

Engineering the spatial organization of metabolic payhways -a new approach for optimization of cell factories

Buron, Line Due; Mortensen, Uffe Hasbro

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):

Albertsen, L., & Mortensen, U. H. (2010). Engineering the spatial organization of metabolic payhways: -a new approach for optimization of cell factories. Kgs. Lyngby, Denmark: Technical University of Denmark (DTU).

DTU Library Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Preface and acknowledgements

This thesis describes the primary results obtained during my PhD project carried out at Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark from September 2005 - November 2008. The project has been financed by Forskningsrådet and for this they are gratefully acknowledged.

First and foremost, I would like to thank my supervisor Associate Professor Uffe Mortensen for making this PhD study such an exciting experience. His enthusiasm, optimism, excellent guidance and passion for research have made this project a great experience with very few hard times. Moreover, I would like to thank him for being such a rich source of inspiration and for spending time on commenting this thesis.

All the present and former colleagues at CMB also deserve thanks for making this department such a great place to work. Particularly, I would like to thank Anne Egholm Pedersen, Jerome Maury and Mohammad Asadollahi for introducing me to the sesquiterpene system and for letting me in on the secrets of running two phase fermentations, Swee Hallwyl Jensen and Gäelle Lettier for showing me how to handle yeast, Jesper Mogensen and Kristian Fog Nielsen for invaluable help with GC-analysis, Ana Rita and Kiran Patil for fruitful discussions and introduction to the vanillin system, Iben Plate for introducing me to the nuclear transport project and Jakob Blæsbjerg and Jerome for reading and giving very useful feedback on the thesis. My students, Jimmy Hoffman, Lars Bach, Martin Kornholt and Stig Rattleff also deserve thanks for being wonderfully enthusiastic and productive students. All of the people in the DNA repair group of which I have not mentioned Elvira Chapka, Lene Christiansen and Michael Lyng Nielsen also deserve warm thanks for helping me with various things, for being great company and for giving great feedback at the weekly group meetings. Finally, I would also like to thank my boyfriend Jakob for his love and support throughout the project – particularly for technical support during the intense writing-up process.

Lyngby, January 2009

Line Albertsen

Summary

Sequentially acting metabolic enzymes are often positioned in close proximity by the cell using three different methods: attachment to cellular structures, up-concentration in intracellular compartments or formation of large enzyme complexes. Positioning sequential enzymes in close proximity may accelerate reaction rates and this can prevent the loss of intermediates through diffusion, degradation or competing reactions. Most studies on the effect of enzyme proximity have been carried out *in vitro*. These studies have demonstrated that reaction rates may be increased by a factor of 1.5-2.5 when two sequential catalytic sites are positioned in the vicinity of each other by either fusing the genes encoding the enzymes or by attaching the enzymes to some kind of scaffold. Moreover, when experiments were carried out in viscous solutions designed to mimic *in vivo* conditions, the benefits of enzyme proximity were even greater. Despite these promising results, very few studies have systematically attempted to use enzyme proximity for optimization of metabolic pathways in cell factories.

The aim of this project is to investigate whether two model pathways can be optimized by positioning two sequentially acting enzymes in the vicinity of each other. Two different strategies for positioning enzymes in close proximity are tested. One method is to construct a polypeptide with two catalytic sites by fusing the genes encoding the enzymes. The other is to position enzymes on a protein-based platform that self-assembles *in vivo*. Two industrially relevant model systems leading to either sesquiterpene or vanillin glucoside production in yeast has been selected for testing the concepts.

In the first part of the project, it is demonstrated that both the sesquiterpene and the vanillin glucoside pathway can be optimized by enzyme fusion. With the sesquiterpene system, a two-fold higher concentration of end product is observed when the two sequential enzymes farnesyl diphosphate synthase and patchoulol synthase are fused. Moreover, it is shown to be advantageous to use the fusion strategy in combination with other traditional metabolic engineering tools, as down-regulation of a competing pathway and enzyme fusion combined results in an almost five-fold increase in the concentration of end-product. With the vanillin

glucoside pathway, fusion of O-methyl transferase and UDP-glycosyl transferase can effectively decrease the accumulation of the toxic intermediate vanillin. The decreased accumulation results in a higher growth rate but not in a higher production of end-product. While the effect of enzyme fusion is evaluated, the importance of enzyme orientation within the fusion protein is investigated. Here, it is shown that the benefits of enzyme fusion are largely dependent on the orientation, as no advantages of enzyme fusion are found for some specific configurations of fusions proteins. In connection with the testing of sesquiterpene system, the effect of the linker type is investigated and for this specific model system, no linker types that significantly affect the efficiency of the fusion protein are identified.

The second part of the project focuses on developing and testing a technology that enables assembly of sequential enzymes on a protein-based platform *in vivo*. The concept is in this project termed a nano-platform. Here, it is first tested whether the ring-structure forming DNA repair protein, Rad52, is suitable as a scaffold for forming the nano-platform core. Secondly the developed concept is tested with the two model systems mentioned above. In connection with the evaluation of Rad52's suitability as a scaffold, the mechanism responsible for Rad52's transport to the nucleus is studied in *Saccharomyces cerevisiae*. Two factors required for nuclear localization of Rad52 of *S. cerevisiae* (ScRad52) are identified. One is a nuclear localization signal consisting of three positively charge amino acid residues and a proline residue, the other is ScRad52's ability to self-associate. Based on these findings, an assay for detecting whether Rad52 of mouse (MmRad52) self-associates in yeast is developed. Using this assay, it is demonstrated that MmRad52 is suitable as a scaffold for assembling of enzymes.

In the last part of the project, the nano-platform concept is tested with the two model pathways leading to sesquiterpene and vanillin glucoside production in yeast. For both systems, proximity effects could be demonstrated when the enzymes are positioned on the nano-platform, but in both cases the system is not advantageous production wise over a system consisting of the free counterparts. In the case of the nano-platform tested in this project, the benefits of enzyme proximity cannot compensate for the negative effect of fusing enzymes to MmRad52.

In summary, it is demonstrated using two simple model systems that the production in cell factories can be optimized by enzyme proximity. This illustrates the great potential that the engineering of the spatial organization of pathways holds for metabolic engineering.

Sammendrag

Cellen anbringer ofte enzymer, som katalyserer på hinanden følgende trin i en syntesevej, i nærheden af hinanden ved at fæstne dem på cellulære strukturer, opkoncentrere dem i intracellulære organeller eller samle dem i enzymkomplekser. Ved at anbringe sekventielle enzymer tæt på hinanden er det muligt at øge den samlede reaktionshastighed og herved kan det forhindre at mellemprodukter mistes ved diffusion, nedbrydning eller til konkurrerende reaktioner. Hidtige studier af enzymproximitet er hovedsagelig udført *in vitro*. Disse studier har vist at reaktionshastigheder kan accelereres 1,5-2,5 gange når to katalytiske sites anbringes i nærheden af hinanden ved enten at fusionere generne eller ved at fæstne enzymerne til en form for platform. Ydermere er det vist at når sådanne forsøg udføres i viskøse opløsninger, som er designet til at efterligne *in vivo* forhold, er fordelene ved enzymproximitet endnu større. På trods af disse meget lovende resultater, er der imidlertid meget få, der systematisk har forsøgt at anvende enzymproximitet til optimering af synteseveje i cellefabrikker.

Målet med dette projekt er at undersøge om to modelsynteseveje kan optimeres ved at anbringe to sekventielle enzymer i nærheden af hinanden. To forskellige metoder til at anbringe enzymer i nærheden af hinanden testes. Den ene metode er at konstruere et polypeptid med to katalytiske sites ved at fusionere af generne, som koder for enzymerne. Den anden metode er at anbringe enzymerne på en protein-baseret platform, som samler sig selv *in vivo*. Som modelsystemer blev to industrielt relevante synteseveje valgt. De valgte synteseveje fører til henholdsvis sesquiterpen- og vanillinglukosidproduktion i gær.

I første del af projektet vises det at både sesquiterpen- og vanillinglukosidsyntesevejen kan optimeres ved enzymfusion. For sesquiterpensystemet vises det at koncentrationen af slutprodukt kan øges to gange ved at fusionere de to sekventielle enzymer farnesyldiphosphat syntase og patchoulol syntase. Ydermere bliver det vist at fusionsstrategien med fordel kan anvendes i kombination med mere traditionelle metabolic engineering strategier, idet det vises at nedregulering af en konkurrerende syntesevej og enzymfusion tilsammen øger produktionen næsten 5 gange. For vanillinglukosidsyntesevejen vises det at fusion af de to sekventielle

enzymet O-methyl transferase og UDP-glykosyl transferase er effektiv til at mindske ophobningen af det giftige mellemprodukt vanillin. Den formindskede ophobning af vanillin giver anledning til en højere vækstrate, men ikke en højere produktion af slutprodukt. I forbindelse med at effekten af enzymfusion bliver evalueret, bliver betydningen af enzymløsløsheden indenfor fusionsproteinet også undersøgt. Konklusionen er at fordelene ved enzymfusion afhænger af hvordan proteinerne fusioneres, idet der ikke er nogen fordel ved fusion for nogle af de testede enzymorienteringer. For sesquiterpensystemet bliver effekten af linkertypen også undersøgt. Her findes det at linkertypen ikke har den store betydning for effektiviteten af fusionsproteinet i dette modelsystem.

Anden del af projektet fokuserer på at udvikle og teste en teknologi som muliggør samling af sekventielle enzymer på en proteinbaseret platform *in vivo*. Konceptet betegnes i dette projekt en nano-plattform. I forbindelse hermed bliver det for det første undersøgt om det ringstrukturdannende DNA reparationsprotein Rad52 er egnet til at danne kernen af nano-plattformen og ydermere testes det udviklede koncept med de to ovenfor nævnte modelsystemer. I forbindelse med evalueringen af Rad52's anvendelighed som platformprotein bliver mekanismen, som er ansvarlig for transport af Rad52 til kernen i *S. cerevisiae*, studeret. Her findes det at der kræves to ting for at *S. cerevisiae* Rad52 (ScRad52) sorterer til kernen. Dels kræves et kernelokaliseringssignal bestående af tre positivt ladede aminosyrer samt en proline rest, dels kræves det at ScRad52 multimeriserer. På baggrund af denne viden udvikles et assay som kan benyttes til undersøge om Rad52 fra mus (MmRad52) multimeriserer når dette protein udtrykkes i *S. cerevisiae*. Ved hjælp af dette assay vises det at MmRad52 er egnet til at danne platformkernen.

I sidste del af projektet testes det om syntesevejene førende til sesquiterpen- og vanillinglukosidproduktion i gær kan optimeres ved at anbringe to udvalgte sekventielle enzymer på nano-plattformen. For begge systemer påvises der proksimitetseffekter når enzymerne er anbragt på nano-plattformen, men systemet er ikke konkurrencedygtigt med de frie enzymer rent produktionsmæssigt. Dette skyldes formodentlig at fordelene ved enzymproksimitet ikke kan opveje den negative effekt, som det har på enzymerne at de blev fusioneret til MmRad52.

Sammenfattet bliver det for to simple modelsystemer vist at man ved at anbringe sekventielle enzymer i umiddelbar nærhed af hinanden kan opnå fordele. Dette illustrerer det store potentiale som optimering af den rumlige organisering af synteseveje besidder som værktøj for metabolic engineering af cellefabrikker.

Nomenclature

3DSD	3-dehydro shikimate dehydrase
ACAR	Aromatic carboxylic acid reductase
C-terminal	carboxy-terminal
FPPS	Farnesyl diphosphate synthase
gnDNA	Genomic DNA
HsRad52	Human Rad52
MmRad52	Rad52 of mouse
Nano-platform	A structure consisting of a multimeric protein with metabolic enzymes linked to it.
N-terminal	amino-terminal
OMT	O-methyl transferase encoded by <i>hsOMT</i>
ORF	Open reading frame
PAC	Protocatechuic acid
PAL	Protocatechuic aldehyde
Platform subunit	A self-associating protein fused to a metabolic enzyme
PTS	Patchoulol synthase
P _{GALI}	The <i>GALI</i> promoter of <i>S. cerevisiae</i>
P _{GAL10}	The <i>GAL10</i> promoter of <i>S. cerevisiae</i>
Scaffoldin	A protein based scaffold
ScRad52	Rad52 of <i>S. cerevisiae</i>
UGT	UDP glycosyl transferase encoded by <i>UGT72E2</i>
VALS	Valencene synthase
VG	Vanillin glucoside

Contents

Preface.....	i
Acknowledgements.....	i
Summary.....	ii
Sammendrag.....	v
Nomenclature.....	viii
Introduction.....	1
Objectives.....	2
Outline of the thesis.....	3
Chapter 1: Spatial organization of metabolic enzymes.....	5
1.1. Benefits of enzyme proximity.....	5
1.2. Pathways likely to benefit from enzyme proximity.....	6
1.3. Organization of metabolic pathways in nature.....	8
1.3.1. Multifunctional enzymes.....	9
1.3.2. Enzyme complexes.....	10
1.3.3. Compartmentation.....	11
1.4. Engineering the spatial organization of metabolic pathways.....	12
1.4.1. <i>In vitro</i> approaches applied for studying enzyme proximity.....	12
1.4.2. <i>In vivo</i> studies of the effect of enzyme proximity.....	16
1.4.3. Effect of linker length and composition.....	18
Chapter 2: Enzyme fusion for optimization of sesquiterpene production in <i>S. cerevisiae</i>	20
Chapter 3: Enzyme fusion for optimization of vanillin glucoside production in <i>S. cerevisiae</i>	49
3.1. Introduction.....	49
3.2. Materials and Methods.....	52
3.2.1. Strains and media.....	52

3.2.2. Plasmid construction.....	53
3.3.3. Shake flask experiments	54
3.3. Results.....	55
3.3.1. Enzyme fusion results in decreased intermediate accumulation and increased growth rates.....	55
3.4. Discussion	57
Chapter 4: Introduction to the nano-platform concept.....	59
4.1. Requirements to the scaffoldin.....	60
4.2. Advantages of a self-assembling nano-platform.....	60
4.3. Rad52 as a scaffoldin candidate	61
Chapter 5: Evaluation of Rad52's applicability as a scaffoldin for enzyme assembly <i>in vivo</i>	63
5.1. Transport mutants of ScRad52 as scaffoldin candidates.....	74
5.2. MmRad52 as a scaffoldin candidate for enzyme assembly in <i>S. cerevisiae</i>	74
5.3. Materials and Methods	79
Chapter 6: Effect of assembling metabolic enzymes on the nano-platform	81
6.1. Introduction	81
6.2. Materials and methods	83
6.2.1. Strains and media.....	83
6.2.3. Shake flask experiments	88
6.2.4. Fluorescent microscopy	88
6.3. Results	88
6.3.1. FPPS and PTS are functional <i>in vivo</i> when they are fused to the scaffoldin.....	88
6.3.2. Attachment of both enzymes to the nano-platform results in an increased patchouliol production compared to when only one enzyme is attached	90
6.3.3. Effect of linker type on the functionality of the scaffoldin linked PTS	91

6.3.5. The platform forms aggregates when expressed from a multi copy plasmid in <i>S. cerevisiae</i>	93
6.3.6. Aggregate formation is not related to the patchoulol producing platform	95
6.3.7. Decreasing the expression level provides no additional benefits of enzyme assembly on the platform.....	96
6.3.8. Test of the nano-platform concept with the vanillin glucoside pathway as a model system	97
6.4. Discussion	99
Chapter 7: Conclusion.....	103
7.1. Future perspectives.....	104
References.....	106

Introduction

Nature holds a great treasury of valuable compounds that we use as e.g. medicine, cosmetics and food additives. The production of such valuable compounds by conventional chemical methods or purification from the biological source is often difficult and expensive (Chang and Keasling, 2006). Modern recombinant DNA technology allows genes to be transferred from one organism to another. In this way, biosynthetic pathways can be transferred to more suitable production host organisms such as the yeast *S. cerevisiae*. The production in such cell factories holds several advantages over conventional methods because cell factories are less energy demanding, creates less toxic waste in the form of solvents and metal catalysts, uses renewable “building” materials and facilitates purification (Chemler et al., 2006).

In nature, clusters of biosynthetic enzymes are often found in the vicinity of each other. By positioning enzymes in close proximity the cell can relieve kinetic constraints that result from dilution of intermediates and minimise loss of intermediates through degradation and competing pathways. When novel pathways are introduced into a cell factory, they depend on the enzymatic apparatus of the host as well as on the newly introduced enzymes. In this case no spatial organization can be expected to be in place. Presumably, this contributes to novel pathways suffering low productivity and yields as well as the formation of undesired side-products. Metabolic engineering aims at solving these problems using a rational, directed approach. In this context, modulating the gene expression and activity of enzymes are tools that have been used extensively. Another approach that is gaining interest is the engineering of the spatial organization of pathways. Although this strategy is used extensively by the cell to efficiently regulate metabolism, it has been used to a very limited extent by metabolic engineers. A researcher working in the field did however recently state that ignoring the spatial arrangement of metabolic pathways when constructing novel pathways “is like laying down the pipes without connecting them” (Keasling, 2008), implying that failing to consider the spatial arrangement is neglecting some of the most important options for improving cell factories.

Objectives

This project aims at testing whether engineering the spatial organization of pathways can be used to optimize cell factories by placing enzymes catalyzing sequential steps in close proximity. Two different strategies for positioning enzymes in the vicinity of each other are tested:

- 1) End to end fusion of the genes encoding the enzymes
- 2) Attachment of enzymes to a protein-based scaffold.

As model systems, two industrially relevant pathways were chosen. One that leads to sesquiterpene production and one that leads to vanillin glucoside production in yeast.

Outline of the thesis

The thesis is divided into the following chapters:

Chapter 1: Describes the kinetic benefits of enzyme proximity and highlights the types of pathways that are most likely to gain from enzyme proximity. Secondly, examples of the cells different strategies for spatially arranging pathways are given and lastly the attempts to exploit spatial organization for metabolic engineering purposes are reviewed.

Chapter 2: Demonstrates how enzyme fusion can be used to optimize sesquiterpene production in *S. cerevisiae*. The effect of enzyme fusion on the final titer of sesquiterpenes and farnesol is examined. Moreover, the importance of enzyme orientation and linker type is investigated. This chapter is written in the format of a manuscript and will be submitted to an international journal.

Chapter 3: Illustrates how enzyme fusion may be used to optimize vanillin production in *S. cerevisiae*. The effect of enzyme fusion on intermediate pools, final titers and growth rates are examined.

Chapter 4: Explains the concept that was applied for testing the effect of assembling sequentially acting metabolic enzymes on a protein-based nano-platform *in vivo*.

Chapter 5: Deals with the identification and verification of a nano-platform core. In this chapter the applicability of using two different species of Rad52 as a scaffoldin for assembly of metabolic enzymes *in vivo* is evaluated. A suitable candidate is identified and the concept of nano-platform mediated enzyme assembly is verified using YFP and CFP as model enzymes. In connection with the evaluation of Rad52's suitability as a scaffoldin, the localization and transport of Rad52 is studied. The chapter is mainly based on the publication:

Plate, I., Albertsen, L., Lisby, M., Hallwyl, S., Feng, Q., Seong, C., Rothstein, R., Sung, P., and Mortensen, U. (2008). Rad52 multimerization is important for its nuclear localization in *Saccharomyces cerevisiae*. *DNA Repair*, 7(1):57–66.

Chapter 6: Is dedicated to the testing of the concept developed in Chapter 5 with two model pathways. The pathways leading to either sesquiterpene or vanillin glucoside production in yeast served as model systems. The effect of nano-platform mediated enzyme assembly on growth rates, final titers and intermediate pools were examined and some of the problems encountered when using this approach were identified.

Chapter 7: Concludes the thesis by summarizing the obtained results. This chapter also proposes strategies for further research.

Chapter 1: Spatial organization of metabolic enzymes

The cell often positions sequentially acting enzymes in close proximity. In this way the cell can accelerate reaction rates and thereby gain an extra level of metabolic control. In this chapter it will be explained how reactions can be accelerated by enzyme proximity and moreover the types of pathways that are most likely to gain from this will be highlighted. Furthermore, examples of how the cell employs different strategies to spatially arrange pathways will be given and lastly the engineering attempts carried out so far are reviewed.

1.1. Benefits of enzyme proximity

One immediate consequence of positioning consecutive enzymes in close proximity is that a locally high concentration of substrate builds up in the vicinity of the enzyme catalysing the next step in the reaction (See Fig. 1-1A and B). According to the Michaelis-Menten model of enzyme kinetics, increases in substrate concentration generally accelerate the reaction rate until the enzyme becomes saturated (see Fig. 1-1C).

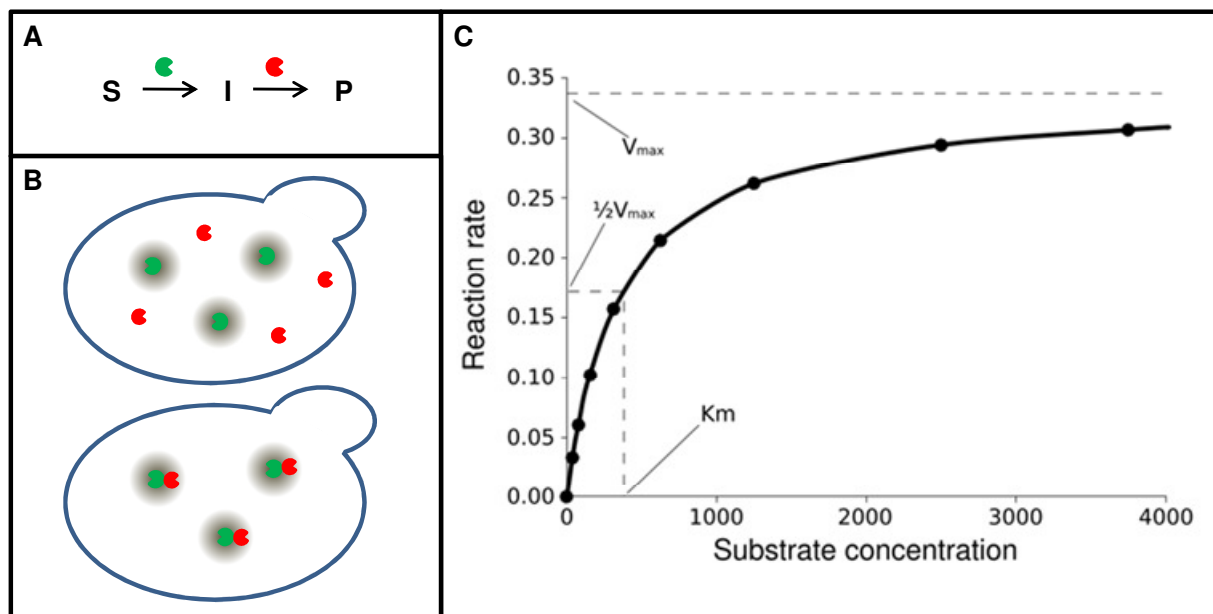


Figure 1-1. Proximity of sequentially acting enzymes increases the substrate concentration and thereby the reaction rate. A) A metabolic pathway where a substrate is converted to a product by two sequential enzymes. B) A locally higher concentration of the intermediate (grey cloud) builds up around the enzyme catalysing the second step in the pathway (shown in red), if the two sequential enzymes are positioned in the vicinity of each other. C) Michaelis-Menten saturation curve. The reaction rate increases with the substrate concentration until a saturation point is

reached.

The Michaelis-Menten constant, K_m , denotes the substrate concentration at which the reaction rate is half of its maximal value (Fig. 1.1C). As physiological substrates concentrations generally are in the range of 1/100 to 1 of K_m (Stryer, 1995) most reactions can be accelerated by increases in substrate concentration. In fact it can be shown that when substrate concentrations are much lower than K_m , reaction rates increase linearly with both $[S]$ and K_{cat}/K_m , where $[S]$ denotes the substrate concentration and K_{cat} denotes the turnover number of the enzyme, i.e. the number of substrate molecules that are turned into product when the enzyme is fully saturated. Several natural enzymes have evolved to have high turnover numbers. Under these conditions the ultimate limit of K_{cat}/K_m is set by the rate at which the enzyme-substrate complex is formed. This rate cannot be faster than the diffusion controlled encounter of enzyme and its substrate (Stryer, 1995). In this case catalytic gains can only come from decreases in diffusion time. One way that diffusion rates can be decreased is by shortening the distance the intermediates have to migrate by positioning enzymes catalyzing sequential steps in close proximity. Accordingly, reactions that have been optimized through perfection of the enzymes capacity for turning over product may be even further accelerated by proximity of sequentially acting enzymes.

1.2. Pathways likely to benefit from enzyme proximity

When a substrate is converted to a product of interest via an intermediate, an increase in the overall reaction rate will generally decrease the transit time of the intermediate and thereby the life-time and accumulation of the intermediate can be reduced. These consequences of enzyme proximity can be envisioned to be particularly advantageous for several types of pathways (Reviewed in Jørgensen et al., 2005). These include pathways with (see also Fig. 1-2):

Competition for intermediate. For pathways with competition for the intermediate, an accelerated conversion of substrate to product can minimize loss of intermediate to competing reactions. Competition may arise from non-enzymatic or enzymatic reactions that convert the intermediate to undesired by-products.

Unstable intermediates. If an intermediate is unstable an accelerated reaction rate can ensure that the intermediate is converted to more stable compound before it is lost through degradation. Intermediates may also be lost through diffusion.

Toxic intermediates. Toxic intermediates can inhibit a process by inhibiting cell growth. When dealing with toxic intermediates, it can be beneficial for the cell to maintain a locally high concentration around the enzymes that convert the intermediate while keeping total cellular levels low, as this will minimize the toxic stress imposed on the cell.

Intermediates that exert inhibitory effects. Intermediates may for example inhibit a pathway by directly or indirectly mediating negative feedback on a preceding step. In this case less accumulation of intermediates can prevent down-regulation of the pathway.

Competition for enzyme. Several cellular enzymes are capable of converting a variety of substrates rather than just one. By maintaining a high concentration of the favoured intermediate around such a promiscuous enzyme, other substrates or compounds exerting an inhibitory effect may be prevented from reaching the active site. In this way the desired reaction may be favoured.

As the metabolism consists of a network of reactions that are widely interconnected, the flux through a metabolic pathway is often affected by several of the above mentioned limitations. Probably, this also explains why the cell has several strategies for dealing with these types of problems.

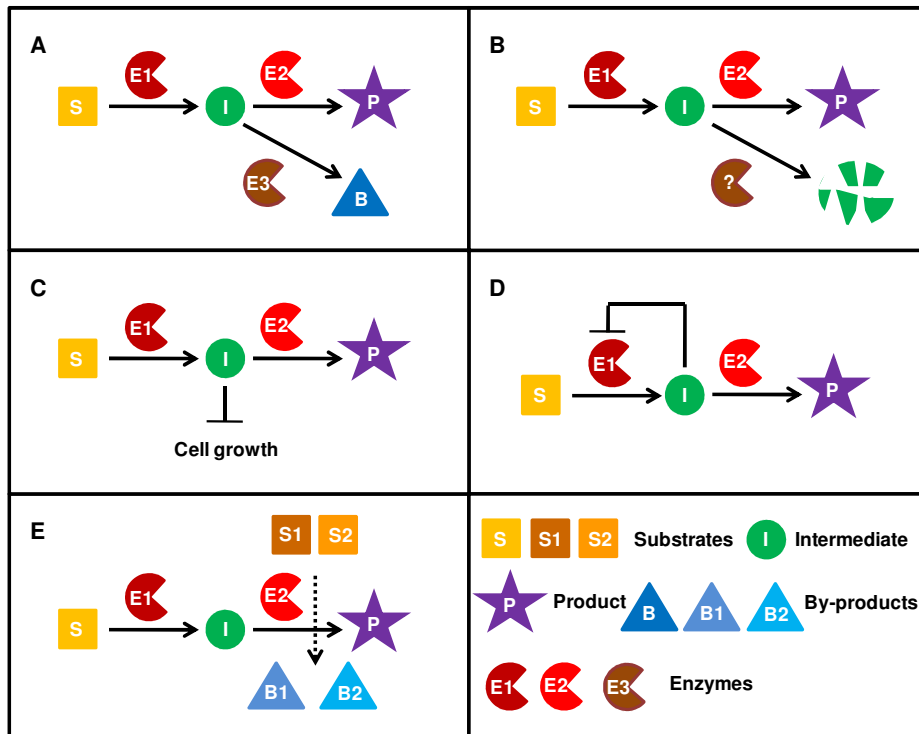


Figure 1-2. Overview of the pathways most likely to benefit from enzyme proximity. A) Pathways with competition for intermediate from either enzymatic or non-enzymatic reactions B) Pathways with an unstable intermediate that is susceptible to degradation. C) Pathways with a toxic intermediate. D) Pathways where the intermediate inhibits the reaction by exhibiting direct or indirect product inhibition on a preceding catalytic step or E) Pathways with competition for enzyme.

1.3. Organization of metabolic pathways in nature

Accumulating evidence suggests that the cell employs spatial organization in order to efficiently regulate metabolism (Conrado et al., 2008). Instead of being randomly distributed within the cell, sequential enzymes often colocalize (Ovadi and Srere, 2000, Ovadi and Saks, 2004).

Colocalization may for example be achieved by sorting enzymes to the same organelle, by docking them on cellular structures or by forming large complexes. Alternatively, proximity of catalytic sites may be ensured by encoding them within the same gene. Here, some of the most remarkable examples across these themes will be described to illustrate nature's intriguing ways of optimizing and regulating metabolism using spatial organization (see also Fig. 1-3).

1.3.1. Multifunctional enzymes.

Multifunctional enzymes that contain one or more catalytic sites are widespread in nature and most often catalyze two or more consecutive reactions in a metabolic pathway (Duncan et al., 1987). Comparative genomics indicate that they at least in some cases have evolved through spontaneous fusions of separate genes. For instance, the five central steps in the shikimate pathway are catalyzed by one large pentafunctional protein, Arom, in yeast (Fig. 1-3A), whereas the same reactions are carried out by five monofunctional enzymes in *E. coli* (Duncan et al., 1987). Sequence comparison of the *E. coli* genes and the yeast gene, showed that there were clear sequence similarities between the five functional domains of the yeast protein and the five separate *E. coli* proteins indicating that the yeast protein have evolved through multiple gene fusions of common ancestral genes (Duncan et al., 1987). The evolutionary driving force behind the creation of multifunctional enzymes is presumably the benefits that can be achieved by positioning catalytic sites in close proximity (See section 1.2.). Furthermore multifunctional enzymes may resemble structures where substrates can be transferred directly from one catalytic site to the next without getting in contact with the cellular matrix – a phenomenon commonly referred to as substrate channelling or metabolic channelling.

The synthase catalyzing the last two steps in the biosynthesis of tryptophan is a remarkable example of an enzyme displaying such a channelling feature. The bifunctional protein resembles a structure where its two catalytic sites are connected by a tunnel with a diameter that exactly matches the size of the intermediate indole (Dunn et al., 2008). This tunnel allows indole to be transferred from one catalytic site to the next within the molecule without any contact with the cellular matrix (Fig. 1-3B). For tryptophan synthase of *Salmonella typhirium*, this type of substrate channelling was shown to increase the reaction rate by 1-2 orders of magnitude (Hyde et al., 1988). This result demonstrates the great benefits that can be achieved by transferring substrate directly from one catalytic site to the next instead of letting substrates migrate by diffusion.

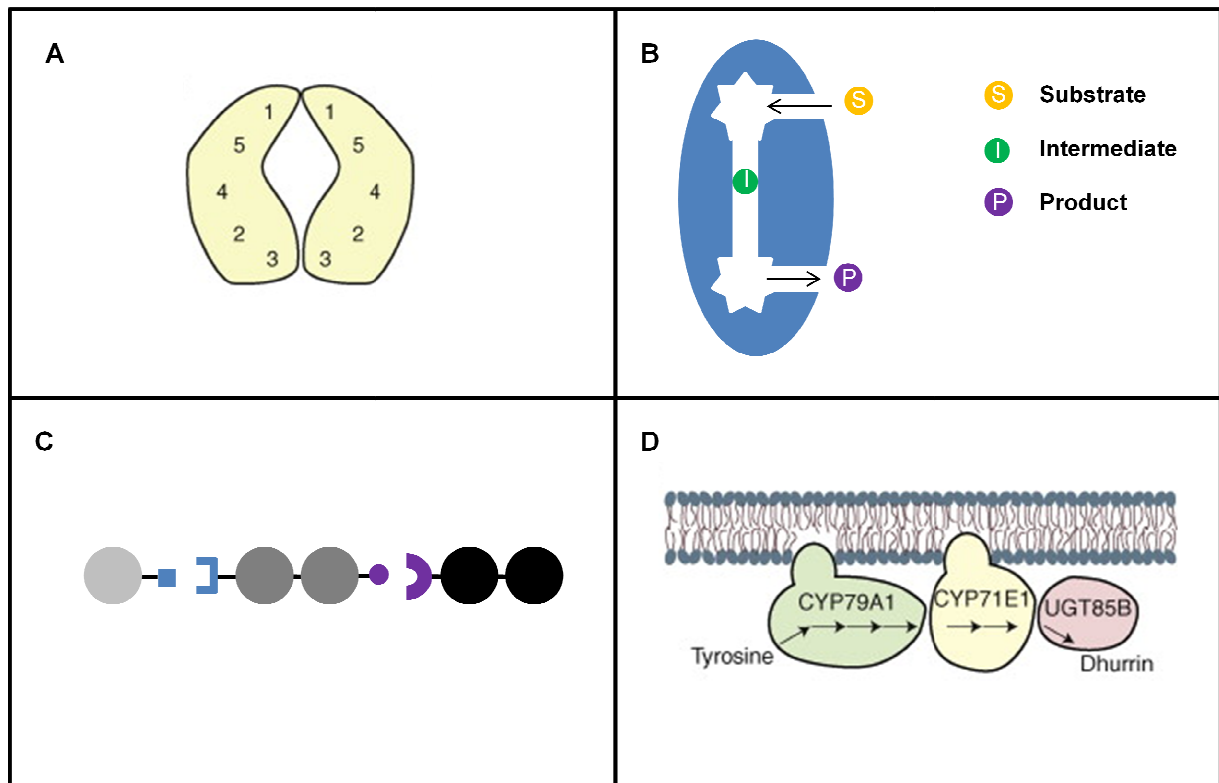


Figure 1-3. The cells strategies for positioning catalytic sites in close proximity. **A)** The Arom dimer with five catalytic sites within the same polypeptide (Conrado et al., 2008). **B)** Loss of intermediate through diffusion is hindered, by direct channelling of substrate between the two catalytic sites of tryptophan synthase (Adapted from Dunn et al., 2008). **C)** In polyketide synthesis, proximity of catalytic domains (represented as large circles) is ensured both by encoding several catalytic sites within the same polypeptide chain and by letting these protein modules associate via specific docking domains into even larger complexes often referred to as assembly lines, **D)** In the dhurrin pathway, proximity of catalytic sites is ensured both by encoding more than one catalytic site within the same polypeptide and by letting sequential enzymes dock on the ER membrane. Arrows indicates reaction steps (Conrado et al., 2008).

1.3.2. Enzyme complexes

A large number of sequential enzymes from both primary and secondary metabolism have been demonstrated to interact and form multifunctional complexes often referred to as metabolons (Ovadi and Srere, 1996, Conrado et al., 2008). The interactions supporting such enzyme complexes may vary in strength and this allows for dynamic pathway regulation (Ovadi & Srere 1996). Biosynthesis of purine is for example carried out by a complex consisting of six enzymes that was found to associate in response to purine depletion and disassociate when purine levels

were sufficient (An et al., 2008). Accordingly, the cell can gain an extra level of metabolic control by assembling catalytic sites in response to cellular signals.

In another type of enzyme complexes, interactions between sequential enzymes ensure that biosynthesis proceeds in an orderly fashion. For example, the synthases involved in the production of polyketides form some of the most well-organized enzyme complexes in nature. These complexes consist of several functional modules that each performs one cycle of chain elongation. The different modules are often encoded by more than one gene, but this does not prevent them from cooperating, as the functional modules in this case associate via docking domains located at their termini to form large assembly lines where the molecule to be synthesized is passed sequentially between functional domains (Weissman, 2006) (Fig. 1-3C). Here, specific recognition between acceptor and donor parts of docking domains ensure that functional domains assemble in an orderly fashion so random chain elongation is prevented.

Enzyme complexes may also assemble on cellular structures such as the cytoskeleton or internal membranes. The pathway leading to synthesis of dhurrin is a good example of a pathway where this kind of spatial organization is applied. Dhurrin is synthesized in seven steps from tyrosine by two multifunctional enzymes that dock on the ER membrane and one soluble enzyme that glycosylates the final intermediate (Nielsen et al., 2008) (Fig. 1-3D). The last enzyme of the pathway is believed to interact with the enzyme catalyzing the preceding step (Nielsen et al., 2008) and in this way, close proximity of all seven catalytic sites are ensured. As the pathway involves a number of labile intermediates, the short distance between the catalytic sites supposedly ensures that intermediates are swiftly converted before degradation can occur.

1.3.3. Compartmentation

In eukaryotic cells, spatial pathway organization may also be achieved by sorting enzymes to a membrane enclosed compartment, such as the mitochondria, ER or the microbodies (Ovadi and Saks, 2004). Such compartmentation may serve a number of purposes. Intermediates can be up-

concentrated, a favourable microenvironments that provide optimal conditions for a specific reaction can be created or intermediates can be protected from unwanted enzymatic conversions or degrading environments. In the terpene pathway of plants, compartmentation is for example applied to segregate the pathway branches that lead to synthesis of various types of terpenes, as the steps following the common terpene precursor IPP is carried out in various subcellular compartments. Monoterpenes are for instance synthesized in the plastids, sesquiterpenes in the cytoplasm, sterols and dolichol in ER or the ER membrane and ubiquinone in the mitochondria (Bouwmeester, 2006). In this case, the cell can presumably regulate the distribution of the common precursors among the different pathway branches by controlling their transport in and out of subcellular compartments.

1.4. Engineering the spatial organization of metabolic pathways

Natures intriguing examples of spatial arrangement of pathways have inspired people to study the effect of enzyme proximity using different approaches. Most of this work has been carried out *in vitro* and several of these studies have demonstrated that reaction rates can be increased by proximity of sequential enzymes. Moreover, *in vitro* studies has identified interesting approaches for positioning enzymes in close proximity of which some easily can be adapted for *in vivo* applications. However, despite these promising results, very few studies have systematically evaluated the effect of enzyme proximity *in vivo*. In this section the different approaches for positioning enzymes in close proximity will be described, and the results obtained with them will be summarized.

1.4.1. *In vitro* approaches applied for studying enzyme proximity

The approaches used for studying enzyme proximity *in vitro* include fusion of enzymes or parts of enzymes, interaction via docking domains or assembly on supports via chemical tethering or docking domains (See Fig. 1-4).

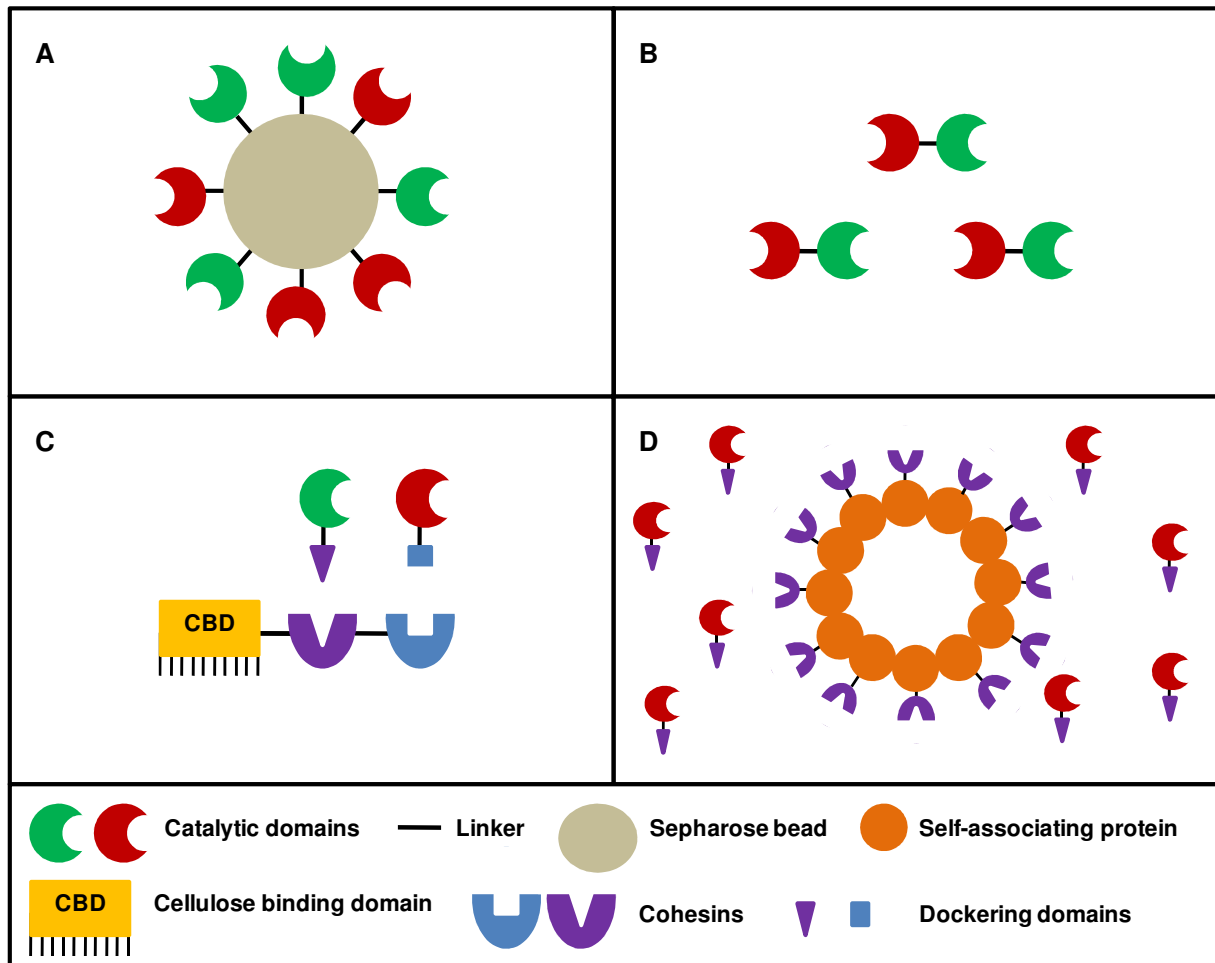


Figure 1-4. Approaches used for studying the effect of enzyme proximity *in vitro*. A) Covalent linking to a scaffold , B) Fusion of enzymes via a linker, C) Assembly on a scaffold protein via docking domains (adapted from Mingardon et al., 2007b) , D) Assembly on a self-associating protein via docking domains.

1.4.1.1. Enzyme assembly on supports via chemical tethering

The first attempts to study enzyme proximity focused on assembling enzymes on solid supports by covalently linking them to polymer beads such as e.g. sepharose beads (Fig. 1-4A). When two different pairs of sequential enzymes were immobilized in this way, the reaction rate was increased 10-140% compared to comparable amounts of free enzymes (See also Table 1) (Mosbach and Mattiasson, 1970, Srere et al., 1973). The advantage of being attached to the beads was largest when the substrate concentrations were low and no positive effect of attachment to the beads was observed when saturating concentrations of substrate were used. In one of these studies the cofactor NAD^+ is consumed in the first reaction step (Srere et al., 1973). Here, an

even greater advantage of enzyme immobilization was observed when a third enzyme that regenerates NAD⁺ was linked to the beads as well. In summary this indicates that the benefits of enzyme assembly on supports are achieved through the establishment of a favourable microenvironment around the enzymes with a locally higher substrate and/or cofactor concentration.

While covalent linking of enzymes to solid supports provides the evidence that pathways can benefit from spatial arrangement of sequential enzymes, the usefulness of the approach *in vivo* is limited. However, more recent experimental work has identified strategies that are more suitable for *in vivo* applications.

Table 1-1: Effect of enzyme proximity when different methods for positioning catalytic sites in close proximity was used. All studies were carried out *in vitro*.

Pathway	Enzymes (organism)	Method	Effect ^a	Reference
Pentose phosphate	HK (yeast)	Covalent linking to scaffold	1.4-2.4	Mosbach & Mattiasson, 1970
Citric acid cycle	G6PD (yeast) MD(pig) CS (pig)	Covalent linking to scaffold	1.1-2.3	Srere et al., 1973
Sesquiterpene synthesis	FPPS (<i>Artemisia annua</i>) eAS (<i>Nicotiana tabacum</i>)	Fusion	2.5	Brodelius et al., 2002
Formaldehyde fixation	HPS (<i>Mycobacterium gastri</i>) PHI (<i>M. gastri</i>)	Fusion	2	Orita et al., 2007
Lactose metabolism	β-gal (<i>Escherichia coli</i>) galDH (<i>Pseudomonas fluorescens</i>)	Fusion	2	Carlsson et al., 1996
Trehalose synthesis	TPS (<i>E. coli</i>) TPP (<i>E. coli</i>)	Fusion	1.6	Seo et al., 2000
Cellulose hydrolysis	Cel6A (<i>Neocallimastix patriciarum</i>) Cel9G (<i>Clostridium cellulolyticum</i>)	Assembly on scaffold via docking domains	2.5	Mingardon et al. 2007a
Cellulose hydrolysis	Cel6A (<i>N. patriciarum</i>) Cel9G (<i>C. cellulolyticum</i>) Cel48F (<i>C. cellulolyticum</i>)	Assembly on scaffold via docking domains	3.1	Mingardon et al. 2007a

^aThe effect of enzyme proximity is measured as the fold increase in the overall reaction rate, when enzymes engineered to be in the vicinity of each other are compared to equal amounts of their separate counterparts.

1.4.1.2. Fusion proteins

The introduction of modern genetic techniques allowed catalytic sites of sequential enzymes to be positioned in close proximity by simply fusing the structural genes and thereby generate a polypeptide with more than one catalytic site. Several *in vitro* studies with purified enzymes have shown that the overall reaction rate can be increased up to 2.5 fold by fusion of sequential enzymes (See Table 1-1). Moreover, experiments carried out with fusion proteins in polyethylene glycol solutions to mimic the crowded environment inside the cell, indicate that the positive effects of enzyme fusion may be even greater *in vivo* (Yilmaz and Bulow, 2002, Prachayasittikul et al., 2006, Bülow, 1987).

1.4.1.3. Assembly on scaffolds via docking domains

In an approach inspired by nature's way of organizing cellulases on their substrate cellulose, enzymes have been assembled on protein-based scaffolds (scaffoldins) via docking domains (Fig. 1-4C). Cellulose degrading organisms often express an arsenal of cellulases (Bayer et al., 2004). In anaerobic microorganisms these enzymes are organized in cellulosomes which consists of a scaffoldin possessing a cellulose binding domain and several cohesin domains (Gilbert, 2007). The cohesin domains are specifically recognized by docking domains that are expressed as part of the cellulases. The recognition between cohesion domains and docking domains is compatible within species, but incompatible among at least some species of cellulose degrading microorganisms (Pages et al., 1997). This can be exploited to position enzymes at specific locations on a chimeric scaffoldin displaying cohesin domains from different species (Fig. 1-4C). When a fungal and a bacterial endoglucanase were positioned on a chimeric scaffoldin by mixing the purified components *in vitro* a 2.5 fold increase in released sugars was observed (Table 1-1) (Mingardon et al., 2007a). Moreover, accelerated substrate breakdown was observed when an additional cellulase displaying endo-processive activity was incorporated in the scaffold. This study was not carried out with sequentially acting enzymes but with enzymes that work simultaneously on their substrate. Thus, the synergistic effect is not thought to arise from kinetic benefits. Instead it is believed to result from the different cellulases ability to enhance each other's access to substrate and thereby enable hydrolysis of additional sites. The system has

been demonstrated to work with various types of cellulases from different organisms (Caspi et al., 2008, Fierobe et al., 2001, Fierobe et al., 2002, Fierobe et al., 2005) and can presumably also be used for other types of enzymes.

In a further development of the cellulosome inspired concept, cohesin domains were fused to the self-associating protein, SP1, capable of forming ring-structures. When SP1 displaying cohesin domains was mixed *in vitro* with a cellulase harbouring a compatible docking domain, large SP1 complexes exhibiting cellulase activity were observed indicating that the cellulase had successfully attached itself to the ring-structure via its docking domain (Fig. 1-4D) (Heyman et al., 2007).

Although, the cellulosome inspired strategies seems to be applicable *in vivo*, no one has reported such experiments. In the study with SP1, purification of the protein fused to cohesin domains was compromised by the fusion protein's tendency to form inclusion bodies indicating that there are problems to be solved before proper assembly and functioning of enzymes *in vivo* is a reality. But in summary, *in vitro* studies have identified interesting strategies that remain to have their potential demonstrated *in vivo*.

1.4.2. *In vivo* studies of the effect of enzyme proximity

Most attempts to optimize cell factories by positioning enzymes in close proximity have been carried out with fusion proteins. Despite the promising results obtained with fusion enzymes *in vitro* (Table 1-1) and the ease with which the technique can be applied *in vivo*, the reported use of fusion enzymes for pathway optimization has been limited. The few systematic studies that have been carried out have generally resulted in decreased titers or yields (Table 1-2). However, one noteworthy study did identify a fusion protein displaying superior properties useful in metabolic engineering (Meynial Salles et al., 2007). This study was carried out with an *E. coli* strain that had been genetically engineered to couple glycerol production to growth. The strain expressed two enzymes from the glycerol pathway of *S. cerevisiae* from an artificial operon

located on a plasmid. When the strain was evolved in a chemostat, a faster growing clone was identified and subsequent sequencing of the plasmid revealed that this clone harboured a spontaneous deletion between gene 1 and 2 that had resulted in an in frame fusion of the genes. Further characterization of the strain expressing the fusion protein showed that for these two enzymes, fusion resulted in a five-fold reduced intermediate concentration and a two-fold increased production rate. This result demonstrates that enzyme fusion may be a useful tool for optimization of pathways in cell factories.

Table 1-2: *In vivo* studies on the effect of enzyme fusion

Pathway	Enzymes (organism)	Host organism	Effect ^a	Reference
Sesquiterpene synthesis	FPPS (avian) PTS (<i>Pogostemon cablin</i>)	<i>N. tabacum</i>	0.2-0.4	Wu et al., 2006
Polyhydroxybutyrate production	PhaA (<i>R. eutropha</i>)	<i>E. coli</i>	0.2	Kourtz et al., 2005
	PhaB (<i>R. eutropha</i>)	<i>A. thaliana</i>	0.7	
Cellulose hydrolysis	XhynX (<i>C. thermocellum</i>)	<i>E. coli</i>	~0.5-1	An et al., 2005
	Cel5Z::Ω (<i>P. chrysanthemi</i>)			
Glycerol production	GPD1 (<i>S. cerevisiae</i>)	<i>E. coli</i>	1 ^b	Meynial Salles et al., 2007
	GPP2 (<i>S. cerevisiae</i>)			

^a Effect is measured as the fold increase in product formed compared to the production from the free enzymes where a value of 1 means no difference.

^b The productivity is approximately 2 times higher because the growth rate is twice as high for the strain expressing the fusion protein. The production per gram biomass is however similar for the fusion and the individual enzymes.

Moreover, fusion proteins have been used with success within a lot of other research fields for e.g. studying localization (Huh et al., 2003) and protein interactions (Fields and Song, 1989), for enhancing solubility and expression (Chatterjee and Esposito, 2006) and for facilitating protein purification (Sheibani, 1999). These studies confirm that proteins often can be fused to other proteins or parts of proteins and remain functional *in vivo*. For the above mentioned applications it is however not necessary for the enzymes to retain 100 percent activity, but this becomes important when the aim is to use fusion enzymes for metabolic engineering purposes as a successful metabolic engineering strategy is dependent on full activity of the enzymes in order to benefit from the fusion strategy. Specific activities of fused enzymes have been demonstrated to be within the range of 0-100% (Bülow and Mosbach, 1991, Carlsson et al., 1996, Hong et al., 2006, Orita et al., 2007). From the *in vitro* studies of fusion proteins, the benefits of enzyme fusion are expected to be within the range of 1.5-2.5 fold (Table 1-1). Accordingly, the fusion

enzymes probably need to retain approximately 50% of their activity in order to observe any benefits and this may explain why attempts carried out to date have been unsuccessful. One factor that has been demonstrated to affect the functionality of fusion enzymes is the sequence used to link the two proteins to be fused.

1.4.3. Effect of linker length and composition

All the approaches suitable for studying the effect of enzyme proximity *in vivo* described above require the enzymes to be fused to either another protein or to a docking domain. When fusing genes or parts of genes a linker sequence is often inserted in between the ORFs. The function of the linker is to serve as spacer between the individual domains to prevent them from interacting and interfering with each other's folding and functionality. Linker length and composition has been demonstrated to affect protein stability, folding rates, specific activities and the ability to interact with other subunits or proteins (Prescott et al., 1999, Carlsson et al., 1996, George and Heringa, 2002). For an artificial fusion enzyme, the specific activity could for example be increased 50% by using longer linkers of 9-13 amino acid residues instead of using 3 amino acid residues (Carlsson et al., 1996). In this study, one of the fused enzymes was capable of forming tetramers and when the long linkers were used, a higher percentage of fusion protein was found in large complexes matching the sizes of multimers, indicating that the longer linkers provide the freedom in movement required for subunit interaction.

A study conducted to identify tendencies among interdomain linkers in natural multifunctional proteins found that natural linkers on average are 10 amino acids long with the majority being between 2 and 12 amino acids in length (George and Heringa, 2002). Natural linkers were also shown to have a preference for certain amino acids with proline being the most preferred. Short proline rich stretches are known to be stiff and proline is the only amino acid with no amide bound hydrogen to donate in hydrogen bonding. Thus, the authors speculated that proline is favoured because it provides a inert spacer that structurally isolates the linker from the rest of the protein (George and Heringa, 2002).

When generating artificial fusion enzymes, two main types of linkers have typically been used. These linkers are either rigid or flexible in secondary structure depending on the amino acid composition. Rigid linkers of variable length may be generated from the helical peptide EAAAK repeated 1-5 times (Arai et al., 2001). Flexible linkers are typically rich in glycine and serine and may for example be generated from the peptide GGGGS that can be repeated if increased length is desired. Experimental data indicates that rigid linkers are more efficient in separating domains, as the distance between fused proteins was shown to increase with the length of rigid linkers, whereas the length of flexible linkers did not affect the distance (Arai et al., 2001).

As the large variation among natural linkers in both composition and size indicates, the linker needed for optimal folding and functioning is probably largely dependent on the individual properties of the proteins to be fused. Some may require rigidity, others flexibility. In conclusion, an optimal linker should be the best compromise between a linker that provides the separation needed for correct folding and the flexibility needed for proper functioning. Lastly, linkers should not be susceptible to degradation by proteases.

Chapter 2: Enzyme fusion for optimization of sesquiterpene production in *S. cerevisiae*

In this chapter and the following chapter, it was tested whether fusion of sequentially acting metabolic enzymes can be used for optimization of metabolic pathways. Two different pathways that both depended on host as well as heterologous enzymes were selected as model systems. One that leads to sesquiterpene production in yeast and one that leads to vanillin glucoside production in yeast. The results obtained with the sesquiterpene pathway as a model pathway is presented in a manuscript with the title “Fusion of host and heterologous enzymes diverts flux towards sesquiterpene production in *S. cerevisiae*”. The results obtained with the vanillin glucoside pathway are presented in Chapter 3.

Fusion of host and heterologous enzymes diverts flux towards sesquiterpene production in *S. cerevisiae*

Running head: Fusion enzyme increases sesquiterpene production

Line Albertsen¹, Jerome Maury¹, Lars Bach¹, Jens Nielsen², Uffe H. Mortensen^{1*}

¹Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, Building 223, DK-2800, Kgs. Lyngby, Denmark.

²Systems Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

* Corresponding author. Phone: +45 45252701. Fax: +45 4588 4148. E-mail: um@bio.dtu.dk

Keywords: bifunctional enzyme, gene fusion, patchoulol synthase, metabolic engineering

Abstract

The ability to transfer metabolic pathways from a normal producer organism to the well-characterized cell factory *Saccharomyces cerevisiae* is well documented. However, as many secondary metabolites often are produced by several collaborating enzymes in a metabolon, metabolite production in yeast may be limited by the inability of the native yeast enzymes to collaborate with the heterologously expressed enzymes. To bypass these problems, fusion proteins consisting of farnesyl diphosphate synthase (FPPS) of yeast origin and patchoulol synthase (PTS) of plant origin (*Pogostemon cablin*) were constructed and expressed in *S. cerevisiae*. Production of the main sesquiterpene produced by PTS, patchoulol, was two-fold higher when FPPS and PTS were fused compared to when they were expressed as individual enzymes from the same promoters. The higher production of patchoulol from the strain expressing the fusion protein FPPS-PTS was not due to a generally higher expression level of the fusion protein, as tagging of the free enzymes and the fusion enzyme with fluorescent proteins showed that expression levels were similar. It was also demonstrated that enzyme fusion can be used in combination with traditional metabolic engineering strategies to increase the production of a desired product. When the fusion strategy was used in combination with *ERG9* repression, these two modifications combined resulted in an almost five-fold up-regulation reaching a final titer of 64mg sesquiterpenes/L media. This work demonstrates that engineering the spatial organization of metabolic enzymes around a branch point is of great potential for diverting flux towards a desired product.

Introduction

In multistep metabolic pathways, the efficiency of forming the final product may be influenced by loss of intermediates through diffusion, degradation or conversion by competitive enzymes (For review see Jørgensen et al., 2005). One way to prevent such losses is to increase the overall turnover rate of intermediate to product. This may be achieved by coordinating the spatial arrangement of enzymes catalyzing consecutive steps in a metabolic pathway. Close proximity of sequentially acting enzymes will speed up intermediate turnover in the pathway, not only by reducing the transit time required for the intermediate to reach the enzyme that catalyzes the next step in the reaction, but also by ensuring that a high local concentration of intermediate builds up in the vicinity of the this enzyme. The latter is important, since the global concentration of a metabolic intermediate in most cases is much lower than K_M of the enzymatic step that converts this intermediate to the next compound (Stryer, 1995). Accordingly, cells often organize enzymes acting in the same pathway in close proximity for example by sorting them to the same organelle, by docking them on cellular structures such as the ER membrane (Nielsen et al., 2008), or by allowing them to form large complexes that are often referred to as metabolons (Ovadi and Srere, 1996).

One of the goals of metabolic engineering is to optimize metabolic fluxes of pathway intermediates towards a desired product. It is therefore an attractive possibility to employ the benefits of enzyme proximity to efficiently control the flow of intermediates through a pathway. To this end, modern genetic techniques allow catalytic sites to be brought in close proximity in several ways. For example, enzymes may be engineered to assemble on supports via docking

domains (Fierobe et al., 2001) or be brought to interact via special communication domains (Hahn and Stachelhaus, 2006). However, the most simple and most studied approach is to simply fuse two genes, which are encoding different enzymatic activities. This allows expression of chimerical proteins where two enzymes are fused end-to-end into a single polypeptide. Unfortunately, *in vivo* expression of enzyme fusions has so far resulted in decreased metabolite production compared to expression of the separate counterparts (An et al., 2005, Kourtz et al., 2005, Wu et al., 2006). However, a number of observations suggest that this strategy is still of interest for improving the flux through a metabolic pathway. Firstly, several protein fusions consisting of sequential enzymes have been characterized *in vitro* and shown to possess superior kinetic properties compared to the individual enzymes. (Brodelius et al., 2002, Carlsson et al., 1996, Orita et al., 2007, Seo et al., 2000). Secondly, experiments carried out with purified fusion proteins in polyethylene glycol solutions to mimic the crowded environment inside the cell, indicate that the positive effects of enzyme fusion may be even greater *in vivo* (Yilmaz and Bulow, 2002, Prachayasittikul et al., 2006, Bülow, 1987). Lastly, *in vivo* evolution of an *Escherichia coli* strain where glucose consumption was coupled to glycerol production, selected a faster growing strain where two plasmid-borne heterologous genes catalyzing sequential steps in the glycerol pathway were spontaneously fused (Meynial Salles et al., 2007), showing that a protein fusion may benefit a process. Encouraged by these results we have explored the possibilities of using fusion of enzymes in metabolic engineering with *Saccharomyces cerevisiae* as a model.

Isoprenoids constitute an important class of natural compounds that has received considerable interest from industry due their potent pharmaceutical, flavor, color and aromatic properties

(Maury et al., 2005). As most of the valuable isoprenoids are produced in plants in low amounts, great efforts have been put into enabling production in microbial cell factories. Previously, *S. cerevisiae* was demonstrated to be a promising host for isoprenoid production, as it was engineered to produce up to 153 g/L of the important sesquiterpene amorphadiene, the precursor for the antimalarial drug artemisinin (Ro et al., 2006). These high titers were achieved using traditional metabolic tools for modulating gene expression, such as over-expression of the enzymes involved in precursor synthesis and down-regulation of enzymes that compete for precursors.

Here we investigated the possibility of using another approach, namely enzyme fusion, to optimize the production of sesquiterpenes in *S. cerevisiae*. In one test case the gene encoding patchoulol synthase (PTS) was fused to the endogenous yeast enzyme farnesyl diphosphate synthase (FPPS) and expressed in *S. cerevisiae*. PTS originates from patchouli (*Pogostemon cablin*) and catalyzes the formation of at least 14 different sesquiterpenes with patchoulol as a major product constituting 37% of the produced sesquiterpenes (Deguerry et al., 2006). The substrate for PTS is farnesyl diphosphate (FPP) produced by FPPS. FPP is also the substrate for a number of endogenous *S. cerevisiae* enzymes, including those that are involved in production of sterols, dolichols, heme A, quinones, farnesol, and farnesylated mating factors (Grabinska and Palamarczyk, 2002) (Fig. 1), suggesting that a fusion of FPPS and PTS could benefit patchoulol production by alleviating some of the competition for FPP. Using this system, we show that metabolic engineering may benefit from the construction of an artificial bifunctional enzyme. Specifically, our best FPPS-PTS fusion improves patchoulol production approximately two-fold compared to free enzymes. Moreover, the effect is additive to improvements obtained by

traditional metabolic engineering, as *ERG9* repression and enzyme fusion combined was demonstrated to result in a five-fold increase in sesquiterpene production reaching a final titer of 64 mg sesquiterpenes/L.

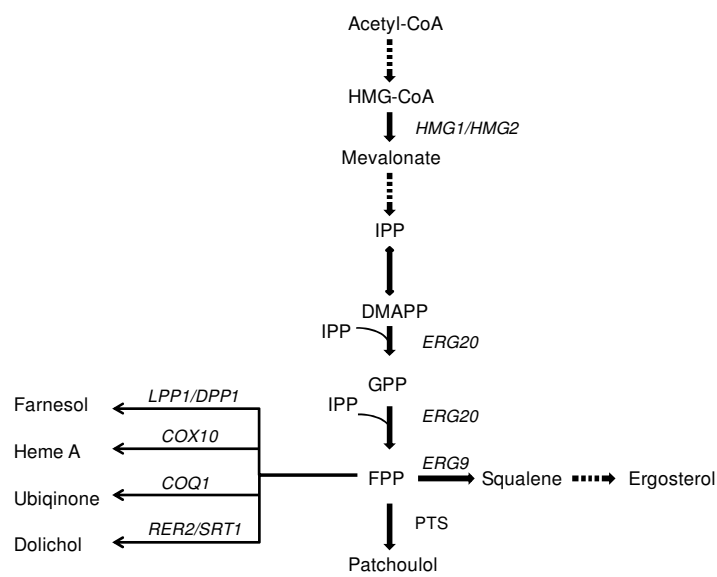


Figure 1. The mevalonate pathway in *S. cerevisiae*. Patchoulol synthase (PTS) has been introduced to enable conversion of FPP to 14 different sesquiterpenes including patchoulol. FPP is also the substrate for several other enzymes including squalene synthase (*ERG9*), heme A:farnesyltransferase (*COX10*), hexaprenyl diphosphate synthetase (*COQ1*), the cis-prenyltransferases (*RER2* and *SRT1*) and possibly the phosphatases encoded by *LPP1* and *DPP1*.

Materials and methods

Construction of *erg20*Δ strains

The genotype and the source of strains used in this study are given in Table 1. To generate a heterozygous *ERG20/erg20*Δ strain auxotrophic for uracil, CEN.PK113-13D and CEN.PK113-5D were crossed and the resulting diploid was transformed with gene targeting substrates designed to knock out *ERG20*. Gene targeting was performed according to the method described by (Erdeniz et al., 1997). HphMX which confers resistance towards Hygromycin B was used as a marker. To generate haploid *erg20*Δ strains expressing PTS and FPPS as separate enzymes or fused, the *ERG20/erg20*Δ strain was transformed with the plasmids pESC-ura (control), pLA001, pLA002 and pLA003. The resulting transformants were transferred to sporulation media containing galactose and dissection of tetrads was carried out on YPgal using a micromanipulator.

Table 1. Yeast strains used in this study.

Strain	Genotype	Source
CEN.PK113-13D	MATalpha MAL2-8 ^c SUC2 ura3-52	Peter Kötter ^a
CEN.PK113-5D	MATa MAL2-8 ^c SUC2 ura3-52	Peter Kötter
YIP-M0-04	MATa <i>erg9::P_{MET3}-ERG9</i> MAL2-8 ^c SUC2 ura3-52	Asadollahi et al., 2008

^aInstitut für Mikrobiologie, der Johan Wolfgang Goethe-Universität, Frankfurt am Main, Germany.

Plasmid construction

A list of all plasmids used in this study is given in Table 2. All DNA sub-cloning steps were performed in *E. coli* DH5 alpha using standard methods described in (Sambrook et al., 2001). Correct sequence of the cloned genes was verified by sequencing (MWG-Biotech AG).

A plasmid expressing PTS and FPPS as separate enzymes from the *GAL1* and the *GAL10* promoter, respectively, was constructed by insertion of and *ERG20* fragment into an *EcoRI-ClaI*

vector fragment of pIP029. The *ERG20* fragment was amplified from genomic DNA obtained from CEN.PK113-5D with the primers E3 and E2. To construct a plasmid expressing the fusion protein FPPS-PTS, a fragment encoding FPPS-PTS was constructed by separate amplification of the genes *ERG20* and *PatTps177* using the primers E10, E12, P5 and P4. The first round of PCR was followed by a second round of PCR in which the two fragments were fused using E10 and P4. A fragment containing PTS-FPPS was generated in a similar way except that the primers were P6, P7, E8 and E11. The resulting fusion PCR fragments were cut with *Bam*HI and *Xho*I and inserted into a *Bam*HI-*Xho*I vector fragment of pESC-ura. Flexible, rigid and stable linkers were inserted in between *ERG20* and *PatTps177* by inserting PCR fragments of *PatTps177* generated with one of the following primers as forward primers: P-FL1, P-RL1, P-RL2 or P-SL and P4 as reverse primer into a *Bsp*EI-*Xho*I vector fragment of pLA002. To insert CFP as a linker, CFP was amplified with G15 and G16 and inserted into a *Bsp*EI vector fragment of pLA002.

A plasmid expressing *ERG20* and *PatTps177* tagged N-terminally with YFP and CFP respectively was constructed in two steps. Firstly, a fragment encoding YFP-*ERG20* was amplified with Y2 and E13 and inserted into an *Eco*RI-*Spe*I fragment of pESC-ura. Secondly, CFP-*PatTps177* was amplified using G11 and P4 and inserted into a vector *Bam*HI-*Xho*I vector fragment of the plasmid expressing YFP-*ERG20*. To N-terminally tag the fusion protein FPPS-PTS with YFP and CFP, PCR fragments of YFP and CFP were generated with the primers G11 and G17 and inserted into a *Bsp*EI vector fragment of pLA002.

All oligonucleotide sequences are listed in supplementary materials, Table 1.

Table 2. Plasmids used in the fusion study.

Plasmid name	Genotype	Source /reference
pESC-ura	2 μ URA3	Stratagene
pIP029	2 μ URA3 P _{GALI} -PatTps177	Asadollahi et al., 2008
pLA001	2 μ URA3 P _{GALI} -PatTps177 P _{GALI10} -ERG20	This study
pLA002	2 μ URA3 P _{GALI} -ERG20-SF-PatTps177	This study
pLA003	2 μ URA3 P _{GALI} -PatTps177-SF-ERG20	This study
pLA004	2 μ URA3 P _{GALI} -ERG20-LF-PatTps177	This study
pLA005	2 μ URA3 P _{GALI} -ERG20-SR-PatTps177	This study
pLA006	2 μ URA3 P _{GALI} -ERG20-LR-PatTps177	This study
pLA007	2 μ URA3 P _{GALI} -ERG20-S-PatTps177	This study
pLA011	2 μ URA3 P _{GALI} -ERG20-CFP-PatTps177	This study
pLA010	2 μ URA3 P _{GALI} -CFP-PatTps177 P _{GALI10} -YFP-ERG20	This study
pLA013	2 μ URA3 P _{GALI} -CFP-ERG20-PatTps177	This study
pLA014	2 μ URA3 P _{GALI} -YFP-ERG20-PatTps177	This study

Media for genetic manipulations

All media for genetic manipulations of yeast were prepared as described by (Sherman et al., 1986) with minor modifications as the synthetic medium contained twice the amount of leucine (60mg/L) and the sporulation media contained galactose instead of glucose. A media allowing selection of hygromycin resistant colonies was prepared by adding hygromycin B to YPD to a final concentration of 300mg/L media.

2.2.4. Shake flask experiments

Shake flask experiments were carried out essentially as described by Asadollahi et al., 2008. Briefly described, cultures were grown in 500 ml baffled Erlenmeyer flasks with 100 ml Delft mineral media containing 20g/L galactose. In experiments with the strain YIP-M0-04, *ERG9* was repressed by supplementing the media with 2mM filter sterilized methionine. The pH of the mineral media was adjusted to 6.5 by adding NaOH. Shake flasks were incubated in a 30°C shaker running at 150rpm. When cell densities reached OD₆₀₀ = 0.5-1, an overlay of 10 ml dodecane was added to the flasks to collect the volatile sesquiterpenes. When cultures entered stationary phase, the fermentation broth was centrifuged to separate the organic phase from the water phase. To determine the sesquiterpene and farnesol content, samples from the organic layer was analyzed on a Finnigan Focus GC-MS with a split/splitless injector used in the splitless mode.

2.2.5. Fluorescence microscopy and quantification

Fluorescent imaging was carried out essentially as described by (Plate et al., 2008). In brief, cells were grown O/N at 23 °C in SC-Ura media with galactose, spun down and immobilized on a glass slide by mixing appropriate amounts of cells with a 1.2% solution of low melting agarose (NuSieve 3:1 from FMC). Live cell images were captured with a cooled Evolution QEi monochrome digital camera (media Cybernetics Inc.) mounted on a Nikon Eclipse E1000 camera (Nikon). Images were captured using a Plan-Fluor 100X, 1.3 numerical aperture lens. The light source was a mercury arc lamp (Osram, Germany). The fluorophores CFP and YFP were visualized with the band pass filters D436/20, 455DCLP, D480/40 and HQ500/20, Q515LP, HQ535/30, respectively (Chroma technology).

Results

FPPS-PTS and PTS-FPPS fusions are functional *in vivo*

A successful metabolic process based on fusion proteins requires that the chimerical enzyme is expressed and soluble, and that its two enzymatic activities are functional. We therefore determined the *in vivo* functionality and solubility of the two possible fusion-configurations, FPPS-PTS and PTS-FPPS.

First, we tested whether FPPS could be fused to PTS and retain activity. As the gene encoding FPPS, *ERG20*, is essential for viability of *S. cerevisiae*, the functionality of FPPS can be evaluated by investigating whether expression of the fusion enzymes can support growth of a strain harboring an *ERG20* deletion. A heterozygous diploid strain, *ERG20/erg20Δ*, was individually transformed with four different versions of a 2μ-based plasmid that contains a bidirectional *GALI/GAL10* promoter: one that serves as an “empty” control vector, one that expresses PTS and FPPS as free enzymes from the *GALI/GAL10* promoter, and two that express fusion enzymes, either FPPS-PTS or PTS-FPPS, from the *GALI* promoter. In the latter two cases, the enzymes are fused via a short flexible linker, Gly-Ser-Gly. In agreement with *ERG20* being an essential gene, the strain transformed with empty plasmid produced tetrads containing two viable spores (Fig. 2A). In contrast, the three other strains occasionally produced tetrads containing four viable spores. Hence, expression of free FPPS, FPPS-PTS or PTS-FPPS, rescue the viability of *erg20Δ* spore-clones. For these strains we also observed tetrads containing less than four viable spores. In these cases, the *erg20Δ* spore(s) likely did not inherit a plasmid. Hence, FPPS can be fused N- and C-terminally to PTS and retain activity.

Secondly, it was tested whether PTS remains active after fusion to FPPS. As patchouliol is a non-native yeast compound, we used patchouliol as a reporter metabolite to demonstrate PTS activity. Accordingly, haploid *ERG20* strains were transformed with the four plasmids described above. The four transformed strains were grown in shake flasks using a two-phase fermentation system in which dodecane is used as solvent layer to collect volatile compounds such as patchouliol. The dodecane phase was analyzed for reaction products on a GC-MS. The peak corresponding to patchouliol was identified and confirmed by comparing to the retention-time and mass-spectra of a patchouliol standard (Fig. 2B). No production of patchouliol was observed when the strain harbored the empty plasmid. In contrast, patchouliol production was observed for strains expressing both FPPS-PTS and PTS-FPPS. PTS can therefore be extended N- and C-terminally with FPPS and retain PTS activity *in vivo*.

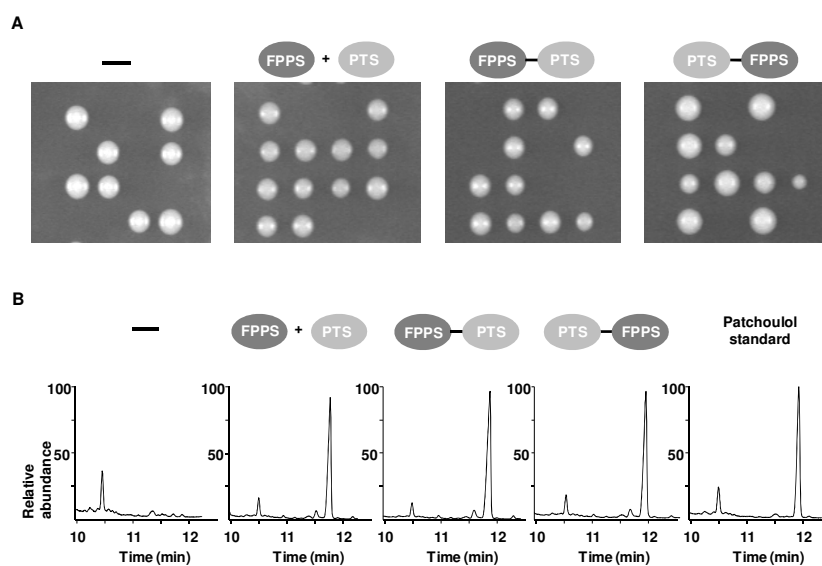





Figure 2. Functionality FPPS-PTS and PTS-FPPS when they are expressed in *S. cerevisiae*. A) Tetrad analysis of a diploid *ERG20/erg20Δ* strain harbouring an empty plasmid or plasmids expressing FPPS and PTS as free enzymes or fused. B) GC spectra for *ERG20* strains expressing the same plasmids. The spectre of a patchouliol standard is also shown.

Enzyme fusion results in an increased production of patchoulol in *S. cerevisiae*.

To investigate whether enzyme fusions may divert the flux more efficiently towards sesquiterpene production compared to the corresponding free enzymes, *ERG20* strains were individually transformed with the four plasmids described above. The resulting strains were inoculated into shake flasks in triplicates and grown until stationary phase. At this point, patchoulol production was evaluated as described above. The strain transformed with the plasmid expressing PTS-FPPS produced patchoulol at the same level as the strain expressing free PTS and FPPS (Table 3). More interestingly, the strain transformed with the plasmid expressing FPPS-PTS produced patchoulol at a two-fold higher level compared to the strain expressing free PTS and FPPS (Table 3), thus, showing that that a fusion protein may benefit a metabolic process.

Table 3: Patchoulol (PT) production from *ERG20* strains expressing FPPS and PTS as individual enzymes or fused from 2 μ -based plasmids.

Proteins expressed	Linker type (composition)	Final PT titer [mg/L]	Final PT yield [mg/L/OD ₆₀₀]
	-	5.8 ± 1.2	0.209 ± 0.004
	Short flexible (GSG)	9.5 ± 0.6	0.327 ± 0.007
	Short flexible (GSG)	5.7 ± 1.2	0.210 ± 0.056

Cells were grown in 500ml shake flasks containing 100ml mineral Delft media with 20g/L galactose. Final titers are averages of triplicates ± standard deviations.

The observed benefits of enzyme fusion is not due to a higher expression level

To test whether the increased patchoulol production from the strain expressing FPPS-PTS was due to a generally higher expression level of this protein, the expression levels of free and fused enzymes was assessed by N-terminally fusing FPPS, PTS or FPPS-PTS to either cyan- or yellow fluorescent protein (CFP or YFP, respectively) This was made by extending the relevant ORFs in the 2 μ -based plasmids described above by the *CFP* or *YFP* sequences. The plasmids were transformed into an *ERG20* strain and inspected by fluorescent microscopy. This analysis showed that all protein species produced fluorescent products that were spread evenly in the cytoplasm (Fig. 3). This indicates that all protein species are soluble, folded and sorts into the same compartment. However, we note that the fluorescent enzyme content varied substantially among individual cells. Possibly, this reflects that the population of 2 μ plasmids is unevenly divided between mother and daughter cells during division. Although, the variation among cells makes it hard to assess the expression level, we judged from analysis of several images that the expression level of free and fused enzymes was similar. Accordingly, the increased patchoulol from the strain expressing the fusion protein is not due to a higher enzyme concentration in these cells.

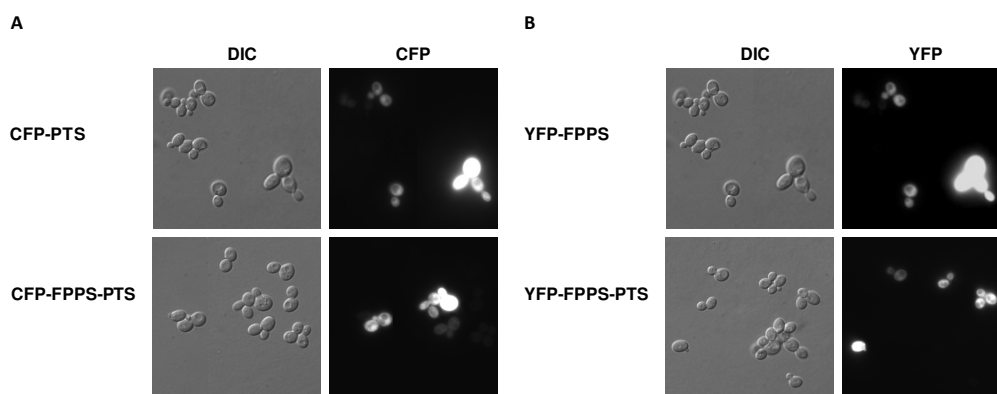



Figure 3. Cellular localization and expression levels of A) PTS and B) FPPS when they are expressed as free enzymes or fused.

Effect of linker length and composition

Since specific linkers can presumably favor beneficial conformations of fusion proteins, (Carlsson et al., 1996, Lu and Feng, 2008, Robinson and Sauer, 1998, van Leeuwen et al., 1997) we proceeded to investigate whether the nature of the linker combining FPPS and PTS can influence patchoulol production. To this end, we focused on the configuration FPPS-PTS, and examined whether patchoulol production could be further improved by altering the linker composition and length. In addition to the short flexible linker, which we already tested, we also determined patchoulol production of *ERG20* strains expressing FPPS-PTS fused by linker types representing: long flexible, short rigid, long rigid, stable, and very long linkers (Table 4). The stable linker is a short linker that should be resistant towards degradation by proteases and the long linker is in fact the sequence of CFP. The sequences of the remaining linkers can be found in Table 4. Strains expressing the five new FPPS-PTS variants were analyzed in triplicates for their ability to produce patchoulol. The results of this experiment show that although the FPPS-

PTS fusion has a slight preference for short linkers, no matter whether these linkers are flexible or rigid in nature, the linker lengths and compositions do not dramatically influence patchoulol production as they all roughly lead to the same patchoulol production. The exception is when CFP acts as the linker to connect FPPS and PTS. In this case the amount of patchoulol is lower than that obtained by the free enzymes (Table 4).

Table 4: Effect of inserting linkers that varied in length and composition in between FPPS and PTS.

Proteins expressed ^a	Linker type (composition)	Effect of linker ^b PT titer _{linker} / PT titer _{short flexible linker}
	Short flexible (GSG)	1,00
	Long flexible (GSGGGGS)	0,83 ± 0,02*
	Short rigid (GSGEAAAK)	0,98 ± 0,19
	Long rigid (GSGEAAAKEAAAK)	0,85 ± 0,03*
	Stable (GSGMGSSSN)	0,82 ± 0,22
	Very long (Sequence of CFP)	0,33 ± 0,03*

Cells were grown in 500ml shake flasks containing 100ml Delft media with 20g/L galactose. Final titers are averages of triplicates ± standard deviations.

^aProteins were in all cases expressed from 2μ based plasmids from P_{GAL1}. In all cases FPPS was also expressed endogenously.

^bThe effect of a linker is given as patchoulol production using this linker relative to production achieved when the short flexible linker was used.




*Patchoulol production using this linker is significantly ($p < 0.05$) different from when the short flexible linker was used.

Enzyme fusion can be used in combination with traditional metabolic engineering strategies to increase the patchoulol production

Previously, we have used different metabolic engineering strategies to increase production of patchoulol (Asadollahi et al., 2008, data not shown). Amongst these, the highest patchoulol production was obtained using an *ERG9* repressed strain, where the flux towards sterol synthesis is reduced (Asadollahi et al., 2008, data not shown). To investigate whether production benefits obtained from a fusion enzyme can be additive to those obtained from traditional metabolic engineering, we transformed a plasmid expressing FPPS fused to PTS via a short flexible linker into a strain in which *ERG9* expression is repressed (Asadollahi et al., 2008). As for comparisons, this strain was also transformed with a plasmid expressing FPPS and PTS as separate enzymes or a plasmid that only expresses PTS.

When the fused enzymes were expressed in the *ERG9* repressed strain, an almost three-fold and five-fold increase in patchoulol production was observed compared to the strain expressing FPPS and PTS as free enzymes and the strain that only expresses PTS, respectively (Table 5). The final titer of patchoulol was 24 mg/L when *ERG9* repression and fusion was combined (Table 5) and as patchoulol only constitutes 37% of the total amount of sesquiterpenes produced, this means that the final titer of sesquiterpenes reached 64mg/L. Surprisingly, the production of another FPP-derived metabolite, farnesol, was also increased two-fold in the strain expressing FPPS-PTS.

Table 5: Patchoulol (PT) and farnesol (FOH) production in *ERG9* repressed strains expressing FPPS and PTS as individual enzymes or fused.

Proteins expressed ^a	Final PTS titer [mg/L]	Final PT yield [mg/L/OD ₆₀₀]	Final FOH titer [mg/L]	Final FOH yield [mg/L/OD ₆₀₀]
	4.9 ± 0.7	0.16 ± 0.03	9.7 ± 1.1	0.31 ± 0.05
	8.5 ± 0.1	0.31 ± 0.02	12.2 ± 5.0	0.45 ± 0.21
	23.7 ± 0.9	0.77 ± 0.05	23.8 ± 7.9	0.78 ± 0.27

Final titers and yields are averages of triplicates ± standard deviations, except that only duplicates were available for the strain expressing the fusion protein FPPS-PTS.

^aIn all cases FPPS was also expressed endogenously.

Discussion:

In the present study, we demonstrate that enzyme fusion can be successfully used to improve the flux through a pathway in the cell factory *S. cerevisiae*. The process likely benefits from a proximity effect of the two active sites to reduce loss of intermediates to other competing pathways. Previous attempts to use this concept *in vivo* have been unsuccessful despite the fact that fusion enzymes have proven superior to free enzymes in several *in vitro* studies (Brodelius et al., 2002, Carlsson et al., 1996, Orita et al., 2007, Seo et al., 2000). At least two parameters should be considered to obtain successful exploitation of fusion proteins *in vivo*. Firstly, the cellular concentration of correctly folded chimerical protein should be similar to the concentrations of the corresponding free enzymes. Secondly, the fusion should not negatively affect the activity of the individual catalytic sites. Unfortunately, with our current knowledge, the impact of a fusion on these two parameters is not easy to predict. For example, despite that fusion of FPPS and PTS can successfully improve patchoulol production, expression of a similar fusion of FPPS and valencene synthase, which at the amino acid residue level is approximately 40% identical to PTS, does not produce additional valencene in *S. cerevisiae* compared to a strain expressing the corresponding two free enzymes (Supplementary materials Table 2). Similarly, when Wu and co-workers expressed a fusion of an FPPS of avian origin and the PTS used in this study, they obtained a 5-6 fold lower production in tobacco plants compared to production obtained with the free enzymes (Wu et al., 2006). It is therefore necessary to optimize fusion of the two enzymes for functionality to benefit from this strategy. For example, the configuration of the fusion may play an important role as demonstrated in our study where only the fusion of FPPS at the amino terminal part of PTS (FPPS-PTS) improves patchoulol production (Table 3). Similarly, since it is well known that correct folding and activity of a

chimerical protein depend on the nature of the linker, it is advisable to create a collection of fusion proteins assembled with different linker-types. In our case, the C-terminus of FPPS and the N-terminus of PTS appear quite flexible as protein acceptors, since strains expressing FPPS-PTS connected by different linkers produce patchoulol in similar amounts. The only exception is when the two enzymes are linked by CFP. In this case, patchoulol production is reduced to the level obtained with free enzymes and we speculate that the distance between the active sites in the FPPS-CFP-PTS fusion enzyme is too large to benefit from the proximity effect.

Two facts suggest that the benefit of *ERG9* repression is not fully exploited. First, the main effect of the mutation is a substantial increase in farnesol production indicating that the increased FPP pool is used for farnesol rather than for sesquiterpene production (Asadollahi et al., 2008). Secondly, it is known that farnesol mediates product inhibition of an early step in the mevalonate pathway by signaling degradation of Hmg2 (Shearer and Hampton, 2005) (see Fig. 1). To overcome these problems, we have previously tried to decrease the flux towards farnesol by deleting the two genes *LPP1* and *DPP1*, which encode the two phosphatases that are believed to be responsible for phosphorylation of FPP to FOH (Faulkner et al., 1999). Furthermore, we have tried to bypass product inhibition of the HMG-CoA reductase step by over-expression of a truncated version of *HMG1* that contained the catalytic domain only (Asadollahi, unpublished results). Surprisingly, neither of these modifications increased patchoulol production in the *ERG9* repressed strain. To take advantage of the higher FPP pool in *ERG9* repressed strains, we therefore decided to investigate the possibility that an FPPS-PTS fusion may redirect the flux towards sesquiterpene production; and in the present study we show that this fusion increases patchoulol almost three-fold in an *ERG9* repressed strain. Surprisingly, farnesol levels were also

increased two-fold in this strain. This is somewhat unexpected, as the reaction catalyzed by the fusion enzyme is thought to be favored at the expense of other pathway branches starting from the FPP node. The fact that farnesol levels are increased is an indication that flux towards FPP is generally increased in the fusion strain rendering the possibility that enzyme fusion somehow bypasses some of the regulatory circuits of the pathway. One explanation for this could be that the fusion of FPPS to PTS blocks FPPS from participating in interactions required for direction of FPP into other pathway branches. Although farnesol levels are increased in the strain expressing the fusion protein we do however note that the partitioning towards patchoulol is increased as the farnesol to patchoulol ratio is decreased approximately 40% when the enzymes are fused compared to when they are expressed as free enzymes. This indicates that the benefits of enzyme fusion are at least partly due to PTS converting a greater share of the FPP pool.

Based on the above, we conclude that an enzyme-fusion strategy can successfully be used in combination with traditional metabolic engineering strategies and contributes to redirecting the metabolic flux into a desired pathway. However, we note that farnesol levels are still high in *ERG9* repressed strains expressing the FPPS-PTS fusion. Accordingly, some FPP may still be converted to farnesol by competing enzymes; or alternatively, the FPPS-PTS itself may in fact produce farnesol. Moreover, free FPPS, encoded by the endogenous *ERG20* locus may also contribute to farnesol production. It is therefore likely that further metabolic engineering of the *ERG9* repressed strains expressing the FPPS-PTS fusion may additionally improve patchoulol production, for example by deleting *ERG20* or by relieving product inhibition the mevalonate pathway.

References

- An, J., Kim, Y., Lim, W., Hong, S., An, C., Shin, E., Cho, K., Choi, B., Kang, J., Lee, S., et al. (2005). Evaluation of a novel bifunctional xylanase-cellulase constructed by gene fusion. *Enzyme and Microbial Technology*, 36(7):989–995.
- Asadollahi, M., Maury, J., Moller, K., Nielsen, K., Schalk, M., Clark, A., and Nielsen, J. (2008). Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: Effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnology and bioengineering*, 99(3):666.
- Bülow, L. (1987). Characterization of an artificial bifunctional enzyme, beta-galactosidase/galactokinase, prepared by gene fusion. *European Journal of Biochemistry*, 163(3):443–448.
- Brodelius, M., Lundgren, A., Mercke, P., and Brodelius, P. (2002). Fusion of farnesyl diphosphate synthase and epi-aristolochene synthase, a sesquiterpene cyclase involved in capsidiol biosynthesis in *Nicotiana tabacum*. *European Journal of Biochemistry*, 269(14):3570–3577.
- Carlsson, H., Ljung, S., and Bülow, L. (1996). Physical and kinetic effects on introduction of various linker regions in beta-galactosidase/galactose dehydrogenase fusion enzymes. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular Enzymology*, 1293(1):154–160.
- Deguerry, F., Pastore, L., Wu, S., Clark, A., Chappell, J., and Schalk, M. (2006). The diverse sesquiterpene profile of patchouli, *Pogostemon cablin*, is correlated with a limited number of sesquiterpene synthases. *Archives of Biochemistry and Biophysics*, 454(2):123–136.

- Erdeniz, N., Mortensen, U., and Rothstein, R. (1997). Cloning-Free PCR-Based Allele Replacement Methods. *Genome Research*, 7(12):1174–1183.
- Faulkner, A., Chen, X., Rush, J., Horazdovsky, B., Waechter, C., Carman, G., and Sternweis, P. (1999). The LPP1 and DPP1 Gene Products Account for Most of the Isoprenoid Phosphate Phosphatase Activities in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 274(21):14831–14837.
- Fierobe, H., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J., and Bayer, E. (2001). Design and production of active cellulosome chimeras selective incorporation of dockerin-containing enzymes into defined functional complexes. *Journal of Biological Chemistry*, 276(24):21257–21261.
- Grabinska, K. and Palamarczyk, G. (2002). Dolichol biosynthesis in the yeast *Saccharomyces cerevisiae*: an insight into the regulatory role of farnesyl diphosphate synthase. *FEMS Yeast Research*, 2(3):259–265.
- Hahn, M. and Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proceedings of the National Academy of Sciences*, 103(2):275–280.
- Jørgensen, K., Rasmussen, A., Morant, M., Nielsen, A., Bjarnholt, N., Zagrobelny, M., Bak, S., and Møller, B. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current Opinion in Plant Biology*, 8(3):280–291.

Kourtz, L., Dillon, K., Daughtry, S., Madison, L., Peoples, O., and Snell, K. (2005). A novel thiolase-reductase gene fusion promotes the production of polyhydroxybutyrate in *Arabidopsis*.

Plant Biotechnology Journal, 3(4):435–447.

Lu, P. and Feng, M. (2008). Bifunctional enhancement of a beta-glucanase-xylanase fusion enzyme by optimization of peptide linkers. *Applied Microbiology and Biotechnology*, 79(4):579–587.

Maury, J., Asadollahi, M., Moller, K., Clark, A., and Nielsen, J. (2005). Microbial isoprenoid production: An example of green chemistry through metabolic engineering. *Biotechnology for the Future*, 100:19–51.

Meynial Salles, I., Forchhammer, N., Croux, C., Girbal, L., and Soucaille, P. (2007). Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. *Metabolic Engineering*, 9(2):152–159.

Nielsen, K., Tattersall, D., Jones, P., and Moller, B. (2008). Metabolon formation in dhurrin biosynthesis. *Phytochemistry*, 69(1):88–98.

Orita, I., Sakamoto, N., Kato, N., Yurimoto, H., and Sakai, Y. (2007). Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. *Applied Microbiology and Biotechnology*, 76(2):439–445.

Ovadi, J. and Srere, P. (1996). Metabolic Consequences of Enzyme Interactions. *Cell Biochemistry and Function*, 14(4):249–258.

- Plate, I., Hallwyl, S., Shi, I., Krejci, L., Muller, C., Albertsen, L., Sung, P., and Mortensen, U. (2008). Interaction with RPA Is Necessary for Rad52 Repair Center Formation and for Its Mediator Activity. *Journal of Biological Chemistry*, 283(43):29077.
- Prachayasittikul, V., Ljung, S., Isarankura-Na-Ayudhya, C., and Bülow, L. (2006). NAD (H) recycling activity of an engineered bifunctional enzyme galactose dehydrogenase/lactate dehydrogenase. *International Journal of Biological Sciences*, 2(1):10.
- Ro, D., Paradise, E., Ouellet, M., Fisher, K., Newman, K., Ndungu, J., Ho, K., Eachus, R., Ham, T., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943.
- Robinson, C. and Sauer, R. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences*, 95(11):5929–5934.
- Sambrook, J., Russell, D., and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. Cold Spring Harbor Laboratory Press.
- Seo, H., Koo, Y., Lim, J., Song, J., Kim, C., Kim, J., Lee, J., and Choi, Y. (2000). Characterization of a Bifunctional Enzyme Fusion of Trehalose-6-Phosphate Synthetase and Trehalose-6-Phosphate Phosphatase of *Escherichia coli*. *Applied and Environmental Microbiology*, 66(6):2484–2490.
- Shearer, A. and Hampton, R. (2005). Lipid-mediated, reversible misfolding of a sterol-sensing domain protein. *The EMBO Journal*, 24:149–159.

Sherman, F., Fink, G., Hicks, J., and Laboratory, C. S. H. (1986). *Laboratory course manual for methods in yeast genetics*. Cold Spring Harbor Laboratory [SI].

Stryer, L. (1995). *Biochemistry*. *New York*, pages 192–196.

van Leeuwen, H., Strating, M., Rensen, M., de Laat, W., and van der Vliet, P. (1997). Linker length and composition influence the flexibility of Oct-1 DNA binding. *The EMBO Journal*, 16:2043–2053.

Wu, S., Schalk, M., Clark, A., Miles, R., Coates, R., and Chappell, J. (2006). Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature Biotechnology*, 24:1441–1447.




Yilmaz, J. and Bulow, L. (2002). Enhanced Stress Tolerance in *Escherichia coli* and *Nicotiana tabacum* Expressing a Betaine Aldehyde Dehydrogenase/Choline Dehydrogenase Fusion Protein. *Biotechnology Progress*, 18(6):1176–1182.

Supplementary Table 1. Oligonucleotides used in this study.

Primer	Sequence ^a	Template
E2	gtacgtagtgatcgatCTATTTGCTTCTCTTGTAAC	<i>ERG20</i>
E3	gtatcgtagtgaattcATGGCTTCAGAAAAAGAAATTAGG	<i>ERG20</i>
E8	ATGGCTTCAGAAAAAGAAATTAGGAGAG	<i>ERG20</i>
E10	ccctatgagaggatccATGGCTTCAGAAAAAGAAATTAGGAGAG	<i>ERG20</i>
E11	gcgaaaaggctcgagTTATTTACTTCTCTTGTAACCTTGTTCAAAAACGC	<i>ERG20</i>
E12	TTTACTTCTCTTGTAACCTTGTTCAAAAACGC	<i>ERG20</i>
E13	gcgataccactagtctaTTTACTTCTCTTGTAACCTTGTTCAAAAACGC	<i>ERG20</i>
P4	catgatgcgactcgagTTAATATGGAACAGGGTGAA	<i>PatTps177</i>
P5	gaacaaggtttacaagagaagtaaagggtccggaATGGAGTTGTATGCCCAAAGTG	<i>PatTps177</i>
P6	ccctatgagaggatccATGGAGTTGTATGCCCAAAGTG	<i>PatTps177</i>
P7	ctctcctaattcttttctgaagccattccggagccATATGGAACAGGGTGAAGGTAC	<i>PatTps177</i>
P-FL1	gatgatagattccggagcggtggtccATGGAGTTGTATGCCCAAAGTG	<i>PatTps177</i>
P-RL1	gatgatagattccggagaagctcggcaaaaATGGAGTTGTATGCCCAAAGTG	<i>PatTps177</i>
P-RL2	gatgatagattccggagaagctcggcaaaagaagcagcggctaaaATGGAGTTGTATGCCCAAAGTG	<i>PatTps177</i>
P-SL	gatgatagatccggaatgggagctctcgaatATGGAGTTGTATGCCCAAAGTG	<i>PatTps177</i>
G2	cgatgagctgacgaattcATGAGTAAAGGAGAAGAAGCTTTTCACTG	YFP and CFP
G11	ctgtagtggaggatccATGAGTAAAGGAGAAGAAGCTTTTCACTG	YFP and CFP
G15	gagcgactgtccggaATGAGTAAAGGAGAAGAAGCTTTTCACTG	YFP and CFP
G16	ctagtgcgttccggaTTTGTATAGTTCATCCATGCCATGTG	YFP and CFP
G17	ctagtgcaccgatccTTTGTATAGTTCATCCATGCCATG	YFP and CFP

^aThe parts of the nucleotide sequences shown in capitals are those that specifically anneal to the template DNA. The parts of the oligonucleotides that contain restriction sites and adaptamers are shown in lower-case.

Supplementary Table 2. Fusion of FPPS and valencene synthase does not increase the sesquiterpene production.

Proteins expressed from 2 μ -based plasmids ^a	Linker type (composition)	Final valencene titer [mg/L]
	-	0.29 \pm 0.12
	Short flexible (GSG)	0.25 \pm 0.05
	Short flexible (GSG)	0.04 \pm 0.02

Cells were grown in 500ml shake flasks containing 100ml Delft media with 20g/L galactose. Final titers are averages of triplicates \pm standard deviations.

^aIn all cases FPPS was also expressed endogenously.

Chapter 3: Enzyme fusion for optimization of vanillin glucoside production in *S. cerevisiae*.

3.1. Introduction

Vanillin is one of the most important aromatic compounds produced by industry with widespread applications within industries producing food, beverages, perfumes, cosmetics and pharmaceuticals (Barghini et al., 2007, Priefert et al., 2001). The worldwide production exceeds 12,000 tons per year of which most are chemically synthesized (Yoon et al., 2007). Only approximately 0.2% (20-50 tons) of the produced vanillin derives from the botanical source (Krings and Berger, 1998). An increased focus on the environmental hazards connected with chemical synthesis combined with consumer-driven trends towards producing foods with natural flavours rather than synthetic flavours, has enlarged the demand for natural vanillin. The production in the natural host is however compromised by limited availability of suitable cultivation areas, the plants susceptibility to diseases and the labour intense production method (Priefert et al., 2001). One attractive alternative to production in the natural host is to develop microbial production platforms for synthesis of natural vanillin. Most of the systems developed so far have focused on biotransformation of eugenol or ferulic acid (Priefert et al., 2001). These methods suffer from the relatively high price of natural ferulic acid, which is due to its limited accessibility from lignin by biological means (Bhathena et al., 2007, Walton et al., 2000). The alternative use of synthetic derived substrates removes the label "natural" from the vanillin produced, hereby lowering the price considerably. In an attempt to produce vanillin from the much cheaper substrate glucose, *de novo* vanillin pathways have been engineered into *E. coli* (Li and Frost, 1998) and *S. cerevisiae* (Hansen et al., 2009). The *de novo* vanillin pathway engineered into yeast is the focus of this study. Here, vanillin is produced in three catalytic steps from 3-dehydroshikimic acid (3DSA) which is produced by the endogenous shikimate pathway of yeast (Fig. 3-1). The enzymes catalyzing these three steps were 3-dehydroshikimate dehydrase (3DSD) of *Podospira pauciseta*, aromatic carboxylic acid reductase (ACAR) of *Nocardia ssp* and human O-methyl transferase (OMT). Moreover, ACAR requires activation by a phosphopanteyl transferase (PPTase) in *S. cerevisiae* (Hansen et al., 2009). This was achieved by expressing the PPTase encoded by the *E. coli* gene *entD*. As vanillin is toxic for yeast, even at low concentrations (Larsson et al., 2000), a fourth step was inserted to detoxify vanillin by

glycosylating it. This was accomplished by expressing a UDP- glycosyl transferase (UGT) of *Arabidopsis thailana*.

Generally, production of vanillin in microbial cell factories is hampered by the cells tendencies to rapidly convert vanillin to less toxic compounds such as vanillic acid or vanillyl alcohol (Barghini et al., 2007). Metabolic engineering aims at solving such problems using a rational approach. One traditional approach involves the identification and knock-out of enzymes catalyzing unwanted side-reactions (Chotani et al., 2000). In the case of the vanillin pathway, two unwanted reactions were identified. One is the conversion of vanillin to vanillyl alcohol – a process that presumably can be carried out by the natural alcohol dehydrogenases of yeast (Fig. 3-1). To inhibit this process, several different alcohol dehydrogenase mutants were tested for their ability to reduce the conversion of vanillin to vanillyl alcohol (Hansen et al., 2009). The screen showed that deletion of the alcohol dehydrogenase encoded by *ADH6* could reduce formation of vanillyl alcohol without impairing growth (Hansen et al., 2009). Another unwanted reaction is catalyzed by the β -glucosidase encoded by *BGLI*. This enzyme rapidly reverses the reaction catalyzed by UGT and converts vanillin glucoside (VG) back to vanillin. To avoid these unwanted reactions, a strain carrying deletions of *ADH6* and *BGLI* was used in this study.

Another approach for pathway optimization that has been gaining increasing interest recently is the engineering of the spatial organization. In chapter 2, it was shown that flux could be diverted towards patchoulol production by positioning the catalytic sites of two consecutive enzymes in close proximity by enzyme fusion. Here, it is tested whether this approach also can be applied to optimize VG production in *S. cerevisiae*. To this end, the enzymes catalyzing last two steps in the pathway, namely OMT and UGT were positioned in close proximity by enzyme fusion. There are several reasons why the vanillin pathway is particularly likely to benefit from proximity of these two enzymes. First, vanillin is known to be toxic and has been reported to inhibit growth of yeast in relatively low concentrations (Larsson et al., 2000). By keeping OMT and UGT in close proximity, reaction rates can be increased and this may prevent vanillin accumulation and thereby minimize growth inhibition. Secondly, UGT is believed to work on a number of substrates in yeast, including the early intermediates of the vanillin pathway. Accordingly, there is competition for enzyme. Moreover, if early intermediates such as PAC and

3.2. Materials and Methods

3.2.1. Strains and media

All media for genetic manipulations of yeast were prepared as described by Sherman (Sherman et al., 1986) with minor modifications as the synthetic medium contained twice the amount of leucine (60 mg/L).

For shake flasks experiments Delft mineral media composed of 7.5 g/L (NH₄)₂SO₄, 14.4 g/L KH₂PO₄, 0.5 g/L MgSO₄ *7H₂O, 2 mL/L trace metal solution, 1 mL/L vitamin solution and 50 µL/L synperonic antifoam was used. In all cases Delft media was supplied with galactose (20g/L) and methionine (21g/L). The pH of mineral medium was adjusted to 6.50 by adding 2 M NaOH. Vitamin and amino acids were sterile filtered and added after autoclavation.

To construct a strain producing PAL, VAN265 was transformed with pJH674 which had been linearized in the TPI1 promoter with *SphI*. Transformants were plated on YPD based media containing aureobasidin A (Takara Bio Inc.) and hygromycin B (Sigma Aldrich) to a final concentration of 0.5 mg/L and 300 mg/L media, respectively.

Table 3-1. Strains used in this study

Strain name	Genotype	Source
VAN265	Mata, <i>adh6::LEU2</i> , <i>bgl1::KanMX4</i> , <i>his3Δ1</i> , <i>leu2Δ0</i> , <i>met15Δ0</i> , <i>ura3Δ0</i> , P _{TPI1} ::3DSD [AurC]	Hansen et al., 2009
VAN265JH	Mata, <i>adh6::LEU2</i> , <i>bgl1::KanMX4</i> , <i>his3Δ1</i> , <i>leu2Δ0</i> , <i>met15Δ</i> , <i>ura3Δ0</i> , P _{TPI1} ::3DSD- [AurC]::ACAR [HphMX]	This study

3.2.2. Plasmid construction

A list of all plasmids used in this study is given in Table 3-2. All DNA sub-cloning steps were performed in *E. coli* DH5 alpha using standard methods described in (Sambrook et al., 2001).

To construct a plasmid expressing OMT and UGT as free enzymes from P_{GAL1} and P_{GAL10} respectively, OMT was amplified by PCR with the primers O1 and O3. The resulting PCR fragment was cut with *Bam*HI and *Xho*I and inserted into a *Bam*HI-*Xho*I vector fragment of pESC-his. Subsequently, a PCR fragment of UGT was amplified with the primers U1 and U3, and inserted into a *Cla*I-*Pac*I of the OMT containing vector. To generate a DNA fragment encoding OMT-UGT, the two genes were separately amplified using the primers O1, O4, U8 and U7 and subsequently they were fused in a second round of PCR with the primers O1 and U7. The resulting OMT-UGT fragment was digested with *Bam*HI and *Xho*I and inserted into a *Bam*HI-*Xho*I vector fragment of pESC-his. The plasmid expressing UGT-OMT was generated in an equivalent way except the primers were U5, U6, O5 and O6.

The sequence of all oligonucleotides used for plasmid construction is given in Table 3-3.

Table 3-2. Plasmids used in this study

Strain name	Genotype ^a	Source
PJH674	P _{TPI1} ACAR, hphMX	Hansen et al., 2009
pJH589	<i>CEN</i> , <i>URA3</i> , <i>EntD</i>	Hansen et al., 2009
pJH543	P _{TPI1} hsOMT, NatMX	Hansen et al., 2009
pJH665	<i>CEN</i> , <i>URA3</i> , P _{TPI1} UGT72E2	Hansen et al., 2009
pESC-his	<i>HIS3</i>	Stratagene
pESC-his-OMT+UGT	2μ, <i>HIS3</i> , P _{GAL1} hsOMT, P _{GAL10} UGT72E2	This study
pESC-his-OMT-UGT	2μ, <i>HIS3</i> , P _{GAL1} hsOMT-UGT72E2	This study
pESC-his-UGT-OMT	2μ, <i>HIS3</i> , P _{GAL1} UGT72E2-hsOMT	This study

^aThe genes encoding ACAR and HsOMT had been codon optimized for yeast.

3.3.3. Shake flask experiments

Shake flasks experiments were carried out in 500 ml baffled Erlenmeyer flasks with 100 ml Delft mineral media containing 20g/L galactose and 21g/L methionine. Shake flasks were inoculated with precultures to an initial OD₆₀₀ of 0.05. Samples for OD₆₀₀ measurement and HPLC analysis were taken at time points 0, 13.5, 19.5, 28, 35, 62.5 hours after inoculation. For HPLC analysis an Agilent ChemStation LC30 was used and concentrations of metabolic products were determined by comparing them to standard curves.

Table 3-3. Oligonucleotides used in this study.

Primer name	Sequence ^a	Template
O1	agcatgacttaggatccATGGGTGACACTAAGGAGC	<i>hsOMT</i>
O3	tctatcatggctcgagTTATGGACCAGCTTCAGAACC	<i>hsOMT</i>
O4	TGGACCAGCTTCAGAACCTG	<i>hsOMT</i>
O5	ggacttgctcacgtggtgccgggtccggaATGGGTGACACTAAGGAGC	<i>hsOMT</i>
O6	ctaagataggccgcggTTATGGACCAGCTTCAGAACC	<i>hsOMT</i>
U1	ttcaacatccatgatcATGCATATCACAAAACCCACACG	<i>UGT72E2</i>
U3	tgatagatccttaattaaCTAGGCACCACGTGACAAG	<i>UGT72E2</i>
U5	ccagtagtgggatccATGCATATCACAAAACCCACACG	<i>UGT72E2</i>
U6	GGCACCACGTGACAAGTCC	<i>UGT72E2</i>
U7	agtactaggcccgcggCTAGGCACCACGTGACAAG	<i>UGT72E2</i>
U8	caggttctgaagctggtccagggtccggaATGCATATCACAAAACCCACACG	<i>UGT72E2</i>

^aThe parts of the nucleotide sequences shown in capitals are those that specifically anneal to the template DNA. The parts of the oligonucleotides that contain restriction sites and adaptamers are shown in lower-case.

3.3. Results

3.3.1. Enzyme fusion results in decreased intermediate accumulation and increased growth rates

To investigate whether the *de novo* vanillin pathway could be optimized by fusion of the enzymes catalyzing the last two steps in the pathway, namely OMT and UGT, a strain producing the substrate for OMT, protocatechuic aldehyde (PAL), was constructed. This was done by integrating ACAR into the genome of a strain that already expressed 3DSD from the genome. The strain also harboured deletions of *BGL1* and *ADH6* to avoid unwanted side-reactions. To activate ACAR the strain was transformed with a CEN-based plasmid expressing the PPTase encoded by *EntD*. In a preliminary experiment both ACAR and 3 DSD were shown to be active in this strain, as PAL accumulated in the media of an O/N culture (data not shown). It was however noted that large amounts of the substrate for ACAR, protocatechuic acid (PAC), also accumulated in the media indicating that ACAR possibly is not fully activated by the PPTase from *E. coli*.

To evaluate the effect of enzyme fusion, the PAL producing strain was transformed with 2 μ -based plasmids expressing OMT and UGT as free enzymes or fused in the two possible configurations, OMT-UGT or UGT-OMT. The strains were grown in shake flasks until stationary phase while growth and the formation of metabolic products was followed by measuring OD₆₀₀ and taking samples for HPLC analysis at time points 0, 13.5, 19.5, 28, 35, 62.5 hours after inoculation.

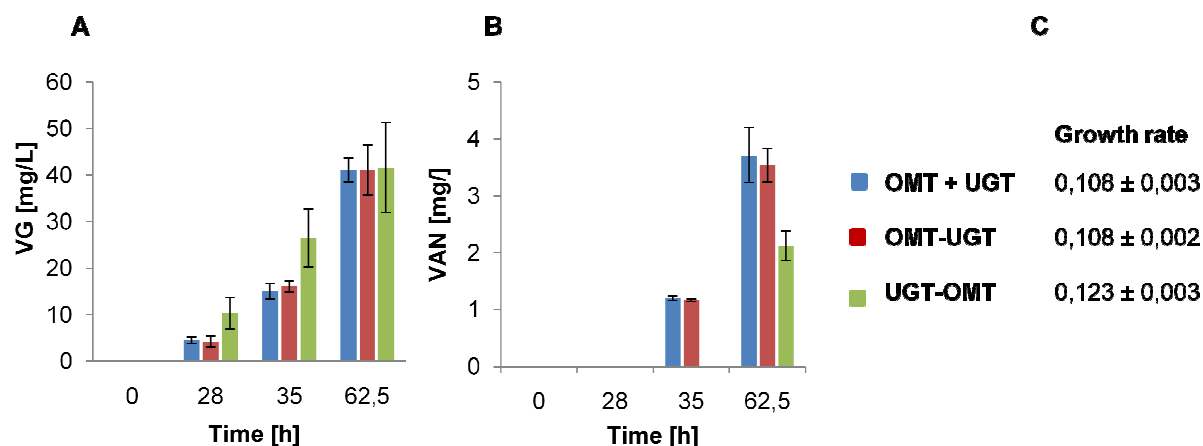


Figure 3-1. Summary of shake flask results for strains expressing OMT and UGT as free enzymes (OMT+UGT) or fused in the two possible configurations (OMT-UGT and UGT-OMT). Concentrations of A) vanillin glucoside (VG) and B) vanillin (VAN) in the media at time points 0, 28, 35 and 62.5 hours after inoculation. C) Growth rates of the same three strains given as averages of triples ± standard deviations.

The fact that VG is produced by all three strains, demonstrates that both OMT and UGT are functional when they are fused in the two possible configurations. Although the final concentration of VG was similar for all three strains (Fig. 3-1A), clear differences among the tested strains were observed. Generally, the strain expressing OMT-UGT behaved similarly to the strain expressing the free enzymes, whereas the strain expressing UGT-OMT differed from the two others. In the strain expressing UGT-OMT, VG seems to build up faster compared to the other strains (Fig. 3-1A). Furthermore, less vanillin accumulation was observed for the strain expressing UGT-OMT, as the vanillin concentration was approximately 40% decreased compared the two other strains in the final sample and could not be observed after 35 hours (Fig 3-1B). In contrast, vanillin accumulation was observed after 35 hours for both the strain expressing the free enzymes and the strain expressing OMT-UGT, although these strains grew slower and at this point only had reached OD_{600} 2.5-3, whereas the strain expressing UGT-OMT had reached OD_{600} 5.5-6 after 35 hours (Fig. 3-1B and data not shown). The strain expressing UGT-OMT also differed when it came to growth rates, as it had a 14% higher growth rate than the other two strains (Fig. 3-1C).

3.4. Discussion

Here, it is shown that accumulation of the toxic intermediate vanillin can be significantly decreased by fusion of the two enzymes OMT and UGT, when they are fused in the configuration UGT-OMT. This is likely due to fusion induced proximity effects, as proximity of catalytic sites can reduce intermediate transit times and thereby minimised accumulation of intermediates. Moreover, the strain expressing UGT-OMT seems to benefit from the reduced accumulation of vanillin as the strain displays a 14% higher growth rate. Presumably, the higher growth rate also explains why VG production occurs faster in this strain. The final titer of VG is however not increased in the strain expressing UGT-OMT. Possibly, greater benefits of fusion could be observed under less optimized conditions. In this study, a strain that has already been partly optimized for VG production by deletion of *ADH6* which catalyzes a competing reaction was used. Moreover, although vanillin accumulation was observed in this study, concentrations are very low compared the minimal inhibitory concentrations reported by others (Klinke et al., 2004) rendering the possibility that greater benefits of enzyme fusion could be observed under conditions where accumulation of vanillin is more pronounced.

When vanillin is present in the fermentation media it has been demonstrated to inhibit growth in concentrations as low as 1mM (Larsson et al., 2000). In this study vanillin concentrations in the media were at all times less than 4 mg/L (Fig 3-1B) which is equal to less than 0.03mM. Accordingly, the concentrations observed in this study are much lower, than the concentrations that have previously been demonstrated to inhibit growth. We do however measure the vanillin concentration in the media and can therefore not rule out that concentrations inside cells were larger.

In all three strains flux towards vanillin seems to be limited by ACARs low turnover of PAC to PAL, as a substantial accumulation of PAC was observed in all strains (data not shown). Possibly, this is because ACAR is not optimally activated by the PPTase of *E. coli* which was used in this study. In a recent study, the activity of ACAR was increased 20-fold in *E. coli* when it was co-expressed with the PPTase that derives from a *Nocardia* species (Venkitasubramanian

et al., 2007). This suggests that the PPTase of *Nocardia* is more efficient in activating ACAR than the *E.coli* PPTases. Accordingly, flux towards vanillin can possibly be increased by expressing another PPTase than the one used in this study. Once this bottleneck has been removed, greater benefits of enzyme fusion could perhaps be demonstrated, because vanillin accumulation under these circumstances would be expected to be more pronounced.

Chapter 4: Introduction to the nano-platform concept

In Chapter 2 and 3, it was demonstrated that cell factory engineering can benefit from positioning catalytic sites in close proximity by enzyme fusion. This inspired the development of another more flexible approach for exploiting proximity effects *in vivo*. The second strategy was to use a self-associating protein to guide the assembly of metabolic enzymes. If metabolic enzymes are fused to a scaffold protein (scaffoldin) that is able to self-associate and form a multimeric structure such as e.g. a ring-structure, the enzymes will automatically be positioned in close proximity once the scaffoldin self-associates *in vivo* (Fig. 4-1). Here, the term nano-platform is used to describe a structure consisting of a multimeric scaffoldin with metabolic enzymes linked to each of its subunits. Each subunit of the nano-platform can be expressed separately and subsequently the nano-platform is assembled post-translationally. For simplicity, the concept is shown here with three metabolic enzymes but several enzymes could be added to the nano-platform. The second part of this project aimed at developing the technology for constructing such a self-assembling nano-platform.

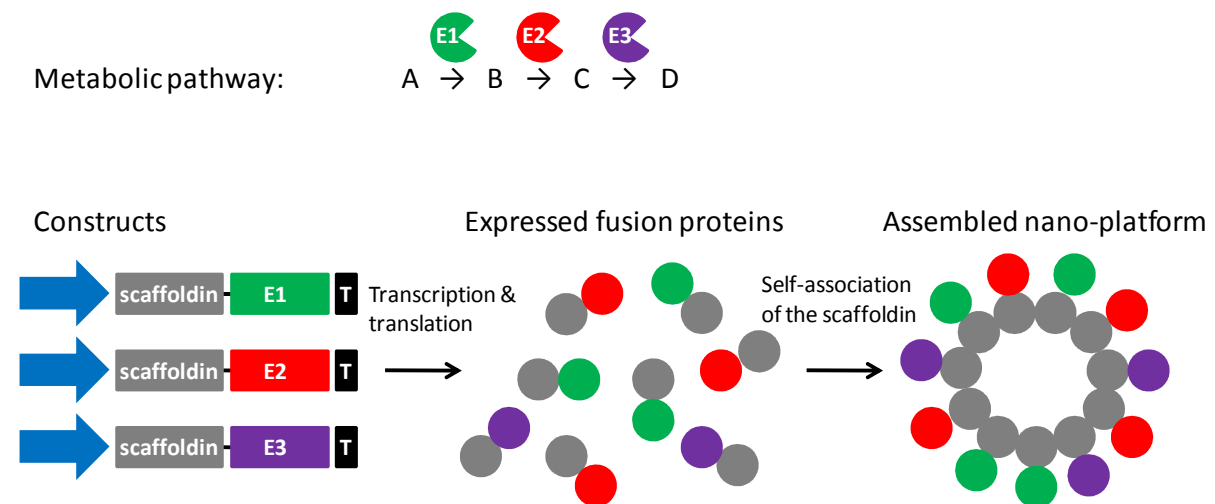


Figure 4-1. Nano-platform concept. When fusion proteins consisting of a self-associating protein linked to three different metabolic enzymes are expressed, the self-associating protein will assemble and form a nano-platform with three metabolic enzymes attached to it.

4.1. Requirements to the scaffoldin

In order to serve as an ideal custom nano-platform core, the scaffoldin must possess several properties. Most importantly, the scaffoldin should be able to self-associate and form a stable multimeric structure that does not affect metabolism negatively. Moreover, it should be possible to tag the scaffoldin both C- and N-terminally without interfering with its ability to self-associate, as this allows the metabolic enzymes to be fused to the scaffoldin via their preferred terminal. Furthermore, the scaffoldin's localization and mobility is important since metabolic processes takes place in various organelles. To provide the most universal application of the concept, the scaffoldin should preferably have the ability to sort to any organelle of interest if tagged with a proper signal peptide. Lastly, the scaffoldin's ability to be highly expressed and properly folded are important parameters. When two enzymes are fused, misfolding can be a problem particularly if the resulting fusion proteins are large in size. For this reason it is of interest to minimize the size of the scaffoldin to ensure high functionality.

4.2. Advantages of a self-assembling nano-platform

There are several advantages of assembling metabolic enzymes on a nano-platform instead of directly fusing them:

- 1) **Several enzymes can be attached to the nano-platform without having to express large proteins.** The number of enzymes that can be attached to the nano-platform is theoretically only limited by the number of subunits in multimeric structure formed by the scaffoldin. As several proteins are known to form multimers with more than 12 subunits (Jones and Thornton, 1996), nano-platforms with more than 12 enzymes attached to it could theoretically be constructed. If on the other hand, 12 catalytic sites were to be positioned in close proximity by direct fusion of the genes, the resulting fusion protein would most likely display severe folding problems.
- 2) **Ratios of the expressed metabolic enzymes can be tuned.** When enzymes are fused to each other the ratio will always be 1 to 1, but when enzymes are fused to a scaffoldin it is possible to change the ratio between the enzymes by expressing the various nano-

platform subunits from different promoters. This may for example be an advantage if one of the enzymes has a much lower activity than the other.

- 3) **Folding problems can be minimized.** Misfolding is a general problem when working with fusion proteins. In the case where enzymes are fused to a scaffoldin instead of each other, a scaffoldin that folds well may be selected as the fusion partner and this may reduce misfolding.
- 4) **Enzymes can be fused at their preferred terminal.** Some proteins are only functional when they are tagged at a specific terminal (Orita et al., 2007, Hong et al., 2006). If wanting to coordinate two enzymes that are functional only when tagged via their N-terminal, it would be impossible to construct a functional fusion protein as one protein inevitably has to be fused via the C-terminal and the other via the N-terminal. In this case fusion to the scaffoldin, that allows the enzymes to be tagged at their preferred terminal, would be the only solution.

4.3. Rad52 as a scaffoldin candidate

Initially, Rad52 was identified as a good scaffoldin candidate because this protein possesses the following properties:

- 1) Rad52 forms ring-structures with 7-11 subunits in the ring depending on whether the full-length protein or a C-terminal truncation is expressed (Shinohara et al., 1998, Ranatunga et al., 2001b, Stasiak et al., 2000)
- 2) The crystal structure has been solved for a C-terminal truncation of human Rad52 (HsRad52- Δ 212) (Kagawa et al., 2002, Singleton et al., 2002). The fact that the crystal structure is known may facilitate further protein engineering of the scaffoldin.
- 3) Tagging with other proteins does not seem to impair the functionality of Rad52 of *S. cerevisiae* (ScRad52), as ScRad52-YFP is biologically functional and can repair DNA almost as efficiently as the wild type protein (Lisby et al., 2001).
- 4) The functional domains involved in self-association, DNA-binding and association to other proteins have been mapped (Cortes-Ledesma et al., 2004, Hays et al., 1998, Krejci et al., 2002, Mortensen et al., 1996, Mortensen et al., 2002, Park et al., 1996, Ranatunga

et al., 2001b, Shen et al., 1996). This may greatly facilitate the construction of inert mutants.

- 5) The rings-structure has been demonstrated to be remarkably stable as it is extremely resistant to heat (Ranatunga et al. 2001)

In summary, Rad52 fulfils several of the requirements to an ideal scaffoldin candidate. Little is however known about Rad52's mobility. GFP-tagging has demonstrated that it localizes to the nucleus, but the mechanism of transport is unknown. As mentioned in the requirements to the ideal scaffoldin, the scaffoldin should preferably be mobile. Initially, a cytosolic localization would however be preferred, as several of the processes of interest to cell factory engineering takes place in the cytosol and as the cytosol generally serves as a starting point for sorting of proteins to other organelles. For this reason it was of interest to learn more about how Rad52 is transported to the nucleus and if possible, identify a nuclear transport mutant that resides in the cytosol. In the following chapter the results leading to the identification of a suitable scaffoldin is presented and in the subsequent chapter the concept is tested with two model pathways.

Chapter 5: Evaluation of Rad52's applicability as a scaffoldin for enzyme assembly *in vivo*.

This chapter presents the results of the experiments that were carried out to identify a scaffoldin that allows the assembly of metabolic enzymes on a nano-platform *in vivo*. The study was carried out in *S. cerevisiae* and the native yeast protein ScRad52 served as a starting point for the identification of a suitable scaffoldin. Except for its localization, ScRad52 fulfils the requirements to an ideal scaffoldin. In an effort to learn more about ScRad52's mobility, the mechanism of nuclear transport was studied.

The first part of the results in this chapter describes the biological mechanism underlying nuclear transport of ScRad52. This part has been published in separate publication:

Plate, I., Albertsen, L., Lisby, M., Hallwyl, S., Feng, Q., Seong, C., Rothstein, R., Sung, P., and Mortensen, U. (2008). Rad52 multimerization is important for its nuclear localization in *Saccharomyces cerevisiae*. *DNA Repair*, 7(1):57–66.

In the second part of this chapter, the work published in Plate et al., 2008 is interpreted with respect to the applicability of ScRad52 as a scaffoldin and it is concluded that ScRad52 is not a good scaffoldin candidate. Lastly, Rad52 of mouse (MmRad52) is evaluated as a scaffoldin candidate and found to be suitable backbone for enzyme assembly in *S. cerevisiae*.

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/dnarepair

Rad52 multimerization is important for its nuclear localization in *Saccharomyces cerevisiae*

Iben Plate^a, Line Albertsen^a, Michael Lisby^b, Swee C.L. Hallwyl^a, Qi Feng^c, Changhyun Seong^d, Rodney Rothstein^c, Patrick Sung^d, Uffe H. Mortensen^{a,*}

^a Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Bldg. 223, DK-2800 Kgs. Lyngby, Denmark

^b Department of Molecular Biology, University of Copenhagen, Ole Maaløesvej 5, DK-2200 Copenhagen N, Denmark

^c Department of Genetics & Development, Columbia University Medical Center, 701 West 168th Street, New York, NY 10032-2704, USA

^d Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520, USA

ARTICLE INFO

Article history:

Published on line 20 September 2007

Keywords:

DNA repair

Homologous recombination

Rad52

Nuclear localization

NLS

S. cerevisiae

ABSTRACT

Rad52 is essential for all homologous recombination and DNA double strand break repair events in *Saccharomyces cerevisiae*. This protein is multifunctional and contains several domains that allow it to interact with DNA as well as with different repair proteins. However, it has been unclear how Rad52 enters the nucleus. In the present study, we have used a combination of mutagenesis and sequence analysis to show that Rad52 from *S. cerevisiae* contains a single functional pat7 type NLS essential for its nuclear localization. The region containing the NLS seems only to be involved in nuclear transport as it plays no role in repair of MMS-induced DNA damage. The NLS in Rad52 is weak, as monomeric protein species that harbor this NLS are mainly located in the cytosol. In contrast, multimeric protein complexes wherein each subunit contains a single NLS_{Rad52} sort efficiently to the nucleus. Based on the results we propose a model where the additive effect of multiple NLS_{Rad52} sequences in a Rad52 ring-structure ensures efficient nuclear localization of Rad52.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

The integrity of the genome is constantly challenged by DNA damage induced by reactive metabolic intermediates and environmental agents. Among the different types of DNA lesions that can occur, DNA double strand breaks are particularly dangerous, as they may cause cell death or provoke genomic rearrangements. In *Saccharomyces cerevisiae*, DNA double strand breaks are mainly repaired by pathways that involve homologous recombination (HR). HR depends on the genes of the RAD52 epistasis group, RAD50, RAD51, RAD52, RAD54, RAD55, RAD57, RAD59, RDH54, RFA1, MRE11 and XRS2 [1]. Among these, mutations in RAD52 show the most severe phenotype, reflecting the involvement of this gene in multiple HR pathways. The importance of Rad52

is stressed by the fact that it is conserved from yeast to human.

Several biochemical properties of Rad52 are germane for its HR role, including DNA binding and an ability to interact with the Rad51 recombinase, Rad59 protein, and the single-strand DNA binding protein RPA. These attributes enable Rad52 to promote the annealing of RPA-coated ssDNA and to function with Rad51 in the displacement of RPA from ssDNA [2–9]. The highly conserved N-terminus of Rad52 contains domains that allow it to self-associate and form ring-structures, to bind Rad59, to bind DNA and to facilitate DNA annealing (Fig. 1) [7,2,8–11]. The middle- and C-terminal regions of yeast and human Rad52 proteins have been shown to contain the RPA and Rad51 interaction domains, respectively, but are otherwise not well conserved in primary sequence [12,13,2,14,15]

* Corresponding author. Tel.: +45 4525 2701; fax: +45 4588 4148.

E-mail address: um@biocentrum.dtu.dk (U.H. Mortensen).

1568-7864/\$ – see front matter © 2007 Elsevier B.V. All rights reserved.

doi:10.1016/j.dnarep.2007.07.016

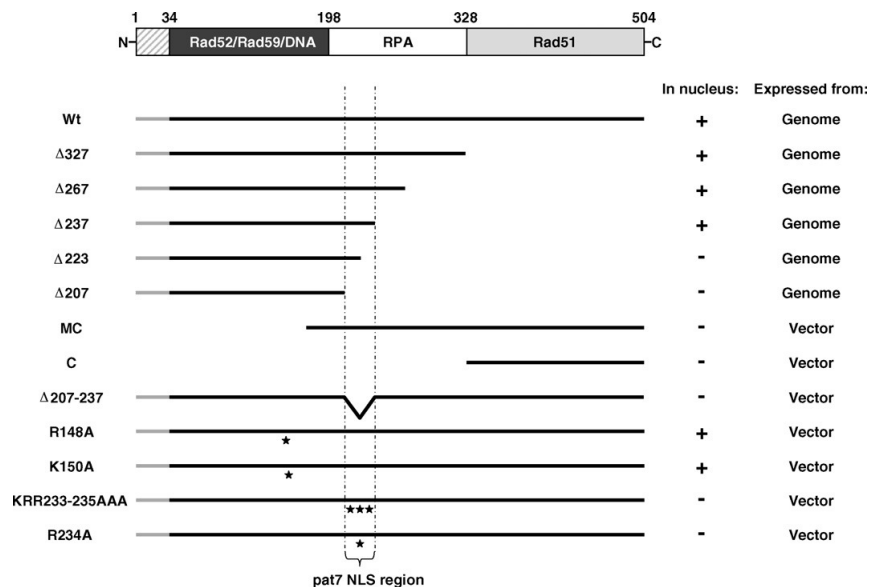


Fig. 1 – Functional map of *Saccharomyces cerevisiae* Rad52 and an overview of the cellular localization of all Rad52 mutants. A schematic representation of Rad52 from *S. cerevisiae* is presented in the top. The hatched region covering aa residues 1–34 is not expressed. The dark region spanning aa residues 34–198 corresponds to the region of Rad52 that is highly evolutionary conserved. The regions in Rad52 that are involved in protein–protein interactions and in binding to DNA are indicated. A diagram showing all individual Rad52 deletion and mutated species relative to wild-type Rad52 is presented below. All Rad52 species are C-terminally extended by YFP (not shown in the figure). Rad52-YFP (Wt), Rad52-Δ327-YFP (Δ327), Rad52-Δ267-YFP (Δ267), Rad52-Δ237-YFP (Δ237), Rad52-Δ223-YFP (Δ223), Rad52-Δ207-YFP (Δ207), Rad52MC-YFP (MC), Rad52C-YFP (C), Rad52-Δ207-237-YFP (Δ207-237), Rad52-R148A-YFP (R148A), Rad52-K150A-YFP (K150A), Rad52-KRR233-235AAA-YFP (KRR233-235AAA) and Rad52-R234A-YFP (R234A). The position of single aa residue substitutions are indicated with stars and the pat7 NLS region is indicated below the figure. To the right, the cellular localization of individual Rad52 species and whether it is expressed from the genomic *RAD52* locus or from a vector is shown. Nuclear and cytosolic localization of a given Rad52 species is indicated by (+) and (–), respectively.

(Krejci et al., submitted). Notably, it has remained unclear how Rad52 is transported into the nucleus.

Most nuclear proteins larger than 40–60 kDa require active transport to enter the nucleus [16]. This transport is facilitated by nuclear transport receptors, importins, which recognize nuclear localization signals, NLSs, which are typically composed by clusters of basic amino acid (aa) residues [16]. The complex of a cargo protein and a nuclear transport receptor is then shuttled from the cytosol into the nucleus through the nuclear pore complex by forming transient interactions with nucleoporins that line the channel of the pore. In a previous search for NLS motifs in DNA repair proteins, no putative NLS in *S. cerevisiae* Rad52 was identified and it was proposed that Rad52 is escorted into the nucleus via an interaction with another protein factor that harbors such a transport signal [17]. We have located the region in Rad52 required for its nuclear localization. Combining these domain mapping results with a complementary sequence analysis, we have identified a single “pat7” type NLS in Rad52 and shown that it is essential and sufficient for efficient Rad52 transport into the nucleus. Interestingly, the functionality of this NLS seems to be dependent on Rad52 oligomerization being mediated by the N-terminus of the protein.

2. Materials and methods

2.1. Genetic methods and strains

All media were prepared as described by Sherman [18] with minor modifications as the synthetic medium contained twice the amount of leucine (60 mg/L). All strains are isogenic to W303 [19] except they are *RAD5* [20,21], and *ADE2* (see Table 1). Integrated *RAD52* mutants were constructed and fused to YFP using the cloning-free PCR-based allele replacement method previously described by Erdeniz and colleagues [22,23]. Correct integration of the mutations were verified by PCR and sequencing (MWG-Biotech AG).

2.2. Plasmid construction

2.2.1. Plasmids expressing *MmRad52*-YFP and *KlRad52*-YFP

Plasmids were constructed from the CEN6-based plasmid pWJ1213 [24] by replacing *S. cerevisiae* *RAD52* with *RAD52* from *Kluyveromyces lactis* and *Mus musculus* preserving the *S. cerevisiae* *RAD52* promoter. Both genes are lacking stop codons

Table 1 – Yeast strains used in this study

Strain ^a	Genotype
UM74-3B	MATa bar1::LEU2 his3-11, 15 leu2-3, 112 trp1-1 ura3-1 RAD52-YFP
UM94-9C	MATa his3-11, 15 leu2-3, 112 trp1-1 ura3-1 rad52-Δ327-YFP
UM93-12D	MATa his3-11, 15 leu2-3, 112 trp1-1 ura3-1 rad52-Δ267-YFP
UM227-9D	MATalpha his3-11, 15 leu2-3, 112 lys2Δ ura3-1 rad52-Δ237-YFP
UM262-3B	MATa his3-11, 15 leu2-3, 112 trp1 ura3-1 rad52-Δ223-YFP
UM261-9C	MATalpha his3-11, 15 leu2-3, 112 lys2Δ ura3-1 Rad52-Δ207-YFP
UM69-1A	MATa bar1::LEU2 can1-100 his3-11, 15 leu2-3, 112 trp1-1 ura3-1 rad52-Δ207-CFP
UM94-5D	MATa his3-11, 15 leu2-3, 112 trp1-1 ura3-1 RAD52
UM263-7C	MATalpha his3-11, 15 leu2-3, 112 lys2Δ ura3-1 RAD52-YFP
UMR128	MATalpha bar1::LEU2 his3-11, 15 trp1-1 ura3-1 rad52-R234A-YFP
W3849-10C	MATa bar1::LEU2 his3-11, 15 ura3-1 RAD52-CFP
UM101-15B	MATa his3-11, 15 leu2-3, 112 trp1-1 ura3-1 rad52::HIS5

^a All strains are derivatives of W303-1A and W303-1B. In addition to the genotype listed above all strains are RAD5 and ADE2.

to ensure fusion to YFP. The new vectors, pPL2 and pPL3, harboring RAD52-YFP from either *K. lactis* or *M. musculus* were transformed into a *rad52Δ* strain (UM101-15B) and nuclear localization visualized by fluorescent microscopy.

2.2.2. Plasmids expressing Rad52-YFP

A series of plasmids expressing Rad52-YFP with mutations in areas predicted to be involved in nuclear transport of Rad52 were constructed. All plasmids were constructed by inserting an AgeI-SphI digested RAD52 fragment into an AgeI-SphI vector fragment of pWJ1213. Plasmids harboring *rad52-R148A-YFP* and *rad52-K150A-YFP* were constructed from two plasmids (p52mut-R148A and p52mut-K150A) previously constructed by Mortensen et al. [25]. To create fragments encoding the mutations *rad52-KRR233-235AAA*, *rad52-R234A* and *rad52-Δ207-37*, sequences flanking the mutation were amplified by PCR. The primers used to construct the RAD52 plasmids are listed in [supplementary material](#). The PCR fragments were inserted into pWJ1213 with AgeI and SphI to generate the plasmids pWJ1213-*rad52-KRR233-235AAA-YFP*, pWJ1213-*rad52-R234A-YFP* and pWJ1213-*rad52-Δ207-37-YFP*.

2.2.3. Plasmids expressing NLS-tagged Rad52-YFP mutants

To tag *rad52-Δ207-237-YFP* with the NLS_{SV40} sequence [26,27], a PCR fragment encoding *rad52-Δ207-237-YFP-NLS* was constructed by PCR using pWJ1213-*rad52-Δ207-237* as template. The PCR fragment was inserted into vector pWJ1213 to generate pWJ1213-*rad52-Δ207-37-YFP-NLS* using AgeI and XhoI. A plasmid harboring *rad52-R234A-YFP-NLS* was constructed by digestion of pWJ1213-*rad52-Δ207-37-YFP-NLS* with SphI and XhoI. The resulting fragment was inserted into a SphI-XhoI

vector fragment of pWJ1213-*rad52-R234A-YFP* to generate the plasmid pWJ1213-*rad52-R234A-YFP-NLS*.

The CEN6-based RAD52-YFP expression vector, pWJ1213, and the 2-micron-based RAD52-YFP vector pWJ1214 were used to clone *rad52MC-YFP* and *rad52C-YFP* (primers are listed in [supplementary material](#)).

2.2.4. Plasmids expressing mono- and tetrameric DsRed tagged with the NLS of Rad52

Plasmids expressing either monomeric DsRed (pCupmono-DsRed-NLS_{Rad52}) or tetrameric DsRed (pCup-tetra-DsRed-NLS_{Rad52}) tagged at the C-terminus with NLS_{Rad52} (PNKRRQL) were constructed from pCM1513 (A kind gift from C. Müller) leading to a protein under the control of the inducible Cup1 promoter. A PCR fragment harboring monomeric DsRed-NLS_{Rad52} was constructed by PCR using pCM1513 as a template (primers are listed in [supplementary material](#)), and a fragment harboring tetrameric DsRed-NLS_{Rad52} using pPgpD-DsRed as a template [28]. Both the monomeric and the tetrameric DsRed containing PCR fragments were digested with HpaI-HindIII and ligated into a HpaI-HindIII vector fragment of pCM1513. Plasmids were transformed into UM94-5D and nuclear localization was visualized by fluorescence microscopy.

2.3. MMS assay

To assess sensitivity of mutant strains to methyl methanesulfonate (MMS) (M4016 from Sigma), plasmids harboring wild type RAD52, *rad52-Δ207-37-YFP*, *rad52-Δ207-37-YFP-NLS_{SV40}*, *rad52-R234A-YFP*, *rad52-R234A-YFP-NLS_{SV40}* and an empty plasmid were transformed into a *rad52Δ* strain (UM101-15B). Cells were grown overnight at 30 °C in selective media (SC-His), washed with sterile water and resuspended in an appropriate volume. Subsequently, six 10-fold dilutions were made of cell suspensions containing 10⁸ cells per ml and 5 μl of each dilution was spotted on SC-His plates containing no or 0.0025% MMS. Plates were incubated at 30 °C for 2 days before examination [29].

2.4. Fluorescence microscopy and imaging

Microscopy was essentially performed as previously described [30]. Cells were grown in SC medium prior to microscopy, except when RAD52 molecules were expressed from a plasmid. In this case, cells were grown in SC-His or SC-Leu medium to select for the plasmid. When fluorescent molecules were expressed under the control of the Cup1 promoter, CuSO₄ was added to the inoculation media to a final concentration of 0.2 mM Cu²⁺ to induce expression. In all experiments cells were grown at 23 °C to allow efficient formation of the chromosome.

DNA in living cells was stained for visualization by adding 10 μg/ml DAPI to the culture 30 min prior to imaging. Selected strains were made ρ⁰ (mitochondrial DNA negative) before staining to eliminate any signal from mitochondrial DNA.

2.5. Co-expression of RAD52 strains

UM69-1A was crossed with UM261-9C, UM227-9D, and UM263-7C to generate diploids *rad52-Δ207-CFP/rad52-Δ207-YFP*,

rad52-Δ207-CFP/rad52-Δ237-YFP and *rad52-Δ207-CFP/RAD52-YFP*. Diploid cells were grown overnight in SC media and the subcellular localization of the proteins examined by using fluorescence microscopy.

2.6. Purification of Rad52MC

(His)₆-RAD52MC was constructed in pRSET-c and purified from *E. coli* strain Rosetta (Novagen) as described (Krejci et. al., submitted).

2.7. Gel filtration analysis of Rad52MC

(His)₆-tagged Rad52MC (100 μg) was fractionated in a 36-ml Sephacryl S300 column in K buffer containing 150 mM KCl, collecting 0.4 ml fractions at 0.1 ml/min. The column fractions were analyzed by 10% SDS-PAGE and staining with Coomassie Blue. The column was calibrated with thyroglobulin (663 kDa), catalase (223 kDa) and ovalbumin (43 kDa).

3. Results

3.1. Nuclear localization of Rad52 is mediated by a region in the middle section of the protein

Rad52 of *S. cerevisiae* is predominantly localized in the nucleus in all phases of the cell cycle, even in the absence of genotoxins [23] and Fig. 2. To delimit the region in Rad52 required for its nuclear localization, a series of five mutant strains expressing YFP tagged C-terminally truncated Rad52 species expressed from the endogenous RAD52 locus were constructed (Fig. 1). These truncation alleles terminate at aa residue positions ranging from 207 to 327, compared to wild-type Rad52, which terminates at position 504. In addition, two N-terminally truncated species, one starting from aa residue 169 (Rad52MC-YFP) and one from aa residue 327 (Rad52C-YFP) were constructed. All mutant strains were subjected to fluorescence microscopy to determine the cellular localization of the fusion proteins. Of these truncations, Rad52-Δ327-YFP was expected to sort into the nucleus because the MMS sensitivity of a *rad52-Δ327* strain has previously been shown to be fully suppressed by over-expression of Rad51 [13,31]. In agreement with this, we find that Rad52-Δ327-YFP localizes in the nucleus. Interestingly, we find that the truncation species Rad52-Δ237-YFP and Rad52-Δ267-YFP also sort correctly into the nucleus whereas the two shortest truncations, Rad52-Δ207-YFP and Rad52-Δ223-YFP mainly localize in the cytosol (Figs. 1 and 2 supplementary material). In addition, the N-terminal truncations Rad52MC-YFP and Rad52C-YFP also fail to sort efficiently to the nucleus. To rule out the possibility that mis-sorting of the Rad52 truncation species could be explained by their nuclear transport being compromised by the YFP moiety of the fusion protein, we constructed a Rad52-YFP species containing an internal deletion, Rad52-Δ207-237-YFP, and expressed it from a single copy plasmid in a *rad52Δ* strain. This species also fails to sort correctly to the nucleus (Fig. 1). Together these results indicate that a region involved in nuclear localization is located between or close to aa residues 207-237, a region of the protein that has not previously been assigned any Rad52

function. However, the fact that Rad52MC-YFP, which contains aa residues 207-237, accumulates in the cytosol (Figs. 1 and 2) suggests that nuclear localization of Rad52 may also depend on other Rad52 features, see below.

3.2. A nuclear localization signal is present in the middle of Rad52

Next, we investigated the possibility that the domain responsible for nuclear localization of Rad52 is sufficiently conserved to retain interspecies functionality. Hence, *K. lactis* and *M. musculus* Rad52 species (KlRad52 and MmRad52) tagged with YFP were individually expressed under the control of the *S. cerevisiae* RAD52 promoter from single copy plasmids in *S. cerevisiae rad52Δ* strains. When these strains were examined by fluorescence microscopy, we observed that KlRad52-YFP locates in the nucleus whereas MmRad52-YFP mainly remains in the cytosol (Fig. 3A) suggesting that a common mechanism ensures nuclear localization of Rad52 and KlRad52, but not MmRad52, in *S. cerevisiae*. In agreement with this view, sequence comparisons of the aa residues 207-237 in Rad52 to the corresponding Rad52 sequences from *K. lactis* and *M. musculus* show that only Rad52 and KlRad52 share identical stretches of aa residues in this region (Fig. 3B). Importantly, we find a proline residue at position 231 in Rad52 that is situated close to a stretch of three positively charged aa residues (PNKRR). At the corresponding position, a similar motif is present in KlRad52 (PSLKRR), but not in MmRad52. Both motifs qualify as pat7 NLS sequences [32] and therefore constitute sequences that may be involved in nuclear localization of KlRad52 and Rad52 in *S. cerevisiae*. The fact that the putative NLS sequence in Rad52 (and in KlRad52) only contains three basic aa residues explains why it was not recognized in the more stringent NLS search performed previously by Boulikas [33,17] who defined a monopartite NLS motif as a cluster of at least four basic residues within a hexapeptide. The existence of a pat7 type NLS in Rad52 prompted us to perform a new sequence search for other putative NLS sequences in the entire Rad52 primary sequence using two different algorithms PSORT II and Predict NLS [32,34]. Using PSORT II, one additional candidate sequence, a pat4 sequence (RRKP) [32] at position 148-151 was identified. However, this motif is not conserved in KlRad52 and MmRad52 (Fig. 3B).

3.3. A single NLS ensures nuclear localization of yeast Rad52

Three basic aa residues (K233-R234-R235) constitute the core of the predicted pat7 NLS identified in Rad52. To verify that these aa residues in fact are part of an NLS, the triple mutant Rad52-KRR233-235AAA-YFP and the single mutant Rad52-R234A-YFP were individually expressed from single copy plasmids in *rad52Δ* strains. In agreement with this motif being an NLS, both Rad52-KRR233-235AAA-YFP and Rad52-R234A-YFP were observed to remain mainly in the cytosol (Figs. 1 and 2). To rule out the possibility that mis-sorting was due to the Rad52 mutants being expressed from a plasmid, the R234A mutation was introduced into the genomic version of RAD52-YFP. In the resulting strain, Rad52-R234A-YFP was also observed to localize in the cytosol (unpublished

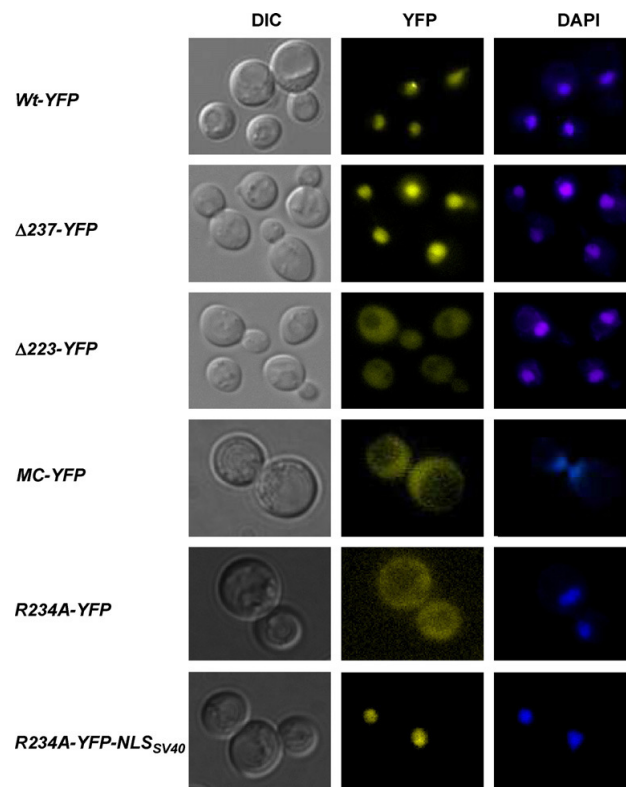


Fig. 2 – Cellular localization of Rad52 mutant proteins. Microscopy of cells expressing Rad52-YFP or corresponding mutant derivatives as indicated. Pictures shown are pseudocolored monochrome images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

data). Next, we investigated the role of the pat4 NLS motif spanning aa residues 148–151 in nuclear sorting. Two mutant Rad52 species, Rad52-R148A-YFP and Rad52-K150A-YFP, were expressed from plasmids in *rad52Δ* strains and their cellular location determined. Both mutants were found to be present in the nucleus showing that this pat4 NLS motif plays little or no role in the nuclear localization of Rad52 (Fig. 1). In fact, this result was expected as two Rad52 mutant strains, *rad52-R148A* and *rad52-R149A*, were previously shown to repair γ -ray-induced damage and to perform mitotic homologous recombination at wild-type levels [25] showing that sufficient Rad52 protein must reach the nucleus to maintain these functions. We therefore conclude that a single NLS of the pat7 type spanning aa residues 231–235 is responsible for targeting Rad52 to the nucleus.

3.4. The NLS_{Rad52} can direct a heterologous cytosolic protein to the nucleus

To demonstrate that the PNKRR motif in Rad52 acts as a functional NLS, we extended the C-terminus of monomeric DsRed, which is normally located in the cytosol with an NLS_{Rad52}, and asked whether the presence of this motif would direct

it to the nucleus. Surprisingly, this species fails to sort efficiently to the nucleus (Fig. 4). In contrast, if monomeric DsRed is C-terminally fused to the well-characterized NLS sequence from SV40 virus, NLS_{SV40} [26,27], then monomeric DsRed localizes in the nucleus (data not shown). This result suggests that a single NLS_{Rad52} is insufficient to mediate the transport of monomeric DsRed to the nucleus. Since some proteins contain several NLS sequences to ensure efficient nuclear sorting, we fused NLS_{Rad52} to the C-terminus of tetrameric DsRed to determine whether this protein complex that contains a total of four NLS_{Rad52} sequences would localize to the nucleus. Unlike monomeric DsRed-NLS_{Rad52} tetrameric DsRed-NLS_{Rad52} is highly concentrated in the nucleus (Fig. 4) showing that the NLS_{Rad52} can act as a functional NLS if it is present in one copy in each of the subunits of a homomultimeric protein complex.

3.5. The role of Rad52 multimerization in nuclear localization of Rad52

Rad52 exists as a heptameric ring-structure and the observation that the NLS_{Rad52} promotes efficient nuclear localization of tetrameric, but not monomeric, DsRed suggests that Rad52

62

DNA REPAIR 7 (2008) 57-66

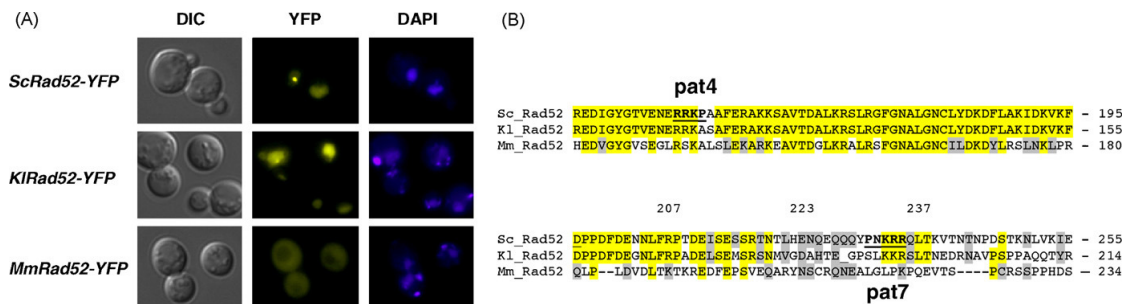


Fig. 3 – Localization of Rad52, KlRad52 and MmRad52 in *S. cerevisiae* cells. (A) *rad52Δ* cells expressing *S. cerevisiae*, *Kluyveromyces lactis* or *Mus musculus* Rad52-YFP fusion proteins from plasmids as indicated. Pictures shown are pseudocolored monochrome images. (B) A sequence comparison of the region of Rad52, which is important for nuclear targeting, to the corresponding regions of KlRad52 and MmRad52. Identical or similar aa residues are highlighted by yellow and grey, respectively. The positions of the two predicted NLS sequences, pat4 and pat7, in *S. cerevisiae* Rad52 are indicated as bold underscored aa residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

multimerization may also be important for the nuclear localization of Rad52. It would therefore be interesting to test whether a Rad52 mutant that fails to form ring-structures, yet contains the NLS_{Rad52}, sorts to the nucleus. In fact, the fragment Rad52MC-YFP described above may represent such a mutation as it lacks the N-terminal self-association domain responsible for ring-structure formation, but contains the NLS_{Rad52} sequence. Importantly, this Rad52 fragment fails to concentrate in the nucleus (Figs. 1 and 2). Since the C-terminus of human Rad52 contains a self-association domain that allows Rad52 ring structures to further multimerize [12], we tested whether Rad52MC exists as a monomer or a multimer as judged by a gel-filtration experiment. As shown in Fig. 5A, Rad52MC elutes at a volume corresponding to that expected if it forms a tetra- or pentamer. To investigate whether this C-terminal self-association domain contributes to the nuclear localization of Rad52, we co-expressed Rad52MC-YFP and Rad52-CFP in the same strain. Of these two protein species, only Rad52-CFP concentrates in the nucleus (Fig. 5B) indicating that the interaction between Rad52MC-YFP and Rad52-CFP

is not sufficiently strong to ensure Rad52-CFP mediated nuclear localization of Rad52MC-YFP. Similarly, MC-NLS_{SV40} fails to mediate nuclear localization of Rad52-R234A-YFP (data not shown). Finally, we asked whether the ability of a sorting defective Rad52 mutant to participate in a multimeric ring-structure could promote its nuclear localization if it is co-expressed with a sorting proficient Rad52 species. This possibility was tested by co-expressing Rad52-Δ207-CFP with Rad52-Δ237-YFP and by co-expressing Rad52-Δ207-CFP together with wild-type Rad52-YFP. In both cases, the NLS defective Rad52 mutants were mainly located in the nucleus (Fig. 5C) showing that Rad52 multimerization may facilitate nuclear targeting.

3.6. The NLS region of Rad52 is not involved in the DNA DSB repair functions of Rad52

Finally, we investigated whether the region of Rad52 containing the NLS motif participates directly in DNA repair in addition to its role in nuclear transport. To address this pos-

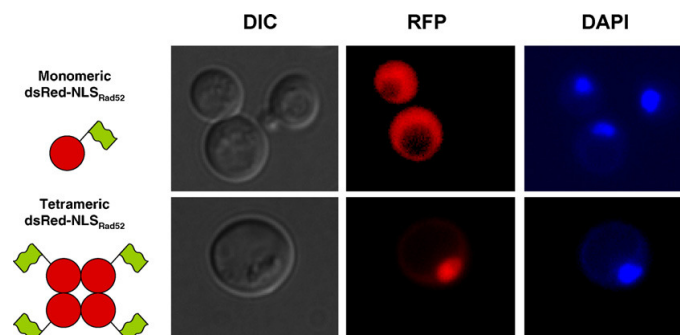


Fig. 4 – The pat7 NLS sequence in Rad52 is a functional NLS. The cellular localization of monomeric and tetrameric DsRed tagged with NLS_{Rad52}. DsRed is depicted as a filled red circle and the NLS_{Rad52} sequences as green flags. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

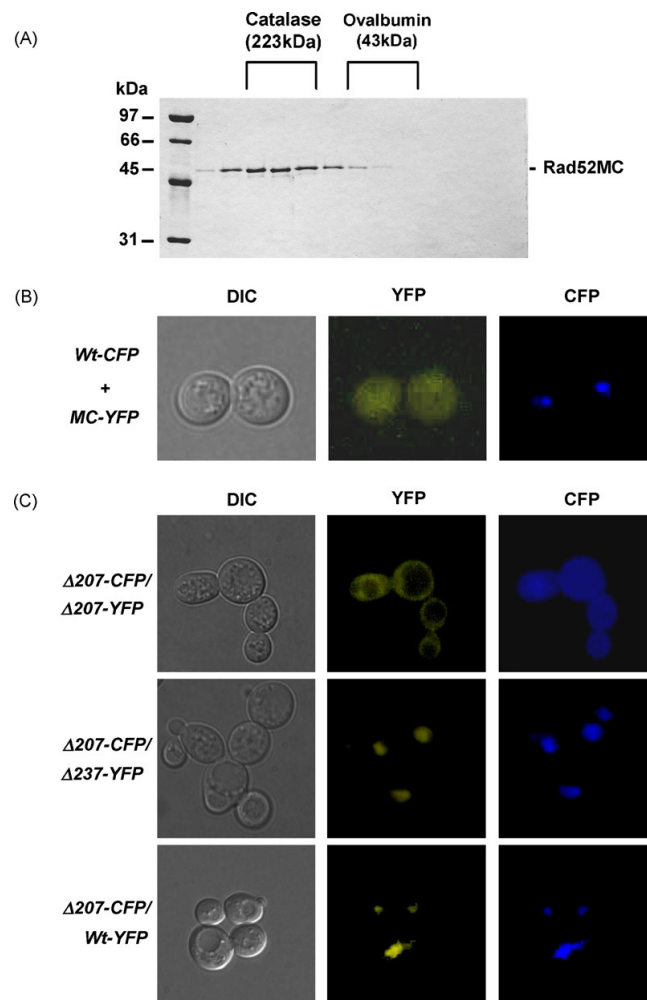


Fig. 5 – Rad52 self-association mediates nuclear transport. (A) Elution profile of Rad52MC after gel filtration. Fractions eluting at the same volume as the marker proteins catalase and ovalbumin are indicated. (B) Expression of Rad52MC-YFP from a plasmid in a *RAD52-CFP* strain. Co-localization of Rad52 species tagged with either CFP or YFP. (C) Microscopy of heterozygous diploid strains co-expressing Rad52- $\Delta 207$ -CFP and Rad52- $\Delta 207$ -YFP, Rad52- $\Delta 207$ -CFP and Rad52- $\Delta 237$ -YFP and Rad52- $\Delta 207$ -CFP and Rad52-YFP.

sibility, we tagged Rad52-R234A-YFP C-terminally with the NLS_{SV40} to mediate its nuclear transport by a non-Rad52 sequence. As expected, this fusion protein is indeed concentrated in the nucleus (Fig. 2). The ability of the resulting strain, *rad52-R234A-YFP-NLS_{SV40}*, to perform Rad52 functions during repair of MMS-induced damage was determined and compared to that of *rad52-R234A-YFP*. As expected *rad52-R234A-YFP* strains, like *rad52 Δ* strains, are sensitive to MMS reflecting the absence of Rad52 in the nucleus. In contrast, *rad52-R234A-YFP-NLS_{SV40}* strains, like wild-type strains, were not affected by this treatment (Fig. 6). In fact, it is possible to delete the entire region, aa residues 207–237, and still maintain efficient DNA repair, as *rad52 Δ* strains transformed

with a plasmid expressing Rad52- $\Delta 207$ -237-YFP-NLS_{SV40} are resistant to MMS whereas *rad52 Δ* strains expressing Rad52- $\Delta 207$ -237-YFP are not (supplementary material). These results show that the NLS region in Rad52 does not contribute directly to repair of MMS-induced DNA damage.

4. Discussion

In the present study, we have explored the mechanism of Rad52 nuclear localization by using a combination of mutagenesis and sequence analysis. We have identified a single pat7 type NLS, aa residues 231–235 (PNKRR), which is required

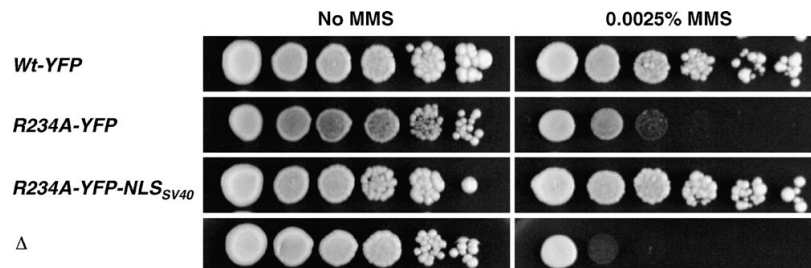


Fig. 6 – A mutation in the pat7 NLS sequence in Rad52 does not affect its ability to repair MMS-induced DNA damage. A spot assay of *rad52Δ* strains transformed with plasmids expressing *RAD52-YFP*, *rad52-R234A-YFP* or *rad52-R234A-YFP-NLS_{SV40}* as indicated. A strain transformed with an empty plasmid is also included in the analysis. Serial 10-fold dilutions of each transformed strain were spotted on selective media containing either no or 0.0025% MMS. Pictures were captured after 2 days.

and sufficient for efficient nuclear localization of Rad52. This NLS_{Rad52} is a functional NLS sequence as tetrameric DsRed tagged with NLS_{Rad52} is sorted efficiently to the nucleus. In some cases, an NLS sequence may provide a dual function in a protein. For example, in many proteins the NLS sequences overlap nucleic acid binding domains [35]. We observed, that Rad52 mutants, where the NLS_{Rad52} has been eliminated, efficiently repair MMS-induced DNA damage when they are tagged with NLS_{SV40}. This suggests that the NLS_{Rad52} only contributes to nuclear sorting of Rad52.

The observation that the NLS_{Rad52} only contains three basic aa residues suggests that it may be a weak signal for nuclear transport. In agreement with this, we find that a single NLS_{Rad52}, in contrast to NLS_{SV40}, does not mediate efficient nuclear localization of monomeric DsRed. In addition, the presence of four weak NLS_{Rad52} sequences in tetrameric DsRed-NLS_{Rad52} efficiently mediates nuclear transport of this protein complex. This result predicts that individual Rad52 subunits also do not accumulate in the nucleus. In agreement with this view, we observed that the Rad52 N-terminal truncation, Rad52MC-YFP, fails to sort to the nucleus despite the fact that it contains the NLS_{Rad52} (Fig. 2). This Rad52 mutant protein does not contain the N-terminal protein oligomerization domains responsible for ring-structure formation [7–9,15,36] and the self-association of the MC part of Rad52 observed *in vitro* ([12] and this study) does not seem to be sufficient for nuclear localization *in vivo*. Taken together, we conclude that Rad52 multimerization is an essential aspect of the nuclear localization mechanism of Rad52 because the NLS_{Rad52} is a weak NLS.

Two models can explain how Rad52 multimerization drives efficient nuclear localization (see Fig. 7). In the first model, the additive NLS signal model (Fig. 7A), the heptameric ring is formed in the cytosol, and the additive effect of seven weak NLS sequences within a single Rad52 ring-structure drives efficient nuclear targeting by, e.g. increasing the frequency of interactions to the nuclear transport receptors. Rad52 monomers remain in the cytoplasm because each monomer only contains a single weak NLS of Rad52, which is insufficient to mediate any significant transport into the nucleus. In the second model, the nuclear retention model (Fig. 7B), the heptameric ring is formed in the nucleus. In this model, the weak

Rad52-NLS allows Rad52 to enter the nucleus at a slow rate as a monomer. Moreover, in this model, Rad52 monomers exit the nucleus at a similar rate preventing monomers from accumulating in the nucleus. In contrast, Rad52 subunits, which are incorporated into heptameric ring-structures, are trapped in the nucleus, e.g. due to the larger size of this complex compared to monomers. A similar model has been proposed to explain nuclear localization of the pathogenic fusion protein CBF β -SMMHC (core binding factor beta fused to smooth muscle myosin heavy chain) [37].

The effects predicted by the two models may both contribute to the nuclear localization of Rad52. However, we note that the nuclear retention model does not easily explain the observation that Rad52- Δ 207-CFP and Rad52- Δ 237-YFP both localize in the nucleus when they are co-expressed. For example, in the nuclear retention model, Rad52- Δ 207-CFP enters the nucleus independently of Rad52- Δ 237-YFP. Hence, Rad52- Δ 207-CFP must be able to enter the nucleus independently of the NLS_{Rad52} sequence. This is unlikely as the size of this fusion protein is approximately 47 kDa. Moreover, since Rad52- Δ 207-CFP does not accumulate in the nucleus when it is expressed on its own, the retention model predicts that Rad52- Δ 207-CFP fails to multimerize in the nucleus since it is not retained. This seems unlikely as the nuclear retention model at the same time predicts that Rad52- Δ 207-CFP must be efficiently incorporated into multimers together with Rad52- Δ 237-YFP to ensure its nuclear retention in the co-expression experiment. In contrast, in the additive NLS model, efficient co-localization of Rad52- Δ 207-CFP and Rad52- Δ 237-YFP in the nucleus is easily explained. In this model Rad52- Δ 207-CFP and Rad52- Δ 237-YFP both form ring-structures in the cytosol, but the Rad52- Δ 207-CFP ring-structures do not enter the nucleus, as they do not contain any NLS sequences. However, when the two species are co-expressed they form chimeric multimers in the cytosol that enter the nucleus via NLS sequences present in the Rad52- Δ 237-YFP molecules. In this context it is important to note that the DsRed-NLS_{Rad52} tetramer efficiently sorts to the nucleus showing that four NLS_{Rad52} sequences are adequate for efficient nuclear transport. Since wild-type Rad52 rings contain seven [36], and the Rad52 truncations may even contain a higher number of subunits [12,38,39] then most chimeric complexes should have enough NLS_{Rad52} sequences

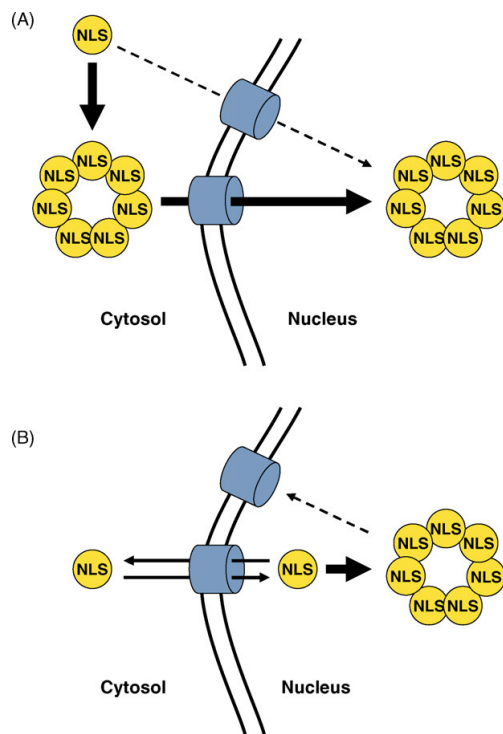


Fig. 7 – The illustrations show two models explaining how Rad52 multimerization may facilitate nuclear localization of Rad52. Rad52 monomers are depicted as filled yellow circles. The fact that each Rad52 molecule contains a single NLS sequence is indicated by the presence of a single “NLS” in each Rad52 monomer. The nuclear pore complex is shown as blue cylinders. (A) The additive NLS model. (B) The nuclear retention model. The models are explained in detail in the text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

to relocate efficiently to the nucleus. Finally, the nuclear retention model would likely depend on a mechanism preventing Rad52 from multimerizing in the cytoplasm to ensure efficient sorting. A similar level of complication is not required if nuclear transport occurs according to the additive NLS model. For these reasons, we favour a scenario where the mechanism predicted by the additive NLS signal model provides the driving force of Rad52 nuclear targeting.

The ability of Rad52 to form ring-like structures is evolutionary conserved from yeast to humans [40,9,12], which raises the question whether the nuclear sorting mechanism is also conserved. A pat7 NLS sequence similar to the one in *S. cerevisiae*, is present at the same position in Rad52 from *K. lactis*, *A. gossypii* and *C. glabrata* suggesting that the mechanism of nuclear transport of Rad52 is conserved in these yeasts. In fact, we find that KlRad52 is sorted to the nucleus of *S. cerevisiae*. This result was expected as it has previously been

demonstrated that expression of KlRad52 in *S. cerevisiae* partially suppresses the severe phenotype of *rad52Δ* strains [13]. On the other hand, Rad52 homologs from higher eukaryotes, e.g. chicken, zebra fish, mouse and human do not contain an NLS at this position, but rather they have a single putative NLS sequence in their C-terminal end suggesting that the Rad52 nuclear sorting mechanism may have diverged during evolution. Whether efficient nuclear targeting of Rad52 in higher eukaryotes is dependent on Rad52 multimerization therefore remains elusive. Lastly, *S. cerevisiae* contains a Rad52 homolog, Rad59 [41]. Inspection of the Rad59 sequence by PSORT II and Predict NLS did not reveal any NLS (data not shown). Interestingly, it has previously been observed that Rad59 fails to localize to the nucleus in the absence of Rad52 [42], and since Rad52 and Rad59 interacts physically [10] it is likely that Rad52 and Rad59 are transported to the nucleus as a complex.

Acknowledgements

This work was supported by the Danish Research Council for Technology and Production Sciences (U.H.M.), The Alfred Benzon Foundation (U.H.M.), The Hartmann Foundation (U.H.M.), The Technical University of Denmark (S.C.L.H.), The Danish Natural Science Research Council (M.L.), The Villum Kann Rasmussen Foundation (M.L.), and NIH grants ES07061 (P.S.), GM50237 and GM67055 (R.R.). We thank Elvira Chapka for excellent technical help and members of the Mortensen laboratory for comments on this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.dnarep.2007.07.016.

REFERENCES

- [1] L.S. Symington, Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair, *Microbiol. Mol. Biol. Rev.* 66 (2002) 630–670, table.
- [2] U.H. Mortensen, C. Bendixen, I. Sunjevaric, R. Rothstein, DNA strand annealing is promoted by the yeast Rad52 protein, *Proc. Natl. Acad. Sci. U.S.A.* 93 (1996) 10729–10734.
- [3] T. Sugiyama, J.H. New, S.C. Kowalczykowski, DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA, *Proc. Natl. Acad. Sci. U.S.A.* 95 (1998) 6049–6054.
- [4] W. Kagawa, H. Kurumizaka, S. Ikawa, S. Yokoyama, T. Shibata, Homologous pairing promoted by the human Rad52 protein, *J. Biol. Chem.* 276 (2001) 35201–35208.
- [5] P. Sung, Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase, *J. Biol. Chem.* 272 (1997) 28194–28197.
- [6] J.H. New, T. Sugiyama, E. Zaitseva, S.C. Kowalczykowski, Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A, *Nature* 391 (1998) 407–410.
- [7] Z. Shen, S.R. Peterson, J.C. Comeaux, D. Zastrow, R.K. Moyzis, E.M. Bradbury, et al., Self-association of human RAD52 protein, *Mutat. Res.* 364 (1996) 81–89.
- [8] A. Shinohara, M. Shinohara, T. Ohta, S. Matsuda, T. Ogawa, Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing, *Genes Cells* 3 (1998) 145–156.

- [9] D.E. Van, N.M. Hajibagheri, A. Stasiak, S.C. West, Visualisation of human rad52 protein and its complexes with hRad51 and DNA, *J. Mol. Biol.* 284 (1998) 1027-1038.
- [10] A.P. Davis, L.S. Symington, The yeast recombinational repair protein Rad59 interacts with Rad52 and stimulates single-strand annealing, *Genetics* 159 (2001) 515-525.
- [11] F. Cortes-Ledesma, F. Malagon, A. Aguilera, A novel yeast mutation, rad52-L89F, causes a specific defect in Rad51-independent recombination that correlates with a reduced ability of Rad52-L89F to interact with Rad59, *Genetics* 168 (2004) 553-557.
- [12] W. Ranatunga, D. Jackson, J.A. Lloyd, A.L. Forget, K.L. Knight, G.E. Borgstahl, Human RAD52 exhibits two modes of self-association, *J. Biol. Chem.* 276 (2001) 15876-15880.
- [13] G.T. Milne, D.T. Weaver, Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52, *Genes Dev.* 7 (1993) 1755-1765.
- [14] L. Krejci, B. Song, W. Bussen, R. Rothstein, U.H. Mortensen, P. Sung, Interaction with Rad51 is indispensable for recombination mediator function of Rad52, *J. Biol. Chem.* 277 (2002) 40132-40141.
- [15] M.S. Park, D.L. Ludwig, E. Stigger, S.H. Lee, Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells, *J. Biol. Chem.* 271 (1996) 18996-19000.
- [16] D. Gorlich, I.W. Mattaj, Nucleocytoplasmic transport, *Science* 271 (1996) 1513-1518.
- [17] T. Bouliskas, Nuclear import of DNA repair proteins, *Anticancer Res.* 17 (1997) 843-863.
- [18] F.G.H.J. Sherman F., *Methods in Yeast Genetics*, Cold Spring Harbor Lab. Press, 1986.
- [19] B.J. Thomas, R. Rothstein, Elevated recombination rates in transcriptionally active DNA, *Cell* 56 (1989) 619-630.
- [20] H.Y. Fan, K.K. Cheng, H.L. Klein, Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant hpr1 delta of *Saccharomyces cerevisiae*, *Genetics* 142 (1996) 749-759.
- [21] H. Zou, R. Rothstein, Holliday junctions accumulate in replication mutants via a RecA homolog-independent mechanism, *Cell* 90 (1997) 87-96.
- [22] N. Erdeniz, U.H. Mortensen, R. Rothstein, Cloning-free PCR-based allele replacement methods, *Genome Res.* 7 (1997) 1174-1183.
- [23] M. Lisby, R. Rothstein, U.H. Mortensen, Rad52 forms DNA repair and recombination centers during S phase, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 8276-8282.
- [24] Q. Feng, L. Durning, A.A. de Mayolo, G. Lettier, M. Lisby, N. Erdeniz, et al., Rad52 and Rad59 exhibit both overlapping and distinct functions, *DNA Repair (Amst)* 6 (2007) 27-37.
- [25] U.H. Mortensen, N. Erdeniz, Q. Feng, R. Rothstein, A molecular genetic dissection of the evolutionarily conserved N terminus of yeast Rad52, *Genetics* 161 (2002) 549-562.
- [26] D. Kalderon, W.D. Richardson, A.F. Markham, A.E. Smith, Sequence requirements for nuclear location of simian virus 40 large-T antigen, *Nature* 311 (1984) 33-38.
- [27] D. Kalderon, A.E. Smith, In vitro mutagenesis of a putative DNA binding domain of SV40 large-T, *Virology* 139 (1984) 109-137.
- [28] L. Mikkelsen, S. Sarrocco, M. Lubeck, D.F. Jensen, Expression of the red fluorescent protein DsRed-Express in filamentous ascomycete fungi, *FEMS Microbiol. Lett.* 223 (2003) 135-139.
- [29] L. Prakash, S. Prakash, Isolation and characterization of MMS-sensitive mutants of *Saccharomyces cerevisiae*, *Genetics* 86 (1977) 33-55.
- [30] M. Lisby, M.A. ntunez de, U.H. Mortensen, R. Rothstein, Cell cycle-regulated centers of DNA double-strand break repair, *Cell Cycle* 2 (2003) 479-483.
- [31] E.N. Asleson, R.J. Okagaki, D.M. Livingston, A core activity associated with the N terminus of the yeast RAD52 protein is revealed by RAD51 overexpression suppression of C-terminal rad52 truncation alleles, *Genetics* 153 (1999) 681-692.
- [32] G.R. Hicks, N.V. Raikhel, Protein import into the nucleus: an integrated view, *Annu. Rev. Cell Dev. Biol.* 11 (1995) 155-188.
- [33] T. Bouliskas, Nuclear localization signals (NLS), *Crit. Rev. Eukaryot. Gene Expr.* 3 (1993) 193-227.
- [34] M. Cokol, R. Nair, B. Rost, Finding nuclear localization signals, *EMBO Rep.* 1 (2000) 411-415.
- [35] E.C. Lacasse, Y.A. Lefebvre, Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins, *Nucleic Acids Res.* 23 (1995) 1647-1656.
- [36] A.Z. Stasiak, E. Larquet, A. Stasiak, S. Muller, A. Engel, E. Van Dyck, et al., The human Rad52 protein exists as a heptameric ring, *Curr. Biol.* 10 (2000) 337-340.
- [37] T. Kummalue, J. Lou, A.D. Friedman, Multimerization via its myosin domain facilitates nuclear localization and inhibition of core binding factor (CBF) activities by the CBFbeta-smooth muscle myosin heavy chain myeloid leukemia oncoprotein, *Mol. Cell Biol.* 22 (2002) 8278-8291.
- [38] W. Kagawa, H. Kurumizaka, R. Ishitani, S. Fukai, O. Nureki, T. Shibata, et al., Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form, *Mol. Cell* 10 (2002) 359-371.
- [39] M.R. Singleton, L.M. Wentzell, Y. Liu, S.C. West, D.B. Wigley, Structure of the single-strand annealing domain of human RAD52 protein, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 13492-13497.
- [40] A. Shinohara, T. Ogawa, Stimulation by Rad52 of yeast Rad51-mediated recombination, *Nature* 391 (1998) 404-407.
- [41] Y. Bai, L.S. Symington, A Rad52 homolog is required for RAD51-independent mitotic recombination in *Saccharomyces cerevisiae*, *Genes Dev.* 10 (1996) 2025-2037.
- [42] M. Lisby, J.H. Barlow, R.C. Burgess, R. Rothstein, Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins, *Cell* 118 (2004) 699-713.

5.1. Transport mutants of ScRad52 as scaffoldin candidates

The work published in Plate et al., 2008a demonstrates that ScRad52 possesses several properties that makes it useful as a nano-platform scaffoldin. Most importantly, ScRad52 forms ring-structures that assemble in the cytosol of *S. cerevisiae*. Secondly, these ring-structures remain in the cytosol, if the nuclear localization signal (NLS) is removed and thirdly, ScRad52 seems to be mobile, as the protein can be redirected to the nucleus if it is tagged with the consensus NLS of SV40 virus. However, we also observe that when nuclear transport mutants are co-expressed with the wild type protein, both species sort to the nucleus. Accordingly, if a nuclear transport mutant of ScRad52 is used as a scaffoldin in a *RAD52* strain, the platform will likely sort to the nucleus and not the cytoplasm. Therefore, experiments with a ScRad52 based nano-platform would have to be carried out in a *rad52Δ* strain in order to ensure cytosolic localization of the nano-platform. Unfortunately, a *rad52Δ* strain is likely to be a poor cell factory, as *rad52Δ* strains display growth defects and elevated mutation rates (Petes et al., 1991). As a consequence, I judged that nuclear transport mutants of ScRad52 are not suitable nano-platform scaffoldins.

5.2. MmRad52 as a scaffoldin candidate for enzyme assembly in *S. cerevisiae*

In the analysis of ScRad52's transport mechanism, it was also investigated whether the mechanism of transport was conserved between species. Interestingly it was observed that the transport mechanism seems to be conserved between yeasts but not between yeast and higher eukaryotes, as Rad52 of *Kluyveromyces lactis* sorted to the nucleus in a *rad52Δ* strain of *S. cerevisiae* whereas MmRad52 did not (Plate et al., 2008a). The fact that MmRad52 does not sort to the nucleus in *S. cerevisiae*, renders the possibility that MmRad52 could be applied as a scaffoldin in yeast. However, given that missorting species of ScRad52 are piggybacked to the nucleus by sorting species, MmRad52 could be piggybacked to the nucleus by ScRad52 in a similar way. To test this possibility, MmRad52-YFP was expressed in an *S. cerevisiae* strain expressing ScRad52-CFP from the genome. As a reference, the nuclear transport mutant ScRad52-R234A-YFP was also expressed in the same *S. cerevisiae* strain. Co-localization in the nucleus was observed for the nuclear transport mutant derived from ScRad52, but not for MmRad52, indicating that MmRad52 and ScRad52 do not interact in *S. cerevisiae* (Fig. 5-1). This prompted a further evaluation MmRad52's ability to serve as scaffoldin for assembly of enzymes.

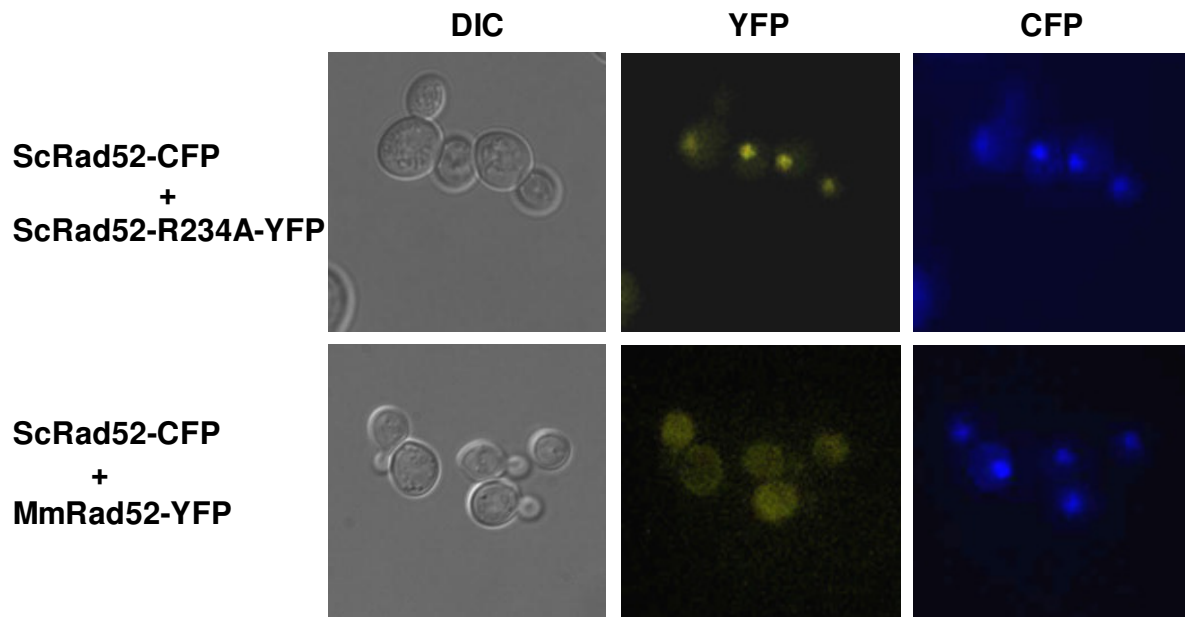


Figure 5.1. Co-expression of ScRad52 and either the nuclear transport mutant ScRad52-R234A-YFP or MmRad52.

Given the model described in Plate et al., 2008a, there could be two reasons for MmRad52 not sorting to the nucleus in *S. cerevisiae*. Either the NLS ensuring transport of MmRad52 in its native host is not recognized by *S. cerevisiae* or MmRad52 does not self-associate in *S. cerevisiae*. There is some evidence against both of these theories. An argument against the theory of a non-recognised NLS is that MmRad52 is predicted to contain an NLS consisting of 4 positively charged amino acid residues in its C-terminal. Such an NLS may be categorized as a typical consensus NLS (Boulikas, 1996) and should therefore be functional in *S. cerevisiae*. An argument against the theory that MmRad52 does not self-associate in *S. cerevisiae* is that MmRad52 has been demonstrated to self-associate with the yeast two hybrid assay (Krejci et al., 2000). The validity of this result can however be questioned because the number of false positives identified by the yeast two hybrid assay has been estimated to be as high as 50% for heterologous proteins (Deane et al., 2002).

As none of the above options could clearly be ruled out and as the ability to self-associate is an essential property of the scaffoldin, it was of interest to demonstrate that MmRad52 self-associates in *S. cerevisiae*. For this study a C-terminal truncation consisting of only the first 212 amino acid residues of MmRad52 was used. The crystal structure has been solved for this exact truncation (Kagawa et al., 2002) and it appears that this N-terminal fragment of the protein is sufficient for ring-structure formation.

To investigate whether MmRad52- Δ 212 self-associates in *S. cerevisiae*, it was exploited that molecules may be actively transported to the nucleus in two ways. They may contain an NLS within their own peptide sequence or they may interact with a molecule that contains an NLS (Boulikas, 1997). Based on this, an assay for determining whether MmRad52 self-associates in the cytosol of yeast was developed (Fig. 5-2). First, MmRad52 is tagged with either YFP or CFP and a consensus NLS. When these fusion proteins are expressed alone, the protein tagged with only YFP will reside in the cytosol, whereas the protein that has been tagged with CFP-NLS will sort to the nucleus. However, in the case of co-expression, the proteins localization will depend on whether they interact or not. If the proteins interact both proteins are expected to sort to the nucleus, whereas only the NLS-bearing molecule is expected to sort to the nucleus if no interaction occurs (Fig. 5-2).

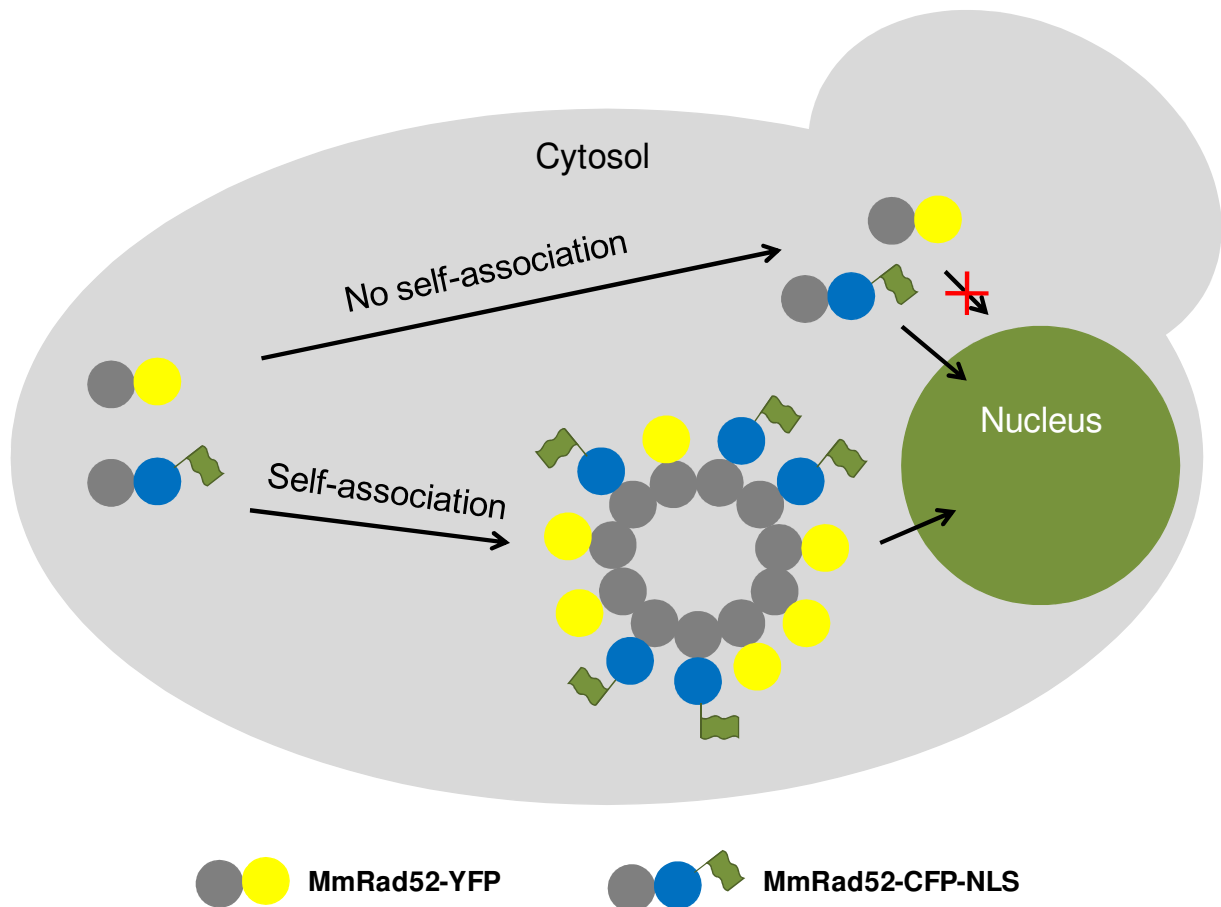


Figure 5-2. Principle of nuclear transport assay. Both the YFP and the CFP bearing molecule will localize to the nucleus if MmRad52 self-associates, whereas only the CFP bearing molecule will localize to the nucleus if no self-association occurs.

Initially, the localization of the separately expressed truncated species of MmRad52 was demonstrated to be as expected. When expressed alone MmRad52- Δ 212 tagged with YFP resided in the cytosol no matter whether the protein was tagged C- or N-terminally, whereas tagging of MmRad52- Δ 212 with the strong consensus NLS of SV40 virus and CFP ensured nuclear localization of this protein (data not shown).

But when co-expressed, both the nuclear sorting molecule MmRad52- Δ 212-CFP-NLS and the otherwise cytosolic proteins MmRad52- Δ 212-YFP and YFP-MmRad52- Δ 212 localized to the nucleus (Fig. 5-3). The co-localization suggests that MmRad52- Δ 212 interacts in the cytosol of *S. cerevisiae* and forms heteromeric ring-structures containing both the CFP and the YFP tagged

MmRad52 species. The fact that the co-localization was observed both when MmRad52- Δ 212 was tagged C- and N-terminally with YFP indicates that neither N-terminal nor C-terminal tagging interferes with the ability to form ring-structures.

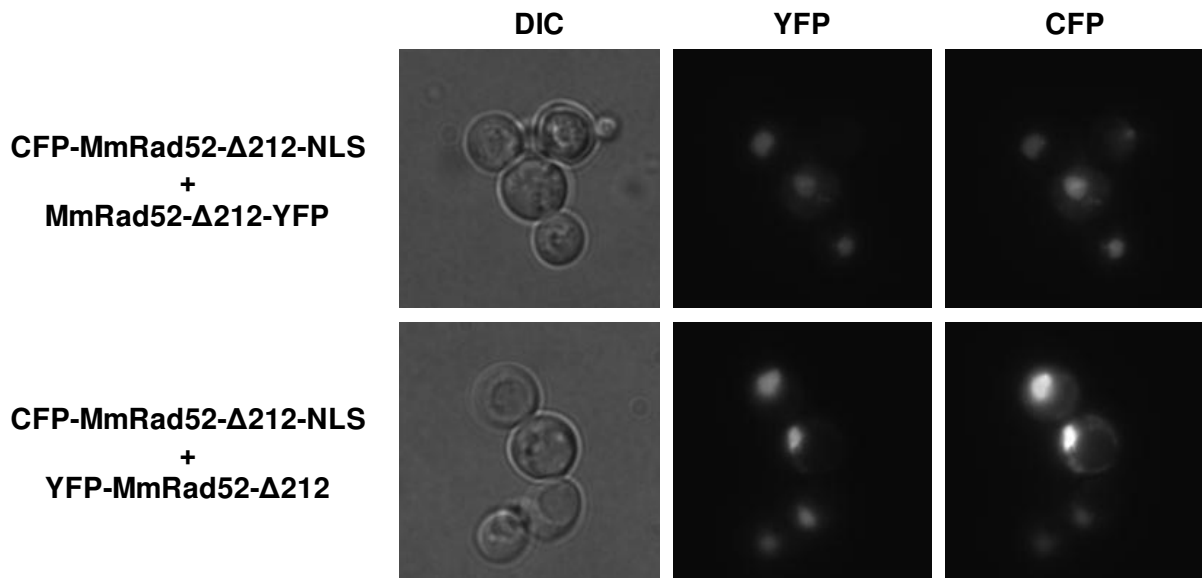


Figure 5-3. Co-expression of the nuclear sorting molecule, MmRad52-CFP-NLS with the otherwise cytosolic proteins MmRad52-YFP and YFP-MmRad52. Both the NLS bearing and the NLS lacking molecule co-localizes in the nucleus.

In summary, an assay that possibly can be generally applied to determine protein interactions *in vivo* was developed. Using this nuclear transport assay, it was demonstrated that MmRad52- Δ 212 self-associates in the cytoplasm of yeast. The result suggests that MmRad52 can be applied as a scaffoldin for assembly of at least two different enzymes, here exemplified by YFP and CFP. In the last chapter, it will be tested whether cell factories can benefit from enzyme assembly on an MmRad52-based nano-platform.

5.3. Materials and Methods

5.3.1. Strains and plasmids

The genotype and the source of the strains and plasmids used in this study are given in Table 5-1. To construct a plasmid expressing MmRad52- Δ 212-YFP, MmRad52- Δ 212 was PCR amplified with the primers M1 and M2, and YFP was PCR amplified with G1 and Y4. Subsequently, the two fragments were fused in a second round of PCR with M1 and Y4. Fragments encoding YFP-MmRad52- Δ 212 and CFP-MmRad52- Δ 212-NLS were generated in a similar way except the primers were: G5, G6, M7 and M8 or C11, G6, M7 and M12. The fragments encoding MmRad52- Δ 212-YFP and YFP-MmRad52- Δ 212 were digested with *ClaI* and *SacI* and inserted into a *ClaI-SacI* vector fragment of pESC-his. The fragment encoding CFP-MmRad52- Δ 212-NLS was inserted into an *EagI-BglIII* vector fragment of pESC-ura.

The sequence of all the primers used in this study is given in Table 5-2.

Table 5-1. Strains and plasmids used in this study.

Strains	Genotype	Source
W3849-10C	MATa bar1::LEU2 his3-11,15 ura3-1 RAD52-CFP	U. Mortensen strain collection
CEN.PK113-11C	MATa MAL2-8c SUC2 ura3-52 his3 Δ 1	Peter Kötter ^a
Plasmids	Genotype	Source
pIPL3	<i>CEN HIS3 P_{RAD52}MmRad52-YFP</i>	Plate et al., 2008a
pWJ1213-rad52-R234A	<i>CEN HIS3 P_{RAD52}Rad52-R234A-YFP</i>	Plate et al., 2008a
pESC-ura	<i>2μ URA3</i>	Stratagene
pF1-2	<i>2μ URA3 P_{GAL10}MmRad52-Δ212-YFP</i>	This study
pF2-2	<i>2μ URA3 P_{GAL10}YFP-MmRad52-Δ212</i>	This study
pESC-his	<i>2μ HIS3</i>	Stratagene
pF4-8	<i>2μ HIS3 P_{GAL10}CFP-MmRad52-Δ212-NLS</i>	This study

^aInstitut für Mikrobiologie, der Johan Wolfgang Goethe-Universität, Frankfurt am Main, Germany.

Table 5-2. Oligonucleotides used in this study.

Primer name	Sequence ^a	Template
M1	gtatcgtagcatcgatATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
M2	cagtgaaaagttcttctcttactcatcccATTCTGTCGGCAGCTGTTGT	<i>MmRad52</i>
M7	cacatggcatggatgaactatacaaggATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
M8	ggagcatcggggagctcctaATTCTGTCGGCAGCTGTTGT	<i>MmRad52</i>
M12	ggagcatcgggagatctctactcgacttccgctttttctcggagatgcATTCTGTCGGCAGCTGTTGT	<i>MmRad52</i>
G1	ATGAGTAAAGGAGAAGAAGACTTTTCACTG	<i>YFP/CFP</i>
Y4	ggagcatcggggagctcttaTTTGTATAGTTCATCCATGCCATGTGTAATCCC	<i>YFP/CFP</i>
G5	gtatcgtagcatcgatATGAGTAAAGGAGAAGAAGACTTTTCACTG	<i>YFP/CFP</i>
G6	TTTGTATAGTTCATCCATGCCATGTGTAATCCC	<i>YFP/CFP</i>
C11	gtatcgtagccggccgATGAGTAAAGGAGAAGAAGACTTTTCACTG	<i>YFP/CFP</i>

^aThe parts of the nucleotide sequences shown in capitals are those that specifically anneal to the template DNA. The parts of the oligonucleotides that contain restriction sites and adaptamers are shown in lower-case.

5.3.2. Co-expression of different Rad52 species in *S. cerevisiae*

To investigate whether MmRad52 and ScRad52 interacts in *S. cerevisiae*, MmRad52-YFP was expressed in a strain expressing ScRad52-CFP. This was done by transforming and expressing the plasmid pIPL3 in the yeast strain (w3849-10C).

To investigate whether MmRad52- Δ 212 self-associates in *S. cerevisiae*, an NLS tagged MmRad52 species was co-expressed with two different cytosolic species of MmRad52. This was done by transforming CEN.PK113-11C with pF4-8 and either pF1-2 or pF2-1. In this way strains expressing either CFP-MmRad52- Δ 212-NLS and MmRad52- Δ 212-YFP or CFP-MmRad52- Δ 212-NLS and YFP-MmRad52- Δ 212 were constructed.

5.3.3. Fluorescence microscopy

Fluorescence microscopy was carried out as described in Albertsen et al., 2009.

Chapter 6: Effect of assembling metabolic enzymes on the nano-platform

6.1. Introduction

The cell may employ several different strategies for positioning sequentially acting enzymes in close proximity, as enzymes may sort to the same organelle, associate to form multifunctional complexes or dock on cellular structures such as the cytoskeleton or internal membranes. This spatial organization of enzymes has important implications for the efficiency, specificity and regulation of metabolic pathways (For reviews see Conrado et al., 2008, Ovadi and Srere, 2000, Winkel, 2004). Accordingly, it is of great interest to mimic nature's way of organizing enzymes, in order to efficiently optimize novel metabolic pathways in cell factories. In Chapter 2 and 3, it was demonstrated that two novel pathways could be optimized by fusion of two sets of sequentially acting enzymes. As an alternative to enzyme fusion, catalytic sites can be brought in close proximity by assembling them on a scaffold. As early as the 1970's, it was shown that reaction rates can be increased approximately two-fold by cross-linking consecutive enzymes to polymer-based scaffolds (Mosbach and Mattiasson, 1970, Srere et al., 1973). This technique can however not be easily applied *in vivo*. More recently a technique that seems easy to adapt for *in vivo* applications has been developed. This approach is inspired by nature's way of organizing cellulases on their substrate cellulose. To ensure efficient targeting of cellulases to their substrate, simultaneously acting cellulases are assembled via docking domains on a scaffoldin that also contains a cellulose binding domain (See also section 1.4.1.3.). In this way all the enzymes required for cellulose breakdown are positioned in close proximity near their substrate. Inspired by this fascinating system for enzyme assembly, chimeric scaffoldins containing cohesin domains from different organisms has been generated (Fierobe et al., 2002, Fierobe et al., 2005). When these chimeric scaffoldins were mixed *in vitro* with cellulases containing different docking domains that specifically recognise a particular cohesin, it has been demonstrated that the cellulases can be incorporated at specific positions within the scaffoldin (Fierobe et al., 2005). Moreover, substrate breakdown could be accelerated when simultaneously acting cellulases from different species were assembled on the chimeric scaffoldin (Mingardon et al., 2007a). Inspired by these results, we decided to test whether scaffoldin mediated assembly of sequentially acting enzymes could be used to optimize two novel pathways *in vivo*.

In this study, it was chosen to use Rad52 as a scaffoldin for assembly of metabolic enzymes. This protein holds great potential for enzyme assembly because it is capable of forming a remarkably stable heptameric ring-structure (Ranatunga et al., 2001a, Stasiak et al., 2000). If metabolic enzymes are fused to such a multimeric protein, the multimeric protein's ability to self-associate can be used to guide the assembly of several metabolic enzymes on a ring-shaped platform (See Fig. 4-1). In this study, a structure consisting of metabolic enzymes linked to a ring-shaped platform is termed a nano-platform.

Rad52 is a well-characterized protein due to its important role in DNA repair. As a consequence, several of the protein's features have been investigated in detail. In the context of the nano-platform, it is important to note that the domain responsible for ring-structure formation has been mapped to the N-terminal, more precisely amino acid residues 65-165 for HsRad52 (Shen et al., 1996). Moreover, the crystal-structure has been solved for the C-terminal truncation of HsRad52 (HsRad52- Δ 212) (Kagawa et al., 2002). The crystal structure revealed that this truncation of Rad52 forms a ring-structure with 11 subunits in the ring. As it is of interest to use a scaffoldin of minimal size for enzyme assembly, it was chosen to use a truncated version of Rad52 to form the core of the nano-platform. More specifically, MmRad52- Δ 212 rather than HsRad52 was chosen. Previously, we have worked with MmRad52 and shown that this protein is soluble and localizes to the cytosol when it is expressed in *S. cerevisiae* (Plate et al., 2008a). Moreover, as MmRad52- Δ 212 is 90% similar to HsRad52- Δ 212, MmRad52 is likely to form a structure similar to the one elucidated for HsRad52 and generally possess similar features.

To investigate whether metabolic pathways could be optimized by enzyme assembly on a MmRad52-based nano-platform in the cytosol of *S. cerevisiae*, two model pathways were chosen. One that leads to sesquiterpene production and one that leads to vanillin glucoside production in yeast. In Chapter 2 and 3, it was demonstrated that these two pathways could benefit from having catalytic sites positioned in close proximity by enzyme fusion. Accordingly, they both provide excellent model pathways for testing whether pathways can benefit from scaffoldin mediated enzyme assembly.

To determine whether the sesquiterpene pathway can benefit from MmRad52 guided enzyme assembly, the effect of attaching FPPS and PTS to MmRad52- Δ 212 was examined by evaluating the functionality, the solubility and the patchoulol production from the nano-platform linked enzymes when they were expressed in *S. cerevisiae*. Next, it was investigated whether vanillin glucoside production could be optimized in yeast by positioning OMT and UGT on the nano-platform. Here, the effect was evaluated by measuring growth rate, selected intermediate concentrations and the final concentrations of the end-product, vanillin glucoside.

6.2. Materials and methods

6.2.1. Strains and media

The genotype of all the strains used in this study is given in Table 6-1. All media for genetic manipulations of yeast were prepared as described by (Sherman et al., 1986) with minor modifications as the synthetic medium contained twice the amount of leucine (60mg/L). A media allowing selection of hygromycin resistant colonies was prepared by adding hygromycin B to YPD to a final concentration of 300mg/L media.

To construct *erg20* Δ strains expressing free FPPS, MmRad52-FPPS or FPPS-MmRad52, CEN.PK113-5D were transformed with the plasmids pLA001, pLA030 and pLA033. The resulting transformants were streak purified and retransformed with fragments designed to knock out *ERG20*. Fragments were prepared as described in Albertsen et al., 2009.

Table 6-1. Yeast strains used in this study.

Strain	Genotype	Plasmid supplementation	Source
CEN.PK113-5D	MATa MAL2-8 ^c SUC2 ura3-52		Peter Kötter ^a
CEN.LA101	MATa erg20::hphMX MAL2-8 ^c SUC2 ura3-52	pLA001	This study
CEN.LA130	MATa erg20::hphMX MAL2-8 ^c SUC2 ura3-52	pLA030	This study
CEN.LA133	MATa erg20::hphMX MAL2-8 ^c SUC2 ura3-52	pLA033	This study
VAN265JH	Mata, adh6::LEU2, bgl1::KanMX4, his3Δ1, leu2Δ0, met15Δ, ura3Δ0, P _{TPI1} ::3DSD- [AurC]::ACAR [HphMX]		This study (See section 3.2.1.)

^aInstitut für Mikrobiologie, der Johan Wolfgang Goethe-Universität, Frankfurt am Main, Germany.

6.2.2. Plasmid construction

The name and genotype of all plasmids used in this study are given in Table 6-2. The sequences of all oligonucleotides used for constructing plasmids are given in Table 6-3.

FPPS and PTS constructs

To fuse PTS to MmRad52-Δ212, fragments encoding these enzymes were amplified separately using R3, R2, P1 and P4. Subsequently, the two fragments were fused in a second round of PCR. The resulting fusion PCR fragment was inserted into a *Bam*HI-*Xho*I fragment of pESC-ura. FPPS and MmRad52-Δ212 were fused to each other in the two possible configurations in a similar way and inserted into an *Eco*RI-*Cla*I vector fragment of the vector expressing MmRad52-Δ212-PTS.

To insert different linkers in between MmRad52-Δ212 and PTS, fragments encoding PTS attached to the different linkers were obtained by digesting vectors expressing FPPS and PTS fused by different linkers with *Bsp*EI and *Xho*I (See Albertsen et al., 2009). The Linker-PTS fragments were inserted into a *Bsp*EI-*Xho*I vector fragment of pLA033. To insert the gene encoding CFP in between MmRad52-Δ212 and PTS, a PCR fragment of CFP was amplified with G15 and G16 and inserted into a *Bsp*EI vector fragment of pLA033. To insert the gene encoding YFP in between the genes encoding MmRad52-Δ212 and FPPS, a PCR fragment of YFP was inserted into a *Bss*HIII fragment of the pLA033 derived vector expressing MmRad52-Δ212-FPPS and MmRad52-CFP-PTS.

OMT and UGT constructs

To construct plasmids expressing OMT and UGT fused to MmRad52- Δ 212 in all possible combinations, *OMT* and *UGT* were fused to *Mmrad52- Δ 212* using fusion PCR and the oligonucleotides given in Table 6-3. Generally, OMT encoding fragments were inserted into *Bam*HI-*Xho*I vector fragments of pESC-his, whereas UGT encoding fragments were inserted into *Cla*I-*Pac*I vector fragments of pESC-his derived vectors encoding OMT.

The correct sequence of all cloned plasmids was verified by sequencing at MWG-biotech.

Table 6-2. Plasmids used in this study

Name	Genotype ^a	Source
pWJ669	MmRad52	Rodney Rothstein ^a
pESC-ura	2 μ URA3	Stratagene
pIP029	2 μ URA3 P _{GAL1} -PatTps177	Asadollahi et al., 2008
pLA001	2 μ URA3 P _{GAL1} PatTps177 P _{GAL10} -ERG20	Albertsen et al., 2009
pLA030	2 μ URA3 P _{GAL10} ERG20-MmRad52- Δ 212 P _{GAL1} -MmRad52- Δ 212-SF-PatTps177	This study
pLA031	2 μ URA3 P _{GAL10} -ERG20 P _{GAL1} MmRad52- Δ 212- SF-PatTps177	This study
pLA032	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} PatTps177	This study
pLA033	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 212- SF-PatTps177	This study
pLA034	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 212- LF-PatTps177	This study
pLA035	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 212- SR-PatTps177	This study
pLA036	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 212- LR-PatTps177	This study
pLA037	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 212- S-PatTps177	This study
pLA038	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 222-PatTps177	This study
pLA039	2 μ URA3 P _{GAL10} MmRad52- Δ 212-ERG20 P _{GAL1} -MmRad52- Δ 212- CFP-PatTps177	This study
pLA040	2 μ URA3 P _{GAL10} MmRad52- Δ 212-YFP-ERG20 P _{GAL1} -MmRad52- Δ 222-CFP-PatTps177	This study
pLA041	URA3 P _{GAL10} MmRad52- Δ 212-YFP-ERG20 P _{GAL1} -MmRad52- Δ 222- CFP-PatTps177	This study
pJH589	CEN URA3 EntD	Hansen et al., 2009
pJH543	P _{TPI1} hsOMT NatMX	Hansen et al., 2009
pJH665	CEN URA3 P _{TPI1} UGT72E2	Hansen et al., 2009
pESC-his	2 μ HIS3	Stratagene
pLA050	2 μ HIS3 P _{GAL1} MmRad52- Δ 212-hsOMT P _{GAL10} MmRad52- Δ 212-UGT72E2	This study
pLA051	2 μ HIS3 P _{GAL1} MmRad52- Δ 212-hsOMT P _{GAL10} UGT72E2-MmRad52- Δ 212	This study
pLA052	2 μ HIS3 P _{GAL1} hsOMT-MmRad52- Δ 212 P _{GAL10} MmRad52- Δ 212-UGT72E2	This study
pLA053	2 μ HIS3 P _{GAL1} hsOMT-MmRad52- Δ 212 P _{GAL10} UGT72E2-MmRad52- Δ 212	This study

^aDepartment of Genetics and Development, Columbia University, USA.

Table 6-3. Oligonucleotides used in this study.

Primer name	Sequence ^a	Template
E1	caacagctgccgacagaatggcgcgctaATGGCTTCAGAAAAAGAAAT	<i>ERG20</i>
E2	gtacgtagtgatcgatCTATTTGCTTCTCTTGTAAC	<i>ERG20</i>
E4	gtatcgtagtcggccgATGGCTTCAGAAAAAGAAATTAGGAGAG	<i>ERG20</i>
E12	TTTACTTCTCTTGTAACCTTGTTCAAAAACGC	<i>ERG20</i>
O1	agcatgacttaggatccATGGGTGACACTAAGGAGC	<i>OMT</i>
O2	caacagctgccgacagaatgggtccggaATGGGTGACACTAAGGAGC	<i>OMT</i>
O3	tctatcatggctcgagTTATGGACCAGCTTCAGAACC	<i>OMT</i>
O5	ggactgtcacgtggtgccgggtccggaATGGGTGACACTAAGGAGC	<i>OMT</i>
P1	caacagctgccgacagaattccggaATGGAGTTGTATGCCCAAAG	<i>PatTps177</i>
P4	catgatgcgactcgagTTAATATGGAACAGGGTGAA	<i>PatTps177</i>
R1	gtatcgtagtgattcATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
R2	ATTCTGTCGGCAGCTGTTGT	<i>MmRad52</i>
R3	ctgtagctggtgatccATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
R4	gtatcgtagcatcgatATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
R5	gtatcgtagtcggccgATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
R11	caggttctgaagctggtccaggcgcggatATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
R12	ggactgtcacgtggtgccgggtccggaATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
R13	tctagtagcctcgagCTAATTCTGTTCGGCAGCTGTTG	<i>MmRad52</i>
R14	agtactaggcttaattaaCTAATTCTGTTCGGCAGCTGTTG	<i>MmRad52</i>
R17	cgttttgaacaagtttacaagagaagtaaaggcgcgctaATGGCTGGGCCTGAAGAAG	<i>MmRad52</i>
U2	caacagctgccgacagaatggagccggcATGCATATCACAAAACCACACG	<i>UGT</i>
U3	tgatagatccttaattaaCTAGGCACCACGTGACAAG	<i>UGT</i>
U4	ccagtagttgatcgatttcATGCATATCACAAAACCACACG	<i>UGT</i>
U6	GGCACCACGTGACAAGTCC	<i>UGT</i>
G15	gagcgactgtccggaATGAGTAAAGGAGAAGAACTTTTCACTG	<i>CFP</i>
G16	ctagtgcgttccggaTTTGTATAGTTCATCCATGCCATGTG	<i>CFP</i>

^aThe parts of the nucleotide sequences shown in capitals are those that specifically anneal to the template DNA. The parts of the oligonucleotides that contain restriction sites and adaptamers are shown in lower-case.

6.2.3. Shake flask experiments

Shake flask experiments with the sesquiterpene producing strains were carried out as described in Albertsen et al. 2009, whereas shake flask experiments with vanillin glucoside producing strains were carried out as described in section 3.3.3. In all cases final titers of intermediates or end-products are averages of triplicates \pm standard deviations.

6.2.4. Fluorescent microscopy

Microscopy was carried out as described in (Plate et al., 2008b).

6.3. Results

6.3.1. FPPS and PTS are functional *in vivo* when they are fused to the scaffoldin

To benefit from MmRad52 guided enzyme assembly, the metabolic enzymes have to remain functional when fused to MmRad52. To test whether FPPS remains functional when it is fused to MmRad52- Δ 212 in the two possible configurations: MmRad52- Δ 212-FPPS and FPPS-MmRad52- Δ 212, it was exploited that the gene encoding FPPS, *ERG20*, is essential for viability of *S. cerevisiae*. The functionality of FPPS can therefore be assessed by investigating whether these fusion proteins can sustain growth of a strain harbouring an *ERG20* deletion. To this end, a haploid *ERG20* strain was transformed with 4 different 2 μ -based plasmids. One was an “empty” vector that served as a negative control, one was a plasmid that expresses FPPS as a free enzyme from P_{GAL10} and the last two were vectors expressed the fusion proteins MmRad52- Δ 212-FPPS and FPPS-MmRad52- Δ 212 from P_{GAL10}. The resulting transformants were streak purified and retransformed with gene targeting fragments specifically designed to knock out *ERG20*. In agreement with *ERG20* being an essential gene, no transformants were obtained when the strain harbouring the empty plasmid was transformed with the *ERG20* knock out fragments (data not shown). In contrast, transformants were obtained for the three strains harbouring plasmids expressing FPPS fused to MmRad52- Δ 212 in the two possible configurations or FPPS as a free enzyme. Hence, it is demonstrated that FPPS can sustain growth of a Δ *erg20* strain both when it is linked to MmRad52- Δ 212 via its C- and N-terminal. It was however noted that the *erg20* Δ strains grew slower when they harboured the plasmids expressing MmRad52- Δ 212-FPPS or

FPPS-MmRad52- Δ 212 compared to when FPPS was expressed as a free enzyme suggesting that FPPS does not retain full activity when fused to the MmRad52- Δ 212 (Fig 6-1A).

Next, it was tested whether PTS remains active after fusion to MmRad52- Δ 212. As patchouliol is not naturally produced by *S. cerevisiae*, patchouliol production was used as a measure to demonstrate that PTS is active. PTS can be fused to MmRad52- Δ 212 in two possible configurations, but only one of the possible configurations was tested here. Accordingly, a haploid *ERG20* strain was transformed with an “empty” plasmid (negative control) or plasmids expressing MmRad52- Δ 212-PTS or PTS where the latter serves as a positive control. To assess whether the expression of MmRad52- Δ 212-PTS resulted in patchouliol production, the three strains described above were cultured in shake flasks according to the method described in materials and methods (See section 6.2.3.). Patchouliol production was observed both for the strain expressing MmRad52- Δ 212-PTS and for the one expressing PTS as a free enzyme. Thus, demonstrating that PTS is functional when fused to MmRad52- Δ 212 via its N-terminal (Fig. 6-1B). As expected, no production of patchouliol was observed when the strain harboured an “empty” plasmid. Accordingly, PTS retains activity at least when fused to the scaffoldin via its N-terminal.

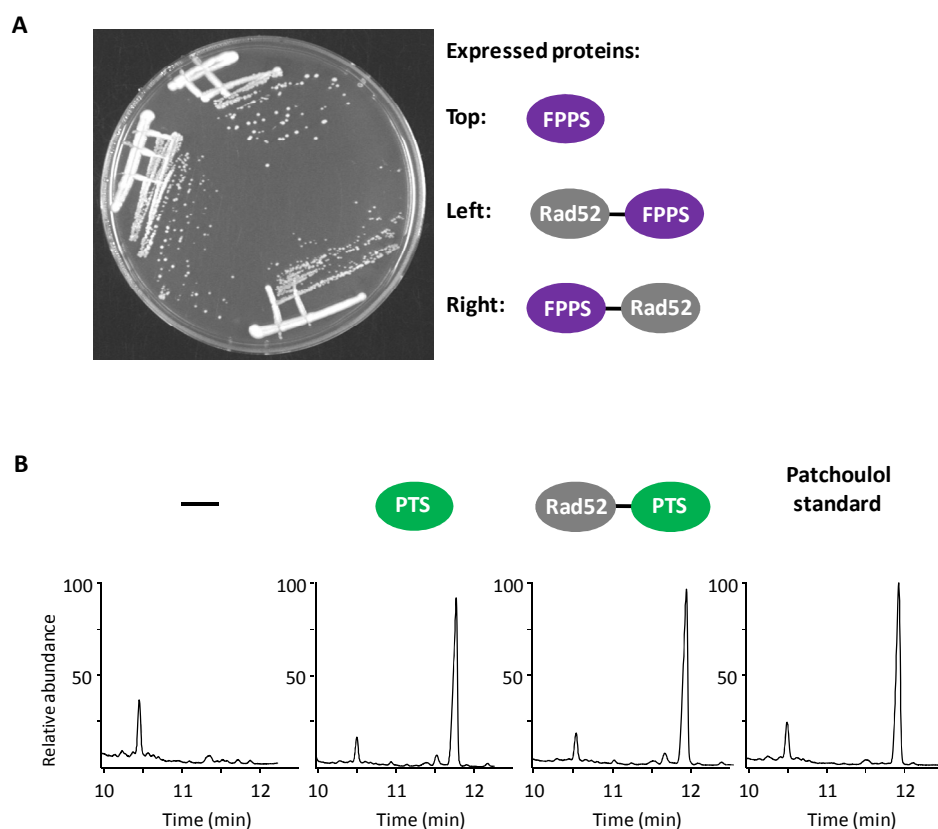






Figure 6-1. Functionality of FPPS and PTS when fused to MmRad52- Δ 212. A) Expression of FPPS (top), MmRad52- Δ 212-FPPS (left) and FPPS –MmRad52- Δ 212 (right) supports growth of a *erg20* Δ strain. B) Patchoulol is produced both when PTS is expressed as a free enzyme and when it is fused to MmRad52- Δ 212.

6.3.2. Attachment of both enzymes to the nano-platform results in an increased patchoulol production compared to when only one enzyme is attached

To investigate whether enzyme assembly of the two consecutive enzymes, FPPS and PTS on the MmRad52-based scaffoldin can increase the patchoulol production, *ERG20* strains expressing PTS and FPPS as free enzymes or fused to MmRad52- Δ 212 were grown in shake flasks until they reached stationary phase. To determine how fusion to MmRad52- Δ 212 affects the activity of PTS and FPPS, strains expressing either FPPS or PTS fused to MmRad52- Δ 212 and the other enzyme as a free enzyme was also tested for patchoulol production (see Table 6-4).

Table 6-4. Patchoulol (PT) production from *ERG20* strains expressing FPPS and PTS from 2 μ -based plasmids as individual enzymes or fused to MmRad52- Δ 212.

Proteins expressed ^a	Final PT titer [mg/L]	Final PT yield [mg/L/OD ₆₀₀]
 FPPS + PTS	5.1 \pm 0.3	0.190 \pm 0.022
 Rad52—FPPS + PTS	3.7 \pm 0.4	0.122 \pm 0.021
 Rad52—PTS + FPPS	0.7 \pm 0.1	0.024 \pm 0.004
 Rad52—FPPS + Rad52—PTS	1.2 \pm 0.1	0.044 \pm 0.001

^aFPPS and PTS containing proteins were always expressed from P_{GAL10} and P_{GAL1}, respectively. In all cases FPPS was also expressed endogenously.

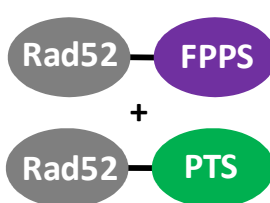
An 80% higher patchoulol production is observed when both FPPS and PTS are fused to MmRad52- Δ 212, compared to when only PTS is fused to MmRad52- Δ 212 (Table 6-4). The fusion of PTS to MmRad52- Δ 212 did however generally affect patchoulol production negatively, as a 7-fold lower patchoulol production was observed when PTS was fused to MmRad52- Δ 212 compared to when it was expressed as a free enzyme. This suggests that PTS is not fully active when fused to MmRad52- Δ 212. The effect of fusing FPPS to MmRad52- Δ 212 is harder to assess because all experiments were carried out in an *ERG20* strain which means that FPPS in all cases was also expressed endogenously as a free enzyme.

6.3.3. Effect of linker type on the functionality of the scaffoldin linked PTS

Since linker length and composition have been demonstrated to affect the specific activity (Carlsson et al., 1996), the catalytic efficiency (Lu and Feng, 2008), the stability (Prescott et al., 1999) and the folding rates (Robinson and Sauer, 1998) of fusion proteins, it was next investigated whether the sesquiterpene synthase activity of the fusion protein, MmRad52- Δ 212-PTS, could be increased by optimizing the linker between MmRad52- Δ 212 and PTS. In the experiment described in section 6.3.2., a short flexible linker consisting of Ser-Gly was used. To

further evaluate the effect of linker length and composition, 6 additional linkers were tested. One linker was in fact a 10 amino acid residue extension of MmRad52- Δ 212 (Rad52- Δ 222). The rest of the linkers were designed to be long flexible, short rigid, long rigid, stable or very long. The stable linker should be resistant towards degradation by proteases (Lu and Feng, 2008, Xue et al., 2004) and the very long linker was in fact the sequence CFP. The amino acid residue sequences of the rest of the linkers are given in Table 6-5.

Table 6-5: Effect of inserting linkers that varied in length and composition in between MmRad52- Δ 212 and PTS. The linker between MmRad52- Δ 212 and FPPS was in all cases a short linker (Gly-Ala-Leu).

Proteins expressed ^a	Linker (composition)	Effect of linker ^b PT titer _{linker} / PT titer _{short flexible linker}
	Short flexible (SG)	1
	Long flexible (SGGGGS)	0.99 ± 0.14
	Short rigid (SGEAAAK)	1.20 ± 0.03
	Long rigid (SGEAAAKEAAAK)	1.36 ± 0.33
	Stable (SGMGSSSN)	0.61 ± 0.17*
	Extended scaffoldin (EALGLPKPQESG)	0.69 ± 0.26
	Very long (Sequence of CFP)	0.37 ± 0.07*

^aAll subunits of the platform were expressed from 2 μ -based plasmids. In all cases FPPS and PTS containing proteins were expressed from P_{GAL10} and P_{GAL1}, respectively.

^bThe effect of a linker is given as patchoulol production using this linker relative to production achieved when the short flexible linker was used.

*Patchoulol production using this linker is significantly different (p<0.05) from when the short flexible linker was used.

No linkers capable of significantly increasing the patchoulol production beyond the short flexible linker were identified (Table 6-5). In contrast, two linkers that resulted in a significantly lower production of patchoulol compared to the short flexible linker were identified. These two linkers were the stable linker and the very long linker consisting of CFP. When the extended version of MmRad52- Δ 212 (MmRad52- Δ 222) was used as a scaffoldin no increases in production were achieved. Hence, truncating MmRad52 at another “random” position could not improve the performance of the scaffoldin linked PTS.

6.3.5. The platform forms aggregates when expressed from a multi copy plasmid in *S. cerevisiae*

To investigate whether the low production from the patchoulol producing nano-platform could be due to decreased expression levels, solubility or mislocalization of the nano-platform, the subcellular localization and the solubility of FPPS and PTS when they are fused to MmRad52- Δ 212 was determined. To visualize the platform subunits, a plasmid expressing MmRad52- Δ 212-YFP-FPPS and MmRad52- Δ 212-CFP-PTS was constructed (See Fig. 6-2A). The plasmid was constructed in two versions. One was a 2 μ -based plasmid that allows proteins to be highly expressed and the other was an integrative plasmid that allows expression of a single copy of the construct from the genome. According to our previous observations, the expressed proteins are expected to self-assemble and form a nano-platform in the cytosol of yeast that displays CFP, YFP, FPPS and PTS activity (See Fig. 6-2B).

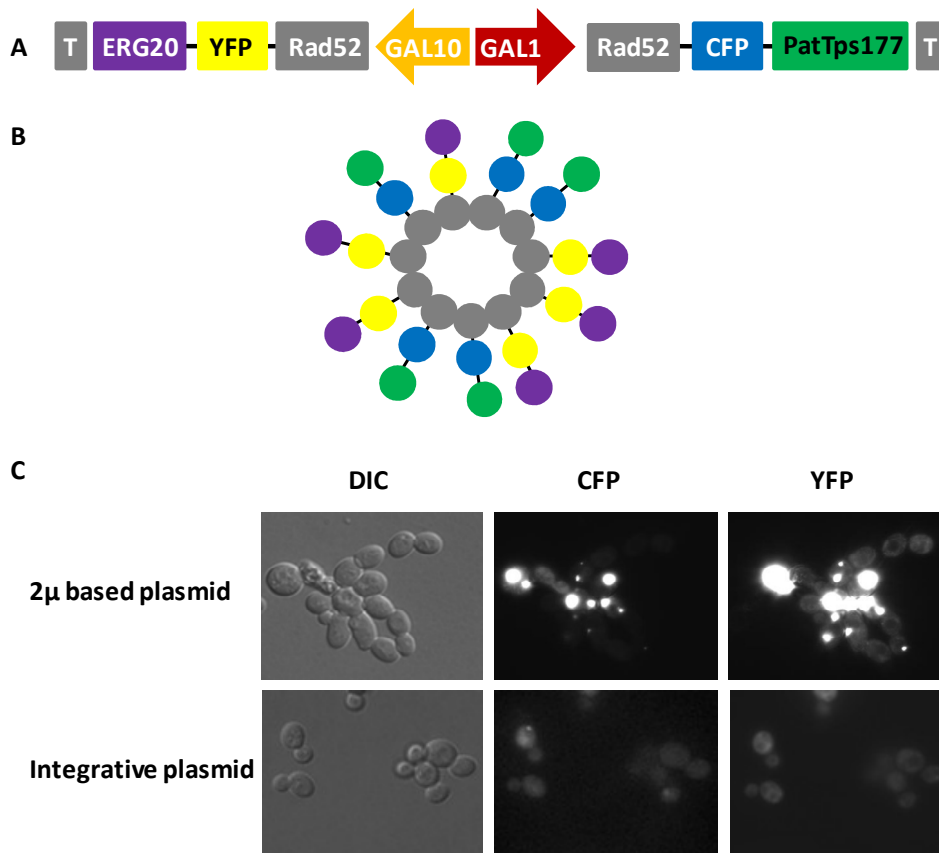


Figure 6-2. Fluorescent microscopy of the patchouliol producing nano-platform tagged with YFP and CFP. A) Construct used for fluorescent microscopy. B) Expected structure of the nano-platform once it has assembled. C) Fluorescent microscopy of strains that expressed the fluorescently tagged nano-platform from either a 2 μ -based plasmid or from a single copy on the genome.

When the platform subunits were expressed from a 2 μ -based plasmid, fluorescent aggregates displaying CFP and YFP activity were observed (Fig. 6-2C). The aggregates containing CFP and YFP often co-localize indicating that YFP and CFP containing subunits at least on some occasions are present in the same aggregate. To investigate whether the aggregate formation was related to over-expression, the nano-platform was also visualized when it was expressed from a single genomically integrated gene-cassette. When expression occurs from the genome, the expression level and the amounts of visible aggregates was severely reduced (Fig. 6-2C).

It is also noted that the localization of the platform in all cases is cytosolic and that an even distribution of fluorescent signal is observed when expression occurs from the genome (Fig. 6-2C). In contrast the expression levels vary greatly among individual cells when

expression occurs from the 2μ -based plasmid.

6.3.6. Aggregate formation is not related to the patchouliol producing platform

To investigate whether aggregate formation was related to the patchouliol producing nano-platform, the tendency to form aggregates was tested for a nano-platform with no metabolic enzymes linked to it. To this end, MmRad52- Δ 212 was tagged with YFP and expressed from P_{GALI} from a 2μ -based plasmid. To regulate the expression level, strains were grown in galactose media containing different amounts of glucose which is known to repress P_{GALI} .

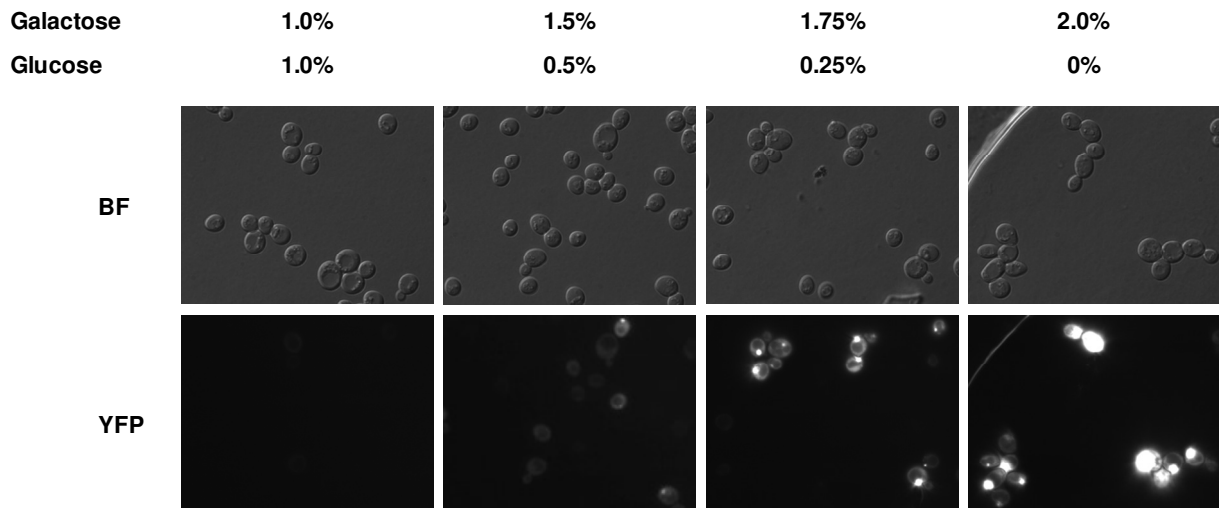


Figure 6-3. Expression of MmRad52- Δ 212-YFP from P_{GALI} , when strains were grown in media with decreasing glucose concentrations and increasing galactose concentrations.





Aggregates were also observed when no enzymes except YFP were linked to the MmRad52- Δ 212 (Fig. 6-3). The size of these aggregates increases with the expression level, and interestingly aggregates were also observed even when the expression level was relatively low. The same tendencies were observed when the expression of MmRad52-YFP was varied using the regulatable promoters CUP1 and MET25 (data not shown). It was also investigated whether aggregate formation is an effect of MmRad52- Δ 212 being a truncated protein. When we previously expressed full length MmRad52-YFP from the weak constitutive Rad52 promoter of *S. cerevisiae*, no aggregates were observed (Plate et al., 2008a). Accordingly, the observed aggregates could be due to the truncation. To rule out this possibility, MmRad52- Δ 212-YFP,

MmRad52-Δ222-YFP and MmRad52-YFP were all expressed from p_{GAL1} from a 2μ -based plasmid and subjected to fluorescence microscopy. Aggregates were observed for all three proteins and accordingly, the aggregation does not seem to be a consequence of the truncation (data not shown).

6.3.7. Decreasing the expression level provides no additional benefits of enzyme assembly on the platform

To investigate whether the low production of patchouliol from the assembled platform compared to the free enzymes is due to the nano-platform subunits tendency to form aggregates when it is highly expressed, the patchouliol production was determined when expression levels were decreased. In section 6.3.5., it was shown that expression levels can be substantially reduced by expressing the platform subunits from a single copy on the genome instead of from a 2μ -based plasmid. Accordingly integrative versions of the plasmids described in section 5.3.2. were constructed and the production of patchouliol was determined as previously described.

Table 6-6. Patchouliol (PT) production from *ERG20* strain expressing FPPS and PTS from the genome as individual enzymes or fused to the MmRad52-Δ212.

Proteins expressed ^a	Final PT titer	[mg/L]
	1.06 ± 0.07	
	0.94 ± 0.13	
	0.10 ± 0.01	
	0.16 ± 0.02	

Cells were grown in 500ml shake flasks containing 100ml Delft media with 20g/L galactose. Final titers are averages of triplicates ± standard deviations.






^a FPPS and PTS containing proteins were always expressed from P_{GAL10} and P_{GAL1} , respectively. In all cases FPPS was also expressed endogenously.

When expression levels were decreased by expressing the nano-platform subunits from the genome, a 60% increase in patchoulol production is observed when both FPPS and PTS are linked to MmRad52- Δ 212, compared to when only PTS is MmRad52- Δ 212 linked (Table 6-6). It is also noted that the production from the scaffoldin linked enzymes is substantially lower than when the enzymes are expressed as free enzymes from the same promoters.

6.3.8. Test of the nano-platform concept with the vanillin glucoside pathway as a model system

The results above suggest that both FPPS and PTS are not fully functional when they are fused to MmRad52- Δ 212 (Fig. 6-1A, Table 6-4 and 6-6). To test whether the substantially reduced production from the patchoulol producing nano-platform compared to the free enzymes is due to PTS and FPPS being particularly sensitive to being fused to MmRad52- Δ 212, another model system was tested. To test a second model system, the enzymes catalyzing the last two steps in the pathway leading to vanillin glucoside, namely OMT and UGT were linked to MmRad52- Δ 212. From the results presented in chapter 3, it is known that when these two enzymes are fused to each other, the configuration of the fusion enzymes can influence the functionality (see section 3.3.1). Therefore OMT and UGT were fused to MmRad52- Δ 212 in all the possible configurations, to test whether the vanillin glucoside pathway could benefit from having OMT and UGT positioned on a nano-platform. The constructs were transformed in a strain producing the PAL, the substrate for OMT (See Chapter 3 for introduction to the vanillin pathway).

Table 6-7: Growth rate and final titers of vanillin glucoside (VG) and vanillin (VAN) when strains express OMT and UGT as individual enzymes or fused to the scaffoldin.

Expressed proteins ^a	Growth rate	Final VG titer [mg/L]	Final VAN titer [mg/L]
	0.108 ± 0.003	41.0 ± 2,5	3.7 ± 0.5
	0.127 ± 0.009*	21.5 ± 3.6*	0.9 ± 0.8*
	0.123 ± 0.003*	26.5 ± 2.4*	1.3 ± 0.1*
	0.117 ± 0.005	7.5 ± 0.2*	2.1 ± 0.1*
	0.111 ± 0.003	9.9 ± 0.7*	2.0 ± 0.1*

^a OMT and UGT containing proteins were always expressed from PGAL1 and PGAL10, respectively.

*A significantly ($p < 0.05$) different growth rate or titer was observed for this strain compared to the strain expressing the free enzymes.

Generally, the final titer of VG was lower when the strains expressed OMT and UGT fused to MmRad52-Δ212 compared to when the enzymes were expressed as free enzymes from the same promoters. The VG production for the strains expressing OMT and UGT linked to MmRad52-Δ212 in the different configurations did however vary substantially and for the highest producing nano-platform the final titer of VG was only 35% lower than the production achieved with the free enzymes. It is also evident that OMT clearly prefers to be fused to MmRad52-Δ212 via its N-terminal whereas UGT prefers to be fused via its C-terminal.

Interestingly, a significantly ($p < 0.05$) lower vanillin accumulation was observed for all the strains expressing the nano-platforms compared to the strains expressing the free enzymes. Moreover, a significantly higher growth rate was observed for the two nano-platforms that produced the highest amount of VG.

6.4. Discussion

At the first glance, there appears to be no advantages of attaching FPPS and PTS to the MmRad52 based nano-platform, as the patchoulol production was 4-6 fold lower when the enzymes were attached to the nano-platform compared to when they were expressed as free enzymes (Table 6-4 and Table 6-6). However, by attaching one enzyme at a time to the nano-platform and express the other enzyme as a free enzyme, it is evident that PTS is affected negatively by being linked to MmRad52- Δ 212. Once this is taken into account when evaluating the effect of attaching enzymes to the nano-platform, a positive effect of linking FPPS and PTS to the MmRad52-based nano-platform is observed, as an up to 80% increase in patchoulol production was found when both enzymes were attached to the nano-platform compared to when only PTS was. This provides the first proof that benefits may be achieved through positioning consecutive enzymes in close proximity on the self-assembling scaffoldin such as the one tested here.

There could be at least two explanations for the decreased production from the nano-platform attached FPPS and PTS compared to their individual counterparts. First, the expression level could be lower when enzymes are fused to MmRad52- Δ 212 compared to when they are expressed as free enzymes. Visual assessment of the expression levels of fluorescently tagged enzymes that were either expressed as individual enzymes or fused to the MmRad52- Δ 212, indicates that differences in expression levels do not explain all of the observed difference as expression levels appear to be relatively similar (data not shown). Secondly, the enzymes may not retain full activity when they are fused to the scaffoldin. As already mentioned data suggest that PTS is less functional when it is fused to MmRad52- Δ 212. Furthermore, the activity of FPPS seems to be decreased as well, as *erg20* Δ strain expressing MmRad52- Δ 212-FPPS and FPPS-MmRad52- Δ 212 grew slower than the strain expressing FPPS as a free enzyme (Fig 6-1A and data not shown). As it has been shown that the activity of FPPS can be decreased to about 40% of wild type levels without affecting growth (Karst et al., 2004), the reduced growth of the *erg20* Δ strains observed when MmRad52-linked FPPS was highly expressed indicates that the activity of the scaffoldin linked enzyme is substantially decreased. The negative effect of fusing FPPS to MmRad52 does however not appear to be due to a general tagging problem of FPPS, as

a two-fold higher patchoulol production was observed when FPPS was fused to PTS via its C-terminal (Albertsen et al., 2009). Thus, indicating that in this case fusion does not impair the functionality of the enzyme. FPPS is however known to function as a dimer (Eberhardt and Rilling, 1975). When FPPS is linked to the nano-platform it is possibly more constrained in its movement than when it is fused to a monomeric protein such as PTS. Therefore, it is possible that fusion to a monomeric protein allows dimer formation and hence does not result in decreased activity whereas fusion to the nano-platform subunit does not allow dimer formation and this leads to a decreased activity of MmRad52 linked FPPS.

Tagging of MmRad52 with fluorescent proteins revealed that the protein has a strong tendency to form large aggregates particularly when it is highly expressed. Aggregate formation is often observed when heterologous genes are over-expressed in *E. coli* (Garcia-Fruitos et al., 2005) and although aggregates are less common in yeast, they have also been observed (Binder et al., 1991, Hartig et al., 1990, Kamasawa et al., 1999). Aggregates have long been believed to consist of misfolded proteins that are formed because over-expression of proteins overloads the folding machinery (Gonzalez-Montalban et al., 2007). The theory is that if proteins are over-expressed a large amount of nascent protein chains will exit the ribosomes at the same time (Hartl and Hayer-Hartl, 2002). Under these circumstances there are not sufficient amounts of chaperones to keep them separated till they are correctly folded and this results in aggregate formation because the nascent chains associate via hydrophobic interactions before they are completely folded. Recently, several studies have however observed that proteins in aggregates are biologically active (Ventura and Villaverde, 2006). In one case, the biological activity of the protein in aggregates was even increased compared to when the protein was not in aggregates (Garcia-Fruitos et al., 2005). It is unknown whether the aggregates observed in this study affect the functionality of the proteins that are fused to MmRad52. The fact that they fluoresce indicates that the at least the fluorophore part of the molecule is correctly folded. Moreover, the benefits of attaching enzymes to the nano-platform were actually less pronounced when the platform was expressed in low levels where aggregates are expected to be substantially reduced compared to when they were highly expressed and expected to be in aggregates (Compare table 6-4 and 6-6). Furthermore, when the nano-platform concept was tested with another set of enzymes, namely

OMT and UGT, the production from the best producing nano-platform was only 35% lower than when the free enzymes were expressed indicating that most of the proteins in the aggregates are active. In summary, nothing indicates that the reduced activity of the nano-platform linked enzymes is due to the observed aggregates.

But why are the aggregates formed? MmRad52 appears to be particularly aggregation prone, as aggregates are observed even when expression levels are low (see e.g. Fig. 6-3). The fact that MmRad52 possess the ability to self-associate renders the possibility that in the case of MmRad52 the aggregates are not due to hydrophobic interactions between nascent chains but rather they are a result of Rad52's ability to self-associate. It has been shown that the size of the Rad52 protein affects the number of subunits in the ring, as expression of the C-terminal truncation of HsRad52 results in the formation of ring-structures with 11 subunits in the ring (Kagawa et al., 2002, Singleton et al., 2002), whereas the expression the full length protein results in 7 subunits in the ring (Stasiak et al., 2000). Accordingly, extending the MmRad52 ORF with genes encoding other proteins, as it was done in this study, could affect the number of subunits in the ring. One possibility is that when the subunits become too big the ring is unable to close and this results in long polymer fragments that curl up on each other and form the observed aggregates.

Generally enzyme fusion was shown to result in greater benefits than MmRad52 mediated enzyme assembly, as enzyme fusion was shown to be superior with respect to all the parameters measured in this study for both of the tested systems (Compare results presented in chapter 2 and 3 with the result presented in this chapter). One of the advantages of assembling enzymes on the nano-platform compared to directly fusing them is the increased flexibility that the nano-platform system provides. For example several enzymes may be attached to the nano-platform, enzyme ratios may be optimized and the metabolic enzymes may be fused via their preferred terminal (see also section 4.2). In this study the flexibility of the nano-platform system was not fully exploited as relatively simple model systems consisting of only two enzymes were tested. The fact that production from the nano-platform was within the same range when OMT and

UGT were attached to the nano-platform (Table 6-7) is a promising result. Both the studies where OMT and UGT were fused to each other and when they were fused to MmRad52- Δ 212 revealed that these two enzymes preferred to be fused via a specific terminal. In the case of UGT the highest production was observed when it was fused via its C-terminal, whereas OMT preferred to be fused via its N-terminal. Accordingly, the proteins preferred tagging in different terminals and therefore a fusion protein where both proteins were tagged at their preferred terminal could be constructed. However, if the proteins had preferred to be fused via the same terminal, attachment to the nano-platform could possibly have been advantageous over the fusion strategy.

Considering that a first proof of concept was demonstrated here with two relatively simple model systems where several parameters can be optimized, the nano-platform concept or a further development of the concept should have great potential for optimization of metabolic pathways in cell factories.

Chapter 7: Conclusion

In order to develop new tools for optimization of cell factories, the effect of positioning sequentially acting enzymes in close proximity was studied *in vivo*. Two strategies for engineering catalytic sites to be in close proximity were tested. One strategy was to simply fuse the genes encoding the enzymes and in this way generate a bifunctional enzyme. The other strategy was to assemble metabolic enzymes on a nano-platform via fusion to a self-associating protein. Both strategies were tested with two model systems that each consisted of two sequentially acting enzymes. The model systems were chosen because they for different reasons were particularly likely to benefit from enzyme proximity. The first model system consisted of the enzymes catalysing the last two steps in the pathway leading to patchoulol synthesis in yeast, namely FPPS and PTS. In this pathway, flux towards sesquiterpenes is believed to be limited because there is competition for the intermediate FPP and because FPP mediates negative feedback on a preceding pathway step. The second model system consisted of OMT and UGT - the enzymes that catalyse the last two steps in a novel pathway leading to VG synthesis in *S. cerevisiae*. The efficiency of this pathway is believed to be compromised for at least three reasons. First, accumulation of the toxic intermediate vanillin inhibits cell growth. Secondly, competition for enzyme leads to less efficient glycosylation of the target intermediate. Thirdly, the intermediate is lost through degradation and diffusion.

In the first part of the project, the effect of fusing either FPPS and PTS or OMT and UGT was examined (Chapter 2 and 3). The effect of fusion on final titers, intermediate pools and growth rates were investigated. When FPPS and PTS were fused, enzyme fusion resulted in a two-fold higher patchoulol production compared to the separate counterparts. Moreover, the strategy could be used in combination with a traditional metabolic engineering strategy, as *ERG9* repression and enzyme fusion combined could increase patchoulol production almost five-fold. When OMT and UGT were fused, enzyme fusion could efficiently decrease the accumulation of the toxic intermediate vanillin and presumably this resulted in a higher growth rate and an increased production rate. The benefits that could be achieved by enzyme fusion were shown to depend largely on the orientation of the enzymes within the fusion protein and to a certain degree

on the linkers that were used to combine the individual proteins. In summary the first part of the project demonstrated that both of these systems could be optimized by enzyme proximity and accordingly these two pathways were suitable model systems for testing the second strategy for positioning catalytic sites in close proximity.

In the second part of the project, a new tool for positioning enzymes in close proximity was developed and tested (Chapter 4-6). The developed tool involves the assembly of metabolic enzymes on a protein-based scaffold *in vivo* – here termed a nano-platform. The assembly occurs post-translationally and is guided by a self-associating protein. In Chapter 5, MmRad52- Δ 212 was identified as a suitable scaffoldin for forming the nano-platform core, as it was demonstrated to self-associate in the cytosol of yeast when it was tagged with two model enzymes. In Chapter 6, the effect of attaching metabolic enzymes to the MmRad52-based scaffoldin was investigated using the two model systems described above. For both the sesquiterpene and the vanillin system, proximity effects were observed when the enzymes were attached to the nano-platform, but in both cases the system was not competitive with expression of the free enzymes with respect to final titers of end-product. Presumably, the lower production is at least partly due to decreased functionality of the enzymes when they are linked to the scaffoldin.

7.1. Future perspectives

In a further development of the nano-platform concept, enzymes could be attached to the nano-platform via compatible cohesin and docking domains instead of being directly fused to the nano-platform subunits. In this case a universal nano-platform core displaying different types of cohesins could be constructed. The attachment of enzymes to such a nano-platform would only require that the metabolic enzymes are fused to a docking sequence that is specifically recognised by one of the cohesins displayed on the nano-platform. One of the advantages of this system over the approach used in this study would be that it would require less work to attach enzymes to the nano-platform. Possibly, fusion to a docking sequence will also affect protein folding and expression less than attachment to a whole protein.

The field engaged in engineering of the spatial organization of pathways is an emerging research field that is likely to benefit from advances within the field of protein structure prediction. In this project no attempts were made to predict the distance between catalytic sites when enzymes were engineered to be in close proximity by either enzyme fusion or nano-platform mediated assembly. Here, all engineering attempts were carried out on a trial and error basis. The experiments clearly showed that the efficiency of the strategies is dependent on the orientation of the enzymes and the linker sequences used to connect the fusion proteins. By predicting the structure of chimeric proteins as the ones generated in this project, it may be possible to envision how enzymes should be fused to each other or to the nano-platform in order to minimise the distance between catalytic sites. In this way, chimeric structures that allow pathways to benefit utmost from proximity effects may be generated using a more rational approach.

References

- An, J., Kim, Y., Lim, W., Hong, S., An, C., Shin, E., Cho, K., Choi, B., Kang, J., Lee, S., et al. (2005). Evaluation of a novel bifunctional xylanase-cellulase constructed by gene fusion. *Enzyme and Microbial Technology*, 36(7):989–995.
- Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering*, 14(8):529.
- Barghini, P., Di Gioia, D., Fava, F., and Ruzzi, M. (2007). Vanillin production using metabolically engineered *Escherichia coli* under non-growing conditions. *Microbial Cell Factories*, 6(1):13.
- Bayer, E., Belaich, J., Shoham, Y., and Lamed, R. (2004). The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annual Review of Microbiology*, 58(1):521–554.
- Bhathena, J., Kulamarva, A., Urbanska, A., Martoni, C., and Prakash, S. (2007). Microencapsulated bacterial cells can be used to produce the enzyme feruloyl esterase: preparation and in-vitro analysis. *Applied Microbiology and Biotechnology*, 75(5):1023–1029.
- Binder, M., Schanz, M., and Hartig, A. (1991). Vector-mediated overexpression of catalase A in the yeast *Saccharomyces cerevisiae* induces inclusion body formation. *European Journal of Cell Biology*, 54(2):305–12.
- Bülow, L. (1987). Characterization of an artificial bifunctional enzyme, beta-galactosidase/galactokinase, prepared by gene fusion. *European Journal of Biochemistry*, 163(3):443–448.
- Bülow, L. and Mosbach, K. (1991). Multienzyme systems obtained by gene fusion. *Trends in biotechnology(Regular ed.)*, 9(7):226–231.
- Boulikas, T. (1996). Nuclear import of protein kinases and cyclins. *Journal of Cellular Biochemistry*, 60:61–82.

- Boulikas, T. (1997). Nuclear import of DNA repair proteins. *Anticancer research*, 17(2 A):843–864.
- Bouwmeester, H. (2006). Engineering the essence of plants. *Nature Biotechnology*, 24:1359–1361.
- Brodelius, M., Lundgren, A., Mercke, P., and Brodelius, P. (2002). Fusion of farnesyl diphosphate synthase and epi-aristolochene synthase, a sesquiterpene cyclase involved in capsidiol biosynthesis in *Nicotiana tabacum*. *European Journal of Biochemistry*, 269(14):3570–3577.
- Carlsson, H., Ljung, S., and Bülow, L. (1996). Physical and kinetic effects on introduction of various linker regions in beta-galactosidase/galactose dehydrogenase fusion enzymes. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular Enzymology*, 1293(1):154–160.
- Caspi, J., Irwin, D., Lamed, R., Li, Y., Fierobe, H., Wilson, D., and Bayer, E. (2008). Conversion of *Thermobifida fusca* free exoglucanases into cellulosomal components: Comparative impact on cellulose-degrading activity. *Journal of Biotechnology*, 135(4):351–357.
- Chang, M. and Keasling, J. (2006). Production of isoprenoid pharmaceuticals by engineered microbes. *Nature Chemical Biology*, 2:674–681.
- Chatterjee, D. and Esposito, D. (2006). Enhanced soluble protein expression using two new fusion tags. *Protein Expression and Purification*, 46(1):122–129.
- Chemler, J., Yan, Y., and Koffas, M. (2006). Biosynthesis of isoprenoids, polyunsaturated fatty acids and flavonoids in *Saccharomyces cerevisiae*. *Microbial Cell Factories*, 5(1):20.
- Chotani, G., Dodge, T., Hsu, A., Kumar, M., LaDuca, R., Trimbur, D., Weyler, W., and Sanford, K. (2000). The commercial production of chemicals using pathway engineering. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular Enzymology*, 1543(2):434–455.
- Conrado, R., Varner, J., and DeLisa, M. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Current Opinion in Biotechnology*.

- Cortes-Ledesma, F., Malagon, F., and Aguilera, A. (2004). A Novel Yeast Mutation, rad52-L89F, Causes a Specific Defect in Rad51-Independent Recombination That Correlates With a Reduced Ability of Rad52-L89F to Interact With Rad59. *Genetics*, 168(1):553–557.
- Deane, C., Salwinski, L., Xenarios, I., and Eisenberg, D. (2002). Protein Interactions Two Methods for Assessment of the Reliability of High Throughput Observations*. *Molecular & Cellular Proteomics*, 1(5):349–356.
- Eberhardt, N. and Rilling, H. (1975). Prenyltransferase from *Saccharomyces cerevisiae*. Purification to homogeneity and molecular properties. *Journal of Biological Chemistry*, 250(3):863–866.
- Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246.
- Fierobe, H., Bayer, E., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., and Belaich, J. (2002). Degradation of Cellulose Substrates by Cellulosome Chimeras Substrate Targeting Versus Proximity of Enzyme Components. *Journal of Biological Chemistry*, 277(51):49621–49630.
- Fierobe, H., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J., and Bayer, E. (2001). Design and production of active cellulosome chimeras selective incorporation of dockerin-containing enzymes into defined functional complexes. *Journal of Biological Chemistry*, 276(24):21257–21261.
- Fierobe, H., Mingardon, F., Mechaly, A., Belaich, A., Rincon, M., Pages, S., Lamed, R., Tardif, C., Belaich, J., and Bayer, E. (2005). Action of Designer Cellulosomes on Homogeneous Versus Complex Substrates: Controlled Incorporation of Three Distinct Enzymes Into a Defined Trifunctional Scaffoldin. *Journal of Biological Chemistry*, 280(16):16325.
- Garcia-Fruitos, E., Gonzalez-Montalban, N., Morell, M., Vera, A., Ferraz, R., Aris, A., Ventura, S., and Villaverde, A. (2005). Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb Cell Fact*, 4:27.

- George, R. and Heringa, J. (2002). An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering*, 15(11):871.
- Gilbert, H. (2007). Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Molecular Microbiology*, 63(6):1568–1576.
- Gonzalez-Montalban, N., Garcia-Fruitos, E., and Villaverde, A. (2007). Recombinant protein solubility - does more mean better? *Nature Biotechnology*, 25(7):718.
- Hansen, E., Møller, B., Kock, G., Büchner, C., Kristensen, C., Jensen, O., Okkels, C., Olsen, C., Motawia, M., and Hansen, J. (2009). De novo biosynthesis of vanillin in fission yeast and baker's yeast. *Applied and Environmental Microbiology*, (12):In press.
- Hartig, A., Ogris, M., Cohen, G., and Binder, M. (1990). Fate of highly expressed proteins destined to peroxisomes in *Saccharomyces cerevisiae*. *Current Genetics*, 18(1):23–27.
- Hartl, F. and Hayer-Hartl, M. (2002). Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science*, 295(5561):1852–1858.
- Hays, S., Firmenich, A., Massey, P., Banerjee, R., and Berg, P. (1998). Studies of the Interaction between Rad52 Protein and the Yeast Single-Stranded DNA Binding Protein RPA. *Molecular and Cellular Biology*, 18(7):4400–4406.
- Heyman, A., Barak, Y., Caspi, J., Wilson, D., Altman, A., Bayer, E., and Shoseyov, O. (2007). Multiple display of catalytic modules on a protein scaffold: Nano-fabrication of enzyme particles. *Journal of Biotechnology*, 131(4):433–439.
- Hong, S., Lee, J., Cho, K., Math, R., Kim, Y., Hong, S., Cho, Y., Kim, H., and Yun, H. (2006). Assembling a novel bifunctional cellulase–xylanase from *Thermotoga maritima* by end-to-end fusion. *Biotechnology Letters*, 28(22):1857–1862.
- Huh, W., Falvo, J., Gerke, L., Carroll, A., Howson, R., Weissman, J., and O'Shea, E. (2003). Global analysis of protein localization in budding yeast. *Nature*, 425:686–691.

- Hyde, C., Ahmed, S., Padlan, E., Miles, E., and Davies, D. (1988). Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *Journal of Biological Chemistry*, 263(33):17857–17871.
- Jones, S. and Thornton, J. (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences*, 93(1):13–20.
- Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T., and Yokoyama, S. (2002). Crystal Structure of the Homologous-Pairing Domain from the Human Rad52 Recombinase in the Undecameric Form. *Molecular Cell*, 10(2):359–371.
- Kamasawa, N., Yoshida, T., Ueda, M., Tanaka, A., and Osumi, M. (1999). Three-dimensional analysis of protein aggregate body in *Saccharomyces cerevisiae* cells. *Journal of Electron Microscopy*, 48(2):173–176.
- Karst, F., Plochocka, D., Meyer, S., and Szkopinska, A. (2004). Farnesyl diphosphate synthase activity affects ergosterol level and proliferation of yeast *Saccharomyces cerevisiae*. *Cell Biology International*, 28(3):193–197.
- Keasling, J. (2008). Cheaper medicines. *Engineering and Technology Magazine*, 13(6 Suppl).
- Klinke, H., Thomsen, A., and Ahring, B. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Applied Microbiology and Biotechnology*, 66(1):10–26.
- Kourtz, L., Dillon, K., Daughtry, S., Madison, L., Peoples, O., and Snell, K. (2005). A novel thiolase-reductase gene fusion promotes the production of polyhydroxybutyrate in *Arabidopsis*. *Plant Biotechnology Journal*, 3(4):435–447.
- Krejci, L., Song, B., Bussen, W., Rothstein, R., Mortensen, U., and Sung, P. (2002). Interaction with Rad51 Is Indispensable for Recombination Mediator Function of Rad52. *Journal of Biological Chemistry*, 277(42):40132–40141.
- Krejci, L., Thomsen, B., Duno, M., Westergaard, O., and Bendixen, C. (2000). Homomeric interaction of the mouse Rad52 protein. *Molecular Biology Reports*, 27(1):55–59.

- Krings, U. and Berger, R. (1998). Biotechnological production of flavours and fragrances. *Applied Microbiology and Biotechnology*, 49(1):1–8.
- Larsson, S., Quintana-Sáinz, A., Reimann, A., Nilvebrant, N., and Jönsson, L. (2000). Influence of lignocellulose-derived aromatic compounds on oxygen-limited growth and ethanolic fermentation by *Saccharomyces cerevisiae*. *Applied Biochemistry and Biotechnology*, 84(1):617–632.
- Li, K. and Frost, J. (1998). Synthesis of Vanillin from Glucose. *Journal-American Chemical Society*, 120:10545–10546.
- Lisby, M., Rothstein, R., and Mortensen, U. (2001). Rad52 forms DNA repair and recombination centers during S phase. *Proceedings of the National Academy of Sciences*, 98(15):8276–8282.
- Lu, P. and Feng, M. (2008). Bifunctional enhancement of a beta-glucanase-xylanase fusion enzyme by optimization of peptide linkers. *Applied Microbiology and Biotechnology*, 79(4):579–587.
- Meynial Salles, I., Forchhammer, N., Croux, C., Girbal, L., and Soucaille, P. (2007). Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. *Metabolic Engineering*, 9(2):152–159.
- Mingardon, F., Chanal, A., López-Contreras, A., Dray, C., Bayer, E., and Fierobe, H. (2007a). Incorporation of Fungal Cellulases in Bacterial Minicellulosomes Yields Viable, Synergistically Acting Cellulolytic Complexes? *Applied and Environmental Microbiology*, 73(12):3822–3832.
- Mingardon, F., Chanal, A., Tardif, C., Bayer, E., and Fierobe, H. (2007b). Exploration of New Geometries in Cellulosome-Like Chimeras? *Applied and Environmental Microbiology*, 73(22):7138–7149.
- Mortensen, U., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. *Proceedings of the National Academy of Sciences*, 93(20):10729–10734.
- Mortensen, U., Erdeniz, N., Feng, Q., and Rothstein, R. (2002). A Molecular Genetic Dissection of the Evolutionarily Conserved N Terminus of Yeast Rad52. *Genetics*, 161(2):549–562.

- Mosbach, K. and Mattiasson, B. (1970). Matrix-bound enzymes. II. Studies on a matrix-bound two-enzyme-system. *Acta Chem Scand*, 24(6):2093–100.
- Nielsen, K., Tattersall, D., Jones, P., and Møller, B. (2008). Metabolon formation in dhurrin biosynthesis. *Phytochemistry*, 69(1):88–98.
- Orita, I., Sakamoto, N., Kato, N., Yurimoto, H., and Sakai, Y. (2007). Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. *Applied Microbiology and Biotechnology*, 76(2):439–445.
- Ovadi, J. and Saks, V. (2004). On the origin of intracellular compartmentation and organized metabolic systems. *Molecular and Cellular Biochemistry*, 256(1):5–12.
- Ovadi, J. and Srere, P. (1996). Metabolic Consequences of Enzyme Interactions. *Cell Biochemistry and Function*, 14(4):249–258.
- Ovadi, J. and Srere, P. (2000). Macromolecular compartmentation and channeling. *Int Rev Cytol*, 192:255–80.
- Pages, S., Bélac, A., Bélac, J., Morag, E., Lamed, R., Shoham, Y., and Bayer, E. (1997). Species-Specificity of the Cohesin-Dockerin Interaction Between *Clostridium thermocellum* and *Clostridium cellulolyticum*: Prediction of Specificity Determinants of the Dockerin Domain. *Proteins: Structure, Function, and Genetics*, 29:517–527.
- Park, M., Ludwig, D., Stigger, E., and Lee, S. (1996). Physical Interaction between Human RAD52 and RPA Is Required for Homologous Recombination in Mammalian Cells. *Journal of Biological Chemistry*, 271(31):18996.
- Petes, T., Malone, R., and Symington, L. (1991). Recombination in yeast. *The molecular and cellular biology of the yeast Saccharomyces*, 1:407–521.
- Plate, I., Albertsen, L., Lisby, M., Hallwyl, S., Feng, Q., Seong, C., Rothstein, R., Sung, P., and Mortensen, U. (2008a). Rad52 multimerization is important for its nuclear localization in *Saccharomyces cerevisiae*. *DNA Repair*, 7(1):57–66.

- Plate, I., Hallwyl, S., Shi, I., Krejci, L., Muller, C., Albertsen, L., Sung, P., and Mortensen, U. (2008b). Interaction with RPA Is Necessary for Rad52 Repair Center Formation and for Its Mediator Activity. *Journal of Biological Chemistry*, 283(43):29077.
- Prachayasittikul, V., Ljung, S., Isarankura-Na-Ayudhya, C., and Bülow, L. (2006). NAD (H) recycling activity of an engineered bifunctional enzyme galactose dehydrogenase/lactate dehydrogenase. *International Journal of Biological Sciences*, 2(1):10.
- Prescott, M., Nowakowski, S., Nagley, P., and Devenish, R. (1999). The Length of Polypeptide Linker Affects the Stability of Green Fluorescent Protein Fusion Proteins. *Analytical Biochemistry*, 273(2):305–307.
- Priefert, H., Rabenhorst, J., and Steinbüchel, A. (2001). Biotechnological production of vanillin. *Applied Microbiology and Biotechnology*, 56(3):296–314.
- Ranatunga, W., Jackson, D., II, R., and Borgstahl, G. (2001a). Human RAD52 Protein Has Extreme Thermal Stability. *Biochemistry*, 40(29):8557–8562.
- Ranatunga, W., Jackson, D., Lloyd, J., Forget, A., Knight, K., and Borgstahl, G. (2001b). Human RAD52 Exhibits Two Modes of Self-association. *Journal of Biological Chemistry*, 276(19):15876–15880.
- Robinson, C. and Sauer, R. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences*, 95(11):5929–5934.
- Sambrook, J., Russell, D., and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. Cold Spring Harbor Laboratory Press.
- Seo, H., Koo, Y., Lim, J., Song, J., Kim, C., Kim, J., Lee, J., and Choi, Y. (2000). Characterization of a Bifunctional Enzyme Fusion of Trehalose-6-Phosphate Synthetase and Trehalose-6-Phosphate Phosphatase of *Escherichia coli*. *Applied and Environmental Microbiology*, 66(6):2484–2490.
- Sheibani, N. (1999). Prokaryotic Gene Fusion Expression Systems and their Use in Structural and Functional Studies of Proteins. *Preparative Biochemistry and Biotechnology*, 29(1):77–90.

- Shen, Z., Peterson, S., Comeaux, J., Zastrow, D., Moyzis, R., Bradbury, E., and Chen, D. (1996). Self-association of human RAD52 protein. *Mutation Research-DNA Repair*, 364(2):81–89.
- Sherman, F., Fink, G., Hicks, J., and Laboratory, C. S. H. (1986). *Laboratory course manual for methods in yeast genetics*. Cold Spring Harbor Laboratory [SI].
- Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S., and Ogawa, T. (1998). Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells*, 3(3):145–156.
- Singleton, M., Wentzell, L., Liu, Y., West, S., and Wigley, D. (2002). Structure of the single-strand annealing domain of human RAD52 protein. *Proceedings of the National Academy of Sciences*, 99(21):13492–13497.
- Srere, P., Mattiasson, B., and Mosbach, K. (1973). An Immobilized Three-Enzyme System: A Model for Microenvironmental Compartmentation in Mitochondria. *Proceedings of the National Academy of Sciences*, 70(9):2534–2538.
- Stasiak, A., Larquet, E., Stasiak, A., Müller, S., Engel, A., Van Dyck, E., West, S., and Egelman, E. (2000). The human Rad52 protein exists as a heptameric ring. *Current Biology*, 10(6):337–340.
- Stryer, L. (1995). *Biochemistry*. New York, pages 192–196.
- Venkatasubramanian, P., Daniels, L., and Rosazza, J. (2007). Reduction of Carboxylic Acids by Nocardia Aldehyde Oxidoreductase Requires a Phosphopantetheinylated Enzyme. *Journal of Biological Chemistry*, 282(1):478.
- Ventura, S. and Villaverde, A. (2006). Protein quality in bacterial inclusion bodies. *Trends in Biotechnology*, 24(4):179–185.
- Walton, N., Narbad, A., Faulds, C., and Williamson, G. (2000). Novel approaches to the biosynthesis of vanillin. *Current Opinion in Biotechnology*, 11(5):490–496.
- Weissman, K. (2006). The Structural Basis for Docking in Modular Polyketide Biosynthesis. *ChemBioChem*, 7(3):485.

-
- Winkel, B. (2004). Metabolic Channeling in Plants. *Annual Review of Plant Biology*, 55:85–107.
- Wu, S., Schalk, M., Clark, A., Miles, R., Coates, R., and Chappell, J. (2006). Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature Biotechnology*, 24:1441–1447.
- Xue, F., Gu, Z., and Feng, J. (2004). LINKER: a web server to generate peptide sequences with extended conformation. *Nucleic Acids Research*, 32(Web Server Issue):W562.
- Yilmaz, J. and Bulow, L. (2002). Enhanced Stress Tolerance in *Escherichia coli* and *Nicotiana tabacum* Expressing a Betaine Aldehyde Dehydrogenase/Choline Dehydrogenase Fusion Protein. *Biotechnology Progress*, 18(6):1176–1182.
- Yoon, S., Lee, E., Das, A., Lee, S., Li, C., Ryu, H., Choi, M., Seo, W., and Kim, S. (2007). Enhanced Vanillin Production from Recombinant *E. coli* Using NTG Mutagenesis and Adsorbent Resin. *Biotechnology Progress*, 23(5):1143–1148.