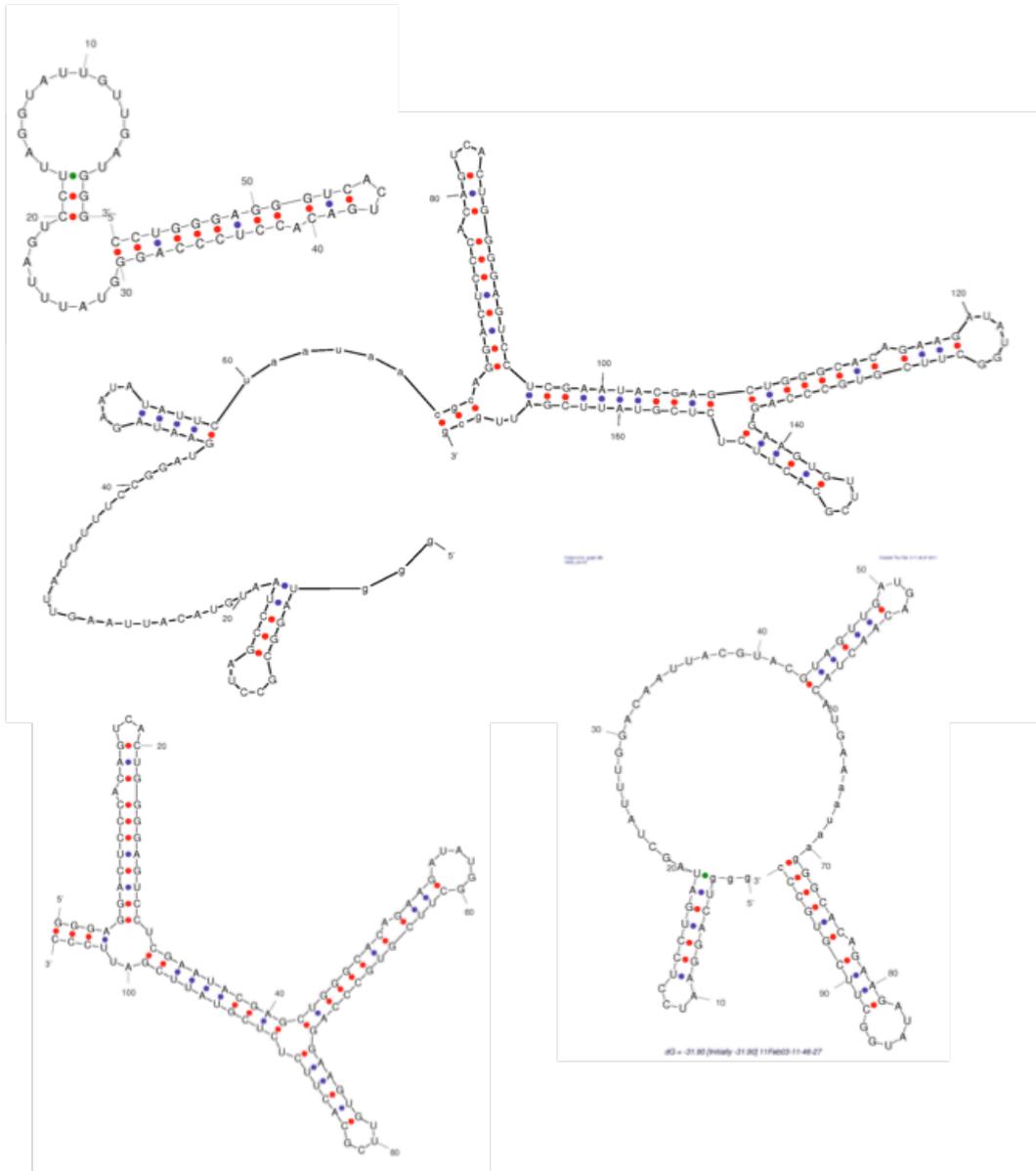


Figure 30: Secondary structures of the RNA building blocks do, di, d2', and d2''.

4. Probing the stability of RNA scaffolds

1. Considerations about RNA half life in *E.coli*

RNA strands have a half-life that ranges between 40 seconds and 60 minutes (Selinger 2003). A number of parameters can be taken into account when designing our RNA modules to minimize degradation rates. The degradosome (degradation complex) contains RNaseE, PNPase, RhlB and an enolase, and is primarily responsible for the degradation of RNA (Richards et al. 2008). Studies reveal that stem loops play a key role in regulating the half-life of RNA, and they could be used to inhibit initiation of the process that results in degradation (Coburn & Mackie 1996). Additionally, the RNaseE complexes require a minimum of 4 single-stranded bases (5') for binding, while RNaseII and PNPase require a minimum of 6-10 single-stranded bases (3') for binding (Richards et al. 2008). We took both of these points into consideration when designing our RNA scaffolds, and insured that no RNA module had any single-stranded overhangs in either direction.

2. Developing a qPCR strategy for short and highly complex ncRNAs

To best assess the relative stability of our RNA scaffolds *in vivo*, we monitored the total intracellular RNA levels as a function of time and assembly. This is challenging as our RNA modules are small and highly complex molecules to PCR. To overcome this problem, we adapted a qPCR kit developed for miRNA. The main trick relies on using a universal sequence present in all of our scaffold, the rho-independent T7 terminator,

to bind a forward primer and adding a poly-A tail during retro-transcription to provide ground for a reverse primer to bind.

Briefly, cultures were grown to mid log-phase, induced with 0.1 mM IPTG, and sampled at 0, 20, 40, 60, and 70 minutes. RNA was purified using Norgen Total RNA kits. cDNA was prepared using NCode™ VILO™ miRNA cDNA synthesis kits, which was selected because it allowed for efficient cDNA synthesis from short RNA strands by incorporating a polyA tail. 0.5 µg of total RNA was elongated at 37°C (60 min) and 55°C (30 min), and terminated at 95°C (5 min). RT-PCR was performed using 7.5 µL of SYBR Green Supermix (Ambion), 4.5 µL of 10 fold-diluted cDNA and 3 µL of a 2 nM solution of forward (T7RNAF) and reverse primers (polyARNAR). The incorporation of a polyA tail into all of our constructs during cDNA synthesis allows for the use of the same reverse primer. The use of the T7 terminator, which is also present in all of our RNA constructs, also allows for the use of the same forward primer. We were thus able to use the same forward/reverse primer-set when analyzing any of our RNA constructs, which enables similar RT-PCR amplification efficiencies. Each reaction was performed in triplicate. The relative amount of RNA was normalized against GapA, which was selected as an internal control for accurate profiling of our RNA modules using geNorm. gapA mRNA consistently passed geNorm's gene-stability tests against Mdhe, and has been successfully used in other studies as a control. The primer set used was gapAF and gapAR. The concentration of each RNA, relative to the gapA mRNA internal reference, was calculated using the qPCR software Biogazelle Plus. Quantitative PCR and data collection was performed in an Eppendorf Mastercycler ep realplex thermal cycler with cycles as follows (Figure 32): reactions were initially held

at 50°C for 2 minutes then 95°C for 5 min to activate the enzyme. PCR cycles were 15 sec at 95°C, 30 sec at 60°C and 30 sec at 72°C. PCR product purity was systematically verified by melt curve analysis. A standard curves was obtained following the serial dilution of *in vitro* transcribed Do RNA modules using an Ambion Megascript Kit, followed by its quantification using a Nanodrop 1000 spectrophotometer and conversion into cDNA and analysis using RT-PCR as described above.

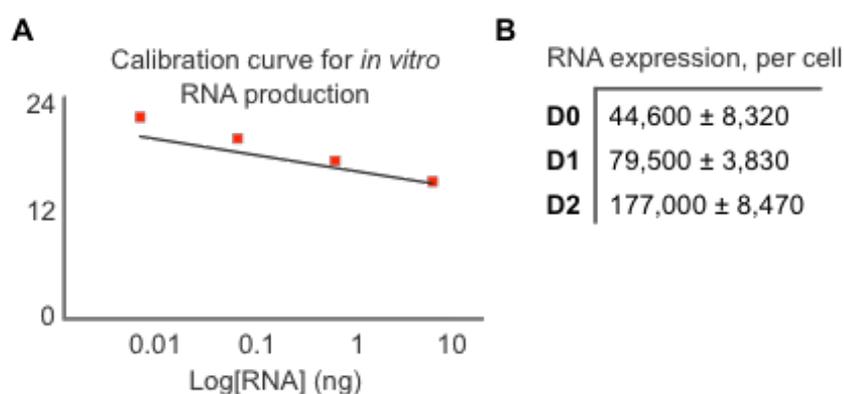


Figure 32: qRT-PCR of RNA modules. (A) A calibration curve was used to determine the (B) amount of RNA produced in cells expressing D0, D1 and D2.

5. Purifying RNA scaffolds

1. *In vitro* transcription

Our expression system as described in Chapter V and Chapter VI is based on T7 Duet vectors and was designed to facilitate the *in vitro* expression and characterization of RNA scaffolds.

In vitro transcription of T7 driven gene is easily realized with commercially available kits. We used a MEGAscript T7 kit (Applied Biosystems) according to the manufacturer's recommendations. 500 ng of plasmid, containing our RNA was incubated overnight (16 hours) at 37°C. RNA was purified using a Norgen total RNA purification kit, according to the manufactures instructions. Total RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2. Developing a pull-down protocol for *in vivo* purification

We developed a new DNA-based purification (DP) approach for isolating our *in vivo* RNA assemblies. The real challenge of purification here is to do a pull-down without denaturing the sensitive RNA structures. Hence, no cross-linking and harsh solvents could be used. The approach developed here uses DNA probes to pull down the RNA scaffolds. These probes are designed so that they can be un-zipped from their target by fully complementary DNA probes hence releasing the RNA.

Briefly, scaffold capturing was conducted using a capture probe (DPCP), while scaffold release was conducted using a release strand (DPR). DPCP is a biotinylated DNA strand, purchased from IDT. The sequence of DPCP is divided into a 10 base long region that binds the T7 terminator present in any of our RNA modules, and an additional region allowing for its controlled release using DPR. The T_m of the T7 binding domain is 21°C, while the T_m of the entire releasing probe DPCP is 34°C. The experiment involved taking 500 μ L of in-

duced mid-log phase cells, and pelleting them at 5,000g for 5 minutes. The cells were then placed on ice and lysed using 200 μ L of ice cold Bacterial Protein Extraction Reagent (B-PER; Pierce). 1 nM of DPCP was then added to the cell lysate, and allowed to interact at room temperature with the RNA assemblies for 5 minutes. 1 mg of Dynabeads M-280 Streptavidin (Invitrogen) with the binding capacity of about 650 to 900 pmol of free biotin was added to the lysate, and allowed to interact for 5 minutes on ice. The lysate was then placed on a magnet, and the beads were rinsed twice with a buffer composed of 150 mM KCl, 0.1 mM dithiothreitol, 0.1 mM EDTA, 10 mM Tris, pH 7.4. Finally, the beads were resuspended in 10 μ L of the same buffer, complemented with 10 nM of the release strand DPR, and put at 37°C for 5 minutes with gentle shaking. The release strand is fully complementary to the capture probe DPCP. It is expected to first bind to the 10 base overhang of DPCP, followed by full binding, causing strand displacement and release of the captured RNA assemblies.

This approach was tested using a model system of three strands that possess a capture probe DPCP (20 bases), a substrate DPS (40 bases), and a release strand of different length DPR (30 bases) (Figure 18D, Chapter V).

6. Imaging the Scaffolds

1. Using fluorophores

The GFP was split according to a previously reported method by Valencia-Burton (Valencia-Burton et al. 2007). Briefly, we PCR-amplified fragments A

and B (FA and FB) from pEGFP (Clontech), with BB primers (FAF/FAR and FBF/FBR). We then constructed the full combinatorial library of chimeric genes of the two halves (FA or FB) with aptamer proteins (MS2 or PP7) separated by a linker (10aa glycine serine "2x" or 2aa cloning scar sites)

For expression, PP7 containing chimeras were cloned with NotI-Spe into MCS1 of the pCDFDuet-BB vector, and MS2 containing chimeras with EcoRI-PstI into MSC2 of the pColaDuet-BB vector.

Initial tests were done with scaffold D0. Briefly, cells were induced at mid-log phase using 1 M IPTG, for two hours, placed between a glass slide and 2% agarose pads, and analyzed using a Nikon TE-2000 microscope (100X, 1.4 numerical aperture objective, ORCA-ER charge-coupled device camera, FITC channel).

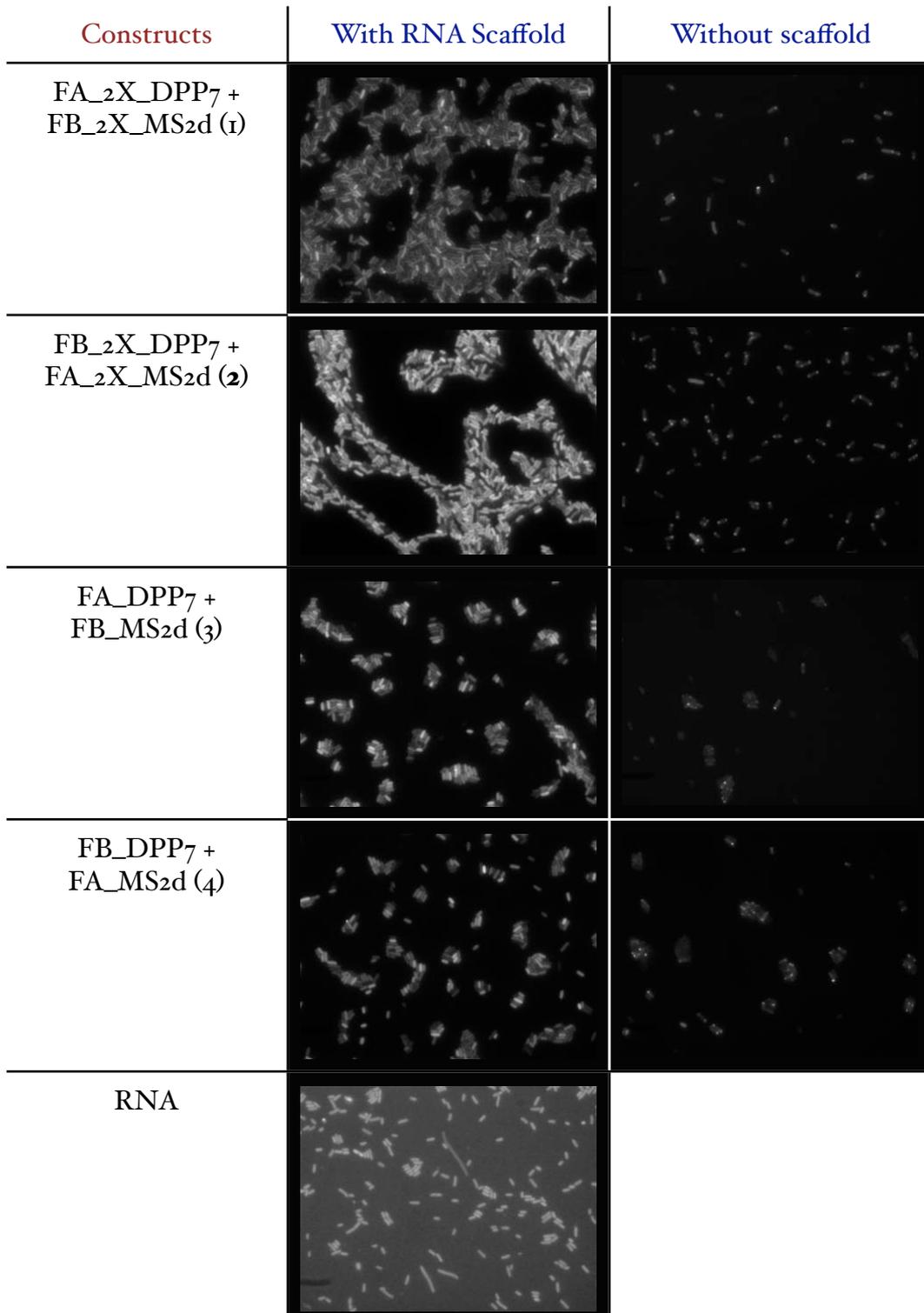


Figure 33: Initial experiments of RNA scaffold assisted fluorescence complementation.

A solubilization and complementation effect induced by RNA scaffold expression was observed. The presence of a longer linker and protein order onto the scaffold influences complementation and fluorescence output. For later experiment we choose the best performing FB_2X_DPP7 / FA_2X_MS2d pair, reduced induction time to 1 hour and lowered IPTG concentration to 0.2mM.

2. Using Electron Microscopy

In-situ visualization of RNA structures by transmission electron microscopy (TEM) is a real challenge and a technique had yet to be developed. When it comes to proteins, contrast is usually provided by post-embedding immunogold labeling. While, this could have been an indirect way of observing the structures via immunodetection of bound aptamer proteins, a number of drawbacks to this solution do exist. It is a rather indirect way at observing the structures and it lacks definition as only a fraction of the target present at the very surface of a micro-section will bind the antibody and be detected (Sosinsky et al. 2007). A genetically encoded tag erases all of these problems and we used the metal-binding proteins methalothioines fused to aptamers to visualize our RNA structures in-situ.

To visualize our RNA assemblies in cells using TEM, we co-expressed D1 and D2 with the metal-binding protein tag PAu. PAu possesses cysteine residues that are capable of binding several gold atoms by a reaction that is akin to that used in gold nanocluster formation (Diestra et al. 2009). The gene coding for

the metal binding domain of PAu (i.e. the MT domain) was PCR amplified from a plasmid generously provided by C. Risco (Diestra et al. 2009) with the primer set MTF / MTR . The purified and biobricked product was cloned with PP7 and a (Gly₄Ser)₂ linker into pCola-BB yielding plasmid pCJDMT. Log-phase cells expressing either D₁ or D₂ were grown in LB media and induced with 0.1mM IPTG supplemented with 10 mM AuCl₂. After one hour, cells were pelleted, washed twice with ice-cold PBS, and analyzed by TEM. TEM preparation involved freezing cells using a Leica high-pressure freezer, followed by freeze substitution in 0.2 % glutaraldehyde and 0.1% uranyl, in acetone, for 72 hours at -90° C (Leica AFS). The temperature was increased to room temperature over 22 hours and the samples were embedded in LR White (EMS). 60 nm sections were cut using a Reichert Ultracut-S microtome and imaged using a Tecnai™ G² Spirit BioTWIN transmission electron microscope with an AMT 2k CCD camera.

Gold elemental mappings provides a way to resolve gold elemental constitution of nanoparticles. Gold elemental mapping enabled us to resolve D₂ structures and confirm that indeed the observed structures co-localize with gold. Unfortunately, high quality elemental maps were not obtained for D₁ due to the resolution limit of the technique. Gold elemental maps were calculated as jump ratios of pre and post O_{2,3} edge images, taken with the 8 eV slit centered on 48 and 64 eV, respectively. Images were acquired on a Cs-corrected MC Zeiss Libra 200, tuned to 80 keV at the Center for Nanoscale Systems (CNS) at Har-

vard University. These images were acquired with the generous help of Dr. Mike Strauss.

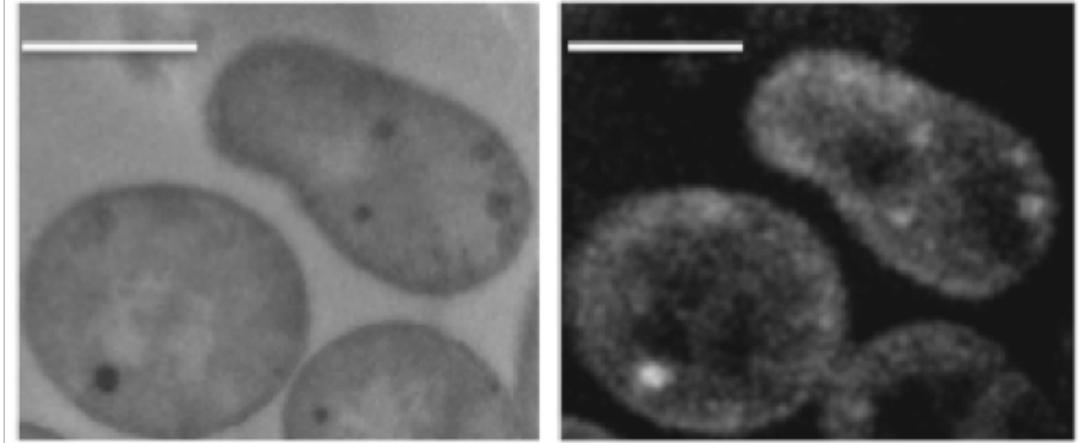


Figure 34: Gold elemental mapping of RNA scaffold *in vivo*. D2 Scaffolds colocalize with gold (Left) - zero-loss image (Right). (Bar corresponds to 0.5 μ m)



VIII - DISCUSSION

1. Small synthetic ncRNA as tools to study protein-protein interactions

Understanding the effect of spatial organization in multi-enzymatic systems is fundamentally interesting but also important from a synthetic biology stand-point. Engineers and biologist are striving for solutions to increase yields of economically important biochemical reactions *in vivo*. This is where we would like to start the discussion with some preliminary work we have been doing. We first started by developing a new reporting system for *in vivo* scaffolding and then use modified versions of our simple discrete RNA scaffold to study enzyme position, orientation and complex size *in vivo*.

1. Building a new *in vivo* scaffolding reporter system

Bacterial Luciferase Pathway - The bacterial lux system is an interesting reporter system widely used in molecular biology, especially as a molecular reporter system for gene expression studies or as a biosensor (Szittner et al. 2003). The bacterial luciferase operon comprises up to seven different genes, but luminescence can be obtained by only expressing the genes coding for the luciferase itself, *luxAB*, and with the exogenous addition of a long-chain aldehyde such as decanal (Blouin et al. 1996; Tu & Hastings 2003; Becvar & Hastings 1975).

In this light-emitting reaction, flavin mononucleotide (FMN) is reduced using reduced nicotinamide adenine dinucleotide (NADH) as an electron donor by the NAD-FMN oxido-reductase enzyme (NFOR). FMNH₂ then binds to the

luciferase and the luciferase oxidizes decanal emitting in the process a blue-green light (Figure 35A).

The kinetics of light emission follows the response profile shown in Figure xx B. There are two distinct phases. The primary phase corresponds to the initial peak intensity when decanal is injected (time $t=0$). It requires the presence of FMN and some incubation time to allow the formation of the luciferase-FMNHOOH complex which in turn quickly reacts upon decanal injection. The secondary phase exhibits three distinct “subphases”: (i) the intensity increases as luciferase-FMNHOOH complexes form back; (ii) it reaches a maximum when the complex forming rate and decanal use equilibrates; and (iii) it decreases again as the decanal concentration drops further and becomes rate limiting (Blouin et al. 1996).

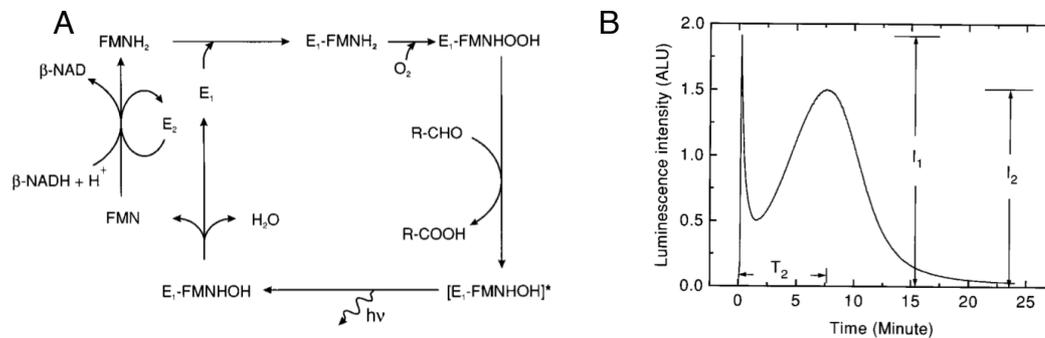


Figure 35: Simplified bacterial luciferase biochemical reaction (A) giving rise to chemoluminescence. (B) (Adapted from Blouin et al).

When decanal is supplied exogenously and dissolved O₂ is non-limiting, FMNH₂ is the only limiting substrate directly required for the luciferase-

catalyzed chemoluminescence (Blouin et al. 1996). It makes this reporting system especially interesting to study scaffolding effects, especially since FMNH₂ has a relatively short half-life and is susceptible to auto-oxidation. It has been previously used *in vitro* on DNA scaffolds (Niemeyer et al. 2002).

For all our experiments, we measure the total amount of light emitted before decanal becomes rate limiting (T₂ in Figure 35B). It captures the most interesting part of the reaction corresponding to the substrate channeling (FMNH₂) in between NFOR and Luciferase (subphase (i)).

Studying yield improvement upon scaffolding - To study the kinetic effect of scaffolding on the bacterial luciferase biochemical reaction, we decided to use our simplest, previously characterized scaffold, Do (see Chapter V and Chapter VI). It is a simple discrete RNA scaffold with a PP7 and a MS2 aptamer binding sites, effectively providing support for two different proteins. We use the same previously describe IPTG inducible duet vector system and heterogeneously express the two genes sufficient for luminescence upon decanal addition, luxAB and NFOR as fusion proteins with PP7 and MS2 aptamer proteins.

The effect of scaffolding is followed as a function of induction time, decanal concentration, enzyme position on the scaffold and linker length in between enzymes and aptamer proteins. We use a plate-reader (Tecan Infinite F500) with a one second acquisition time. Results are normalized to account for optical density.

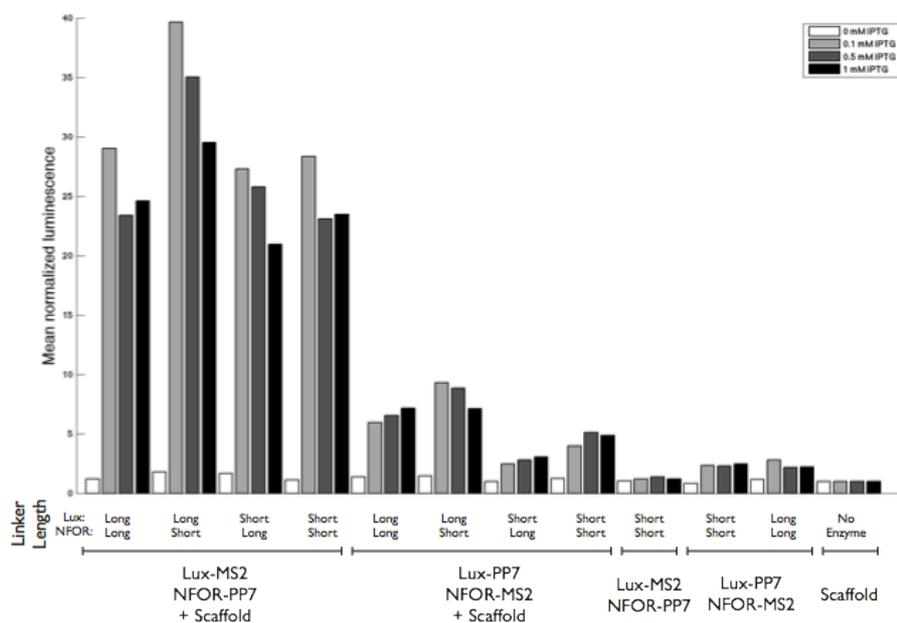


Figure 36: Bioluminescence assay of scaffolded vs. unscaffolded luciferase pathway. After one hour of induction with the specified amount of IPTG (0 to 1mM), 10 μ M decanal was added to the medium and readings were acquired for 10min. Linker length (“long” vs. “short”) and enzyme position (Lux-MS2 & NFOR-PP7 vs. Lux-PP7 & NFOR-MS2) on the scaffold was investigated.

Scaffolding clearly enhances the bacterial luciferase reaction by up to 15 fold, probably by limiting FMNH₂ auto-oxidation via substrate channeling. Interestingly, enzyme placement on the scaffold is of great importance (Lux-MS2 & NFOR-PP7 giving much higher kinetic improvements) while linker size affects the kinetics to a lesser extent. We demonstrate here that bacterial luciferase is

a powerful yet simple *in vivo* reporter system to test our scaffolds' ability to organize multi-enzyme complexes and follow kinetic improvements.

2. Controlling protein relative distance, orientation and complex sizes

Controlling protein relative distance and orientation - We made variations using the Do scaffold design as a platform. This construct has two binding sites for two different proteins (aptamer PP7 and MS2) separated by one double stranded rigid RNA helix. This region that can be finetuned to change aptamer orientation and relative enzyme distance.

The Do variant library was constructed using RNA designer (Andronescu et al. 2003) and folding was checked with NuPack (Zadeh et al. 2010) according to the previously described method (see Chapter VI).

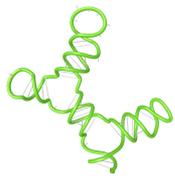
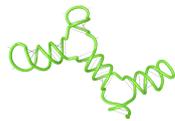
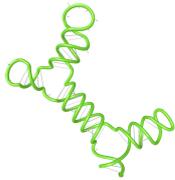
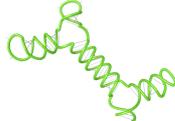
| Name | RNA Helix Number | Aptamer Distance (nm) | Aptamer Orientation | Helicity |
|-------------|------------------|-----------------------|---------------------|---|
| D0 | 1 | 3.5 | same |  |
| D1.5 | 1.5 | 5.25 | opposite |  |
| D2 | 2 | 7 | same |  |
| D2.5 | 2.5 | 8.75 | opposite |  |

Figure 37: Discrete RNA scaffold variant library. Relative aptamer distance and orientation is indicated.

Controlling complex sizes - The Do scaffold can also be used to make large constructs of controllable size, to study the effect of enzymatic complex sizes on overall yields. There are two major challenges to be addressed. First, large DNA constructs with multiple repeats are very hard to make with the current synthesis state of the art. We relied here on a PCR-based solution in which the primers used to PCR Do add complementary overhangs. Thus, after the first cycle the construct can amplify itself and giving a smear of different scaffold complex sizes that can be then cloned. The second challenge lies in the fact that RNA folding software do not account for the linearity of transcription thus wrongly predicting the most favorable folded thermodynamic state (Mahren et al. 2010).

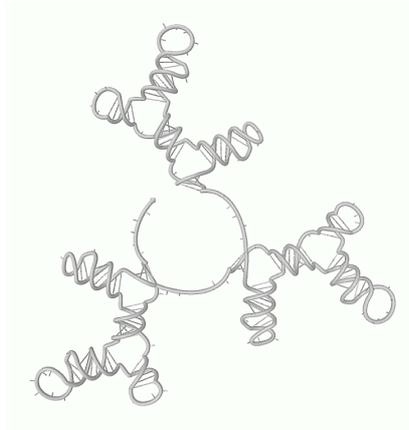


Figure 38: Computed secondary structure of a large discrete scaffold. Nupack folding of a large scaffold comprised of a triple Do repeat

We are currently testing all of these constructs with our new luciferase-based reporting system. Precisely characterizing substrate channeling as a function of

protein interaction *in vivo* will provide a new understanding on the modularity and usefulness of RNA-based scaffolding.

3. *E. coli* DsrA ncRNA – a naturally assembling RNA?

E.coli DsrA ncRNA - The ability to engineer extended functionalized RNA nanostructures *in vivo* shines a new light on natural non-coding RNAs - and whether higher order assemblies and scaffolding could be among their functions. In fact, natural RNA-RNA interactions are not known to promote extended self-assemblies. This is in sharp contrast with the many examples of large supramolecular structures made of proteins, such as microtubules or virus capsids (Yeates & Padilla 2002).

Recently, a number of reports suggest that RNA might naturally used for scaffolding – laying out a new ground for engineering possibilities (Xiao et al. 2005; Weinberg et al. 2009; Zappulla 2004). Among them, the ncRNA DsrA of *E. coli* has been shown to assemble *in vitro* and offers interesting engineering opportunities (Cayrol et al. 2009).

DsrA is a 87-nt noncoding RNA of *E. coli* with regulatory properties. Interestingly, it regulates both transcription, by overcoming transcriptional silencing by the nucleoid-associated H-NS proteins, and translation, by promoting efficient translation of the stress sigma factor, RpoS (Majdalani et al. 1998; Sledjeski & Gottesman 1995). Both the spontaneous formation of dimers and extended

filamentous DsrA assemblies has been reported *in vitro* (Cayrol et al. 2009). Self-assembly predictions, experimental evidence and detailed structural models demonstrating the formation and disruption of a hierarchy of nanostructures made by DsrA have been reported.

Three contiguous self-complementary regions along the DsrA sequence are responsible for the formation of these novel nanostructures. Interestingly, the self-assembly of DsrA nanostructures enables the collective formation of weakly paired 12-bp duplexes that overlap both with the cleavage site (5'-AAUUU-3') of a single-strand endoribonuclease (RNase E) and the functional region of DsrA, that is known to bind several target mRNAs.

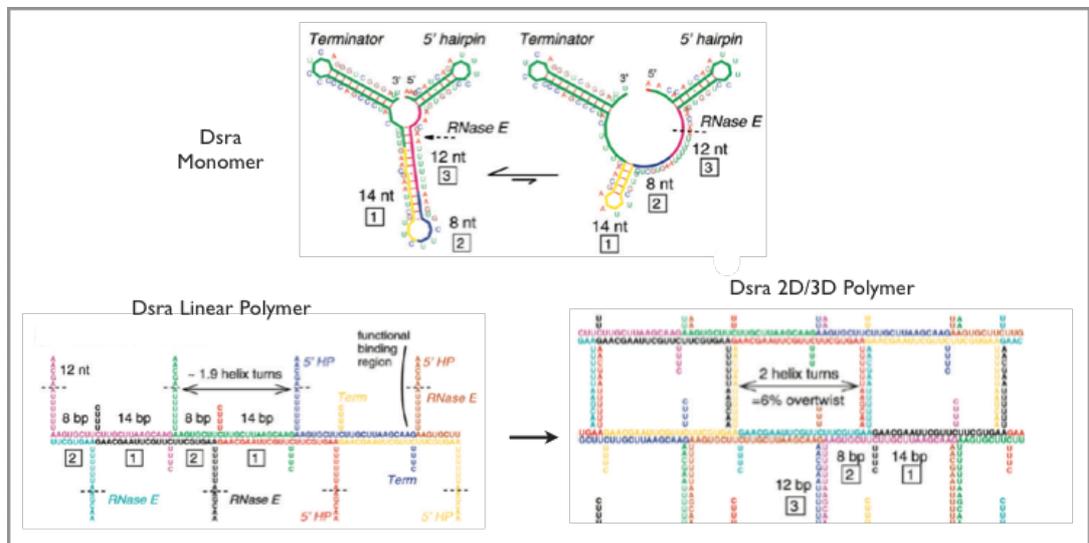


Figure 39: DsrA RNA polymers. DsrA monomers can self-polymerize *in vitro* to form 1D and 2D nanostructures. (adapted from Cayrol et al. 2009)

Constructing DsrA Analogues - DsrA structure is comprised of 3 hairpins. The 5' hairpin ("hairpin I") is the "functional" domain where the Hfq protein binds. It is not involved in the assemblies and is replaced by aptamer PP7 or MS2 to enable the targeted binding of heterologous proteins. The second hairpin ("hairpin II") is made of the sequences involved in the assembly. A 8nt stretch and a 14nt stretch are involved in the 1D polymerization while a 12nt stretch is involved in the 2D polymerization. Computer assisted forward engineering of the sequences enable the control of polymerization. The third hairpin ("hairpin III") is comprised of the transcription terminator. It is switched for a T7 terminator.

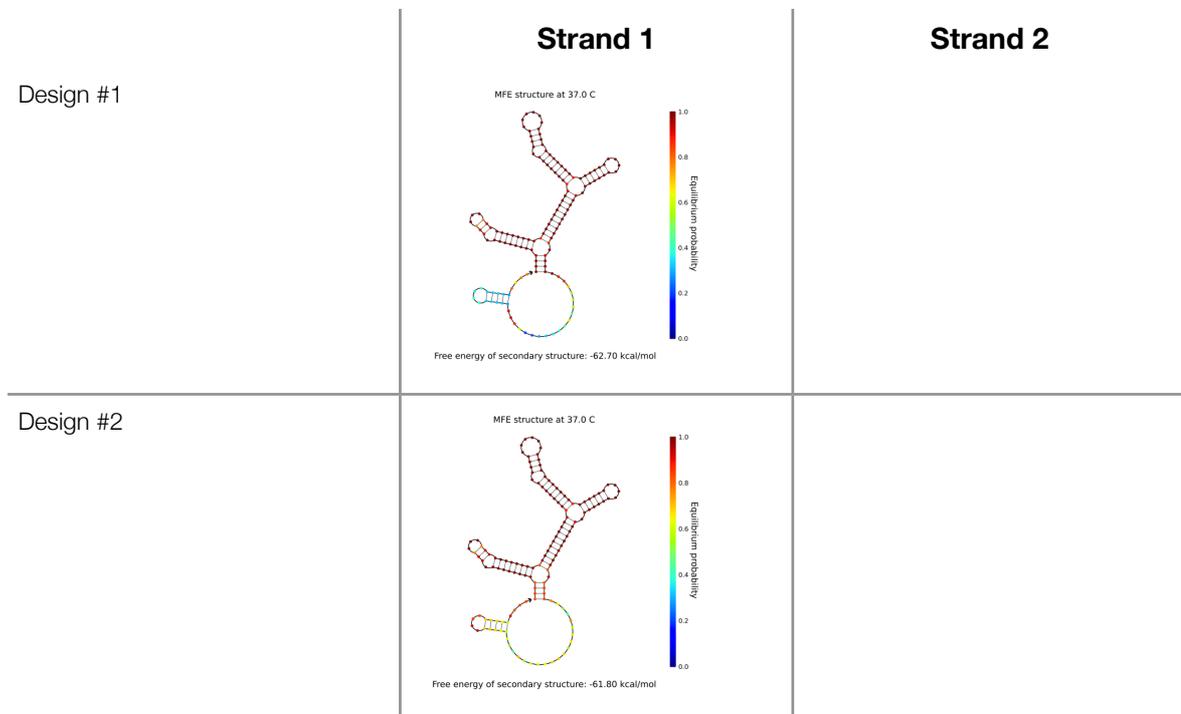
Expression of these RNA is done using a IPTG-inducible T7 promoter duet vector in which the expression cassette is switched for the RNA expression cassette as reported previously. Six design variants have been made to test 1D vs. 2D polymerization:

- #1 bears aptamers MS2 and PP7 instead of hairpin I, the polymerisation sequence of hairpin II is unchanged;
- #2 bears aptamers MS2 and PP7 instead of hairpin I, the polymerisation sequence of hairpin II is modified to strengthen 2D polymerization;
- #3 is a 2-member system where dsra_1 bears PP7 and dsra_2 bears MS2. The polymerisation sequence of hairpin II is unchanged;

- #4 is a 2-member system where dsra_1 bears PP7 and dsra_2 bears MS2. The polymerisation sequence (2) is modified to strengthen 2D polymerization;

- #5 is a 2-member system where dsra_1 bears PP7 and dsra_2 bears MS2. The polymerisation sequence of hairpin II is modified to strengthen 2D polymerization and to enforce a 1 to 1 dsra_1 / dsra_2 stoichiometry during polymerization.

- #6 is a 2-member system where dsra_1 bears PP7 and dsra_2 bears MS2. The polymerisation sequence of hairpin II is modified to prevent 2D polymerization;



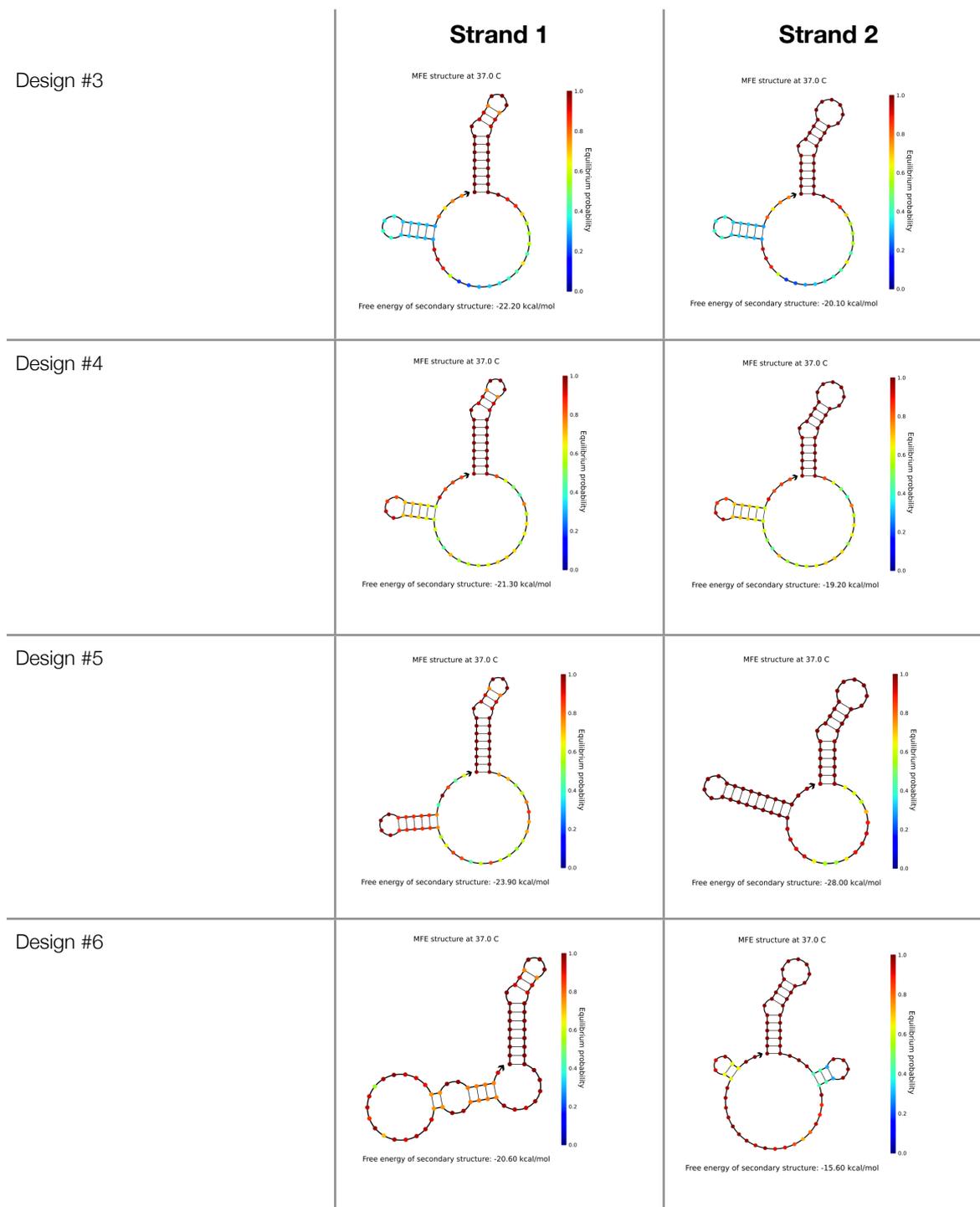


Figure 40: Table of DsrA design variants secondary structures. Color scale represents the equilibrium probability. Computed with NuPack.

This is still an ongoing process as we are currently testing all of these constructs with our new luciferase-based reporting system. *In Vivo* polymerization of DsrA, if demonstrated, will shine a new light on small bacterial RNAs where polymerization could play an essential role in genetic regulation. Upon temperature change, the DsrA transcript could depolymerize and activate RpoS.

2. *In vivo* Assembly

There are many different possible routes to further explore *in vivo* assembly of nucleic acids including using DNA, working with better software, and using directed evolution. Early on, we ruled out the use of DNA as a building material *in vivo* for a number of reasons including bio-availability. Certain types of DNA molecule however can be produced in high quantity, including plasmids (Conrado et al. 2012) and phagemid vectors (Lin et al. 2008). In the former, dsDNA sequences of classical DNA assemblies (non-polymerizing holiday junction, and cross-over motifs) were cloned into phagemid vectors and transformed into *E. coli* cells. The infection by a helper phage initiated an *in vivo* rolling circle replication of ssDNAs encoding the nanostructures. *In vivo* assembly or polymerization was not attempted, but purified ssDNA could assemble *in vitro* following a classical denaturation and cool-down protocol. Fairly large amounts of ssDNA could thus be produced by bacteria. The next question would be what for? A recent trend in DNA / RNA nanotechnology has been the development of hybrid nanostructures (Ko et al. 2010). DNA/RNA hybrid nanostructures are indeed likely to

have a synergistic potential that combines the predictability of DNA assembly and the *in vivo* functional diversity of RNA.

Gaining access to a comprehensive set of rules for *in vivo* assembly would be an enormous step forward. The development of new assembly algorithms specifically tuned for this purpose would simplify the process, increase structural complexity and improve the fidelity and yields of the assembly process. Lately, there has been a number of new RNA folding software released, and a few can compute simple RNA-RNA interactions (Zadeh et al. 2010). An interesting new approach on computing RNA folding comes from the Das lab and their online RNA folding game EteRNA. By crowdsourcing online RNA folding and testing in lab solutions gamers come up with, it aims at understanding and bridging the gap between current computational models and reality. While better folding algorithms is certainly part of the solution, it would be very useful if these software were to take into account the sequentiality of transcription and be combined with assembly algorithms.

Some responses and comments about our work inspired further thinking. One of my favorites comes from Professor Ellington, an eminent scientist in the DNA nanotechnology field. He writes in his blog:

“This is especially true given the recent glorious demonstration [...] that organized, functional RNA structures could be generated inside of cells [...]. But in the end, the cellular operating system is as foreign to many of the concepts of DNA nanotechnology as the latter is to electronics. Cells are evolutionary machines that have crafted operating systems that work without design. To now impose design is akin to what the Tea Party seeks to do with the US Constitution in ignoring hundreds of years

of court interpretations. It can be done, but it would be pretty primitive compared to where we are now.”

There might not be a comprehensive, easily accessible set of rules for *in vivo* assembly as the intracellular environment is so complex. As Professor Ellington is commenting, rational design and engineering may not be the most relevant approach to this problem as cells are evolutionary machines. A logical follow-up work would then be to explore the power of evolution to solve this problem for us and make a great variety of assembling nano-structures *in vivo*.

We envision a directed evolution approach to evolve de novo non-coding RNA sequences assemblies as new tools for systems and synthetic biology. RNA presents a great advantage over protein here as new functions can be generated through an *in vitro* selection strategy, systematic evolution of ligands by exponential enrichment (SELEX) (Ellington & Szostak 1990; Tuerk & Gold 1990). The selection would be divided into two steps, first an *in vitro* SELEX step and a second *in vivo* “validation” step that would reduce the pool potential assembling molecules to the ones doing so under physiological conditions. This second *in vivo* step is extremely important and, to the best of my knowledge, has never been implemented. Folding conditions greatly differ *in vivo* vs. *in vitro* and the last rounds of selection need to happen *in vivo* for selected RNA to perform correctly. The whole process would happen as described below:

***In vitro* SELEX** - Our starting RNA library (typically 10^{14} to 10^{15} molecules) is based on a backbone consisting of a randomized 45-nt region flanked at 5' T7 promoter and the MS2 aptamer and 3' T7 rho-independent terminator. Given that each variant in the pool is present in approximately 10^4 copies, 0.1% (i.e. 10 copies on av-

erage) will be 3' biotinylated (Pierce RNA 3' Biotinylation Kit) and bound to streptavidin-coupled magnetic beads and re-mixed with the entire library. The pulled-down recovered RNA mix will be amplified through RT-PCR and used as the starting library for the next round of selection. Iterative rounds of selection are performed, and the selection stringency and counter selections can be tailored to enrich for polymerizing RNA sequences.

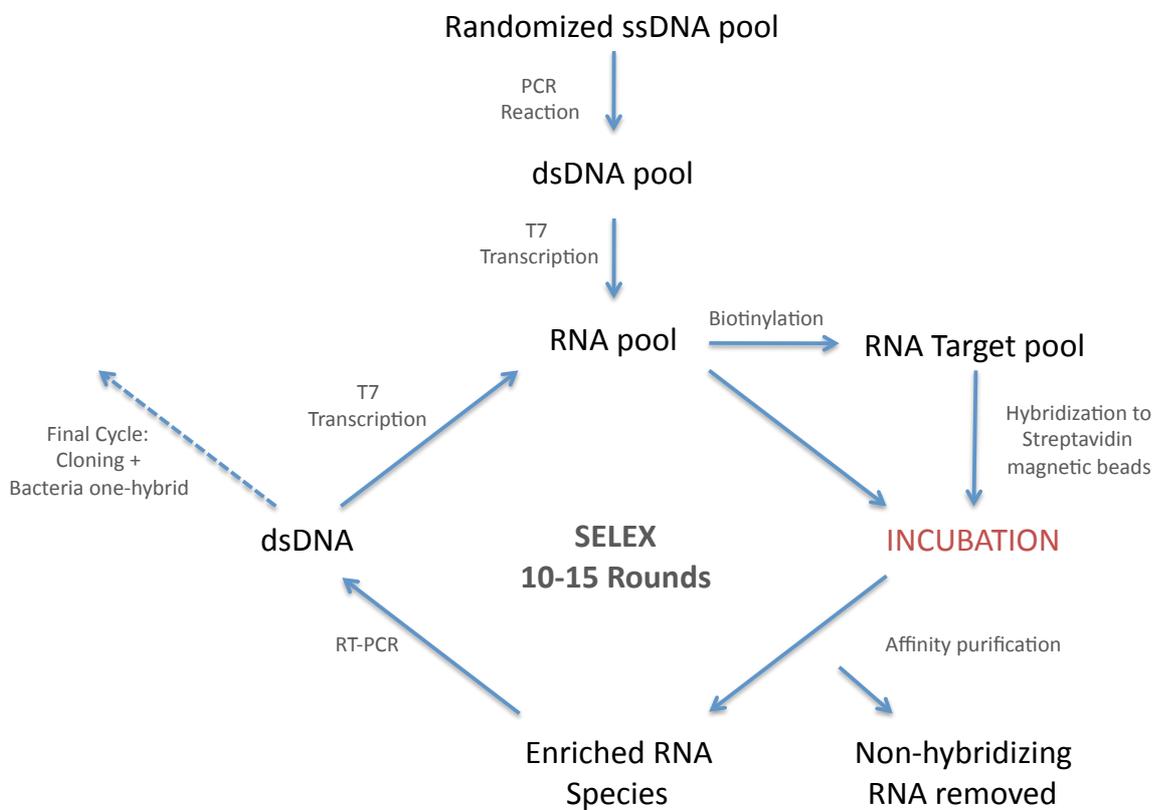


Figure 41: Evolving assembling RNA - First step. Schematic of the tailored SELEX process.

***In vivo* one-hybrid selection system** - We propose to perform a secondary screen on *in vitro*-enriched RNA component libraries by using oligomerizing RNA sequences to recover histidine auxotrophs in an adapted bacterial one-hybrid system (Meng &

Wolfe 2006). Our starting material will be the SELEX enriched library. Each selected sequence will be cloned with the original MS2 aptamer. Successful oligomerizing RNA modules will be linking a MS2-fused transcription factor (MS2-TF) together with a MS2-fused α -subunit of RNA polymerase (RpoA) to yield expression from a weak lac promoter downstream of the TF binding site thus enhancing the expression of HIS3 to relieve his3 cells from auxotrophic starvation in appropriate minimal media (i.e., absence of histidine and presence of 3-amino-triazole, a competitive inhibitor of HIS3). To further increase the stringency of our selection system, HisB, pyrF, and rpoZ genes can be knocked out thus eliminating the cell's ability to synthesize histidine, pyrF, and the omega-subunit of RNA polymerase, respectively. Finally, to further enrich for multimer assembly as opposed to simple dimerization, the distance in between the TF binding site and the reporter vector promoter can be tuned so that only increasing size of multimers will be able to close the gap.

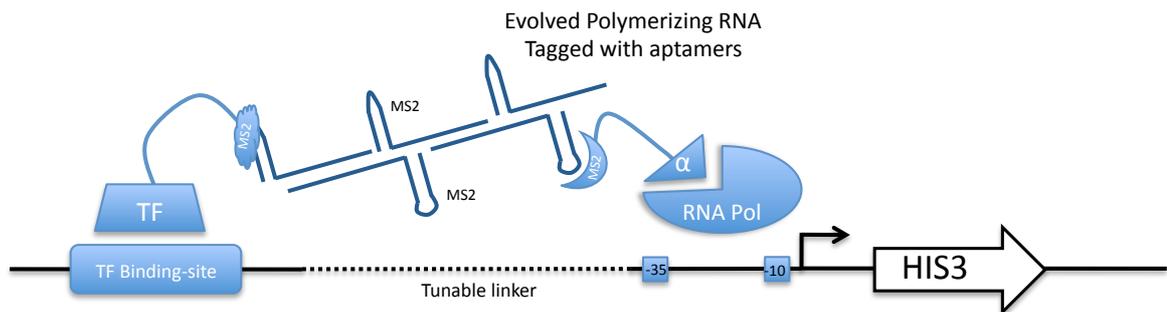


Figure 42: Evolving Assembling RNAs - Second step. Selection of *in vivo* polymerizing RNAs from initial SELEX pool via rescue of an histidine auxotrophy.

3. Dark matter RNA

Students were once taught, and most of them still probably are, that there are three kinds of RNA: messenger RNA, ribosomal RNA and transfer RNA all required for protein synthesis. While the list is much more extensive now, there is still a huge amount of transcribed genomic material, about which researchers have no idea.

This is especially true of eukaryotic organisms where more than 95% of the genome is referred to as “junk DNA” (G. S. L. A. P. Kapranov 2012) and where “dark matter RNA” makes up the majority of non-ribosomal RNA by mass (P. Kapranov et al. 2010). The function of RNA from these non-coding regions of the genome is the subject of many debates, and I would like to argue that one of the functions might well be structural. In this view, a new class of non-coding RNA could take a central role in structuring and organizing the intra-cellular machinery, especially in the nucleus.

A number of arguments support this hypothesis, specifically the abundance and conservation of intronic RNA and long non-coding RNAs (lincRNA). In eukaryotic cells, while intronic RNA have long been thought to be piece of pre-RNA junk en route to degradation (J. M. Johnson et al. 2005), this view is changing. It was recently shown that intronic RNAs represents the major component of the mammalian transcriptome (St Laurent et al. 2012) and this begs the question of what potential functions could intronic RNAs carry. An attractive answer would be scaffolding that could bridge distal DNA loci in the 3-dimensional space of the nucleus. LincRNA acting as scaffolds were recently shown to be essential to maintain pluripotency of mouse and human

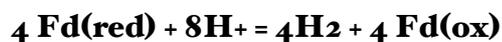
embryonic stem cells (Ng et al. 2011; Guttman et al. 2011). The molecular circuitry associated with pluripotency was long thought to be protein-centric, but these new essential players interact with chromatin modifiers and regulate gene expression in cis by controlling local chromatin structure (Dinger et al. 2008). Last but not least, the scaffolding role of RNA in eukaryotic cells is not restricted to intronic RNA and long non-coding RNAs. Yeast telomerase RNA TLC_I has been shown to act as a flexible scaffold for protein subunits (Zappulla 2004).

There is no reason to think that the structural role of non-coding RNAs would be limited to eukaryotic cells (Gottesman & Storz 2011). DsrA RNA is one example of a small bacterial RNA (sRNA) shown to polymerize *in vitro* (Cayrol et al. 2009) and thought to do so *in vivo* as well. Other examples include the maturation of RNase E complex through sRNA mediated mRNA-Hfq-RNaseE association (Morita 2005). An extremely interesting follow-up work would be to set up a screen for assembling non-coding RNAs. Briefly, total mammalian, yeast and bacterial RNA would be purified a specified amount of time after blocking transcription. Long-lived transcripts would then be retro-transcribed and cDNA would be cloned into a plasmid containing a T7 promoter, an MS2 aptamer and a rho-independent terminator. The plasmid library would then be transformed into our tailored bacterial-one hybrid strain described in Figure 42 and only bacteria with polymerizing RNA would survive. The purified plasmid can then be sequenced and the RNA identified. Such a screen would be very insightful and could potentially open our eyes to a prevalent new class of non-coding RNAs.

4. Saving the world with hydrogen?

Fermentative bio-hydrogen research has seen an explosion in the number of research papers published lately and is becoming a very popular field. It is one of the most promising bio-hydrogen pathways and some of the techniques have brought fermentative hydrogen production to a point where production rates with real substrates like waste waters and under realistic conditions are approaching practical levels (Hallenberg & Ghosh 2009). It is still early to exactly tell what a “practical” production levels would be, but first estimates based on comparison with lignocellulosic ethanol production plants (50 to 100kj/l/h) would call for a hydrogen production rate of 5 to 10 l/l/h (Aden et al. 2002).

So where are we at exactly with our hydrogen production rates? Let’s first look and understand the theoretical maximal rate of the PFOR fermentative bio-hydrogen pathway that we use in Chapter V. When bacteria degrade organic substrates, electrons are produced which need to be disposed of to maintain electrical neutrality. In anoxic environments, the PFOR pathway utilizes protons as electron acceptors to produce molecular hydrogen. In this reaction, pyruvate is oxidized to acetyl-CoA thereby reducing ferredoxin (Fd). Reduced Fd is oxidized by hydrogenase which generates Fd(ox) and releases electrons to produce molecular hydrogen:



With one mole of glucose, this pathway catalyzes the production of up to 4 moles of gaseous hydrogen. Our scaffolding approach improves the work done by Agapakis et al. by over 12 fold, bringing us from a 3% of the maximal theoretical yields to about 36% (about 1.4 moles of H₂ per mole of glucose). This could be further enhanced as *E.coli* tends to keep a fairly high rate of other fermentative processes producing lactate, succinate, ethanol, acetate and carbon dioxide. A background strain with deletion of pyruvate metabolism genes such as pyruvate dehydrogenase, pyruvate formate lyase as well as other metabolic enzymes degrading pyruvate has proven to lead to high amount of pyruvate accumulation (Zhu et al. 2008), and would be useful to our purpose. Additionally, this strain could have deletions to improve electron flux through our pathway and improve hydrogenase metallic cluster biogenesis. For example, *ydbK* is an *E. coli* analogue of PFOR which can divert electron away from our pathway. Its deletion was previously shown to improve hydrogen production (Agapakis et al. 2010). *iscR* is a negative regulator of iron sulfur cluster biogenesis shown to limit hydrogen production from heterologously expressed [FeFe]-hydrogenases (Akhtar & P. R. Jones 2008).

Would that still be enough to make fermentative bio-hydrogen production economically viable? Studies have shown that reaching maximal theoretical yields of the PFOR pathway would still be not sufficient (DAS & VEZIROGLU 2008). However, different strategies could still make this process economically plausible. A hybrid system using both fermentative and photosynthetic bacteria could be envisioned in which photosynthetic bacteria would further reduce the acid fermentation wastes from the fermentative bacteria to drive its photosynthetic processes and further evolve bio-

hydrogen (Basak & Das 2007; TAO et al. 2007). Another solution, and an interesting direction for synthetic biology, is to recognize that the quest for a universal chassis is probably an utopia and to start working at the microbial community scale (Brenner et al. 2008). Microbial consortia could here be an interesting solution to the bio-hydrogen production process as the community metabolic range would far exceed the one of any individual members (Li & Fang 2007). Here, mutual interdependence would have to be engineered to keep the consortia stable and spatial organization of the community could also be enforced to keep anaerobic processes at its core (Kim et al. 2008). There is still a long way for fermentative bio-hydrogen to become economically viable, however it is still a very dynamic field of research full of hope and where practical solutions will come from concomitant advances in synthetic biology, fermentation and bio-reactor engineering.

APPENDIX



Appendix I - Engineering *Ana-* *baena's* Heterocysts

1. Heterocysts - an attractive natural compartmentalization strategy to engineer.
1. Cyanobacterial heterocysts, a striking example of prokaryotic cell differentiation

Many cyanobacteria are capable of nitrogen fixation; however, they face the challenge of resolving an inherently oxygen-evolving photosynthetic process with oxygen inactivated nitrogenases (Tamagnini et al. 2002). Most cyanobacteria resolve this issue by separating the two metabolic processes in time through their circadian clock. This is the case in *Synechococcus sp.*, a popular chassis in synthetic biology (Huang et al. 1990). A few others, however, separate the two processes in space, and protect their nitrogenase from atmospheric and photosynthetic oxygen in differentiated cells called heterocysts. Heterocysts are terminally differentiated nitrogen fixing cells and are probably one of the most striking examples of cell differentiation in prokaryotes (Flores & Herrero 2009).

Heterocysts are morphologically and metabolically distinguishable from vegetative cells. They are larger, more rounded in shape, and have a distinct pigmentation composition compared to vegetative cells. The differentiated cells undergo profound metabolic and morphological changes. Notably, the oxygen evolving photosystem II seems to be dismantled, and two additional membrane-associated layers are added: an inner glycolipidic layer and an outer poly-saccharide layer (Flores & Herrero 2009). Heterocysts and vegetative cells are mutually interdependent under conditions of nitrogen deprivation. Vegetative cells supply fixed carbon as sucrose to heterocysts (Cumino et al. 2007) but also glutamate. Glutamate is converted to glutamine and other amino-acids by heterocysts which are shared with the vegetative cells (Martín-Figueroa et al. 2000).

Combined nitrogen availability



Combined nitrogen limitation

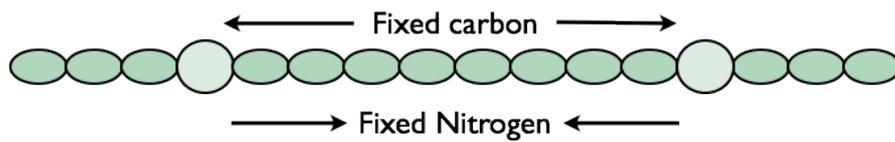


Figure 43: *Anabaena PCC7120* and heterocyst development. In the absence of combined nitrogen, vegetative cells (green) differentiate into heterocysts (pale green). Photosynthetically fixed carbon is exchanged for fixed nitrogen in between cells across the filament.

Of these filamentous heterocyst developing cyanobacteria, *Anabaena PCC 7120* is the best studied (Kumar et al. 2010). In the presence of combined nitrogen, *Anabaena* grows as long filaments spanning several hundreds of vegetative cells. In the absence of any combined nitrogen, every ten to twenty cells one vegetative cell differentiates into a heterocyst.

Therefore, we chose *Anabaena PCC 7120* as the candidate of choice to develop a new synthetic biology chassis to protect oxygen-sensitive synthetic metabolic reactions such as the PFOR hydrogen evolving pathway as heterocyst offer an anoxic protecting environment. To control both spatial and temporal expression of heterologous in *Anabaena*, we aimed at developing and characterizing a promotor library. To do so, it is

first important to understand gene expression and morphogenesis during heterocyst differentiation.

2. Gene expression and morphogenesis during heterocyst development.

The initial step of heterocyst development is sensing combined-nitrogen starvation. In *Anabaena*, it is an intermediate of the Krebs cycle, 2-oxaloglutarate, that gives this specific signal (Laurent et al. 2005). It is indeed the primary carbon skeleton for incorporation of ammonium and is considered as the metabolic junction between carbon and nitrogen metabolism (M. I. Muro-Pastor et al. 2001). The intracellular level of 2-oxaloglutarate conditions the correct expression of the transcriptional regulator NtcA. At the core of this process, the DNA binding activity of NtcA is enhanced in the presence of 2-oxoglutarate, and 2-oxoglutarate is necessary for transcriptional activation by NtcA (Tanigawa et al. 2002). In *Anabaena*, NtcA is absolutely necessary for the expression of the genes in pathways for ammonium and nitrate assimilation, as well as heterocyst development and is induced soon after combined nitrogen step-down (Wei et al. 1994).

HetR is the central signal processor of heterocyst differentiation and is one of the earliest genes induced in differentiating cells (Buikema & Haselkorn 1991), induced partly through NtcA (A. M. A. Muro-Pastor et al. 2002). HetR and NtcA reinforce the expression of one-another through simple transcriptional positive feedback. Both HetR and NtcA are also auto-regulated but this positive regulatory loop is central to hetero-

cyst development and determines the entry into the later stages of differentiation (A. M. A. Muro-Pastor et al. 2002).

An impressive regulatory cascade ensues that takes the early heterocysts into further stages of differentiation until the mature heterocyst, capable of nitrogen fixation (Golden & Yoon 2003) (Zhang et al. 2005) (Kumar et al. 2010). Of particular interest is the mechanism behind pattern formation and maintenance. The *patS* gene in *Anabaena* PCC 7120 is essential to normal pattern formation. It is transcribed early during heterocyst development, encoding a small peptide that diffuses from differentiating cells to neighboring cells. It is thought to bind and inhibit HetR and thus inhibits the NtcA-HetR feedforward loop necessary for entry into heterocyst differentiation (Kumar et al. 2010) (Zhang et al. 2005).

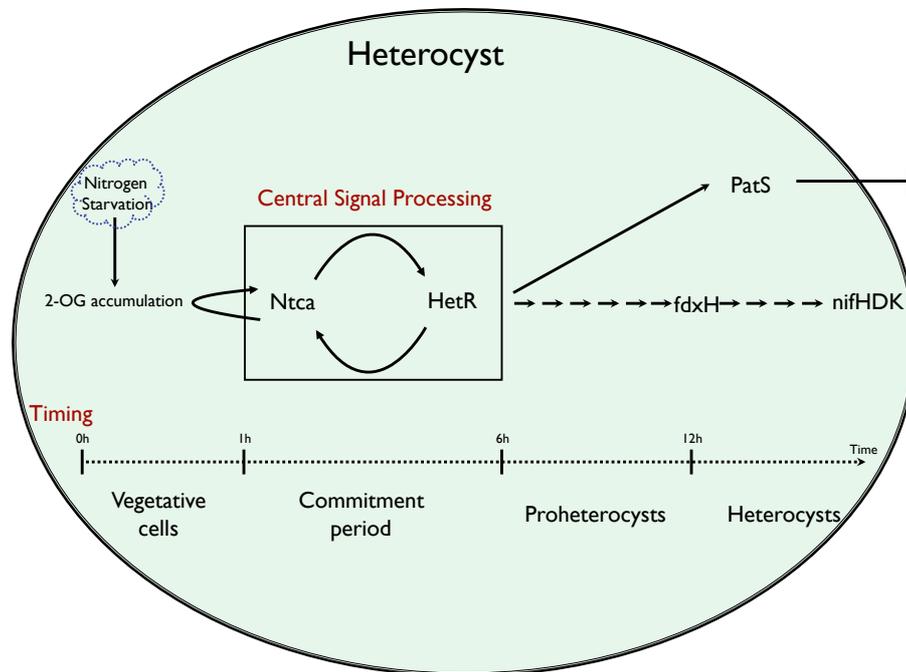


Figure 44: Heterocyst differentiation and pattern formation.

2. Challenges in engineering *Anabaena PCC 7120*

1. *Anabaena* cultures

Anabaena PCC7120 strain was kindly provided by Peter Wolk from Michigan State University. A first step consisted of characterizing and optimizing growth conditions before developing a suitable transformation protocol to engineer *Anabaena*.

Anabaena cultures were inoculated in 15mL transparent culture tubes in BG11 media supplemented with either 30mM fructose, 1 g/liter HEPES (pH 8.9; Sigma) or BG110 (nitrate depleted). Cultures were grown in temperature-controlled (35°C) and CO₂-controlled (2%) Multitron Infors HT incubators. Light intensity (150W, Gro-Lux bulbs; Sylvania) at the growth surface was measured at 65 $\mu\text{E m}^2 \text{ s}^{-1}$ (LI-250A Light Meter LI-COR with LI-190SA Quantum Sensor or US-SQS Spherical Micro Quantum Sensor; Walz). Optical density was followed at 750nm.

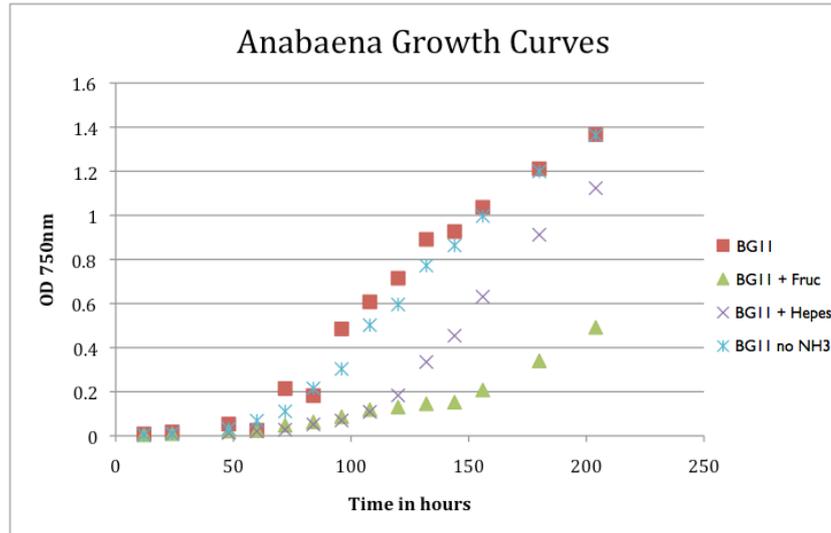


Figure 45: *Anabaena* growth assays. *Anabaena* inoculum were grown in Multitron In-fors HT incubators in BGII, BGII with fructose, BGII with Hepes or BGIIo (without nitrate). Optical Density was measured at 750nm.

Interestingly, fructose supplementation appeared to be toxic to the culture, despite contradictory reports (Ungerer et al. 2008). Fructose has been used to help maintain growth during conjugation protocols involving a dark incubation for some cyanobacteria strains. However, we eliminated fructose in our conjugation protocol as a consequence of these results. Buffering BGII with HEPES slowed the growth of *Anabaena* during the first few days, accordingly we did not buffer our growth media in subsequent experiments. Finally, combined nitrogen deprivation only slightly slowed the growth of *Anabaena* during the first few days of the culture.

2. Developing a working transformation protocol

Conflicting reports also exist concerning transformation protocols. Genetic transformation of *Anabaena* by tri-parental or bi-parental conjugation (Elhai & Wolk 1988) and electroporation (Thiel & Poo 1989; Thiel & Poo 1989) have been reported. Interested in making biology easier to engineer, we started by testing electroporation protocols without success. . We suspect this is probably due to the many restriction enzymes known to be present in *Anabaena*, capable of rapidly degrading unprotected exogenous DNA (Elhai et al. 1997).

To circumvent this problem we adapted a conjugation protocol originally developed by Wolk and colleagues (Elhai et al. 1997). We settled on using a bi-parental conjugation approach with *E. coli* conjugal donor strain AM1359. This strain contains the ampicillin resistant conjugal plasmid pRL443 (an RP4 derivative plasmid) and helper chloramphenicol resistant plasmid pRL623 (protects DNA through methylation against the known isoschizomers of AvaI, AvaII and AvaIII present in *Anabaena*). Overnight culture of AM1359 containing the plasmid to be conjugated were diluted 1:20 and allowed to grow for approximately 2 hours until mid-log phase. 1.5mL of this culture was washed twice with 1mL LB to wash away any traces of antibiotic. Aliquots of *Anabaena* cultures containing approximately 3×10^7 cells were added to the *E. coli* cells and mixed gently. The mixture were spread onto petri dishes made of BG11 1% agar with 5% LB and the plates were allowed to dry before incubation under *Anabaena* growth conditions for 24 hours. The plates were then washed with 1mL BG11 and the resulting suspension diluted to 1:100 and 150uL was spread onto new BG11 plates with the appropriate antibiotic (25ug Neomycin per mL). The plates were incubated under *Anabaena*

growth conditions for 1 to 3 weeks in temperature-controlled (35°C) and CO₂-controlled (2%) Multitron Infors HT incubator until the formation of visible *Anabaena* colonies.

This simplified transformation protocol for *Anabaena* is greatly inspired by the work of Elhai and Wolk (Elhai et al. 1997). It only requires one *E. coli* parental strain (strain AM1359) which carries both the conjugation and methylation machinery. We reduced the overall time needed to complete a transformation by doing the conjugation directly on plates instead of filters and by using whole *Anabaena* filaments (no need to sonicate to fragment the filaments to obtain individual cells). Another important improvement is the use of a CO₂ enriched incubator for plates. Overall, this protocol can now take as little as ten days.

3. Controlling the expression of heterologous proteins in time, space and intensity

1. Developing a promoter library

Many individual papers report spatial and temporal expression of individual genes involved in heterocyst-differentiation in *Anabaena* and have attempted to determine the associated promoter sequences. These reports frequently use reporters cloned next to the promoter element of interest in order to elucidate the pattern of gene expression along a filament. However, no solidified database exists and all these reports use dif-

ferent characterization protocols, different expression vectors and reporter systems. In order to maximally utilize *Anabaena* as a platform to house oxygen-sensitive or other synthetic pathways, it will be necessary to have more detailed information on the kinetics and relative expression driven from each promoter. We attempt here to carefully characterize a library of working promoters for the expression of heterologous proteins in *Anabaena* in both space (heterocyst vs. vegetative), time (at different heterocyst maturation points) and intensity.

The Figure 46 is a table compiling all the promoters PCRRed from *Anabaena* for used in heterologous protein expression. It results from an intense literature search looking for candidate promoters. It includes the name of the promoter, its expression location, its expression timing, the primer used to PCR it (Bgl Biobrick format) and the associated literature reference. All promoters were PCRRed from *Anabaena* and the library was cloned into p125 replicative plasmid (Elhai et al. 1997) in front of a YFP gene using a Gibson cloning strategy (Gibson et al. 2009). Resulting plasmids were transferred into *E. coli* strain AM1359 and conjugated into *Anabaena* according to the previously described protocol.

| Promoter Name | Spacial Expression | Expression timing | PCR Primers | Reference |
|---------------|--------------------|-------------------|--|----------------|
| pNif | Het | Late | CCAACCAATTGCAG- GAAAAGAGAACA GGATCC TAA CTCGAG GAAGA TCTTC CTCGAG TTA GGATCC TGTTCTCTTTTCCCTG CAATTGGTTGG | {Elhai:1990wu} |

| Promoter Name | Spacial Expression | Expression timing | PCR Primers | Reference |
|---------------|--------------------|-------------------|--|---------------------|
| pHetC | Het | Early | GTTGG GAATTC ATG AGATCTTTTAGT ACA TCG GTG AGG GG TCTTC CTCGAG TTA GGATCC AGTT- TAATTTCTGTTTGGT GTGTAAAC | {Valladares:2008ju} |
| pHetR | Het | Early | GTTGG GAATTC ATG AGATCT TGGTATTGGCAAATA CAAATCC TCTTC CTCGAG TTA GGATCC ATTA- CAAATAGTTGAATAG- CACGC | {Black:1993dj} |
| pPatB | Het | Mid | GTTGG GAATTC ATG AGATCT AATA- CATCTGCCACAACCG TCTTC CTCGAG TTA GGATCC ATAACTTTCTTCC- CACCTAATCG | {Jones:2003bs} |
| pCox-BACII | Het | Late | ACTGG GAATTC ATG AGATCT CTAA- GAACTGCTACACACA- CAAC TCTTC CTCGAG TTA GGATCC ACCACC- TACTCATTACT- TATCG | {Jones:2002wy} |
| pGlnA | All cells | Constitutive | ACTGG GAATTC ATG AGATCT AG- TAGCGTAGCGCAGA- TAGTAGTCC TCTTC CTCGAG TTA GGATCC TGTTACTCCTTCTCTG CCAATTTC | {Elhai:1990wu} |

| Promoter Name | Spacial Expression | Expression timing | PCR Primers | Reference |
|---------------|--------------------|-------------------|---|---------------------|
| pRbc1 | Veg | Constitutive | GCTGG GAATTC ATG AGATCT TGTTGGTGATGGTGC ATTAGTG TCTTC CTCGAG TTA GGATCC ATC- TATCCTTCCAA- GATGTCACTC | {Elhai:1993kn} |
| pNtcA | Het | Early | GCTGG GAATTC ATG AGATCT TATCG- GAAAAAATCTGTAA- CATGAG TCTTC CTCGAG TTA GGATCC TATT- TACCTCCTTTATAGA- GAGATAC | {MuroPastor:1999ti} |

Figure 46: Table of all the characterized *Anabaena* promoters with associated PCR primer used and references. Het=Heterocyst; Veg= Vegetative cells; Early= 1-6h Mid = 7-14h and late= 15-24h after Nitrogen deprivation.

2. FACS characterization

To the best of our knowledge, flow cytometry has never been used as a tool to study *Anabaena*. However, the very large morphological and optical differences in between vegetative cells and heterocysts (larger, extra membrane layers, lack of chlorophyll) makes it a very interesting tool to study the bacterium and characterize our promoter library.

We show here early experiments that represent an important step forward towards developing molecular tools and protocols to better study and work with *Anabaena*. Prior to FACS characterization, filaments were disrupted via sonication resulting in population of individual cells. After sonication, filaments were checked via optical microscopy for correct disruption. Cells were then centrifuged and washed once with fresh BG11. FACS characterization was done on a LSRII flow cytometer according to the manufacturer's recommendations.

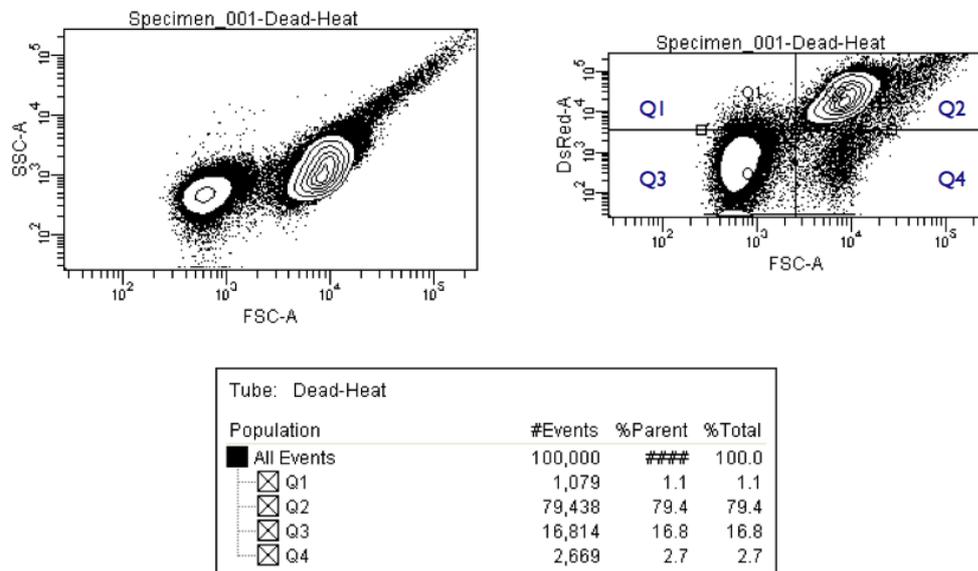


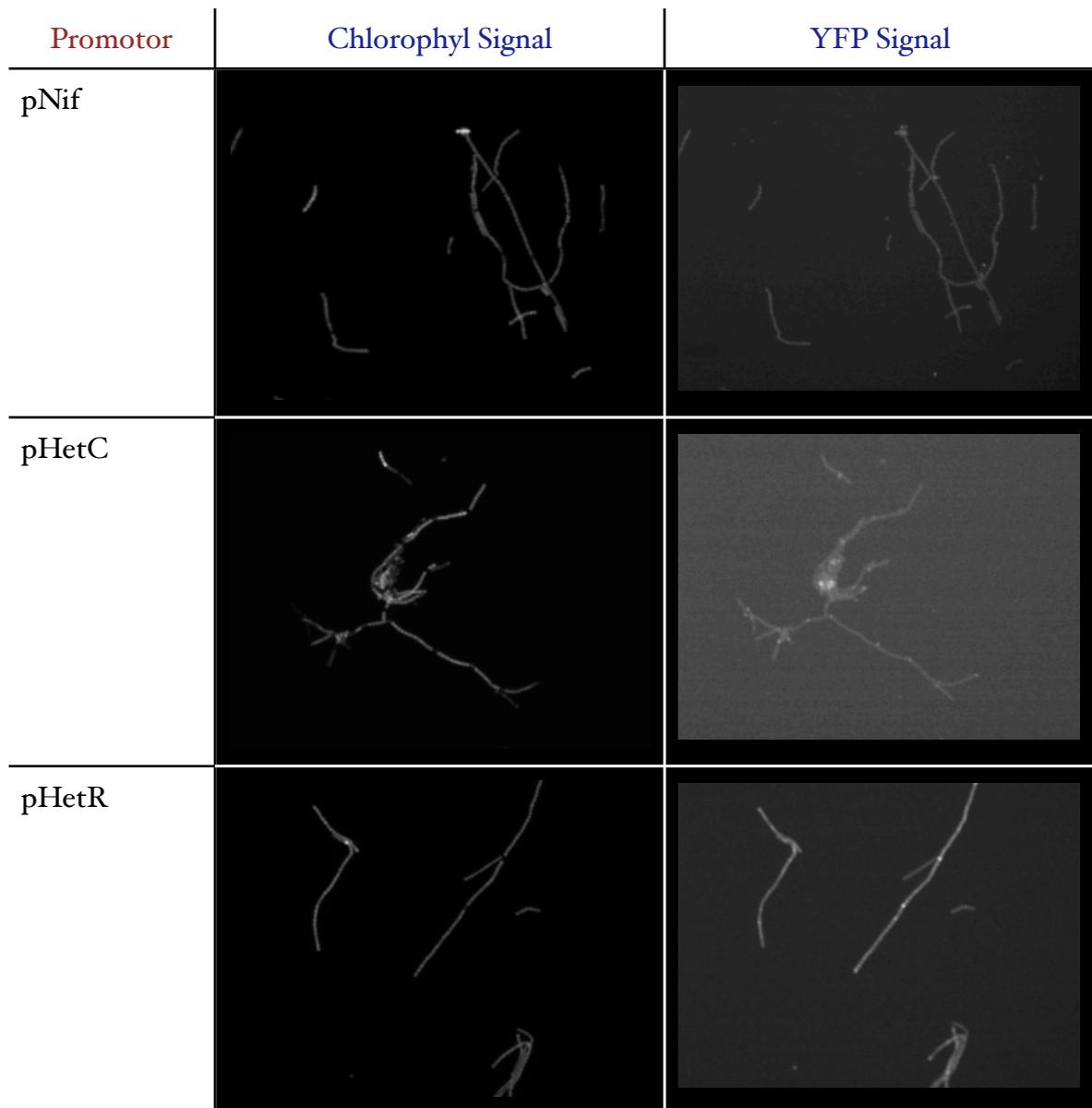
Figure 47: FACS Characterization of wild-type fragmented *Anabaena* PCC 7120. Two different cells population can be observed differentiated using Forward Scatter (FSC, measure of cell size), Side Scatter (SSC, measure of granularity, *i.e.* differences in membrane light diffraction) or red fluorescence (DsRed, measure of chlorophyll concentration).

A clear separation in between heterocysts and vegetative cells can be seen based on forward scatter, side scatter and DsRed signal (fluorescent signal in the red channel is proportional to chlorophyll levels). It is an interesting first result suggesting FACS can be a useful tool to study *Anabaena*. The quality of the data could however greatly benefit from improved filament fragmentation methods. Most of the “noise” in the experiment comes from cells fragment or non dissociated small filaments, observed here as the tail in the FACS data. Improving and fine tuning sonication is difficult, but a nano-filtration step could greatly improve the data quality. The goal in a first step would be to remove non separated filaments by using a filter with a pore size of approximately 10 μ m. A second nanofiltration would remove small cell fragments with a filter with a pore size 0.4 μ m. This first set of data holds promises as of the utility of FACS as a characterization tools for molecular work on *Anabaena* by being able to clearly separate cell types.

3. Optical Microscopy characterization of an endogenous promoter library

For initial promotor characterization, *Anabaena* was grown in 15mL BG11 cultures in Multitron Infors HT incubators until saturation. Cultures were then washed and diluted to 1:20 in BG110 (without combined nitrogen) and allowed to grow for another 36 hours. This 36 hour time frame for initial characterization of the promoters comes from the fact that for metabolic engineering purposes, we are not interested in promoters turning off after heterocyst differentiation.

For observation, cells were placed between a glass slide and 2% agarose pads, and analyzed using a Nikon TE- 2000 microscope (60X, 1.4 numerical aperture objective, ORCA-ER charge-coupled device camera, DsRed 100ms exposure time and FITC 2500ms exposure time channels).



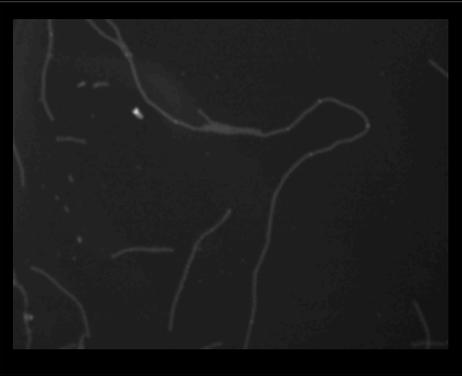
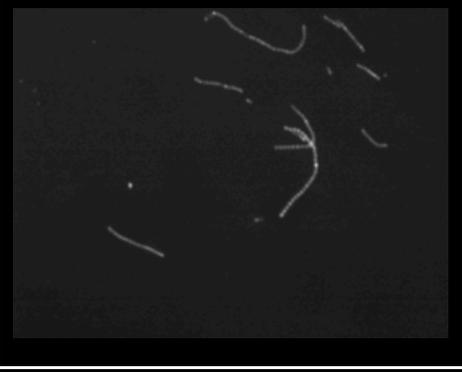
| Promotor | Chlorophyl Signal | YFP Signal |
|-----------|---|--|
| pPatB |  |  |
| pCoxBACII |  |  |
| pGlnA |  |  |
| pRbcL |  |  |

Figure 48: Spatiotemporal characterization of the *Anabaena* promoter library via fluorescence imaging. *Anabaena* cells express YFP under the control of our library of promoters. First column corresponds to the red chlorophyll signal where only vegetative cells fluoresce, black spots within filaments correspond to heterocysts. Second column corresponds to the promoter driven YFP signal. Fluorescence signal where a black spot was observed in the red channel corresponds to heterocysts expression.

Based on these early results, YFP expression intensity varies widely between the different promoters. pHetC, pHetR, pPatB, pCoxBACII are indeed expressed preferentially in heterocysts. pGlnA is expressed constitutively and pRbxI is uniquely expressed in vegetative cells. pCoxBACII seems like a very interesting candidate for strong heterocyst specific expression. Its expression is highly heterocyst specific and remains strong 36h after nitrogen deprivation.

The very high expression levels observed with pCoxBACII is consistent with our current understanding of heterocyst's metabolism. 40% higher rate of respiration has been observed in isolated heterocysts (Fay & Walsby 1966) thus providing extra ATP necessary to fix atmospheric nitrogen. Higher respiration rates are also believed to be one of the ways anoxic environment is maintained in heterocysts (Murry et al. 1984). The CoxBacII operon codes for a cytochrome c oxidase operon (K. M. K. Jones & Haselkorn 2002) which could be a major player in dissipating oxygen as well as providing ATP for the cell by being a terminal mitochondrial electron acceptor.

Promoters' intensity after complete heterocyst differentiation has never been observed to the best of our knowledge. We present here early results that should be interesting for the metabolic engineering community as a whole. Further characterization of the promoter and relative expression level across time and space will enable us to develop a highly useful toolbox for synthetic biology by opening the doors to implementing fully functional oxygen-sensitive pathways in oxygen evolving cyanobacteria. As such, it paves our way towards engineering *Anabaena* for hydrogen production, an ongoing work.



Appendix II - Science Paper

Organization of Intracellular Reactions with Rationally Designed RNA Assemblies

Camille J. Delebecque,^{1,2,3,4} Ariel B. Lindner,^{3,4*} Pamela A. Silver,^{1,2*} Faisal A. Aldaye^{1,2}

The rules of nucleic acid base-pairing have been used to construct nanoscale architectures and organize biomolecules, but little has been done to apply this technology *in vivo*. We designed and assembled multidimensional RNA structures and used them as scaffolds for the spatial organization of bacterial metabolism. Engineered RNA modules were assembled into discrete, one-dimensional, and two-dimensional scaffolds with distinct protein-docking sites and used to control the spatial organization of a hydrogen-producing pathway. We increased hydrogen output as a function of scaffold architecture. Rationally designed RNA assemblies can thus be used to construct functional architectures *in vivo*.

In cells, multienzymatic pathways are often physically and spatially organized onto scaffolds or clusters or into microcompartments (1). Spatial organization helps substrates flow between interacting proteins, limits cross-talk between signaling pathways, and increases yields of sequential metabolic reactions (1, 2). The abil-

ity to organize protein complexes and biological pathways spatially presents a strategy to engineer cells (3, 4).

The spatial organization of biomolecules has been the focus of DNA nanotechnology (5–8). This approach makes use of DNA's base-pairing to generate one-, two-, and three-dimensional (1D,

2D, 3D) assemblies. DNA structures have largely remained limited to *in vitro* applications (9). RNA provides a compatible material for *in vivo* nucleic acid-based construction (10). It can be produced via the transcription machinery and forms stable interactions. RNA has been used to build higher-order assemblies *in vitro* (11, 12) and can potentially be used *in vivo* to engineer the intracellular environment.

In this work, we engineered synthetic RNA modules that assemble into functional discrete, 1D, and 2D scaffolds *in vivo*, and we used them to control the spatial organization of bound proteins (Fig. 1A) (see supporting online materials and methods). Scaffold D0 was constructed from a single RNA module d0, which folded into a duplex with PP7 and MS2 aptamer domains that bind PP7 and MS2 fusion proteins (Fig. 1B) (13).

¹Harvard Medical School, Department of Systems Biology, 200 Longwood Avenue, Boston, MA 02115, USA. ²Wyss Institute for Biologically Inspired Engineering, 3 Blackfan Circle, Boston, MA 02115, USA. ³Institut National de la Santé et de la Recherche Médicale, Unité 1001, F-75015 Paris, France. ⁴Faculty of Medicine, Paris Descartes University, F-75015 Paris, France.

*To whom correspondence should be addressed. E-mail: pamelasilver@hms.harvard.edu (P.A.S.); ariel.lindner@inserm.fr (A.B.L.)

We developed an approach for the *in vivo* isothermal assembly of extended RNA scaffolds by constructing sequence-symmetric RNA building blocks (Fig. 1, C and D) inspired by 2D DNA analogs (13, 14). These RNA strands possess dimerization domains (DDs) and polymerization domains (PDs). To prevent the formation of ill-defined networks, it was necessary to disfavor the collapse of the palindromic regions (15) and control assembly order by insuring tile formation before polymerization. We achieved this by designing PDs that fold intramolecularly into kinetically protected hairpin structures (Fig. 1D, step i). The stem of these hairpins is an overlapping shared domain with the DD that discourages collapse (Fig. 1D, red segments), allowing the DD to activate the PD upon self-binding (Fig. 1D, step iii). We further destabilized the collapsed state by incorporating wobble pairs and mismatches (figs. S1 to S4).

The 1D RNA assembly D1 was derived from a single RNA d1 with PP7 and MS2 binding domains (Fig. 1E). d1 assembled into d1-1 (step i), which self-assembled into d1-2 (step iii). The torsion in d1-2 induced folding into an RNA nanotube capable of growing into the 1D scaffold D1 (step iv). The 2D RNA assembly D2 was formed from d2' and d2'', each carrying a distinct PP7 and MS2 aptamer (Fig. 1F). The dor-

mant tile d2' spontaneously generated the pro-tile d2-1 (step i), which interacted with d2'' to generate tile d2-2 (step ii). d2-2 then self-assembled into the 2D RNA scaffold D2 with PP7 and MS2 binding domains (step iv).

We used atomic force microscopy (AFM) to characterize *in vitro* transcribed RNA modules d1 and d2'/d2''. d1 formed 1D RNA fibers (D1), whereas d2'/d2'' assembled into 2D extended RNA fibers (D2) (Fig. 2A). The width of D1 (~5 nm, a few tiles wide) is smaller than that of its DNA analog (14) and might also correspond to 1D ribbons (D1₂) constructed from a continuous line of single tiles (Fig. 1E, step ii). Given that D2 preferentially grows in a single direction when compared with its DNA analog (13), it might also correspond to RNA nanotubes (D2₂) that are relatively wider than D1 (Fig. 1D, step iii). To confirm the validity of our assemblies, we used analogs of d1 and d2' with a poly-T stretch in place of the DD incapable of assembling; d1_T and d2'_T did not generate extended assemblies (Fig. 2B).

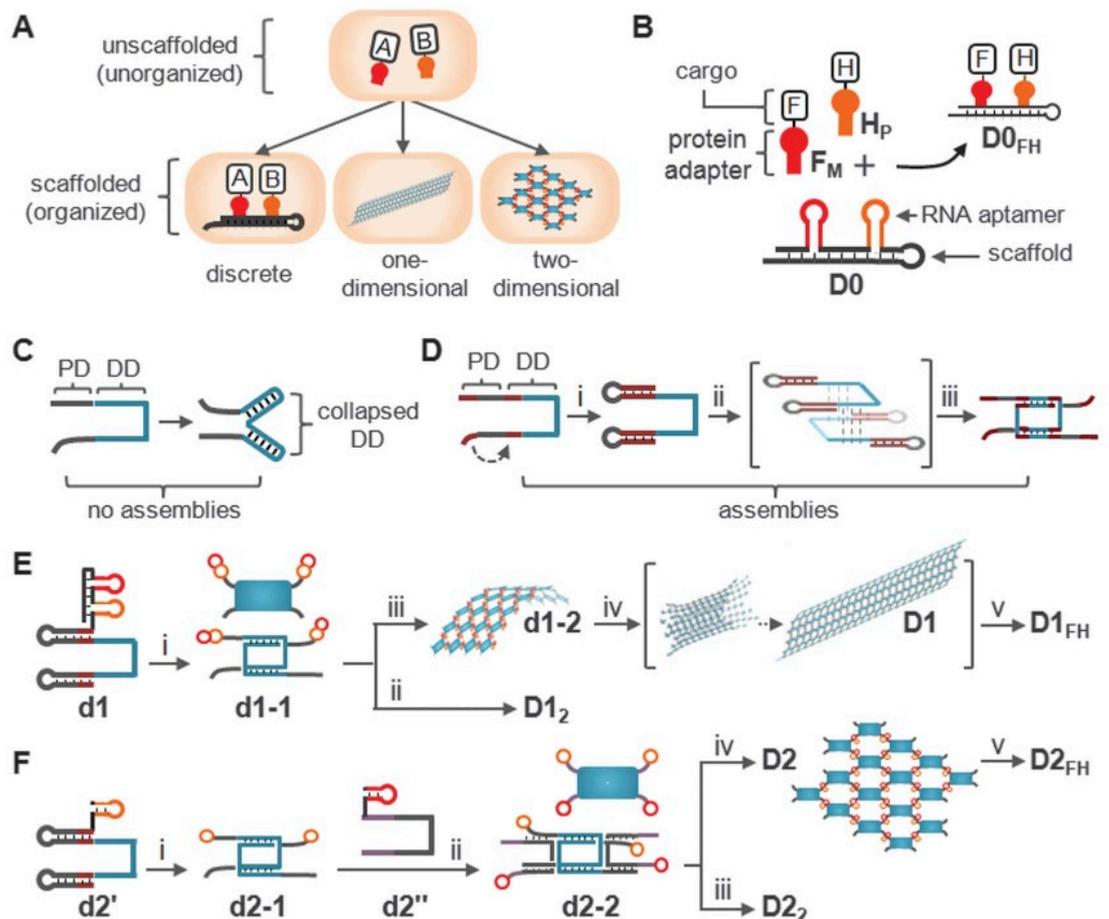
We developed a DNA-based precipitation (DP) method to purify our RNA assemblies from cells. Streptavidin-coated magnetic beads with a biotinylated DNA capture probe (DP_C) were added to cell lysates. The capture domain of DP_C binds the T7 terminator in our RNA molecules (Fig. 2C,

step i). The RNA assemblies were released upon addition of DP_R that bound the release domain of DP_C (Fig. 2C, step ii). We were able to capture and release RNA (Fig. 2D).

In vivo synthesized D1 and D2 revealed extended 1D and 2D assemblies (Fig. 2E and fig. S5). Cross-sectional height analysis showed D1 to have two populations of distinct height (3 and 6 nm), which is characteristic of open versus closed nanotubes. *In vivo* D2 assembled into 2D structures that are smaller and somewhat different than their *in vitro* counterparts, suggesting that the assembly process in cells is of lower fidelity. To confirm that the assemblies formed *in vivo*, we engineered a set of inhibitory strands (ISs) that bound the trigger domains of d1' and d2'. The inhibition by these strands was confirmed *in vitro* (Fig. 2F). The purification of D1/D2 in the presence of excess ISs did not eliminate the observed 1D and 2D assemblies (Fig. 2G), confirming the formation of D1 and D2 in cells pre-lysis.

Transmission electron microscopy (TEM) analysis of whole bacterial cells expressing D1 or D2 confirmed their assembly in cells. The RNA assemblies were tagged with gold-binding metallothionin-PP7 fusion proteins (P_{Au}) that form clusters (Fig. 2H) (16). Cells coexpressing P_{Au} and D1 formed thin filaments with lengths

Fig. 1. Design of RNA modules to organize proteins. **(A)** Proteins A and B scaffolded onto discrete, 1D, and 2D RNA assemblies. **(B)** D0 is a RNA strand that folds into a duplex with PP7 and MS2 sites. Ferredoxin/MS2 (F_M) and hydrogenase/PP7 (H_P) bind D0 to generate D0_{FH}. **(C and D)** RNA with DDs and PDs initiates the formation of extended assemblies. Capping the palindromic sequences in DDs with PDs prevents its collapse (i) and allows for self-assembly (ii) into functioning tiles (iii). **(E)** D1 is constructed from a RNA strand d1 bearing PP7 and MS2, and it assembles into tile d1-1 (i). d1-1 assembles into a ribbon D1₂ (ii) or into a nanotube d1-2 (iii) that grows into D1 (iv). D1 organizes F_M and H_P into D1_{FH} (v). **(F)** D2 is constructed from d2' and d2'' bearing PP7 and MS2, respectively. d2' assembles into the pro-tile d2-1 (i) and interacts with d2'' to generate d2-2 (ii). d2-2 self-assembles into a nanotube D2₂ (iii) or the 2D D2 (iv). D2 organizes F_M and H_P into D2_{FH} (v).



of 200 to 300 nm, whereas cells coexpressing D2 formed compact spherelike structures ~100 nm in diameter. D0, D1, or D2 does not affect cell growth (fig. S12). Cells carrying the D1 and D2 scaffolds had higher RNA levels relative to cells expressing mutated poly-T RNA analogs (Fig. 2I), consistent with the formation of degradation resistant assemblies. Thus, d1 and d2'/d2'' assembled *in vivo* into D1 and D2.

We used fluorescence complementation to detect protein assembly on our RNA scaffolds (Fig. 3). Green fluorescent protein (GFP) split into two halves (F_A and F_B) fused to the PP7 or MS2 aptamer binding proteins was used (Fig. 3A). Cells expressing F_A and F_B alone (Fig. 3B) or D0, D1, or D2 without the split GFPs displayed little fluorescence. However, the co-expression of D0, D1, or D2 with the split GFPs showed increased fluorescence (Fig. 3C).

Thus, our RNA scaffolds served as docking sites to promote protein-protein interactions in cells.

Biological hydrogen production has both fundamental and practical implications. Coexpression of [FeFe]-hydrogenase and ferredoxin catalyzes the reduction of protons to hydrogen through electron transfer (17). We used this system to assess the potential of our RNA scaffolds to constrain flux through spatial organization. We fused the hydrogenase to a single copy of PP7 (H_P) and ferredoxin to a dimer of MS2 (F_M), and we conducted electrophoretic gel-shift analysis of the binding of F_M and H_P to D0 (Fig. 4A). Addition of H_P to D0 resulted in a single product termed $D0_H$. The addition of F_P to D0 resulted in the formation of $D0_F$. The addition of H_P and F_M to D0 resulted in a single product assigned to the protein-RNA assembly $D0_{FH}$.

H_P and F_M assembled onto D0 in cells to form $D0_{FH}$ (Fig. 4B).

To determine whether our RNA scaffolds increased hydrogen biosynthesis, we used gas chromatography to analyze cells expressing the hydrogen-producing pathway, along with the different RNA assemblies (fig. S11). The relative levels of F_M and H_P expression in D0, D1, and D2 cells were comparable (fig. S8). D0, D1, and D2 assembled F_M and H_P into $D0_{FH}$, $D1_{FH}$, and $D2_{FH}$ (Fig. 1). D0 resulted in a 4.0 ± 1.3 -fold increase in hydrogen production compared with unscaffolded H_P and F_M (Fig. 4C). Hydrogen output with the extended assemblies D1 and D2 resulted in a 11 ± 2.8 - and 48 ± 1.5 -fold increase in hydrogen production (Fig. 4C). When normalized against the amount of RNA in cells (Fig. 2I and fig. S7), D0, D1, and D2 resulted in a 4.0-, 6.2-, and 24-fold increase. The increase with D2 is consistent with

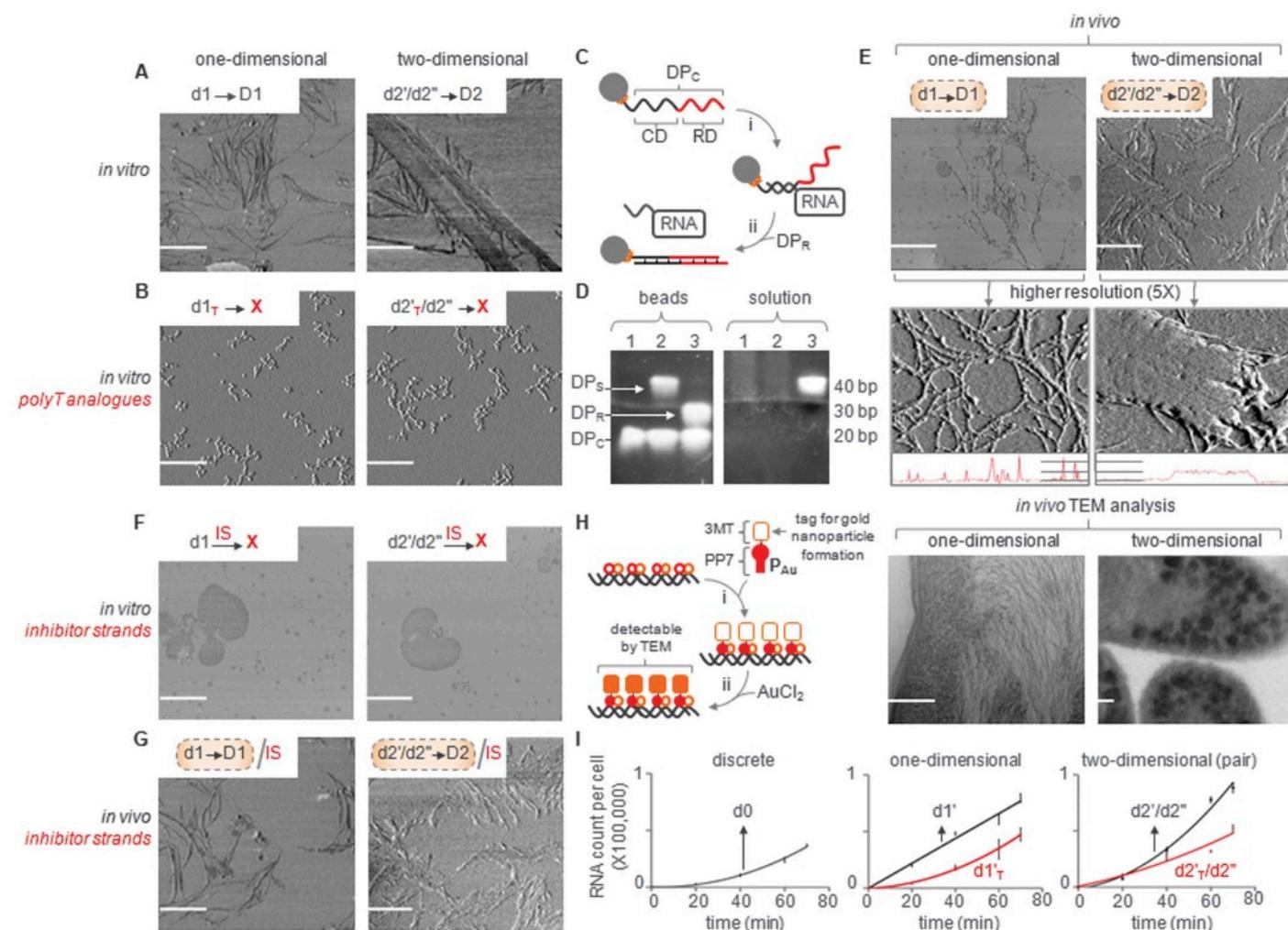


Fig. 2. Characterization of RNA assemblies. (A) *In vitro* transcribed d1 and d2'/d2'' assemble into D1 and D2 (AFM; phase images; scale bars, 0.25 μ m). (B) *In vitro* transcribed mutated RNA d1_T and d2'_T/d2''_T do not assemble. (C) DNA-based precipitation of *in vivo* RNA assemblies uses DP_C (i) and a release probe (DP_R) for recovery (ii). (D) Capture and release of substrate DP_S (left gel, beads; right gel, solution). Lane 1, conjugation of DP_C to streptavidin-coated magnetic beads; lane 2, capture of DP_S; lane 3, release of DP_S using DP_R. (E)

AFM analysis of purified assemblies. (F) ISs bind the DDs of d1 and d2' to prevent their assembly into D1 or D2 (circular structures are drying artifacts). (G) When used during the purification of d1 and d2'/d2'', D1 and D2 assemblies are still observed. (H) TEM analysis revealed the formation of 1D assemblies for D1 and 2D aggregates for D2 (scale bars, 100 nm). (I) Quantitative real-time fluorescence polymerase chain reaction analysis of *in vivo* RNA production levels. Error bars indicate SEM.

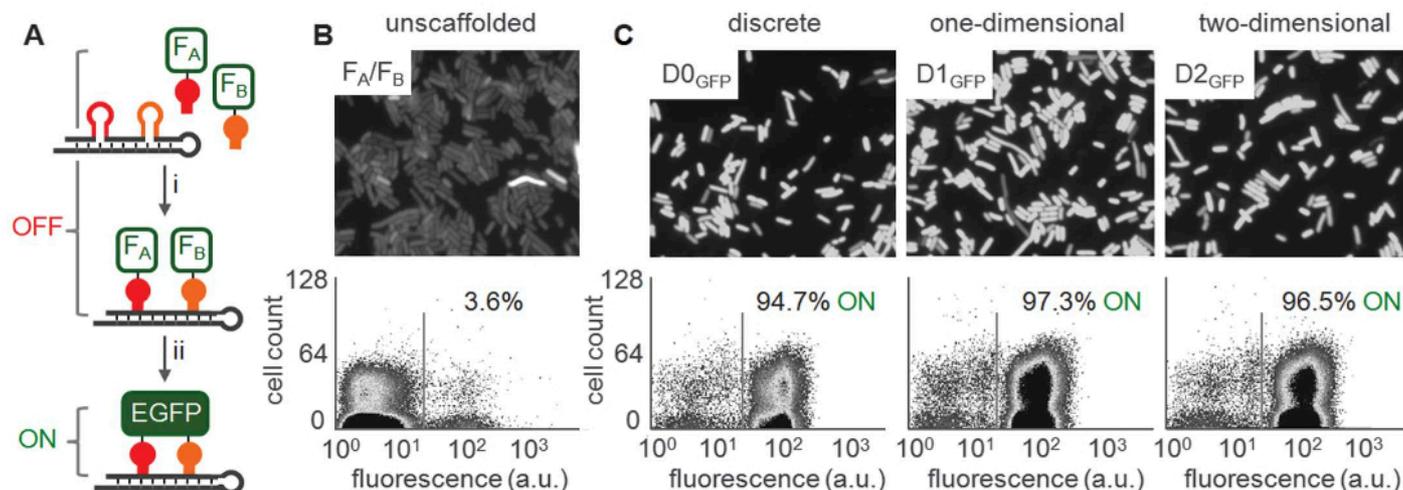


Fig. 3. Fluorescence protein complementation in vivo. (A) GFP split into two halves, each of which is fused to PP7 or MS2 (F_A and F_B). F_A and F_B bind their respective aptamers (i) and reconstruct functional fluorescent GFP (ii). EGFP, enhanced green fluorescent protein. (B) Fluorescence microscopy imaging of

cells expressing F_A and F_B revealed little to no fluorescence (scale bars, 10 μm). a.u., arbitrary units. (C) Cells coexpressing F_A and F_B with D0, D1, or D2 reveal an increase in fluorescence, indicating that D0, D1, and D2 scaffold PP7 and MS2 protein chimeras. Gray lines in flow cytometry plots separate OFF and ON cells.

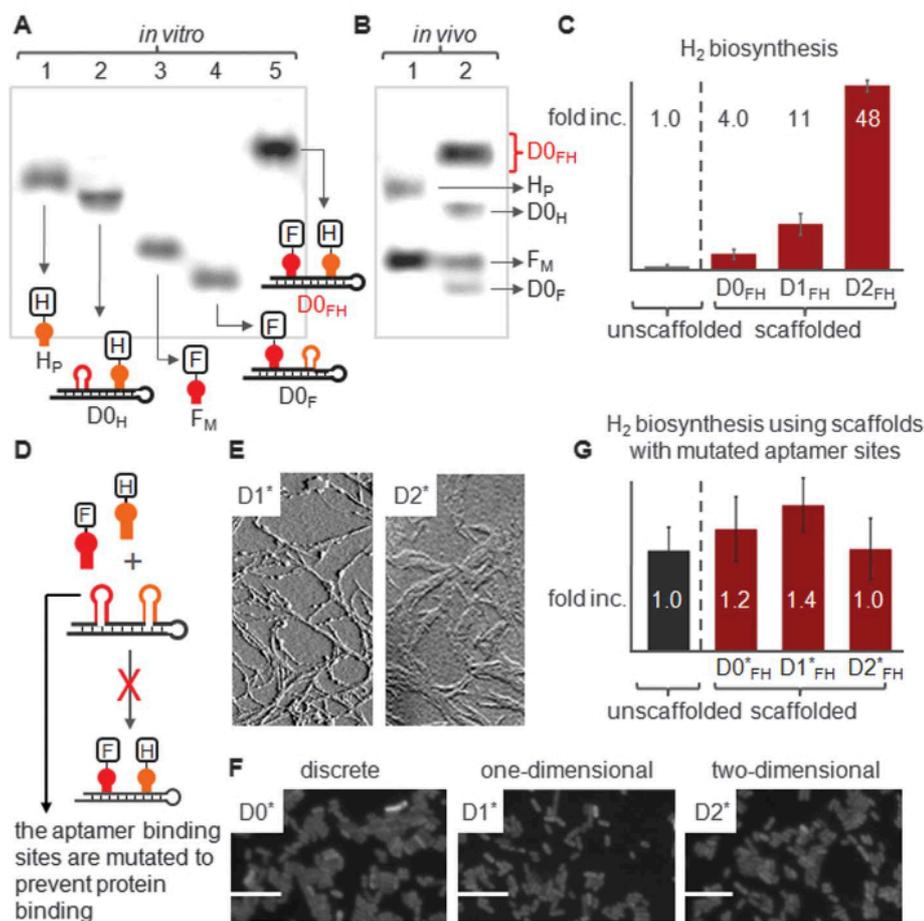


Fig. 4. Scaffolding hydrogen production. (A) In vitro gel shift of H_P (lane 1) binds D0 to form $D0_H$ (lane 2). F_M (lane 3) binds D0 to form $D0_F$ (lane 4). H_P and F_M bind D0 to form $D0_{FH}$ (lane 5). (B) In vivo gel shift of H_P and F_M (lane 1) and H_P and F_M in the presence of D0 (lane 2). (C) Hydrogen biosynthesis as a function of scaffold, normalized to unscattered cells expressing H_P and F_M . (D) Mutating aptamer binding sites (E) do not affect self-assembly, (F) but do prevent protein binding (scale bars, 10 μm) and (G) hydrogen production. Error bars indicate SEM. Dashed lines in (C) and (G) denote separation between scaffolded and unscattered proteins.

its assembly in vivo into “organelle-like” structures effective at concentrating proteins and their products (Fig. 2H). Mutating the PP7 and MS2 binding sites prevented protein scaffolding (Fig. 4, D to G). Thus, RNA can be used to organize enzymatic pathways in vivo to increase output as a function of architecture.

We controlled the spatial organization of proteins in cells using RNA molecules that are sequence-programmed to isothermally assemble into predefined discrete, 1D, and 2D structures in vivo. These assemblies scaffolded proteins and were used to organize a hydrogen-producing biosynthetic pathway. Hydrogen production was optimized as a function of scaffold architecture. Unlike protein-based approaches (3, 4, 17), RNA-based scaffolds allow for the formation of complex multidimensional architectures with nanometer precision. In vivo RNA assemblies can thus be used to engineer biological pathways through spatial constraints (18, 19).

References and Notes

- W. R. Burack, A. S. Shaw, *Curr. Opin. Cell Biol.* **12**, 211 (2000).
- D. F. Savage, B. Afonso, A. H. Chen, P. A. Silver, *Science* **327**, 1258 (2010).
- J. E. Dueber et al., *Nat. Biotechnol.* **27**, 753 (2009).
- S.-H. Park, A. Zarrinpar, W. A. Lim, *Science* **299**, 1061 (2003); 10.1126/science.1076979.
- P. W. K. Rothmund, *Nature* **440**, 297 (2006).
- N. C. Seeman, *Annu. Rev. Biochem.* **79**, 65 (2010).
- C. Lin, Y. Liu, H. Yan, *Biochemistry* **48**, 1663 (2009).
- E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **394**, 539 (1998).
- C. Lin et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17626 (2008).
- P. Guo, *Nat. Nanotechnol.* **5**, 833 (2010).
- L. Jaeger, A. Chworos, *Curr. Opin. Struct. Biol.* **16**, 531 (2006).
- B. Cayrol et al., *J. Am. Chem. Soc.* **131**, 17270 (2009).

REPORTS

13. H. Liu, Y. He, A. E. Ribbe, C. D. Mao, *Biomacromolecules* **6**, 2943 (2005).
14. H. Liu, Y. Chen, Y. He, A. E. Ribbe, C. D. Mao, *Angew. Chem. Int. Ed. Engl.* **45**, 1942 (2006).
15. P. Yin, H. M. T. Choi, C. R. Calvert, N. A. Pierce, *Nature* **451**, 318 (2008).
16. E. Diestra, J. Fontana, P. Guichard, S. Marco, C. Risco, *J. Struct. Biol.* **165**, 157 (2009).
17. C. M. Agapakis *et al.*, *J. Biol. Eng.* **4**, paper no. 3 (2010).
18. F. J. Isaacs, D. J. Dwyer, J. J. Collins, *Nat. Biotechnol.* **24**, 545 (2006).
19. M. N. Win, C. D. Smolke, *Science* **322**, 456 (2008).

Acknowledgments: This work was supported by from the Enerbio-Tuck Foundation and the Institut Français du Pétrole Energies Nouvelles (to C.J.D.); Agence Nationale de la Recherche France, Institut National de la Santé et de la Recherche Médicale–Institut National de Recherche en Informatique et en Automatique projet d’envergure, and Axa Foundation Chair on Longevity (to A.B.L.); Natural Sciences and Engineering Research Council of Canada (to F.A.A.); and the Wyss Institute for Biologically Inspired Engineering. We thank the Harvard Medical School Electron Microscopy Center for assistance

in electron microscopy, and we thank the reviewers for their insight.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1206938/DC1
Materials and Methods
Figs. S1 to S13
References (20–35)

13 April 2011; accepted 6 June 2011

Published online 23 June 2011;

10.1126/science.1206938



Appendix III - Nature Protocol Paper

Designing and using RNA scaffolds to assemble proteins *in vivo*

Camille J Delebecque¹⁻⁴, Pamela A Silver^{1,2} & Ariel B Lindner^{3,4}

¹Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. ²Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts, USA. ³Institut National de la Santé et de la Recherche Médicale, Paris, France. ⁴Faculty of Medicine, Paris Descartes University, Paris, France. Correspondence should be addressed to P.A.S. (pamela_silver@hms.harvard.edu) or A.B.L. (ariel.lindner@inserm.fr).

Published online 6 September 2012; doi:10.1038/nprot.2012.102

RNA scaffolds are synthetic noncoding RNA molecules with engineered 3D folding harnessed to spatially organize proteins *in vivo*. Here we provide a protocol to design, express and characterize RNA scaffolds and their cognate proteins within 1 month. The RNA scaffold designs described here are based on either monomeric or multimeric units harboring RNA aptamers as protein docking sites. The scaffolds and proteins are cloned into inducible plasmids and expressed to form functional assemblies. RNA scaffolds find applications in many fields in which *in vivo* organization of biomolecules is of interest. RNA scaffolds provide extended flexibility compared with DNA or protein scaffolding strategies through programmed modulation of multiple protein stoichiometry and numbers, as well as the proteins' relative distances and spatial orientations. For synthetic biology, RNA scaffolds provide a new platform that can be used to increase yields of sequential metabolic pathways.

INTRODUCTION

Synthetic RNA scaffolds are engineered noncoding RNAs (ncRNAs) designed to bind and organize specific proteins¹. They rely on a simple principle: a designed gene is transcribed *in vivo* into an ncRNA whose secondary structure enables the controlled scaffolding of heterologous proteins in *Escherichia coli* via aptamer domains. In the present context, aptamers are small, naturally occurring RNA secondary structures that bind specific protein targets. Aptamers can also be evolved *in vitro* to bind a specific target by SELEX (systematic evolution of ligands by exponential enrichment)^{2,3}. As aptamers have numerous applications in the pharmaceutical industry, the library of available protein-aptamer pairs is extensive and well characterized⁴.

Engineered RNA molecules serve as a more versatile, rationally programmable alternative to protein-based scaffolding strategies, allowing a new level of access and control over spatial organization of proteins¹. The scaffold's primary sequence can be rationally designed to control both the distance and orientation between bound proteins, as well as their stoichiometry and the size of the overall complexes. Engineered RNA scaffolds are widely applicable in basic research, nanotechnology and biotechnology.

Natural and engineered ncRNA

Natural ncRNA molecules derive their diverse range of behaviors from their unique ability to fold into complex tertiary structures with recognition and even catalytic properties⁵. These properties in turn inspired the RNA hypothesis for the origins of life⁵, and much effort has been devoted to evolving novel RNA functions by SELEX^{2,3}. Recently, RNA molecules were rationally designed to perform specific tasks both *in vitro* and *in vivo*⁶⁻⁹, with scaffolding being a compelling new application¹. As our understanding about the causal relationship between primary sequence, secondary structure and function grows, RNA is now viewed as a modular molecule with extensive engineering potential.

Advantages and applications of synthetic RNA scaffolds

Scaffolding is widely used in nature. For example, multienzyme pathways are often physically and spatially organized onto clusters

through protein domain interactions¹⁰, microcompartments^{11,12} or natural RNA scaffolds¹³⁻¹⁵. Spatial organization helps direct substrate flow between interacting enzymes, limiting cross-talk and increasing the yields of sequential metabolic reactions¹⁶⁻²².

Previously, synthetic protein scaffolds were developed to direct flux in synthetic metabolic pathways (e.g., improving titers of mevalonate in *E. coli*¹⁷). These scaffolds were built by fusing three eukaryotic protein-protein interaction domains (PDZ, SH3 and GBD) and coexpressing proteins to be scaffolded as fusions with their cognate binding domains. This strategy allows for the localization and stoichiometric control of a limited number of proteins and was successfully applied to improve yields of hydrogen and glucaric acid synthesis in *E. coli*^{20,21}. Plasmid DNA was also recently used for scaffolding and improving the titer of resveratrol, 1,2-propanediol and melvanoate²².

The attraction of RNA scaffolds is their ability to be rationally programmed using the rules of base pairing. This offers access to larger scaffolds, in which hundreds of proteins are gathered to work together, and confers the ability to control not only stoichiometry but also the distance and orientation between interacting proteins. With hundreds of different, orthogonal, characterized aptamer domains⁴, and thus an expansive range of different binding domains, RNA scaffolds may bring together large complex pathways.

RNA scaffolds are applicable in many areas in which the spatial organization of biomolecules is desirable. In the synthetic biology realm, RNA scaffolds bring an added level of control by offering a tunable platform to control the spatial organization of proteins¹. However, current tested designs are limited to the assembly of two distinct aptamers, but can be, in principle, enlarged either by using the discrete system (see Experimental design) or by mixing different aptamers in given ratios. Scaffold and assembly size can also be further controlled by adding nonpolymerizing 'ends' or by using the discrete system as described in the Experimental design. Furthermore, RNA scaffold libraries will also be expanded by applying future advances in the RNA nanotechnology field²³⁻²⁵ or by inspiration from naturally existing structures (e.g., Dsra RNA¹⁴), as well as by designing directed evolution selections for functional ncRNA-mediated assemblies.

PROTOCOL

Experimental design

The general workflow for the design, induction and analysis of RNA scaffold expression is illustrated in **Figure 1**.

RNA scaffold design: RNA Designer (Steps 1–8). RNA scaffold design and optimization uses RNA folding software that relies on a free energy minimization algorithm to predict sequences that are amenable to *in vivo* scaffolding (**Table 1**). RNA Designer²⁶ (<http://www.rnasoft.ca/cgi-bin/RNAsoft/RNAdesigner/rnadesign.pl>), for example, can be used to design the primary sequence of an RNA molecule that folds into a desired secondary structure (**Fig. 2**).

This software computes an RNA sequence that folds into the specified secondary structure given a number of optional sequence constraints. In the present context, the sequence constraints to be specified consist of the sequence of aptamers, terminator and restriction enzyme sites that will be used. The secondary structure specified depends on the RNA folding scheme to be used. The simulation should be run at physiological temperature (e.g., 37 °C for *E. coli*) and with a target GC percentage matching the genomic content of the organism in which the RNA will be expressed (e.g., about 51% for *E. coli*).

RNA scaffold design: choosing aptamers to tether proteins to the RNA scaffolds (Steps 1 and 2). To choose the aptamers for the scaffold to be designed, a number of parameters should be considered: (i) the binding affinity between the aptamer and its binding (adaptor) protein should be as high as possible (e.g., nanomolar range); (ii) the binding should also be as specific as possible, leading to mutually orthogonal aptamers; and (iii) the binding protein sequence should be optimized for bacterial expression in terms of codon use and stability.

One such set of mutually orthogonal and extensively studied aptamers are those from MS2 and PP7 bacteriophages^{27,28} with dissociation constants of about 82 nM (F6 aptamer) and 1 nM, respectively^{29,30} (**Fig. 3**). This set of two aptamers will enable repetitive scaffolding of two different proteins from a chosen pathway (e.g., [Fe-Fe] hydrogenase and ferredoxin¹).

RNA scaffold design: choosing a folding scheme (Steps 3–8). RNA scaffolds can be expressed *in vivo* as discrete (**Fig. 4a,b**) or polymerizing molecules (**Fig. 4c**). The latter requires a more complex approach to both design and characterization but allows for more complex architectures and RNA-RNA interactions to be studied *in vivo* (see Step 3B). Discrete scaffolds are easier to engineer and can also be used as tags for mRNA expression studies^{31,32}.

For metabolic engineering purposes, the scaffolds should be modular in length and number of docking sites, as well as easy to characterize. For this application, we suggest using an initial discrete scaffold. This ncRNA contains multiple copies of the chosen aptamer flanked by spacers that define the relative distance and orientation between the folded aptamers, as successfully used in previous work aiming at mRNA tagging^{31,32} (**Fig. 4a**). This design can be serially cloned to reach a desired scaffold length (typically 96-mer or more; **Fig. 4b**).

To achieve more complex scaffold geometries, scaffolds can instead be made of polymerizing RNA molecules. This relies on the molecular cross-assembly of RNA strands on the basis of principles from the toolbox of RNA and DNA nanotechnology, including symmetry and kinetic considerations, as well as Watson-Crick and

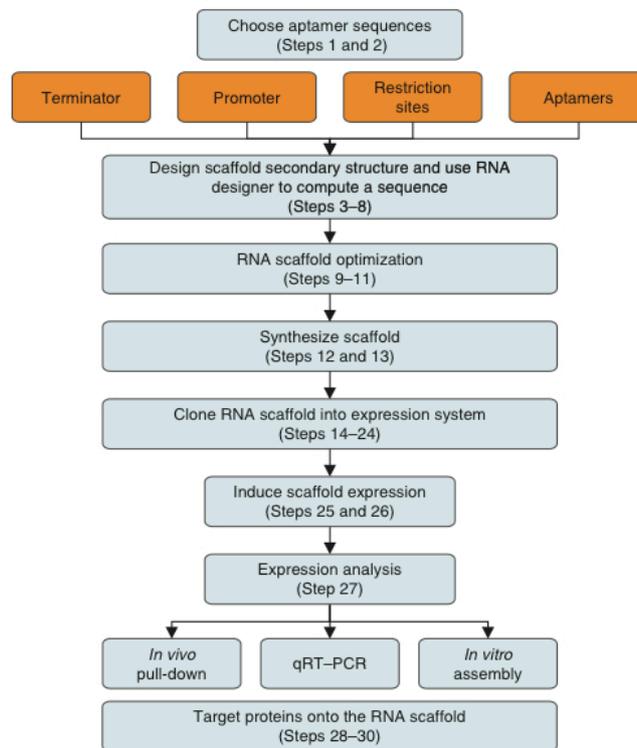


Figure 1 | General workflow for design, induction, expression and experimental testing of the RNA scaffold. Orange boxes represent sequence constraints to be input into RNA Designer.

noncanonical interactions^{24,33,34}. Polymerizing scaffolds are made of short RNA strands cross-polymerizing to create extended structures. Much exploration remains to be done to better understand and further expand the library of assembling RNAs and assembly schemes. Here, we give recommendations on how to use both symmetry and kinetic assembly, which are key to our published assembly scheme¹.

Our polymerizing scaffold design strategy¹ relies on two principles. First, the use of palindromic sequences minimizes the number of different interacting strands necessary to form the extended structures. By using sequence symmetry, it is possible to design nanotubes or two-dimensional sheets with only one or two different polymerizing strands, respectively^{24,35,36}. Second, kinetic assembly pathways³⁷ enable the assembly to occur isothermally. In our design, RNA molecules assemble in a two-step process in which all noninteracting regions are locked into metastable assembly-intermediate hairpins. They only unfold upon cross-interaction through their dimerization region, leading to the polymerization of the RNA into extended structures¹.

RNA scaffold optimization (Steps 9–11). Often a given sequence output from RNA Designer folds into a number of alternative secondary structures. To further optimize the structure, these should be analyzed using another RNA secondary structure program such as mfold³⁸ or NUPACK³⁹ (**Table 1**). By screening the top RNA Designer outputs, one should pick the sequence that optimizes the thermodynamic gap between the desired folding and the next most favorable structure. Finally, the candidate sequences should be

TABLE 1 | List of useful online RNA design tools.

| Tool | Developer | Website | Summary |
|-----------------------|---|---|--|
| mfold | University of Albany | http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form | RNA folding software; folding temperature and ionic conditions are fixed |
| NUPACK | California Institute of Technology | http://www.nupack.org/ | RNA software suite for design and folding analysis with the option of designing RNA reaction pathways |
| RNA Designer | University of British Columbia | http://www.rnasoftware.ca/cgi-bin/RNAsoft/RNAdesigner/rnadesign.pl | RNA design tool using the dot-bracket format; temperature and GC content are adjustable |
| RBS Calculator | Penn State University | https://salis.psu.edu/software/ | Predicts translation initiation rate in bacteria; takes into account RNA secondary structures for predictions |
| Nucleotide BLAST | National Center for Biotechnology Information | http://blast.ncbi.nlm.nih.gov/Blast.cgi | BLAST compares nucleotide sequences to sequence database and calculates the statistical significance of any match |
| Primer-BLAST | National Center for Biotechnology Information | http://www.ncbi.nlm.nih.gov/tools/primer-blast/ | Uses the popular primer3 engine to design primers; results are submitted to BLAST to check for unwanted endogenous match |
| BioNumbers | Harvard Medical School | http://bionumbers.hms.harvard.edu/ | Registry of useful biological numbers, including genomic GC contents |
| genom ^{PLUS} | Biogazelle | http://www.biogazelle.com/genomplus/ | Algorithm to determine the most stable reference genes from a set of tested candidate reference genes in a given qPCR sample panel |

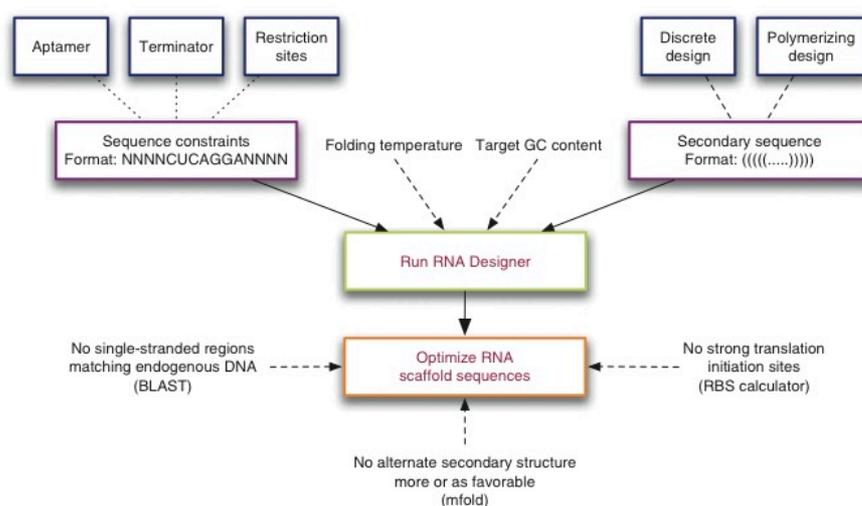
screened against any ribosome binding sites using an RBS calculator⁴⁰ (<https://salis.psu.edu/software/>) in order to avoid translation and against complete match of free single-stranded regions with any endogenous mRNA using the Basic Local Alignment Search Tool (BLAST) server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The workflow for RNA scaffold design and optimization is illustrated in **Figure 2**.

Cloning the designed RNA scaffold into an expression system (Steps 12–24). We have taken a modular cloning approach in which the key elements—promoters, assembly tags, aptamer-protein binding sites and terminators—are separated by unique restriction

sites to allow for tuning of the fundamental properties of the scaffold through quick and precise changes (**Fig. 5a**). This cloning strategy allows users to switch between the two different scaffold architectures: (i) the concatenation of the aptamer module to create large discrete scaffolds and (ii) assembly by polymerization of discrete ncRNA molecules (**Fig. 5b**). Once the construct is verified, it is cloned into commercial plasmids (e.g., pETDuet-1, EMD Chemicals) for *in vivo* expression.

Expression systems (Steps 25–27). For initial characterization of a candidate RNA scaffold, we recommend using a T7 expression system. This well-characterized expression system can be used for

Figure 2 | Designing and optimizing RNA scaffolds using RNA Designer. The design process starts with choosing both the sequence constraints and the secondary structure folding scheme. Parameters to input are linked through dashed lines: for the sequence constraints, aptamer, terminator and restriction enzyme site sequences need to be input, whereas the secondary structure should be input in dot-bracket format and will vary according to whether a discrete or polymerizing folding scheme is chosen. Run RNA Designer taking into account the folding temperature and a target GC content. Optimizing the RNA scaffold sequence output is a crucial part of the design process, and key parameters to consider are annotated with dashed lines. Bioinformatic programs used to check these parameters are shown in brackets.



PROTOCOL

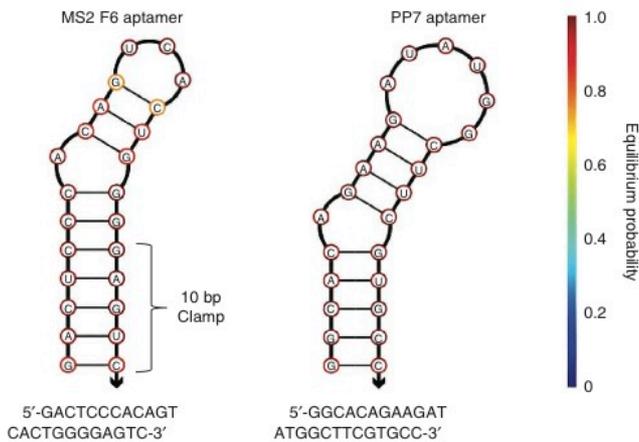


Figure 3 | Compatible aptamer pair. Left, primary sequence and secondary structure of the MS2 F6 aptamer and PP7 aptamers. Both aptamers are orthogonal to each other and are usable for RNA scaffolds. A 10-bp clamp (as indicated on figure) is added to the stem of the 14-bp MS2 F6 aptamer to adjust its length to that of the PP7 aptamer. Right, visualization with NUPACK³⁵: the color-coded scale depicts the base-pairing probabilities.

preliminary *in vitro* expression studies using commercially available IVT systems (e.g., MEGascript T7 Kit from Invitrogen), and can be tuned for a wide range of expression levels, yielding very high expression levels upon full induction. We thus suggest cloning the RNA scaffold expression cassette into a pET-based vector (e.g., pETDuet-1, EMD Chemicals) and transforming it into BL21 DE3 *E. coli* cells. Scaffold concentration is tunable *in vivo* by using different amounts of the inducing molecule (e.g., isopropyl β -D-1-thiogalactopyranoside (IPTG) or arabinose). A number of compatible T7-based plasmids have been developed and can be used for the simultaneous expression of up to six different proteins in addition to the RNA scaffold (e.g., EMD Chemicals Duet Vectors system). It should be noted that the T7 system is the only expression system tested in ref. 1, and other systems may also be useful.

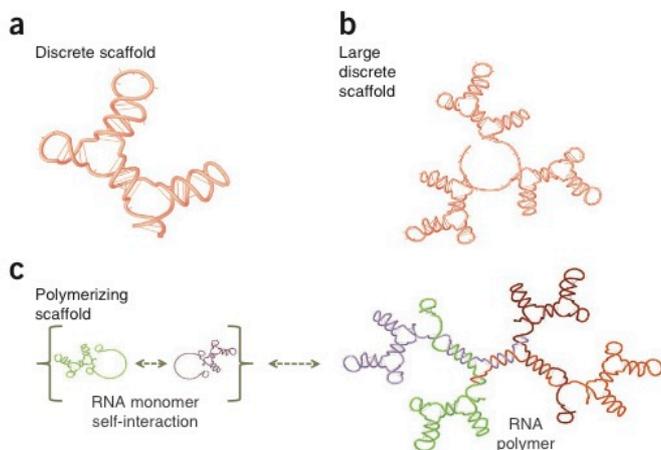


Figure 4 | Examples of RNA scaffold folding scheme designs. (a) Computed secondary structure of a small discrete scaffold comprising two aptamer-protein binding sites. (b) Secondary structure of an extended discrete scaffold expressed as a repeat of the small discrete scaffold. (c) Proposed secondary structure for a polymerizing scaffold. Small RNA modules interact through single-stranded domains and polymerize into the multimeric scaffold. Visualization with NUPACK³⁵.

DNA capture probes for scaffold expression analysis by *in vivo* pull-down assay (Step 27B(iii)). The DNA capture probe has two functional regions with well-defined melting temperatures (T_m). The binding domain is designed to interact with an ~ 10 -bases-long constant region that is single stranded and nonassembled (e.g., the region between RE4 and the start of the terminator, see Fig. 5). The annealing temperature of this region should be ~ 20 °C. The release region of the DNA capture probe should be designed to be about 10 bases. These extra 10 bases should be designed so that they do not form hairpins with the binding domain and bring the T_m of the 20 bases full DNA capture probe to ~ 35 °C. The probe should then be ordered as a 5'-biotinylated oligo.

Primer design for quantitative reverse-transcription PCR (qRT-PCR; Step 27C(iii)). qRT-PCR of small RNAs with complex secondary structures can be complicated. One strategy used by commercial kits is to add a poly-A tail to the RNA scaffold during retrotranscription⁴¹ (see Step 27C(ii)). This provides a specific site to anneal a reverse primer for the qPCR. A specific forward primer is equally important and we recommend designing three of them with a dedicated software program using a target T_m of 60 °C (e.g., primer-BLAST; Table 1), hybridizing between the RE4 and the end of the terminator stem and assessing their performance by incorporating a melting curve analysis step at the end of the PCR program (refer to instrument manual for specific programming). Choose a primer giving a clear single peak in the melting curve graph.

Stability considerations. Different strategies can be considered to enhance the stability of the RNA scaffold. RNA turnover is a natural component of gene expression. Half-lives of most bacterial RNAs range from 40 s to 60 min (ref. 42). Steady-state transcript concentrations are a result of degradation and synthesis rate⁴³. Strategies to enhance RNA scaffold level include the use of highly efficient expression systems (e.g., T7-based expression system), but also the implementation of design considerations to minimize decay.

In *E. coli*, 5'- and 3'-end accessibility is of particular importance in initiating the decay process. For example, RNase E requires

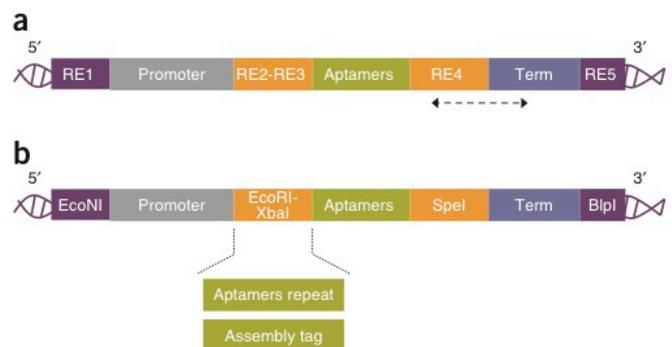


Figure 5 | Modular cloning approach to the design of RNA scaffolds. (a) Separation of promoter (gray box), aptamer (green box) and terminator (term; blue box) sequences by five unique restriction sites (RE; purple and yellow boxes). The dotted arrow indicates the suggested region to hybridize probes for the pull-down assay (see Step 27B(iii)). (b) Suggested restriction enzymes: RE1 = EcoNI and RE5 = BlnI enable cloning into the pETDuet-1 vector (EMD Chemicals); RE2 = EcoRI, RE3 = XbaI and RE4 = SpeI enable repeated cloning of the aptamer domain to create large discrete scaffolds or the cloning of an assembly tag (see Step 3B(i)) to create polymerizing RNA molecules.

5' single-stranded RNA of at least four nucleotides in length for efficient binding, whereas RNase II and PNPase are unable to bind substrates with fewer than six to ten unpaired bases at the 3' end⁴³. Therefore, minimizing single-stranded regions and locking both ends of the RNA scaffold with hairpins is an important part of the design strategy⁴⁴ (Fig. 4). In addition, using RNase E knockout strains for expression (e.g., BL21 DE3 Star, Invitrogen) may also help further stabilize the RNA transcripts⁴³.

Targeting proteins onto the RNA scaffold (Steps 28–30). The chosen aptamer-binding proteins are fused to the proteins to be scaffolded. For this step, we recommend exploring the space of possible linker length and fusion orientation so as to optimize both

scaffolding and protein interactions. It is also useful to examine three-dimensional structures of the protein domains to be used to determine whether an N or C terminus can be used as a fusion junction without interfering with the function of the protein.

Controls. Impairing RNA polymerization (Step 27A(iii)): short DNA oligos are designed to match and hybridize to the dimerization region of the polymerizing scaffold to prevent RNA polymerization. qRT-PCR control (Step 27C(v)): *GapA* or *MreB* *E. coli* housekeeping genes¹ are used as internal references to assess the relative RNAs given their scaffold concentration after induction. Impairing protein binding (Step (30)): RNA scaffolds with aptamers mutated to poly-T and poly-A sequences are designed to prevent protein binding.

MATERIALS

REAGENTS

▲ **CRITICAL** All solutions and buffers should be kept sterile and be stored according to the manufacturer's recommendation.

Chemicals and solvents

- DNA ladder, 1 kb Plus (Invitrogen, cat. no. 10787-018)
- TBE-urea gels (6% (wt/vol), Novex gels; Invitrogen, cat. no. EC68652BOX)
- Acetic acid (Fisher Scientific, cat. no. A35-500) ! **CAUTION** Acetic acid is corrosive. Avoid inhalation and exposure to skin and eyes.
- Ampicillin (Sigma-Aldrich, cat. no. A9518)
- Bacterial peptone (Fisher Scientific, cat. no. BP1420-2)
- Bacterial protein extraction reagent (Pierce, cat. no. 78243)
- ddH₂O, sterile
- Dithiothreitol (Invitrogen, cat. no. 15508-013)
- Dynabeads M-280 streptavidin (Invitrogen, cat. no. 112-05D)
- EDTA (Invitrogen, cat. no. AM9261)
- Ethidium bromide (1% (wt/vol) solution; Sigma-Aldrich, cat. no. E-8751)
- ! **CAUTION** Ethidium bromide is mutagenic; always use nitrile gloves.
- Glycerol (Sigma-Aldrich, cat. no. G5516-100ML)
- Isopropyl β-D-1-thiogalactopyranoside (IPTG; Anatrace, cat. no. I1002)
- Lysozyme (Sigma-Aldrich, cat. no. L6876-5G)
- Potassium chloride (KCl; Sigma-Aldrich, cat. no. P5405)
- SeaKem LE Agarose (Lonza Cologne, cat. no. 50002)
- Sodium chloride (Fisher Scientific, cat. no. S671-500)
- Tris (tris(hydroxymethyl)aminomethane; Fisher Scientific, cat. no. BP152-5)
- Yeast extract (VWR, cat. no. 97063-370)
- Luria-Bertani broth (LB; Sigma-Aldrich, cat. no. L2542-500ML)

Bacterial strains and vectors

- *E. coli* cloning strain (Turbo competent *E. coli*; NEB, cat. no. C2984H)
- *E. coli* expression strain (One Shot BL21 Star (DE3) chemically competent *E. coli*; Invitrogen, cat. no. C6010-03)
- pETDuet-1 T7 expression vector (EMD Chemicals, cat. no. 71146-3)
- pACYCDuet-1 T7 expression vector (EMD Chemicals, cat. no. 71147-3)
- pCOLADuet-1 T7 expression vector (EMD Chemicals, cat. no. 71406-3)
- pCDFDuet-1 T7 expression vector (EMD Chemicals, cat. no. 71340-3)

Kits

- MEGAscript T7 kit (Invitrogen, cat. no. AM1334M or Applied Biosystems)
- NCode VIL0 miRNA cDNA synthesis kit (Invitrogen, cat. no. A11193-050)
- ▲ **CRITICAL** This kit allows for the detection and quantification of small RNAs by qRT-PCR by adding a poly-A tail before the retrotranscription step.

- QIAprep spin miniprep kit (Qiagen, cat. no. 27104)
- QIAquick gel extraction kit (Qiagen, cat. no. 28704)
- SYBR Green Supermix (Invitrogen, cat. no. 4309155)
- Total RNA purification 96-well kit (Norgen, cat. no. 24300)
- ▲ **CRITICAL** This column-based purification kit allows for the purification of total RNA without a size cutoff.

Enzymes

- Restriction endonucleases: EcoNI and BlnI (Fermentas, cat. nos. FD1304 and FD0094, respectively)
- T4 DNA ligase (Promega, cat. no. M1801)

Buffers

- Assembly buffer: 150 mM KCl, 0.1 mM dithiothreitol, 0.1 mM EDTA, 10 mM Tris, in ddH₂O; the buffer is made fresh and adjusted to pH 7.4

Primers and probes

- PCR primers (Integrated DNA Technologies) ▲ **CRITICAL** All PCR primers should be designed using appropriate software (e.g., primer-BLAST; Table 1), in order to avoid the formation of self- or heterodimers, as well as to avoid complementarity to genomic DNA (in order to avoid nonspecific amplifications). Primers should be diluted to 100 μM and kept at –20 °C until use.
- DNA oligos (Integrated DNA Technologies)
- Biotinylated DNA capture probes (Integrated DNA Technologies)

EQUIPMENT

- Asylum MFP-3D atomic force microscope (Asylum)
- Cuvettes (Semi-Micro cuvettes; BrandTech Scientific, cat. no. 2711010)
- Eppendorf centrifuge (Eppendorf AG, cat. no. 5702R)
- Eppendorf Mastercycler ep realplex (Eppendorf, cat. no. 6302 000.601)
- Etched silicon cantilever (Olympus, cat. no. OMCL-AC160TS)
- Highest-grade V1 mica, 25 × 76 mm (Ted Pella, cat. no. 56)
- Incubation facilities for bacterial culture
- Liquid nitrogen
- Spectrophotometer (Ultraspec 3100 pro; GE Healthcare, cat. no. 80-2112-32)

Bioinformatics RNA Design tools (see Table 1)

- RNA Designer (ref. 33—<http://www.rnasoft.ca/cgi-bin/RNAsoft/RNAdesigner/rnadesign.pl>)
- mfold (ref. 34—<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>)
- NUPACK (ref. 35—<http://www.nupack.org/>)
- RBS calculator (ref. 36—<https://salis.psu.edu/software/>)
- Biogazelle geNorm Plus (<http://www.biogazelle.com/genormplus/>)

PROCEDURE

RNA scaffold design ● TIMING 2 h

1| Start by choosing an appropriate set of aptamers (e.g., MS2 and PP7, see Experimental design and Fig. 3).

2| If the aptamer stem loops are of different lengths, add complementary bases at both sides of the shorter stem's root in order to equilibrate their length (see Fig. 4; use RNA Designer).

probability of alternative secondary-structure formation by comparing the free energy of predicted alternative structures with that of the desired structure (see Experimental design).

10| Select the most stable primary sequence with the desired secondary structure, which optimizes the thermodynamic gap between the desired folding and the next most favorable structure.

? TROUBLESHOOTING

11| Confirm that the selected primary sequence is lacking any strong ribosome binding sites using, for example, the RBS calculator web server (**Table 1**).

Cloning the designed RNA scaffold into an expression system ● TIMING 12–15 d

12| Add the features enabling expression and cloning to the selected primary sequence from Step 11: the cloning site (RE1) and full promoter on the 5' end, and the chosen cloning site (RE5) after the terminator on the 3' end.

13| Order the full construct as a synthetic gene cloned into a high-copy cloning vector.

? TROUBLESHOOTING

14| By using standard protocols⁴⁵, transform competent bacterial cells (e.g., heat shock-competent *E. coli* Turbo cells from NEB) with about 0.1 ng of the high-copy plasmid containing the scaffold.

15| Spread the transformed bacteria on LB plates (100–150 μ l of bacterial suspension per plate) containing the appropriate antibiotic for selection (e.g., 100 mg ml⁻¹ ampicillin for high-copy-number *E. coli* plasmids of the pUC or pBluescript series) and incubate them overnight at 37 °C.

16| Inoculate a colony into an individual aliquot of 5 ml of LB medium supplemented with the correct antibiotic (e.g., 100 μ g ml⁻¹ ampicillin) and incubate the culture overnight at 37 °C.

17| Isolate the plasmid using a plasmid isolation kit (e.g., Qiagen miniprep kit) according to the manufacturer's instructions.

18| Digest the plasmid using a suitable combination of restriction enzymes flanking the cassette (e.g., Fermentas FastDigest EcoNI and BspI restriction enzymes; **Fig. 5b**) according to the manufacturer's instructions.

19| Separate the restriction enzyme-digested samples by electrophoresis in 1.5–2% (wt/vol) agarose gels, using standard protocols⁴⁵. Verify the fragment size by using a suitable DNA ladder.

20| Purify the cassette-containing DNA using an agarose gel purification kit according to the manufacturer's instructions.

? TROUBLESHOOTING

21| Digest the chosen expression vector (e.g., EMD Chemicals pETDuet-1 vector) with restriction enzymes corresponding to the 5' and 3' overhangs of the cassette (e.g., Fermentas FastDigest EcoNI and BspI restriction enzymes) according to the manufacturer's instructions.

22| Ligate the purified cassette from Step 20 into the digested expression vector from Step 21, using a T4 DNA ligase (e.g., Promega T4 DNA ligase) using standard protocols⁴⁵.

23| Transform into competent bacterial cells (e.g., Invitrogen heat shock-competent *E. coli* BL21 DE2-star for the suggested T7-based expression system) using standard protocols⁴⁵.

24| Identify correct clones by running purified and digested plasmids in a gel electrophoresis to confirm insert size corresponding to your scaffold using standard protocols⁴⁵. This should be further confirmed by sequencing.

? TROUBLESHOOTING

■ PAUSE POINT Correct clones can be kept as glycerol stocks at –80 °C.

Induction of RNA scaffold expression ● TIMING 1–2 d

25| Inoculate one colony of the scaffold-bearing strain from Step 23 into liquid growth medium (e.g., LB or a minimal growth medium for *E. coli*) with the appropriate antibiotic (e.g., 100 mg ml⁻¹ ampicillin) in a 5-ml culture tube.

PROTOCOL

26| After overnight culture at 37 °C, make a 1:20 dilution of the culture with fresh LB medium and the appropriate antibiotic and let it grow until an optical density (OD) of ~0.3 at 600 nm (as measured by a spectrophotometer) is reached (about 2 h, see ref. 45) and induce scaffold production with the appropriate molecule (e.g., with 0.1–1 mM IPTG).

▲ CRITICAL STEP Induction conditions are crucial. Cells should be incubated at the temperature at which the RNA scaffold was designed (i.e., 37 or 30 °C). Induce the cells when they reach the beginning of exponential phase (i.e., OD 0.2–0.3).

Expression analysis

27| Proceed with expression analysis using 27 or more of options A–C, depending on whether the RNA scaffold is designed to polymerize. Option A is suitable for studying RNA polymerization *in vitro* and relies on the choice of a T7, T3 or SP6 promoter in the design. Option B is suitable for purifying *in vivo*-produced RNA samples or RNA-protein complexes. It is a modified pull-down assay in which cross-linking is not necessary and biological samples are not denatured. Option C enables the precise quantification of *in vivo* RNA scaffolds.

(A) *In vitro* assembly ● TIMING 2–3 d

(i) Set up an *in vitro* transcription reaction using an appropriate kit (e.g., MEGAscript T7 kit, Applied Biosystems). Use 500 ng of linearized purified plasmid (e.g., digested with Fermentas FastDigest PstI and purified with the QIAquick PCR purification kit, according to the manufacturers' protocols) containing the RNA scaffold expression cassette from Step 24. Perform *in vitro* transcription overnight at 37 °C according to the manufacturer's instructions.

▲ CRITICAL STEP Circular plasmid templates will generate extremely long heterogeneous RNA transcripts because RNA polymerases are processive and rho-independent terminators may not provide efficient termination *in vitro*. Thus, it is preferable to digest the plasmid with an appropriate restriction enzyme downstream of the RNA scaffold cassette. It is worthwhile to also gel-purify the linearized plasmid.

(ii) Purify the *in vitro*-transcribed RNA scaffold using the Norgen RNA purification kit according to the manufacturer's instructions.

■ PAUSE POINT The purified *in vitro*-transcribed RNA scaffold may be stored at –20 °C for 2–3 d. It is recommended that samples be placed at –80 °C for long-term storage.

(iii) Deposit ~25 ng of the RNA onto freshly cleaved mica (2.0 cm²), and allow it to dry for 20 min at the appropriate assembly temperature (e.g., 30 or 37 °C). As an appropriate negative control, allow for assembly with a tenfold molar excess of DNA oligos designed to impair RNA interactions. These oligos should be fully complementary to the interacting regions of the RNA molecules (e.g., dimerization region in **Fig. 6**).

(iv) Visualize the presence of RNA assemblies (e.g., RNA polymers of length 10–100 nm) using an atomic force microscope (e.g., Asylum MFP-3D). Perform acquisition in air, at room temperature (~20 °C), using an etched silicon cantilever with a resonance frequency of ~300 kHz, a spring constant of ~42 N m⁻¹ and a tip radius of ~10 nm.

(B) *In vivo* pull-down assay ● TIMING 2–3 d

(i) Take 500 ml of induced cells from Step 26 and centrifuge the cells at 5,000*g* for 5 min at room temperature.

(ii) Remove the supernatant and place the cells on ice. Perform cell lysis by adding 200 µl of ice-cold bacterial protein extraction reagent and gently pipette up and down until the cell suspension is homogenous. Leave the suspension on ice for 5 min.

(iii) Add 1 nmol of a DNA-biotinylated capture probe designed to specifically hybridize with the scaffold (see Experimental design) and allow for interaction with the RNA scaffold at room temperature for 5 min.

(iv) Add 1 mg of streptavidin-coated Dynabeads M-280 and allow for interaction with the sample for 5 min on ice.

(v) Place the lysate on a magnet and rinse the beads twice with ice-cold assembly buffer made fresh.

(vi) Resuspend the beads in 10 ml of assembly buffer complemented with 10 nM of a DNA probe fully complementary to the capture probe. Incubate at 37 °C for 5 min with intermittent gentle hand shaking.

(vii) Run on a 6% (wt/vol) TBE-urea gel and visualize the RNA scaffold with ethidium bromide staining according to the manufacturer's instructions. The purified scaffold should be visible as a clear single band.

? TROUBLESHOOTING

(C) *In vivo* expression analysis—qRT-PCR ● TIMING 2–3 d

(i) Purify RNA from 1 ml of induced cells from Step 26 and from 1 ml of noninduced cells from Step 23 using the Norgen kit according to the manufacturer's instructions.

(ii) Prepare cDNA from 0.5 mg of purified total RNA using the NCode VILO kit according to the manufacturer's instructions.

■ PAUSE POINT cDNA may be kept in the provided buffer at –20 °C.

(iii) To a 4.5-ml aliquot of tenfold-diluted cDNA in ddH₂O, add 3 ml of a 2 nM solution of a specific forward primer, the provided poly-A annealing reverse primer (according to the manufacturer's instructions) and 7.5 ml of SYBR Green Supermix.

(iv) Set up the PCR reaction using SYBR Green Supermix according to the manufacturer's instructions and perform the qPCR on an appropriate thermocycler (e.g., Eppendorf Mastercycler ep realplex) with the following program: 50 °C for 2 min, and then 95 °C for 5 min to activate the enzyme. The PCR cycles are then as follows: 15 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C, repeated 45 times.

- (v) Calculate the relative concentration of the RNA scaffold by using a stable mRNA internal reference (e.g., GapA or MreB for *E. coli*⁴⁶): assess total RNA samples from Step 27C(i) for the concentration of the internal reference RNA. Next, calculate the relative concentration of each RNA sample (induced versus noninduced from Step 27C(i)) relative to the mRNA internal reference by using the Biogazelle qPCR software geNorm Plus (**Table 1**) according to the manufacturer's instructions.

▲ **CRITICAL STEP** Good-quality, stable housekeeping genes are essential for high-quality RNA quantification by qRT-PCR⁴⁶. We recommend designing primer pairs for at least three candidate housekeeping genes and assessing their stability in the assay conditions using dedicated software (e.g., geNorm; **Table 1**).

? TROUBLESHOOTING

Targeting proteins onto the RNA scaffold ● TIMING 2–3 d

28 | Make fusion proteins between the chosen aptamer proteins and the proteins to be scaffolded according to standard protocols (e.g., using the biobrick cloning system; see <http://dspace.mit.edu/handle/1721.1/32535>).

29 | Clone the genes into compatible expression plasmids (e.g., Duet vector, EMD Chemicals) with appropriate restriction enzymes according to standard protocols⁴⁵.

30 | Co-transform the RNA scaffold plasmid from Step 23 alongside the protein-coding plasmid from Step 29 into competent bacterial cells (e.g., heat shock-competent *E. coli* BL21 DE2-star). Appropriate negative controls here are RNA scaffolds with mutated aptamers designed to prevent protein binding (e.g., modifying PP7 or MS2 stem loops to a stretch of T and A bases). Plate the transformed bacteria, pick colonies and identify correct clones according to standard protocols⁴⁵.

■ **PAUSE POINT** Correct clones can be kept as glycerol stocks at $-80\text{ }^{\circ}\text{C}$ and aliquots can be used for further experiments.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

| Step | Problem | Possible reason | Solution |
|------|--|--|---|
| 8 | RNA Designer cannot find a sequence folding into the desired secondary structure | Sequence constraints are too restrictive | Extend the double-stranded region of the desired secondary structure and add more complementary bases using RNA Designer to promote its formation |
| | RNA Designer does not allow the design of symmetric sequence (((((()))))) | This function is not supported by the software | As exemplified in Figure 6b , 'trick' RNA Designer by designing a fake hairpin: add a few unpaired bases in between the symmetrical sequences and remove them after the results ((((((...)))))) Alternatively, you can use the NUPACK design package |
| 10 | Prediction of several alternative structures | The program computes not only the most favorable structure, but also less favorable structures with relatively lower free energy | Compare the different free energies. If necessary, pick another RNA Designer result |
| 13 | Gene-encoding RNA scaffold cannot be synthesized | Long sequence with complicated secondary structures and repeats | Divide the sequence into two genes. In the case of extended discrete scaffolds, a solution is to synthesize the discrete RNA module with biobrick restriction sites and serially clone it until desired scaffold size is reached (Fig. 5b) |
| 20 | Low DNA recovery | Fuzzy DNA band on the agarose gel | Increase the agarose concentration up to 3% (wt/vol) |
| 24 | No correct clone | Low ligation efficiency | To ensure maximum ligation efficiency, dephosphorylate the vector and adjust the molar ratio between vector and insert to approximately 1:3 |

(continued)

PROTOCOL

TABLE 2 | Troubleshooting table (continued).

| Step | Problem | Possible reason | Solution |
|----------|---------------------------|--------------------------|--|
| 27B(vii) | No RNA scaffold recovered | Low pull-down efficiency | Screen for a number of DNA capture probes |
| | | RNase contamination | Make sure to use standard practice when working with RNA. Use gloves, filter tips and a decontaminated workspace |
| 27C(v) | Poor qPCR readings | Nonspecific primers | Check the primer melting curve. Design and try a new pair of primers |

● TIMING

Steps 1–8, RNA scaffold design: 2 h

Steps 9–11, RNA scaffold optimization: 1 h

Steps 12–24, cloning the designed RNA scaffold into an expression system: 12–15 d (Step 13 takes most of this time, typically 5–10 business days)

Steps 25 and 26, induction of RNA scaffold expression: 1–2 d

Step 27, expression analysis: 2–3 d

Steps 28–30, targeting proteins onto the RNA scaffold: 2–3 d (typically Steps 28 and 29 can be done while waiting for the scaffold synthesis in Step 13)

ANTICIPATED RESULTS

The described protocol results in a number of RNA scaffolds that are cloned and expressed *in vivo* and are capable of binding target proteins with prominent affinity and selectivity. Both polymerizing and discrete scaffolds can be designed, although, as discussed above, characterization is slightly more complex for polymerizing scaffolds as assembly efficiency should be assessed isothermally *in vitro* first (Step 27A). *In vitro* assembling candidates can then be evaluated for *in vivo* assembly and scaffolding efficiency by purifying them from cells (Step 27B). Polymerizing RNA scaffolds can be expected to reach tens of nanometers in size and gather hundreds of proteins. Discrete scaffolds can also gather hundreds of proteins depending on design choice (i.e., the number of aptamers). Aptamer occupancy was estimated to be of at least 70% for a repeat of 96 MS2 aptamers by Golding *et al.*³¹. It is expected to be equal or higher here because of the complementary 3' region stabilizing the aptamers (Step 3A(iii)). Finally, depending on the expression system, fully induced cells can be expected to produce tens of thousands of RNA molecules without excessive metabolic burden¹.

ACKNOWLEDGMENTS We are indebted to F. Aldaye who had a leading role in this work. This work was supported by the Enerbio-Tuck Foundation and the Institut Français du Pétrole Energies Nouvelles (to C.J.D.); by the Agence Nationale de la Recherche France, Institut National de la Santé et de la Recherche Médicale (Unité 1001)–Institut National de Recherche en Informatique et en Automatique projet d'envergure, and an Axa Foundation Chair on Longevity (to A.B.L.); and by the Wyss Institute for Biologically Inspired Engineering and support from the Department of the Army W911NF-09-1-00226 (to P.A.S.).

AUTHOR CONTRIBUTIONS All authors participated extensively in developing the protocol described in this paper.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nprot.2012.102>. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Delebecque, C.J., Lindner, A.B., Silver, P.A. & Aldaye, F.A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).
- Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
- Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
- Bunka, D.H.J. & Stockley, P.G. Aptamers come of age - at last. *Nat. Rev. Micro.* **4**, 588–596 (2006).
- Eddy, S.R. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919–929 (2001).
- Isaacs, F.J., Dwyer, D.J. & Collins, J.J. RNA synthetic biology. *Nat. Biotechnol.* **24**, 545–554 (2006).
- Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
- Callura, J.M., Dwyer, D.J., Isaacs, F.J., Cantor, C.R. & Collins, J.J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc. Natl. Acad. Sci. USA* **107**, 15898–15903 (2010).
- Isaacs, F.J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841–847 (2004).
- Conrado, R.J., Varner, J.D. & DeLisa, M.P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* **19**, 492–499 (2008).
- Savage, D.F., Afonso, B., Chen, A.H. & Silver, P.A. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* **327**, 1258–1261 (2010).

12. Burack, W.R. & Shaw, A.S. Signal transduction: hanging on a scaffold. *Curr. Opin. Cell Biol.* **12**, 211–216 (2000).
13. Zappulla, D.C. & Cech, T.R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl. Acad. Sci. USA* **101**, 10024–10029 (2004).
14. Cayrol, B. *et al.* A nanostructure made of a bacterial noncoding RNA. *J. Am. Chem. Soc.* **131**, 17270–17276 (2009).
15. Shevtsov, S.P. & Dunder, M. Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* **13**, 167–173 (2011).
16. Adam, G. & Delbrück, M. Reduction of dimensionality in biological diffusion processes. *Struct. Chem. Mol. Biol.* 198–215 (1968).
17. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
18. Lee, H., DeLoache, W.C. & Dueber, J.E. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* **14**, 242–251 (2012).
19. Park, S.-H., Zarrinpar, A. & Lim, W.A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061–1064 (2003).
20. Agapakis, C.M. *et al.* Insulation of a synthetic hydrogen metabolism circuit in bacteria. *J. Biol. Eng.* **4**, 3–15 (2010).
21. Moon, T.S., Dueber, J.E., Shiue, E. & Prather, K.L.J. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* **12**, 298–305 (2010).
22. Conrado, R.J. *et al.* DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* **40**, 1879–1889 (2011).
23. Shu, D., Moll, W.-D., Deng, Z., Mao, C. & Guo, P. Bottom-up assembly of RNA arrays and superstructures as potential parts in nanotechnology. *Nano Lett.* **4**, 1717–1723 (2004).
24. Guo, P. The emerging field of RNA nanotechnology. *Nat. Nanotechnol.* **5**, 833–842 (2010).
25. Afonin, K.A. *et al.* *In vitro* assembly of cubic RNA-based scaffolds designed in silico. *Nat. Nanotechnol.* **5**, 676–682 (2010).
26. Andronescu, M., Aguirre-Hernandez, R., Condon, A. & Hoos, H.H. RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res.* **31**, 3416–3422 (2003).
27. Chao, J.A., Patskovsky, Y., Almo, S.C. & Singer, R.H. Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* **15**, 103–105 (2008).
28. Convery, M.A. *et al.* Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat. Struct. Biol.* **5**, 133–139 (1998).
29. Lim, F., Downey, T.P. & Peabody, D.S. Translational repression and specific RNA binding by the coat protein of the *Pseudomonas* phage PP7. *J. Biol. Chem.* **276**, 22507–22513 (2001).
30. Parrott, A.M. *et al.* RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. *Nucleic Acids Res.* **28**, 489–497 (2000).
31. Golding, I. & Cox, E.C. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA* **101**, 11310–11315 (2004).
32. Valencia-Burton, M., McCullough, R.M., Cantor, C.R. & Broude, N.E. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat. Methods* **4**, 421–427 (2007).
33. Aldaye, F.A., Palmer, A.L. & Sleiman, H.F. Assembling materials with DNA as the guide. *Science* **321**, 1795–1799 (2008).
34. Seeman, N.C. An overview of structural DNA nanotechnology. *Mol. Biotechnol.* **37**, 246–257 (2007).
35. Liu, H., Chen, Y., He, Y., Ribbe, A.E. & Mao, C. Approaching the limit: can one DNA oligonucleotide assemble into large nanostructures? *Angew. Chem. Int. Ed. Engl.* **45**, 1942–1945 (2006).
36. Liu, H., He, Y., Ribbe, A.E. & Mao, C. Two-dimensional (2D) DNA crystals assembled from two DNA strands. *Biomacromolecules* **6**, 2943–2945 (2005).
37. Yin, P., Choi, H.M.T., Calvert, C.R. & Pierce, N.A. Programming biomolecular self-assembly pathways. *Nature* **451**, 318–322 (2008).
38. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
39. Zadeh, J.N. *et al.* NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
40. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
41. Benes, V. & Castoldi, M. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods* **50**, 244–249 (2010).
42. Selinger, D.W. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* **13**, 216–223 (2003).
43. Richards, J., Sundermeier, T., Svetlanov, A. & Karzai, A.W. Quality control of bacterial mRNA decoding and decay. *Biochimica et Biophysica Acta.* **1779**, 574–582 (2008).
44. Molinaro, M. & Tinoco, I. Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res.* **23**, 3056–3063 (1995).
45. Tolia, N.H. & Joshua-Tor, L. Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* **3**, 55–64 (2006).
46. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–12 (2002).



Supporting Material

The following supporting material is relevant to Chapter V.

1. Plasmid and strains

All cloning was done using BioBrick standard assembly (BB). Initial cloning was performed into the ampicillin-resistant BB vector pV₀₁₂₀ or pV_{0120-2X} [pV₀₁₂₀ into which we cloned a (Gly₄Ser)₂ linker (2X)], using Turbo cells (NEB), followed by their transfer into T7 BB-compatible duet expression vectors (pACYCDuet-BB, pCDFDuet-BB, pCOLADuet-BB). 2X means that the linker used contains two units of Gly₄Ser (i.e. it is Gly₄SerGly₄Ser). All RNA modules were cloned into pETDuet

(Novagen). Expression assays were conducted in BL21-star (DE3) cells, which possess a mutation in the RNaseE (*rne131*) gene to limit RNA degradation.

The T7 duet expression vectors pACYCDuet, pCDFDuet, pCOLADuet and pETDuet have respective copy numbers of 10^{-12} , 20^{-40} , 20^{-40} , and ~ 40 . This system was selected based on a contribution by Tolia and Joshua-Tor (Tolia & Joshua-Tor 2006). The pETDuet vector was reserved for expressing the RNA scaffolds because of its high copy number, while the proteins and the maturation factors were expressed using the relatively lower copy number vectors pACYCDuet, pCDFDuet and pCOLADuet.

2. RNA scaffolds: aptamer domains

The use of aptamer is discussed at length in previous Chapters. We used the SELEX MS2 F6, and the wild-type PP7 aptamers. Both bind their respective proteins well (dissociation constants in the lower nanomolar regime), and are orthogonal to each other (Lim et al. 2001; Valencia-Burton et al. 2007; Werstuck & Green 1998). The PP7's binding site is 26 bases long, while the MS2's binding site is 14 bases long. To make both aptamers of relatively similar length, we increased MS2's stem to 26 bases.

3. RNA scaffolds: synthesis

All of RNA modules were synthesized by IDT with a T7 promoter, a T7 terminator, and EcoNI/PstI cloning sites. In the case of d2' and d2'', both RNA strands were synthesized jointly enabling their expression from a single plasmid. These tailored expression systems were cloned into pETDuet to yield pCJDD₀, pCJDD₁ and pCJDD₂.

4. RNA scaffolds: mutations to prevent self-assembly

The trigger domains of d_1 and d_2' were mutated into poly-T stretches to prevent self-assembly, and were designated d_1T and $d_2'T$. Mutagenesis was done in two steps. In the first step, the d_1 construct was PCR amplified into the two overlapping halves $A(d_1T)$ and $B(d_1T)$. We used the primer sets AF / $A(d_1T)R$ and $B(d_1T)F$ / BR. This step introduces a poly-T stretch in place of the trigger domain. In the second step, the overlapping regions of $A(d_1T)$ and $B(d_1T)$ are annealed, and extended by PCR to generate the full length construct d_1T . d_1T is purified using Qiagen PCR purification kits, and cloned into the pETDuet vector for expression to form pCJDD $_1T$. Mutagenesis of d_2' into $d_2'T$ is similarly achieved and yielded plasmid pCJDD $_2T$. The two halves $A(d_2'T)$ and $B(d_2'T)$ are generated using the primer set AF / $A(d_2'T)R$ and $B(d_2'T)F$ / BR.

5. RNA scaffolds: mutations to prevent protein binding

The sites of mutagenesis were determined according to the crystal structure of each protein, and according to their interactions with RNA $_{2-4}$ (and references therein). Mutagenesis of d_1 was done in four steps. In the first step, the d_1 construct was PCR amplified into the two overlapping halves $A(d_1PP_7)$ and $B(d_1PP_7)$. We used the primer sets AF / $A(d_1PP_7)R$ and $B(d_1PP_7)F$ / BR. This step introduces the poly-T / poly-A stretch in place of the PP $_7$ binding domain. In the second step, the overlapping regions of $A(d_1T)$ and $B(d_1T)$ were annealed and extended by PCR to generate the full

length construct d1pre-mut. Mutagenesis of the d1 MS2 aptamer binding domain was achieved similarly using d1pre-mut (for the initial PCR amplification step) and AF / A(d1ms2)R and B(d1ms2)F / BR, yielding d1mut. d1mut was purified using Qiagen PCR purification kits, and cloned into the pETDuet vector for expression to form pCJDD1mut. Mutagenesis of d2' and d2'' into d2'mut and d2''mut was similarly achieved and yielded plasmid pCJDD2mut. We used the primer set AF / A(d2''pp7)R and B(d2'pp7)F / BR to modify the PP7 aptamer and AF / A(d2'''ms2)R and B(d2'''ms2)F / BR to modify the MS2 aptamer. domut was ordered as an ultramer-oligo and cloned into the pETDuet vector for expression to form pCJDDomut.

6. Atomic force microscopy

Atomic force microscopy was conducted using an Asylum MFP-3D. 10 μ L of the assembled RNA constructs were deposited onto freshly cleaved mica (2.0 cm²), allowed to dry for 20 minutes, and analyzed within 24 hours to minimize sample degradation. Data acquisition was performed in air, at room temperature, and using an etched silicon cantilever with a resonance frequency of ~300 kHz, a spring constant of ~42 N/m, and a tip radius of ~10 nm. Samples were either imaged for *in vitro* samples directly after *in vitro* transcription and RNA purification, or for *in vivo* samples directly after purification using our DNA-precipitation approach. In both cases, 0.05 ODs was loaded on 2.0 cm² of mica, allowed to dry to completeness, and imaged within 24 hours to minimize sample degradation.

7. RNA scaffolds: inhibitory strands

Inhibitory oligonucleotide strands that are 24 bases long, and that are complementary to the trigger domains of d1 and d2' were used to bind and inhibit the assembly of the one- and two-dimensional scaffolds. This occurs by preventing the initial formation of tile d1-1 and pro-tile d2-1. These strands are called d1In and d2'In, and were purchased from Integrated DNA Technology (IDT). In the case of *in vitro* transcription, we used 1 µg of either d1In or d2'In during the transcription process (this translates to an excess that is greater than 10-fold), and stopped the reaction after 16 hours. For the *in vivo* characterization of assemblies in the presence of the inhibitor strands, cells were lysed in the presence of excess d1In or d2'In.

8. PP7 / MS2 proteins

Targeted attachment of proteins onto our scaffolds was conducted using the protein/RNA aptamer sets MS2 and PP7. PP7 was obtained from Singer (Chao et al. 2007) as an aggregation deficient mutant (PP7 dIFG or dPP7), while MS2 was obtained from Golding (Golding et al. 2005) as a fused dimer of the wild type protein (MS2). Both genes were PCR-amplified using the primer sets MS2F/MS2R and PP7F/PP7R, which incorporate the BB restriction sites necessary for cloning into pV0120-0 or pV0120-2X.

9. Protein production levels

The level of HP and FM protein content inside cells was analyzed in cells expressing no scaffold, and in cells expressing D0, D1 and D2. As seen from the α -myc Western blot in Figure 49, comparable levels of FM/HP production are observed in all cases.

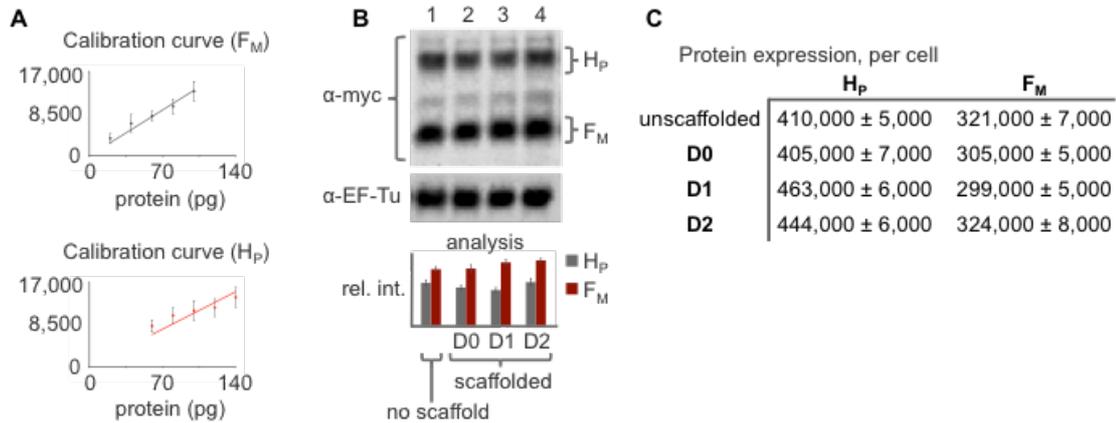


Figure 49: Protein production levels. (A) Calibration curves for FM and HP. (B) α -myc Western blot analysis of the relative amounts of FM and HP. Lane 1 corresponds to no scaffolds, and lanes 2-4 correspond to cells co-expressing D0, D1 and D2, respectively. (C) Protein levels.

10. Gel-shift assays

In vitro assays involved the use of the purified and quantified discrete RNA scaffold D0, along with the purified and quantified protein chimeras FM and HP. FM and HP are myc-tagged, forming mycFM and mycHP, which allows for their detection using α -

myc Western blot analysis. Myc tagging was achieved using a PCR amplification step followed by re-cloning and yielded respectively pCJDFMmyc and pCJDHPmyc. The PCR primers mycFM-F/ mycFM-R and mycHP-F/mycHP-R were used to incorporate the myc tags into the respective proteins. To determine the appropriate binding ratios between each of the respective proteins, and the scaffold, we performed titration assays using Do and mycFM or mycHP. Both the scaffold and the protein were combined in the appropriate molar ratio, in PBS buffer, and allowed to incubate at room temperature for 5 minutes before analysis. A total protein amount of 50 ng was used in each lane. Do was added on molar 0, 1, 2, and 3 molar ratios (lanes 1-4, respectively). Complete binding occurs for HP in 1:2 a molar ratio, and FM in a 1:1 molar ratio. Given that we expressed MS2 domain in FM as a dimer, this suggests that both PP7 and MS2 bind their respective aptamers completely as dimers (Figure 50) – which correlates well with the crystallographic data.

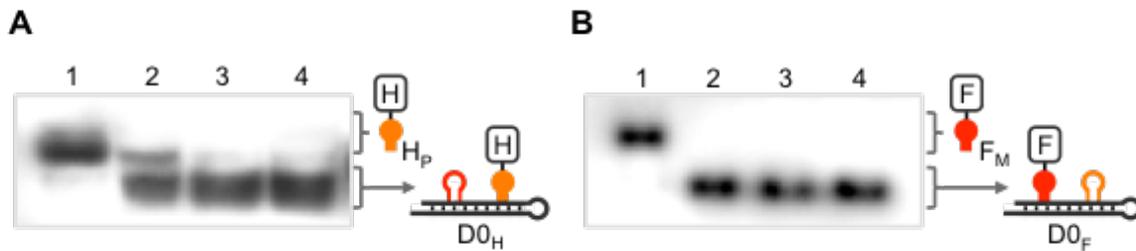


Figure 50: Protein binding assay onto RNA scaffold via reverse gel shift assay. (A) Do is added to HP in a 0, 1, 2, and 3 molar ratios (lanes 1-4, respectively). (B) Do is added to FM in a 0, 1, 2, and 2 molar ratios (lanes 1-4, respectively).

Gel-shift assays involved mixing the scaffold Do with mycFM (1:1 molar ratio) and/or mycHP (1:2 molar ratio), followed by analysis using α -myc Western blot (incubation for 2 hours at room temperature; AbCam). We used 4-20% gradient gels. *In vivo* assays

using mycFM and mycHP cells, with and without the discrete RNA scaffold Do, were also conducted, and involved lysing cells with B-PER® Bacterial Protein.

11. Hydrogen constructs

We used the *Clostridium acetobutylicum* Hydrogenase (HydA) and the *Spinacia oleracea* Ferredoxin I (Fd) (Agapakis et al. 2010). We constructed the chimeric proteins Fd-2X-MS₂ (FM) and HydA-2X-PP7 (HP) by cloning Fd and HydA directly into pV_{0120-2X} vectors containing the respective aptamer proteins. For expression, FM was cloned with NotI/SpeI into the MCS₁ of pColaDuet-BB forming pCJD_{FM}, while HP was cloned with EcoRI/PstI into the MCS₁ of the original pACYCDuet. MCS₂ of the same pACYCDuet vector is then used to clone with NdeI-AvrII the pyruvate ferredoxin oxido-reductase (PFOR) from *Desulfovibrio africanus*. We call this plasmid pCJD_{HHP}. The maturation factors for our *Clostridium acetobutylicum* hydrogenase, HydE (MCS₁) and HydFG (MCS₂) in pC_{DFDuet}, were provided by Agapakis (Agapakis et al. 2010).

12. Hydrogen production assays

Hydrogen production experiments were conducted as follows. BL21-star (DE₃) cells were transformed with the plasmids containing FM and HP, PFOR and the maturation factors, and with the appropriate scaffold. Bacterial cultures were grown aerobically in 20 mL of LB (50 µg/mL ampicillin, 25 µg/mL spectinomycin, 25 µg/mL kana-

mycin, and 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol), in 40 mL Suba sealed glass vials (Sigma), until mid-log phase ($\text{OD}_{600} = 0.4$). The vials were septa sealed, and incubated for an additional hour. At this point, cultures were supplemented with 0.5% glucose, and induced with 0.1 mM of IPTG. Hydrogen production was allowed proceed for 16 hours. Cultures were then quenched using 1 mL of 100% methanol. The headspace gas composition was analyzed using gas chromatography (Shimadzu GC14A), equipped with a TCD detector, a ShinCarbon ST column (Restek Corporation), at 40°C. All hydrogen production values were normalized to the measured cell number at OD_{600} , and to BL21-star (DE3) cells expressing the protein components without any scaffold. Examples of actual chromatograms are illustrated in Figure 51.

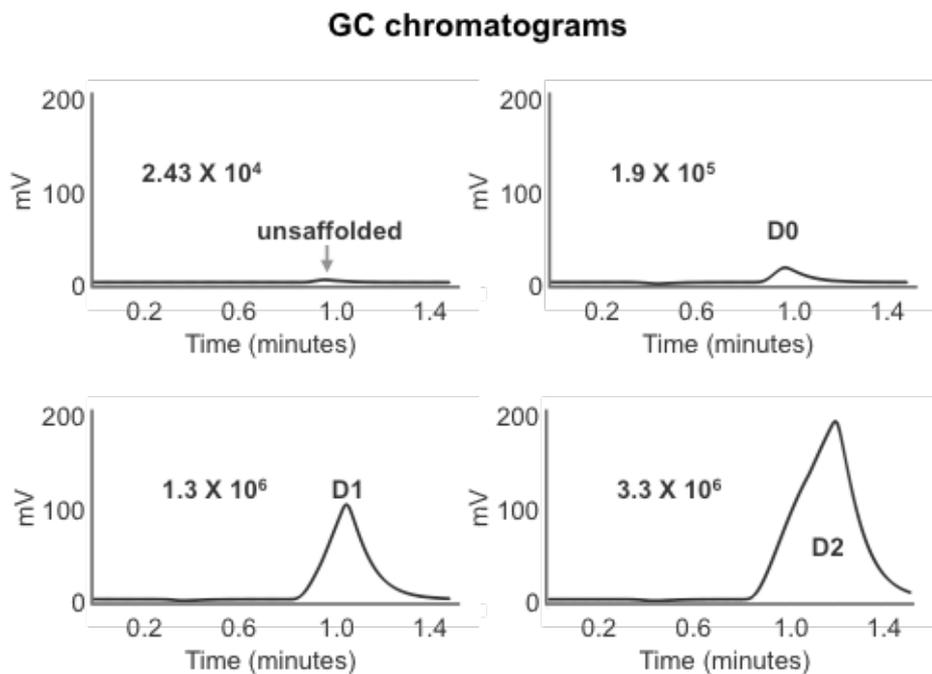


Figure 51: GC chromatograms of hydrogen production. Hydrogen levels were measured for unscaffolded cells, and for cells co-expressing the RNA scaffolds D0, D1 and D2. Numbers correspond to areas under the peaks.

In the case of the unscaffolded, D₀, D₁ and D₂ systems, the respective levels of hydrogen production correspond to 0.8, 3.2, 8.8 and 38.4% of gas volume (20mL in total). This translates to a 4.0, 11 and 48-fold increase for the scaffolded cells D₀, D₁ and D₂ when compared to unscaffolded cells. In all cases, hydrogen was measured after 16 hours, the volume occupied by the gas was 20 mL, and the number of cells producing hydrogen was 0.6×10^9 . This translates to hydrogen production levels for the unscaffolded, D₀, D₁ and D₂ assemblies of 2.6×10^{-21} , 10.4×10^{-21} , 28.6×10^{-21} and 124×10^{-21} moles of hydrogen per moles of cells per hour.

13. Growth curves

We compared the growth rates of cells expressing D₀FH, D₁FH and D₂FH versus cells expressing just the unscaffolded protein components FM and HP after addition of IPTG as shown in Figure 52. We found that cells co-expressing the RNA scaffolds and the protein components grow at similar rates (maximal doubling time of 40-50 minutes) when compared to cells expressing just the protein components. We also observed that the growth rate of all the cells co-expressing both the RNA and protein components is similar – thus the type of scaffold does not affect cell growth. We also determined the growth rates of cells expressing the protein components FM and HP along with the product of what is generated from an empty pETDuet vector. Cells expressing both the protein and the ‘random’ RNA strands grow at a rate similar to the cells co-expressing proteins and either of D₀, D₁ or D₂ (Figure 52). Taken together, the production of our assemblies does not affect cell growth. Growth curves were obtained using a PerkinElmer VICTOR3 plate reader. Briefly, cells co-expressing FM and

HP along with D₀, D₁ or D₂ were grown to mid-log phase, induced with 0.1 mM IPTG, and monitored over four hours (Figure XX).

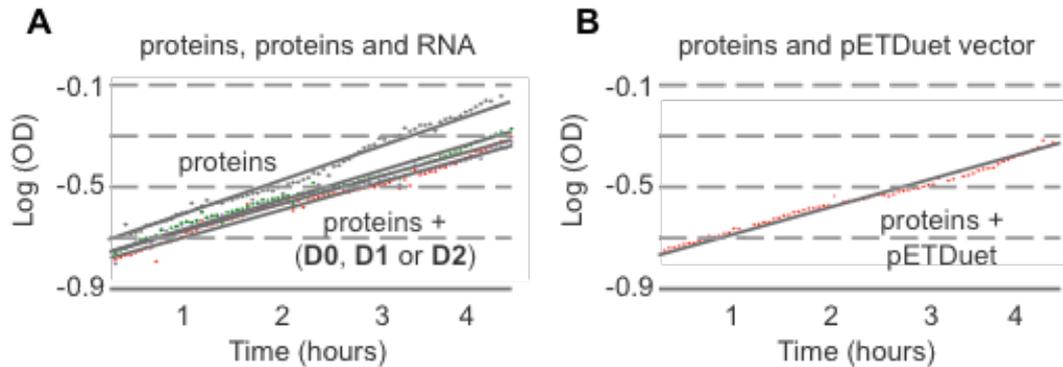


Figure 52: Growth curves of *E. coli* cells expressing RNA scaffolds. (A) Cells expressing just the proteins (i.e. FM and HP), and cells co-expressing proteins with D₀, D₁ or D₂. (B) Cells co-expressing FM and HP with the pETDuet vector (included to account for the effect of an extra antibiotic).

We conducted growth curves, as well as RT-PCR data on the RNA levels of cells expressing the D₀, D₁ and D₂ assemblies in the BL21 *E. coli* strain (i.e. expressing RNase E). We observed little to no difference in growth rates, and in the *in vivo* RNA levels of production (Figure 53).

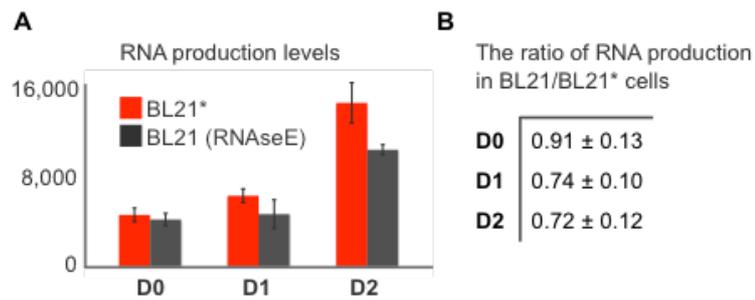


Figure 53: Comparison of RNA production levels in different BL21 cells. (A) RT-PCR data show that RNA levels in BL21 cells (expressing RNaseE) is comparable to the RNA levels in BL21* cells (RNase E knock-off). (B) Values.

14. Strains, plasmids, sequences

| <i>E. coli</i> Strain | Relevant genotype |
|-----------------------|---|
| Turbo cells | Host strain for plasmid construction (NEB) |
| BL21-Star (DE3) | Host strain for T7 expression with a deletion of <i>rne131</i> (Invitrogen) |

| Name | Plasmid | Resistance |
|-----------------------|--|------------|
| pV ₀₁₂₀ | Biobrick Shuttle vector | Amp |
| pV _{0120-2X} | Biobrick Shuttle vector with (Gly ₄ Ser) ₂ | Amp |
| pETDuet | Novagen T7 vector | Amp |
| pCOLADuet | Novagen T7 vector | Amp |
| pACYCDuet-BB | Novagen T7 vector | Amp |
| pCDFDuet-BB | Novagen T7 vector | Amp |
| pCOLADuet-BB | Novagen T7 vector | Amp |
| pCJDD ₀ | d ₀ RNA expression cassette cloned into EcoNI and PstI | Amp |
| pCJDD ₁ | d ₁ RNA expression cassette cloned into EcoNI and PstI | Amp |
| pCJDD ₂ | d ₂ RNA expression cassette cloned into EcoNI and PstI | Amp |
| pCJDD ₁ T | d ₁ T mutated RNA expression cassette cloned in EcoNI and PstI | Amp |
| pCJDD ₂ T | d ₂ T mutated RNA expression cassette cloned in EcoNI and PstI | Amp |
| pCJD _{FB} | F _{B-2X} -PP ₇ was cloned with EcorI-PstI into MCS ₂ of the pColaDuet-BB vector | Kan |
| pCJD _{FA} | F _{A-2X} -MS ₂ with NotI-SpeI into MSC ₁ of the pCDFDuet-BB vector | Spe |
| pCJD _{FM} | F _{d-2X} -MS ₂ cloned with NotI/SpeI into the MCS ₁ of pColaDuet-BB | Kan |

| Name | Plasmid | Resistance |
|-----------|---|------------|
| pCJDHP | HydA-2X-PP7 cloned with EcoRI/PstI into the MSC ₁ and PFOR cloned with NdeI-AvrII into the MSC ₂ of pACYCDuet. | Cm |
| pCJDHE | HydE (MCS ₁) and HydFG (MCS ₂) in pCDFDuet | Spe |
| pCJDFMmyc | myc tagged Fd-2X-MS2 cloned with NotI/SpeI into the MCS ₁ of pColaDuet-BB | Kan |
| pCJDHPmyc | myc tagged HydA-2X-PP7 cloned with EcoRI/PstI into the MSC ₁ and PFOR cloned with NdeI-AvrII into the MSC ₂ of pACYCDuet. | Cm |

| Sequences | Oligo Sequence 5' - 3' |
|--|--|
| $2X$ (Gly ₄ Ser) ₂ | GGAGGAGGAGGATCAGGAGGAGGAGGATCA |
| do | TTAATACGACTCACTATAGGGAGGACTCCCACAGTCACTG GGGAGTCCTCGAATACGAGCTGGGCACAGAAGATATGGC TTCGTGCCCAGGAAGTGTTCGCACTTCTCTCGTATTCTGA TTCCCCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCT TGAGGGGTTTTTTG |
| dr | TTAATACGACTCACTATAGGGTAGGCGCCTAGCCTAATGT ACATTAAGTTATTTTTCCGGATGAATAGAATATATTCTAAT AACGCAGGACTCCCACAGTCACTGGGGAGTCCTCGAATA CGAGCTGGGCACAGAAGATATGGCTTCGTGCCCAGGAAG TGTTTCGCACTTCTCTCGTATTTCGATTGCGACTAGTCTAGC ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTT TT |
| dz' / dz'' | AGGTAAATACGACTCACTATAGGGTCAGGAATCCTCCTGA TAGCTATTTGGACAATTACGTACGTAGTTGATGACAACTA CATGAAAATAAGGGCACAGAAGATATGGCTTCGTGCCCTC TAGACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTT GAGGGGTTTTTTGGCTAAGCATCGATGAATTCTTAATAC GACTCACTATAGGGACGCATTTTCTCCCTTAGCATTA ACTACACCTGCCACAGTCACTGGGCAGGTGTACTAGTCTAGC ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTT TTTGC |
| dr _T | TTAATACGACTCACTATAGGGTAGGCGCCTAGCCTAATTT TTTTTTTTTTTTTTTTTTTTTTTGAATAGAATATATTCTAA TAACGCAGGACTCCCACAGTCACTGGGGAGTCCTCGAAT ACGAGCTGGGCACAGAAGATATGGCTTCGTGCCCAGGAA GTGTTTCGCACTTCTCTCGTATTTCGATTGCGACTAGTCTAG CATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT TTTTG |
| dz' _T | TTTAGATCTCCGGGACCTGCATTAGGTTAATACGACTCA CTATAGGGTCAGGAATCCTCCTGATTTTTTTTTTTTTTTTT TTTTTTTTTTAGTTGATGACAACTACATGAAAATAAGGGCA CAGAAGATATGGCTTCGTGCCCTCTAGACTAGCATAACCC CTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGC TAAGCATCGATGAATTCTTAATACGACTCACTATAGGGAC GCATTTTCTCCCTTAGCATTA ACTACACCTGCCACAGTCA CTGGGCAGGTGTACTAGTCTAGCATAACCCCTTGGGGCC TCTAAACGGGTCTTGAGGGGTTTTTTGCTGCAGGCATGC TTT |
| A _F | TGTTTGATGGTGGTTAACGGC |
| A(dr _T) _R | ATTCAAAAAAAAAAAAAAAAAAAAAAAAAATTAGGCTAGGC GCCTACCCTATAGTGAGTCG |
| A(dz' _T) _R | CCTAATTTTTTTTTTTTTTTTTTTTTTTTGAATAGAATA TATTCTAATAACGCAGGACTCCC |

| Sequences | Oligo Sequence 5' - 3' |
|-----------------------------------|--|
| B(d ₁ T) _F | CAACTAAAAAAAAAAAAAAAAAAAAAAAAAATCAGGAGGAT TCCTGACCCTATAGTG |
| B(d ₂ 'T) _F | CTGATTTTTTTTTTTTTTTTTTTTTTTTAGTTGATGACA ACTACATGAAAATAAGGGC |
| B _R | GTGACTGGTGAGTACTCAACCAAGTC |
| DP _C | /5Biosg/TGTTATGCTAGGTTGTCGGGA |
| DP _R | TCCCGACAACCTAGCATAACA |
| T ₇ RNA _F | AGCATAACCCCTTGGGGGCT |
| polyARNA _R | Invitrogen proprietary primer |
| GapA _F | ACTGACTGGTATGGGGTTCC |
| GapA _R | AGGTTTAACGGCAGCTTTGA |
| MS _{2F} | TTTGAATTCGCGCCGCTTCTAGAATGGCTTCTAACTT- TACT |
| MS _{2R} | TTTCTGCAGCGCCGCTACTAGTTTAGTAGATGCCG- GAGTT |
| PP _{7F} | TTTGAATTCATGTCCAAAACCATCGTTCTTTC |
| PP _{7R} | TTTCTGCAGTTATCAACGGCCCAGCGG |
| F _A F | CCTTGAATTCGCGCCGCTTCTAGAATGGTGAG- CAAGGGCG |
| F _A R | AAGGCTGCAGCGCCGCTACTAGTCTGCTTGTCGGC- CATGATATA |
| F _B F | CCTTGAATTCGCGCCGCTTCTAGAATGGGCAAGAACGG CATCAAGGTGA |
| F _B R | AAGGCTGCAGCGCCGCTACTAGTCTTGACAG- CTCGTCCATGC |
| MS ₂ Aptamer | CCACAGUCACUGGG |
| PP ₇ Aptamer | GGCACAGAAGAUUUGGCUUCGUGCC |

Figure 54: Plasmid, sequences and strain tables.

References

- Aden, A. et al., 2002. Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis For Corn Stover. pp.1-154.
- Adam, G. & Delbruck, M., 1968. Reduction of dimensionality in biological diffusion processes. In A. Rich & N. Davidson, eds. *Structural chemistry and molecular biology*. Structural chemistry and molecular biology. W. H. Freeman and Co. San Francisco, pp. 198-215.
- Adams, D., 1979. *The Hitchhiker's Guide to the Galaxy*, Harmony Books.
- Afonin, K.A. et al., 2010. In vitro assembly of cubic RNA-based scaffolds designed in silico. *Nature nanotechnology*, 5(9), pp.676-682.
- Agapakis, C.M. & Silver, P.A., 2010. Modular electron transfer circuits for synthetic biology: insulation of an engineered biohydrogen pathway. *Bioengineered Bugs*, 1(6), pp.413-418.
- Agapakis, C.M. et al., 2010. Insulation of a synthetic hydrogen metabolism circuit in bacteria. *Journal of biological engineering*, 4(1), p.3.
- Agapakis, C.M. et al., 2011. Towards a Synthetic Chloroplast D. M. Ojcius, ed. *PLoS ONE*, 6(4), pp 77-88.

- Aigrain, L., Pompon, D., Truan, G., Role of the interface between the FMN and FAD domains in the control of redox potential and electronic transfer of NADPH-cytochrome P450 reductase. *Biochemical Journal*, 435 (1) pp. 197-206.
- Akhtar, M.K. & Jones, P.R., 2008. Deletion of *iscR* stimulates recombinant clostridial Fe-Fe hydrogenase activity and H₂-accumulation in *Escherichia coli* BL21(DE3). *Applied microbiology and biotechnology*, 78(5), pp.853-862.
- Aldaye, F.A., Palmer, A.L. & Sleiman, H.F., 2008. Assembling Materials with DNA as the Guide. *Science (New York, NY)*, 321(5897), pp.1795-1799.
- Aliverti, A. & Zanetti, G., 1997. A three-domain iron-sulfur flavoprotein obtained through gene fusion of ferredoxin and ferredoxin-NADP⁺ reductase from spinach leaves. *Biochemistry*, 36(48), pp.14771-14777.
- Allaby, R.G., Fuller, D.Q. & Brown, T.A., 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences*, 105(37), pp.13982-13986.
- Allison, M.J.M., Hammond, A.C.A. & Jones, R.J.R., 1990. Detection of ruminal bacteria that degrade toxic dihydroxypyridine compounds produced from mimosine. *Applied and environmental microbiology*, 56(3), pp.590-594.
- Alon, U., 2006. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. 1st ed, Chapman and Hall/CRC.
- Alon, U., 2003. Biological networks: the tinkerer as an engineer. *Science (New York, NY)*, 301(5641), pp.1866-1867.
- Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp.403-410.
- An, S. et al., 2008. Reversible Compartmentalization of de Novo Purine Biosynthetic Complexes in Living Cells. *Science (New York, NY)*, 320(5872), pp.103-106.

- Andrianantoandro, E. et al., 2006. Andrianantoandro et al. - 2006 - Synthetic biology new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2, p.2006.0028.
- Andronescu, M. et al., 2003. RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic acids research*, 31(13), pp.3416–3422.
- Basak, N. & Das, D., 2007. The Prospect of Purple Non-Sulfur (PNS) Photosynthetic Bacteria for Hydrogen Production: The Present State of the Art. *World Journal of Microbiology and Biotechnology*, 23(1), pp.31–42.
- Bashor, C.J. et al., 2008. Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics. *Science (New York, NY)*, 319(5869), pp.1539–1543.
- Bayer, T.S. et al., 2009. Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *Journal of the American Chemical Society*, 131(18), pp.6508–6515.
- Becvar, J.E. & Hastings, J.W., 1975 - Bacterial Luciferase Requires One Reduced Flavin for Light Emission. *Proceedings of the National Academy of Sciences of the United States of America*, 72(9), pp.3374–3376.
- Benenson, Y., 2012. Synthetic biology with RNA: progress report. *Current opinion in chemical biology*, pp.1–7.
- Benes, V. & Castoldi, M., 2010. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods*, 50(4), pp.244–249.
- Benner, S.A., Ellington, A.D. & Tauer, A., 1989. Modern metabolism as a palimpsest of the RNA world. *Proceedings of the National Academy of Sciences of the United States of America*, 86(18), pp.7054–7058.
- Bingham, A.S., Smith, P.R. & Swartz, J.R., 2012. Evolution of an [FeFe] hydrogenase with decreased oxygen sensitivity. *International Journal of Hydrogen Energy*, 37(3), pp.2965–2976.

- Blackburn, E.H. & Collins, K., 2011. Telomerase: An RNP Enzyme Synthesizes DNA. *Cold Spring Harbor Perspectives in Biology*, 3(5), pp.3558.
- Blouin, K. et al., 1996. Blouin et al. - 1996 - Characterization of In Vivo Reporter Systems for Gene Expression and Biosensor Applications Based on luxAB Luciferase Genes. *Applied and environmental microbiology*, 62(6), pp.2013-2021.
- Bobik, T.A., 2006. Polyhedral organelles compartmenting bacterial metabolic processes. *Applied microbiology and biotechnology*, 70(5), pp.517-525.
- Bonacci, W. et al., 2012. Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences of the United States of America*, 109(2), pp.478-483.
- Boyle, P.M. & Silver, P.A., 2009. Harnessing nature's toolbox: regulatory elements for synthetic biology. *Journal of The Royal Society Interface*, 6(Suppl 4), pp.S535-S546.
- Breaker, R.R., 2012. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(2), pp.3566.
- Brenner, K., You, L. & Arnold, F.H., 2008. Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology*, 26(9), pp.483-489.
- Buikema, W.J. & Haselkorn, R., 1991. Characterization of a gene controlling heterocyst differentiation in the cyanobacterium *Anabaena 7120*. *Genes & Development*, 5(2), pp.321-330.
- Bunka, D.H.J. & Stockley, P.G., 2006. Aptamers come of age – at last. *Nature Reviews Microbiology*, 4(8), pp.588-596.
- Burack, W.R. & Shaw, A.S., 2010. Signal transduction: hanging on a scaffold. *Current opinion in cell biology*, 12(2), pp.211-216.
- Burbank, L., 1909. Another mode of species forming. *Journal of Heredity*, (1), pp.40-43.
- Burbank, L., 1923. Popular Science Monthly.

- Callura, J.M. et al., 2010. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), pp.15898–15903.
- Campos, L., 2009. *Synthetic Biology* M. Schmidt et al., eds., Dordrecht: Springer Netherlands.
- Carlson, R., 2009. The changing economics of DNA synthesis. *Nature Biotechnology*, 27(12), pp.1091–1094.
- Cayrol, B. et al., 2009. A nanostructure made of a bacterial noncoding RNA. *Journal of the American Chemical Society*, 131(47), pp.17270–17276.
- Cech, T.R., 2012. The RNA Worlds in Context. *Cold Spring Harbor Perspectives in Biology*, 4(7), pp.6742.
- Renewable Energy Policy Network for the 21st century. *Renewables 2011 Global Status Report*, DIANE Publishing.
- Chao, J.A. et al., 2007. Structural basis for the coevolution of a viral RNA-protein complex. *Nature structural & molecular biology*, 15(1), pp.103–105.
- Chowdhury, S.S. et al., 2006. Molecular basis for temperature sensing by an RNA thermometer. *The EMBO Journal*, 25(11), pp.2487–2497.
- Cluis, C.P. et al., 2011. Identification of bottlenecks in Escherichia coli engineered for the production of CoQ10. *Metabolic Engineering*, 13(6), pp.733–744.
- Coburn, G.A. & Mackie, G.A., 1996. Differential sensitivities of portions of the mRNA for ribosomal protein S20 to 3'-exonucleases dependent on oligoadenylation and RNA secondary structure. *The Journal of biological chemistry*, 271(26), pp.15776–15781.
- Coetsee, T., Herschlag, D. & Belfort, M., 1994. Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes & Development*, 8(13), pp.1575–1588.

- Conrado, R.J. et al., 2012. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic acids research*, 40(4), pp.1879–1889.
- Conrado, R.J., Varner, J.D. & DeLisa, M.P., 2008. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Current opinion in biotechnology*, 19(5), pp.492–499.
- Convery, M.A. et al., 2004. Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nature structural biology*, 5(2), pp.133–139.
- Cumino, A.C. et al., 2007. Carbon Cycling in *Anabaena* sp. PCC 7120. Sucrose Synthesis in the Heterocysts and Possible Role in Nitrogen Fixation. *Plant Physiology*, 143(3), pp.1385–1397.
- Danna, K.K. & Nathans, D.D., 1971. Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proceedings of the National Academy of Sciences of the United States of America*, 68(12), pp.2913–2917.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection Or The Preservation of Favoured Races in the Struggle for Life*.
- Darwin, C., 1868. *The Variation of Animals and Plants Under Domestication*.
- DAS, D. & VEZIROGLU, T., 2008. Advances in biological hydrogen production processes. *International Journal of Hydrogen Energy*, 33(21), pp.6046–6057.
- Dasgupta, C.N. et al., 2010. Recent trends on the development of photobiological processes and photobioreactors for the improvement of hydrogen production. *International Journal of Hydrogen Energy*, pp.1–21.
- de Kok, A.A. et al., 1998. The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1385(2), pp.353–366.
- Dekker, J.P. & Boekema, E.J., 2005. Supramolecular organization of thylakoid membrane proteins in green plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1706(1–2), pp.12–39.

- Delebecque, C.J. et al., 2011. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science (New York, NY)*, 333(6041), pp.470-474.
- Delebecque, C.J., Silver, P.A. & Lindner, A.B., 2012. Designing and using RNA scaffolds to assemble proteins in vivo. *Nature Protocols*, 7(10), pp.1797-1807.
- 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.
- Diestra, E. et al., 2009. Visualization of proteins in intact cells with a clonable tag for electron microscopy. *Journal of Structural Biology*, 165(3), pp.157-168.
- Dinger, M.E. et al., 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research*, 18(9), pp.1433-1445.
- Drubin, D.A., Way, J.C. & Silver, P.A., 2007. Genes & Development 2007 Drubin-I. *Genes & Development*, 21(3), pp.242-254.
- Dueber, J.E. et al., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology*, 27(8), pp.753-759.
- Dutta, D. et al., 2005. Hydrogen Production by Cyanobacteria. *Microbial Cell Factories*, 4(1), p.36.
- Eddy, S.R.S., 2001. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2(12), pp.919-929.
- Eigen, M. et al., 1981. The origin of genetic information. *Scientific American*, 244(4), pp.88-92- 96- et passim.
- Elhai, J. & Wolk, C.P., 1988. Conjugal transfer of DNA to cyanobacteria. *Methods in enzymology*, 167, pp.747-754.
- Elhai, J.J. et al., 1997. Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. *Journal of Bacteriology*, 179(6), pp.1998-2005.

- Ellington, A.D. & Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), pp.818–822.
- Ellis T., Adle T., Baldwin G. S., 2011. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology*, 3, pp 109–118.
- Endres, R.G.R., 2011. Intracellular chemical gradients: morphing principle in bacteria. *BMC Biophysics*, 5, pp.18–18.
- Endy, D., 2005. Foundations for engineering biology. *Nature*, 438(7067), pp.449–453.
- Erkelenz, M., Kuo, C.-H. & Niemeyer, C.M., 2011. DNA-Mediated Assembly of Cytochrome P450 BM3 Subdomains. *Journal of the American Chemical Society*, 133(40), pp.16111–16118.
- Famulok, M. & Ackermann, D., 2010. RNA nanotechnology: inspired by DNA. *Nature nanotechnology*, 5(9), pp.634–635.
- Fay, P.P. & Walsby, A.E.A., 1966. Metabolic activities of isolated heterocysts of the blue-green alga *Anabaena cylindrica*. *Nature*, 209(5018), pp.94–95.
- Flores, E. & Herrero, A., 2009. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology*, 8(1), pp.39–50.
- Fu, J. et al., 2012. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *Journal of the American Chemical Society*, 134(12), pp.5516–5519.
- Gantt, E.E. & Conti, S.F.S., 1969. Ultrastructure of blue-green algae. *Journal of Bacteriology*, 97(3), pp.1486–1493.
- Gibson, D.G. et al., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), pp.343–345.
- Gilardi, G.G. et al., 2001. Molecular Lego: design of molecular assemblies of P450 enzymes for nanobiotechnology. *Biosensors and Bioelectronics*, 17(1-2), pp.133–145.
- Gilbert, W., 1986. Origin of life: The RNA world. *Nature*, 319(6055).

- Golden, J.W. & Yoon, H.-S., 2003. Heterocyst development in *Anabaena*. *Current opinion in microbiology*, 6(6), pp.557-563.
- Golding, I. & Cox, E.C., 2009. RNA dynamics in live *Escherichia coli* cells. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), pp.11310-11315.
- Golding, I. et al., 2005. Real-time kinetics of gene activity in individual bacteria // Supplement. *Cell*, 123(6), pp.1025-1036.
- Gottesman, S. & Storz, G., 2011. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harbor Perspectives in Biology*, 3(12), pp.a003798-a003798.
- Gralla, J.J. & Crothers, D.M.D., 1973. Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *Journal of Molecular Biology*, 73(4), pp.497-511.
- Guerrier-Takada, C. et al., 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2), pp.849-857.
- Guo, P., 2010. The emerging field of RNA nanotechnology. *Nature nanotechnology*, 5(12), pp.833-842.
- Guttman, M. et al., 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, 477(7364), pp.295-300.
- Hallenbeck, P.C. & Ghosh, D., 2009. Advances in fermentative biohydrogen production: the way forward? *Trends in Biotechnology*, 27(5), pp.287-297.
- Hallenbeck, P.C., Abo-Hashesh, M. & Ghosh, D., 2012. Strategies for improving biological hydrogen production. *Bioresource Technology*, 110(C), pp.1-9.
- Hartl, F.U. & Hayer-Hartl, M., 2009. Converging concepts of protein folding in vitro and in vivo. *Nature structural & molecular biology*, 16(6), pp.574-581.
- Heijne, von, G., 1990. The signal peptide. *Journal of Membrane Biology*, 115(3), pp.195-201.

- Hillson, N.J., Rosengarten, R.D. & Keasling, J.D., 2012. j5 DNA Assembly Design Automation Software. *ACS Synthetic Biology*, 1(1), pp.14–21.
- Hiscox, J.A., 2007. RNA viruses: hijacking the dynamic nucleolus. *Nature Reviews Microbiology*, 5(2), pp.119–127.
- Howe, C.J. et al., 2008. The origin of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1504), pp.2675–2685.
- Huang, T.C.T. et al., 1990. Circadian Rhythm of the Prokaryote *Synechococcus* sp. RF-1. *Plant Physiology*, 92(2), pp.531–533.
- Huh, W.-K. et al., 2003. Global analysis of protein localization in budding yeast. *Nature*, 425(6959), pp.686–691.
- Isaacs, F.J. et al., 2004. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7), pp.841–847.
- Isaacs, F.J., Dwyer, D.J. & Collins, J.J., 2006. RNA synthetic biology. *Nature Biotechnology*, 24(5), pp.545–554.
- Jacob, F. & MONOD, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), pp.318–356.
- Jaeger, L. & Chworos, A., 2006. The architectonics of programmable RNA and DNA nanostructures. *Current opinion in structural biology*, 16(4), pp.531–543.
- Johnson, I.S., 1983. Human Insulin from Recombinant DNA Technology. *Science (New York, NY)*, 219(4), pp.632–637.
- Johnson, J.M. et al., 2005. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends in Genetics*, 21(2), pp.93–102.
- Jones, K.M.K. & Haselkorn, R.R., 2002. Newly identified cytochrome c oxidase operon in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120 specifically induced in heterocysts. *Journal of Bacteriology*, 184(9), pp.2491–2499.

- Jr, I.T. & Bustamante, C., 1999. How RNA folds. *J. Mol. Biol.*, 293, pp.271–281.
- Kapranov, G.S.L.A.P., 2012. Dark matter RNA: an intelligent scaffold for the dynamic regulation of the nuclear information landscape. pp.1–11.
- Kapranov, P. et al., 2010. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. *BMC Biology*, 8(1), p.149.
- Keiichi I. et al., 1977. Expression in Escherichia Coli of a Chemically Synthesized gene for the hormone somatostatin. *Science*, 198 pp 1056–1063.
- Khalil, A.S. & Collins, J.J., 2010. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5), pp.367–379.
- Khorana, H.G.H., 1979. Total synthesis of a gene. *Science (New York, NY)*, 203(4381), pp.614–625.
- Khosla, C., Kapur, S. & Cane, D.E., 2009. Revisiting the modularity of modular polyketide synthases. *Current opinion in chemical biology*, 13(2), pp.135–143.
- Kim, H.J. et al., 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47), pp.18188–18193.
- Kirchhoff, H. et al., 2008. Protein Diffusion and Macromolecular Crowding in Thylakoid Membranes. *Plant Physiology*, 146(4), pp.1571–1578.
- Ko, S.H. et al., 2010. Synergistic self-assembly of RNA and DNA molecules. *Nature Chemistry*, 2(12), pp.1050–1055.
- Kortmann, J. et al., 2011. Translation on demand by a simple RNA-based thermosensor. *Nucleic acids research*, 39(7), pp.2855–2868.
- Kruger, K. et al., 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1), pp.147–157.
- Kuhn, T.S., 1996. *The Structure of Scientific Revolutions* 3rd ed, University of Chicago Press.

- Kumar, K., Mella-Herrera, R.A. & Golden, J.W., 2010. Cyanobacterial Heterocysts. *Cold Spring Harbor Perspectives in Biology*, 2(4), pp.a000315–a000315.
- Kuzyk, A., Laitinen, K.T. & Törmä, P., 2009. DNA origami as a nanoscale template for protein assembly. *Nanotechnology*, 20(23), p.235305.
- Agence de l'Environnement et de la Maîtrise de l'Énergie, 2010. Life cycle assessments applied to first generation biofuels used in France. pp.1–229.
- Laurent, S. et al., 2005. Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in *Anabaena* sp. PCC 7120. *Proceedings of the National Academy of Sciences of the United States of America*, 102(28), pp.9907–9912.
- Lee, H., DeLoache, W.C. & Dueber, J.E., 2011. Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering*, 14(3), pp.242–251.
- Leontis, N.B., Lescoute, A. & Westhof, E., 2006. The building blocks and motifs of RNA architecture. *Current opinion in structural biology*, 16(3), pp.279–287.
- Leroux, F. et al., 2010. Is engineering O₂-tolerant hydrogenases just a matter of reproducing the active sites of the naturally occurring O₂-resistant enzymes? *International Journal of Hydrogen Energy*, 35(19), pp.10770–10777.
- Li, C. & Fang, H.H.P., 2007. Fermentative hydrogen production from wastewater and solid wastes by mixed cultures. *Critical Reviews in Environmental Science and Technology*, 37, pp.1–39.
- Liang, J.C., Bloom, R.J. & Smolke, C.D., 2011. Engineering Biological Systems with Synthetic RNA Molecules. *Molecular Cell*, 43(6), pp.915–926.
- Lim, F. & Peabody, D.S., 2002. RNA recognition site of PP7 coat protein. *Nucleic acids research*, 30(19), pp.4138–4144.
- Lim, F., Downey, T.P. & Peabody, D.S., 2001. Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7. *Journal of Biological Chemistry*, 276(25), pp.22507–22513.

- Lin, C. et al., 2008. In vivo cloning of artificial DNA nanostructures. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), pp.17626–17631.
- Lin, C., Liu, Y. & Yan, H., 2009. Designer DNA Nanoarchitectures †. *Biochemistry*, 48(8), pp.1663–1674.
- Liu, H. et al., 2006. Approaching The Limit: Can One DNA Oligonucleotide Assemble into Large Nanostructures? *Angewandte Chemie International Edition*, 45(12), pp.1942–1945.
- Liu, H. et al., 2005. Two-Dimensional (2D) DNA Crystals Assembled from Two DNA Strands. *Biomacromolecules*, 6(6), pp.2943–2945.
- Lucks, J.B. et al., 2011. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(21), p.8617.
- Lütke-Eversloh, T. & Stephanopoulos, G., 2008. Combinatorial pathway analysis for improved L-tyrosine production in *Escherichia coli*: Identification of enzymatic bottlenecks by systematic gene overexpression. *Metabolic Engineering*, 10(2), pp.69–77.
- Mahen, E.M. et al., 2010. mRNA Secondary Structures Fold Sequentially But Exchange Rapidly In Vivo M. Wickens, ed. *PLoS biology*, 8(2), p.e1000307.
- Majdalani, N. et al., 1998. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), pp.12462–12467.
- Marchesini, S., 2003. Futile Cycling of Intermediates of Fatty Acid Biosynthesis toward Peroxisomal α -Oxidation in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 278(35), pp.32596–32601.
- Martín-Figueroa, E.E., Navarro, F.F. & Florencio, F.J.F., 2000. The GS-GOGAT pathway is not operative in the heterocysts. Cloning and expression of glsF gene from the cyanobacterium *Anabaena* sp. PCC 7120. *FEBS letters*, 476(3), pp.282–286.

- Melis, A.A. et al., 1999. Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*. *Plant Physiology*, 122(1), pp.127–136.
- Meng, X. & Wolfe, S.A., 2006. Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nature Protocols*, 1(1), pp.30–45.
- Miles, E.W., 2001. Tryptophan synthase: a multienzyme complex with an intramolecular tunnel. *Chemical record (New York, N.Y.)*, 1(2), pp.140–151.
- Miles, E.W., Rhee, S. & Davies, D.R., 1999. The molecular basis of substrate channeling. *The Journal of biological chemistry*, 274(18), pp.12193–12196.
- Molinaro, M.M. & Tinoco, I.I., 1995. Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic acids research*, 23(15), pp.3056–3063.
- Momirlan, M. & Veziroglu, T.N., 2002. Current status of hydrogen energy. *Renewable and Sustainable Energy Reviews*, 6(1), pp.141–179.
- Moon, T.S. et al., 2010. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metabolic Engineering*, 12(3), pp.298–305.
- Moore, P.B. & Steitz, T.A., 2011. The Roles of RNA in the Synthesis of Protein. *Cold Spring Harbor Perspectives in Biology*, 3(11), pp.a003780–a003780.
- Moose, S.P. & Mumm, R.H., 2008. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. *Plant Physiology*, 147(3), pp.969–977.
- Morita, T., 2005. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & Development*, 19(18), pp.2176–2186.
- Muro-Pastor, A.M.A. et al., 2002. Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. *Molecular Microbiology*, 44(5), pp.1377–1385.

- Muro-Pastor, M.I., Reyes, J.C. & Florencio, F.J., 2001. Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *The Journal of biological chemistry*, 276(41), pp.38320–38328.
- Murry, M.A.M., Horne, A.J.A. & Benemann, J.R.J., 1984. Physiological Studies of Oxygen Protection Mechanisms in the Heterocysts of *Anabaena cylindrica*. *Applied and environmental microbiology*, 47(3), pp.449–454.
- Narayanaswamy, R.R. et al., 2009. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), pp.10147–10152.
- Neupert, J. & Bock, R., 2009. Designing and using synthetic RNA thermometers for temperature-controlled gene expression in bacteria. *Nature Protocols*, 4(9), pp.1262–1273.
- Ng, S.-Y., Johnson, R. & Stanton, L.W., 2011. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *The EMBO Journal*, 31(3), pp.522–533.
- Niemeyer, C.M.C., Koehler, J.J. & Wuerdemann, C.C., 2002. DNA-directed assembly of bienzymic complexes from in vivo biotinylated NAD(P)H:FMN oxidoreductase and luciferase. *ChemBioChem*, 3(2-3), pp.242–245.
- Nordon, R.E., Craig, S.J. & Foong, F.C., 2008. Molecular engineering of the cellulosome complex for affinity and bioenergy applications. *Biotechnology Letters*, 31(4), pp.465–476.
- Noree, C. et al., 2010. Identification of novel filament-forming proteins in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. *The Journal of Cell Biology*, 190(4), pp.541–551.
- Ogawa, A. & Maeda, M., 2008. An Artificial Aptazyme-Based Riboswitch and its Cascading System in *E. coli*. *ChemBioChem*, 9(2), pp.206–209.
- Ozimek, P. et al., 2005. Alcohol oxidase: A complex peroxisomal, oligomeric flavoprotein. *FEMS Yeast Research*, 5(11), pp.975–983.

- Page, C.C.C. et al., 1999. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature*, 402(6757), pp.47-52.
- Park, S.-H., Zarrinpar, A. & Lim, W.A., 2003. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science (New York, NY)*, 299(5609), pp.1061-1064.
- Parrott, A.M. et al., 2000. RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. *Nucleic acids research*, 28(2), pp.489-497.
- Peabody, D.S., 1993. The RNA binding site of bacteriophage MS2 coat protein. *The EMBO Journal*, 12(2), pp.595-600.
- Pitera, D.J. et al., 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metabolic Engineering*, 9(2), pp.193-207.
- Pörschke, D., 1974. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophysical Chemistry*, 1(5), pp.381-386.
- Richards, J. et al., 2008. Quality control of bacterial mRNA decoding and decay. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1779(9), pp.574-582.
- Roberts, R.J.R., 2005. How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences of the United States of America*, 102(17), pp.5905-5908.
- Rothemund, P.W.K., 2006. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082), pp.297-302.
- Saito, H. & Inoue, T., 2009. Synthetic biology with RNA motifs. *The international journal of biochemistry & cell biology*, 41(2), pp.398-404.
- Salis, H.M., Mirsky, E.A. & Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10), pp.946-950.

- Savage, D.F. et al., 2010. Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science (New York, NY)*, 327(5970), pp.1258–1261.
- Savage, D.F.D., Way, J.J. & Silver, P.A.P., 2008. Defossilizing fuel: how synthetic biology can transform biofuel production. *ACS Chemical Biology*, 3(1), pp.13–16.
- Schrader, M. & Fahimi, H.D., 2004. Mammalian peroxisomes and reactive oxygen species. *Histochemistry and Cell Biology*, 122(4), pp.383–393.
- Schroeder, R.R. et al., 2002. RNA folding in vivo. *Current opinion in structural biology*, 12(3), pp.296–300.
- Schwarz, W.H., 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied microbiology and biotechnology*, 56(5-6), pp.634–649.
- Seeman, N.C., 2007. An Overview of Structural DNA Nanotechnology. *Molecular Biotechnology*, 37(3), pp.246–257.
- Seeman, N.C., 2010. Nanomaterials Based on DNA. *Annual review of biochemistry*, 79(1), pp.65–87.
- Seeman, N.C., 1982. Nucleic acid junctions and lattices. *Journal of Theoretical Biology*, 99(2), pp.237–247.
- Selinger, D.W., 2003. Selinger et al. - 2003 - Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation. *Genome Research*, 13(2), pp.216–223.
- SEMRAD, K., 2004. RNA chaperone activity of large ribosomal subunit proteins from Escherichia coli. *RNA (New York, NY)*, 10(12), pp.1855–1860.
- Shevtsov, S.P. & Dunder, M., 2011. Nucleation of nuclear bodies by RNA. *Nature Publishing Group*, 13(2), pp.167–173.
- Shiota, N. et al., 2000. Expression of human cytochromes P450 1A1 and P450 1A2 as fused enzymes with yeast NADPH-cytochrome P450 oxidoreductase in transgenic tobacco plants. *Bioscience, Biotechnology, and Biochemistry*, 64(10), pp.2025–2033.

- Shu, D. et al., 2004. Bottom-up Assembly of RNA Arrays and Superstructures as Potential Parts in Nanotechnology. *Nano Letters*, 4(9), pp.1717-1723.
- Siddiqui, M.S. et al., 2012. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Research*, 12(2), pp.144-170.
- Sinha, J., Reyes, S.J. & Gallivan, J.P., 2010. Reprogramming bacteria to seek and destroy an herbicide. *Nature Methods*, 6(6), pp.464-470.
- Sledjeski, D.D. & Gottesman, S.S., 1995. A small RNA acts as an antisilencer of the HNS-silenced rcsA gene of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 92(6), pp.2003-2007.
- Smith, H.O.H. & Wilcox, K.W.K., 1970. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *Journal of Molecular Biology*, 51(2), pp.379-391.
- Sosinsky, G.E. et al., 2007. Markers for correlated light and electron microscopy. *Methods in cell biology*, 79, pp.575-591.
- St Laurent, G. et al., 2012. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics*, 13(1), p.504.
- Stapleton, J.A. & Swartz, J.R., 2010. Development of an In Vitro Compartmentalization Screen for High-Throughput Directed Evolution of [FeFe] Hydrogenases F. Romesberg, ed. *PLoS ONE*, 5(12), p.e15275.
- Stockley, P.G. et al., 1995. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic acids research*, 23(13), pp.2512-2518.
- Szittner, R. et al., 2003. Bright stable luminescent yeast using bacterial luciferase as a sensor. *Biochemical and Biophysical Research Communications*, 309(1), pp.66-70.
- Tamagnini, P. et al., 2002. Tamagnini et al. - 2002 - Hydrogenases and Hydrogen Metabolism of Cyanobacteria. *Microbiology and molecular biology reviews*, 66(1), pp.1-20.

- Tan, S.J. et al., 2011. Building plasmonic nanostructures with DNA. *Nature nanotechnology*, 6(5), pp.268–276.
- Tanigawa, R. et al., 2002. Transcriptional activation of NtcA-dependent promoters of *Synechococcus* sp. PCC 7942 by 2-oxoglutarate in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), pp.4251–4255.
- TAO, Y. et al., 2007. High hydrogen yield from a two-step process of dark- and photo-fermentation of sucrose. *International Journal of Hydrogen Energy*, 32(2), pp.200–206.
- Teller, C. & Willner, I., 2010. Organizing protein-DNA hybrids as nanostructures with programmed functionalities. *Trends in Biotechnology*, 28(12), pp.619–628.
- Thiel, T.T. & Poo, H.H., 1989. Transformation of a filamentous cyanobacterium by electroporation. *Journal of Bacteriology*, 171(10), pp.5743–5746.
- Tolia, N.H. & Joshua-Tor, L., 2006. Tolia, Joshua-Tor - 2006 - Strategies for protein coexpression in *Escherichia coli*. *Nature Methods*, 3(1), pp.55–64.
- Tran, L. et al., 2010. Insights into Protein-Protein and Enzyme-Substrate Interactions in Modular Polyketide Synthases. *Chemistry & biology*, 17(7), pp.705–716.
- Tsuji, S. et al., 2009. RNA aptamer binding to polyhistidine-tag. *Biochemical and Biophysical Research Communications*, 386(1), pp.227–231.
- Tu, S.C. & Hastings, J.W., 2003. Tu, Hastings - 1980 - Physical interaction and activity coupling between two enzymes induced by immobilization of one. *Proceedings of the National Academy of Sciences of the United States of America*, 77(1), pp.249–252.
- Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, NY)*, 249(4968), pp.505–510.
- Uhlenbeck, O.C., 1995. Keeping RNA happy. *RNA (New York, NY)*, 1(1), pp.4–6.

- Ungerer, J.L., Pratte, B.S. & Thiel, T., 2008. Regulation of Fructose Transport and Its Effect on Fructose Toxicity in *Anabaena* spp. *Journal of Bacteriology*, 190(24), pp.8115–8125.
- Valencia-Burton, M. et al., 2007. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature Methods*, 4(5), pp.421–427.
- Valenzano, D.R. et al., 2006. Resveratrol Prolongs Lifespan and Retards the Onset of Age-Related Markers in a Short-Lived Vertebrate. *Current Biology*, 16(3), pp.296–300.
- van Hylckama Vlieg, J.E.J. et al., 2000. Detoxification of reactive intermediates during microbial metabolism of halogenated compounds. *Current opinion in microbiology*, 3(3), pp.257–262.
- Vogel, U.U. & Jensen, K.F.K., 1994. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *Journal of Bacteriology*, 176(10), pp.2807–2813.
- Volpe, T. & Martienssen, R.A., 2011. RNA Interference and Heterochromatin Assembly. *Cold Spring Harbor Perspectives in Biology*, 3(9), pp.a003731–a003731.
- Waks, Z. & Silver, P.A., 2009. Engineering a Synthetic Dual-Organism System for Hydrogen Production. *Applied and environmental microbiology*, 75(7), pp.1867–1875.
- Walsh, C.T.C., 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science*, 303(5665), pp.1805–1810.
- Wang, Y. & Yu, O., 2012. Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells. *Journal of Biotechnology*, 157(1), pp.258–260.
- Weeks, K.M., 2010. Advances in RNA structure analysis by chemical probing. *Current opinion in structural biology*, 20(3), pp.295–304.
- Wei, T.F., Ramasubramanian, T.S. & Golden, J.W., 1994. *Anabaena* sp. strain PCC 7120 *ntcA* gene required for growth on nitrate and heterocyst development. *Journal of Bacteriology*, 176(15), pp.4473–4482.

- Weinberg, Z. et al., 2009. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature*, 462(7273), pp.656–659.
- Werstuck, G. & Green, M.R., 1998. Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions. *Science (New York, NY)*, 282(5387), pp.296–298.
- Whitmarsh, A.J. & Davis, R.J., 1998. Structural organization of MAP-kinase signaling modules by scaffold proteins in yeast and mammals. *Trends in biochemical sciences*, 23(12), pp.481–485.
- Will, C.L. & Luhrmann, R., 2011. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, 3(7), pp.a003707–a003707.
- Williams, B.A.R. et al., 2007. Self-Assembled Peptide Nanoarrays: An Approach to Studying Protein–Protein Interactions. *Angewandte Chemie*, 119(17), pp.3111–3114.
- Wilner, O.I. et al., 2009. Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature nanotechnology*, 4(4), pp.249–254.
- Win, M.N.M. & Smolke, C.D.C., 2008. Higher-order cellular information processing with synthetic RNA devices. *Science (New York, NY)*, 322(5900), pp.456–460.
- Winfree, E. et al., 1998. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394(6693), pp.539–544.
- Winkler, W.C. & Breaker, R.R., 2003. Genetic Control by Metabolite-Binding Riboswitches. *ChemBioChem*, 4(10), pp.1024–1032.
- Witherell, G.W., Gott, J.M. & Uhlenbeck, O.C., 1991. Specific interaction between RNA phage coat proteins and RNA. *Progress in nucleic acid research and molecular biology*, 40, pp.185–220.
- Wolpert, L., 1969. Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, 25(1), pp.1–47.
- Woodson, S.A., 2000. Recent insights on RNA folding mechanisms from catalytic RNA. *Cellular and molecular life sciences : CMLS*, 57(5), pp.796–808.

- Xiao, F. et al., 2005. Binding of pRNA to the N-terminal 14 amino acids of connector protein of bacteriophage phi29. *Nucleic acids research*, 33(8), pp.2640–2649.
- Xu, P. et al., 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metabolic Engineering*, pp.1–10.
- Yeates, T.O., Crowley, C.S. & Tanaka, S., 2010. Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution. *Annual Review of Biophysics*, 39(1), pp.185–205.
- Yeates, T.O.T. & Padilla, J.E.J., 2002. Designing supramolecular protein assemblies. *Current opinion in structural biology*, 12(4), pp.464–470.
- Yin, P. et al., 2008. Programming biomolecular self-assembly pathways. *Nature*, 451(7176), pp.318–322.
- Zadeh, J.N. et al., 2010. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1), pp.170–173.
- Zappulla, D. & Cech, T.R.T., 2004. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proceedings of the National Academy of Sciences of the United States of America*, pp.1–6.
- Zaug, A.J.A. & Cech, T.R.T., 1986. The intervening sequence RNA of Tetrahymena is an enzyme. *Science (New York, NY)*, 231(4737), pp.470–475.
- Zhang, C.-C. et al., 2005. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Molecular Microbiology*, 59(2), pp.367–375.
- Zhu, Y. et al., 2008. High Glycolytic Flux Improves Pyruvate Production by a Metabolically Engineered Escherichia coli Strain. *Applied and environmental microbiology*, 74(21), pp.6649–6655.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, 31(13), pp.3406–3415.