

***In vitro* Untersuchungen zu Proteinbiosynthese-
assoziierten Prozessen:**

**Methylierungsabhängige Genregulation und
Konkurrenz zwischen kanonischen und nicht-
kanonischen Aminosäuren beim Einbau in
Proteine**

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Marc Finkler

Saarbrücken

2022

Tag des Kolloquiums: 18.07.2022

Dekan: Prof. Dr. Jörn Erik Walter

Berichterstatter: Prof. Dr. Albrecht Ott

Jun.-Prof. Dr. Laura Aradilla-Zapata

Prof. Dr. Hans-Günther Döbereiner

Vorsitz: Prof. Dr. Karin Jacobs

Akad. Mitarbeiter: Dr. Jens Neunzig

Danksagung

An erster Stelle bedanke ich mich bei meinem Doktorvater Herrn Prof. Dr. Albrecht Ott für die Möglichkeit der Bearbeitung eines solch interessanten Themas. Dadurch, dass er mir immer mit Rat und Tat zur Seite stand, hat er maßgeblich am Gelingen dieser Dissertation beigetragen. Durch die vielen wissenschaftlichen Diskussionen konnte ich eine Vielzahl neuer Erkenntnisse erlangen. Vielen Dank!

Zudem möchte ich Frau Jun.-Prof. Dr. Laura Aradilla-Zapata für die Zweitkorrektur meiner Arbeit danken.

Ich danke den Mitgliedern des Arbeitskreises von Herrn Prof. Dr. Albrecht Ott für ihre Hilfsbereitschaft, ihre Freundlichkeit und die vielen gemeinsamen Stunden, in denen wir zusammen forschen konnten.

Ein besonderer Dank gilt Emanuel Worst, meinem wissenschaftlichen Vorgänger. Durch seine anfängliche Betreuung, sein Know-how und seine ständige Hilfsbereitschaft war es mir möglich, seine Arbeit am Lehrstuhl erfolgreich fortzuführen.

Auch Philip Hartz möchte ich an dieser Stelle besonders danken, da er mir mit seinem biologischen Fachwissen und seiner enthusiastischen Art stets zur Seite stand.

Ich möchte meinen Freunden und Verwandten danken, die mir immer einen Ausgleich zum Studium geboten haben.

Ein besonderer Dank gilt meinen Eltern Ingrid und Klaus Finkler, meinem Bruder Björn Finkler sowie meiner Ehefrau Yonlada Finkler. Sie waren während meiner Promotion immer an meiner Seite und gaben mir Unterstützung, wenn ich sie brauchte.

Wissenschaftliche Arbeiten

Diese kumulative Dissertation basiert auf fünf Originalpublikationen, welche in Abschnitt 3 abgedruckt sind. Die Einbindung der Veröffentlichungen innerhalb dieser Dissertation wurde von den entsprechenden Verlägen genehmigt:

- IOP Publishing (Physical Biology): 3.1 Schenkelberger *et al.* (2017)
- ACS Publications (ACS Synthetic Biology): 3.2 Worst *et al.* (2021)
- Future Science Group (BioTechniques): 3.3 Finkler & Ott (2019)
- Elsevier (Journal of Biotechnology): 3.4 Finkler *et al.* (2020)
- Elsevier (Bioorganic & Medicinal Chemistry): 3.5 Finkler *et al.* (2021)

3.1 Schenkelberger *et al.* (2017)

Schenkelberger, M., Shanak, S., **Finkler, M.**, Worst, E. G., Noireaux, V., Helms, V., and Ott, A. (2017) Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system. *Phys. Biol.* 14, 1–11. <https://doi.org/10.1088/1478-3975/aa5d37>

Marc Finkler bestätigte den vorgeschlagenen Mechanismus der epigenetischen Regulation der Genexpression anhand durchgeführter Experimente.

Zudem schlug Marc Finkler das einfache kinetische Modell für das Binden von MBD und/oder σ^{70} an den BDNF-Promotor vor, welches in dieser Arbeit vorgestellt wurde. Er bestätigte dieses Modell durch Anpassen der experimentellen Daten mit der aus dem Modell resultierenden Gleichung.

Marc Finkler war beim Schreiben des Manuskripts involviert.

3.2 Worst *et al.* (2021)

Worst, E. G.†, **Finkler, M.**†, Schenkelberger, M., Kurt, Ö., Helms, V., Noireaux, V., and Ott, A. (2021) A Methylation-Directed, Synthetic Pap Switch Based on Self-Complementary Regulatory DNA Reconstituted in an All *E. coli* Cell-Free Expression System. *ACS Synth. Biol.* 10, 2725–2739. <https://doi.org/10.1021/acssynbio.1c00326>

† Geteilte Erstautorenschaft

Marc Finkler war an der Generierung der experimentellen Daten bzgl. der methylierungsabhängigen Genregulation beteiligt.

Marc Finkler designte Experimente, welche der Untersuchung der Sekundärstruktur des vorgestellten synthetischen Pap epigenetischen Schalters dienten und führte diese durch.

Marc Finkler war beim Schreiben des Manuskripts involviert.

3.3 Finkler & Ott (2019)

Finkler, M., and Ott, A. (2019) Bead-based assay for spatiotemporal gene expression control in cell-free transcription–translation systems. *BioTechniques* 66, 29–33. <https://doi.org/10.2144/btn-2018-0097>

Marc Finkler hat zu allen Beitrags-Kategorien beigetragen.

3.4 Finkler *et al.* (2020)

Finkler, M., Kurt, Ö., Grimm, F., Hartz, P., and Ott, A. (2020) A bead-based method for the removal of the amino acid lysine from cell-free transcription-translation systems. *J. Biotechnol.* 324S, 1–6. <https://doi.org/10.1016/j.btex.2020.100024>

Marc Finkler hat zu allen Beitrags-Kategorien beigetragen.

3.5 Finkler *et al.* (2021)

Finkler, M., Ravanbodshirazi, S., Grimm, F., Hartz, P., and Ott, A. (2021) Full incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in green fluorescent protein. *Bioorg. Med. Chem.* 41, 1–6. <https://doi.org/10.1016/j.bmc.2021.116207>

Marc Finkler hat zu allen Beitrags-Kategorien beigetragen.

Weitere wissenschaftliche Arbeiten

Die in Finkler *et al.* (2020) präsentierten Ergebnisse wurden zuvor auf dem Preprint-Server BioRxiv veröffentlicht (<https://doi.org/10.1101/2020.01.15.907360>). Dies führte zu einem Online-Artikel der Zeitschrift „Laborjournal“, welcher unter <https://www.laborjournal.de/editorials/1923.php> aufgerufen werden kann.

Neben den eingebunden Publikationen wurde ein weiteres Manuskript publiziert, welches jedoch in dieser Dissertation nicht näher diskutiert wird:

Khangholi, N.†, **Finkler, M.**†, Seemann, R., Ott, A., and Fleury, J.-B. (2021) Photoactivation of Cell-Free Expressed Archaerhodopsin-3 in a Model Cell Membrane. *Int. J. Mol. Sci.* 22, 1–8. <https://doi.org/10.3390/ijms222111981>

† Geteilte Erstautorenschaft

Inhaltsverzeichnis

Inhaltsverzeichnis.....	vii
Abkürzungsverzeichnis.....	ix
Zusammenfassung.....	xiii
Abstract.....	xiv
1 Einleitung.....	1
2 Grundlagen.....	4
2.1 Nukleinsäuren.....	4
2.2 Proteine.....	8
2.3 Proteinbiosynthese.....	10
2.4 Expressionssysteme.....	13
2.5 Regulation der Genexpression.....	17
2.6 Einbau nicht-kanonischer Aminosäuren in Proteine.....	21
2.6.1 Ortsspezifischer Einbau nicht-kanonischer Aminosäuren.....	22
2.6.2 Seitenkettenspezifischer Einbau nicht-kanonischer Aminosäuren.....	24
2.7 Fehlerreduktion auf Translationsebene durch „Kinetic Proofreading“.....	27
2.8 Fluoreszenz.....	30
2.8.1 Grundlagen.....	30
2.8.2 Grün fluoreszierendes Protein.....	34
2.9 Massenspektrometrie.....	37
3 Originalpublikationen.....	41
3.1 Schenkelberger <i>et al.</i> (2017).....	41
3.2 Worst <i>et al.</i> (2021).....	64
3.3 Finkler & Ott (2019).....	94
3.4 Finkler <i>et al.</i> (2020).....	103
3.5 Finkler <i>et al.</i> (2021).....	113
4 Diskussion.....	126
4.1 Schenkelberger <i>et al.</i> (2017).....	126
4.1.1 Hintergrund.....	126

Inhaltsverzeichnis

4.1.2 Ergebnisse	126
4.1.3 Zusammenfassung und Ausblick.....	129
4.2 Worst <i>et al.</i> (2021).....	131
4.2.1 Hintergrund	131
4.2.2 Ergebnisse	133
4.2.3 Zusammenfassung und Ausblick.....	139
4.3 Finkler & Ott (2019).....	141
4.3.1 Hintergrund	141
4.3.2 Ergebnisse	141
4.3.3 Zusammenfassung und Ausblick.....	144
4.4 Finkler <i>et al.</i> (2020).....	146
4.4.1 Hintergrund	146
4.4.2 Ergebnisse	146
4.4.3 Zusammenfassung und Ausblick.....	150
4.5 Finkler <i>et al.</i> (2021).....	151
4.5.1 Hintergrund	151
4.5.2 Ergebnisse	151
4.5.3 Zusammenfassung und Ausblick.....	156
5 Weitere Ergebnisse.....	160
5.1 Spektroskopische Methode zur Untersuchung des Aminosäure-Einbaus	160
5.2 Massenspektrometrische Methode zur Untersuchung des Aminosäure-Einbaus.....	165
6 Zusammenfassung und Ausblick.....	169
Anhang	170
Literaturverzeichnis.....	174

Abkürzungsverzeichnis

AaRS(s)	Aminoacyl-tRNA-Synthetase(n)
ACN	Acetonitril
AGE(s)	Advanced glycation end product(s) (engl.)
AMP	Adenosinmonophosphat
AQUA	Absolute quantification (engl.)
ATP	Adenosintriphosphat
BDNF	Brain-derived neurotrophic factor (engl.)
BHQ1	Black Hole Quencher-1 (engl.)
bp	Basenpaar(e)
BS	Bindestelle(n)
bzgl.	bezüglich
bzw.	beziehungsweise
ca.	circa (lat. „ungefähr“)
cAMP	Cyclisches Adenosinmonophosphat
Can	Canavanin
CAP	Catabolite activator protein (engl.)
CML	Carboxy-Methyl-Lysin
CRISPR/Cas	Clustered Regulatory Interspaced Short Palindromic Repeats/ CRISPR associated proteins (engl.)
crRNA	CRISPR-RNA

Abkürzungsverzeichnis

Dam	DNA-Adenin-Methyltransferase
DNA	Desoxyribonukleinsäure
<i>E. coli</i>	<i>Escherichia coli</i>
EF	Elongationsfaktor
EGFP	Enhanced GFP (engl.)
EI	Elektronenstoßionisation
ESI	Elektronenspray-Ionisation
ESPT	Excited state proton transfer (engl.)
<i>et al.</i>	et alii (lat. „und andere“)
etc.	et cetera (lat. „und die übrigen (Dinge)“)
FA	Ameisensäure
6-Fam	6-Carboxyfluorescein
FRET	Förster-Resonanzenergietransfer
FT-ICR	Fourier-Transformations-Ionenzyklotronresonanz
GC	Gaschromatographie
GFP	Grün fluoreszierendes Protein
ggf.	gegebenenfalls
HLDP	High-Lysine-Dummy-Protein (engl.)
HPLC	Hochleistungsflüssigkeitschromatographie
Hyl	Hydroxylysin
IMAC	Immobilisierte-Metallionen-Affinitätschromatographie

IPTG	Isopropyl- β -D-thiogalactopyranosid
i. V. m.	in Verbindung mit
LC	Flüssigchromatographie
Lrp	Leucine-responsive regulatory protein (engl.)
MALDI	Matrix-unterstützte Laser-Desorption/Ionisation
MBD	Methyl-CpG Bindedomäne
MCP	Mikrokanalplatten
mRNA	Boten-Ribonukleinsäure
MS	Massenspektrometrie
NIST	National Institute of Standards and Technology (engl.)
NkAA	Nicht-kanonische Aminosäure
Ori	Replikationsursprung
PAGE	Polyacrylamid-Gelelektrophorese
Pap	<i>Pyelonephritis</i> -assoziierte Pili
PCR	Polymerase-Kettenreaktion
PURE	Protein synthesis using recombinant elements (engl.)
RF	Release factor (engl.)
RNA	Ribonukleinsäure
SDS	Natriumlaurylsulfat
SEV	Sekundärelektronenvervielfacher
tRNA	Transfer-Ribonukleinsäure

Abkürzungsverzeichnis

u. a.	unter anderem
UPEC	Uropathogene <i>E. coli</i>
UTR	Untranslatierte Region
w.E.	Willkürliche Einheit
z. B.	zum Beispiel

Zusammenfassung

Obwohl eine Vielzahl an Prozessen, die während der Proteinbiosynthese ablaufen, bereits ausführlich untersucht wurden, sind deren zugrunde liegende molekulare Mechanismen oft nicht vollständig verstanden. Prominente Beispiele sind methylierungsabhängige Genregulationsmechanismen sowie die Fehlerreduktion beim Einbau von Aminosäuren unter kompetitiven Bedingungen. Da zellfreie Expressionssysteme im Vergleich zu *in vivo* Systemen eine genauere Kontrolle diverser biochemischer Parameter erlauben, eignen sich solche Systeme für die Untersuchung der zuvor genannten Prozesse.

Im Rahmen dieser Dissertation wurden unter Verwendung eines auf *E. coli* basierenden, zellfreien Expressionssystems zunächst zwei verschiedene methylierungsabhängige Genregulationsmechanismen untersucht: die MeCP2-assoziierte Genregulation sowie die Regulation des Pap epigenetischen Schalters. Bei beiden Untersuchungsobjekten ergaben sich Hinweise, dass eine veränderte Sekundärstruktur der DNA ein essenzieller Bestandteil der methylierungsabhängigen Genregulation ist.

Das gleiche *in vitro* System wurde auch für Untersuchungen der Konkurrenz zwischen der kanonischen Aminosäure Lysin und deren nicht-kanonischem Analogon Hydroxylysin beim Einbau in Proteine verwendet. In diesem Zusammenhang wurden neue Methoden entwickelt, die den vollständigen seitenkettenspezifischen Einbau von Hydroxylysin als Lysin-Ersatz unter den bestehenden kompetitiven Bedingungen ermöglichten.

Abstract

Although a variety of processes involved in protein biosynthesis are well examined, often there is a lack of knowledge regarding underlying molecular mechanisms. Prominent examples are methylation dependent gene regulation mechanisms, or the error reduction of amino acid incorporation in competitive situations. Because cell-free gene expression systems allow to control various biochemical parameters more precisely compared to *in vivo* systems, they are well suited for the investigation of the earlier mentioned processes.

In this dissertation, at first, two different methylation dependent gene regulation mechanisms were investigated using an *E. coli* based, cell-free gene expression system: the MeCP2-associated gene regulation as well as the regulation of the pap epigenetic switch. Both objects of investigation agree well with the idea of an altered secondary structure of the DNA being an essential part of methylation dependent gene regulation.

The same *in vitro* system was also used for the investigation of the concurrence between the canonical amino acid lysine and its noncanonical analog hydroxylysine during their incorporation into proteins. In this context, new methods were developed to achieve the complete residue-specific incorporation of hydroxylysine as a lysine surrogate under such competitive situations.

1 Einleitung

In die Proteinbiosynthese sind eine Vielzahl von verschiedenen Prozessen involviert, welche u. a. der Genregulation, der Fehlerreduktion oder der Proteinmodifikation dienen¹⁻⁴. Aufgrund dieser Komplexität ist es von besonderem Interesse, die einzelnen Prozesse und deren zugrunde liegende Mechanismen zu verstehen, um ein besseres Gesamtverständnis der Proteinbiosynthese zu generieren^{1,4,5}. Dieses kann helfen, krankheitsassoziierte Veränderungen in der Proteinbiosynthese zu erkennen oder zu behandeln, die rekombinante Produktion von modifizierten Proteinen mit neuen Funktionen durchzuführen oder neue biotechnologisch wertvolle Genregulationsmechanismen zu etablieren⁶⁻⁹.

Zur Untersuchung der einzelnen Prozesse können *in vivo* oder *in vitro* Expressionssysteme verwendet werden^{10,11}. Expressionssysteme ermöglichen unter Verwendung eines geeigneten DNA-Konstruktes die kontrollierte Proteinbiosynthese eines bestimmten Proteins in Form einer Genexpressionsreaktion¹¹⁻¹⁴. Das DNA-Konstrukt muss lediglich über das zu exprimierende Gen sowie zu dessen Expression benötigte regulatorische DNA-Elemente verfügen^{15,16}. Mittels einfacher molekularbiologischer Methoden kann ein solches DNA-Konstrukt leicht modifiziert werden¹⁷. Dies ermöglicht das Einbringen von regulatorischen DNA-Elementen in nativer sowie in modifizierter Form. Deren Auswirkung auf die Genexpression kann anschließend in einer Genexpressionsreaktion untersucht werden. Zusätzlich können in das System Komponenten eingebracht werden, welche einen Effekt auf die Genexpression bzw. Proteinbiosynthese haben oder von denen dies vermutet wird^{18,19}.

Diesbezüglich weisen *in vivo* Expressionssysteme Einschränkungen auf^{11,12}. So wird bei solchen Systemen die Aufnahme der Komponenten in die Zellen vorausgesetzt und zudem dürfen die Komponenten keine negativen Effekte auf den Organismus ausüben, die eine verminderte Überlebensfähigkeit zur Folge haben¹⁹⁻²². Diese Einschränkungen gelten nicht für *in vitro* Expressionssysteme. Zusätzlich ermöglichen *in vitro* Systeme eine besser kontrollierbare biochemische Umgebung²³⁻²⁵. Somit sind *in vitro* Expressionssysteme unter bestimmten Umständen zu bevorzugen^{23,26,27}. Dies gilt auch für die innerhalb dieser Dissertation durchgeführten Studien, da die zu untersuchenden Prozesse zur genaueren Untersuchung definierte biochemische Parameter erforderten und die verwendeten Substanzen in einem Organismus toxisch wirken konnten. Aus diesen Gründen wurde für die durchgeführten Experimente stets ein auf *E. coli* basierendes, zellfreies Expressionssystem verwendet^{28,29}.

Die Genregulation ist ein essenzieller Bestandteil der Proteinbiosynthese, da dadurch die Menge an synthetisiertem Protein reguliert werden kann³⁰⁻³². Die Regulation kann durch verschiedene Mechanismen auf genetischer oder auf epigenetischer Ebene stattfinden³³⁻³⁶. Hierbei bezeichnet Epigenetik alle Mechanismen, die eine vererbare Genregulation ohne Veränderung der DNA-Sequenz veranlassen^{34,36,37}. Eine Gruppe von epigenetischen Genregulationsmechanismen umfasst die methylierungsabhängigen Genregulationsmechanismen^{37,38}. Bei diesen wird die DNA an spezifischen

1 Einleitung

Stellen mit einer Methylgruppe versehen, was in Verbindung mit verschiedenen Transkriptions-Regulatoren zu einer veränderten Genexpression führt^{37,39,40}. Obwohl eine Vielzahl an methylierungsabhängigen Genregulationsmechanismen bereits in der Literatur beschrieben wurden, sind die zugrunde liegenden molekularen Mechanismen oft nur teilweise verstanden^{33,41,42}. Aus diesem Grund wurden im Rahmen der vorliegenden Dissertation zwei verschiedene methylierungsabhängige Genregulationsmechanismen näher untersucht. Erstes Untersuchungsobjekt war die durch den humanen Transkriptions-Regulator MeCP2 verursachte eukaryotische epigenetische Genregulation in Verbindung mit der methylierten „Brain-derived neurotrophic factor“ (BDNF)-Promotor-Sequenz^{43–49} (siehe Schenkelberger *et al.* (2017), Abschnitt 3.1). Hierbei wurde lediglich die Methyl-CpG Bindedomäne (MBD) von MeCP2 betrachtet und deren Einfluss auf die Genexpression eines unter Kontrolle des BDNF-Promotors befindlichen Gens in Abhängigkeit des BDNF-Methylierungszustandes untersucht. Das zweite Untersuchungsobjekt war der Phasenvariationsmechanismus des Pap epigenetischen Schalters^{50–53} (siehe Worst *et al.* (2021), Abschnitt 3.2). Dieser kann in Abhängigkeit des Methylierungszustandes zwischen zwei möglichen Phasen „schalten“, in denen die *pap*-Gene exprimiert oder reprimiert werden^{40,53}. Auf Grundlage des Pap epigenetischen Schalters wurde eine synthetische Variante generiert, bei der je nach Methylierungszustand unterschiedliche Gene exprimiert werden. Dies ermöglichte eine bessere Untersuchung des Phasenvariationsmechanismus des Pap epigenetischen Schalters.

Die Fehlerreduktion ist ein anderer essenzieller Bestandteil der Proteinbiosynthese^{2,54,55}. Diverse darin involvierte Prozesse gewährleisten, dass das synthetisierte Protein die genetisch vorgegebene Aminosäuresequenz aufweist^{2,54,56}. Ein Beispiel für einen solchen Prozess ist das so genannte „Kinetic Proofreading“, welches für eine starke Präferenz der „richtigen“ gegenüber „falschen“ Aminosäuren bei der Aminoacyl-tRNA-Synthetase vermittelten Beladung der tRNA verantwortlich ist^{2,22,57,58}. In diesem Zusammenhang kann „Kinetic Proofreading“ auch dem Einbau von nicht-kanonischen Aminosäuren in Proteine als Ersatz einer kanonischen Aminosäure entgegenwirken^{59,60}. Da solche Situationen in der klassischen Betrachtung des „Kinetic Proofreading“-Prozesses^{54,58,61,62} nicht berücksichtigt werden, können entsprechende Experimente zu einer erweiterten Betrachtung führen. Dies kann wiederum in Zukunft den Einbau nicht-kanonischer Aminosäuren in Proteine erleichtern. Aus diesen Gründen wurde im Rahmen der vorliegenden Dissertation die Konkurrenz zwischen Lysin und dessen nicht-kanonischem Analogon Hydroxylysin beim Einbau in Proteine untersucht. In der Literatur wurde bereits gezeigt, dass der vollständige seitenkettenspezifische Einbau von Hydroxylysin als Lysin-Ersatz durch geringe Mengen an im Expressionssystem befindlichem Lysin verhindert wurde⁵⁷. Folglich musste der Gehalt von Lysin auf eine vernachlässigbare Menge reduziert werden, um einen Einbau von Hydroxylysin zu ermöglichen. Zu diesem Zweck wurde zunächst eine Methode entwickelt, die eine räumlich und zeitlich begrenzte Genexpression erlaubte (siehe Finkler & Ott (2019), Abschnitt 3.3). Mit dieser Methode wurde die räumlich und zeitlich begrenzte Genexpression eines Dummy-Proteins mit hohem Lysin-Gehalt durchgeführt, um den Lysin-Gehalt zu reduzieren (siehe Finkler *et al.* (2020),

Abschnitt 3.4). Hierbei konnte zusätzlich festgestellt werden, dass eine effektive Lysin-Entfernung auch durch die elektrostatische Interaktion zwischen Lysin und DNA⁶³⁻⁶⁵ erfolgen kann. Mit diesen neuen Erkenntnissen konnte ein Expressionssystem mit reduziertem Lysin-Gehalt generiert werden, mit dem Untersuchungen zum Einbau von Hydroxylysin unter kompetitiven Bedingungen ermöglicht wurden (siehe Finkler *et al.* (2021), Abschnitt 3.5).

In Abschnitt 2 werden zunächst die wichtigsten Grundlagen erläutert, welche für ein tieferes Verständnis der vorliegenden Dissertation benötigt werden. Nach Betrachtung der Grundlagen, werden die fünf zuvor beschriebenen Studien aufgelistet (Abschnitt 3) und diskutiert (Abschnitt 4). Darauf aufbauend wurden im Rahmen dieser Dissertation weitere Experimente durchgeführt, welche in Abschnitt 5 näher beschrieben werden. Im Anschluss folgt eine Zusammenfassung der publizierten Studien und ein Ausblick, in dem deren zukünftiger Nutzen beschrieben wird (Abschnitt 6).

2 Grundlagen

Dieser Abschnitt behandelt die notwendigen Grundlagen, welche für ein tieferes Verständnis der vorliegenden Dissertation benötigt werden. Hierbei erfolgt zunächst eine Betrachtung der Komponenten, welche in die Proteinbiosynthese involviert sind, sowie eine Betrachtung des eigentlichen Prozesses der Proteinbiosynthese. Anschließend werden verschiedene mögliche Expressionssysteme beschrieben und verglichen, gefolgt von einer Übersicht über die gängigen Methoden zur Genregulation. Im Anschluss erfolgt eine Betrachtung der möglichen Methoden für den Einbau von nicht-kanonischen Aminosäuren. Da in allen durchgeführten Experimenten fluoreszierende Proteine als Modellproteine verwendet wurden, deren Produktion durch Messung der Fluoreszenz-Intensität visualisiert werden konnte, wird zudem der Begriff der Fluoreszenz näher beschrieben. Die Massenspektrometrie war eine essenzielle Methode für den Nachweis von synthetisiertem Protein mit eingebautem Hydroxylysin, weshalb diese Methode in einem Abschnitt gesondert beschrieben wird.

2.1 Nukleinsäuren

Da Nukleinsäuren eine wichtige Rolle in der Proteinbiosynthese spielen, soll an dieser Stelle zunächst der Aufbau der beiden Nukleinsäuren DNA und RNA beschrieben werden^{3,66}. Die Grundbausteine der DNA sind die Desoxyribonukleotide, die der RNA Ribonukleotide (siehe Abbildung 1)⁶⁷⁻⁶⁹.

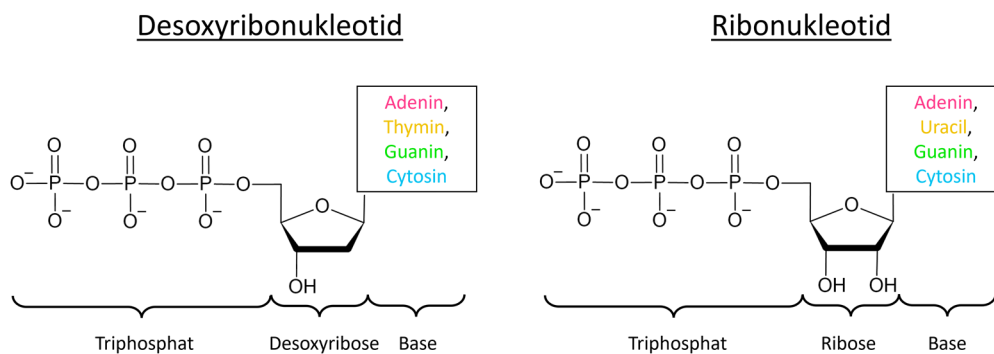


Abbildung 1 Allgemeiner Aufbau eines Desoxyribonukleotids (links) und eines Ribonukleotids (rechts). Beide Nukleotide sind aus einem Triphosphat, einem Zucker (Desoxyribose bei Desoxyribonukleotiden und Ribose bei Ribonukleotiden) sowie einer von vier Basen (Adenin, Guanin, Cytosin, Thymin nur bei Desoxyribonukleotiden, Uracil nur bei Ribonukleotiden) aufgebaut. Das Triphosphat am 5'-C-Atom des Zuckers sowie die Hydroxygruppe am 3'-C-Atom des Zuckers sind für die Polymerisation der (Desoxy)ribonukleotide entscheidend^{70,71}.

Durch Polymerisation dieser (Desoxy)ribonukleotide entstehen Einzelstränge an DNA bzw. RNA (siehe Abbildung 2)^{62,70,72,73}. Die Abfolge der einzelnen (Desoxy)ribonukleotide wird als Primärstruktur bezeichnet⁷⁰.

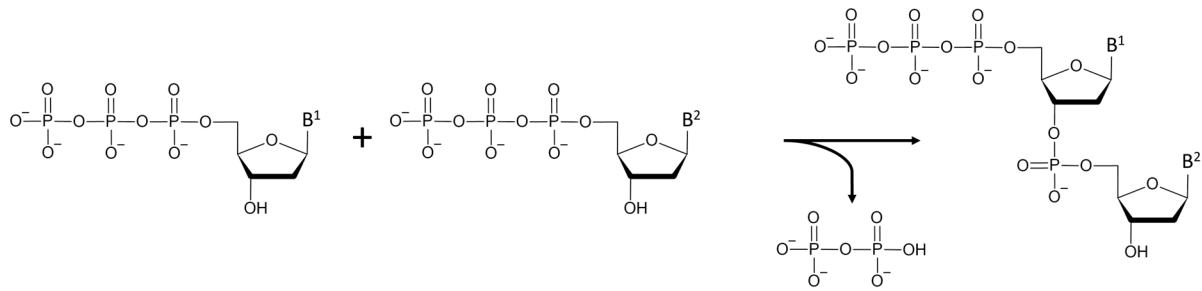


Abbildung 2 Polymerisation von Desoxyribonucleotiden. Unter Pyrophosphat-Abspaltung entsteht zwischen zwei Desoxyribonucleotiden eine Phosphodiesterbindung^{62,70,72,73}. Polymerisieren mehrere Desoxyribonucleotide, so entsteht eine alternierende Abfolge aus Desoxyribose (mit anhängender Base B) und Phosphat. Diese wird als Zucker-Phosphat-Rückgrat bezeichnet, wobei die Phosphate der DNA ihre negative Ladung verleihen⁷⁴. Die Polymerisation von Ribonucleotiden verläuft analog.

Während die RNA im Allgemeinen einzelsträngig vorliegt, formen zwei antiparallele, komplementäre DNA-Einzelstränge in der Regel eine rechtsgängige Doppelhelix mit einer Ganghöhe von 3,4 nm (10,4 - 10,5 bp) und einem Durchmesser von ca. 2 nm (B-Konformation)^{66,67,71,75}. Hierbei entstehen eine große (2,2 nm) sowie eine kleine Furche (1,2 nm)⁶⁷. Neben der B-Konformation können auch andere Konformationen ausgebildet werden, wie z. B. die A- oder Z-Konformation^{66,67}. Für die Paarung beider Stränge sind Wasserstoffbrückenbindungen, welche sich zwischen den komplementären Basen ausbilden, verantwortlich^{66,76}. Hierbei werden zwischen Adenin und Thymin zwei und zwischen Guanin und Cytosin drei Wasserstoffbrücken ausgebildet (siehe Abbildung 3)^{66,67,76}. Des Weiteren wechselwirken benachbarte Basen durch so genannte „Stacking“-Interaktionen, wie z. B. sterische Abstoßungen oder π - π Wechselwirkungen, was eine Stabilisierung der helikalen Struktur bewirkt^{76,77}.

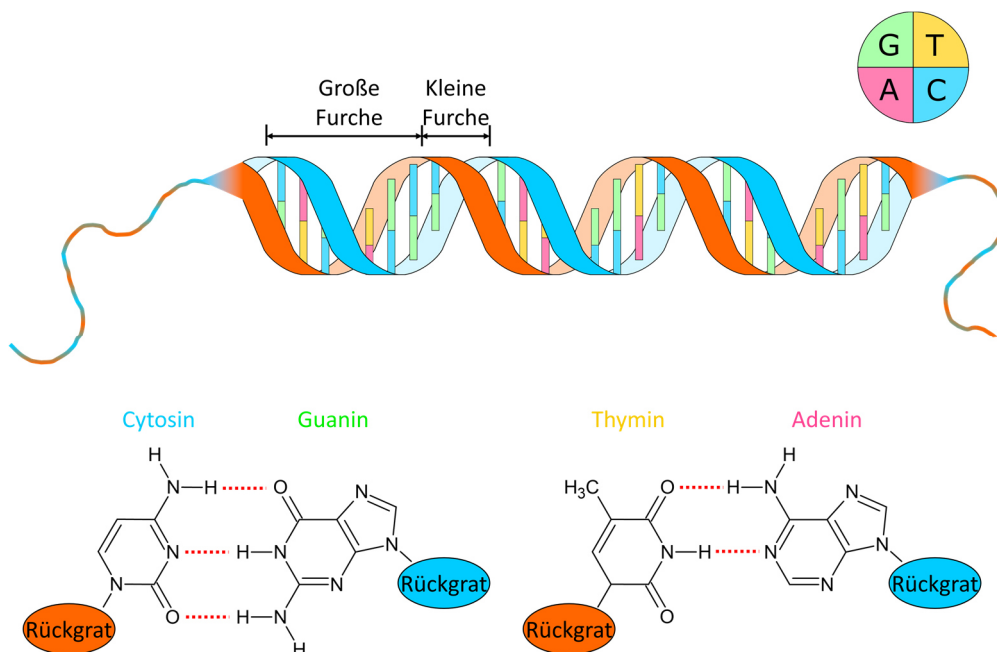


Abbildung 3 Ausbildung der B-Konformation der DNA. Durch Wasserstoffbrückenbindungen paaren sich Cytosin und Guanin (drei Wasserstoffbrückenbindungen) sowie Thymin und Adenin (zwei Wasserstoffbrückenbindungen). Diese Wechselwirkungen bewirken zusammen mit „Stacking“-Interaktionen die Ausbildung einer DNA-Doppelhelix, welche über eine große und eine kleine Furche verfügt.

2 Grundlagen

Durch die helikale Struktur der DNA sowie durch die stark negative Ladung des Zucker-Phosphat-Rückgrats werden Interaktionen mit anderen Komponenten erleichtert bzw. ermöglicht ^{74,78,79}.

So spielt die helikale Struktur bzw. das Vorhandensein der großen und der kleinen Furche eine maßgebliche Rolle bei der Bindung diverser DNA-Bindeproteine an die DNA ⁷⁹⁻⁸². Beispiele sind Helix-Helix Interaktionen verschiedener Transkriptions-Regulatoren ^{79,80,82,83}. Des Weiteren ermöglicht die negative Ladung der DNA Interaktionen mit Proteinen, welche über positiv geladene Areale (durch eine hohe Anzahl an den positiven Aminosäuren Arginin, Histidin oder Lysin) verfügen ^{63,64,78,84,85}. Ein prominentes Beispiel sind Histone, welche über eine Vielzahl an positiven Aminosäuren verfügen und maßgeblich an der Kondensation der DNA beteiligt sind ^{78,86}.

Mittels spezifischer Enzyme können die DNA-ständigen Basen Adenin sowie Cytosin methyliert werden (siehe Abbildung 4), was eine Auswirkung auf die Proteinbiosynthese haben kann ^{30,42,87-91}.

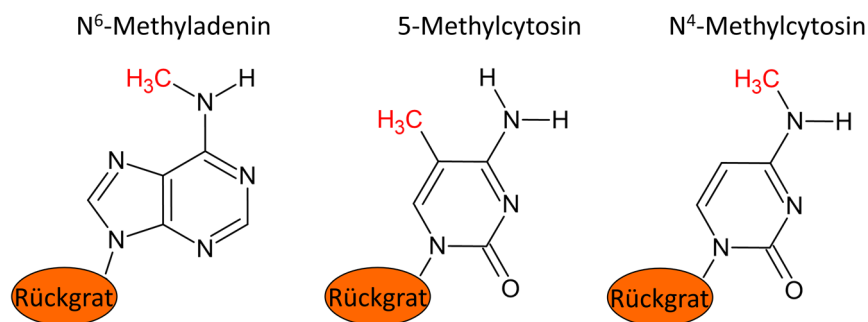


Abbildung 4 Mögliche methylierte DNA-Basen. N⁶-Methyladenin kommt vornehmlich in Bakterien vor und ist in diesen in verschiedene Prozesse, wie z. B. DNA-Replikation, DNA-Reparatur, Transkriptionsregulation etc., involviert ^{41,92}. Dennoch konnte die Präsenz dieser modifizierten Base auch in Eukaryoten nachgewiesen werden, wobei deren Funktion in diesen Organismen weitgehend unbekannt ist ^{41,90}. Im Gegensatz dazu kommt 5-Methylcytosin vermehrt in Eukaryoten vor und erfüllt dort Funktionen bzgl. der Genexpression, Zelldifferentiation etc. ^{88,89}. In Prokaryoten ist diese methylierte Base eher selten vorhanden, wobei sie dennoch in die Genregulation involviert ist ^{42,92}. N⁴-Methylcytosin ist eine weitere methylierte Base, welche nur in Bakterien vorhanden ist ^{88,90}. Diese modifizierte Base erfüllt Funktionen bzgl. der DNA-Replikation, DNA-Reparatur, Genexpression etc. ⁸⁸. Methylierte DNA-Basen sind somit in eine Vielzahl an Prozessen involviert und in einer nicht vernachlässigbaren Menge in den Genomen von Organismen vorhanden. So liegen ca. 1,5 % aller Adenine und ca. 0,75% aller Cytosine im *E. coli* Genom methyliert vor ⁹².

Sowohl bei der DNA als auch bei der RNA können sich Sekundärstrukturen durch Wechselwirkungen der Basen innerhalb eines Stranges ausbilden ⁹³⁻⁹⁶. Hierbei paaren sich komplementäre Basen, weshalb die Ausbildung von Sekundärstrukturen z. B. durch palindromische Sequenzen gefördert wird ^{70,97}. Dies ermöglicht die Ausbildung verschiedener Sekundärstrukturen, wie z. B. einer Haarnadelstruktur oder einer kreuzförmigen DNA-Konformation (siehe Abbildung 5) ^{95,96,98}.

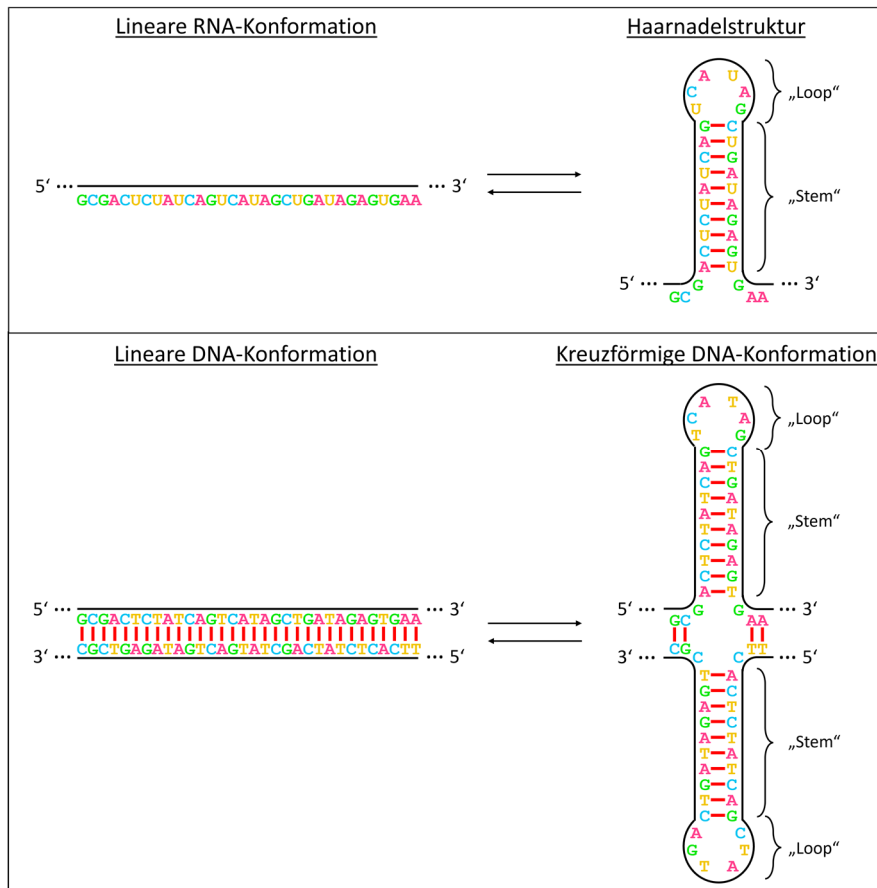


Abbildung 5 Mögliche Sekundärstrukturen von RNA (oben) und DNA (unten; zur besseren Überschaubarkeit wird die DNA nicht helikal dargestellt). Durch intramolekulare Basenpaarungen kann RNA so genannte Haarnadelstrukturen als Alternative zur linearen RNA-Konformation ausbilden. Hierfür erfordert es selbstkomplementäre Sequenzen. Bei Ausbildung einer Haarnadelstruktur paaren sich die selbstkomplementären Basen zu einem so genannten „Stem“. Liegt zwischen den beiden komplementären Teilsequenzen eine nicht selbstkomplementäre Sequenz, resultiert daraus ein ungepaarter Bereich, welcher als „Loop“ bezeichnet wird. Solche „Loops“ können bei anderen möglichen Sekundärstrukturen auch innerhalb eines „Stems“ liegen. Bezüglich der DNA können beide Einzelstränge analoge Sekundärstrukturen ausbilden. Daraus resultiert eine kreuzförmige DNA-Konformation als mögliche Alternative zur linearen DNA-Konformation^{95–98}.

Sekundärstrukturen spielen eine wichtige Rolle in der Proteinbiosynthese^{95,96}. So ist die Ausbildung von Haarnadelstrukturen für die Funktionalität von tRNAs und Ribosomen essenziell^{94,99,100}. Des Weiteren können solche Strukturen die Stabilität der Nukleinsäure sowie deren Interaktion mit verschiedenen Proteinen beeinflussen und somit die Genexpression regulieren^{95,96,98}.

2.2 Proteine

Das Produkt der Proteinbiosynthese sind Proteine³. Deren Grundbausteine sind die L-Aminosäuren (siehe Abbildung 6)¹⁰¹. Neben der L-Konfiguration können Aminosäuren auch in der D-Konfiguration vorliegen^{101,102}. Da solche aber in dieser Dissertation keine Rolle spielen, wird auf D-Aminosäuren nicht näher eingegangen und fortan der Begriff Aminosäuren als Synonym für L-Aminosäuren verwendet.

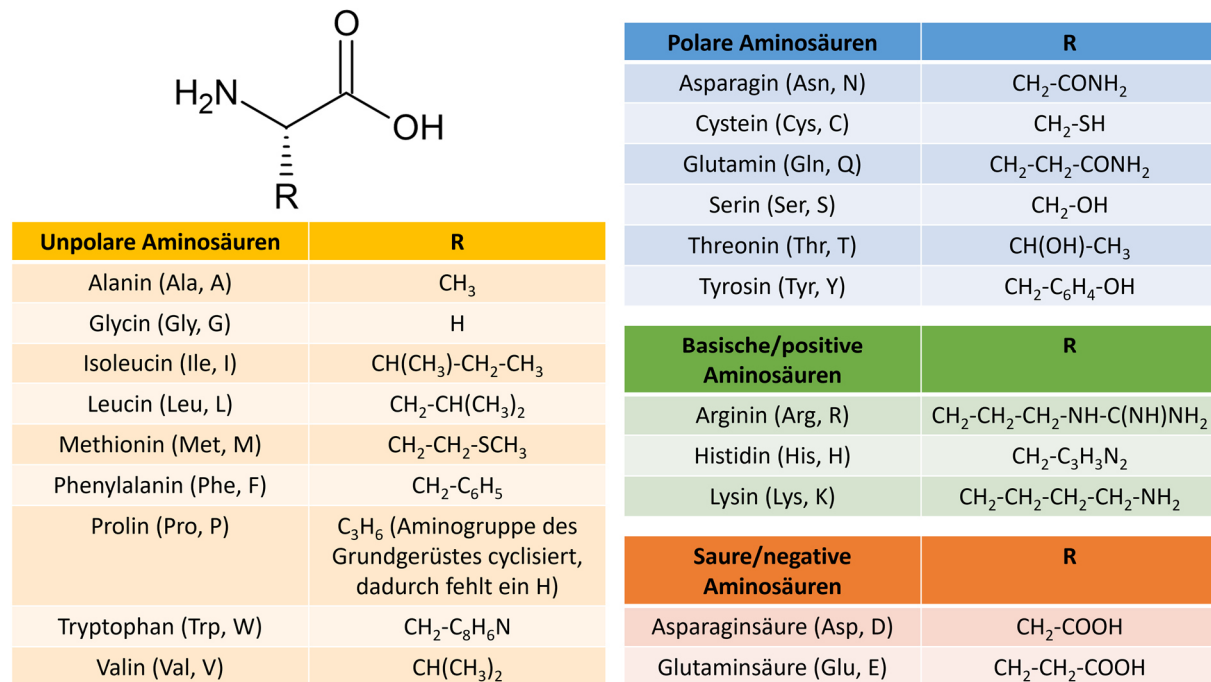


Abbildung 6 Allgemeine Struktur der L-Aminosäuren (links oben). Diese bestehen aus einem zentralen Kohlenstoffatom, an welches eine Aminogruppe, eine Carboxygruppe, ein Wasserstoffatom sowie eine Restgruppe gebunden ist. Die Restgruppe bestimmt die chemischen Eigenschaften der Aminosäure und ist ein Unterscheidungsmerkmal für die Art der Aminosäure. Somit können die 20 kanonischen Aminosäuren in verschiedene Gruppen unterteilt werden (links unten und rechts)¹⁰²⁻¹⁰⁵.

Wie in Abbildung 6 zu sehen, besitzen alle Aminosäuren das gleiche Grundgerüst, wobei die Restgruppe die Art der Aminosäure und deren chemische Eigenschaften (Polarität, Ladung, pKs-Wert) bestimmt^{84,102,103,106}.

Während lediglich 20 Aminosäuren bei der natürlichen Proteinbiosynthese direkt in Proteine eingebaut werden, können weitere Aminosäuren in Proteinen vorhanden sein^{57,102,107-109}. Erstere werden als kanonische Aminosäuren bezeichnet^{57,109}. Die weiteren Aminosäuren, welche in Proteinen vorhanden sein können, entstehen prinzipiell durch posttranslationale Modifikationen der kanonischen Aminosäuren, wie z. B. Glykosylierung, Acetylierung, Phosphorylierung, Hydroxylierung etc.^{59,102,110}. Einen Sonderfall stellen die Aminosäuren Selenocystein und Pyrrolysin dar, welche von speziellen Organismen auch im Rahmen der natürlichen Proteinbiosynthese direkt in Proteine eingebaut werden und somit als spezielle kanonische Aminosäuren klassifiziert werden können¹⁰⁷⁻¹⁰⁹.

Proteine entstehen durch die Verknüpfung von Aminosäuren unter Ausbildung einer Peptidbindung (siehe Abbildung 7) ^{111,112}. Die Aminosäuresequenz eines Proteins wird als Primärstruktur bezeichnet ^{105,112}.

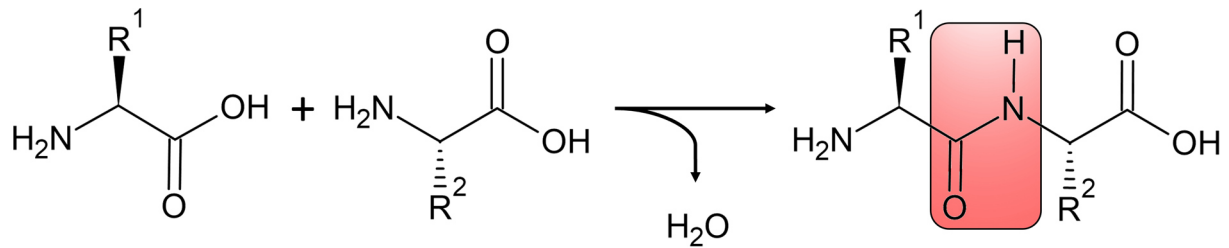


Abbildung 7 Polymerisation von Aminosäuren. Unter Wasserabspaltung entsteht zwischen zwei Aminosäuren eine Peptidbindung (rot). Die endständige Carboxygruppe des entstandenen Dipeptids fungiert als Ansatzpunkt für die Polymerisation weiterer Aminosäuren ^{112,113}. Diese Reaktion kann vielfach stattfinden, wobei Aminosäureketten entstehen, die in Abhängigkeit ihrer Länge als Peptide oder Proteine ($n(\text{Aminosäuren}) > \sim 100$) bezeichnet werden ¹¹⁴.

Durch die unterschiedlichen chemischen Eigenschaften der in Proteine eingebauten Aminosäuren sind diese maßgeblich an den chemischen Eigenschaften des gesamten Proteins beteiligt ^{84,85,104,106,109}. Des Weiteren weisen bestimmte Aminosäuresequenzen hohe Bindungsaffinitäten für spezifische Enzyme bzw. Proteine oder Substanzen auf ^{115,116}. Beispiele sind Streptavidin, welches hochspezifisch an Biotin binden kann und eine der stärksten nicht-kovalenten Bindungen ausbildet oder der Polyhistidin-Tag (His-Tag), welcher eine hohe Affinität zu verschiedenen Metallionen aufweist ¹¹⁶⁻¹¹⁹.

Durch intramolekulare Wechselwirkungen der Aminosäuren in einem Peptid bzw. Protein können sich verschiedene Sekundärstrukturen (α -Helices und β -Faltblätter durch Wasserstoffbrückenbindungen) und Tertiärstrukturen (korrekte räumliche Faltung durch beispielsweise die Oxidation von zwei Cysteinen zu Cystin, woraus eine quervernetzende Disulfidbrücke resultiert) ausbilden ^{105,112,120,121}. Lagern sich mehrere Proteine zu einem geordneten Proteinkomplex zusammen, wird dies als Quartärstruktur bezeichnet ^{105,112}.

2.3 Proteinbiosynthese

Der Prozess, bei dem ein Protein in einem biologischen System synthetisiert wird, nennt sich Proteinbiosynthese. Die Aminosäuresequenz des zu synthetisierenden Proteins wird durch einen definierten Abschnitt auf der DNA (Gen) codiert, weshalb die Synthese auch als Genexpression bezeichnet werden kann ^{3,11}. Dennoch benötigt die Synthese die Einbeziehung einer mRNA ³. Aus diesem Grund kann die Proteinbiosynthese in zwei essenzielle Schritte untergliedert werden: Die Transkription und die Translation ^{3,67}.

Zunächst erfolgt die Transkription. Bei diesem Prozess fungiert zunächst ein DNA-Konstrukt als Informationsträger, welcher das Gen des zu synthetisierenden Proteins enthält ⁶⁷. Von diesem wird im Rahmen der Transkription eine mRNA-Kopie angefertigt ³. Damit eine Transkription des Gens stattfinden kann, werden weitere Elemente auf dem DNA-Konstrukt benötigt ¹²²⁻¹²⁴. Für die Initiation der Transkription ist ein so genannter Promotor essenziell ^{30,125}. Ein Promotor fungiert als Bindestelle für die RNA-Polymerase und eine Vielzahl an Transkriptions-Regulatoren ¹²⁵⁻¹²⁷. Letztere können neben dem Promotor auch an Sequenzen in der Nähe des Promotors oder direkt an die RNA-Polymerase binden und somit die Bindung zwischen Promotor und RNA-Polymerase verstärken (Aktivator) oder abschwächen (Repressor) ^{30,128-130}. Je nach Organismus existieren diverse Promotoren, RNA-Polymerasen und Transkriptions-Regulatoren ^{125,129}.

Bindet die RNA-Polymerase an den Promotor, so wird die helikale Struktur der DNA stromabwärts des Promotors aufgelöst ^{122,124,131}. Die DNA wird in zwei Stränge, den codogenen Strang sowie den nichtcodogenen Strang aufgeteilt, wobei letzterer die genetische Information der zu transkribierenden mRNA darstellt, während der codogene Strang als Template für die RNA-Polymerase fungiert ^{70,124}. Die RNA-Polymerase wandert entlang des codogenen Strangs von 3'- in 5'-Richtung ⁷⁰. Hierbei wird eine mRNA des zu transkribierenden Gens durch Anhängen komplementärer Nukleotide von 5'- in 3'-Richtung synthetisiert (Ribonukleotide, Uracil fungiert als RNA-spezifisches Analogon für Thymin) ^{70,131}. Nachdem das komplette Gen transkribiert wurde, erfordert es für die Transkriptions-Beendigung einen so genannten Terminator ^{124,132,133}. Dieser veranlasst eine Beendigung der Transkription durch Dissoziation der RNA-Polymerase von der DNA und der Freisetzung der synthetisierten mRNA ¹³³.

Während bei Eukaryoten vielfältige Prozessierungsschritte der mRNA vor Durchführung der Translation notwendig sind, kann die synthetisierte mRNA bei Prokaryoten direkt als Template für die Translation fungieren ¹³⁴⁻¹³⁶. Da im Rahmen dieser Dissertation lediglich prokaryotische Expressionssysteme verwendet wurden, wird auf die eukaryotische mRNA-Prozessierung und die darauffolgende eukaryotische Translation nicht näher eingegangen.

Im zweiten Schritt der Proteinbiosynthese, der Translation, fungiert die synthetisierte mRNA als Template für die Aminosäuresequenz des zu synthetisierenden Proteins ^{3,67}. Im Rahmen der Translation findet eine Interaktion zwischen der synthetisierten mRNA und den Ribosomen bzw. deren

2 Grundlagen

die in der A-Stelle befindliche Aminosäure angeknüpft^{143,144}. Anschließend wird durch das Enzym Peptid-Deformylase die Formylgruppe vom Formylmethionin des entstandenen Dipeptids abgespalten^{3,154,155}. Der Elongationsfaktor G (EF-G) bewirkt, dass das Ribosom entlang der mRNA weiterwandert, wodurch die „leere“ Formylmethionyl-tRNA an die E-Stelle gelangt und von der mRNA dissoziiert^{149,150}. Die tRNA mit dem angehängten Dipeptid gelangt wiederum an die P-Stelle, was die Bindung einer weiteren, passend beladenen tRNA an die A-Stelle ermöglicht, an deren Aminosäure in einem nächsten Schritt das Dipeptid gebunden wird¹⁴⁹. Dieser Prozess wiederholt sich so lange, bis das Basentriplett in der A-Stelle durch die Sequenzen „UAA“, „UAG“ oder „UGA“ gegeben ist¹⁵⁰. Diese so genannten Stopcodons, für die keine passende tRNA vorhanden ist, fungieren als Bindestelle für Release-Faktoren (RF), welche eine Dissoziation des Ribosoms von der mRNA sowie des synthetisierten Proteins von der letzten an die mRNA gebundenen tRNA veranlassen^{59,150,156}.

Abbildung 9 liefert ein Schema der zuvor beschriebenen Prozesse der Proteinbiosynthese.

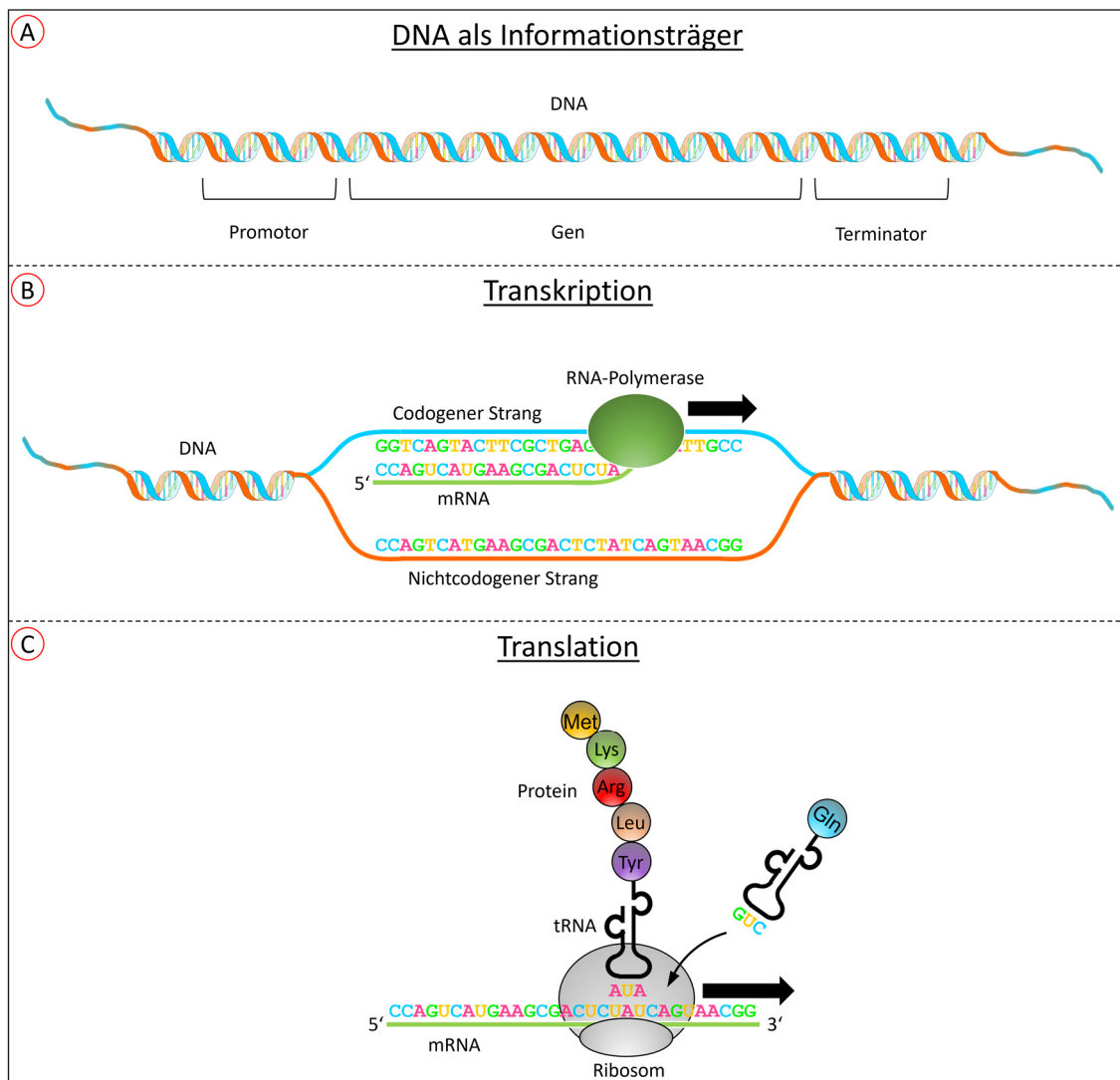


Abbildung 9 Graphische Darstellung der Proteinbiosynthese. Die Grundvoraussetzung ist das Vorhandensein einer geeigneten DNA-Sequenz, welche die Information eines zu exprimierenden Gens enthält (A). Im Rahmen der Transkription entsteht eine mRNA des Gens (B), welche die Synthese des codierten Proteins durch den Prozess der Translation veranlasst (C).

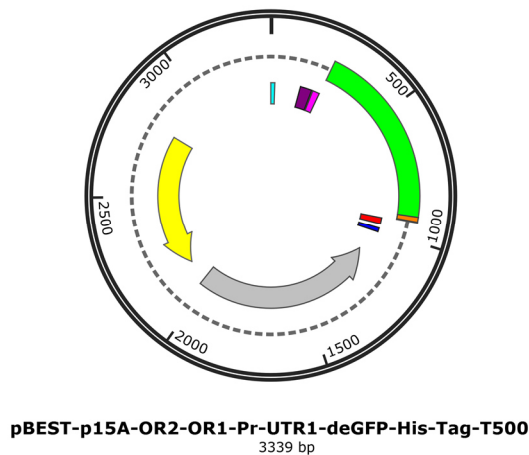
2.4 Expressionssysteme

Neben der zuvor beschriebenen Proteinbiosynthese können Proteine auch chemisch hergestellt werden¹⁵⁷⁻¹⁵⁹. Die Synthese rekombinanter Proteine kann jedoch besser kontrolliert und durchgeführt werden^{158,160}. Des Weiteren entstehen bei der chemischen Synthese unter Umständen ungewollte Nebenprodukte¹⁵⁸. Aus diesem Grund ist die Proteinbiosynthese zur Erzeugung rekombinanter Proteine in vielen Situationen die präferierte Methode^{158,160,161}.

Soll eine Proteinbiosynthese in einem Experiment durchgeführt werden, geschieht dies im Rahmen einer Genexpressionsreaktion¹¹⁻¹⁴. Für die Durchführung einer Genexpressionsreaktion existiert eine Vielzahl an möglichen Expressionssystemen. Grob können diese Expressionssysteme in *in vivo* Systeme und in *in vitro* Systeme unterteilt werden^{10,11}.

In vivo Expressionssysteme umfassen lebende Organismen. Klassische Modellorganismen sind z. B. die Prokaryoten *Escherichia coli* und *Bacillus subtilis* oder der Eukaryot *Saccharomyces cerevisiae*^{14,162,163}. Wird in diese Organismen DNA eingebracht, welche eine Genexpression durch die endogene Transkriptions-Translations-Maschinerie zulässt, veranlasst dies die Synthese rekombinanter Proteine^{16,59,164,165}. Zu diesem Zweck werden meist Plasmide verwendet^{166,167}. Plasmide sind zirkuläre, autonom replizierende DNA-Konstrukte, welche durch verschiedene Methoden in Organismen eingebracht werden können^{168,169}. Neben einem Selektionsmarker, wie z. B. einer Antibiotika-Resistenz, mit dem die Präsenz des Plasmids im Organismus gewährleistet werden kann, enthält ein Plasmid einen Replikationsursprung (Ori), der dessen Replikation ermöglicht^{166,167,169}. Sofern das Plasmid für eine Genexpressionsreaktion verwendet werden soll, enthält dieses zusätzlich ein Gen und die für dessen Expression benötigten regulatorischen Elemente^{163,170,171}. Hierbei muss darauf geachtet werden, dass die endogene Transkriptions-Translations-Maschinerie diese regulatorischen Elemente erkennt^{17,172}. Abbildung 10 zeigt ein Beispiel für ein Plasmid²⁸, welches in dieser Dissertation eine wichtige Rolle einnimmt.

2 Grundlagen



Element	Funktion	Farbe
Seq1	Bindestelle für Sequenzierungsprimer Seq1	Türkis
Or2-Or1-Pr	Promotor	Violett
UTR1	Untranslatierte Region als Teil der ribosomalen Bindestelle	Pink
eGFP-Del6-229 (deGFP)	<i>degfp</i> -Gen	Grün
His-Tag	Erlaubt His-Tag-Purifikation von deGFP	Orange
T500	Terminator	Rot
Seq4	Bindestelle für Sequenzierungsprimer Seq4	Blau
p15A ori	Replikationsursprung	Grau
AmpR	Vermittelt Ampicillin-Resistenz	Gelb

Abbildung 10 Plasmidstruktur des Plasmids „pBEST-p15A-OR2-OR1-Pr-UTR1-deGFP-His-Tag-T500“²⁸. Die Grafik zeigt den zirkulären Aufbau des Plasmids mit allen vorhandenen Elementen*. Auf die einzelnen Elemente sowie deren Funktionen wird in der dargestellten Tabelle eingegangen.

Dennoch unterliegen *in vivo* Expressionssysteme diversen Einschränkungen^{11,12}. So sind die Konzentrationen verschiedener biochemischer Komponenten sowie verschiedene biochemische Parameter in einem Organismus schwer zu bestimmen oder gar unbekannt^{23–25,173}. Des Weiteren können sich bestimmte Substanzen, welche für die Genexpression benötigt werden, negativ auf die Reproduktions- oder Lebensfähigkeit eines Organismus auswirken, wodurch eine Genexpressionsreaktion unter Umständen nicht erfolgreich durchgeführt werden kann^{19–22}. Ein Beispiel ist die nicht-kanonische Aminosäure Canavanin, deren Einbau in Proteine von besonderem Interesse ist^{57,174}. In Organismen ist diese Substanz aber jedoch toxisch¹⁹.

Die Verwendung von *in vitro* Expressionssystemen ermöglicht eine höhere Flexibilität, da es sich hierbei um keinen lebenden Organismus handelt und somit einige der zuvor genannten Einschränkungen für solche Systeme nicht zutreffen. *In vitro* Expressionssysteme sind biochemische Gemische, in denen die Transkriptions-Translations-Maschinerie eines Organismus vorhanden ist und welche lediglich auf die Genexpression spezialisiert sind^{22,175}. Diverse Komponenten können dem *in vitro* Expressionssystem in definierten Konzentrationen hinzugegeben werden^{24,29,176}. Des Weiteren erlaubt ein solches System eine bessere Kontrolle diverser Parameter sowie die Präsenz bestimmter Substanzen, die für einen lebenden Organismus toxisch sind^{13,23–25}.

Für eine Genexpressionsreaktion muss zu einem *in vitro* Expressionssystem zunächst ein geeignetes DNA-Konstrukt, welches von der Transkriptions-Translations-Maschinerie akzeptiert wird, hinzugegeben werden^{29,177}. Neben der zuvor erwähnten Form als Plasmid, erlauben *in vitro*

* Für die Generierung der Grafik wurde der SnapGene Viewer (SnapGene software (von Insightful Science; verfügbar auf snapgene.com)) verwendet.

Expressionssysteme auch die Verwendung alternativer DNA-Konstrukte als DNA-Template für die Genexpression, wie z. B. linearisierte DNA^{21,29,178,179}. In *in vivo* Systemen würde eine solche DNA durch spezielle Desoxyribonukleasen als fremd erkannt und verdaut werden^{180–182}. Diese Enzyme können auch in einem *in vitro* Expressionssystem (auf Grund dessen Präparation) vorhanden sein^{183,184}. Somit verlangt die Verwendung linearer DNA als DNA-Template für eine *in vitro* Genexpressionsreaktion unter Umständen die Hinzugabe von speziellen Desoxyribonukleasen-Inhibitoren, wie z. B. GamS, welches RecBCD hemmt und folglich die Degradation linearer DNA verhindert^{12,178,179,185}. Neben der DNA müssen dem Expressionssystem auch Aminosäuren, welche in Proteine eingebaut werden, und ggf. Pufferlösungen, die eine optimale Genexpression ermöglichen, hinzugefügt werden^{29,177}. Da die Zusammensetzung der Aminosäuren sowie deren Konzentrationen frei wählbar sind, bietet ein *in vitro* Expressionssystem eine optimale Voraussetzung für den Einbau nicht-kanonischer Aminosäuren in rekombinante Proteine^{57,174,186}.

Die beste Kontrolle einer Genexpressionsreaktion kann durch die Verwendung des so genannten „Protein synthesis using recombinant elements“ (PURE)-Systems erreicht werden^{187,188}. Dieses von Ueda *et al.*^{187,188} entwickelte System enthält lediglich die für die Genexpression notwendigen Komponenten, welche zuvor einzeln rekombinant produziert und purifiziert wurden^{22,187,189}. Hieraus resultiert eine definierte biochemische Umgebung des Expressionssystems. Dennoch ist dieses System auf einer Transkription durch die RNA-Polymerase des Bakteriophagen T7 aufgebaut¹⁸⁷, was eine Einschränkung bzgl. der möglichen Regulationsmechanismen für die Transkription bedeutet^{24,177}. Zudem lassen sich aus dem PURE-System gewonnene Erkenntnisse nur begrenzt auf die *in vivo* Situation übertragen, da dieses ein minimalistisches System darstellt^{22,189}. Diverse in einem Organismus vorhandene Komponenten, welche die Genexpression beeinflussen können, sind in einem solchen System nicht präsent. Des Weiteren ist das PURE-System auf Grund der aufwendigen Präparation sehr kostenintensiv^{22,189}.

Einen Lösungsansatz bieten *in vitro* Expressionssysteme, welche auf Zell-Lysaten basieren. Diese können als Bindeglied zwischen *in vivo* Expressionssystemen und dem PURE-System gesehen werden¹⁷⁷. Neben einer kostengünstigeren Produktion^{22,186,189} besitzen solche Systeme eine biochemische Umgebung, die der innerhalb der zugrunde liegenden Zellen entspricht^{29,175,183}. Somit lassen sich experimentelle Ergebnisse besser auf die *in vivo* Situation übertragen^{175,183}. Dies ermöglicht eine genauere Untersuchung natürlicher Genregulationsmechanismen innerhalb lebender Organismen. Dennoch muss berücksichtigt werden, dass in einem auf Zell-Lysat basierenden Expressionssystem auch Komponenten vorhanden sein können, welche die Genexpression nachteilig beeinflussen (wie z. B. RecBCD, was die Hinzugabe von GamS bei Verwendung linearer DNA als DNA-Template erfordert^{12,178,179,183–185}). Abbildung 11 zeigt schematisch die Präparation eines auf *E. coli* basierenden, zellfreien Expressionssystems sowie dessen Verwendung für eine Genexpressionsreaktion^{28,29}. Das dargestellte, von Noireaux *et al.*^{28,29} entwickelte *in vitro* Expressionssystem („all *E. coli* cell-free expression system“)

2 Grundlagen

wurde in dieser Dissertation für alle durchgeführten Experimente verwendet, weshalb dieses System im Folgenden näher betrachtet wird.

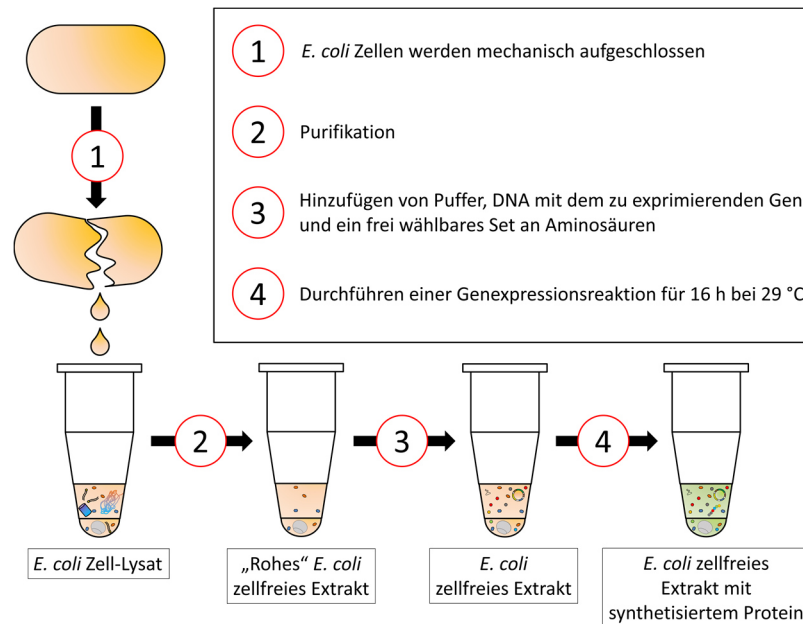


Abbildung 11 Präparation des „all *E. coli* cell-free expression system“ sowie dessen Verwendung für die Durchführung einer Genexpressionsreaktion ^{28,29}.

Für die Präparation des dargestellten „all *E. coli* cell-free expression system“ werden zunächst *E. coli* Zellen mechanisch aufgeschlossen, um ein *E. coli* Zell-Lysat zu erhalten. Da in diesem restliche Zellbestandteile vorhanden sind, wie z. B. genomische DNA, Lipide, Zellwandbruchstücke, diverse Proteine bzw. Enzyme etc., ist eine Purifikation erforderlich. Hierfür werden diverse Zentrifugationsschritte sowie eine Dialyse durchgeführt. Die erhaltene Lösung fungiert als „rohes“ *E. coli* zellfreies Extrakt. Werden diesem Puffer (enthält Energieträger, Nukleotide, Salze etc.), ein geeignetes DNA-Konstrukt mit dem zu exprimierenden Gen sowie Aminosäuren hinzugegeben, so kann eine Genexpressionsreaktion gestartet werden. Während der Genexpressionsreaktion, welche typischerweise über 16 h bei 29 °C abläuft ^{28,29}, wird im *E. coli* zellfreien Extrakt Protein synthetisiert.

Dem Extrakt können auch spezifische Substanzen vor dem Start der Genexpressionsreaktion hinzugefügt werden. Dadurch können diverse Substanzen auf mögliche Genexpressions-assoziierte Effekte untersucht werden. In diesem Zusammenhang kann die Genexpression auch gezielt, z. B. durch Transkriptionsfaktoren, beeinflusst werden. Transkriptionsfaktoren können direkt im Extrakt synthetisiert oder in rekombinanter Form hinzugefügt werden. Somit kann das „all *E. coli* cell-free expression system“ neben der Durchführung von Genexpressionen auch für die genauere Untersuchung von Genregulationsmechanismen verwendet werden. Dennoch ist eine direkte Übertragbarkeit auf die *in vivo* Situation wie bei allen *in vitro* Expressionssystemen nur bedingt möglich. Des Weiteren reagiert das „all *E. coli* cell-free expression system“ in Bezug auf die *in vivo* Situation sensibler auf veränderte Reaktionsbedingungen ^{28,29}.

2.5 Regulation der Genexpression

Da in einem Organismus eine Vielzahl an verschiedenen biochemischen Prozessen stattfinden, müssen diese diversen Regulationsmechanismen unterliegen^{190,191}. Dies gilt auch für die Genexpression. Je nach Situation müssen verschiedene Gene in einem Organismus exprimiert werden^{33,40,122,192}. Zu diesem Zweck existieren diverse Genregulationsmechanismen^{193,194}, welche grob in genetische und epigenetische Mechanismen untergliedert werden können^{33–36}. Bei genetischen Mechanismen bewirkt eine veränderte DNA-Sequenz in regulatorischen DNA-Bereichen eine vererbare Änderung der Genexpression³⁶. Im Gegensatz dazu findet bei epigenetischen Mechanismen eine vererbare Änderung der Genexpression als Folge von DNA-Modifikationen, wie z. B. durch DNA-Methylierung, jedoch ohne Veränderung der DNA-Sequenz statt^{34,36,37}. Im Folgenden sollen verschiedene Möglichkeiten zur Regulation der Genexpression näher beschrieben werden. Hierbei wird auf Genregulationsmechanismen, welche in einem Organismus stattfinden, sowie auf solche, welche in experimentellen Genexpressionsreaktion verwendet werden können, eingegangen.

Die Regulation der Genexpression kann auf verschiedenen Ebenen stattfinden^{193,194}. Zunächst werden Regulationsmechanismen bzgl. der Transkriptionsinitiation betrachtet, wobei die involvierten Elemente je nach Expressionssystem unterschiedlich sein können^{125,129}. Bei der Initiation der Transkription ist die Bindung der RNA-Polymerase an den Promotor von entscheidender Bedeutung^{122,125,195}. Die Affinität zwischen RNA-Polymerase und Promotor wird neben der Promotor-Sequenz maßgeblich durch diverse Proteine, so genannte Transkriptions-Regulatoren, beeinflusst^{82,196–199}. Transkriptions-Regulatoren können direkt an die RNA-Polymerase oder den Promotor binden, oder aber an andere spezifische DNA-Bereiche (Operatoren (Prokaryoten), „Silencer“ oder „Enhancer“ (Eukaryoten))^{30,82,123,128–130}. Weitere Bindungsmöglichkeiten stellen andere, in die Transkription involvierte Proteine, spezielle Sekundärstrukturen der DNA oder methylierte DNA-Basen dar^{95,128,200,201}. Durch die Bindung von Transkriptions-Regulatoren können verschiedene Prozesse stattfinden, die die Veränderung der Affinität zwischen RNA-Polymerase und Promotor bewirken^{82,196}. Mögliche Ansatzpunkte dieser Prozesse sind die Zugänglichkeit des Promotors bzw. der gesamten DNA, sterische Hinderung der RNA-Polymerase oder aber die Veränderung der Anzahl an möglichen Interaktionen^{30,52,53,201,202}. So können Transkriptions-Regulatoren die Affinität zwischen der RNA-Polymerase und dem Promotor erhöhen (Aktivator) oder verringern (Repressor) und somit Gene verstärken oder abschwächen^{30,128–130}. Des Weiteren erlauben solche Regulatoren positive oder negative Rückkopplungen ihrer Gene (siehe Abbildung 12)^{203–205}.

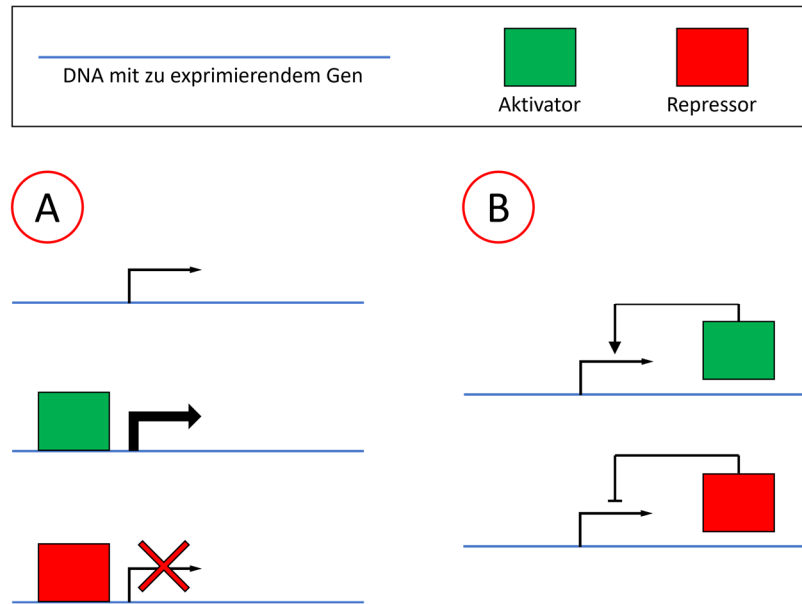


Abbildung 12 Regulation der Genexpression durch Transkriptions-Regulatoren. Während diverse Gene ohne Einwirken von Transkriptions-Regulatoren exprimiert werden, können Aktivatoren die Expression verstärken oder Repressoren die Expression abschwächen (A) ^{30,128–130}. Diesbezüglich existieren Regulationsmechanismen, bei denen die Expression eines Transkriptions-Regulator codierenden Gens durch Präsenz des Transkriptions-Regulators verstärkt (positive Rückkopplung) oder abgeschwächt wird (negative Rückkopplung) (B) ^{203–205}.

Auch spezifische Substanzen (z. B. diverse Saccharide oder Hormone) oder Reize (z. B. Licht) können bei der Transkriptionsinitiation mitwirken ^{18,113,130,206–210}. In diesen Fällen wird von induzierbaren Promotoren gesprochen ^{211,212}. Im Gegensatz dazu werden Promotoren, die unabhängig von externen Signalen eine Genexpression zulassen, als konstitutive Promotoren bezeichnet ^{211,212}. Ein Beispiel bzgl. der induzierbaren Genexpression ist das Lac-Operon ^{213,214}. Dieses ist in *E. coli* Zellen für die Verstoffwechslung von vorhandener Lactose in Abwesenheit von Glucose als Energielieferant verantwortlich ^{129,214}. Ist Lactose vorhanden, so wird diese zunächst in Allolactose umgewandelt, welche wiederum an den Lac-Repressor bindet, der die Genexpression Lactose-verwertender Proteine durch sterische Hinderung reprimiert ^{150,213,215}. Durch die Interaktion zwischen Lac-Repressor und Allolactose ändert sich die Lac-Repressor-Konformation, wodurch eine Dissoziation des Lac-Repressors vom Lac-Operator veranlasst wird ^{213,215,216}. Hierdurch kann die RNA-Polymerase an den Lac-Promotor binden und eine Genexpression der Lactose-verwertenden Proteine veranlassen ¹⁵⁰. Ist die Lactose verstoffwechselt, bindet der Lac-Repressor wieder an den Lac-Operator und reprimiert die Genexpression ²¹⁵.

Die Stabilität der transkribierten mRNA ist eine weitere Ebene, die an der Regulation der Genexpression beteiligt ist, da von dieser abhängt, wie lange die Translation stattfinden kann ^{217–219}. Ribonukleasen veranlassen eine Degradation transkribierter mRNA ^{218–220}. Die Zeit, welche für die Degradation benötigt wird, fungiert als charakteristische Größe für die Stabilität der mRNA, welche u. a. mit deren Sekundärstruktur korreliert ^{221,222}.

Die Sekundärstruktur der mRNA spielt auch bei der Regulation der Genexpression auf Ebene der Translation eine wichtige Rolle^{137,171,221,223}. Diesbezüglich soll die zuvor erwähnte Shine-Dalgarno-Sequenz betrachtet werden, welche in Prokaryoten für die Bindung der mRNA an die Ribosomen von entscheidender Bedeutung ist¹³⁷⁻¹⁴⁰. Durch Konformationsänderungen der mRNA kann die Zugänglichkeit dieser Region und somit die Translationsrate verändert werden^{140,224}. Solche Konformationsänderungen können durch mRNA-Elemente mit regulatorischen Funktionen verursacht werden^{140,224,225}. Beispiele für regulatorische mRNA-Elemente sind Thermosensoren sowie „Riboswitches“. Während bei Thermosensoren eine Konformationsänderung der mRNA als Antwort auf eine Temperaturänderung erfolgt, werden Konformationsänderungen bei „Riboswitches“ durch Bindung spezifischer Substanzen an die Aptamer-Region des „Riboswitches“ induziert^{140,224,225}.

In Experimenten existieren verschiedene Möglichkeiten für die gezielte Regulation der Genexpression²²⁶. So kann durch die Koexpression von spezifischen RNA-Polymerasen bzw. Transkriptions-Regulatoren oder deren Hinzugabe in rekombinanter Form in einer Genexpressionsreaktion die Genexpression reguliert werden^{21,24,29,176}. Eine weitere Möglichkeit bietet die Verwendung induzierbarer Promotoren, welche durch externe Signale (spezifische Substanzen oder Reize) eine Genexpression veranlassen^{18,113,130,206-210}. Diesbezüglich findet das zuvor erwähnte Lac-Operon in der Molekularbiologie eine breite Anwendung. So ermöglicht die Verwendung von IPTG (Isopropyl-β-D-thiogalactopyranosid), welches als Analogon für Lactose fungiert aber nicht verstoffwechselt wird, eine dauerhafte Induktion der Genexpression^{160,227}. Dennoch kann die Genexpression bei Verwendung solcher Konstrukte nicht vollständig abgeschaltet werden, da eine basale Transkription auch in Abwesenheit von Transkriptions-Regulatoren oder in Abwesenheit externer Signale erfolgt^{30,160,206,209,228}.

Die Methode CRISPR/Cas („Clustered Regulatory Interspaced Short Palindromic Repeats/ CRISPR associated proteins“) bietet eine Möglichkeit zur effektiven An- und Abschaltung spezifischer Gene^{35,229-231}. Zu diesem Zweck wird eine crRNA generiert, welche Cas9 zu einer gewünschten DNA-Sequenz dirigiert^{210,230,232}. Dies kann die Inaktivierung eines Gens durch sterische Hinderung der Polymerase oder aber die Aktivierung eines Gens durch an Cas9 gebundene Transkriptions-Aktivatoren bewirken^{35,230}. Des Weiteren können mit CRISPR/Cas gezielt Doppelstrangbrüche eingeführt werden^{231,233}. Wird zusätzlich ein DNA-Konstrukt mit homologen Enden zum Doppelstrangbruch hinzugefügt, kann dieses durch homologe Rekombination in die Bruchstelle eingebaut werden, was die Aktivierung eines Gens bewirken kann²³³. Da die Verwendung von CRISPR/Cas eine Modifikation des Expressionssystems erfordert sowie unter Umständen schwer umsetzbar ist, eignet sich CRISPR/Cas nur bedingt zur Regulation der Genexpression^{232,234-236}.

In vitro Expressionssysteme erlauben weitere Möglichkeiten für die Regulation der Genexpression²²⁶. So kann eine Genexpression lediglich durch Hinzufügen eines geeigneten DNA-Konstrukts gestartet werden^{29,177}. Durch Hinzugabe von Desoxyribonukleasen kann das DNA-Konstrukt verdaut und somit

2 Grundlagen

die Transkription gezielt beendet werden ²³⁷. Ein alternativer Ansatz ist die Transkriptions-Beendigung durch die Hemmung der RNA-Polymerase mit Hilfe spezifischer Substanzen (wie z. B. TagetinTM oder Rifampicin) ^{177,221,238,239}. Sofern direkt eine Translation im Expressionssystem durchgeführt werden soll, kann dem System auch eine geeignete mRNA hinzugefügt werden ^{240,241}. Hierbei muss beachtet werden, dass unter Umständen Ribonukleasen im Expressionssystem vorhanden sein können, die die Translation nach einer bestimmten Zeit durch den Verdau der mRNA beenden ^{24,187,238}. Die Translations-Beendigung kann auch gezielt durch Hinzugabe von Ribonukleasen (wie z. B. MazF) erreicht werden ^{177,221,238,239}. Dennoch verhindern die genannten Methoden zur gezielten Beendigung der Transkription bzw. Translation die Durchführung weiterer Genexpressionsreaktionen im gleichen Expressionssystem, weshalb die Entwicklung neuer Methoden zur regulierten Genexpression von besonderem Interesse ist.

2.6 Einbau nicht-kanonischer Aminosäuren in Proteine

Bei der natürlichen Proteinbiosynthese werden Proteine biologisch hergestellt. Als Grundbausteine solcher Proteine fungieren die 20 kanonischen Aminosäuren^{57,102,107,108}. Obwohl mit diesen eine Vielzahl an Proteinen mit verschiedensten Funktionen hergestellt werden kann, ist durch die begrenzte Anzahl an kanonischen Aminosäuren auch der funktionelle Umfang der herstellbaren Proteine limitiert^{57,59,109,189}. Durch den Einbau von nicht-kanonischen Aminosäuren können Proteine mit neuen Funktionen hergestellt werden^{22,106,174,186,242}.

Für die Synthese von Proteinen mit eingebauten nicht-kanonischen Aminosäuren existieren verschiedene Möglichkeiten. Wie bereits in Abschnitt 2.4 erwähnt, gestaltet sich die chemische Synthese von Proteinen als schwierig^{157–159}, weshalb die Proteinbiosynthese zur Erzeugung von rekombinanten Proteinen in vielen Situationen bevorzugt wird^{158,160,161}. Dies gilt auch für die Synthese von Proteinen mit eingebauten nicht-kanonischen Aminosäuren¹⁸⁹.

Um den Einbau von nicht-kanonischen Aminosäuren in rekombinante Proteine durchzuführen, existieren zwei verschiedene Methoden^{22,57,106}. Die erste Methode ist der ortsspezifische Einbau, bei dem die nicht-kanonische Aminosäure z. B. durch Suppression eines Stopcodons während der Translation in die wachsende Aminosäurekette eingebaut wird (siehe Abschnitt 2.6.1)^{109,243}. Die zweite Methode ist der seitenkettenspezifische Einbau, bei dem die nicht-kanonische Aminosäure durch Abwesenheit der zugehörigen kanonischen Aminosäure anstelle dieser in die wachsende Aminosäurekette eingebaut wird (siehe Abschnitt 2.6.2)^{106,109}. Somit können mit der gleichen DNA bzw. deren transkribierter mRNA prinzipiell verschiedene Proteine synthetisiert werden (siehe Abbildung 13).

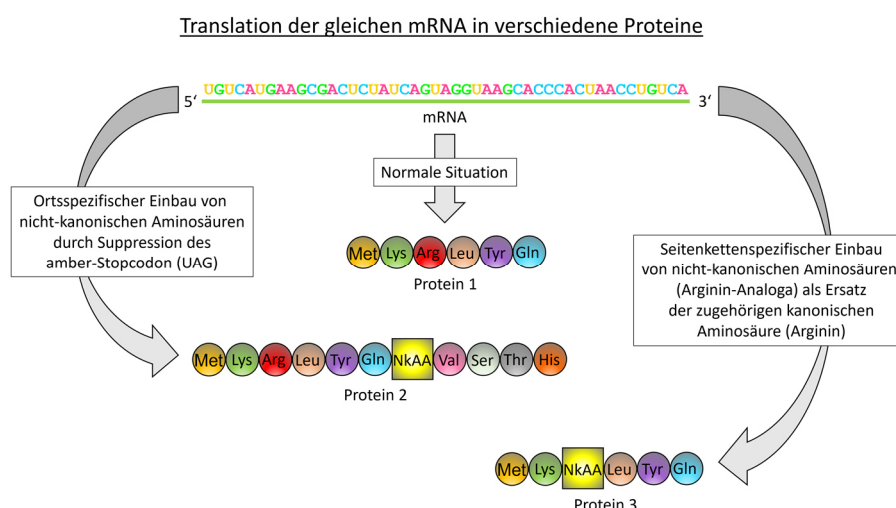


Abbildung 13 Translation einer spezifischen mRNA in verschiedene Proteine. Während eine codierende mRNA unter normalen Bedingungen lediglich die Synthese eines Proteins veranlasst, können durch den ortsspezifischen oder den seitenkettenspezifischen Einbau nicht-kanonischer Aminosäuren (NkAA) andere Proteine aus der gleichen mRNA translatiert werden. Dargestellt ist die Suppression des amber-Stopcodons (ortsspezifischer Einbau) sowie der Einbau eines Arginin-Analogons als Arginin-Ersatz (seitenkettenspezifischer Einbau).

2 Grundlagen

Unabhängig von der Methode zum Einbau von nicht-kanonischen Aminosäuren in Proteine müssen diese zunächst an eine geeignete tRNA gebunden werden^{2,59,109}. Dies geschieht über die so genannten Aminoacyl-tRNA-Synthetasen (AaRSs), welche spezifische Aminosäuren erkennen und an die zur AaRS zugehörige tRNA binden^{2,59,244}. Hierbei wird durch Prozesse, wie z. B. „Kinetic Proofreading“ (siehe Abschnitt 2.7), eine hohe Spezifität bei der Beladung der tRNA mit Aminosäuren erreicht^{54,56,58,61,62}. Dennoch können nicht-kanonische Aminosäuren durch strukturelle Nähe zur zugehörigen kanonischen Aminosäure von den nativen AaRSs erkannt und an tRNA gebunden werden, was einen Einbau in Proteine ermöglicht (seitenkettenspezifischer Einbau)^{22,57,189,245,246}. Ist eine Erkennung der einzubauenden nicht-kanonischen Aminosäure durch die nativen AaRSs nicht gegeben, so erfordert deren Einbau in Proteine ein geeignetes zusätzliches, nicht-natives (orthogonales) Paar aus AaRS und tRNA (ortsspezifischer Einbau)^{106,186,189}.

Eine weitere wichtige Rolle beim Einbau von nicht-kanonischen Aminosäuren in Proteine spielt der in Abschnitt 2.3 erwähnte EF-Tu^{60,153,186}. Dieser weist neben der Funktion der Förderung des Eintritts von Aminoacyl-tRNAs in die Ribosomen^{60,153} auch „Proofreading“-Eigenschaften bzgl. des Einbaus von Aminosäuren in die wachsende Aminosäurekette auf, da falsch beladene tRNAs von EF-Tu prinzipiell nicht erkannt werden und folglich nicht in die Ribosomen eintreten können⁶⁰. Somit kann der Einbau von nicht-kanonischen Aminosäuren, welche bereits an eine tRNA gebunden wurden, durch EF-Tu, welcher die entsprechende Aminoacyl-tRNA unter Umständen nicht akzeptiert, verhindert werden^{60,153,186}.

2.6.1 Ortsspezifischer Einbau nicht-kanonischer Aminosäuren

Der ortsspezifische Einbau nicht-kanonischer Aminosäuren in Proteine erfordert ein orthogonales Paar aus AaRS und tRNA und die Auswahl einer mRNA-Sequenz, welche von der tRNA erkannt werden soll (Codon)^{22,57,247-249}.

Während die orthogonale tRNA so designt wird, dass diese über eine spezielle Anticodon-Sequenz verfügt, welche das ausgewählte zugehörige Codon erkennt, veranlasst die orthogonale AaRS die Bindung der einzubauenden nicht-kanonischen Aminosäure an diese tRNA²⁴⁷. Mögliche Codons für orthogonale tRNAs sind Stopcodons oder vier-Basen-Codons^{22,243,247,249}. Hierbei gilt anzumerken, dass die Auswahl eines der drei möglichen Stopcodons (UAA, UAG, UGA) die Beendigung der Translation an diesem Stopcodon supprimiert (Stopcodon-Suppression) und somit eines der beiden anderen Stopcodons zur Translations-Beendigung verwendet werden muss^{60,186,189}. Wird ein vier-Basen-Codon ausgewählt, so führt dies zu einem so genannten „Frameshift“ auf DNA-Ebene, weshalb in diesem Fall die DNA-Sequenz des zu exprimierenden Gens modifiziert werden muss^{106,189}.

Unabhängig von der Auswahl des Codons erlaubt der ortsspezifische Einbau die Einführung einer oder mehrerer (n) zusätzlicher nicht-kanonischer Aminosäuren zu dem kanonischen Set an Aminosäuren

^{109,250}. Somit werden bei der Proteinbiosynthese 20 kanonische und n nicht-kanonische Aminosäure in Proteine eingebaut. Abbildung 14 zeigt eine mögliche modifizierte Version des genetischen Codes, welche beim ortsspezifischen Einbau nicht-kanonischer Aminosäuren vorliegen kann. Dargestellt ist das Beispiel der Suppression des amber-Stopcodons (UAG).

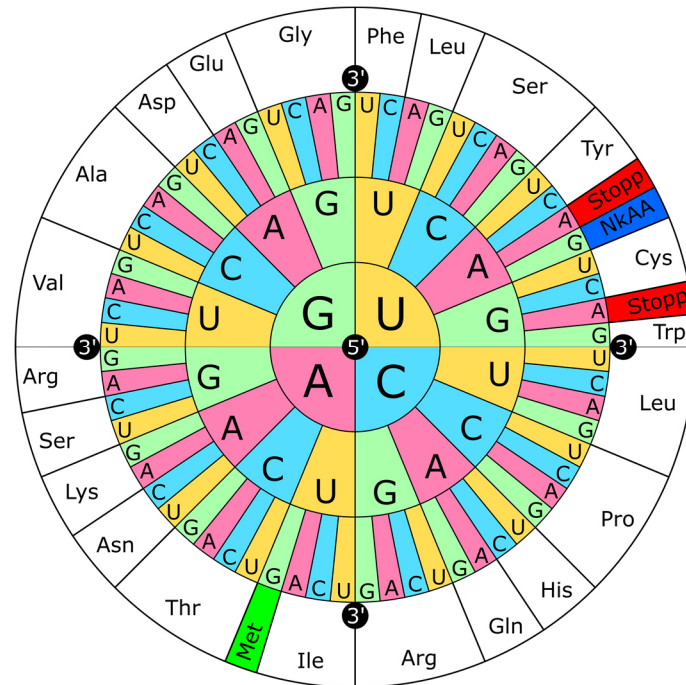


Abbildung 14 Modifizierte Version der in Abbildung 8 dargestellten Code-Sonne. Die Abbildung zeigt den neuen genetischen Code, der bei der Suppression des amber-Stopcodons (UAG) vorliegt. Hierbei wird dieses Codon in eine nicht-kanonische Aminosäure codiert (blau markiert).

Ein großer Nachteil des ortsspezifischen Einbaus nicht-kanonischer Aminosäuren in Proteine ist, dass sich das Design und die Produktion eines funktionellen orthogonalen Paares aus AaRS und tRNA für jede einzubauende nicht-kanonische Aminosäure als schwierig erweisen kann ^{57,106}. Des Weiteren ist es schwierig, mit dieser Methode Proteine mit mehr als einer eingebauten nicht-kanonischen Aminosäure zu erhalten ^{106,109}.

Dennoch konnten durch den ortsspezifischen Einbau bereits eine Vielzahl an verschiedenen nicht-kanonischen Aminosäuren, wie z. B. diverse Tyrosin- und Pyrrolysin-Analoga, in Proteine eingebaut werden ^{109,243,248}. Dies ermöglichte u. a. die Synthese von Proteinen mit hochreaktiven bzw. funktionellen Gruppen ^{243,248}.

2.6.2 Seitenkettenspezifischer Einbau nicht-kanonischer Aminosäuren

Im Gegensatz zum ortsspezifischen Einbau verlangt der seitenkettenspezifische Einbau nicht-kanonischer Aminosäuren in Proteine keine orthogonalen Komponenten¹⁰⁶. Beim seitenkettenspezifischen Einbau wird das Expressionssystem forciert, die einzubauende nicht-kanonische Aminosäure als Ersatz für eine kanonische Aminosäure zu akzeptieren^{57,245}. Zu diesem Zweck wird eine Genexpressionsreaktion in Abwesenheit der zu ersetzenden kanonischen Aminosäure, aber in Präsenz der nicht-kanonischen Aminosäure, welche als Ersatz fungiert, durchgeführt^{22,106,109}. Die Grundvoraussetzung für den seitenkettenspezifischen Einbau ist, dass die nicht-kanonische Aminosäure von der endogenen Translations-Maschinerie des Expressionssystems akzeptiert wird^{22,57,189}. Aus diesem Grund sind nur nicht-kanonische Aminosäuren für diese Methode geeignet, welche ein Analogon einer kanonischen Aminosäure mit nicht allzu großen Modifikationen sind, um eine Erkennung durch die zugehörige native AaRS zu gewährleisten^{245,246}. Abbildung 15 veranschaulicht dies anhand einer schematischen Darstellung.

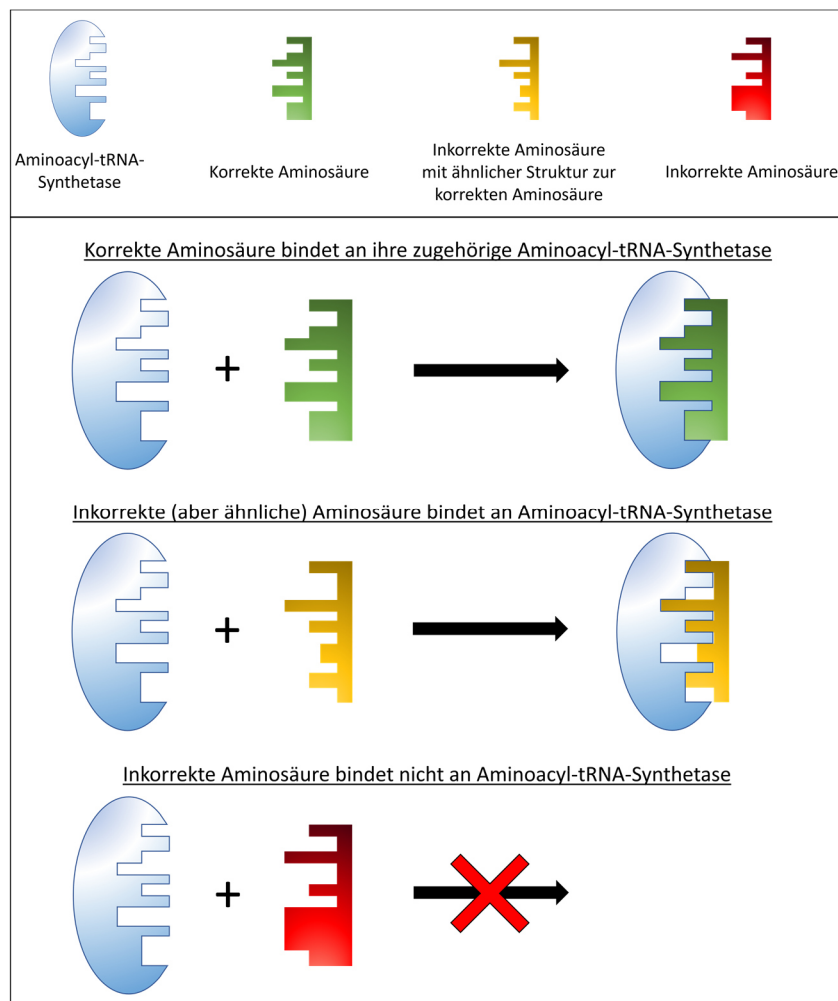


Abbildung 15 Erkennung verschiedener Aminosäuren durch eine AaRS. Eine AaRS kann spezifisch ihre zugehörige Aminosäure binden, während andere Aminosäuren prinzipiell nicht gebunden werden. Weist eine andere Aminosäure aber eine strukturelle Ähnlichkeit zur zugehörigen Aminosäure auf, so kann auch diese an die AaRS binden.

Somit ermöglicht der seitenkettenspezifische Einbau die Ersetzung einer oder mehrerer (n) spezifischer kanonischer Aminosäuren durch n nicht-kanonische Aminosäuren im kanonischen Set an Aminosäuren^{109,250}. Bei der Proteinbiosynthese werden 20-n kanonische und n nicht-kanonische Aminosäuren als Ersatz für die fehlenden n kanonischen Aminosäuren in Proteine eingebaut. Eine mögliche modifizierte Version des genetischen Codes, welche beim seitenkettenspezifischen Einbau nicht-kanonischer Aminosäuren vorliegen kann, wird in Abbildung 16 gezeigt. Dargestellt ist das Beispiel des Einbaus eines Arginin-Analogons als Arginin-Ersatz.

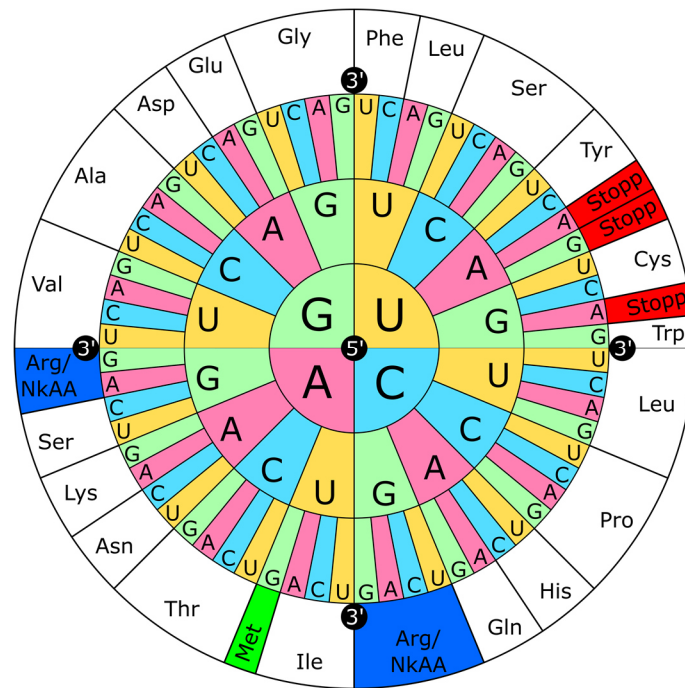


Abbildung 16 Modifizierte Version der in Abbildung 8 dargestellten Code-Sonne. Die Abbildung zeigt den neuen genetischen Code unter Berücksichtigung eines Arginin-Analogons, welches als Arginin-Ersatz fungieren kann. Hierbei fungieren die für Arginin codierenden Codons auch als Codons für das nicht-kanonische Arginin-Analogon (blau markiert).

Beim seitenkettenspezifischen Einbau nicht-kanonischer Aminosäuren in Proteine benötigt das Expressionssystem keine zusätzlichen Komponenten neben den endogenen¹⁰⁶. Dennoch wird eine effektive Erkennung der nicht-kanonischen Aminosäure durch eine native AaRS vorausgesetzt und es kann unter Umständen notwendig sein, dass eine komplette Abwesenheit der zu ersetzenden kanonischen Aminosäure gewährleistet wird^{22,57,106,189,245}. Dies soll im Folgenden näher erläutert werden.

In der Literatur wurde bereits gezeigt, dass K_M von Arginin und K_i von dessen nicht-kanonischen Analogon Canavanin bzgl. der Arginyl-tRNA-Synthetase in der gleichen Größenordnung liegen^{174,251}. Dies ermöglichte den seitenkettenspezifischen Einbau von Canavanin an allen Arginin-Positionen eines Proteins^{57,174}. Zu diesem Zweck wurden Canavanin und alle kanonischen Aminosäuren außer Arginin zu einem auf *E. coli* basierenden, zellfreien Expressionssystem hinzugegeben. In diesem konnten

2 Grundlagen

Proteine mit eingebautem Canavanin aber ohne eingebautes Arginin synthetisiert werden, wobei geringe Mengen an „restlichem“ Arginin, welches aus der Präparation des Expressionssystems stammte, den Canavanin-Einbau nicht nachteilig beeinflussten^{57,174}. Dennoch waren die Ausbeuten an Protein mit eingebautem Canavanin an allen Arginin-Positionen viel geringer als die maximal mögliche Proteinausbeute, wenn dem Expressionssystem Arginin anstatt Canavanin hinzugefügt wurde^{28,174}.

Im Gegensatz dazu erfordert der seitenkettenspezifische Einbau anderer nicht-kanonischer Aminosäuren unter Umständen vernachlässigbare Mengen der zugehörigen kanonischen Aminosäure^{22,57,106,189,245}. Selbst geringe Mengen der kanonischen Aminosäure, welche aus der Präparation des Expressionssystems stammen, können von der Translations-Maschinerie präferiert werden und wirken somit dem seitenkettenspezifischen Einbau zugehöriger nicht-kanonischer Analoga in Proteine entgegen^{22,57}. In der Literatur wurde dies bereits für den seitenkettenspezifischen Einbau des Lysin-Derivates Hydroxylysin (L-5-Hydroxylysin) gezeigt⁵⁷. Ein vollständiger seitenkettenspezifischer Einbau von Hydroxylysin in Proteine konnte ohne Weiteres nicht erreicht werden, da „restliches“ Lysin, welches aus der Präparation des Expressionssystems stammte, für den Einbau stark präferiert wurde.

Als wahrscheinlichster Grund für diese Diskriminierung der „falschen“ Aminosäure kann eine hohe Spezifität der Translations-Maschinerie gesehen werden^{2,22,54,56–58,61,62}. Diese wird u. a. durch den „Kinetic Proofreading“-Prozess, welcher im anschließenden Abschnitt näher beschrieben wird, gewährleistet^{54,58,61,62}.

2.7 Fehlerreduktion auf Translationsebene durch „Kinetic Proofreading“

Generell weisen viele Enzyme eine hohe Substratspezifität auf ^{61,72,252–254}. Dennoch existieren Situationen, in denen solche Enzyme auch Substanzen, welche eine gewisse Ähnlichkeit zum eigentlichen Substrat aufweisen, akzeptieren ^{250,255}. Beispiele sind nicht-kanonische Aminosäuren, welche im Rahmen des zuvor erwähnten seitenkettenspezifischen Einbaus nicht-kanonischer Aminosäuren in Proteine von den nativen AaRSs erkannt werden ^{22,57,189,245,246}. Dennoch wird die zugehörige kanonische Aminosäure generell stark präferiert ^{22,57}. Somit stellt sich die Frage, welcher Mechanismus für diese hohe Spezifität verantwortlich ist.

Ein in der Literatur ausführlich beschriebener Mechanismus bzgl. der AaRS vermittelten Beladung der tRNA mit Aminosäuren ist das „Kinetic Proofreading“ ^{54,58,61,62}. Dieser Prozess wurde zunächst in zwei unabhängigen Studien, einerseits von Hopfield und andererseits von Ninio, beschrieben ^{54,55,61}, wobei in der vorliegenden Dissertation lediglich die Studie von Hopfield betrachtet wird. Dadurch, dass die geringen Fehlerraten bei der Beladung der tRNA mit deren zugehörigen Aminosäuren von 10^{-4} nicht durch die Differenz zwischen der freien Energie der korrekten und einer inkorrekten (aber ähnlichen) Aminosäure erklärt werden konnten, schlug Hopfield zunächst ein einfaches zweistufiges kinetisches Modell vor (siehe Abbildung 17 A und B, Schritte 1-4) ^{54,58,256}. Hierbei wurde die Fehlerreduktion auf die Schritte bis zur tRNA-Bindung an die mit der Aminosäure beladene AaRS reduziert ⁵⁴. Später wurden diesem Modell weitere Details hinzugefügt, die den gesamten Prozess bis zum Erhalt der mit einer Aminosäure beladenen tRNA in die Fehlerreduktion involvieren ⁵⁸. Insgesamt sind in allen Modellen lediglich 3 Schritte für die Fehlerreduktion durch „Kinetic Proofreading“ essenziell: Die Ausbildung eines initialen Michaelis-Menten-Komplexes, ein durch Phosphat-Hydrolyse verursachter irreversibler Schritt und die Ausbildung eines hochenergetischen Intermediates ^{54,58,62}. Während vergleichbare „On“-Raten für die Beladung der tRNA mit der korrekten oder einer inkorrekten Aminosäure angenommen werden, wird die Spezifität durch unterschiedliche „Off“-Raten gewährleistet ^{54,55,62,252}. Mehrere Reaktionsschritte implizieren multiplikative „Off“-Raten ^{257–259}, was Fehlerraten bis zu 10^{-4} ermöglicht, selbst wenn die Diskriminierung der AaRS gegen die inkorrekte Aminosäure in den einzelnen Schritten nur bei 1:10 bis 1:100 liegt ^{2,54,55,58,256}.

Abbildung 17 zeigt ein Schema sowie ein Energiediagramm, anhand derer der „Kinetic Proofreading“-Prozess veranschaulicht wird.

2 Grundlagen

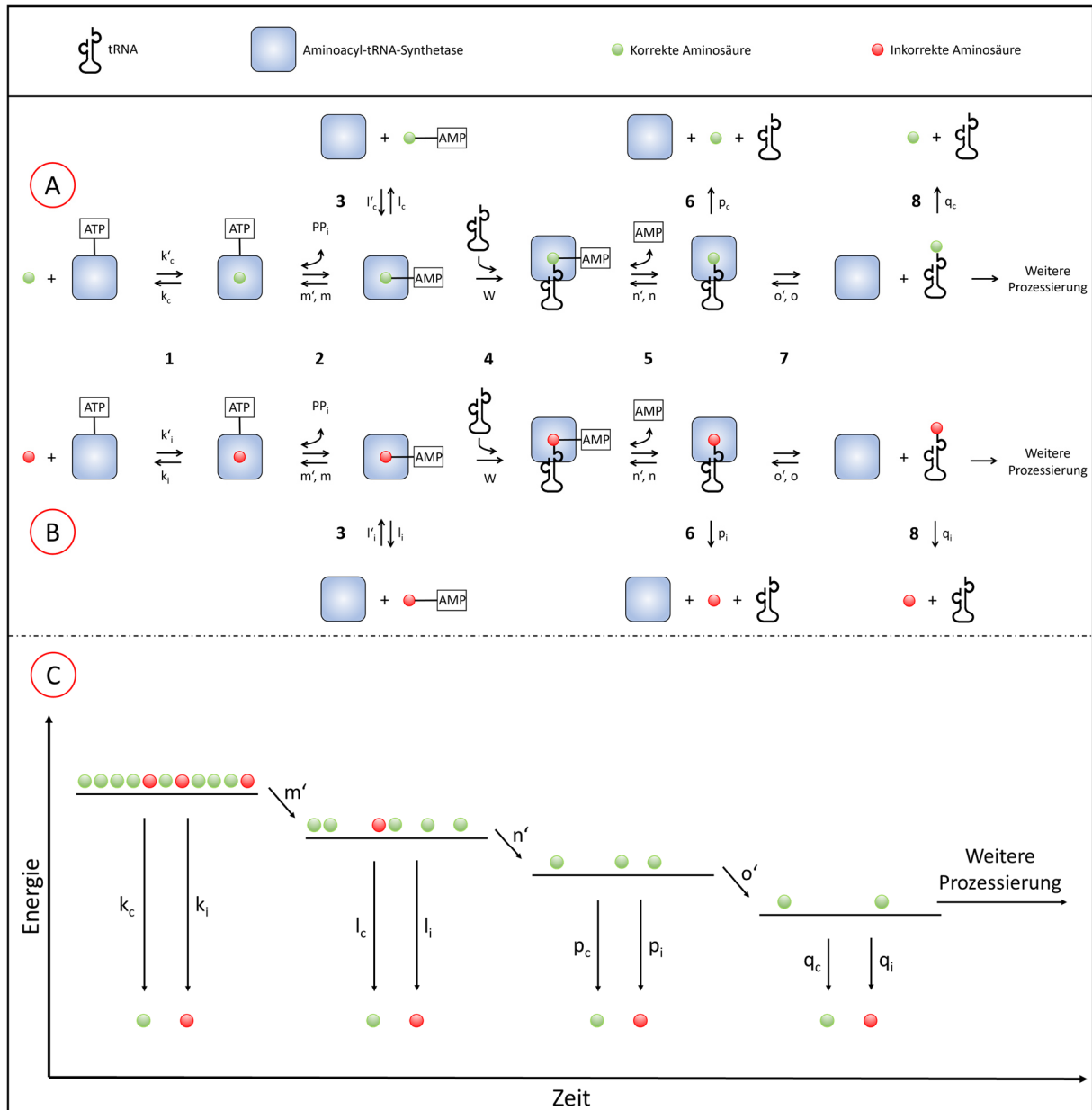


Abbildung 17 Beladung der tRNA mit der korrekten (A) und einer inkorrekten Aminosäure (B). Reaktion 1 zeigt die Bindung einer Aminosäure an einen AaRS-ATP-Komplex, welcher durch die Dissoziation von Pyrophosphat aktiviert wird (2). Während dieser aktivierte Komplex in Aminoacyl-AMP und AaRS dissoziieren kann (3), ermöglicht dessen Interaktion mit tRNA eine Fortsetzung des Gesamtprozesses (4)⁵⁴. Nach der Dissoziation von AMP wird ein Komplex erhalten, in dem die mit der Aminosäure beladene tRNA an die AaRS gebunden ist (5). Dieser Komplex kann in AaRS, Aminosäure und tRNA (6) oder in AaRS und Aminoacyl-tRNA dissoziieren (7)⁵⁸. Letztere werden im Rahmen der Proteinbiosynthese weiter prozessiert, wobei auch eine Deacetylierung der Aminoacyl-tRNA erfolgen kann (8)⁵⁸. Zusätzlich zeigt die Abbildung ein Energiediagramm der AaRS vermittelten Beladung der tRNA mit Aminosäuren (C).

Im dargestellten Schema ist der erste Schritt die Bindung einer Aminosäure an einen AaRS-ATP-Komplex, welche eine reversible und spezifische Reaktion ist. Hopfield schlug vor, dass die Spezifität durch die „Off“-Raten verursacht wird ($k'_c = k'_i$; $k_c < k_i$)⁵⁴. Im Gegensatz dazu ist der zweite Schritt, die Aktivierung des Aminosäure-AaRS-ATP-Komplexes, für geringe Pyrophosphat-Konzentrationen, welche generell vorliegen, irreversibel ($m' \gg m$) und unspezifisch ($m'_c = m'_i = m'$)^{54,58}. Dieser

aktivierter Komplex kann entweder irreversibel und unspezifisch ($W_c = W_i = W$) die zugehörige tRNA binden oder reversibel und spezifisch in AaRS und Aminoacyl-AMP degradiert werden⁵⁴. Die Spezifität in letzterem Fall wird in Analogie zum ersten Schritt durch die „Off“-Raten verursacht ($l'_c = l'_i$; $l_c < l_i$)⁵⁴. Der erhaltene tRNA-Komplex kann durch Dissoziation von AMP erneut aktiviert werden⁵⁸. Dieser Schritt ist bei geringen AMP-Konzentrationen, welche prinzipiell vorliegen, irreversibel ($n' \gg n$) und unspezifisch ($n'_c = n'_i = n'$)^{58,260}. Der aktivierte Komplex kann entweder irreversibel und spezifisch in AaRS, Aminosäure und tRNA dissoziieren ($p_c < p_i$) oder irreversibel ($o' \gg o$) und unspezifisch ($o'_c = o'_i = o'$) in AaRS und Aminoacyl-tRNA⁵⁸. Die Irreversibilität in letzterem Fall resultiert daraus, dass die Aminoacyl-tRNA das Produkt der AaRS vermittelten Reaktion darstellt und somit deren erneute Bindung an die AaRS unwahrscheinlich ist. Neben der weiteren Prozessierung der AaRS sowie der Aminoacyl-tRNA im Rahmen der Proteinbiosynthese ist auch eine irreversible und spezifische Deacetylierung der Aminoacyl-tRNA möglich ($q_c < q_i$)⁵⁸.

Die gesamte Spezifität des „Kinetic Proofreading“-Prozesses wird durch die Diskriminierungen $k_c < k_i$, $l_c < l_i$, $p_c < p_i$ und $q_c < q_i$ verursacht^{54,58}. Selbst wenn diese Diskriminierungen lediglich im Bereich 1:10 bis 1:100 liegen, so können durch die Multiplikation der einzelnen Raten die zuvor erwähnten geringen Fehlerraten von 10^{-4} erhalten werden^{2,54,55,58,256}.

Aufgrund des vorgeschlagenen „Kinetic Proofreading“-Prozesses nahm Hopfield an, dass der Prozess der Fehlerreduktion bei der Beladung der tRNA mit Aminosäuren einen linearen Zusammenhang mit dem Verhältnis der Konzentrationen der inkorrekten und der korrekten Aminosäure aufweist^{58,62}.

2.8 Fluoreszenz

Da der Prozess der Fluoreszenz eine wichtige Rolle im Rahmen der durchgeführten Messungen in der vorliegenden Dissertation spielt, sollen zunächst die Grundlagen dieses Phänomens erklärt werden. Des Weiteren wird das grün fluoreszierende Protein (GFP) genauer betrachtet, da dieses als Modellprotein verwendet wurde.

2.8.1 Grundlagen

Atome oder Moleküle können durch Energieaufnahme vom Grundzustand in einen angeregten Zustand übergehen^{261–265}. Stammt die aufgenommene Energie aus einem Photon, wird der Anregungsprozess als Lichtabsorption bezeichnet²⁶¹. Hierbei entspricht der maximal aufgenommene Energiebetrag der Energie eines Photons E_{ph} , welche durch Gleichung (1) gegeben ist^{266–268}.

$$E_{ph} = h \cdot \nu = h \cdot \frac{c}{\lambda} \quad (1)$$

Hierbei ist h das Planck'sche Wirkungsquantum, ν die Frequenz des Photons, λ dessen Wellenlänge und c die Lichtgeschwindigkeit im Vakuum^{267–269}.

Der Anteil des eintreffenden Lichts, der absorbiert wird, hängt von der Wellenlänge sowie von dem Atom oder Molekül selbst ab^{268,270,271}. Somit ergeben sich unterschiedliche Absorptionsspektren für unterschiedliche Substanzen^{268,270,271}.

Ausgehend vom angeregten Zustand können diverse Prozesse stattfinden, damit das Atom oder Molekül zurück in den Grundzustand gelangt. Abbildung 18 zeigt ein so genanntes Jabłoński-Diagramm, welches die möglichen Prozesse veranschaulicht^{261,266,272–277}.

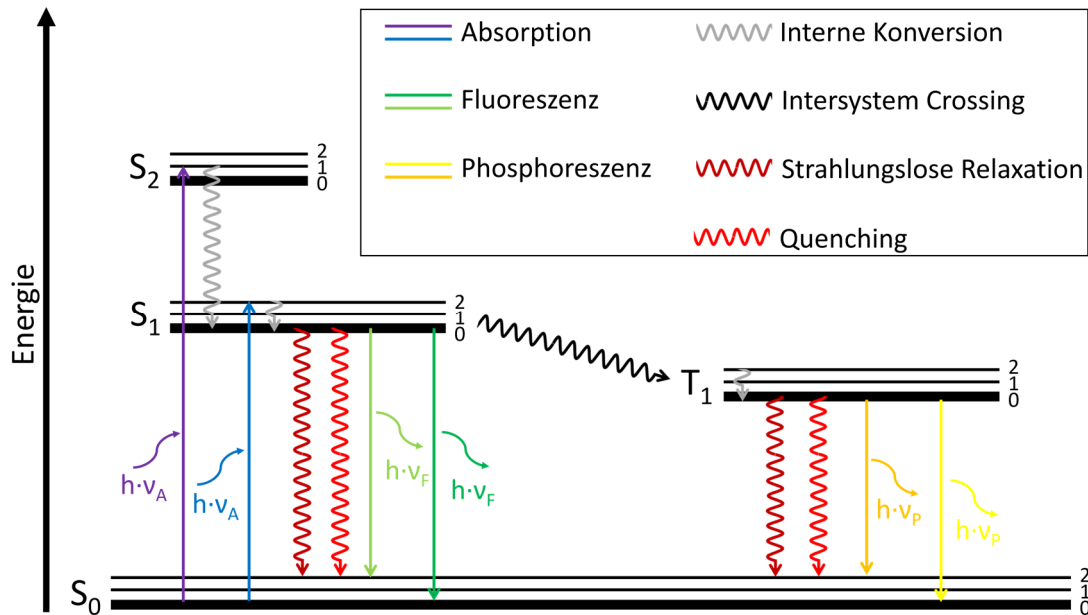


Abbildung 18 Mögliche Darstellungsweise eines Jabłoński-Diagramms. In diesem sind strahlende Übergänge durch gerade Linien und nichtstrahlende Übergänge durch wellenförmige Linien gekennzeichnet.

Wird ein Atom oder Molekül angeregt, so gelangt es vom Grundzustand S_0 in einen angeregten Singulett-Zustand S_n ($n = 1, 2, \dots$), wobei bei diesem verschiedene Schwingungszustände ($0, 1, 2, \dots$) angenommen werden können^{276,278–281}. Innerhalb von ca. 10^{-12} s findet eine Energiedissipation statt, bei der der niedrigste Schwingungszustand von S_1 erreicht wird^{269,276,278}. Dieser Prozess wird „Interne Konversion“ genannt^{269,278}. Ausgehend von diesem Zustand kann der Grundzustand durch Emission eines Photons (Fluoreszenz) oder durch strahlungslose Energiedissipation (Strahlungslose Relaxation oder Quenching) erreicht werden^{261,269,272,277,278,282}. Des Weiteren kann „Intersystem Crossing“ stattfinden, was einen Übergang vom Singulett-Zustand S_1 in den Triplett-Zustand T_1 bedeutet^{269,272,278}. Der anschließende Übergang von T_1 in den Grundzustand kann durch Photonenemission (Phosphoreszenz) oder strahlungslos erfolgen^{276,277}.

Substanzen, welche zur Fluoreszenz befähigt sind, werden als Fluorophore bezeichnet^{103,276,278}. Für Fluorophore lässt sich feststellen, dass deren Absorptions- und Emissionsspektrum meist spiegelverkehrt zueinander sind (siehe Abbildung 19 A am Beispiel von 6-Fam)^{278,283–285}. Wird ein Fluorophor angeregt, so findet eine Energiedissipation durch Interaktionen mit dem Lösungsmittel statt, wodurch ein „relaxierter“ angeregter Zustand S'_1 mit geringerer Energie als S_1 vor der Photonenemission erreicht wird (siehe Abbildung 19 B)^{266,278,286}. Durch diese Energiedissipation findet eine Rotverschiebung des Emissionsspektrums statt, welche als Stokes-Verschiebung bezeichnet wird^{285,287,288}. Die Stokes-Verschiebung ist der Abstand zwischen den 0-0-Übergängen der Absorption und der Emission^{284,285}. Hierbei repräsentieren die beiden Nullen den Zustand S_0 im Schwingungszustand 0 sowie S_1 bzw. S'_1 im Schwingungszustand 0. Abbildung 19 veranschaulicht den Begriff der Stokes-Verschiebung^{283–285}.

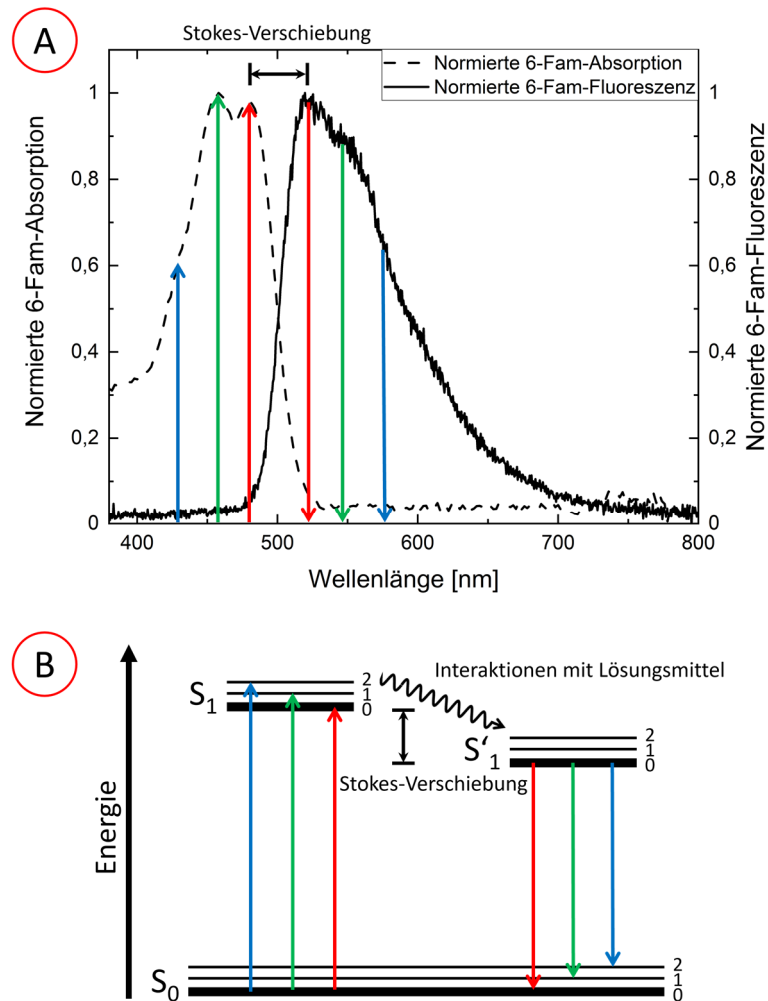


Abbildung 19 Schema zur Stokes-Verschiebung. Dargestellt ist das Absorptionsspektrum sowie das Emissionsspektrum (bei einer Anregungswellenlänge von $\lambda = 480$ nm) von 6-Fam. Mit diesen Spektren wird die Stokes-Verschiebung auf spektraler Ebene veranschaulicht (A). Des Weiteren enthält die Abbildung ein Energiediagramm, welches die Stokes-Verschiebung auf energetischer Ebene veranschaulicht (B). Pfeile gleicher Farbe gehören zu einem Anregungs-Emissions-Vorgang.

Wie bereits zuvor erwähnt, kann ein Übergang vom angeregten Zustand in den Grundzustand auch strahlungslos durch Quenching erfolgen²⁸². Hierbei existieren verschiedene Formen. So kann zwischen dem angeregten Fluorophor und einem Quencher durch Stoßvorgänge eine Energieübertragung stattfinden (dynamisches Quenching) oder der Fluorophor mit einem Quencher einen nicht-fluoreszierenden Komplex bilden (statisches Quenching)^{278,289–291}.

Eine weitere Möglichkeit für einen Quenchingprozess stellt der Förster-Resonanzenergietransfer (FRET) dar^{266,282,292}. Die Grundvoraussetzung für diesen Prozess ist die Überlappung des Emissionsspektrums eines Moleküls (Donor) mit dem Absorptionsspektrum eines anderen Moleküls (Akzeptor) (siehe Abbildung 20 A am Beispiel von 6-Fam und BHQ1²⁹³)^{292,294,295}. Diese Überlappung ermöglicht einen strahlungslosen Energietransfer vom angeregten Donor auf einen Akzeptor im Grundzustand^{278,282,289}. Hierdurch erreicht der Donor den Grundzustand und der Akzeptor den

angeregten Zustand. Der angeregte Akzeptor kann anschließend durch Emission eines Photons oder strahlungslos in den Grundzustand übergehen (siehe Abbildung 20 B) ^{266,282,296,297}.

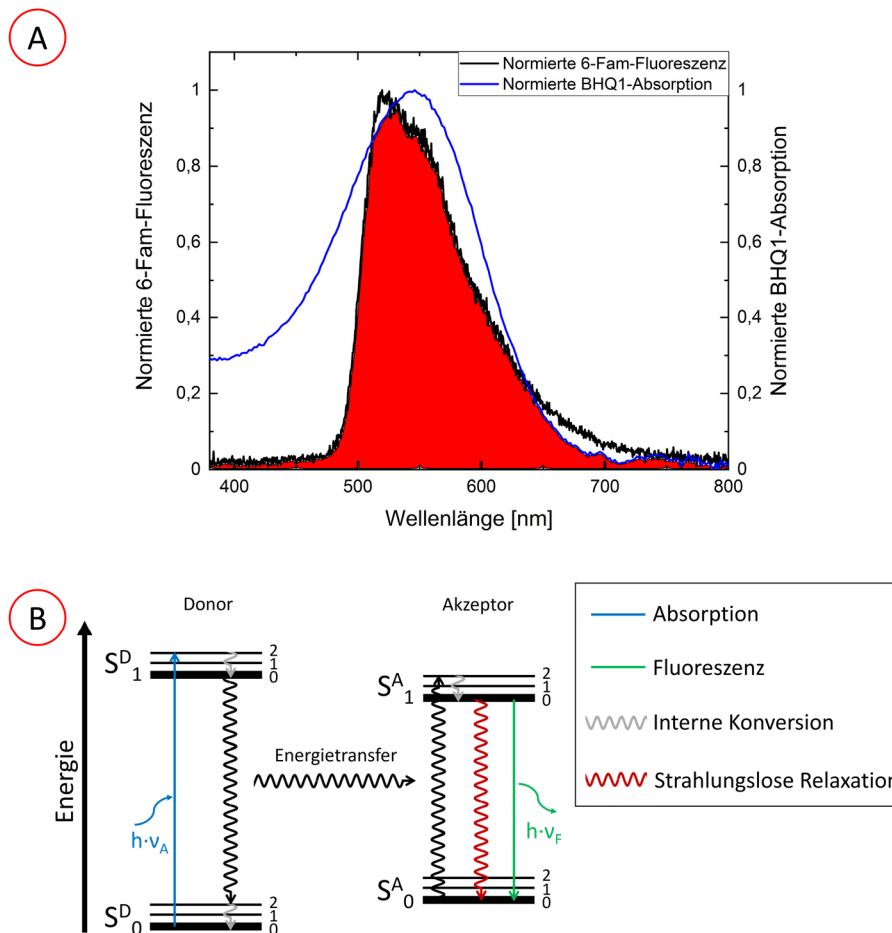


Abbildung 20 Schema zum Förster-Resonanzenergietransfer. Dargestellt ist die Überlappung (rote Fläche) des Emissionsspektrums (bei einer Anregungswellenlänge von $\lambda = 480$ nm) von 6-Fam (Donor; schwarze Linie) mit dem Absorptionsspektrum von BHQ1 (Akzeptor; blaue Linie). Diese ist essenziell für das Auftreten von FRET (A). Zusätzlich zeigt die Abbildung ein Energiediagramm, welches FRET auf energetischer Ebene veranschaulicht (B).

Die Effizienz des Energietransfers E_{FRET} ist durch Gleichung (2) gegeben ^{289,298}.

$$E_{FRET} = \frac{R_0^6}{R_0^6 + r^6} \quad (2)$$

Aus Gleichung (2) folgt, dass der Abstand zwischen Donor und Akzeptor r für die Effizienz von FRET entscheidend ist. Als charakteristische Größe fungiert der Förster-Radius R_0 , bei dem die FRET-Effizienz 50 % beträgt und welcher durch Gleichung (3) gegeben ist ^{278,289,296,299}.

$$R_0 = \left(\frac{9000 \ln 10 \kappa^2 Q_D}{128 \pi^5 N_A n_0^4} J(\lambda) \right)^{\frac{1}{6}} \quad (3)$$

Hierbei ist κ der Orientierungsfaktor der Dipolmomente von Donor und Akzeptor, Q_D die Quantenausbeute des Donors, N_A die Avogadro-Konstante, n_0 der Brechungsindex des Lösungsmittels und $J(\lambda)$ das Überlappungsintegral des Emissionsspektrums des Donors mit dem Absorptionsspektrum des Akzeptors^{278,289,296,299}.

2.8.2 Grün fluoreszierendes Protein

An dieser Stelle wird näher auf das grün fluoreszierende Protein (GFP) eingegangen, da eine mutierte Version dieses Proteins in allen im Rahmen dieser Dissertation durchgeführten Experimenten als Modellprotein verwendet wurde. Durch Messung der Fluoreszenz war es somit möglich, die Genexpression zu visualisieren.

Die verwendete GFP-Variante ist eGFP-Del6-229 (deGFP), welches eine verkürzte Version der GFP-Mutante „Enhanced GFP“ (EGFP) darstellt^{28,238}. Im Gegensatz zum Wildtyp-GFP, bei dem der Fluorophor aus den Aminosäuren Ser-Tyr-Gly gebildet wird, bildet sich der Fluorophor bei EGFP und somit auch bei deGFP aus den Aminosäuren Thr-Tyr-Gly (Strukturformeln der Fluorophore, siehe Abbildung 21)^{294,295,300,301}.

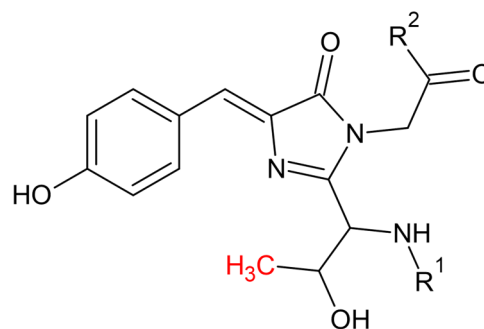


Abbildung 21 Strukturformel des Fluorophors von Wildtyp-GFP (ohne rot markierte Methylgruppe) bzw. deGFP (mit rot markierter Methylgruppe).

Die Bildung des Fluorophors geschieht autokatalytisch durch eine Zyklisierung mit anschließender Wasserabspaltung und eine darauffolgende Oxidation^{57,294,295,300}. Das im Fluorophor vorhandene Tyr kann deprotoniert werden, wodurch der Fluorophor in einer neutralen oder in einer anionischen Form vorliegen kann^{294,302}. Wird der Fluorophor in der neutralen Form angeregt, so geschieht in den meisten Fällen ein „Excited state proton transfer“ (ESPT), durch den der Fluorophor in die anionische Form übergeht³⁰². Somit wird in der Regel nur ein Emissionsmaximum beobachtet, während die

Absorptionsmaxima der neutralen und anionischen Form beide im Absorptionsspektrum sichtbar sind ^{295,302}.

Im Falle von Wildtyp-GFP liegt das Verhältnis zwischen neutraler und anionischer Form im Grundzustand bei 6:1 ^{294,295}. Durch die Substitution von Serin durch Threonin (GFPS65T) erhöht sich die Säurestärke des Fluorophors ^{294,303}. Dadurch liegt der Fluorophor überwiegend in der anionischen Form vor, wodurch sich dessen Absorptionsspektrum ändert ³⁰³. Dies hat u. a. einen ca. 5,5-fach erhöhten Extinktionskoeffizienten (GFPS65T: 39200 M⁻¹cm⁻¹ bei 490nm; Wildtyp-GFP: 7150 M⁻¹cm⁻¹ bei 475 nm) zur Folge ^{294,304}. Somit kann der Fluorophor mit den erwähnten Wellenlängen effizienter angeregt werden, woraus eine erhöhte Fluoreszenz-Intensität resultiert ^{294,304,305}.

Durch andere oder zusätzliche Modifikationen in der Aminosäuresequenz können sich die Fluoreszenzeigenschaften ändern (siehe Tabelle 1) ^{294,301,306}.

Tabelle 1 Absorptions- und Emissionsmaxima verschiedener GFP-Varianten.

GFP-Variante	Mutation(en)	Absorptionsmaximum	Emissionsmaximum	Referenz
		[nm]	[nm]	
Wildtyp-GFP (neutral)	/	395	510	295
Wildtyp-GFP (anionisch)	/	475	510	295
GFPS65T	S65T	489	511	294
EGFP	S65T, F64L	488	507	294,307

Dennoch wurde durch die N- und C-terminale Verkürzung von EGFP zur Generierung von deGFP lediglich dessen Produktionsrate in einem zellfreien Expressionssystem erhöht ²⁸. Die Fluoreszenzeigenschaften blieben gleich ²⁸. Abbildung 22 zeigt das Absorptions- und Emissionsspektrum von deGFP.

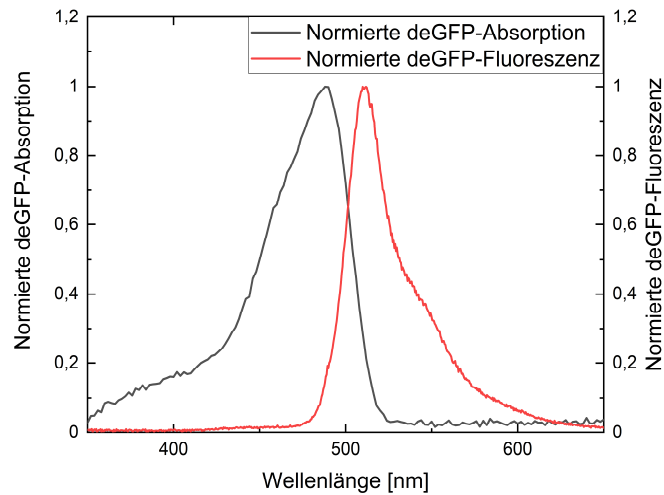


Abbildung 22 Absorptionsspektrum und Emissionsspektrum (bei einer Anregungswellenlänge von $\lambda = 488 \text{ nm}$) von deGFP. Die Maxima der Absorption sowie der Emission entsprechen denen von EGFP.

Alle GFP-Varianten weisen eine analoge Struktur auf. 11 β -Faltblattstrukturen bilden ein zylinderförmiges Konstrukt, in dessen Mitte sich der Fluorophor befindet ^{295,308}. Dieser ist über eine zentrale α -Helix mit dem Zylinder verbunden (siehe Abbildung 23, Bändermodell von EGFP ^{303,309,310}).

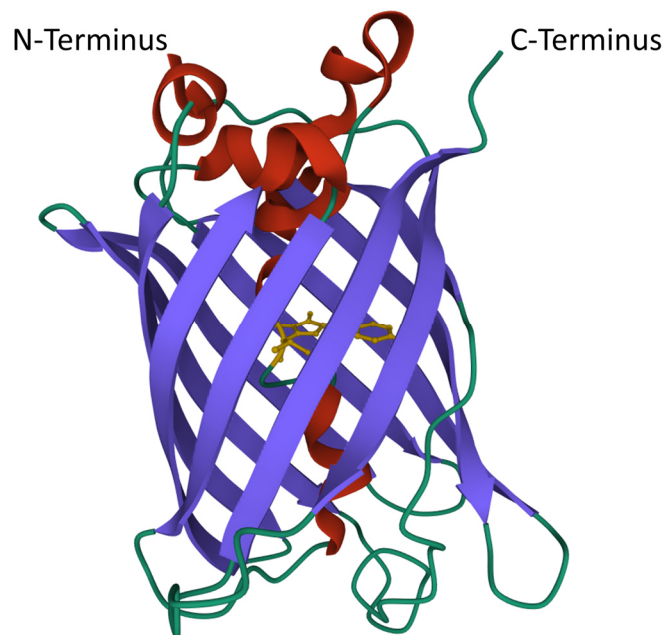


Abbildung 23 Bändermodell von EGFP (PDB ID 2Y0G ^{303,309}, bereitgestellt von RCSB PDB. Für die Generierung der Abbildung wurde der Mol* Viewer ³¹⁰ verwendet). Im Bändermodell sind die β -Faltblätter blau, die α -Helix rot, der Fluorophor gelb und die restliche Polypeptidkette grün gefärbt.

2.9 Massenspektrometrie

Neben der Messung der Fluoreszenz-Intensität von deGFP sowie der Verwendung der Methode Natriumlaurylsulfat-Polyacrylamid-Gelelektrophorese (SDS-PAGE) wurde die Massenspektrometrie als weitere analytische Methode zum Nachweis von deGFP verwendet. Letztere soll im Folgenden näher beschrieben werden.

Mit Hilfe der Massenspektrometrie können diverse Stoffe qualitativ und mit geeigneten Methoden auch quantitativ nachgewiesen werden^{311–313}. Die Grundvoraussetzung ist, dass diese Stoffe ionisiert werden können^{314–316}. Nach der Ionisation können verschiedene Stoffe durch deren unterschiedliches Verhältnis zwischen Masse und Ladung (m/z) separiert werden^{314–316}.

Komplexe Stoff-Gemische erfordern unter Umständen zunächst eine Auftrennung der einzelnen Stoffe vor der Durchführung einer massenspektrometrischen Untersuchung^{317–319}. Dies kann durch geeignete Purifikationsmethoden oder Chromatographieverfahren erfolgen, z. B. durch einen dem Massenspektrometer vorgeschalteten Chromatograph^{314,317,320–326}. Ein solcher verfügt über eine stationäre Phase, welche die Auftrennung verschiedener Stoffe nach unterschiedlichen Parametern, wie z. B. Größe, Ladung, Polarität etc., veranlasst^{317,323,324,327,328}. Dies führt zu unterschiedlichen Retentionszeiten für verschiedene Stoffe^{322–324,329}. Je nachdem, welche mobile Phase für die Chromatographie mit anschließender Massenspektrometrie verwendet wird, nennt sich dies LC-MS (Flüssigkeiten) oder GC-MS (flüchtige Flüssigkeiten oder Gase)^{323,324,327,330}.

Zur Durchführung einer massenspektrometrischen Untersuchung muss der zu untersuchende Stoff, der so genannte Analyt^{313,317,319}, zunächst in der Ionenquelle ionisiert werden^{313,314,316}. Zu diesem Zweck können verschiedene Ionisierungsmethoden verwendet werden, welche grob in harte und weiche Ionisierungsmethoden untergliedert werden können^{331,332}. Während bei den weichen Ionisierungsmethoden (z. B. Elektronenspray-Ionisation (ESI), Matrix-unterstützte Laser-Desorption/Ionisation (MALDI)) Moleküle so gut wie keine Fragmentierung erfahren, werden bei harten Ionisierungsmethoden (z. B. Elektronenstoßionisation (EI)) die Moleküle fragmentiert^{314,316,333–335}. Durch eine Fragmentierung werden für jede Substanz spezifische Fragment-Ionen erhalten, die als einzigartiger „Fingerabdruck“ fungieren, der mit geeigneten Datenbanken (z. B. NIST) abgeglichen werden kann^{314,325,336}. Da große Moleküle, wie z. B. Proteine, mit harten Ionisierungsmethoden nur schwer zu ionisieren sind, ist bei solchen Molekülen eine weiche Ionisierungsmethode von Vorteil^{319,333–335,337,338}. Bei weichen Methoden tritt fast keine Fragmentierung des Analyten (M) auf und es werden fast ausschließlich so genannte Quasi-Molekülionen (Analyt als positiv geladenes Addukt $[M+nX]^{n+}$ ($X = H, NH_4, Na, \dots$) oder deprotonierter Analyt $[M-nH]^{n-}$, wobei $n = 1, 2, \dots$) erhalten^{316,333,337,339}. Somit entstehen lediglich ein- oder mehrfach geladene Moleküle^{335,338}. Anhand der zugehörigen Mess-Signale kann der Analyt direkt oder nach einer so genannten Dekonvolution durch einen Massenabgleich mit der Literatur nachgewiesen werden^{333–335,340}. Bei der Dekonvolution wird durch einen Algorithmus

2 Grundlagen

bestimmt, welche gemessenen m/z -Signale zu dem Analyt gehören und wie viele Ladungen vorhanden sind (n)^{333,335,340}. Aus den Signalen des Analyten mit unterschiedlich vielen Ladungen $(m/z)_n$ kann die Masse des Analyten M_M bestimmt werden (siehe Gleichung (4))³³⁸.

$$\left(\frac{m}{z}\right)_n = \frac{M_M \pm n \cdot M_X}{n} \quad (4)$$

Hierbei ist M_X die Molekülmasse der angehängten Kationen bzw. Protonen (+) oder der abstrahierten Protonen (-).

Abbildung 24 zeigt, wie ein Massenspektrum eines Proteins vor und nach der Dekonvolution aussieht. Des Weiteren liefert die Abbildung eine Tabelle, welche veranschaulicht, wie aus den einzelnen m/z -Signalen die Molekülmasse des Analyten bestimmt werden kann.

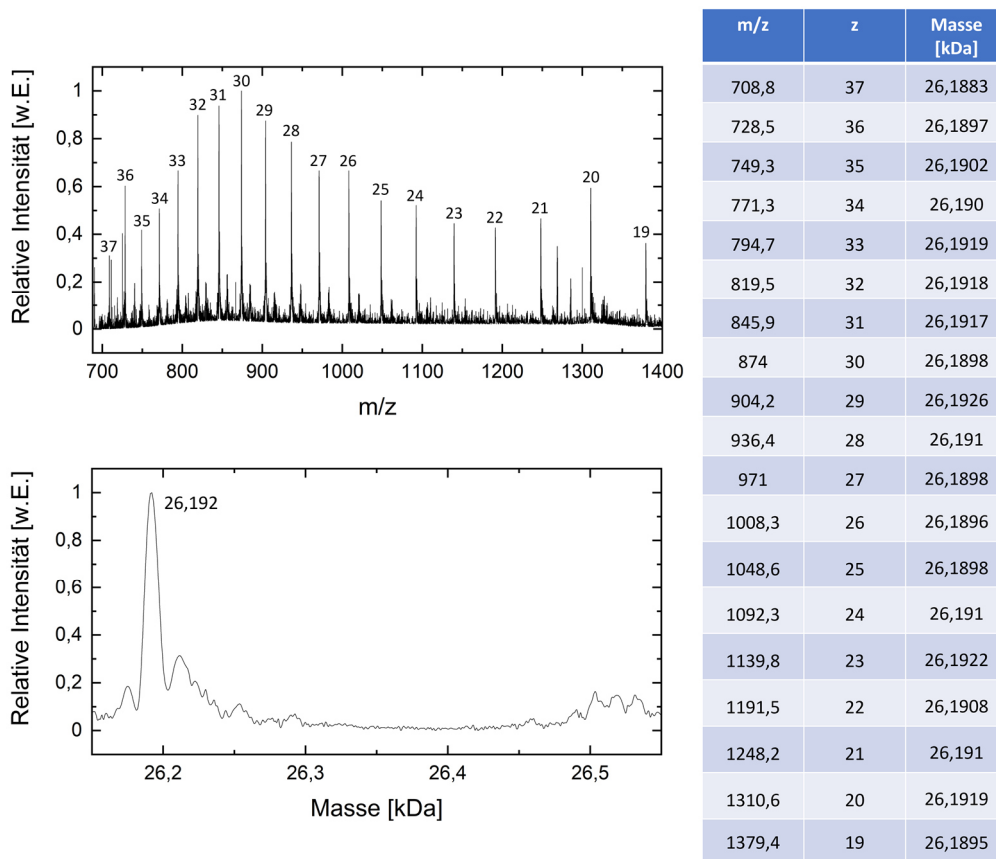


Abbildung 24 Massenspektrum von deGFP. Abgebildet ist das Spektrum, welches zunächst vom Massenspektrometer generiert wird. Dieses zeigt Peaks von verschiedenen Ladungszuständen von deGFP, welche durch das Anhängen von Protonen verursacht werden ($z = 19-37$) (links oben). Mit diesen kann die Molekülmasse von deGFP unter Verwendung von Gleichung (4) errechnet werden. Die errechneten Werte können der Tabelle entnommen werden (rechts). Ein Massenspektrum kann auch weitere Peaks, wie z. B. Isotopenpeaks, enthalten, die eine solche manuelle Berechnung erschweren³⁴⁰. Im Rahmen einer Dekonvolution wird durch einen Algorithmus bestimmt, welche Peaks dem Analyten angehören und welcher Ladungszustand vorliegt^{333,335,340}. Anschließend wird ein dekonvolviertes Spektrum generiert, in dem die Molekülmasse abgelesen werden kann (links unten).

Im Rahmen einer massenspektrometrischen Untersuchung werden die in der Ionenquelle erzeugten Ionen mit Hilfe eines Analysators nach deren Verhältnis aus Masse und Ladung (m/z) separiert^{313,314,316}. Hierfür existieren verschiedene Methoden, die mit unterschiedlichen Analysatoren realisiert werden können^{313,314,316}. Beispiele sind Quadrupol-, Time-of-Flight-, magnetische bzw. elektrostatische Sektorfeld- oder Ionenfallen-Analysatoren^{313,314,316,319,341,342}. Diese können einzeln oder in Kombination (Tandem-Massenspektrometer) verwendet werden^{313,314,316,341}.

Nach deren Erzeugung erhalten die Ionen durch eine Beschleunigungsspannung U zunächst eine von m/z abhängige Geschwindigkeit v (m/z), welche durch Gleichung (5) gegeben ist^{314,316,317,343}.

$$v\left(\frac{m}{z}\right) = \sqrt{\frac{2zeU}{m}} = \sqrt{(2eU) \cdot \left(\frac{m}{z}\right)^{-1}} \quad (5)$$

Hierbei ist e die Elementarladung³⁴³.

Anschließend durchlaufen die Ionen mit der Geschwindigkeit v den Analysator und wechselwirken abhängig von Geschwindigkeit und folglich von m/z mit einem magnetischen oder elektrischen Feld³¹³. Bei den wechselwirkenden Feldern kann es sich um einem elektrischen Quadrupol (Quadrupol-Analysator) oder um ein statisches magnetisches bzw. elektrisches Feld (magnetischer bzw. elektrostatischer Sektorfeld-Analysator) handeln^{313,319,344}. Durch die m/z -abhängige Wechselwirkung gelangen Ionen mit unterschiedlichen m/z auf verschiedene Trajektorien, was eine Separation ermöglicht^{313,315,330}. Eine weitere Möglichkeit zur Separation von Ionen mit unterschiedlichen Geschwindigkeiten v ist die Bestimmung der Zeit, welche die Ionen benötigen, um eine definierte Strecke zurückzulegen (Time-of-Flight-Analysatoren)^{314,345-348}.

Bei Ionenfallen-Analysatoren, wie z. B. einem Fourier-Transformations-Ionenzyklotronresonanz (FT-ICR)-Analysator, werden Ionen zunächst durch Wechselwirkung mit einem elektrischen Quadrupol oder mit einem statischen magnetischen bzw. elektrischen Feld gespeichert^{317,342,349,350}. Die gespeicherten Ionen können durch Änderung der wechselwirkenden Felder oder durch Zuschaltung weiterer Felder aus der Falle gelangen³⁴⁹⁻³⁵³. Je nach Änderung gelangen lediglich Ionen mit einem spezifischen m/z aus der Falle, was eine Separation ermöglicht³⁵⁴. Dies soll beispielhaft für einen FT-ICR-Analysator näher beschrieben werden, da ein solcher im Rahmen dieser Dissertation für diverse Proteinmessungen verwendet wurde.

Bei einem FT-ICR-Analysator werden die Ionen zunächst gespeichert, indem diese durch ein Magnetfeld B auf Kreisbahnen mit geringen Zyklotronradien gezwungen werden, wobei die Umlauffrequenz durch die Zyklotronfrequenz f_{Zyk} gegeben ist (siehe Gleichung (6))³⁴⁹.

$$f_{Zyk}\left(\frac{m}{z}\right) = \frac{zeB}{2\pi m} = \frac{eB}{2\pi} \cdot \left(\frac{m}{z}\right)^{-1} \quad (6)$$

2 Grundlagen

Durch kurzzeitige Zuschaltung eines breitbandigen elektrischen Wechselfeldes (Chirp) finden Resonanzphänomene bei den gespeicherten Ionen statt, da dieses Anregungssignal auch die Zyklotronfrequenzen der Ionen umfasst^{350–352}. Hierbei wird erreicht, dass sich die angeregten Ionen synchron auf einer Kreisbahn mit einem größeren Zyklotronradius bewegen^{349,350,353}. Wird das Anregungssignal abgeschaltet, so bewegen sich die Ionen auf dieser Kreisbahn in Phase weiter, wobei die Umlauffrequenz der einzelnen Ionen durch deren Zyklotronfrequenz gegeben ist^{349,352,353}. Da die Zyklotronfrequenz von m/z abhängt (siehe Gleichung (6)), erfolgt hierdurch eine Separation von Ionen mit unterschiedlichen m/z .

Nach der Separation der Ionen im Analysator müssen diese detektiert werden. Dies geschieht mit Hilfe eines Detektors^{313,320,346}. Hierbei existieren verschiedene Bauweisen, wie z. B. Faraday-Auffänger oder Sekundärelektronenvervielfacher (SEV)^{314,316,355,356}. Letztere werden in Photomultipliern und Mikrokanalplatten (MCP) verwendet^{314,316,356}. Treffen Ionen auf solche Detektoren, so werden zunächst Sekundärelektronen emittiert^{317,357,358}. Diese können direkt weitere Sekundärelektronen erzeugen, wodurch eine Amplifikation des ursprünglichen Signals verursacht wird (MCP-Detektor) oder durch Auftreffen auf einen Szintillator Photonen erzeugen, welche in einem Photomultiplier ein verstärktes Signal auslösen (Daly-Detektor)^{278,344,356–359}.

In Massenspektrometern, welche einen FT-ICR-Analysator aufweisen, fungieren Elektroden als Detektoren^{349,353}. Dadurch, dass die Ionen durch deren Anregung auf einen größeren Zyklotronradius gelangen, kommen diese in die Nähe der radial angeordneten Detektorplatten und lösen dadurch bei jedem Umlauf ein Signal aus^{350,352}. Aufgrund der unterschiedlichen Zyklotronfrequenzen der Ionen entsteht somit ein kumuliertes Mess-Signal, welches als Superposition der einzelnen Ionenbewegungen angesehen werden kann^{349,353}. Durch eine Fourier-Transformation können daraus die verschiedenen m/z der Ionen hergeleitet werden^{349,351,353}.

3 Originalpublikationen

3.1 Schenkelberger *et al.* (2017)

Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system.

Schenkelberger, M., Shanak, S., **Finkler, M.**, Worst, E. G., Noireaux, V., Helms, V., and Ott, A.

Physical Biology (2017), *14*, 1–11.

DOI: 10.1088/1478-3975/aa5d37

<https://doi.org/10.1088/1478-3975/aa5d37>

© IOP Publishing. Reproduced with permission. All rights reserved.

Physical Biology



PAPER

Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL systemRECEIVED
13 September 2016REVISED
9 January 2017ACCEPTED FOR PUBLICATION
31 January 2017PUBLISHED
1 March 2017M Schenkelberger¹, S Shanak^{2,4}, M Finkler¹, E G Worst¹, V Noireaux³, V Helms² and A Ott¹¹ Department of Experimental Physics, Saarland University, Saarbrücken, 66041, Germany² Center for Bioinformatics, Saarland University, Saarbrücken, 66041, Germany³ School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, United States of America⁴ Arab American University-Jenin, Jenin, PO Box 240, PalestineE-mail: albrecht.ott@physik.uni-saarland.de**Keywords:** *Escherichia coli* cell extract, cell-free expression, MeCP2, brain-derived neurotrophic factor (BDNF), methyl-CpG binding domain (MBD), epigenetic regulation, molecular dynamics simulationsSupplementary material for this article is available [online](#)**Abstract**

Cytosine methylation plays an important role in the epigenetic regulation of eukaryotic gene expression. The methyl-CpG binding domain (MBD) is common to a family of eukaryotic transcriptional regulators. How MBD, a stretch of about 80 amino acids, recognizes CpGs in a methylation dependent manner, and as a function of sequence, is only partly understood. Here we show, using an *Escherichia coli* cell-free expression system, that MBD from the human transcriptional regulator MeCP2 performs as a specific, methylation-dependent repressor in conjunction with the BDNF (brain-derived neurotrophic factor) promoter sequence. Mutation of either base flanking the central CpG pair changes the expression level of the target gene. However, the relative degree of repression as a function of MBD concentration remains unaltered. Molecular dynamics simulations that address the DNA B fiber ratio and the handedness reveal cooperative transitions in the promoter DNA upon MBD binding that correlate well with our experimental observations. We suggest that not only steric hindrance, but also conformational changes of the BDNF promoter as a result of MBD binding are required for MBD to act as a specific inhibitory element. Our work demonstrates that the prokaryotic transcription machinery can reproduce features of epigenetic mammalian transcriptional regulatory elements.

Introduction

Methylation of DNA plays an essential role in the epigenetic regulation of mammalian gene expression. The enzymatic addition of a methyl group to the DNA base cytosine takes place at CpG positions. About 40% of mammalian genes contain CpG islands (CGIs), genomic regions with a higher frequency of CpG dinucleotides than elsewhere, in their promoters and exonic regions. The CGI promoters are mostly non-methylated [1]. Cytosine methylation is predominant in eukaryotic genomes while prokaryotes mainly exhibit adenosine methylation. A recent study [2] showed that cytosine methylation can be associated with stationary phase prokaryotic gene expression and has a weak influence on the exponential growth phase.

The human protein MeCP2 [3] belongs to a family of proteins that contain the conserved methyl-CpG

binding domain (MBD). MBD specifically binds to methylated CpG sites, which enables MBD proteins to modulate gene expression as a function of their concentration levels. In binding assays, hyper-methylated promoters become enriched with MBD protein with promoter specific distributions [4]. MeCP2 consists of six domains. Upon its discovery, MeCP2 was understood as a transcriptional repressor, its action based on two domains, MBD and TRD (transcriptional repression domain) [5]. However, it is now well established that MeCP2 has a multifunctional role beyond suppression or activation of transcription [6–8]. MeCP2 was studied by means of a purified *in vitro* transcription system [9]. MeCP2 exhibits high affinity against the promoter III of the brain-derived neurotrophic factor (BDNF), which contains a single central CpG pair [10–12]. Moreover, MeCP2 transcription control is of interest in connection with many neuropsychiatric disorders, among them the Rett syndrome [13].

Physiological or higher salt concentrations are necessary for the MBD to discriminate methylation with high specificity [14]. MBD interacts with the DNA methyl group through strong electrostatic interactions with a tendency to form dimers. Binding assays have shown a clear tendency to bind to one or more symmetrically methylated CpGs, while the non-specific affinity to DNA is negligible [15]. Zou and colleagues [16] conducted conventional molecular dynamics (MD) simulations as well as alchemical free energy perturbation calculations to suggest that the increased hydrophobic interaction in the methylated MBD protein/DNA complexes strengthens binding. While hydrophobic attraction has been shown to play a major role in methylation recognition of the methylation-binding domain of MBD1 [17], crystallographic investigations of the MBD from MeCP2 bound to the methylated BDNF promoter suggest that it makes contact and recognizes methylation through the hydration pattern of the major groove [11]. The recognition process of methylated DNA by MBD involves only little direct interaction at the CpG dinucleotide while interactions with flanking bases remain limited to the DNA backbone including hydrating water molecules [11]. An extensive binding study on MeCP2 revealed the requirement of A/T sequences adjacent to the CpG motif for MBD recognition [12]. This property was traced to amino acids 78–90 of MBD. In contrast, structural studies suggest that this is caused either by an Asx-ST motif at the C terminus, or an AT hook of MeCP2 that is not part of MBD [11]. The increased propeller twist of the AT run could also play a role. However, the AT-run as a prerequisite for MBD binding could not be firmly established on structural grounds [18].

Escherichia coli cell-free expression systems are useful for the synthesis of human proteins, antibiotic peptides, and the incorporation of non-canonical amino acids [19, 20], among many other applications. The molecular machinery of the cytoplasmic extract performs transcription and translation in a single test tube. The molecular composition of a cell-free reaction, DNA template stoichiometry as well as transcriptional elements, can be more tightly controlled and more easily varied than *in vivo*. In particular, these extracts give the experimentalist the opportunity to combine functional biomolecular elements in ways difficult to achieve otherwise. The expression of fluorescent protein enables real time, quantitative measurement of expression dynamics.

Noireaux and colleagues developed a new system that utilizes the endogenous *E. coli* RNA polymerase and sigma factor 70 [21, 22]. This expression system produces recombinant proteins in the micromolar range within a few hours. Compared to the common bacteriophage based transcription systems, such as T7, it presents the advantage of preserving the bacterial transcription toolbox.

In this work, we investigate the role of the methyl-CpG binding domain of MeCP2 as a possible

transcriptional modulator for prokaryotic gene expression in conjunction with the BDNF promoter. To this aim, we use an *E. coli* cell-free expression system with the endogenous *E. coli* RNA polymerase and sigma factor 70 [21, 22]. We combine our experiments with MD simulations that study how MBD binding and methylation affect the structure of the wild-type DNA compared to a mutant.

Methods

Cell-free expression is illustrated in figure 1. The BDNF promoter (see table 1) is cloned into a plasmid in such a way that it overlaps the –35 and –10 consensus sequences of the *E. coli* promoter, specific to the housekeeping transcription factor sigma-70 present in the cell-free reaction. As a function of the methylation of the central CpG site, the BDNF promoter regulates the transcription of a gene that codes for green fluorescent protein (GFP) as a reporter.

Cell-free extract and plasmid preparation

The crude cell-free extract is prepared according to an established protocol [21, 22].

All plasmids stem from pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (Addgene #40019) [22]. Its ColE1 origin of replication is replaced by a p15A. To avoid toxicity due to an over-expression of the recombinant protein during plasmid amplification, the strong lambda phage promoter Pr is flanked by the operons OR1 and OR2 [22]. This enables the temperature sensitive lambda repressor Cl857 to stabilize the plasmid during cell growth and avoid toxicity (*E. coli* strain KL740, Yale CGSC #:4382). The plasmid design involves cloning of the genes and their regulatory parts into the above described plasmid according to standard procedures of molecular biology. The wild type (wt) and several mutant versions of the BDNF promoter are cloned in between restriction sites NheI and SphI. The sequences overlap the –35 and –10 consensus sequence of the promoter specific to the *E. coli* housekeeping transcription factor sigma-70 present in the extract. Here, GFP refers to the enhanced green fluorescent protein (GenBank: CAD97424.1), truncated and modified in the N- and C-terminal [23]. MBD refers to the methylation binding domain of MeCP2 (amino acid 78-162, GenBank: 4202), UTR1 to the untranslated region containing the T7 g10 leader sequence for highly efficient translation initiation (GeneBank: M35614.1 [24];), T500 to the transcription terminator [25], BDNF to the promoter III of the mouse-brain-derived neurotrophic factor in the wild type (wt) and mutated (M2) version, and OR2-OR1-Pr to the lambda repressor Cro promoter (GenBank: J02459.1). Plasmid concentrations are determined using either QuantiFluor (Promega, USA), or Nanodrop 2000c UV-Vis Spectrometer (Thermo Fisher Scientific GmbH, Dreieich, Germany). Plasmid sequences are determined by a sequencing service (Microsynth AG, Switzerland or

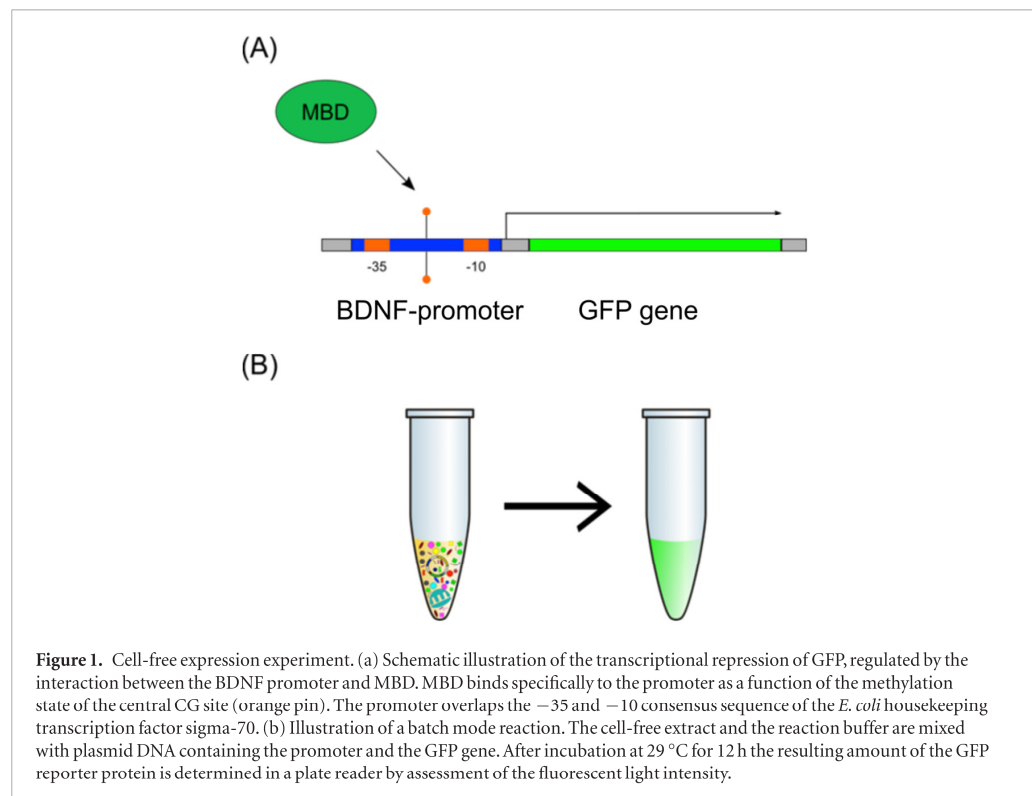


Table 1. Promoter sequences of the studied BDNF promoters, and the plasmid coding for MBD. All promoters contain a central CpG motif (bold), which can either be methylated or non-methylated. In the case of the four mutations of the BDNF promoter (M1–M4), the modified bases are marked as minuscule.

Promoter	Sequence (5'-3')
BDNF-wt	TCTG-GAA-CGG-AAT-TCT-TC
BDNF-M1	TCTG-GAA- CGc -AAT-TCT-TC
BDNF-M2	TCTG-GAA-CGG-Agc-cCT-TC
BDNF-M3	TCTG-Ggg-CGG-AAT-TCT-TC
BDNF-M4	TCTG-Ggg-CGG-Agc-cCT-TC

SEQLAB Sequence Laboratories Göttingen GmbH, Germany). Enzymatic CpG methylation of all plasmids is performed *in vitro* using methyltransferases, *M.SssI* or *HhaI* (New England Biolabs, USA). Complete methylation of the plasmids is analyzed by restriction digest with the methylation dependent endonuclease *HhaI* (New England Biolabs, USA) and subsequent gel electrophoresis (figure S1 (stacks.iop.org/PhysBio/14/026002/mmedia)).

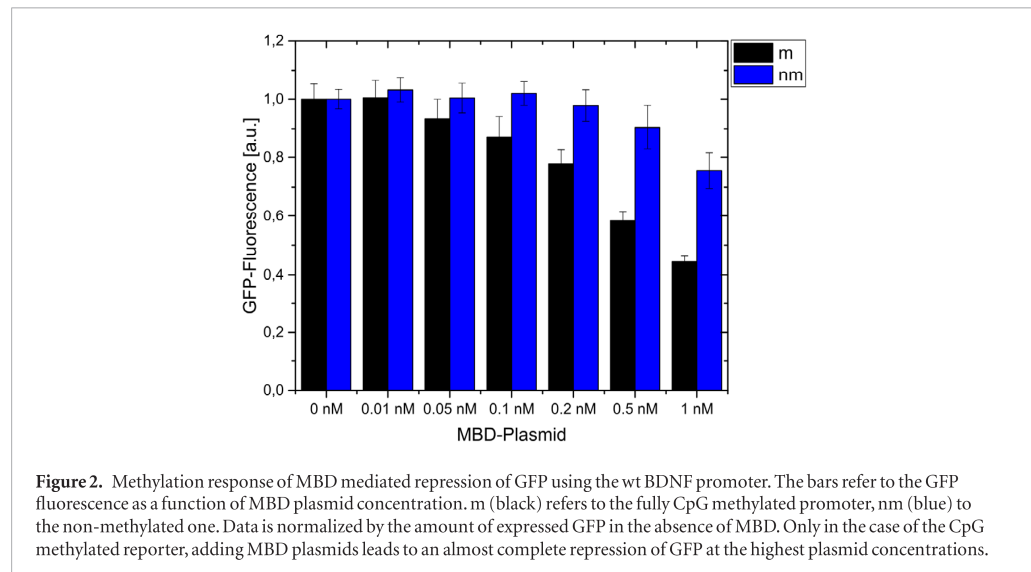
***In vitro* expression experiment**

A typical cell-free reaction consists of 33% (v/v) extract and 67% (v/v) reaction buffer including the DNA templates. The reaction buffer is composed of 50 mM HEPES (pH value 8), 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg ml⁻¹ tRNA *E. coli*, 0.26 mM coenzyme A, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-phosphoglyceric acid (PGA), 1 mM DTT, and 2% (v/v) PEG8000 (all Sigma-Aldrich, Switzerland). Addition of amino

acids (0.5 mM of each of the 20 canonical amino acids, except for Leu that is 0.42 mM), magnesium glutamate (0–10 mM), and potassium glutamate (0 – 120 mM) are adjusted depending on the batch of extract and plasmid concentration. The system is prepared in such a way that the concentrations of these three components can be adjusted independently for any reaction. A typical batch mode reaction is prepared on ice by thoroughly mixing extract, buffer and plasmid DNA at specified concentrations. The solution is split into aliquots of 6 μl . In this case, diffusion of oxygen into the extract is not a limiting factor [22]. The reaction tubes are incubated at 29°C for at least 12 h. After incubation the aliquots are transferred into a single vessel. 10 μl of the solution is transferred into a 384 multi well plate and the fluorescence intensity of the synthesized GFP is determined via plate reader (Polar Star Optima, BMG Labtech, Germany).

Prokaryotic *in vitro* expression of CpG methylated DNA

In bacteria the methylation sensitive restriction system *mcr* is believed to cleave CpG methylated DNA. *Mcr* acts as a primitive immune system [26] in *E. coli*. Figure S2 reveals that the *mcr* system is also present in our cell free extract and degrades CpG methylated reporter plasmids carrying *mcr* recognition motifs. In this case we do not detect any expression of the reporter gene. All CpG methylated reporter plasmids used in this study are devoid of any *mcr* recognition motifs. Their stability is confirmed by the successful expression of GFP from these reporter plasmids (figure S6).



Based on the results in figure S6, we adjusted the reporter plasmid concentration to 5 nM for all experiments. For higher plasmid concentrations, sharing effects of the expression machinery are common [26]. Further, we checked that the presence of recombinant MBD from a different source, or MBD expressed from plasmid DNA within the cell free reaction system lead to a comparable result (figure S3).

MD simulations

As a structural reference for the MeCP2:DNA complex, we use the x-ray structure of the BDNF promoter bound to the methyl-binding domain (PDB: 3C2I). The MD simulations are performed with the GROMACS 4.5.5 package [27] using the CHARMM27 force field [28] and the TIP3P water model [29]. The parameters for 5-methyl-cytosine are used as defined in the CHARMM force field. Systems with unbound DNA duplex strands or protein:dsDNA complexes are placed in a dodecahedral water box of 16 nm box dimensions with 0.10 mol l^{-1} of KCl added, so that the system has an overall zero electrostatic charge. The total size of the simulated systems is 56190 atoms for DNA solvated in a water box, and 56370 atoms for the solvated protein-DNA complex. Boundary conditions are periodic. Long-ranged Coulombic interactions beyond a cutoff of 13 Å are computed by the particle-mesh Ewald (PME) summation method [30]. The non-bonded Lennard-Jones interactions are computed using a smooth cutoff of 13 Å. The integration time step is set to 1 fs.

At first, each simulated system is energy-minimized for 50000 steps using the steepest descent algorithm followed by a second energy minimization for 10000 steps using a quasi-Newtonian algorithm with the low-memory Broyden-Fletcher-Goldfarb-Shanno approach. The tolerance is set to $1.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. After that, the system is heated to 310 K during 4 ps. Then, each system is subjected to 2.0 ns-equilibration in the NVT

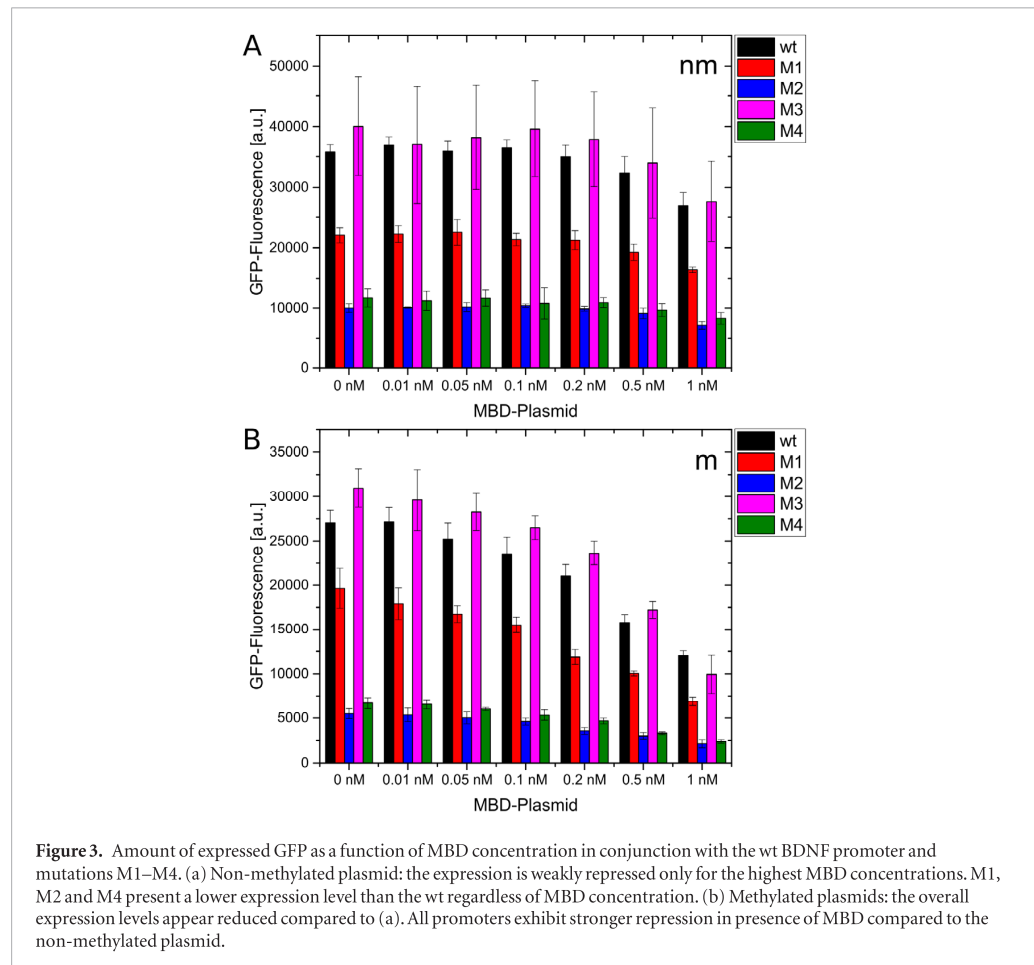
ensemble with harmonic restraints applied to all protein and DNA heavy atoms. The temperature is kept at 310 K by applying leap-frog stochastic dynamics forces with a damping coefficient of 0.1 ps^{-1} . With restraints kept, each system is further equilibrated for 0.5 ns in the NPT ensemble, and then for another 1.0 ns without restraints. MD simulations of 100 ns length are performed for the fully methylated DNA (the wild type), the non-methylated DNA, and methylated and non-methylated versions of the experimentally studied M2 mutant (modeled with 3DNA software [31]), which shows the strongest deviations from the wild type among the four studied mutants. Simulations are conducted in two replicates each. The results of the two replicates are almost indistinguishable. Thus the results for the second replicate are only shown in one table, but not in the figures. For the computation of RMSF fluctuations, the trajectories of the two replicates (100 ns each) are concatenated and fitted to the initial structure.

Results

MBD mediated repression of GFP with the wt version of the BDNF promoter

For the wt BDNF promoter we determine the overall amount of expressed GFP of the cell free reaction mix as a function of MBD plasmid DNA concentration (figure 2).

In our experiments, the final concentration of recombinant proteins attains the micromolar range [22] (figure S3). In case of the CpG methylated reporter, adding MBD plasmids entails repression of GFP expression. In case of the non-methylated reporter the same level of repression requires about $5 \times$ higher plasmid concentrations. If BDNF is replaced by a different promoter that contains a repetition of the CpG recognition motif, this does not modulate GFP expression in a methylation dependent manner (figure S5). We conclude that it is the BDNF promoter that is responsi-



ble for the observed methylation-dependent repression and that MBD recognition is specific to this promoter.

(wt, M1–M4) in the case of CpG methylated and non-methylated reporter plasmids.

Sequence mutations of the BDNF promoter

We experimentally test four mutations in the BDNF promoter (table 1) for changes in transcriptional regulation. M1 and M3 are designed to study the influence of the flanking bases of the central CpG motif where methylation is discriminated. M2 breaks the run of the four AT base pairs downstream of the CpG motif [12]. M4 combines the mutations M2 and M3 within a single promoter. The central CpG motif of the BDNF promoter remains untouched in any case.

Figure 3 shows the expressed GFP concentrations as a function of plasmid methylation state and the type of mutation. Of the non-methylated plasmids, M1 generates lower expression levels than the wt, while M3 tends to exceed this level. M2 and M4 exhibit a strongly reduced expression level to about 20% of wt. If methylated, in presence of the MBD plasmid the expression levels of all constructions appear reduced. In an additional but otherwise identical experiment we normalize the measured GFP fluorescence to the value in absence of the MBD plasmid (figure 4). The results reveal a comparable, relative repression of all constructions

Simple kinetic model

For expression regulation, both, MBD and the sigma factor 70 bind to the BDNF-promoter. MBD binding (k_M^{on}) is limited to the CpG pair that occurs in the wt and M1–M4 promoter only once. In contrast the sigma factor 70 recognizes the -35 and -10 region flanking the promoter. Conformational changes of the promoter by different mutations will influence the association constant of the sigma factor 70 (k_σ^{on}) which has multiple contact points within the promoter. Binding also depends on the presence of MBD because of structural hindrance. When binding to the BDNF promoter in presence of MBD ($k_{\sigma,+}^{\text{on}}$) or in absence ($k_{\sigma,-}^{\text{on}}$), the sigma factor 70 initiates transcription and dissociates, ($k_{\sigma,-}^{\text{on}} \gg k_{\sigma,+}^{\text{on}}$). This is fast, and we can neglect the influence of the presence of the sigma factor 70 on MBD binding. The promoter can occur in 4 states (figure 5).

The expression rate is given by $E = A_- E_- + A_+ E_+$ where $A_- = \frac{[(3)]}{[(3)] + [(4)]}$ and $A_+ = \frac{[(4)]}{[(3)] + [(4)]}$ are the expression rates without (E_-) or with MBD bound (E_+) ($E_- > E_+$). $[(x)]$ denotes the concentration of

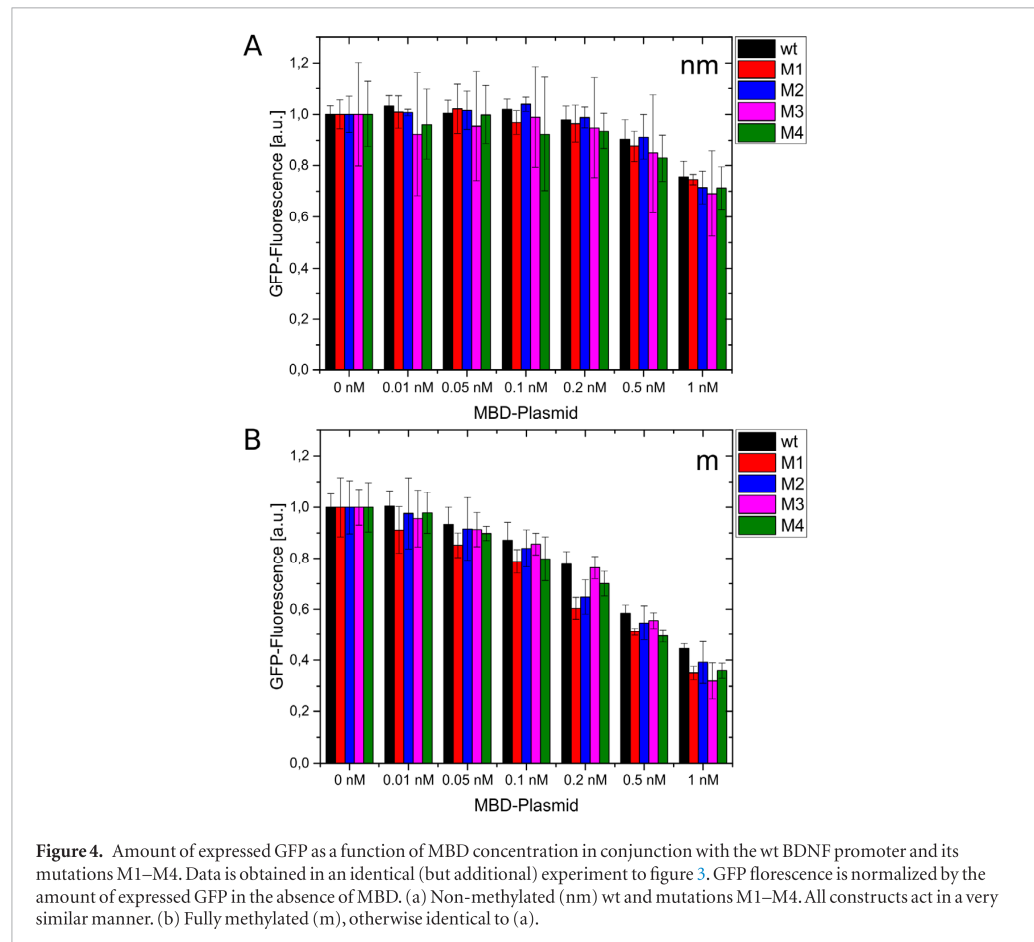


Figure 4. Amount of expressed GFP as a function of MBD concentration in conjunction with the wt BDNF promoter and its mutations M1–M4. Data is obtained in an identical (but additional) experiment to figure 3. GFP fluorescence is normalized by the amount of expressed GFP in the absence of MBD. (a) Non-methylated (nm) wt and mutations M1–M4. All constructs act in a very similar manner. (b) Fully methylated (m), otherwise identical to (a).

the promoter in state x . Using $A_- + A_+ = 1$, and $K_{D,M} = \frac{[^{(3)}][MBD]}{[^{(4)}]}$ yields:

$$E = E_+ + \frac{(E_- - E_+) \cdot K_{D,M}}{K_{D,M} + [MBD]} \quad (1)$$

Figure 6 shows the GFP-Fluorescence as a function of MBD-Plasmid concentration in case of the methylated wt BDNF promoter and the fit with equation (1). The fits of the experimental data of the mutants M1–M4 show similar accuracy (supplementary material, figure S8).

In all cases (wt, M1–M4) E_- corresponds to the measured value $E([MBD\text{-Plasmid}] = 0 \text{ nM})$. The values of E_+ and $K_{D,M}$ were obtained from a fit (Origin). While E_+ is in the range of approximately 1/8 to 1/5 of E_- , $K_{D,M}$ is between 0.29 nM and 0.55 nM for the different mutants (supplementary material, S9). However, here the calculated dissociation constant is given as a function of MBD-Plasmid concentration. Figure S3 relates MBD-Plasmid to MBD-Protein concentrations (0.5 nM MBD-Plasmid $\approx 1 \mu\text{M}$ MBD protein). This leads to an effective dissociation constant of 0.5 μM to 1 μM .

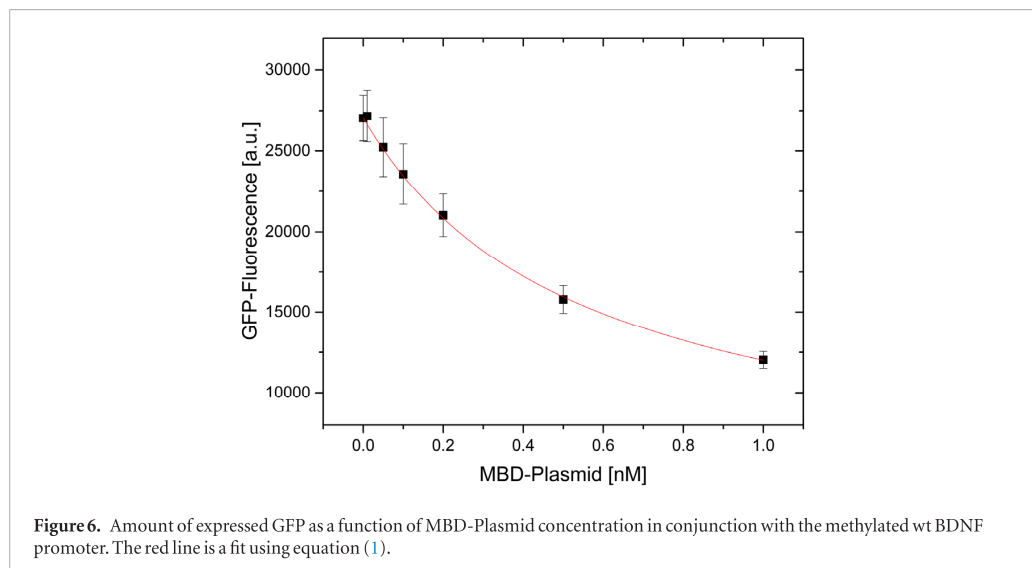
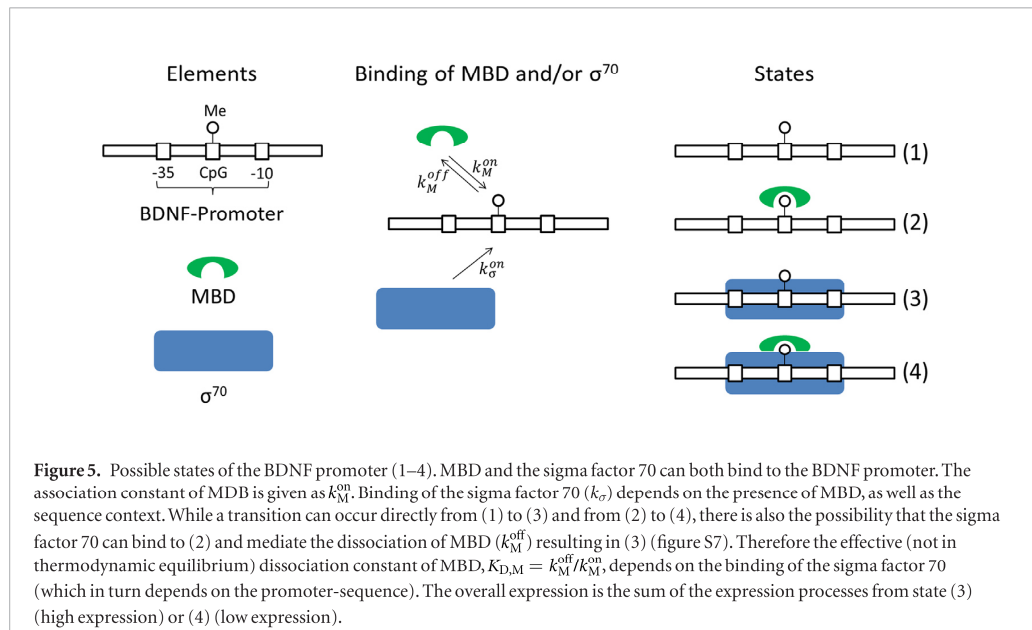
In the case where the BDNF promoter was replaced by the promoter with the CpG-repetitions there is no repression of the expression with increasing MBD-Plasmid concentration. In terms of the model, this is achieved by a higher $K_{D,M}$ and a decreased E_-/E_+ -ratio.

MD simulations

MD simulations of free promoter DNA as well as the MBD:DNA complex reveal stable conformations for methylated and non-methylated DNA (RMSD $\sim 2.0 \text{ \AA}$ from the average structure). Methylated, MBD-bound promoter DNA presents the smallest root mean square fluctuations (RMSF) of the heavy atoms in its backbone whereas non-methylated MBD bound DNA exhibits the largest ones (figure 7), exceeding even those of the unbound DNA, methylated or not.

The simulations show that binding of methylated DNA (wild-type and M2) to MBD induces a widening of the major DNA groove at the binding interface from 15 \AA to about 20 \AA (figure 8). This change is less pronounced when MBD binds to non-methylated DNA (wild-type and M2).

Figure 9 shows the fraction of DNA base pairs adopting a B-DNA conformation (also known as B-fiber ratio). For unbound wt DNA, both in the methylated and non-methylated states, roughly 80% of the base-pairs are in B-DNA conformation on average. Unbound M2-DNA is characterized by lower ratios of 70% ($m\text{DNA}_{M2}$) and 65% ($n\text{DNA}_{M2}$). Upon binding to MBD, the B-fiber ratios either decrease by about 10% to 70% ($m\text{DNA}_{wt}$ and $n\text{DNA}_{wt}$) or by about 5% to 65% ($m\text{DNA}_{M2}$). In the case of non-methylated

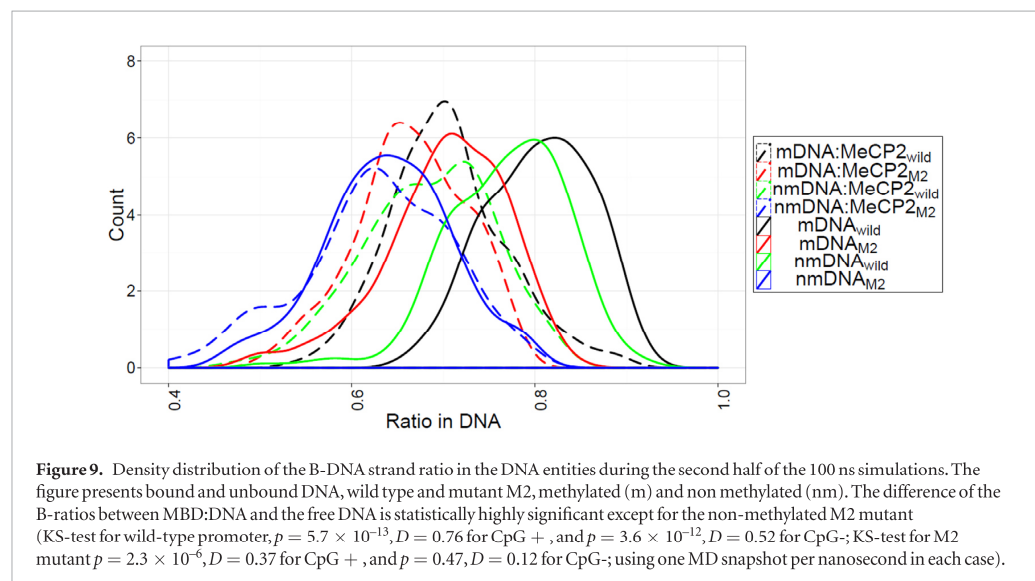
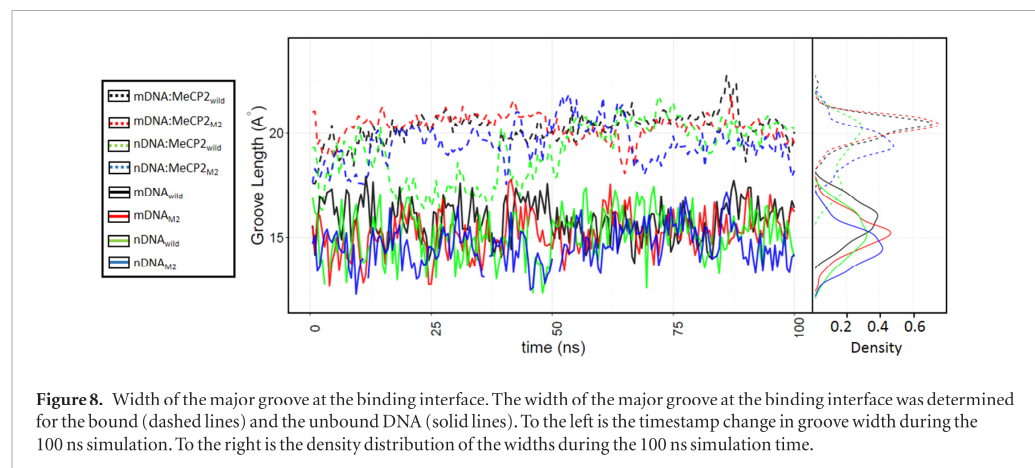
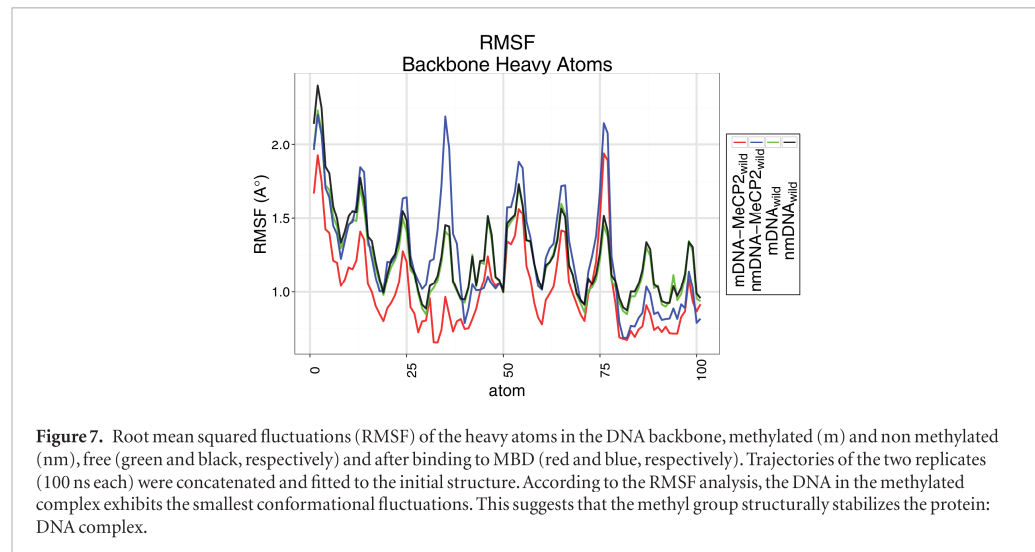


M2, the B-fiber ratio decreases insignificantly to a low value of 62%.

The handedness (see supplementary material and figure S4 for further details on handedness) describes the helical twisting of DNA. It can be used to check for transitions from compact, right-handed B-DNA to the relatively extended, left-handed Z-DNA. The handedness values indicate that upon binding of MBD to wt DNA the CpG recognition motif untwists partially from the perfect B-DNA fiber (table 2, figure 10). This is more pronounced if the DNA is methylated (from 0.552 to 0.446 on average) than for the non-methylated form (from 0.548 to 0.480). M2, however, displays a handedness that is practically independent of methylation while fluctuations are larger in the unbound state. The value of handedness leans towards the methylated wt complex.

Discussion

Since the molecular environment of *E. coli* is tuned for prokaryotic expression, one might suspect eukaryotic transcriptional elements not to function in our experiments. Adenine methylation of the DNA sequence motif 5'-GATC-3' by the enzyme deoxyadenosine methylase (Dam) plays an important role in the timing of initiation of DNA replication in *E. coli*, as well as in the coordination of cellular events, DNA mismatch repair, and gene regulation [32]. At the same time, cytosine methylation is easily recognized as foreign. However, we find that our experiments are not affected by any of the above. We use solely MBD, the recognition domain, instead of the entire MeCP2 protein and find that the expression downstream of the BDNF promoter is strongly repressed by MBD only if



this promoter is methylated (figure 2). This happens at MBD concentrations that are typical for transcription factors in cells [33], however, they are well below the

MeCP2 concentrations found in cell nuclei [34]. An unrelated promoter, inserted at the place of the BDNF sequence, does not exhibit any methylation dependence

Table 2. Handedness of the central dinucleotide steps, namely the (36T-7A-35T-85CM) and the (7DA-35DT-85CM-34DG) in bound and unbound states for wild-type BDNF promoter and for the M2 mutant. The 2 values given in each field belong to 2 independent replicate simulations of 100 ns length each.

	DNA in complex		Unbound DNAs	
	Central step 1 (36T-7A-35T-85CM)	Central step 2 (7DA-35DT-85CM-34DG)	Central step 1 (36T-7A-35T-85CM)	Central step 2 (7DA-35DT-85CM-34DG)
Wild type	0.408 ± 0.057	0.459 ± 0.045	0.524 ± 0.043	0.582 ± 0.038
Methylated	0.441 ± 0.067	0.476 ± 0.063	0.521 ± 0.042	0.580 ± 0.037
Mutant 2	0.402 ± 0.052	0.462 ± 0.038	0.523 ± 0.042	0.579 ± 0.041
Methylated	0.403 ± 0.052	0.462 ± 0.038	0.522 ± 0.044	0.576 ± 0.042
Wild type	0.492 ± 0.055	0.540 ± 0.048	0.518 ± 0.044	0.575 ± 0.042
Non-methylated	0.411 ± 0.084	0.477 ± 0.058	0.523 ± 0.047	0.576 ± 0.043
Mutant 2	0.405 ± 0.071	0.459 ± 0.055	0.526 ± 0.046	0.577 ± 0.039
Non-methylated	0.408 ± 0.048	0.451 ± 0.038	0.452 ± 0.200	0.500 ± 0.214

although it contains a repetition of symmetrical CpG motifs (figure S5) that were shown to considerably enhance MBD affinity in binding assays [15]. All of the above exhibits a high degree of specific repression that is astonishing, given that we are in a prokaryotic molecular environment, and MBD represents a rather small portion of MeCP2 that, to our knowledge, has not been described as a functional eukaryotic regulator *per se*.

We included four mutations into the promoter and examined them for expression efficiency (figures 3 and 4). Both, M1 and M3 carry a mutation of one of the bases flanking the central CpG recognition motif. Based on structural data (14), Ho *et al* showed that direct MBD binding with respect to the flanking bases is limited to the DNA backbone. In agreement, in our experiments the flanking bases only change the expression level of the target gene in the absence of MBD but they do not change the relative repression that is caused by MBD.

Mutant M2 had a broken run of AT bases downstream of the CpG motif. This AT-run was reported to be essential for MBD binding to BDNF [12]. We observe that M2 leads to a lower expression level of the target GFP gene even in the absence of MBD (figure 3). We conclude that the AT-run of the BDNF promoter enhances the expression by the *E. coli* polymerase. For M2, methylation produces repression of comparable magnitude as the wild type. In agreement with the structural analysis [11], the AT-run does not appear as a requirement for MBD binding in our study.

The observations made within the experiments are well described by a simple kinetic model assuming that the BDNF promoter can occur in four states. This Michaelis–Menten type description fits the data well. While the value for the effective dissociation constant of MBD, $K_{D,M}$, from fitting the experimental data is estimated between 0.5 μ M and 1 μ M, the value found in the literature in equilibrium is about 8.5 nM [35]. However, in our experiment there is non-equilibrium, as well as competition between MBD and the sigma factor 70, and possibly many other molecules from the *E. coli* extract. As a result it may not be very useful to compare both values.

In MD simulations, the width of the major groove of the promoter increases upon binding to MBD (figure 8), as with most DNA binding proteins [36]. This improves the accessibility of the functional groups of DNA and favors the specificity of protein–DNA contacts. In parallel to this, the fraction of DNA base pairs in B-DNA conformation decreases (figure 9) as well as the handedness (table 2, figure 10). Both observables are indications of a subtle conformational transition toward Z-DNA. For wtBDNF promoter, MD simulations predict that its conformational fluctuations are lower for the methylated MBD complex as compared to unbound DNA, or to the non-methylated complex. Also, binding of the methylated BDNF promoter induces a larger opening of the major groove than for the non-methylated form. These observations suggest that MBD forms tighter contacts to methylated BDNF (as evidenced by low conformational fluctuations, see RMSF analysis in figure 7). This matches well the efficiency for repression reported in figure 3, as well as the computational results of Zou and co-workers [16].

The degree of opening of the DNA major groove for methylated M2-DNA and the change in handedness of the central dinucleotide steps, are comparable to that for wt DNA (figures 8 and 10, table 2). This suggests that MBD also binds to the mutant M2 although the AT-run is lacking. On the other hand, in simulations of the unbound M2 promoter, a lower fraction of DNA bases adopt B-DNA conformation (figure 9) than for the unbound wt DNA. Also when bound to MBD, the fraction of B-DNA is smaller for M2-DNA than for wt DNA. Moreover, the distribution of handedness of M2 is distinctly broader than for the wt, regardless of methylation. Both observables point to a more floppy DNA conformation of the M2 promoter. At the same time the data above suggests that the decreased expression level of M2 is due to a DNA conformation that, without being methylated, retains structural features of the repressing, methylated wt MBD complex. We suggest that the AT stretch helps to stabilize in particular the non-methylated DNA conformation in such a way that a high expression-level is achieved.

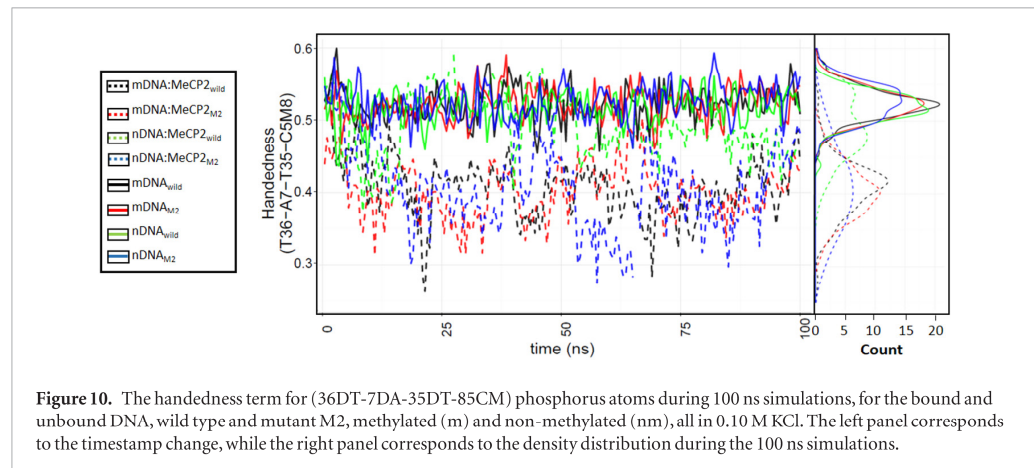


Figure 10. The handedness term for (36DT-7DA-35DT-85CM) phosphorus atoms during 100 ns simulations, for the bound and unbound DNA, wild type and mutant M2, methylated (m) and non-methylated (nm), all in 0.10 M KCl. The left panel corresponds to the timestamp change, while the right panel corresponds to the density distribution during the 100 ns simulations.

As a general observation, the structural transitions upon MBD binding are stronger for the methylated promoter than for the non-methylated one. They extend until the AT stretch, where the mutation of M2 is located. *In vitro* footprinting revealed protection that extends 6 bases from the central motif [15]. We hypothesize that the high degree of cooperativity is a reason for the highly specific BDNF motif recognition that MBD presents in our experiments—in spite of its very localized interaction. In agreement with previous knowledge, our results support the idea that MBD binds to the central CpG pair as a recognition motif. However, we suggest that in order to act as a suppressor, not only steric hindrance plays a role, but also MBD modifies the DNA conformation in such a way that the polymerase binds with reduced affinity. This idea is corroborated by our observation that a repetition of CpGs does not act as an inhibitor, although MBD binds with even higher affinity to this type of sequence [15]. It is clear that a high degree of cooperativity along the DNA BDNF sequence in conjunction with a very localized recognition site implies a certain fragility towards mutations. We speculate that the highly cooperative nature of molecular recognition may contribute to the loss/gain of functions of MBD-proteins, which has been discovered as a result of point mutations in connection with the Rett syndrome [37], a considerable fraction of them identified in the MBD domain. However, a much more detailed study would be required to confirm this hypothesis.

Conclusion

In this study we report the first epigenetic regulation due to a eukaryotic protein and eukaryotic cytosine methylation that is realized in a prokaryotic bacterial cell-free transcription translation system. At first sight, the eukaryotic polymerase and its expression machinery seem to have nothing much in common with prokaryotes. However, the fact that in our experiments MBD alone, as opposed to the entire MeCP2, performs as a specific repressor in conjunction with the BDNF promoter sequence, shows that in our case the situation

is actually quite clear. Whereas binding assays simply detect the formation of MBD-DNA contacts, our cell-free expression system reveals repression of the target gene by MBD that is probably mediated by conformational changes of the connecting DNA stretch. MD simulations suggest subtle cooperative transitions of the DNA conformation. Given the highly localized interaction of MBD, this may well explain the high degree of specificity to the BDNF sequence as observed in our experiments.

Acknowledgments

This work was supported by the collaborative research center SFB 1027 funded by Deutsche Forschungsgemeinschaft (DFG). SS acknowledges a doctoral fellowship from the German Academic Exchange Service (DAAD).

Author contributions

VN, VH, AO, MS, designed research, MF, MS, SS performed research, analyzed data with the help of AO and VH, MF, EGW performed replicate experiments, SS, MS, VH and AO wrote the paper with help from EGW and MF.

References

- [1] Fatemi M, Pao M M, Jeong S, Gal-Yam E N, Egger G, Weisenberger D J and Jones P A 2005 Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level *Nucleic Acids Res.* **33** e176
- [2] Kahramanoglou C, Prieto A I, Khedkar S, Haase B, Gupta A, Benes V, Fraser G M, Luscombe N M and Seshasayee A S N 2012 Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription *Nat. Commun.* **3** 886
- [3] Lewis J D, Meehan R R, Henzel W J, Maurer-Fogy I, Jeppesen P, Klein F and Bird A 1992 Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA *Cell* **69** 905–14
- [4] Lopez-Serra L, Ballestar E, Fraga M F, Alaminos M, Setien F and Esteller M 2006 A profile of methyl-CpG binding domain protein occupancy of hypermethylated promoter CpG

- islands of tumor suppressor genes in human cancer *Cancer Res.* **66** 8342–6
- [5] Nan X, Campoy F J and Bird A 1997 MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin *Cell* **88** 471–81
- [6] Long S W, Ooi J Y Y, Yau P M and Jones P L 2011 A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing *Biosci. Rep.* **31** 333–43
- [7] Chahrour M, Jung S Y, Shaw C, Zhou X, Wong S T C, Qin J and Zoghbi H Y 2008 MeCP2, a key contributor to neurological disease, activates and represses transcription *Science* **320** 1224–9
- [8] Guy J, Cheval H, Selfridge J and Bird A 2011 The role of MeCP2 in the brain *Annu. Rev. Cell Dev. Biol.* **27** 631–52
- [9] Kaludov N K and Wolffe A P 2000 MeCP2 driven transcriptional repression *in vitro*: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery *Nucleic Acids Res.* **28** 1921–8
- [10] Chen W G, Chang Q, Lin Y, Meissner A, West A E, Griffith E C, Jaenisch R and Greenberg M E 2003 Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2 *Science* **302** 885–9
- [11] Ho K L, McNaie I W, Schmiedeberg L, Klose R J, Bird A P and Walkinshaw M D 2008 MeCP2 binding to DNA depends upon hydration at methyl-CpG *Mol. Cell* **29** 525–31
- [12] Klose R J, Sarraf S A, Schmiedeberg L, McDermott S M, Stancheva I and Bird A P 2005 DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG *Mol. Cell* **19** 667–78
- [13] Li W and Pozzo-Miller L 2014 BDNF deregulation in Rett syndrome *Neuropharmacology* **76** 737–46
- [14] Khrapunov S, Warren C, Cheng H, Berko E R, Grealley J M and Brenowitz M 2014 Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity *Biochemistry* **53** 3379–91
- [15] Nan X, Meehan R R and Bird A 1993 Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2 *Nucleic Acids Res.* **21** 4886–92
- [16] Zou X, Ma W, Solov'yov I A, Chipot C and Schulten K 2012 Recognition of methylated DNA through methyl-CpG binding domain proteins *Nucleic Acids Res.* **40** 2747–58
- [17] Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M and Shirakawa M 2001 Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA *Cell* **105** 487–97
- [18] Ho K L 2009 Structural studies of MeCP2 in complex with methylated DNA *Thesis* The University of Edinburgh
- [19] Worst E G, Exner M P, De Simone A, Schenkelberger M, Noireaux V, Budisa N and Ott A 2015 Cell-free expression with the toxic amino acid canavanine *Bioorg. Med. Chem. Lett.* **25** 3658–60
- [20] Worst E G, Exner M P, De Simone A, Schenkelberger M, Noireaux V, Budisa N and Ott A 2016 Residue-specific incorporation of noncanonical amino acids into model proteins using an *Escherichia coli* cell-free transcription-translation system *J. Vis. Exp.* **114** e54273
- [21] Sun Z Z, Hayes C A, Shin J, Caschera F, Murray R M and Noireaux V 2013 Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology *J. Vis. Exp.* **79** e50762
- [22] Shin J and Noireaux V 2010 Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70 *J. Biol. Eng.* **4** 8
- [23] Li X, Zhang G, Ngo N, Zhao X, Kain S R and Huang C C 1997 Deletions of the Aequorea victoria green fluorescent protein define the minimal domain required for fluorescence *J. Biol. Chem.* **272** 28545–9
- [24] Olins P O, Devine C S, Rangwala S H and Kavka K S 1988 The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli* *Gene* **73** 227–35
- [25] Larson M H, Greenleaf W J, Landick R and Block S M 2008 Applied force reveals mechanistic and energetic details of transcription termination *Cell* **132** 971–82
- [26] Shin J and Noireaux V 2012 An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells *ACS Synth. Biol.* **1** 29–41
- [27] Hess B, Kutzner C, van der Spoel D and Lindahl E 2008 GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation *J. Chem. Theory Comput.* **4** 35–447
- [28] Foloppe N and MacKerell A D Jr 2000 All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data *J. Comput. Chem.* **21** 86–104
- [29] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926
- [30] Darden T, York D and Pedersen L 1993 Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems *J. Chem. Phys.* **98** 10089
- [31] Lu X-J and Olson W K 2008 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures *Nat. Protocols* **3** 1213–27
- [32] Braaten B A, Nou X, Kaltenbach L S and Low D A 1994 Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli* *Cell* **76** 577–88
- [33] Milo R, Phillips R and Orme N 2015 *Cell Biology by the Numbers* 1st edn (London: Taylor and Francis)
- [34] Skene P J, Illingworth R S, Webb S, Kerr A R W, James K D, Turner D J, Andrews R and Bird A P 2010 Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state *Mol. Cell* **37** 457–68
- [35] Ghosh R P, Nikitina T, Horowitz-scherer R, Gierasch L M, Uversky V N, Hite K, Hansen J C and Christopher L 2010 Unique physical properties and interactions of the domains of methylated DNA binding protein 2 (MeCP2) *Biochemistry* **49** 4395–410
- [36] Pabo C O and Sauer R T 1984 Protein-DNA recognition *Annu. Rev. Biochem.* **53** 293–321
- [37] Kucukkal T G, Yang Y, Uvarov O, Cao W and Alexov E 2015 Impact of rett syndrome mutations on MeCP2 MBD stability *Biochemistry* **54** 6357–68

Supplementary Material

Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system.

Schenkelberger, M., Shanak, S., **Finkler, M.**, Worst, E. G., Noireaux, V., Helms, V., and Ott, A.

Physical Biology (2017), *14*, 1–11.

DOI: 10.1088/1478-3975/aa5d37

<https://doi.org/10.1088/1478-3975/aa5d37>

© IOP Publishing. Reproduced with permission. All rights reserved.

Supplementary Material

Handedness

The collective variable of handedness was used to characterize the conformational changes of DNA upon MeCP2 binding. It is a good choice for detecting the helical twisting from the right-handed B-DNA to the left-handed Z-DNA. According to the definition by Moradi et al [1], given a set of basepairs, starting at basepair n and ending at basepair m ; the following sequence of atoms is used to describe the collective variable: $P^1_n, P^2_n, P^1_{n+1}, P^2_{n+1}, \dots, P^1_m, P^2_m$; where P^1_n is the atom starting from the 5' position in the n^{th} basepair. As such, the sum can also be started starting at the 5' nucleotide triphosphate from the other end. The total collective variable of handedness for the DNA strand is thus the polynomial sum of the handedness terms, starting at one phosphorus atom each, and ending three bases thereafter (e.g; $P^1_n, P^2_n, P^1_{n+1}, P^2_{n+1} + P^2_n, P^1_{n+1}, P^2_{n+1}, P^1_{n+2} + \dots + P^1_{m-1}, P^2_{m-1}, P^1_m, P^2_m$). Thus, the position of these atoms defines the handedness via:

$$H(p_1 p_2 p_3 \dots p_n) = \sum_{k=1}^{n-3} H(p_i p_{i+1} p_{i+2} p_{i+3})$$

Given a sequence of points A, B, C , and D ; the single handedness term is defined as (figure S4):

$$H(ABCD) = \frac{\overrightarrow{AB}}{|\overrightarrow{AB}|} \times \frac{\overrightarrow{CD}}{|\overrightarrow{CD}|} \cdot \frac{\overrightarrow{EF}}{|\overrightarrow{EF}|}$$

where the points define the vectors \overrightarrow{AB} and \overrightarrow{CD} . The vector \overrightarrow{EF} defines the vector matching the midpoints of \overrightarrow{AB} and \overrightarrow{CD} .

To detect possible features of B-Z transitions in the DNA fibers, the collective variable of handedness was investigated during the 100 ns simulation for the DNA in the bound and free forms (see figure S4). The range of the collective variable for the unbound DNA is between 52-58°. Trajectories used for analysis include the wild type DNA in the bound and unbound forms (CpG+ and CpG-) and the M2 mutant in the bound and unbound forms (CpG+ and CpG-).

Supplementary Figures

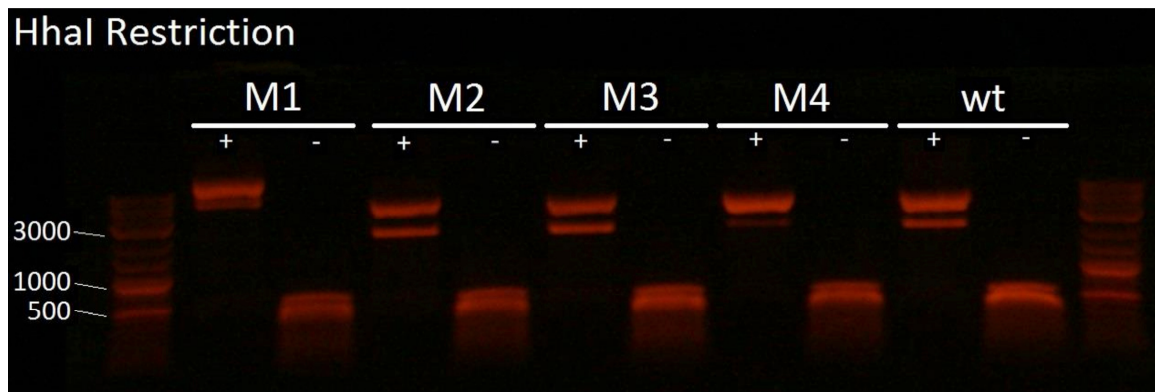


Figure S1. Assessment of the methylation state of the plasmid DNA. Restriction digest of the plasmids pBEST-BDNF-UTR1-GFP-T500 (wt and mutations M1-M4) using the methylation dependent restriction endonuclease HhaI. “+” refers to DNA plasmids previously incubated for three hours with the methyltransferase M.SssI which methylates all CpG sites of the DNA. “-” refers to DNA that has not been incubated with M.SssI. The first and the last lane correspond to a DNA ladder. The size of the three major fluorescent bands are given (number of bases).

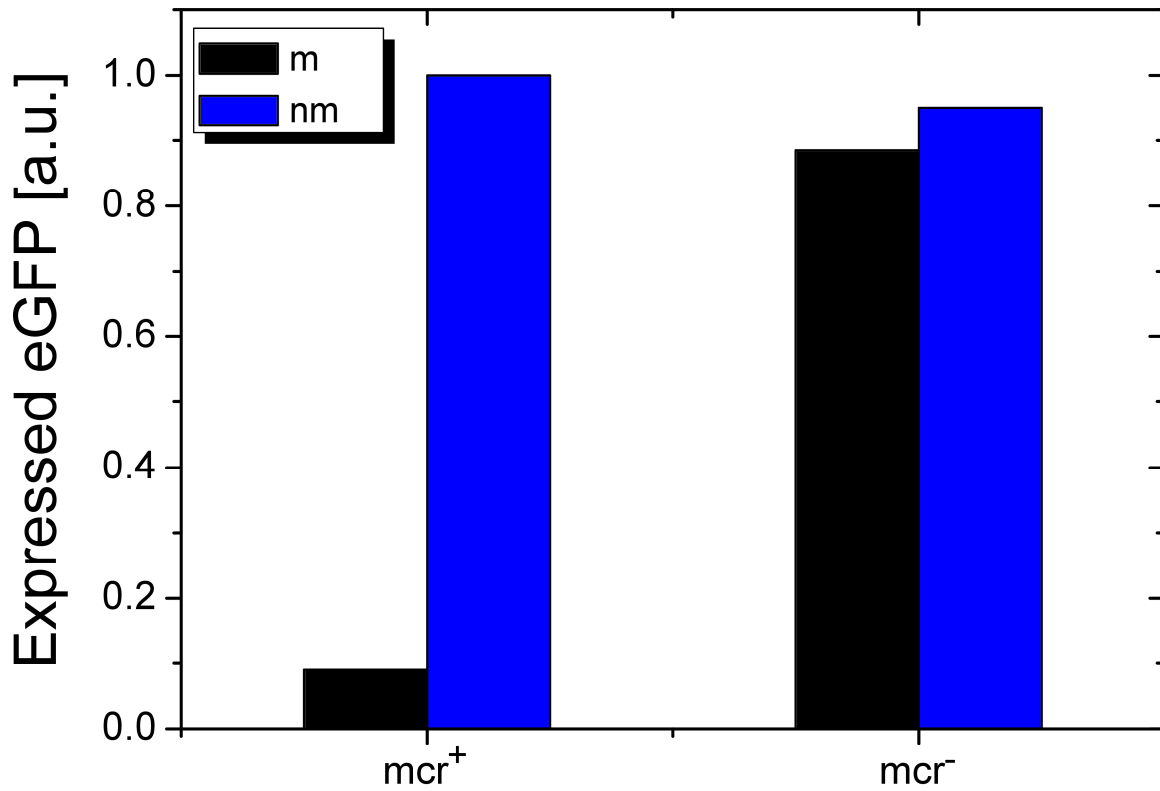


Figure S2. Degradation of CpG methylated DNA in the cell free reaction. The graph shows the GFP expression level using 5 nM of a CpG methylated (black) and unmethylated (blue) *mcr* positive reporter plasmid. Both reporters are expressed with the standard BL21 cell free extract (mcr^+) and a NEB-10-beta extract (mcr^-) not containing the *mcr* restriction system. Data is normalized to the mcr^+/CpG^- case. The concentration of amino acids, magnesium glutamate and potassium glutamate added to the reaction is 0.5 mM, 3 mM, and 50 mM, respectively. CpG methylated DNA is degraded in the mcr^+ extract.

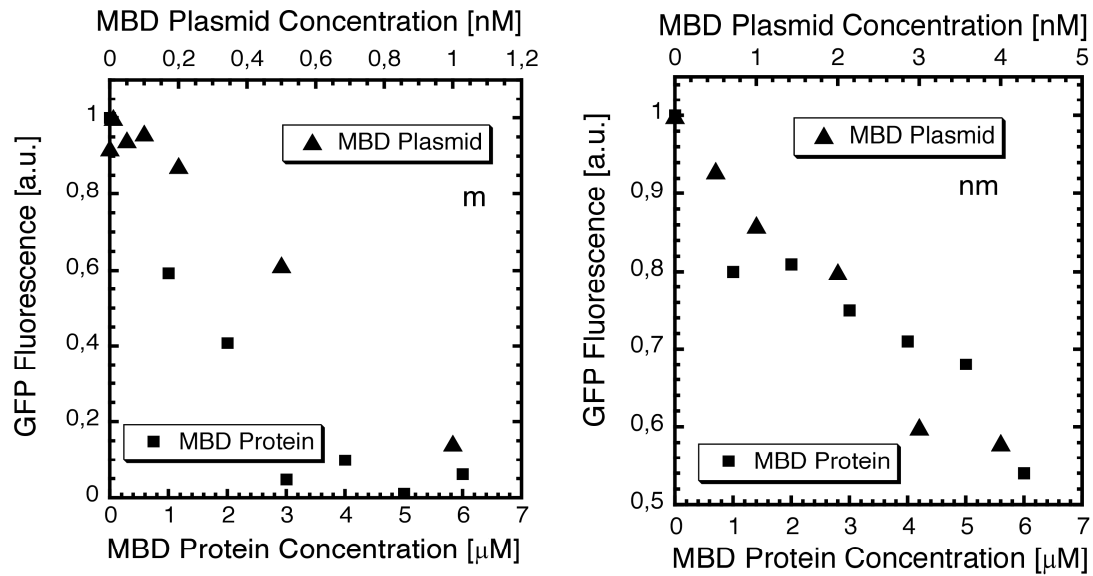


Figure S3. Comparison of GFP expression in the case of the plasmid pBEST-BDNF-wt-UTR1-GFP-T500 for either recombinant MBD protein fragment added to the reaction at a given concentration, or instead MBD expressed from plasmid DNA at a given concentration within the reaction. All data is normalized to the respective amount of GFP expressed in absence of MBD. Left: methylated case, right non methylated. Note the differences in scales.

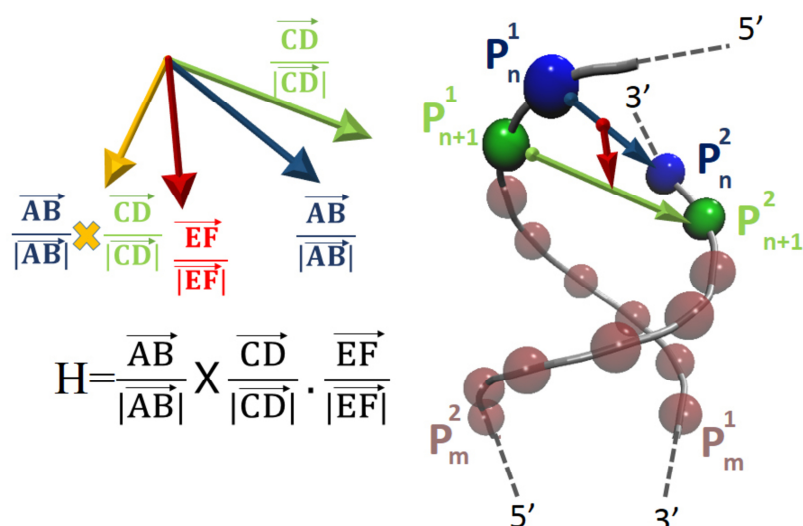


Figure S4. A schematic representation for the collective variable of handedness [1]. On the right-hand side, the figure shows the vectors involved in this definition. P atoms are named in context of the residue number and the strand, with P_n^1, P_n^2 atoms representing the n^{th} basepair P atoms; with 1 and 2 representing the Watson and Crick strands, running in the $5' \rightarrow 3'$ direction, and the $3' \rightarrow 5'$ direction, respectively. For the handedness term including the four atoms $P_n^1, P_n^2, P_{n+1}^1, P_{n+1}^2$, represented in the (ABCD) order here as well, the vectors connecting the atoms and contributing to handedness are defined on the right-hand side, with the blue vector connecting the first two atoms and the green vector connecting the second two atoms. The red vector defines the connection between the two midpoints connecting either vector, in the given direction. On the left-hand side, the definition of handedness is given, in terms of the units vectors of the vectors defined on the right-hand scheme. Vector multiplication, followed by a dot product is assumed, and the final handedness term retrieved. For more details about the definition of the global handedness term in terms of a long run of bases, please refer to the text.

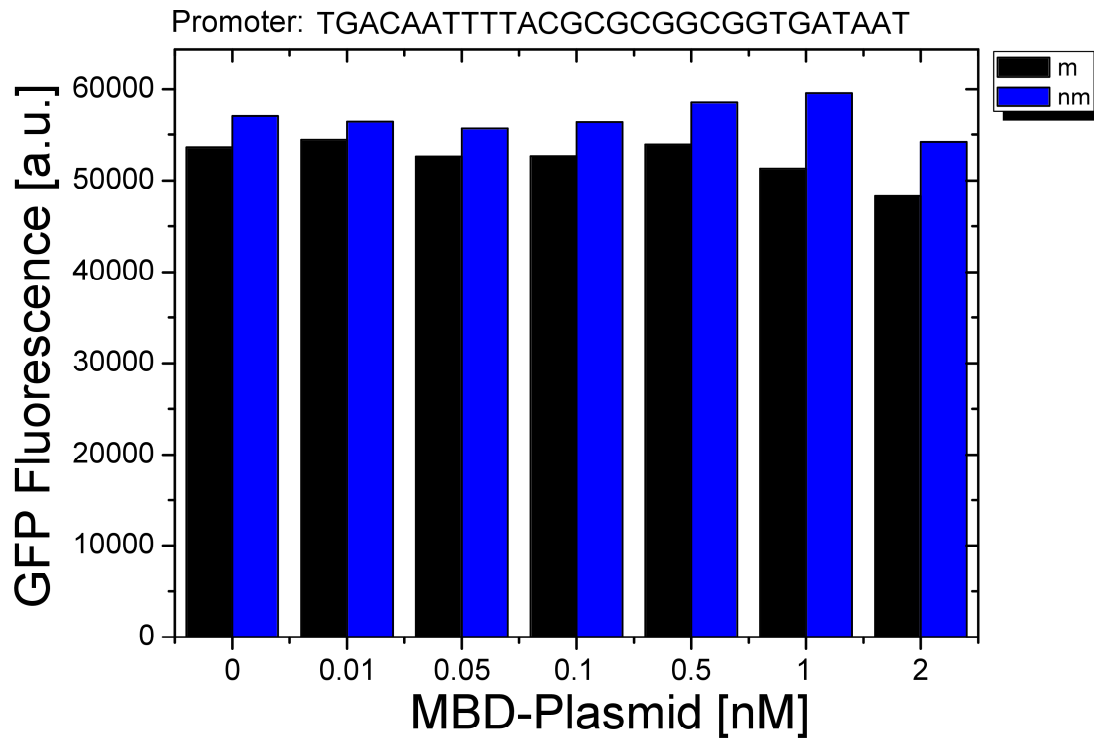


Figure S5. MBD mediated repression of GFP with a different promoter that contains a repetition of the CpG recognition motif. The promoter sequence is given in the figure. The graph shows GFP fluorescence as a function of MBD plasmid concentration for the CpG methylated promoter (black) and the non-methylated one (blue). GFP expression is not modulated in a methylation dependent manner.

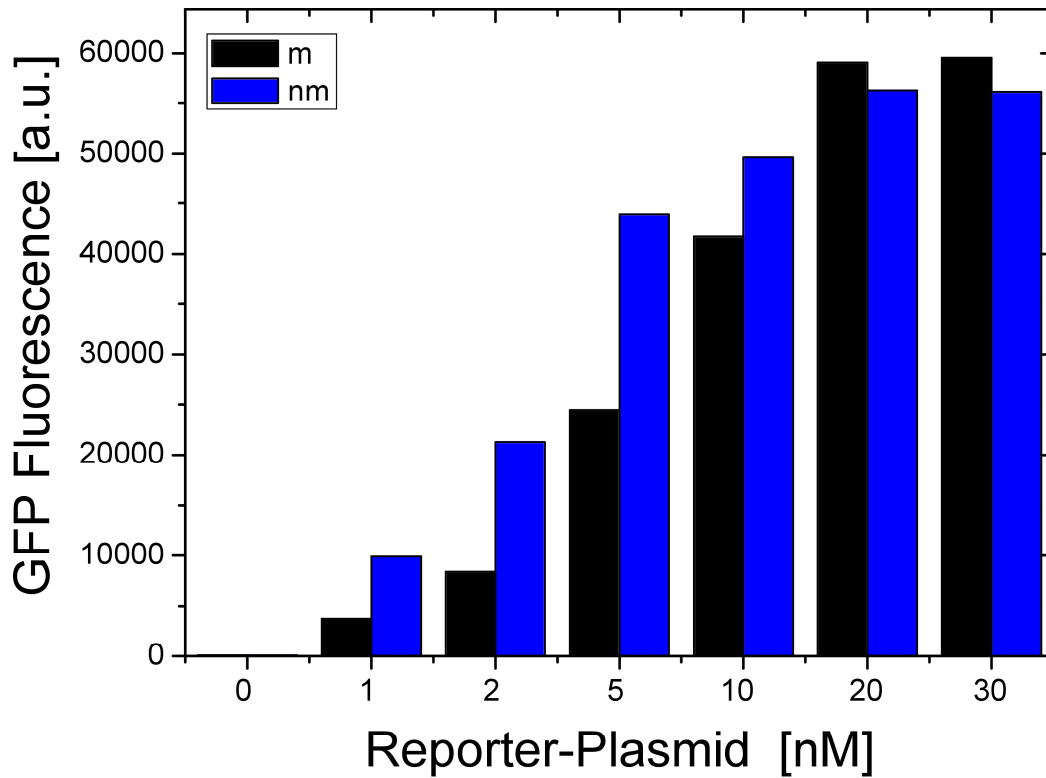


Figure S6. Concentration of recombinant GFP as a function of the concentration of the reporter plasmid pBEST-BDNF-wt-UTR1-GFP-T500 (m: CpG methylated (black); nm: unmethylated, blue). The concentration of amino acids, magnesium glutamate and potassium glutamate added to the reaction is 0.5 mM, 3 mM, and 50 mM, respectively. Expression levels of CpG+ and CpG- plasmids are comparable within a standard error of about 10 %. CpG methylated plasmids are not subject to degradation.

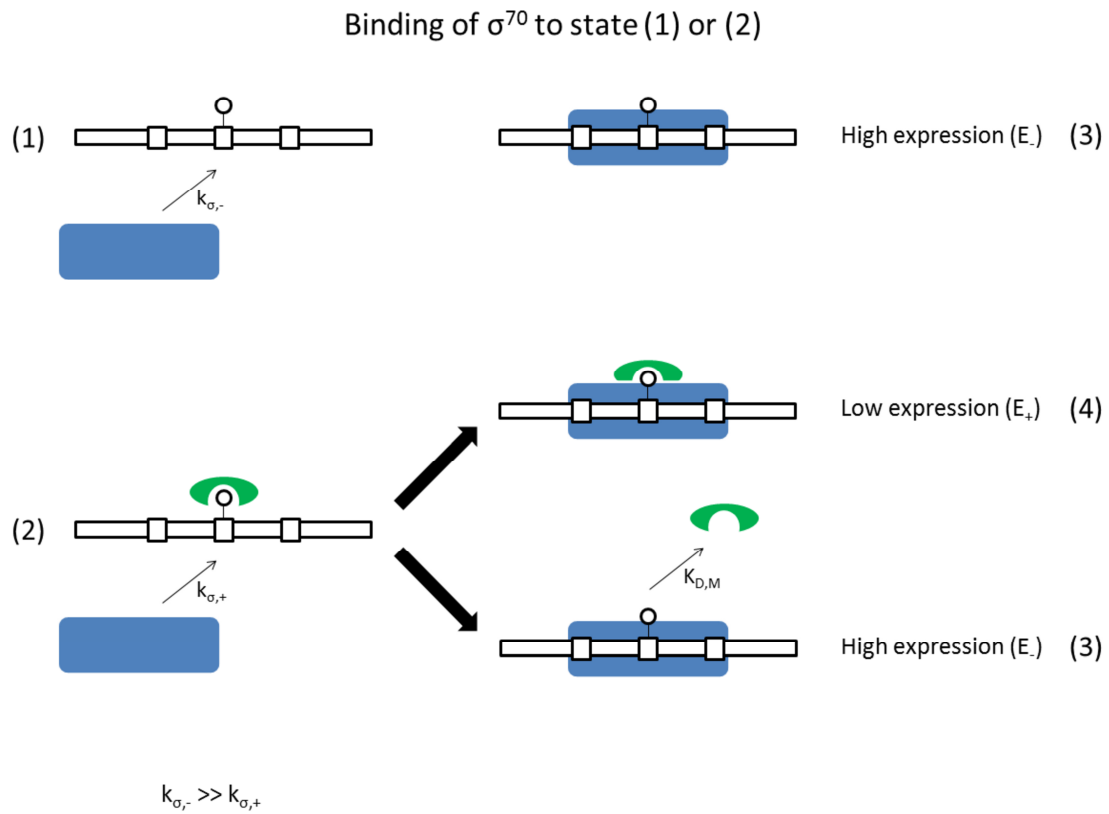


Figure S7. Transition from state [(1) or (2)] to [(3) or (4)], respectively. There are different association constants of the sigma factor 70: either for binding to the free promoter ($k_{\sigma,-}$), or for binding to the promoter with bound MBD ($k_{\sigma,+}$). The difference stems from structural hindrance mediated by MBD ($k_{\sigma,-} \gg k_{\sigma,+}$). If bound to the promoter, the sigma factor 70 can facilitate the dissociation of MBD, resulting in an increased dissociation constant ($K_{D,M}$). At the same time the sigma factor will initiate the expression. Because of this, the transition from (3) to (4) can be neglected.

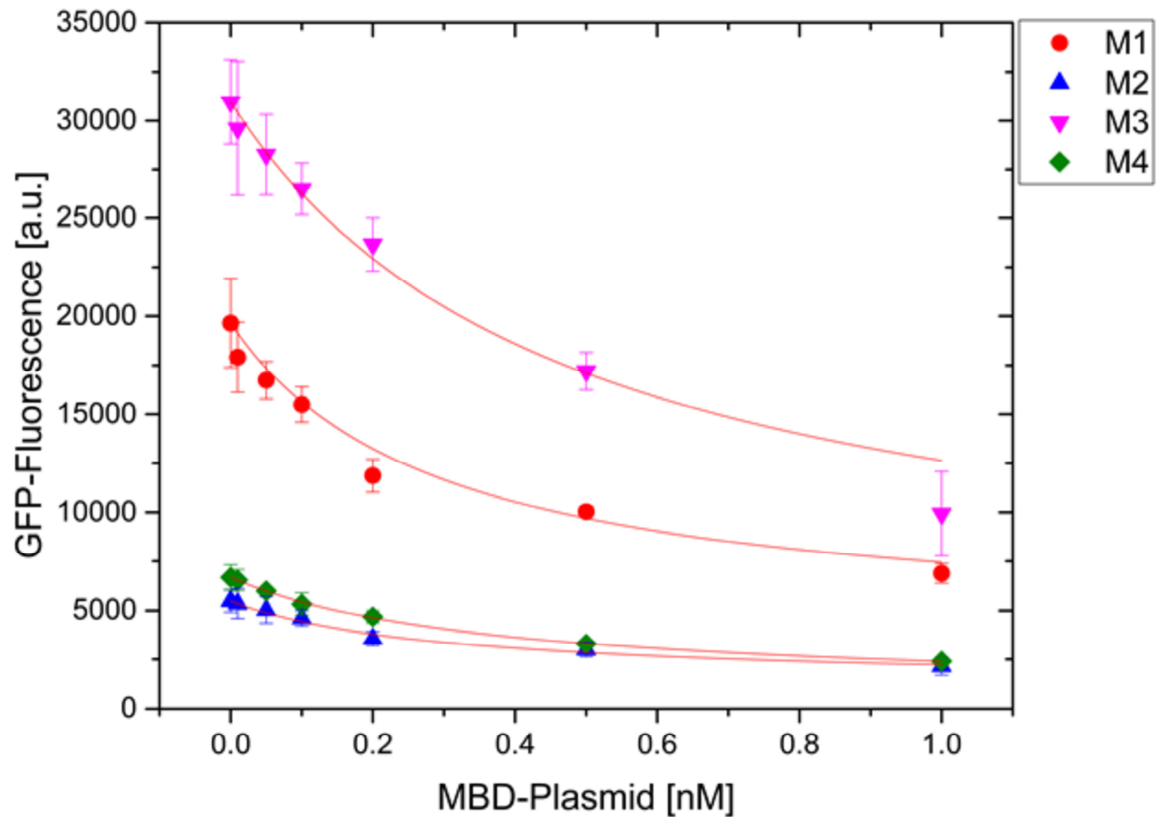


Figure S8. Amount of expressed GFP as a function of MBD-Plasmid concentration in conjunction with the methylated mutants of the BDNF promoter (M1-M4). The red lines represent a fit using equation (1). For M3 the fitting parameter for E_+ became negative, so we set it to a fixed value comparable to wt and the other mutants (see table S9). This results in a curve that fits the data about equally well.

Table S9. Fitting parameters of the plots in figure 6 and figure S8. The values where no error is listed were fixed. E. was fixed to the measured value $E([MBD-Plasmid] = 0 \text{ nM})$. E_+ of M3 had to be fixed to a value that gives a $K_{D,M}$ and a ratio of $E-/E_+$ comparable to wt, M1, M2 and M4 to avoid negative parameters.

	wt	M1	M2	M3	M4
E. [a.u.]	27033.33	19649	5483.33	30968	6680.67
E_+ [a.u.]	3656.23 ± 650.58	3879.71 ± 1490.62	1225.65 ± 395.66	4000	779.45 \pm 72.74
E/E_+	7.39	5.06	4.47	7.74	8.57
$K_{D,M}$ [nM]	0.555 \pm 0.038	0.292 \pm 0.079	0.306 \pm 0.069	0.472 \pm 0.029	0.376 \pm 0.012

References

1. Moradi, M., Babin, V., Roland, C. and Sagui, C. (2013) Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. *Nucleic Acids Research*, **41**, 33-43.

3.2 Worst *et al.* (2021)

A Methylation-Directed, Synthetic Pap Switch Based on Self-Complementary Regulatory DNA Reconstituted in an All *E. coli* Cell-Free Expression System.

Worst, E. G.†, **Finkler, M.†**, Schenkelberger, M., Kurt, Ö., Helms, V., Noireaux, V., and Ott, A.

† Geteilte Erstautorenschaft

ACS Synthetic Biology (2021), *10*, 2725–2739.

DOI: 10.1021/acssynbio.1c00326

<https://doi.org/10.1021/acssynbio.1c00326>

Reprinted with permission from Worst, E. G.†, **Finkler, M.†**, Schenkelberger, M., Kurt, Ö., Helms, V., Noireaux, V., and Ott, A. (2021) A Methylation-Directed, Synthetic Pap Switch Based on Self-Complementary Regulatory DNA Reconstituted in an All *E. coli* Cell-Free Expression System. *ACS Synth. Biol.* *10*, 2725–2739. <https://doi.org/10.1021/acssynbio.1c00326>. Copyright 2021 American Chemical Society.

A Methylation-Directed, Synthetic Pap Switch Based on Self-Complementary Regulatory DNA Reconstituted in an All *E. coli* Cell-Free Expression System

Emanuel G. Worst,[#] Marc Finkler,[#] Marc Schenkelberger, Ömer Kurt, Volkhard Helms, Vincent Noireaux, and Albrecht Ott*

Cite This: *ACS Synth. Biol.* 2021, 10, 2725–2739

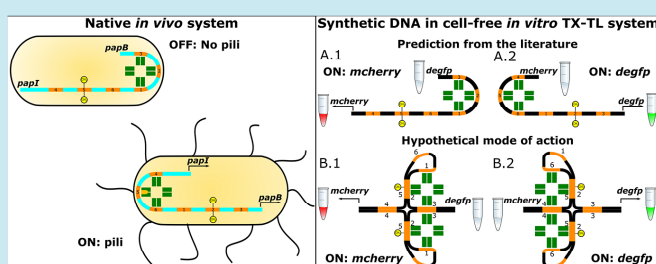
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



ABSTRACT: Pyelonephritis-associated pili (pap) enable migration of the uropathogenic *Escherichia coli* strain (UPEC) through the urinary tract. UPEC can switch between a stable ‘ON phase’ where the corresponding *pap* genes are expressed and a stable ‘OFF phase’ where their transcription is repressed. Hereditary DNA methylation of either one of two GATC motives within the regulatory region stabilizes the respective phase over many generations. The underlying molecular mechanism is only partly understood. Previous investigations suggest that *in vivo* phase-variation stability results from cooperative action of the transcriptional regulators Lrp and PapI. Here, we use an *E. coli* cell-free expression system to study molecular functions of the *pap* regulatory region based on a specially designed, synthetic construct flanked by two reporter genes encoding fluorescent proteins for simple readout. On the basis of our observations we suggest that besides Lrp, the conformation of the self-complementary regulatory DNA plays a strong role in the regulation of phase-variation. Our work not only contributes to better understand the phase variation mechanism, but it represents a successful start for mimicking stable, hereditary, and strong expression control based on methylation. The conformation of the regulatory DNA corresponds to a Holliday junction. Gene expression must be expected to respond if opposite arms of the junction are drawn outward.

KEYWORDS: *E. coli*, pyelonephritis associated pili, *pap* switch, Lrp, methylation patterns, cell-free expression

1. INTRODUCTION

Synthetic biology has applications that span from genomics,^{1,2} transcriptomics, and proteomics^{3,4} to metabolomics.^{5,6} Examples are expression systems engineered to expand metabolic functions⁷ by including noncanonical amino acids into model proteins,^{8–10} opening new directions in medicine. Synthetic oscillators,¹¹ boolean logic gates,^{12,13} and switches¹⁴ constitute examples of functional modules that produce logical operations at the transcriptional level. Synthetic biology is built on the control of gene transcription and protein synthesis. Programming these systems is based on DNA that will unfold its information through gene expression to eventually establish a module.

It is important to identify and characterize new regulatory parts that expand the scope of synthetic modules.^{15,16} Research

in this direction cannot proceed without proper understanding of the molecular interplay which make up the genetic modules.⁷ Conversely, this means that synthetic biology represents a valuable tool to investigate and understand related cooperative molecular actions in unprecedented detail. The use of a cell-free transcription-translation (TX-TL) system offers the opportunity to rapidly prototype and test related constructions.

Received: July 14, 2021

Published: September 22, 2021



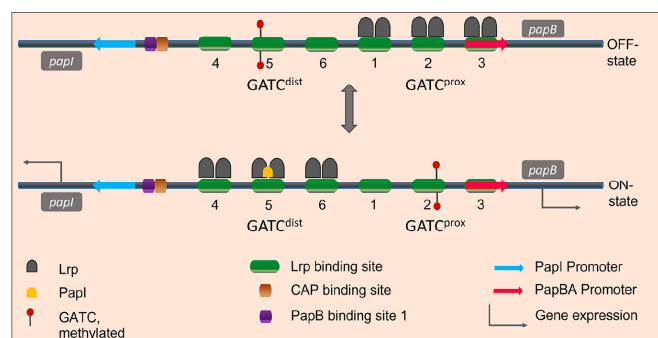


Figure 1. Scheme of the genomic organization of the *pap* regulatory region based on Marinus and Lobner-Olesen.²⁸ The genes *papI* and *papB* (designated by gray quads) at the flanking sites are oriented in opposite directions of transcription. The corresponding promoter regions are highlighted by arrows (blue, *papI* promoter; red, *papBA* promoter). Six Lrp binding sites (green) are located between the genes *papI* and *papB*. The Lrp binding sites 5 and 2 each contain a single DNA adenine methylase (Dam) recognition sequence (GATC). The PapB binding site 1 (purple) and the CAP (cAMP catabolite gene activator protein) binding site (brown) are located within the regulatory region between the *papI* promoter and the Lrp binding site 4. The remaining PapB binding sites 2 and 3 are not depicted. To further simplify, the scheme neither includes further *pap* regulatory proteins such as CAP or PapB, nor the downstream genes coding for proteins involved in the pili forming process, regulated by the “*papBA*” promoter. (Top) the putative OFF-state configuration where the expression of the *papI* and *papB* genes is repressed.³⁸ (Bottom) The ON-state where both genes *papI* and the *papB* are expressed. In the presence of CAP, PapB and PapI amplify their expression by positive, mutual feedback (not shown).^{38–40}

In the expression system, genes must be controlled not only at the level of transcription^{17,18} but of translation as well. Strategies to adjust the amounts of mRNA molecules consist of varying the transcription initiation rate or termination frequency, modulating RNA stability,¹⁹ or active removal of mRNA.^{20,21} At the transcriptional level, control is mostly achieved through transcription factors and inducible promoters. CRISPR/Cas is technically more advanced but offers a much stronger level of expression control.^{22,23} In this work we address methylation as a means to regulate gene expression. DNA methylation patterns are well-known to epigenetically convey hereditary information about genetic regulation in eukaryotes.²⁴ However, the details of epigenetic regulation remain poorly understood.

Recently we investigated the role of the methyl-CpG binding domain (MBD) from the eukaryotic MeCP2 protein, a transcriptional regulator that, in conjunction with the BDNF (brain-derived neurotrophic factor) promoter sequence, acts as a methylation dependent transcriptional modulator in humans.²⁵ The observations made in an *Escherichia coli* (*E. coli*) cell-free expression system accurately reproduced the regulation in spite of its eukaryotic nature.²⁵ We found evidence of MBD acting by deforming the DNA grooves on distances beyond its small footprint of only two basepairs.²⁵ This unexpected way of function (through DNA) may contribute to the well-known fragility of the transcriptional regulator toward point mutations manifesting as the Rett syndrome.²⁶

Methylation based regulatory mechanisms play a role in bacteria as well.²⁷ In the *E. coli* genome, 1.5% of adenine bases are methylated and about 0.75% of cytosine bases are methylation targets.²⁸ In bacteria, DNA methylation influences the time of DNA replication, DNA repair, the distribution of chromosomes to daughter cells, and the time of transposition of plasmids.^{29–32} For these processes, the target DNA

sequence must be hemimethylated so that the related proteins bind to it.³³

There have been successful realizations of synthetic epigenetic modules, conveying for instance inheritable information.³⁴ Epigenetics has potentially important applications in pharmacology because of the orthogonality of bacterial and eukaryotic epigenetic programming.^{35,36}

A prokaryotic example of a methylation dependent genetic switch is the transcription-based phase variation mechanism of uropathogenic *E. coli* (UPEC) that inspired our work. UPEC can switch between the “ON phase”, in which the genes for pathogenic virulence factors (pyelonephritis-associated pili, Pap) are expressed, and an “OFF phase”, in which the expression of these genes is repressed³⁷ (Figure 1). The system is highly cooperative so that the observed, strong genetic switching, stable over generations, is based on the methylation patterns of only two methylation sites. Moreover, the regulatory unit interacts with only a few coregulators. We present the details of the *pap* regulating unit in more detail in what follows.

1.1. Structure of the Pap Regulatory Unit. The genes of the coregulators PapB and PapI are divergently transcribed in opposite directions.⁴¹ The *pap* regulatory region is in between the *papI* gene (upstream) and the *papB* gene (downstream).⁴² Within the *pap* regulatory region there are three binding sites (BS) for PapB.³⁹ As illustrated in Figure 1, PapB BS 1 is located between the *papI* promoter and Lrp BS 4.³⁹ PapB BS 2 overlaps with the RNA polymerase recognition site of the *papBA* promoter and PapB BS 3 is located within the *papB* gene³⁹ (PapB BS 2 and 3 are not depicted not to overload Figure 1. PapB is not part of this study). The *pap* regulatory region includes six leucine-responsive regulatory protein (Lrp) binding sites.⁴² Lrp has a mass of 18.8 kDa.^{43,44} It is a global *E. coli* transcription regulator.^{43,44} The six Lrp BS form two functional units.⁴⁵ Left we have Lrp BS 4, 5, 6 with the recognition sequence for the DNA adenine methylase (Dam)



Figure 2. Scheme of the synthetic pap regulatory DNA construct as designed for our study. The DNA sequence includes all six Lrp BS (green and numbered), Lrp BS according to Hernday *et al.*³⁸). The synthetic sequence corresponds to the native pap regulatory DNA (GenBank: X14471.1, based on Blyn *et al.*³⁷) except for inserted -35 and -10 elements of OR2-ORI-Pr (orange arrows) that enclose the Lrp BS 3 and Lrp BS 4, followed by UTR1 (light blue) and flanked by mcherry and degfp genes. See also Figure S3 for more information. Given the binding pattern as shown in Figure 1, supposing Lrp acting by steric hindrance, one expects degfp expression (green) during phase ON and mcherry expression (red) during phase OFF.

GATC^{dist} as part of the Lrp BS 5, distal to the *papBA* promoter.⁴⁵ On the right we have Lrp BS 1, 2, 3 with the Dam recognition sequence GATC^{prox} as part of the Lrp BS 2, proximal to the *papBA* promoter of which the -35 element and -10 element enclose Lrp BS 3.⁴⁵

1.2. Phases of the Pap Regulatory Unit. Lrp binds to several Lrp BS cooperatively.^{45,46} It was proposed that four Lrp dimers combine to an octameric structure that binds to three Lrp BS.^{47–49} PapI increases the Lrp binding affinity to Lrp BS 2 and Lrp BS 5.³⁸ During the transcriptional regulation of the phase variation,³⁷ methylation of Lrp BS 2 and Lrp BS 5 plays an essential role.⁵⁰ In phase OFF, only GATC^{dist} is methylated, whereas in phase ON only the GATC^{prox} position is methylated.⁵⁰ It was suggested⁴² that during phase ON Lrp cooperatively occupies Lrp BS 4–6 in conjunction with PapI. This way Dam-mediated methylation can only occur at GATC^{prox}.⁴² This inhibits cooperative Lrp binding to Lrp BS 1–3, promoting the expression of *pap* genes.⁴² Binding of PapB to the PapB BS 1 activates *papI* gene expression.³⁹ During phase OFF Lrp cooperatively binds to Lrp BS 1–3,^{38,42,51} sterically inhibiting Dam from methylating GATC^{prox} and, at the same time, preventing RNA polymerase from binding to the corresponding promoter.⁴² This results in a reduced expression rate of genes located downstream of the *pap* BA promoter.⁵² *Mutual exclusive binding* predicts that cooperatively bound Lrp on Lrp BS 1–3 reduces the binding affinity of Lrp binding to Lrp BS 4–6 by roughly a factor of 10, thereby stabilizing the phase OFF in addition to the Lrp-blocking effect of methylation at GATC^{dist}.⁴²

Hernday *et al.*⁴² suggested that the topology of the regulating DNA may well play a role in the phase variation mechanism, a hypothesis supported through computational predictions by Peterson.⁵¹

The binding of the catabolite activator protein (CAP) to the CAP BS (Figure 1) next to PapB BS 1 was found to be a requirement for *pap* gene expression.^{40,53} Binding the Lrp–PapI complex to BS 4–6 may physically bend the pap regulatory DNA, enabling CAP to interact with the RNA polymerase; however, CAP dependent *pap* gene transcription does not result from the interaction between CAP and the *papI* promoter or the coding region.⁴⁰

The Lrp–PapI complex binds more strongly to the hemimethylated than to the fully methylated GATC^{dist}.³⁸

1.3. Phase Switching. A common model for switching from phase OFF to phase ON involves DNA replication leading to hemimethylated GATC^{dist} and dissociating Lrp from BS 1–3.^{38,51,54} Within the pap regulatory region GATC methylation by Dam proceeds nonprocessively, so that methylation of GATC^{prox} does not necessarily lead to subsequent methylation of GATC^{dist} or vice versa.^{54,55} This means that after DNA replication, Lrp and Dam compete for binding to Lrp BS 2 or Lrp BS 5.³⁸ In case Dam binds first, this leads to methylation of GATC^{prox} (or GATC^{dist}) blocking cooperative Lrp binding to Lrp BS 1–3 (or 4–6). In case Lrp

binds first, cooperative Lrp binding to Lrp BS 1–3 (or 4–6) prevents Dam from methylating GATC^{prox} (or GATC^{dist}).³⁸ Suitable papI concentrations after DNA replication favor cooperative binding of Lrp to Lrp BS 4–6, preventing full methylation of hemimethylated GATC^{dist} by Dam.³⁸ This leads to full methylation of GATC^{prox} hindering cooperative Lrp binding to Lrp BS 1–3, enabling RNA polymerase binding and expression of the *pap* genes.³⁸ Upon replication of the hemimethylated GATC^{dist}, a hemimethylated and fully unmethylated GATC^{dist} occur (each with hemimethylated GATC^{prox}). This again favors that Lrp cooperatively binds to Lrp BS 4–6. Full methylation of GATC^{prox} hinders cooperative Lrp binding to Lrp BS 1–3. Both processes stabilize the phase ON and corresponding *pap* gene expression.³⁸

1.4. This Work. In this work we present a designed, synthetic derivative of the transcription-based, epigenetic phase variation mechanism of UPEC. To study its behavior we use an all *E. coli* cell-free expression system obtained from *E. coli* cell lysate.^{56–59} All the necessary elements for gene expression are from *E. coli* only so that transcription and translation are performed by the native *E. coli* molecular machinery. This makes the system an ideal tool for the study of the *E. coli* epigenetic phase variation mechanism. We show that in conjunction with the coregulators Lrp and PapI strong genetic switching can be realized as a response to different methylation patterns. In addition, we present strong evidence that a cruciform DNA conformation of the self-complementary regulatory DNA is involved. We suggest that Lrp binding to the promoter cooperatively diminishes the likelihood of the linear conformation.

The here presented, synthetic pap regulatory DNA can be used for detection of the presence of Lrp or methylase activity. Our experiments suggest that regulation involves DNA conformational changes. Its cruciform conformation represents a Holliday junction. We project that the expression of our construct can be modulated by force dipoles in the pN range applied to opposite arms of the junction⁶⁰ (Figure S9).

2. RESULTS AND DISCUSSION

2.1. A Synthetic Pap Regulatory DNA. To study the epigenetic phase variation mechanism of UPEC in our expression system we designed a synthetic pap regulatory DNA. This DNA preserves the native DNA sequence (GenBank: X14471.1, based on Blyn *et al.*,³⁷ Figure S3) including all six Lrp BS (Lrp BS according to Hernday *et al.*³⁸) except for the -35 and -10 elements of OR2-ORI-Pr⁵⁶ that enclose Lrp BS 3, as well as Lrp BS 4 (Figure 2). These elements initiate the transcription of the corresponding genes as indicated in Figure 2. At each side UTR1 links the promoter to either gene.⁵⁶ UTR1 instead of the native DNA sequences up- as well as downstream of the pap regulatory region, followed by the coding sequences of the fluorescent proteins mCherry and deGFP enables us to monitor gene expression. While the right -35 and -10 elements function as substitute

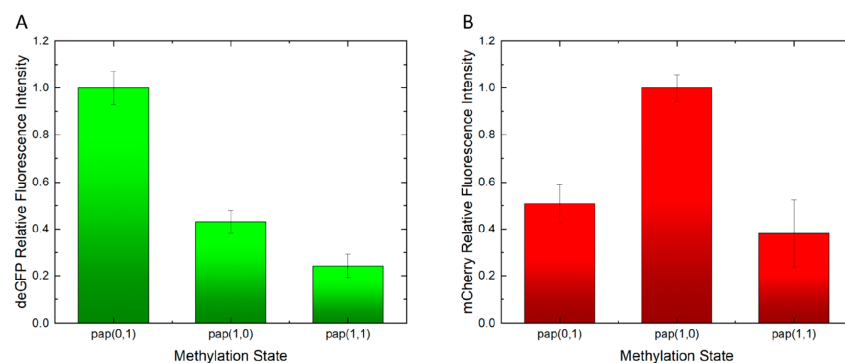


Figure 3. Fluorescence intensities of deGFP (A), mCherry (B) as a function of our synthetic pap regulatory DNA constructs (5 nM) with methylation states (0,1), (1,0) and (1,1). Background fluorescence from the expression system was subtracted. Symbols on the horizontal axis refer to the different methylation states of the DNA construct that were present in the cell-free reactions. (0,1): Lrp BS 5 unmethylated, Lrp BS 2 methylated. (1,0): Lrp BS 5 methylated, Lrp BS 2 unmethylated. (1,1): Lrp BS 2 and 5 are both methylated. Fluorescence intensities were normalized relative to the highest observed intensity of the respective reporter protein. Error bars represent one standard deviation from the mean from five measurements. The relatively high deGFP (mCherry) fluorescence intensity for the methylation states (1,0) ((0,1)) and (1,1) results from leaky gene expression that is common in gene expression systems.^{58,61}

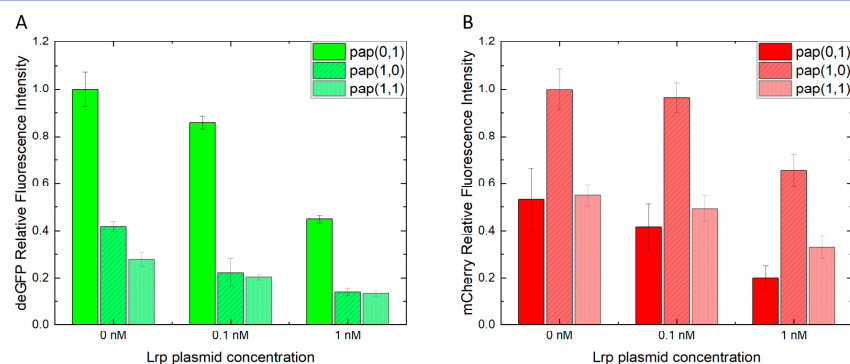


Figure 4. Fluorescence intensities of deGFP (A), mCherry (B) in the presence of our synthetic pap regulatory DNA constructs with methylation states (0,1), (1,0), and (1,1) as a function of varying lrp-coding plasmid concentrations in the absence of papI-coding plasmid. Background fluorescence from the expression system is subtracted. The plotted data correspond to the data presented in Figure S5, where no papI-coding plasmid is added. Symbols on the horizontal axis refer to the different methylation states of the DNA construct that were present in the cell-free reactions. (0,1): Lrp BS 5 unmethylated, Lrp BS 2 methylated. (1,0): Lrp BS 5 methylated, Lrp BS 2 unmethylated. (1,1): Lrp BS 2 and 5 both methylated. Fluorescence intensities were normalized with respect to the highest occurring value. Error bars represent one standard deviation from the mean as gained from two measurements. Gene expression is leaky: even for 1 mM of lrp-coding plasmid the expression is reduced by only about 85% compared to the highest occurring value.

for the native promoter elements, the left -35 and -10 elements are added to the native sequence to enable *mcherry* gene expression.

2.2. Methylation-Dependent Regulation of the Synthetic Pap Regulatory Region. We determined the fluorescence emission intensities of deGFP and mCherry as a function of the methylation state of the synthetic regulatory DNA. The methylation states under examination were (0,1): unmethylated $\text{GATC}^{\text{dist}}$ site and methylated $\text{GATC}^{\text{prox}}$ site; (1,0): methylated $\text{GATC}^{\text{dist}}$ and unmethylated $\text{GATC}^{\text{prox}}$ site; and (1,1): methylated $\text{GATC}^{\text{dist}}$ site and methylated $\text{GATC}^{\text{prox}}$ site.

The methylation pattern (0,1) led to the highest deGFP fluorescence intensity, while (1,1) produced the lowest deGFP fluorescence intensity. Conversely, the highest mCherry fluorescence intensity appeared for the methylation state (1,0). The mean of the mCherry fluorescence intensity in the case of the methylation state (0,1) was slightly higher than for the methylation state (1,1) albeit the errors of the two signals overlap. We understand that a methylated GATC site produced increased fluorescence intensity from its adjacent gene only if the other GATC site of the regulatory unit was unmethylated.

In addition to the fluorescence end point measurements presented in Figure 3, kinetic fluorescence measurements were

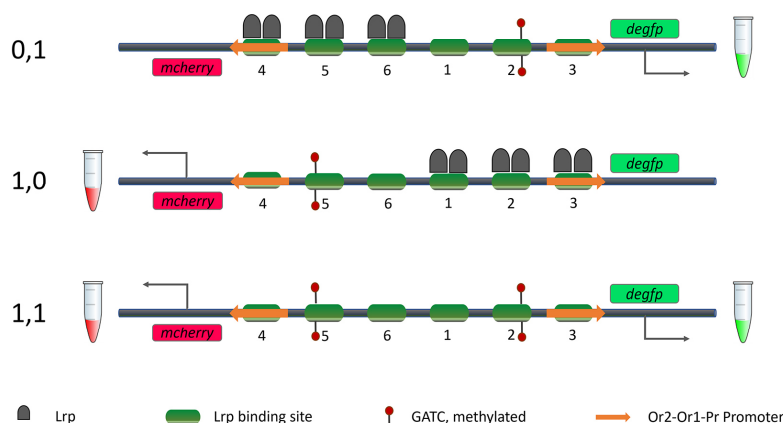


Figure 5. Scheme of the three different methylation states of our synthetic *pap* regulatory DNA construct. While methylation of Lrp BS 2 (0,1) results in *degfp* expression, methylation of Lrp BS 5 (1,0) results in *mcherry* expression. One would expect that methylation of binding sites 2 and 5 (1,1) prevents Lrp binding to all six Lrp BS, enabling coexpression of *degfp* and *mcherry* as shown; however, this is not what we observe.

performed (Figure S4). The results of the end point and the kinetic measurements agreed well (compare Figure S4 and Figure 3). However, the statistical error in the kinetic measurement was increased compared to the end point measurements. We could not find a reason for this observation.

2.3. Methylation-Dependent Regulation of the Synthetic *pap* Regulatory Unit in Conjunction with Cell-Free Synthesis of Lrp and PapI. We determined the fluorescence intensities of deGFP and mCherry emission as a function of the methylation pattern of the synthetic *pap* regulatory DNA in the presence of cell-free synthesis of Lrp and PapI.

At any *lrp*-coding plasmid concentration, the methylation state (0,1) led to the highest deGFP fluorescence intensities compared to the other two methylation states (1,0) and (1,1) (Figure 4A). At any *lrp*-coding plasmid concentration, we observed the highest mCherry fluorescence intensities for (1,0) compared to the methylation states (0,1) and (1,1) (Figure 4B). The methylation pattern (1,1) resulted in rather weak fluorescence intensity signals of both reporter proteins (compared to the methylation states (1,0) or (0,1)) at different *lrp*-coding plasmid concentrations (Figure 4A,B).

The presence of the *papI*-coding plasmid weakly influenced the reporter fluorescence intensities. Only in the case of the methylation pattern (0,1) (Figure S5A), in the absence of the *lrp*-coding plasmid did we observe that increased concentrations of the *papI*-coding plasmid led to increased deGFP fluorescence intensity while mCherry fluorescence intensity was decreased. In the case of methylation state (1,1), again in the absence of the *lrp*-coding plasmid, the analogue behavior of mCherry fluorescence intensity was observed; however, it was less pronounced (Figure S5C).

2.4. Interpretation and Comparison to the Native Regulatory Behavior. Here we interpret increased fluorescence intensities due to increased concentrations of the corresponding fluorescent reporter proteins as a higher cell-free expression rate of the corresponding genes. On the basis of previous data⁵⁶ we expect the concentrations of the coregulators Lrp and PapI, the coding plasmids of which

were simultaneously expressed, to scale as the concentrations of their respective plasmids that were supplemented to the cell-free reaction.

During preparation of the cell-free system, proteins with a molecular weight of less than 10 kDa are removed during the dialysis step (MWCO 10 kDa). Since the molecular weight of Lrp is about 19 kDa, this global *E. coli* transcription regulator^{43,44} is expected to be present in our cell free system. This was confirmed by analytical mass spectrometry (see Supporting Information of ref 62).

We saw that the reporter gene expression varied with the amount of synthesized PapI (Figure S5A). This confirms the presence of Lrp, since PapI is not able to specifically bind to the synthetic *pap* regulatory DNA without Lrp.⁶³ Moreover, *E. coli* strains with knocked down *lrp* genes were unable to express *pap* genes.⁵⁰ At the same time, native PapI cannot be part of the cell free system since the *papI* gene from UPEC is not present in the bacterial genome of the *E. coli* BL21 Rosetta 2 cells used for production of the cell extract. Even if this were not so, given its molecular weight of approximately 8.7 kDa, PapI would have been removed from the cell extract during dialysis.

We observed that the methylation pattern (0,1) resulted in the highest *degfp* expression rate compared to the other methylation states (Figures 3 and 4, and Figures S4 and S5). This finding suggests that GATC^{prox} must be methylated and GATC^{dist} must be unmethylated for *pap* gene expression, in excellent agreement with previous observations from the native system.⁵⁰ Increased PapI concentrations led to decreased *mcherry* expression and increased *degfp* expression (Figure S5A), indicating that PapI sustains the ON state. This again is in excellent agreement with the observed increase of the Lrp binding affinity to Lrp BS 5 due to the presence of PapI.³⁸

Increased concentrations of *lrp*-coding plasmid resulted in decreased reporter gene expression. This decrease was especially pronounced for the synthetic *pap* regulatory DNAs with a single methylated site (Figure 4). We understand that sites that weaken expression upon Lrp binding were far from saturation at the Lrp concentration given by system



Figure 6. Possible Lrp induced conformational changes of the linear synthetic pap regulatory DNA, as inferred from the native pap regulatory DNA in conjunction with the predictions by Peterson.⁵¹ (A.1) Lrp cooperatively binds to Lrp BS 1, 2, and 3 to form an octamer (green rectangles), preventing RNA polymerase from binding to the promoter initiating *degfp* expression. At the same time, $\text{GATC}^{\text{dist}}$ as part of Lrp BS 5 is fully methylated (yellow dots). The methylation prevents cooperative Lrp binding to Lrp BS 4–6 and enables RNA polymerase binding to the promoter initiating *mcherry* expression (representing phase OFF of the native construct). (A.2) The Lrp octamer cooperatively binds to Lrp BS 4, 5, and 6, preventing RNA polymerase from binding to the promoter, initiating *mcherry* expression. At the same time, $\text{GATC}^{\text{prox}}$ as part of Lrp BS 2 is fully methylated. The methylation prevents cooperative Lrp binding to Lrp BS 1–3. This enables the RNA polymerase to bind to the promoter initiating *degfp* expression (representing phase ON of the native construct). However, the idea above implies that the methylation pattern (1,1), which diminishes Lrp binding, must be expected to raise the expression levels of both, *degfp* and *mcherry*, (Figure 5) but this is not what we observe.

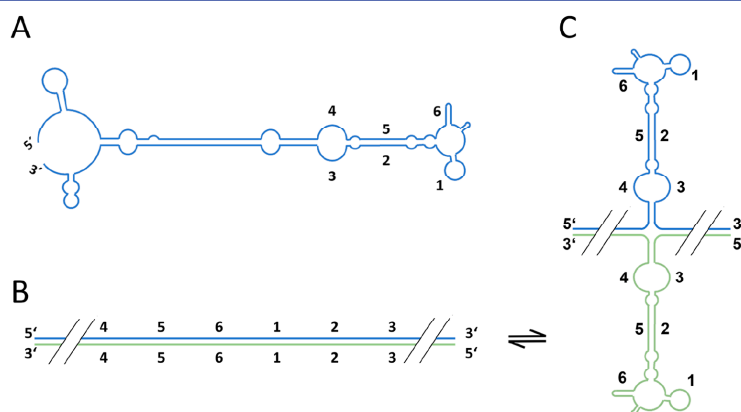


Figure 7. (A) Scheme of the typical secondary structures of the synthetic pap regulatory region DNA (single stranded) as predicted by the mfold web server.⁶⁵ In all predicted structures Lrp BS 1 and 6 as well as 3 and 4 are paired at least partially, moreover there is almost full pairing between $\text{GATC}^{\text{prox}}$ (Lrp BS 2) and $\text{GATC}^{\text{dist}}$ (Lrp BS 5) (for more details on the predictions see Figure S6). According to this prediction the double stranded, linear, synthetic pap regulatory DNA (B) must be suspected to form stable cruciform conformations (C) as well. The conformation shown represents just one possible structure among others with varying arm lengths. In our experiments these structures are likely to coexist with the linear one.

preparation. Accordingly, in the presence of the *lrp*-coding plasmid, the additionally synthesized Lrp exerted the role of a suppressor by occupying these sites.

It is clear that our observations cannot be explained by simple, methylation-dependent Lrp blocking that comes with an occupied GATC site close to the gene of interest. Rather, the finding that methylation of $\text{GATC}^{\text{prox}}$ or $\text{GATC}^{\text{dist}}$ influenced the reporter gene transcription rate (Figures 3 and 4, Figures S4 and S5), although the -35 and -10 elements of both promoters enclose the outer Lrp BS 3 or 4 respectively, sustains the hypothesis that Lrp binds cooperatively to several Lrp BS,^{42,45–49} depending on methylation. We find that in the absence of PapI, contrary to the native system, Lrp alone is sufficient to produce a methylation-dependent interaction between the subunits composed of Lrp BS 1–3 and Lrp BS 4–6 of the pap regulatory region (Figure 5).

Steric blocking of the RNA polymerase⁵² through cooperative binding of Lrp^{38,42,51} to the unmethylated region (Figure 6) is expected to be the mechanism that reduced transcription of the reporter genes in our experiments regarding methylation patterns (1,0) and (0,1). In this

mechanism⁴⁴ Lrp works a repressor as proposed for the native system.⁵¹

Neither from our results (Figure 5) nor from the knowledge about the native system in the literature (Figure 6) is it clear how the (1,1) methylation pattern translates to molecular binding conformations that diminish the expression of both genes on the linear, synthetic DNA regulatory unit. DNA conformation can have an impact on gene expression.⁶⁴ We checked for DNA secondary structures.

2.5. Predictions of DNA Conformations of the Synthetic Pap Regulatory Region. We studied the secondary structures of the synthetic pap regulatory DNA using the mfold web server.⁶⁵ There were only minor differences among all of the three predicted secondary structures (their general structure is shown in Figure 7). In these structures Lrp BS 2 and 5 as well as Lrp BS 3 and 4 and Lrp BS 1 and 6 were paired (at least partially). All structures exhibited almost full pairing of $\text{GATC}^{\text{prox}}$ and $\text{GATC}^{\text{dist}}$.

To check for the presence of these secondary structures, we labeled the regulatory DNA with the quenching pair 6-Fam/BHQ1. We designed two different types of constructs. They

either exhibited quenching for the linear, or for the predicted secondary structures as soon as the single strands located in close vicinity to Lrp BS 5 and BS 2 pair (for further details see section 3 (Methods), incl. Figure 11).

Figure 8 shows the time course of fluorescence intensity from the synthetic regulatory DNA including the quenching

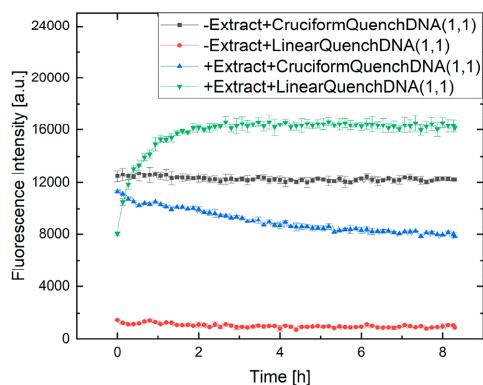


Figure 8. Time course of the fluorescence emission intensity of 6-Fam/BHQ1 labeled DNA constructs ($c = 120$ nM). The methylation state was (1,1) in all cases. The positions of 6-Fam and the corresponding quencher BHQ1 were chosen to show quenching either in the case of the linear conformation (LinearQuenchDNA) or for the predicted DNA cruciform conformation (CruciformQuenchDNA). Experiments were performed in the presence of crude extract and in its absence. We see that in the absence of crude extract the DNA remains linear while in its presence the linear conformation is lost to approach the cruciform conformation. For comparison, the fluorescence intensity of 6-Fam as part of the LinearQuenchDNA in the absence of BHQ1 ($dsDNA_{Ann,Linear(1,1)}$) is $26,859.5 \pm 3513.6$ au ($c(dsDNA_{Ann,Linear(1,1)}) = 120$ nM). The fluorescence intensity of 6-Fam as part of the CruciformQuenchDNA in the absence of BHQ1 ($dsDNA_{Ann,Cruciform(1,1)}$) is $12,357 \pm 229.1$ au ($c(dsDNA_{Ann,Cruciform(1,1)}) = 120$ nM). The difference in intensities is due to different efficiencies in labeling during production.

pair. To test if the DNA conformation depends on the presence of a bacterial environment, experiments were also performed with an expression system lacking crude extract, a main component of the cell-free expression reaction.

The graphs in Figure 8 confirm the formation of the predicted cruciform conformation in the presence of crude extract. In the absence of crude extract the LinearQuenchDNA shows strong quenching of 6-Fam emission (red circles) and the CruciformQuenchDNA shows no quenching (black squares). If crude extract is present, however, the fluorescence intensity of the LinearQuenchDNA increases quickly. This means that the quenching pair separates, in other words the DNA starts to open (green triangles). In addition, the CruciformQuenchDNA exhibits decreasing fluorescence emission. This means that the labeled parts (adjacent to Lrp BS 5 and BS 2) meet, in other words, hairpin formation in excellent agreement with the predicted DNA cruciform conformation (blue triangles). Hairpin formation is slower than the opening of the linear conformation. In the linear case we depart from a well-defined structure: this is fast. The hairpin formation is likely to be one structure among others that may not differ that

much in binding energy. Accordingly it is slow to appear. The observed behavior was very similar for the methylation states (1,0) and (0,1) (Figure S8). Addition of the *lrp*-coding plasmid for Lrp production did not lead to detectable differences (data not shown).

The fluorescence intensity of LinearQuenchDNA in the system does not increase to the level without quencher. We suspect that a fraction of the DNA regulatory region remains in the linear conformation or close to it. CruciformQuenchDNA does not seem to reach full quenching even if we extrapolate the observation to longer time scales. However, this is to be expected. The regulatory region is long so that the simultaneous presence of many secondary structures is likely, including hairpins of different lengths (Figure 9B gives two examples). These conformations do or do not contribute to quenching.

The results above show that the regulatory DNA in the system does not retain the linear conformation, rather the experiments suggest that it forms two hairpins connected through a Holliday junction (Figure 9A and B). Depending on the length of the hairpins, this may or may not lead to expression as shown in Figure 9B. This conformation must be expected to produce the weak expression that we observe in the methylation state (1,1) where both GATC sites are fully methylated so that Lrp does not bind. If the synthetic pap regulatory DNA in the methylation state (0,1) (or (1,0)) folds according to Figure 9A (left or middle), the newly formed GATC and GATC sites will be hemimethylated. Hemimethylated GATC sites have reduced large furrows and are structurally metastable,⁶⁶ resulting in specific interactions with proteins. *In vitro* binding experiments by Hernday *et al.*³⁸ showed that hemimethylation of the GATC^{dist} site has influence on the binding behavior of the coregulators Lrp and PapI. Moreover, the type of hemimethylation also affected their binding behavior.³⁸ These observations apply to the different hemimethylated GATC sites formed from paired nucleotides of Lrp BS 2 and 5 as part of the cruciform conformation (Figure 9A left and middle). One may speculate that in the methylation states (0,1) and (1,0) Lrp promotes and prevents expression of either gene by stabilizing suitable configurations as exemplified in Figure 9C.

Increasing Lrp concentrations did not produce visible changes in the time course of fluorescence emission during the quenching experiments. However, it seems unlikely that Lrp does not play a role. It may well be that the native concentration is already in saturation, regarding DNA conformational changes. Indeed increased Lrp concentrations were found to inhibit expression, leading to the idea that Lrp sterically hinders expression by binding to the BS adjacent to the expressed gene. An equivalent of the above scheme may well play a role in the native system.

3. METHODS

3.1. Standard Molecular–Biological Procedures. Restriction enzymes, Taq DNA Polymerase, Q5 High Fidelity DNA Polymerase and Gibson Assembly Cloning Kits were from New England Biolabs GmbH (Frankfurt am Main, Germany).

Unless otherwise noted, standard PCRs were performed using either Taq DNA Polymerase or Q5 High Fidelity DNA Polymerase in the corresponding reaction buffer. Plasmid DNA or linear DNA was used as a template. All PCR primers were synthesized and desalinated by Metabion international AG

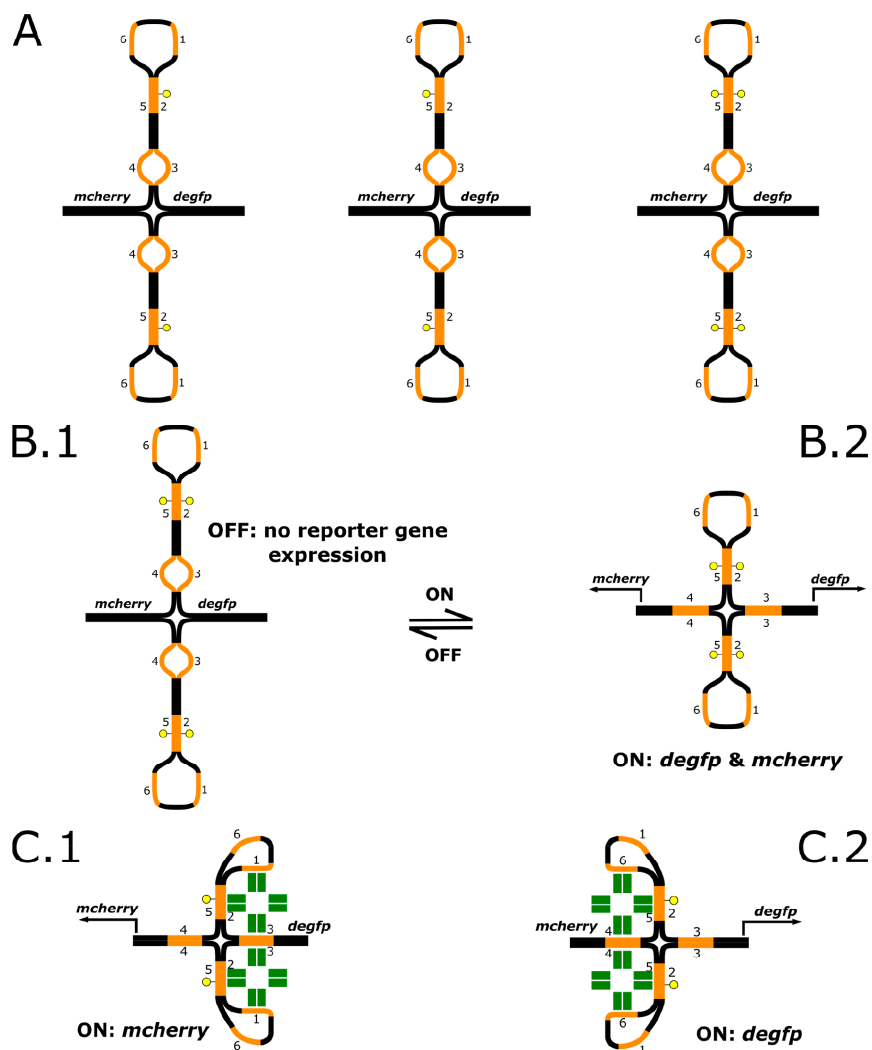


Figure 9. (A) As the methylation states (0,1), (1,0), and (1,1) from the linear regulatory DNA conformation transform to the cruciform conformation, the newly emerging GATC sites have different methylation states. Lrp BS are highlighted in orange. Methylation of a DNA strand is represented by a yellow dot. (Left) Hemimethylation resulting from methylation pattern (0,1) (fully methylated GATC^{prox} in the linear conformation). (Middle) Hemimethylation resulting from methylation pattern (1,0) (fully methylated GATC^{dist} in the linear conformation). (Right) Full methylation resulting from methylation pattern (1,1) (fully methylated GATC^{prox} and GATC^{dist}). (B) In thermodynamic equilibrium, in the absence of binding elements, secondary structures, in which hairpins form to different extent, are about equally likely to form. With such conformations, reporter genes may (B.2) or may not (B.1) be expressed. (C) Hypothetical conformational changes of the pap DNA cruciform conformation with cooperative binding of Lrp as octamers (green rectangles), enabling expression of one gene while suppressing the other. Methylation is represented by a yellow dot.

(Planegg, Germany). PCRs were performed with the Multi-Gene 96-Well Gradient Thermal Cycler (Labnet International Inc., Edison, USA). The hybridization temperature of the primers was predicted using the Tm Calculator (New England Biolabs Inc., Ipswich, USA) based on work by SantaLucia,⁶⁷

including corrections for salt concentrations according to Owczarzy *et al.*⁶⁸ After completion of the PCR, a small amount of the PCR reaction (typically 10 μ L) was used to check the purity of the PCR products by horizontal agarose gel electrophoresis. If only the DNA fragments of expected size

Table 1. Sequences of the Primers Used in This Work. Mutant Nucleotide Bases within the Dam Recognition Sequence are Emphasized by Brackets

primer no.	DNA sequence 5' to 3' direction	forward (F)/reverse (R) primer
1	GAGCTGACTGGGTTGAAGG	F
2	GCTAGCAATAATTTTGTAACTTTAAGAAGGAGATATAC	R
3	GTAAACAAAATTATTGCTAGCTGCAACCATTATGCC	F
4	CACCATCAGCCAGAAAACC	R
5	CGCTATGAATTCCTCGAGTTACTTGTACAGCTC	F
6	CCATGGAAGCTTTTAGATCCCGGGCGGTCA	R
7	CATCAGCCAGAAAACC-GAGCTG-CTGGG-TTGAAG	F
8	AGATTT-AAACGA-[G]CTTT-TAACCCA-CAAAAAC	R
9	TAAAA-G[C]TCG-TTAAAT-CTTGACA-TACAACA-TAAA-AAAC	F
10	CTGACTG-GGTTGA-AGGCAC-CATCA-GCCAG-AAAAC	R
11	CAGCA-TAAAA-GA[G]CGT-CTAAA-TGTTG-ACATAC	R
12	AGACG[C]T-CTTTA-TGCTGT-AAATT-CAAT-TTGC	F

were present in the gel, the remaining sample was purified using the PureLink PCR Purification kit (Fisher Scientific GmbH, Schwerte, Germany), according to manufacturer's instructions, and the DNA was dissolved in ultrapure water. In the case of additional DNA fragments of unexpected sizes, the entire remaining sample was separated by agarose gel electrophoresis. DNA bands corresponding to DNA strands of the expected length were cut out of the gel. The DNA of interest was extracted and purified from the gel using the PureLink Quick Gel Extraction Kit (Fisher Scientific GmbH, Schwerte, Germany), according to manufacturer's instructions, and dissolved in ultrapure water.

For plasmid amplification *E. coli* strain KL740 (Coli Genetic Stock Center (Yale), CGSG no. 4382⁶⁹), was used. Plasmid preparation was executed using the GenElute HP Plasmid Miniprep Kits or GenElute HP Plasmid Midiprep Kits (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany).

Gibson assembly^{70,71} of PCR products was undertaken using the 2X Gibson-Assembly-Master-Mix (New England Biolabs GmbH, Frankfurt am Main). The 20 μ L assembly reactions were prepared and incubated according to manufacturer's instructions. For a reaction, the total concentration of all DNA fragments to be assembled was between 1 nM and 25 nM. The total concentration always consisted of equal concentrations of the different DNA strands to be assembled. If not stated otherwise the prepared reactions were incubated between 15 min and 30 min at 50 °C in a MultiGene 96-Well Gradient Thermal Cycler (Labnet International Inc., Edison, USA).

3.2. DNA Templates, Plasmids and Synthetic Pap Regulatory Construct. For the assembly of the linear synthetic pap regulatory DNA constructs, different DNA templates were used. All plasmids in this study were based on the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-deGFP-T500 that originates from pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (Addgene plasmid no. 40019), but contains the p15a instead of the ColE1 origin of replication.⁵⁶ The vector backbone was pBEST $_{luc}$ (Promega Corporation, Fitchburg, USA), where the P $_{Tlac}$ -Promoter⁷² was substituted through the strong lambda repressor Cro promoter¹⁸ [GenBank: J02459.1] having a single mutation (OR2-OR1-Pr).⁵⁶ The lambda phage promoter Pr is used for strong gene expression. In order to enable amplification by bacteria, this strong promoter is flanked by the operators OR1 and OR2 to avoid toxic overexpression.⁵⁶ *In vivo* the temperature sensitive lambda repressor CI857 causes repression; however, it is not present in our cell-free expression system.²⁵ The untranslated

region of pBEST $_{luc}$ was substituted by UTR1.⁵⁶ UTR1 [GenBank: M35614.1] is an untranslated region that contains the T7 g10 leader sequence that enables a highly efficient translation initiation.⁷³ T500 is a transcription terminator.⁷⁴

To complete the plasmid pBEST-p15a-OR2-OR1-Pr-deGFP-T500, the *luc* gene of pBEST $_{luc}$ was substituted by the *degfp* gene⁵⁶ that codes for a variant of the enhanced green fluorescent protein [GenBank: CAD97424.1] truncated and modified in N- and C-terminal and which is based on that in ref 75. Accordingly, the plasmid pBEST-p15a-OR2-OR1-Pr-mCherry-T500 was created by insertion of the *mcherry* gene [GenBank: AY678264.1, based on work in ref 76]. The plasmids pBEST-p15a-OR2-OR1-Pr-Lrp-T500 and pBEST-p15a-OR2-OR1-Pr-PapI-T500 were created⁷⁷ by replacement of the *degfp* gene by the genes coding for Lrp [GenBank: 949051] or, respectively, PapI [GenBank: 1039535]. The DNA sequences of regulatory elements and genes of the resulting plasmids pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 and pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 can be found in the Supporting Information (Figure S1).

3.3. Construction of Linear Synthetic Pap Regulatory DNA Constructs. Braaten *et al.*⁵⁰ showed that the single mutation of adenine to cytosine does not influence the binding affinity of Lrp to Lrp BS 5 or Lrp BS 2; however, the mutation prevents Dam methylation of the corresponding DNA motif. Accordingly, three linear synthetic pap regulatory DNA constructs were prepared that differed in sequence at the Lrp binding sites 5 or 2 (pap(1,0), pap(0,1), and pap(1,1)). Here, 1 represents the native DNA sequence GATC (possible Dam methylation site), 0 represents the DNA sequence GCTC.

In this paragraph we explain how the linear three pap regulatory DNA constructs were synthesized. In section 3.4 we illustrate the protocol for the subsequent Dam methylation of these products in order to achieve the methylation states ((1,1), (1,0), (0,1)).

3.3.1. Pap(1,1) DNA Construct. Two partial fragments, A and B (PCR products), were covalently linked using the Gibson assembly and subsequently amplified by PCR.

For the PCR synthesis of partial fragment A, the plasmid pBEST-p15a-OR2-OR1-Pr-mCherry-T500 was used as PCR template with the primers 1 and 2 (Table 1).

For the PCR synthesis of partial fragment B, a DNA construct consisting of the transcription–translation regulatory and coding elements 5'-UTR1-synthetic-pap-regulatory-region-UTR1-deGFP-T500 (synthesized and cloned in vector pEX-K4 by Eurofins Genomics GmbH, Ebersberg, Germany,

Table 2. Expected Fragments from DpnI Digestion of the pap Regulatory DNA Constructs with Different Methylation States⁴

Fragment		Methylation state		
Size [bp]	Start & Stop base	1,0	0,1	1,1
525	970-1494	✓	✓	✓
396	234-629	✓	✓	✓
290	630-919	✗	✓	✗
233	1-233	✓	✓	✓
188	630-817	✓	✗	✓
152	818-969	✓	✗	✗
148	1495-1642	✓	✓	✓
102	818-919	✗	✗	✓
50	920-969	✗	✓	✓
38	1643-1680	✓	✓	✓
17	1681-1697	✓	✓	✓

⁴While the regulatory region exhibits just one (methylation states (1,0) and (0,1)) or two cleavage sites (methylation state (1,1)), the genes coding for mCherry and deGFP contain five cleavage sites, resulting in a total of eight (methylation states (1,0) and (0,1)) or nine (methylation state (1,1)) fragments.

sequence in Figure S2) was used with the primers 3 and 4 (Table 1).

As a difference to the standard protocol of the Gibson assembly, for incubation the following scheme was used to assemble the partial fragments A and B: 1 min at 50 °C, 1 min at 45 °C, 1 min at 50 °C, 1 min at 45 °C, 26 min at 50 °C.

Subsequently, a PCR using Taq DNA polymerase (ThermoPol reaction buffer conditions) was executed in order to amplify the assembled linear synthetic pap regulatory DNA using the primers 1 and 4 (Table 1). It was found that this PCR only produced a product of correct length, if 10 μ L of the Gibson assembly reaction as PCR template DNA was directly transferred into the PCR reaction of a final volume of 300 μ L. The annealing temperature was 50 °C, the elongation time was 1 min and 55 s, and the PCR had 18 cycles. However, many byproducts of other lengths were produced at the same time. The DNA strands of correct length were purified by agarose gel extraction. Surprisingly, we did not find any condition to efficiently amplify this purified DNA construct via an additional PCR. Furthermore, we found that the utilization of primers 5 and 6 (Table 1) resulted in higher yields of the pap regulatory DNA construct if we executed a PCR after the Gibson assembly of partial fragments A and B as described above (300 μ L PCR mix containing 10 μ L of the Gibson assembly mix from the previous assembly reaction). Additionally, we increased PCR efficiency using an elongation time of 1 min and 55 s, an annealing temperature of 56 °C and running the PCR for 15 cycles. The sequence of the resulting synthetic linear pap regulatory DNA construct is given in Figure S3; an alternative way to synthesize this DNA construct, using two

different partial fragments, was also successful as described in paragraph 0.4 of the Supporting Information.

3.3.2. *Pap(1,0)* and *pap(0,1)* DNA Construct. We covalently linked in a Gibson assembly reaction and subsequently amplified by PCR: two PCR products of partial fragments C and D to produce the pap regulatory DNA construct pap(1,0); as well as two PCR products of partial fragments E and F to produce the pap regulatory DNA construct pap(0,1).

The PCR-amplified pap regulatory DNA construct from the Gibson assembly (using the primers 1 and 4 as described in paragraph 3.3.1) was the template DNA for the PCR synthesis of partial fragments C and D, or E and F, respectively. For the PCR synthesis of partial fragment C, the primers 7 and 8 (Table 1) were used. For the PCR synthesis of partial fragment D, the primers 9 and 10 (Table 1) were used. For the PCR synthesis of partial fragment E, the primers 7 and 11 (Table 1) were used. For the PCR synthesis of partial fragment F, the primers 12 and 10 (Table 1) were used.

The primers 8 and 9 (Table 1) led to complementary ends of the partial fragments C and D, so that these could be covalently linked in a Gibson assembly reaction. The primers hybridizing to the template DNA on the Lrp binding site 2 region, had a single mutation (illustrated in square brackets in Table 1) leading to the mutation⁵⁰ GCTC described in paragraph 3.3 in both PCR products C and D.

The primer 11 and 12 (Table 1) led to complementary ends of partial fragments E and F, so that they could be covalently linked in a Gibson assembly reaction. The primers hybridizing to the template DNA on the Lrp binding site 5 region, had a single mutation (illustrated in square brackets in Table 1)

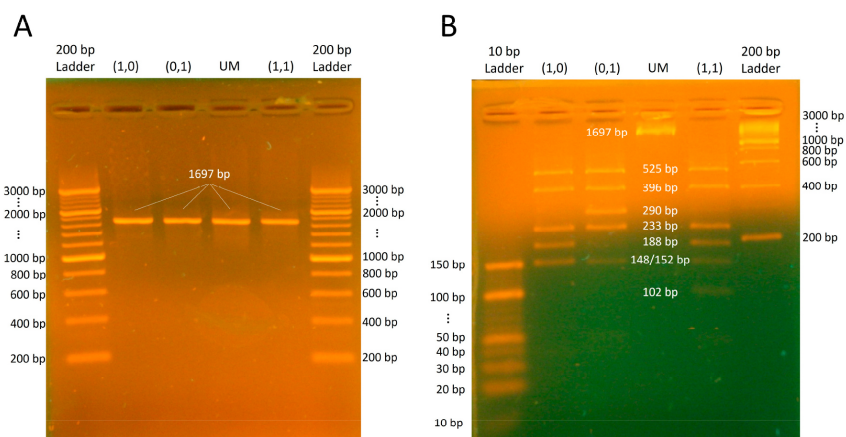


Figure 10. Gel electrophoresis of the methylated pap regulatory DNA with the different methylation states (1,0), (0,1), and (1,1). UM designates the unmethylated pap regulatory DNA (that transforms to (1,1) upon methylation). (A) Agarose gel (0.8%) of the different pap regulatory DNA constructs. A 200 bp ladder (O'RangeRuler 200 bp DNA Ladder, ready-to-use, Thermo Fisher Scientific, Waltham, USA) is present in the first and in the last lane. For all constructs the gel shows just one single band corresponding to the mass of our construct. This indicates high purity. (B) Agarose gel (3%) of different pap regulatory DNA constructs after DpnI digestion. A 10 bp ladder (O'RangeRuler 10 bp DNA Ladder, ready-to-use, Thermo Fisher Scientific, Waltham, USA) is present in the first and a 200 bp ladder (O'RangeRuler 200 bp DNA Ladder, ready-to-use, Thermo Fisher Scientific, Waltham, USA) is present in the last lane. The lane UM shows no traces of digestion of the unmethylated pap regulatory DNA construct. The methylated pap regulatory DNAs with the methylation states (1,0), (0,1), and (1,1) were digested completely (no bands at 1697 bp). We observe the expected fragmentation patterns down to 100bp (for predictions see Table 2). Because of low quantities, bands below 100 bp did not show up in significant intensity. The results indicate that within the limits of gel electrophoresis all of the DNA is methylated by the here employed method.

leading to the mutation⁵⁰ GCTC described in paragraph 3.3 in both PCR products E and F.

Subsequently, partial fragments C and D, as well as E and F, were covalently linked following the Gibson assembly incubation protocol described in paragraph 3.3.1. The PCR amplification using the primers 5 and 6 (Table 1) (elongation time of 1 min and 55 s, an annealing temperature of 56 °C and running the PCR for 15 cycles) and purification by agarose gel extraction of the resulting pap regulatory DNA construct was also executed as described in paragraph 3.3.1.

3.4. Dam Methylation of Pap Regulatory DNA Constructs. DNA adenine methylase (Dam), the supplied 10× reaction buffer and S-adenosylmethionine (SAM) (New England Biolabs GmbH, Frankfurt am Main) were employed for *in vitro* methylation of the purified, unmethylated pap(1,1), pap(1,0), and pap(0,1) constructs (see 3.3). The Dam methylation protocol as used here was established in ref 77, differing from manufacturer's instructions. After Dam methylation, the pap regulatory DNA constructs were purified using the PureLink PCR Purification kit (Fisher Scientific GmbH, Schwerte, Germany) according to manufacturer's instructions, and subsequently dissolved in ultrapure water.

Proper methylation was checked by digestion of the pap regulatory DNA constructs with the restriction enzyme DpnI (New England Biolabs GmbH, Frankfurt am Main, Germany) cleaving at GATC sites only if methylated. This leads to a characteristic fragmentation pattern in case of successful methylation (see Table 2). After digestion, no traces of bands corresponding to undigested synthetic pap regulatory DNAs could be observed.

The expected, characteristic fragmentation patterns for the different methylation states of the pap regulatory DNA could be observed by agarose gel electrophoresis indicating proper methylation (Figure 10).

3.5. Cell-Free Expression and Fluorescence Measurements. All experiments were performed using an all *E. coli* cell-free system described previously.^{56,57} The main component of this system consists of an *E. coli* cell lysate, called crude extract.^{56–59} A ready-to-use version of the all *E. coli* cell-free system is commercially available as the kit myTXTL (Arbor Biosciences, Ann Arbor, USA).

For the experiments presented in Figure 3 and Figure S4, cell-free expression master mix solutions were split into volumes of 9.6 μL . A split reaction volume was completed with 2.4 μL of DNA solution (25 nM) of one of the three pap regulatory DNA constructs pap(1,1), pap(1,0), or pap(0,1) or ultrapure water to a total volume of 12 μL . Accordingly, the concentrations of a pap regulatory DNA construct was 5 nM.

For the experiments presented in Figure 4 and S5, cell-free expression master mix solutions were split into volumes of 9.6 μL . A split reaction volume was completed with 1.2 μL DNA solution of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 and 1.2 μL of DNA solution of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 at the concentrations 0 nM (ultrapure water), 1 nM, and 10 nM. The final concentrations of the plasmids above in the reaction mix were 0 nM, 0.1 nM, or 1 nM. The pap regulatory DNA constructs pap(1,1), pap(1,0), or pap(0,1) in the cell-free reaction mix were at a final concentration of 5 nM.

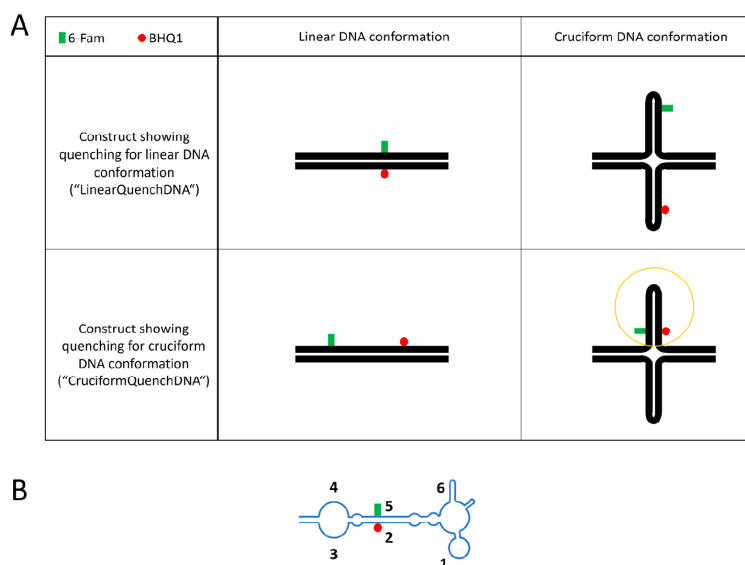


Figure 11. (A) Scheme of the possible DNA conformations for both types of DNA constructs. LinearQuenchDNA shows quenching of 6-Fam fluorescence due to the proximity to BHQ1 if the DNA is linear; CruciformQuenchDNA shows quenching as a cruciform. The orange circle in the cruciform conformation of CruciformQuenchDNA indicates the region that is shown in more detail in part B. (B) More detailed view of the 6-Fam/BHQ1 positions (adjacent to Lrp BS 5 and BS 2) of the CruciformQuenchDNA in a cruciform conformation.

To protect the linear DNA against *E. coli* RecBCD induced degradation, the GamS protein⁷⁸ was added to the cell-free reactions (3.3 μ M).

For cell-free expression, reactions were incubated at 29 °C. Fluorescence measurements were performed using a microplate reader (POLARstar Optima, BMG Labtech, Ortenberg, Germany, observed through bottom optics). For fluorescence measurements, cell-free reactions (10 μ L) were transferred into a 384-well plate (Nunc 384-well optical bottom plate, Thermo Fisher Scientific, Waltham, USA) that was sealed (Nunc sealing tape, Thermo Fisher Scientific, Waltham, USA).

During kinetic measurements, cell-free reactions were incubated in a microplate reader for 16 h. For end point measurements, cell-free reactions were first incubated in 1.5 mL reaction tubes (Eppendorf Safe-Lock Tube 1.5 mL, PCR clean, Eppendorf, Hamburg, Germany) for 16 h and then transferred to the 384-well plate. Cell-free synthesized deGFP was excited at 485 nm and the emission intensity determined at 520 nm. For mCherry, the excitation wavelength was 595 nm while its emission intensity was determined at 620 nm. For all fluorescence intensity measurements the same parameters were used (gains were set to 1100 and 2200 for the deGFP and mCherry channel, respectively). The measured intensities were normalized to the highest deGFP or mCherry fluorescence intensity values (Figures 3 and 4).

Cell-free reactions in which no (pap regulatory) DNA was present served as a blank. The fluorescence intensities of the blank as determined with the deGFP and mCherry filter sets were determined as outlined above. The blank was subtracted to obtain the intensity due to the presence of fluorescent protein.

3.6. Prediction of Pap Regulatory DNA Secondary Structures. Prediction of the secondary structure was performed using the mfold web server.⁶⁵ The underlying DNA sequence corresponds to the sequence of the synthetic pap regulatory region including the upstream and downstream UTR1 sequences and the respective beginnings of the *degfp* and *mcherry* genes (base pairs 685–1053 in sequence presented in Figure S3). The structure was predicted for a temperature of 29 °C (incubation temperature of cell-free reactions) using the standard settings (DNA type, linear DNA; ionic conditions, 1 M Na⁺; correction type, oligomer; percent suboptimality number, 5; upper bound on the number of computed foldings, 50; maximum distance between paired bases, no limit). These conditions were chosen to reflect the observation by Schoen *et al.* that DNA hybridizes much faster in the cell than *in vitro*.⁷⁹ The program Rna Viz⁸⁰ generated Figure S6.

To decide whether the predicted secondary structures are present in our expression system, we labeled the synthetic pap regulatory DNA constructs with the fluorescence marker 6-Fam and the quencher BHQ1 at specific positions. For all three methylation states we designed two different constructs: one showing quenching for the linear conformation, the other for some cruciform conformations (Figure 11).

Because the constructs could not be synthesized as one single construct, we split them in four single stranded DNA oligos. Two pairs of them are complementary to each other and have a complementary overlap with the other pair. The oligos were synthesized by Metabion international AG (Planegg, Germany). Their sequence is given in the Supporting Information (Figure S7).

To obtain the pap regulatory DNA constructs, either pair of complementary single strands was dissolved in annealing buffer (10 mM Tris, pH 7.6; 50 mM NaCl; 1 mM EDTA) and annealed in a thermocycler. The annealing program consists of an initial denaturation for 2 min at 95 °C followed by a stepwise decrease of the temperature by 1.4 °C each minute (50 steps). A small amount of the obtained double stranded DNA constructs was kept as reference to measure the fluorescence intensity of 6-Fam in absence of BHQ1 (dsDNA_{Ann}). The two corresponding double stranded DNA constructs were ligated (Fast-Link DNA Ligation Kit, Lucigen, Middleton, USA). The ligated constructs were purified using a purification kit (DNA Clean and Concentrator 25, Zymo Research Europe, Freiburg, Germany) and subsequently methylated as described in section 3.4.

The obtained constructs were purified a second time and then added to cell-free expression master mix solution aliquots either with or without crude extract (ultrapure water was used as substitute). The concentration of the constructs was 120 nM (increased concentration necessary due to weak fluorescence signal of 6-Fam), and the reaction volume was 12 μ L. The samples were incubated at 29 °C, and the time course of fluorescence intensity was recorded for about 8 h.

4. CONCLUSION

Here, we constructed a synthetic pap DNA regulatory element that controls gene expression upstream and downstream as a function of only two methylation sites. We monitored its gene expression and regulation in a cell-free expression system^{56,57} based on the endogenous *E. coli* transcription–translation machinery. Our results excellently reproduced previous observations by Hernday and colleagues,⁵⁸ showing converse regulatory effects of methylation of either GATC^{prox} or GATC^{dist} sites as part of the phase variation mechanism of UPEC. However, our results for simultaneous methylation of both sites could not be explained based on current literature. Our results suggest that cruciform secondary structures of the pap regulatory DNA are an integral part of the two-state phase variation mechanism. Using magnetic tweezers and Förster resonance energy transfer (FRET), the predicted pap regulatory DNA cruciform conformations could be investigated in more detail in the future. We expect the transcription to respond to pN forces.⁸¹

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssynbio.1c00326>.

Relevant DNA sequences of the DNA constructs used in this work for gene expression; production of the partial fragments A1 and B1 for assembly of the synthetic pap regulatory region pap(1,1) via Gibson assembly reaction; additional figures that support the text (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Albrecht Ott – Universität des Saarlandes, Center for Biophysics, Saarbrücken 66123, Germany; orcid.org/0000-0003-0481-2658; Email: albrecht.ott@physik.uni-saarland.de

Authors

Emanuel G. Worst – Universität des Saarlandes, Center for Biophysics, Saarbrücken 66123, Germany

Marc Finkler – Universität des Saarlandes, Center for Biophysics, Saarbrücken 66123, Germany; orcid.org/0000-0002-8986-0378

Marc Schenkelberger – Universität des Saarlandes, Center for Biophysics, Saarbrücken 66123, Germany

Ömer Kurt – Universität des Saarlandes, Center for Biophysics, Saarbrücken 66123, Germany

Volkhard Helms – Universität des Saarlandes, Center for Bioinformatics, Saarbrücken 66041, Germany; orcid.org/0000-0002-2180-9154

Vincent Noireaux – University of Minnesota, School of Physics and Astronomy, Minneapolis, Minnesota 55455, United States; orcid.org/0000-0002-5213-273X

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acssynbio.1c00326>

Author Contributions

A.O. and E.G.W. designed the research with the help of M.S. and V.N. E.G.W. and M.F. performed the research. A.O. and M.F. designed the fluorescence quenching experiments concerning the investigation of the secondary structure of the synthetic pap switch. E.G.W., M.F., and A.O. analyzed and interpreted the data. E.G.W., M.F., Ö.K., and A.O. wrote the paper. V.N., V.H., and A.O. supervised.

Author Contributions

[#]E.G.W. and M.F. contributed equally.

Notes

The authors declare no competing financial interest. Results excluding the fluorescence quenching experiments concerning the investigation of the secondary structure of our synthetic pap switch were previously published as part of the Ph.D. thesis of E.G.W.⁸²

■ ACKNOWLEDGMENTS

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) within the collaborative research center (CRC/SFB 1027) as well as financial support from the Human Frontiers Science Program Organization (HFSP, RGP0037/2015). We thank Julia Flesch and Lucina Kainka for their assistance with the amplification of DNA templates. We thank Po-IIsien for fruitful discussions.

■ REFERENCES

- (1) Tian, J.; Gong, H.; Sheng, N.; Zhou, X.; Gulari, E.; Gao, X.; Church, G. Accurate Multiplex Gene Synthesis from Programmable DNA Microchips. *Nature* **2004**, *432*, 1050–1054.
- (2) Gibson, D. G. Enzymatic Assembly of Overlapping DNA Fragments. *Methods Enzymol.* **2011**, *498*, 349–361, DOI: [10.1016/B978-0-12-385120-8.00015-2](https://doi.org/10.1016/B978-0-12-385120-8.00015-2).
- (3) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D. Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux. *Nat. Biotechnol.* **2009**, *27*, 753–759.
- (4) Good, M. C.; Zalatan, J. G.; Lim, W. A. Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* **2011**, *332*, 680–686.
- (5) Huang, C. C.; Chen, M. W.; Hsieh, J. L.; Lin, W. H.; Chen, P. C.; Chien, L. F. Expression of Mercuric Reductase from *Bacillus Megaterium* MB1 in Eukaryotic Microalga *Chlorella* Sp. DT: An

- Approach for Mercury Phytoremediation. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 197–205.
- (6) Burgard, A. P.; Parkya, P.; Maranas, C. D. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnol. Bioeng.* **2003**, *84*, 647–657.
- (7) Nash, P. D. Why Modules Matter. *FEBS Lett.* **2012**, *586*, 2572–2574.
- (8) Worst, E. G.; Exner, M. P.; De Simone, A.; Schenkelberger, M.; Noireaux, V.; Budisa, N.; Ott, A. Cell-Free Expression with the Toxic Amino Acid Canavanine. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 3658–3660.
- (9) Worst, E. G.; Exner, M. P.; de Simone, A.; Schenkelberger, M.; Noireaux, V.; Budisa, N.; Ott, A. Residue-Specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an Escherichia Coli Cell-Free Transcription-Translation System. *J. Visualized Exp.* **2016**, *114*, 1–14.
- (10) Finkler, M.; Ravanbodshirazi, S.; Grimm, F.; Hartz, P.; Ott, A. Full Incorporation of the Noncanonical Amino Acid Hydroxylysine as a Surrogate for Lysine in Green Fluorescent Protein. *Bioorg. Med. Chem.* **2021**, *41*, 1–6.
- (11) Elowitz, M. B.; Leibler, S. A Synthetic Oscillatory Network of Transcriptional Regulators. *Nature* **2000**, *403*, 335–338.
- (12) Siuti, P.; Yazbek, J.; Lu, T. K. Synthetic Circuits Integrating Logic and Memory in Living Cells. *Nat. Biotechnol.* **2013**, *31*, 448–452.
- (13) Bonnet, J.; Yin, P.; Ortiz, M. E.; Subsoontorn, P.; Endy, D. Amplifying Genetic Logic Gates. *Science* **2013**, *340*, 599–603.
- (14) Gardner, T. S.; Cantor, C. R.; Collins, J. J. Construction of a Genetic Toggle Switch in Escherichia Coli. *Nature* **2000**, *403*, 339–342.
- (15) Citorik, R. J.; Mimee, M.; Lu, T. K. Bacteriophage-Based Synthetic Biology for the Study of Infectious Diseases. *Curr. Opin. Microbiol.* **2014**, *19*, 59–69.
- (16) Cheng, A. A.; Lu, T. K. Synthetic Biology: An Emerging Engineering Discipline. *Annu. Rev. Biomed. Eng.* **2012**, *14*, 155–178.
- (17) Romanel, A.; Jensen, L. J.; Cardelli, L.; Csikász-Nagy, A. Transcriptional Regulation Is a Major Controller of Cell Cycle Transition Dynamics. *PLoS One* **2012**, *7*, 1–9.
- (18) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*, 4th ed.; W. H. Freeman: New York, 2000.
- (19) Liu, D.; Hoynes-O'Connor, A.; Zhang, F. Bridging the Gap between Systems Biology and Synthetic Biology. *Front. Microbiol.* **2013**, *4*, 1–8.
- (20) Finkler, M.; Ott, A. Bead-Based Assay for Spatiotemporal Gene Expression Control in Cell-Free Transcription–Translation Systems. *BioTechniques* **2019**, *66*, 29–33.
- (21) Finkler, M.; Kurt, Ö.; Grimm, F.; Hartz, P.; Ott, A. A Bead-Based Method for the Removal of the Amino Acid Lysine from Cell-Free Transcription-Translation Systems. *J. Biotechnol.* **2020**, *324S*, 1–6.
- (22) Bikard, D.; Jiang, W.; Samai, P.; Hochschild, A.; Zhang, F.; Marraffini, L. a. Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System. *Nucleic Acids Res.* **2013**, *41*, 7429–7437.
- (23) Sander, J. D.; Joung, J. K. CRISPR-Cas Systems for Editing, Regulating and Targeting Genomes. *Nat. Biotechnol.* **2014**, *32*, 347–355.
- (24) Colot, V.; Rossignol, J. L. Eukaryotic DNA Methylation as an Evolutionary Device. *BioEssays* **1999**, *21*, 402–411.
- (25) Schenkelberger, M.; Shanak, S.; Finkler, M.; Worst, E. G.; Noireaux, V.; Helms, V.; Ott, A. Expression Regulation by a Methyl-CpG Binding Domain in an E. Coli Based, Cell-Free TX-TL System. *Phys. Biol.* **2017**, *14*, 1–11.
- (26) Kucukkal, T. G.; Yang, Y.; Uvarov, O.; Cao, W.; Alexov, E. Impact of Rett Syndrome Mutations on MeCP2/MBD Stability. *Biochemistry* **2015**, *54*, 6357–6368.
- (27) Casadesús, J.; Low, D. Epigenetic Gene Regulation in the Bacterial World. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 830–856.
- (28) Marinus, M. G.; Löbner-Olesen, A. DNA Methylation. *EcoSal Plus* **2013**, *6*, No. 0003, DOI: 10.1128/ecosalplus.ESP-0003-2013.
- (29) Barras, F.; Marinus, M. G. The Great GATC: DNA Methylation in E. Coli. *Trends Genet.* **1989**, *5*, 139–143.
- (30) Modrich, P. Methyl-Directed DNA Mismatch Correction. *J. Biol. Chem.* **1989**, *264*, 6597–6600.
- (31) Palmer, B. R.; Marinus, M. G. The Dam and Dcm Strains of Escherichia Coli - a Review. *Gene* **1994**, *143*, 1–12.
- (32) Wion, D.; Casadesús, J. N6-Methyl-Adenine: An Epigenetic Signal for DNA–Protein Interactions. *Nat. Rev. Microbiol.* **2006**, *4*, 183–192.
- (33) Han, J. S.; Kang, S.; Lee, H.; Kim, H. K.; Hwang, D. S. Sequential Binding of SeqA to Paired Hemi-Methylated GATC Sequences Mediates Formation of Higher Order Complexes. *J. Biol. Chem.* **2003**, *278*, 34983–34989.
- (34) Maier, J. A. H.; Möhrle, R.; Jeltsch, A. Design of Synthetic Epigenetic Circuits Featuring Memory Effects and Reversible Switching Based on DNA Methylation. *Nat. Commun.* **2017**, *8*, 1–10.
- (35) Park, M.; Patel, N.; Keung, A. J.; Khalil, A. S. Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* **2019**, *176*, 227–238.
- (36) Chang, T. Z.; Kuo, J.; Silver, P. A. Beyond the Four Bases: A Home Run for Synthetic Epigenetic Control? *Mol. Cell* **2019**, *74*, 5–7.
- (37) Blyn, L. B.; Braaten, B. A.; White-Ziegler, C. A.; Rolfson, D. H.; Low, D. A. Phase-Variation of Pylonephritis-Associated Pili in Escherichia Coli: Evidence for Transcriptional Regulation. *EMBO J.* **1989**, *8*, 613–620.
- (38) Hernday, A. D.; Braaten, B. A.; Low, D. A. The Mechanism by Which DNA Adenine Methylase and PapI Activate the Pap Epigenetic Switch. *Mol. Cell* **2003**, *12*, 947–957.
- (39) Forsman, K.; Göransson, M.; Uhlin, B. E. Autoregulation and Multiple DNA Interactions by a Transcriptional Regulatory Protein in E. Coli Pili Biogenesis. *EMBO J.* **1989**, *8*, 1271–1277.
- (40) Weyand, N. J.; Braaten, B. A.; van der Woude, M.; Tucker, J.; Low, D. A. The Essential Role of the Promoter-Proximal Subunit of CAP in Pap Phase Variation: Lrp- and Helical Phase-Dependent Activation of PapBA Transcription by CAP from –215. *Mol. Microbiol.* **2001**, *39*, 1504–1522.
- (41) Båga, M.; Göransson, M.; Normark, S.; Uhlin, B. E. Transcriptional Activation of a Pap Pilus Virulence Operon from Uropathogenic Escherichia Coli. *EMBO J.* **1985**, *4*, 3887–3893.
- (42) Hernday, A.; Krabbe, M.; Braaten, B.; Low, D. Self-Perpetuating Epigenetic Pili Switches in Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16470–16476.
- (43) Newman, E. B.; D'Ari, R.; Lin, R. T. The Leucine-Lrp Regulator in E. Coli: A Global Response in Search of a Raison d'Être. *Cell* **1992**, *68*, 617–619.
- (44) Calvo, J. M.; Matthews, R. G. The Leucine-Responsive Regulatory Protein, a Global Regulator of Metabolism in Escherichia Coli. *Microbiol. Rev.* **1994**, *58*, 466–490.
- (45) Nou, X.; Braaten, B.; Kaltenbach, L.; Low, D. A. Differential Binding of Lrp to Two Sets of Pap DNA Binding Sites Mediated by Pap I Regulates Pap Phase Variation in Escherichia Coli. *EMBO J.* **1995**, *14*, 5785–5797.
- (46) Cui, Y.; Midkiff, M. A.; Wang, Q.; Calvo, J. M. The Leucine-Responsive Regulatory Protein (Lrp) from Escherichia Coli. Stoichiometry and Minimal Requirements for Binding to DNA. *J. Biol. Chem.* **1996**, *271*, 6611–6617.
- (47) Leonard, P. M.; Smits, S. H. J.; Sedelnikova, S. E.; Brinkman, A. B.; Vos, W. M.; de Oost, J.; van der Rice, D. W.; Rafferty, J. B. Crystal Structure of the Lrp-like Transcriptional Regulator from the Archaeon Pyrococcus Furiosus. *EMBO J.* **2001**, *20*, 990–997.
- (48) de los Rios, S.; Perona, J. J. Structure of the Escherichia Coli Leucine-Responsive Regulatory Protein Lrp Reveals a Novel Octameric Assembly. *J. Mol. Biol.* **2007**, *366*, 1589–1602.

- (49) Kawamura, T.; Vartanian, A. S.; Zhou, H.; Dahlquist, F. W. The Design Involved in PapI and Lrp Regulation of the Pap Operon. *J. Mol. Biol.* **2011**, *409*, 311–332.
- (50) Braaten, B. A.; Nou, X.; Kaltenbach, L. S.; Low, D. A. Methylation Patterns in Pap Regulatory DNA Control Pyelonephritis-Associated Pili Phase Variation in *E. Coli*. *Cell* **1994**, *76*, 577–588.
- (51) Peterson, S. N.; Reich, N. O. Competitive Lrp and Dam Assembly at the Pap Regulatory Region: Implications for Mechanisms of Epigenetic Regulation. *J. Mol. Biol.* **2008**, *383*, 92–105.
- (52) Weyand, N. J.; Low, D. A. Regulation of Pap Phase Variation. Lrp Is Sufficient for the Establishment of the Phase OFF Pap DNA Methylation Pattern and Repression of Pap Transcription in Vitro. *J. Biol. Chem.* **2000**, *275*, 3192–3200.
- (53) Göransson, M.; Forsman, K.; Nilsson, P.; Uhlin, B. E. Upstream Activating Sequences That Are Shared by Two Divergently Transcribed Operons Mediate CAMP-CRP Regulation of Pilus-Adhesin in *Escherichia Coli*. *Mol. Microbiol.* **1989**, *3*, 1557–1565.
- (54) Peterson, S. N.; Reich, N. O. GATC Flanking Sequences Regulate Dam Activity: Evidence for How Dam Specificity May Influence Pap Expression. *J. Mol. Biol.* **2006**, *355*, 459–472.
- (55) Mashhoon, N.; Carroll, M.; Pruss, C.; Eberhard, J.; Ishikawa, S.; Estabrook, R. A.; Reich, N. Functional Characterization of *Escherichia Coli* DNA Adenine Methyltransferase, a Novel Target for Antibiotics. *J. Biol. Chem.* **2004**, *279*, 52075–52081.
- (56) Shin, J.; Noireaux, V. Efficient Cell-Free Expression with the Endogenous *E. Coli* RNA Polymerase and Sigma Factor 70. *J. Biol. Eng.* **2010**, *4*, 1–9.
- (57) Sun, Z. Z.; Hayes, C. A.; Shin, J.; Caschera, F.; Murray, R. M.; Noireaux, V. Protocols for Implementing an *Escherichia Coli* Based TX-TL Cell-Free Expression System for Synthetic. *J. Visualized Exp.* **2013**, *79*, 1–14.
- (58) Shin, J.; Noireaux, V. An *E. Coli* Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* **2012**, *1*, 29–41.
- (59) Garamella, J.; Marshall, R.; Rustad, M.; Noireaux, V. The All *E. Coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synth. Biol.* **2016**, *5*, 344–355.
- (60) Hohng, S.; Zhou, R.; Nahas, M. K.; Yu, J.; Schulten, K.; Lilley, D. M. J.; Ha, T. Fluorescence-Force Spectroscopy Maps Two-Dimensional Reaction Landscape of the Holliday Junction. *Science* **2007**, *318*, 279–283.
- (61) Becker, N. a.; Peters, J. P.; Lionberger, T. A.; Maher, L. J. Mechanism of Promoter Repression by Lac Repressor-DNA Loops. *Nucleic Acids Res.* **2013**, *41*, 156–166.
- (62) Garenne, D.; Beisel, C. L.; Noireaux, V. Characterization of the All-*E. Coli* Transcription-Translation System MyTXTL by Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2019**, *33*, 1036–1048.
- (63) Kaltenbach, L. S.; Braaten, B. A.; Low, D. A. Specific Binding of PapI to Lrp-Pap DNA Complexes. *J. Bacteriol.* **1995**, *177*, 6449–6455.
- (64) Brázda, V.; Laister, R. C.; Jagelská, E. B.; Arrowsmith, C. Cruciform Structures Are a Common DNA Feature Important for Regulating Biological Processes. *BMC Mol. Biol.* **2011**, *12*, 1–16.
- (65) Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415.
- (66) Bae, S.-H.; Cheong, H.-K.; Cheong, C.; Kang, S.; Hwang, D. S.; Choi, B.-S. Structure and Dynamics of Hemimethylated GATC Sites. Implications for DNA-SeqA Recognition. *J. Biol. Chem.* **2003**, *278*, 45987–45993.
- (67) SantaLucia, J., Jr. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460–1465.
- (68) Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L.; Behlke, M. A.; Walder, J. A. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry* **2004**, *43*, 3537–3554.
- (69) Sussman, R.; Jacob, F. Sur Un Système de Répression Thermosensible Chez Le Bactériophage d'*Escherichia Coli*. *C. R. Hebd. Seances Acad. Sci.* **1962**, *254*, 1517–1519.
- (70) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A., III; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345.
- (71) Gibson, D. G.; Smith, H. O.; Hutchison, C. A., III; Venter, J. C.; Merryman, C. Chemical Synthesis of the Mouse Mitochondrial Genome. *Nat. Methods* **2010**, *7*, 901–903.
- (72) de Boer, H. A.; Comstock, L. J.; Vasser, M. The Tac Promoter: A Functional Hybrid Derived from the Trp and Lac Promoters. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 21–25.
- (73) Olins, P. O.; Devine, C. S.; Rangwala, S. H.; Kavka, K. S. The T7 Phage Gene 10 Leader RNA, a Ribosome-Binding Site That Dramatically Enhances the Expression of Foreign Genes in *Escherichia Coli*. *Gene* **1988**, *73*, 227–235.
- (74) Larson, M. H.; Greenleaf, W. J.; Landick, R.; Block, S. M. Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. *Cell* **2008**, *132*, 971–982.
- (75) Li, X.; Zhang, G.; Ngo, N.; Zhao, X.; Kain, S. R.; Huang, C. C. Deletions of the *Aequorea Victoria* Green Fluorescent Protein Define the Minimal Domain Required for Fluorescence. *J. Biol. Chem.* **1997**, *272*, 28545–28549.
- (76) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y. Improved Monomeric Red, Orange and Yellow Fluorescent Proteins Derived from *Discosoma Sp.* Red Fluorescent Protein. *Nat. Biotechnol.* **2004**, *22*, 1567–1572.
- (77) Schenkelberger, M. Cooperative Biomolecular Binding High Specificity of Competitive Single Stranded DNA. Ph.D. Thesis, Universität des Saarlandes, 2013.
- (78) Murphy, K. C. The λ Gam Protein Inhibits RecBCD Binding to DsDNA Ends. *J. Mol. Biol.* **2007**, *371*, 19–24.
- (79) Schoen, I.; Krammer, H.; Braun, D. Hybridization Kinetics Is Different inside Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21649–21654.
- (80) De Rijk, P.; Wuyts, J.; De Wachter, R. RnaViz 2: An Improved Representation of RNA Secondary Structure. *Bioinformatics* **2003**, *19*, 299–300.
- (81) Swoboda, M.; Grieb, M. S.; Hahn, S.; Schlierf, M. Measuring Two at the Same Time: Combining Magnetic Tweezers with Single-Molecule FRET. *Fluorescent Methods for Molecular Motors. Experientia Supplementum.* **2014**, *105*, 253–276, DOI: 10.1007/978-3-0348-0856-9_12.
- (82) Worst, E. G. Evolution Wachsender, Sich Reproduzierender Polymere. Zellfreie Genexpression Zum Einbau Nicht-Kanonischer Aminosäuren in Proteine Sowie Zur Analyse Des Epigenetischen *Escherichia Coli*-Pili-Phasenvariationsmechanismus. (Evolution of Growing, Reproducing Polymers. Cell-free gene expression for the incorporation of non-canonical amino acids into proteins and for the analysis of the epigenetic *Escherichia coli* pili phase variation mechanism). Ph.D. Thesis, Universität des Saarlandes, 2017.

Supporting Information

A Methylation-Directed, Synthetic Pap Switch Based on Self-Complementary Regulatory DNA Reconstituted in an All *E. coli* Cell-Free Expression System.

Worst, E. G.†, **Finkler, M.†**, Schenkelberger, M., Kurt, Ö., Helms, V., Noireaux, V., and Ott, A.

† Geteilte Erstautorenschaft

ACS Synthetic Biology (2021), 10, 2725–2739.

DOI: 10.1021/acssynbio.1c00326

<https://doi.org/10.1021/acssynbio.1c00326>

Reprinted with permission from Worst, E. G.†, **Finkler, M.†**, Schenkelberger, M., Kurt, Ö., Helms, V., Noireaux, V., and Ott, A. (2021) A Methylation-Directed, Synthetic Pap Switch Based on Self-Complementary Regulatory DNA Reconstituted in an All *E. coli* Cell-Free Expression System. *ACS Synth. Biol.* 10, 2725–2739. <https://doi.org/10.1021/acssynbio.1c00326>. Copyright 2021 American Chemical Society.

SUPPORTING INFORMATION

A methylation-directed, synthetic pap switch based on self-complementary regulatory DNA reconstituted in an all *E. coli* cell-free expression system

Emanuel G. Worst^{1,†}, Marc Finkler^{1,†}, Marc Schenkelberger¹, Ömer Kurt¹, Volkhard Helms², Vincent Noireaux³, and Albrecht Ott^{1,*}

¹Universität des Saarlandes, Center for Biophysics, Saarbrücken, 66123, Germany

²Universität des Saarlandes, Center for Bioinformatics, Saarbrücken, 66041, Germany

³University of Minnesota, School of Physics and Astronomy, Minneapolis, Minnesota 55455, USA

† Authors contributed equally.

* corresponding author. E-Mail: albrecht.ott@physik.uni-saarland.de

DNA sequences

0.1 Relevant DNA sequences of the transcription-translation-regulatory and coding elements of the plasmids pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 and pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500

(A)

```

1   TGAGC TAACA CCGTG CGTGT TGACA ATTTT ACCTC TGGCG GTGAT AATGG TTGCA GCTAG
61  CAATA ATTTT GTTTA ACTTT AAGAA GGAGA TATAC CATGG TAGAT AGCAA GAAGC GCCCT
121 GGCAA AGATC TCGAC CGTAT CGATC GTAAC ATTCT TAATG AGTTG CAAAA GGATG GCGCT
181 ATTTC TAACG TCGAG CTTTC TAAAC GTGTG GACTT TTCCC CAACG CCGTG CCTTG AGCGT
241 GTGCG TCGGC TGGAA AGACA AGGGT TTATT CAGGG CTATA CGGCG CTGCT TAACC CCCAT
301 TATCT GGATG CATCA CTTCT GGTAT TCGTT GAGAT TACTC TGAAT CGTGG CGCAC CGGAT
361 GTGTT TGAAC AATTC AATAC CGCTG TACAA AACTT TGAAG AAATT CAGGA GTGTC ATTTA
421 GTATC CGGTG ATTTT GACTA CCTGT TGAAA ACACG CGTGC CGGAT ATGTC AGCCT ACCGT
481 AAGTT GCTGG GGGAA ACCCT GCTGC GTCTG CCTGG CGTCA ATGAC ACACG GACAT ACGTT
541 GTTAT GGAAG AAGTC AAGCA GAGTA ATCGT CTGGT TATTA AGACG CGCTA ACTCG AGCAA
601 AGCCC GCCGA AAGGC GGGCT TTTCT GT

```

(B)

```

1   TGAGC TAACA CCGTG CGTGT TGACA ATTTT ACCTC TGGCG GTGAT AATGG TTGCA GCTAG
61  CAATA ATTTT GTTTA ACTTT AAGAA GGAGG ATCCA AATGA GTGAA TATAT GAAGA ATGAA
121 ATACT GGAAT TTCTG AACAG GCATG ATGGA GGGAA AACGG CAGAA ATTGC AGAGG CGCTG
181 GCGGT AACGG ATTAC CAGGC CCGCT ATTAC CTGTT ATTAC TGGAA AAAGC AGGTA TGGTT
241 CAGCG CTCAC CATT AAGAC GGGCA TGGCA ACATA CTGGT TTCTG AAAGG AGAGA AGCAG
301 GCGGG GCAGA GCTGT TCTTC CACAA CTTAA CTCTGA GAGTC GACCA AAGCC CGCCG AAAGG
361 CGGGC TTTTC TGT

```

Figure S1. Relevant DNA sequences of the transcription-translation-regulatory and coding elements of the plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-Lrp-T500 (A) and plasmid pBEST-p15a-OR2-OR1-Pr-UTR1-PapI-T500 (B) for the cell-free expression of the genes *lrp* or *papI*, respectively. (A, B) Turquoise: OR2-Or1-Pr with a single mutation. Grey: transcription terminator T500. (A) Purple: *lrp* gene. Yellow: UTR1. (B) Purple: *papI* gene. Yellow: modified UTR1. Details related to transcription-translation-regulatory and coding elements can be found in section 3.2.

0.2 Relevant DNA sequences of the transcription-translation-regulatory and coding elements of the DNA template used for the PCR synthesis of partial DNA fragment B

```

1   GGATAT ATCTC CTTCT TAAAG TTAAA CAAAA TTATT GCTAG CTGCA ACCAT TATGC GAAAA
61  GAAAG TCCGT ATGTC AACAT TTAGA CGATC TTTTA TGCTG TAAAT TCAAT TTGCC ATGAT
121 GTTTT TATCT GAGTA CCCTC TTGCT ATTAG TGTTT TGTTT TAGTT TAATT TTGTT TTGTG
181 GGTTA AAAGA TCGTT TAAAT CTTGA CATAA AACAT AAAAA ACTAA ATAAT GGTTG CAGGA
241 TCCAA TAATT TTGTT TAACT TTAAG AAGGA GATAT ACCAT GGAGC TTTTC ACTGG CGTTG
301 TTCCC ATCCT GGTCG AGCTG GACGG CGACG TAAAC GGCCA CAAGT TCAGC GTGTC CGGCC
361 AGGC GAGGG CGATG CCACC TACGG CAAGC TGACC CTGAA GTTCA TCTGC ACCAC CGGCA
421 AGCTG CCCGT GCCCT GGCCC ACCCT CGTGA CCACC CTGAC CTACG GCGTG CAGTG CTTCA
481 GCCCG TACCC CCACC ACATG AACCA GCACC ACTTC TTCAA CTCCG CCATC CCCC AAGCT
541 ACGTC CAGGA GCGCA CCATC TTCTT CAAGG ACGAC GGCAA CTACA AGACC CGCGC CGAGG
601 TGAAG TTCGA GGGCG ACACC CTGGT GAACC GCATC GAGCT GAAGG GCATC GACTT CAAGG
661 AGGAC GGCAA CATCC TGGGG CACAA GCTGG AGTAC AACTA CAACA GCCAC AACGT CTATA
721 TCATG GCCGA CAAGC AGAAG AACGG CATCA AGGTG AACTT CAAGA TCCGC CACAA CATCG
781 AGGAC GGCAG CGTGC AGCTC GCCGA CCACT ACCAG CAGAA CACCC CCATC GGCGA CGGCC
841 CCGTG CTGCT GCCCG ACAAC CACTA CTTGA GCACC CAGTC CGCCC TGAGC AAAGA CCCCC
901 ACGAG AAGCT CGATC AATG CTAAT GCTGG AGTTC GTGAC CGCCG CCGGG ATCTA ACTCG
961 AGCAA AGCCC GCCGA AAGGC GGGCT TTTCT GTGTC GACCG ATGCC CTTGA GAGGG TTTTC
1021 TGGCT GATGG TG

```

Figure S2. Relevant DNA sequences of the transcription-translation-regulatory and coding elements of the DNA template used for the PCR synthesis of partial DNA fragment B. Yellow: UTR1. Blue: -10 elements. Turquoise: -35 elements. Grey: Lrp binding sites 4 – 6 and 1 – 3 according to Hernday *et al.*¹. Green: *degfp* gene. Black: transcription terminator T500. Details related to transcription-translation-regulatory and coding elements can be found in paragraph 3.2.

0.3 DNA sequence of the linear synthetic pap regulatory DNA products pap(1,1), pap(0,1) and pap(1,0)

```

1   CGCTA TGAAT TCCTC GAGTT ACTTG TACAG CTCGT CCATG CCGCC GGTGG AGTGG CGGCC
61  CTCGG CGCGT TCGTA CTGTT CCACG ATGGT GTAGT CCTCG TTGTG GGAGG TGATG TCCAA
121 CTTGA TGTTG ACGTT GTAGG CGCCG GGCAG CTGCA CGGGC TTCTT GGCCCT TGTAG GTGGT
181 CTTGA CCTCA GCGTC GTAGT GGCCG CCGTC CTTCA GCTTC AGCCT CTGCT TGATC TCGCC
241 CTTCA GGGCG CCGTC CTCGG GGTAC ATCCG CTCGG AGGAG GCCTC CCAGC CCATG GTCTT
301 CTTCT GCATT ACGGG GCCGT CGGAG GGGAA GTTGG TGCCG CGCAG CTTCA CCTTG TAGAT
361 GAACT CGCCG TCCTG CAGGG AGGAG TCCTG GGTC ACGTC ACCAC GCCGC CGTCC TCGAA
421 GTTCA TCACG CGCTC CCACT TGAAG CCCTC GGGGA AGGAC AGCTT CAAGT AGTCG GGGAT
481 GTCGG CGGGG TGCTT CACGT AGGCC TTGGA GCCGT ACATG AACTG AGGGG ACAGG ATGTC
541 CCAGG CGAAG GGCAG GGGCG CACCC TTGGT CACCT TCAGC TTGGC GGTCT GGGTG CCCTC
601 GTAGG GCGCG CCCTC GCCCT CGCCC TCGAT CTCGA ACTCG TGGCC GTTCA CGGAG CCCTC
661 CATGT GCACC TTGAA GCGCA TGAAC TCCTT GATGA TGGCC ATGTT ATCCT CCTCG CCCTT
721 GCTCA CCATG GTATA TCTCC TTCTT AAAGT TAAAC AAAAT TATTG CTAGC TGCAA CCATT
781 ATGCG AAAAG AAAGT CCGTA TGTCA ACATT TAGAC GATCT TTTAT GCTGT AAATT CAATT
841 TGCCA TGATG TTTT ATCTG AGTAC CCTCT TGCTA TTAGT GTTT GTTCT AGTTT AATTT
901 GTTT TGTGG GTTAA AAGAT CGTTT AAATC TTGAC ATACA ACATA AAAAA CTAA TAATG
961 GTTGC AGGAT CCAAT AATTT TGTTT AACTT TAAGA AGGAG ATATA CCATG GAGCT TTTC A
1021 CTGGC GTTGT TCCCA TCCTG GTCGA GCTGG ACGGC GACGT AAACG GCCAC AAGTT CAGCG
1081 TGTC GCGA GGGCG AGGC GATGC CACCT ACGGC AAGCT GACCC TGAAG TTCAT CTGCA
1141 CCACC GCGAA CCGT CCGT CCGT CCGT CCGT CCGT CCGT CCGT CCGT CCGT CCGT CCGT
1201 AGTGC TTCAG CCGCT ACCCC GACCA CATGA AGCAG CACGA CTTCT TCAAG TCCGC CATGC
1261 CCGAA GGCTA CGTCC AGGAG CGCAC CATCT TCTTC AAGGA CGACG GCAAC TACAA GACCC
1321 GCGCC GAGGT GAAGT TCGAG GCGA CACCC TGGTG AACCG CATCG AGCTG AAGGG CATCG
1381 ACTTC AAGGA GGACG GCAAC ATCCT GGGGC ACAAG CTGGA GTACA ACTAC AACAG CCACA
1441 ACGTC TATAT CATGG CCGAC AAGCA GAAGA ACGGC ATCAA GGTGA ACTTC AAGAT CCGCC
1501 ACAAC ATCGA GGACG GCAGC GTGCA GCTCG CCGAC CACTA CCAGC AGAAC ACCCC CATCG
1561 GCGAC GGCCC CGTGC TGCTG CCCGA CAACC ACTAC CTGAG CACCC AGTCC GCCCT GAGCA
1621 AAGAC CCCAA CGAGA AGCGG GATCA CATGG TCCTG CTGGA GTTCG TGACC GCCGC CGGGA
1681 TCTAA AAGCT TCCAT GG

```

Figure S3. DNA sequence of the linear synthetic pap regulatory DNA construct pap(1,1). The DNA sequence of pap(0,1) and pap(1,0) differ by the single mutation of the base adenine to cytosine at position 817 or 919 respectively according to the idea of Braaten *et al.*². Red: *mcherry* gene. Yellow

(732–764): UTR1. Blue (778–782): -10 element. Grey (783–800): Lrp BS 4. Turquoise (801–806): -35 element. Grey (813–830): Lrp BS 5. Grey (844–861): Lrp BS 6. Grey (876–893): Lrp BS 1. Grey (907–924): Lrp BS 2. Turquoise (931–936): -35 element. Grey (937–954): Lrp BS 3. Blue (955–959): -10 element. Yellow (973–1005): UTR1. Green (1008–1685): *degfp* gene. GATC^{dist}: 816–819. GATC^{prox}: 918–921. Further GATC sites: 232–235, 628–631, 968–971, 1493–1496, 1641–1644, 1679–1682. The DNA sequence of the synthetic pap regulatory region (778–959) corresponds the native sequence (GenBank: X14471.1, based on Blyn *et al.*³) except the insertion of the -35 and -10 elements flanking the Lrp binding sites 4 and 3 (778-782: ATTAT instead of TTTCT. 801-806: TGTC AA instead of AAAATT. 931–936: TTGACA instead of AATATT. 955–959: ATAAT instead of ATTTA). All six Lrp BS according to Hernday *et al.*¹. Details related to transcription-translation-regulatory and coding elements can be found in paragraph 3.2.

0.4 Production of the partial fragments A1 and B1 for assembly of the synthetic pap regulatory region pap(1,1) via Gibson assembly reaction

We used the linear synthetic pap regulatory DNA with the sequence shown in Figure S3 as PCR template in order to produce the two alternative partial fragments A1 and B1. The forward primer with the DNA sequence 5'-CGCTATGAATTCCTCGAGTT-3' and reverse primer with the DNA sequence 5'-AGATAAAAACATCATGGCAAATTG-3' were used for the synthesis of A1 in a standard PCR. The forward primer with the DNA sequence 5'-TTTGCCATGATGTTTTTATCTGA -3' and reverse primer with the DNA sequence 5'-CCATGGAAGCTTTTAGATCC -3' were used for the synthesis of B1 in a standard PCR. Assembling these two partial fragments via Gibson assembly reaction as described in paragraph 3.3.1 and subsequent PCR amplification, led again to the regulatory pap DNA construct pap (1,1).

0.5 Kinetic fluorescence measurements. Methylation-dependent regulation of the synthetic pap regulatory DNA without cell-free synthesis of Lrp and PapI

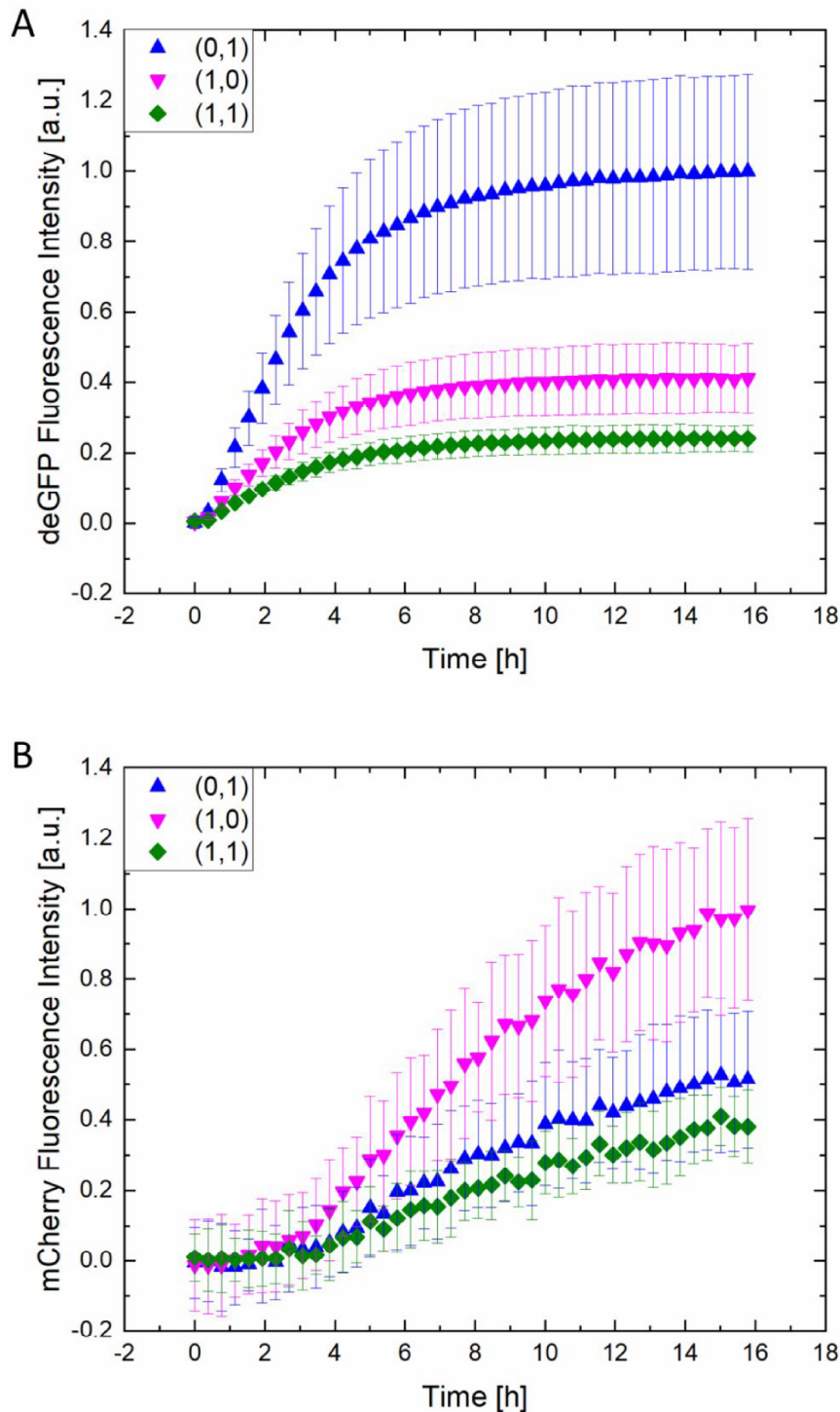


Figure S4. Kinetic measurements of the fluorescence intensities of deGFP (A) and mCherry (B) as a function of time for the different methylation states of the synthetic pap regulatory DNA construct (5 nM). Symbols refer to the different methylation states (see insert). (0,1): Lrp BS 5 unmethylated, Lrp BS 2 methylated. (1,0): Lrp BS 5 methylated, Lrp BS 2 unmethylated. (1,1): Lrp BS 2 and 5 are both methylated. Fluorescence intensities were normalized to the highest occurring value. Error bars represent one standard deviation from the mean as gained from five measurements. To simplify reading, in each kinetic measurement only every seventh data point is visualized.

0.6 Results from endpoint fluorescence measurements in the presence of Lrp and PapI plasmids.

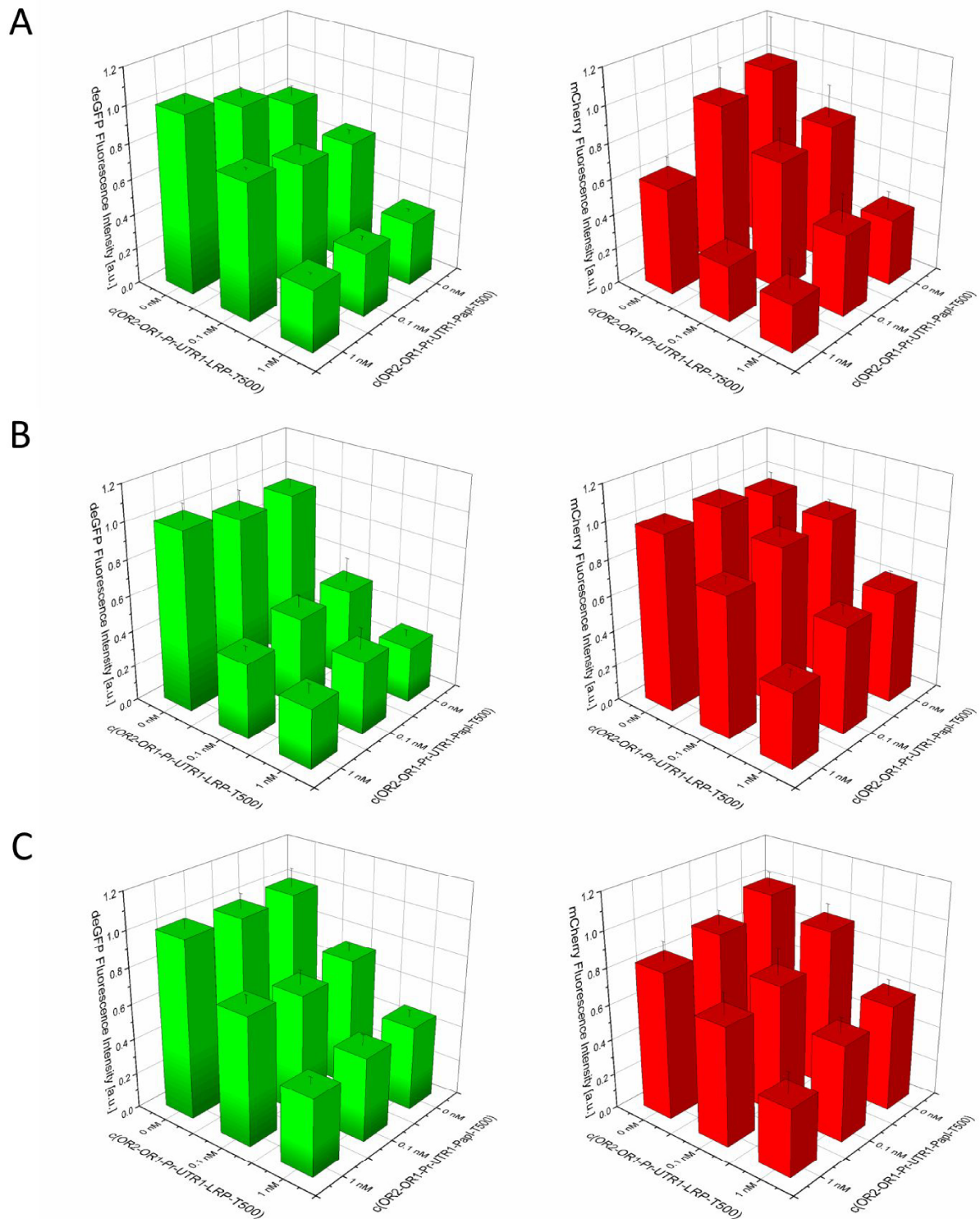


Figure S5. Fluorescence intensities (vertical axis) of deGFP (green bars), mCherry (red bars) in the presence of our synthetic pap regulatory DNA construct (5 nM) with methylation states (0,1) in (A), (1,0) in (B) and (1,1) in (C) as a function of *lrp*- and *papI*-coding plasmid concentrations on the horizontal axes. (0,1), (1,0) and (1,1) refer to the different methylation states of the DNA construct that were present in the cell-free reactions. (0,1): Lrp BS 5 unmethylated, Lrp BS 2 methylated. (1,0): Lrp BS 5 methylated, Lrp BS 2 unmethylated. (1,1): Lrp BS 2 and 5 are both methylated. Fluorescence

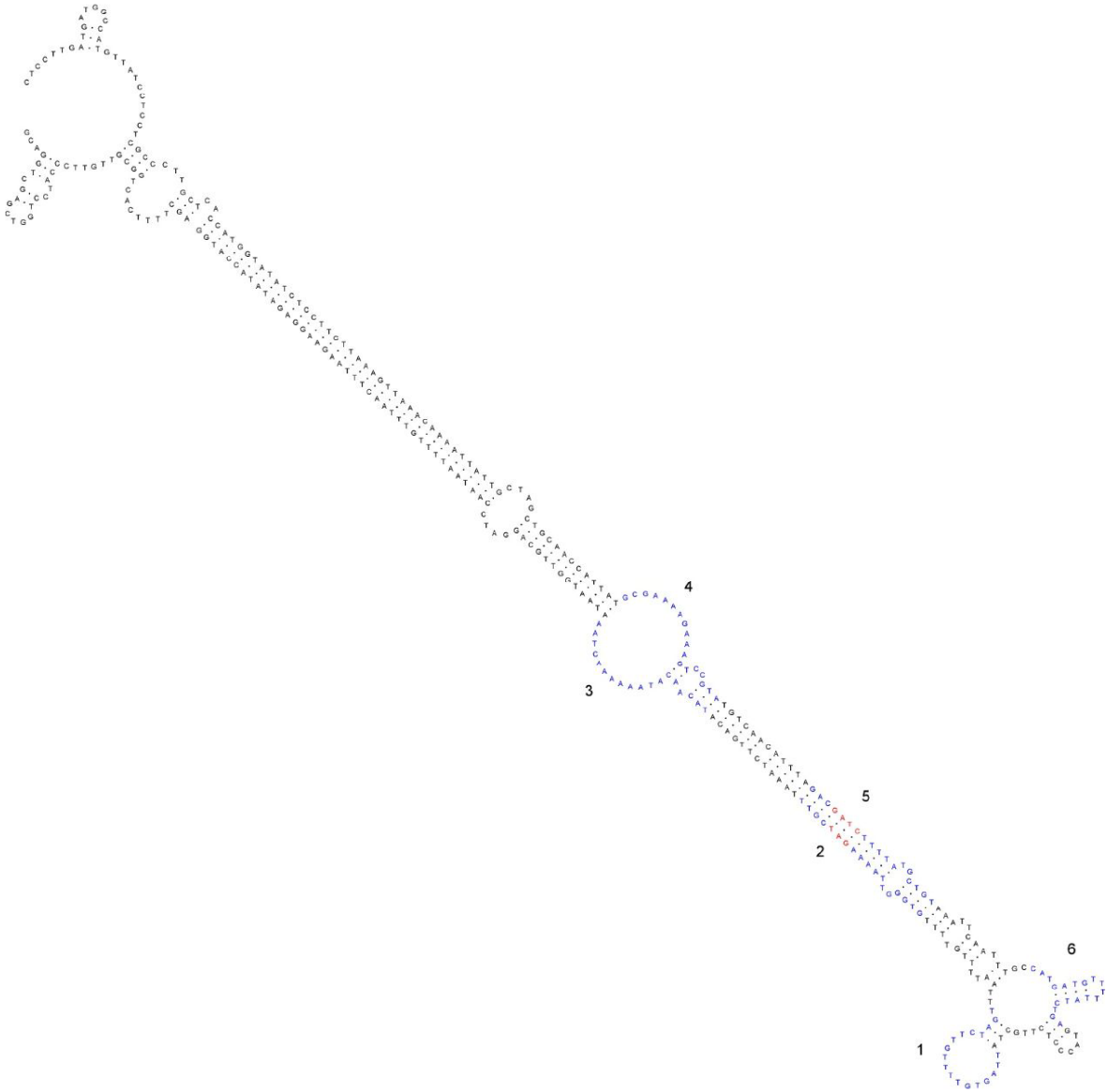
S5

3 Originalpublikationen

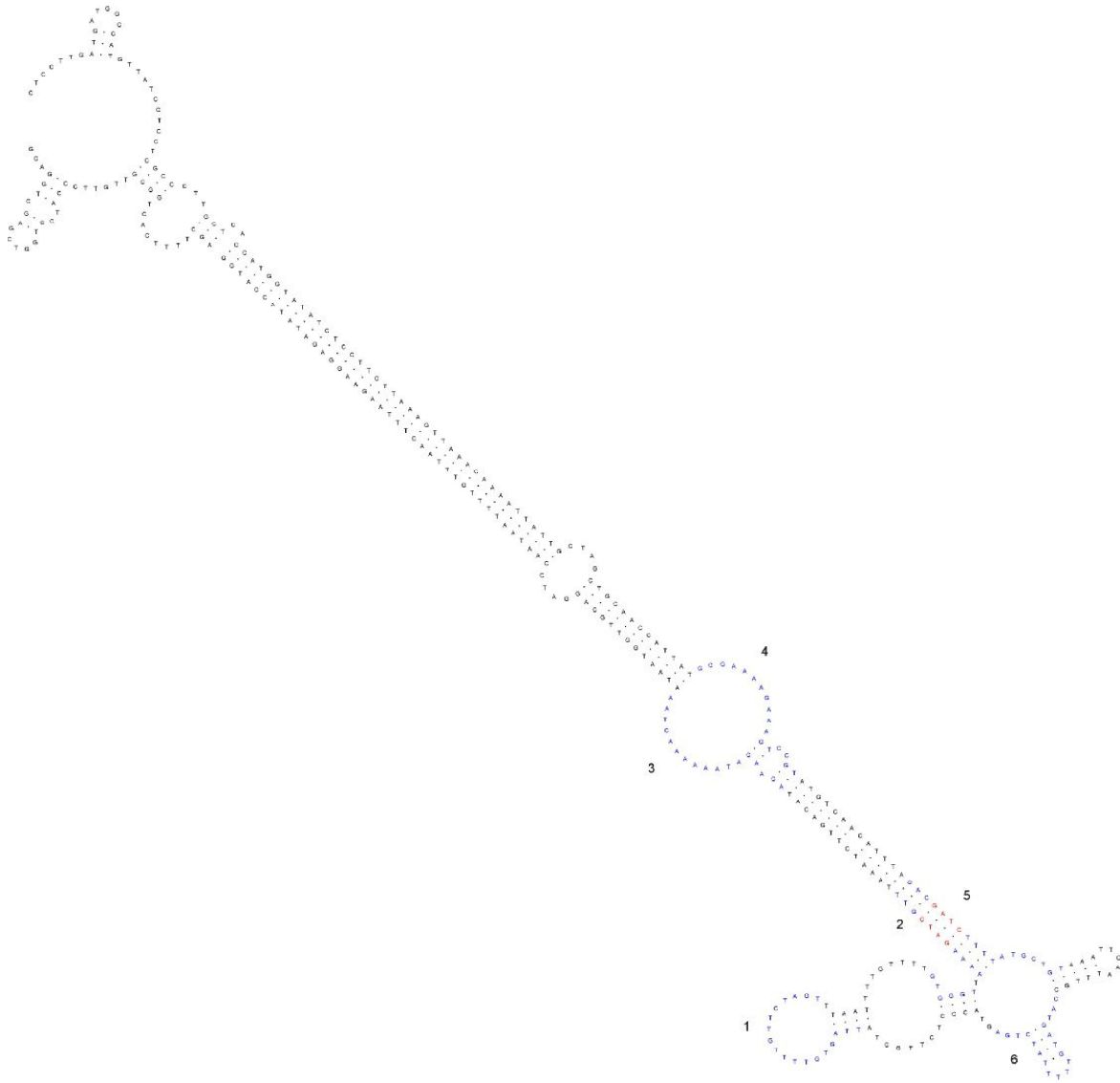
intensities were normalized to the highest occurring value. Error bars represent one standard deviation from the mean gained from two measurements.

0.7 Secondary structures of the synthetic pap regulatory DNA

A



B



C

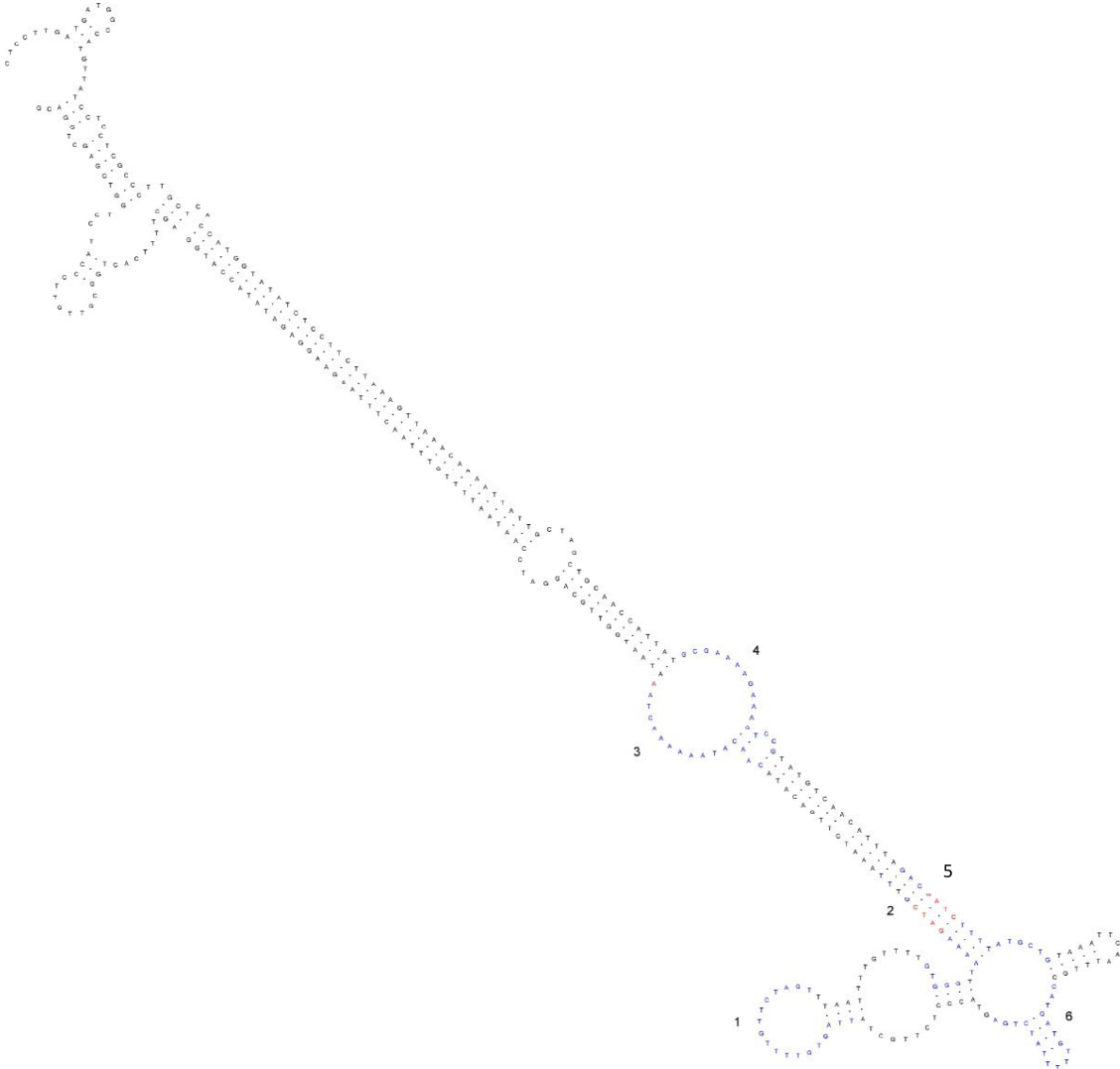


Figure S6. Three most important secondary structures of synthetic pap regulatory DNA as predicted by the mfold web server⁴ (default settings, 29 °C, see paragraph 3.6). The six Lrp BS are highlighted in colour. The figure was generated with the program Rna Viz⁵.

0.8 DNA sequence of the fragments used for investigation of the secondary structure

(I)

1 CGTAT TGCTA GCTGC AACCA TTATG CGAAA AGAAA GTCCG TATGT CAACA TTAG ACGCT
 61 CTTTT ATGCT GTAAA TTCAA TTTGC CATGA TGTTT TTATC TGAGT ACCCT CTTGC TA

(II)

1 TAGT GTTTT GTTCT AGTTT AATTT TGTTT TGTGG GTTAA AAGAT CGTT AAATC TTGAC
 61 ATACA ACATA AAAAA CTAAA TAATG GTTGC AGGAT CCAAT CGC

(III)

1 GCGAT TGGAT CCTGC AACCA TTATT TAGTT TTTTA TGTTG TATGT CAAGA TTTAA ACGAT
 61 CTTTT AACCC ACAA ACAA ATTAA ACTAG AACAA AACAC TAATA GCAAG AGGGT ACT

(IV)

1 CAGAT AAAAA CATCA TGGCA AATTG AATTT ACAGC ATAAA AGAGC GTCTA AATGT TGACA
 61 TACGG ACTTT CTTTT CGCAT AATGG TTGCA GCTAG CAATA CG

(V)

1 CGTAT TGCTA GCTGC AACCA TTATG CGAAA AGAAA GTCCG TATGT CAACA TTTAG ACGCT
 61 CTTTT ATGCT GTAAA TTCAA TTTGC CATGA TGTTT TTATC TGAGT ACCCT CTTGC ATT

(VI)

1 AGTGT TTTGT TCTAG TTTAA TTTTG TTTTG TGGGT TAAAA GATCG TTTAA ATCTT GACAT
 61 ACAAC ATAAA AAAC TTTAA ATGGT TGCAG GATCC AATCG C

(VII)

1 GCGAT TGGAT CCTGC AACCA TTATT TAGTT TTTTA TGTTG TATGT CAAGA TTTAA ACGAT
 61 CTTTT AACCC ACAA ACAA ATTAA ACTAG AACAA AACAC TAA GCAAG AGGGT ACT

(VIII)

1 CGTAT TGCTA GCTGC AACCA TTATG CGAAA AGAAA GTCCG TATGT CAACA TTAG ACGAT
 61 CTTTT ATGCT GTAAA TTCAA TTTGC CATGA TGTTT TTATC TGAGT ACCCT CTTGC TA

(IX)

1 TAGT GTTTT GTTCT AGTTT AATTT TGTTT TGTGG GTTAA AAGCT CGTT AAATC TTGAC
 61 ATACA ACATA AAAAA CTAAA TAATG GTTGC AGGAT CCAAT CGC

(X)

1 GCGAT TGGAT CCTGC AACCA TTATT TAGTT TTTTA TGTTG TATGT CAAGA TTTAA ACGAG
 61 CTTTT AACCC ACAA ACAA ATTAA ACTAG AACAA AACAC TAATA GCAAG AGGGT ACT

(XI)

1 CAGAT AAAAA CATCA TGGCA AATTG AATTT ACAGC ATAAA AGATC GTCTA AATGT TGACA
 61 TACGG ACTTT CTTTT CGCAT AATGG TTGCA GCTAG CAATA CG

(XII)

1 CGTAT TGCTA GCTGC AACCA TTATG CGAAA AGAAA GTCCG TATGT CAACA TTTAG ACGAT
 61 CTTTT ATGCT GTAAA TTCAA TTTGC CATGA TGTTT TTATC TGAGT ACCCT CTTGC ATT

(XIII)

1 AGTGT TTTGT TCTAG TTTAA TTTTG TTTTG TGGGT TAAAA GCTCG TTTAA ATCTT GACAT
 61 ACAAC ATAAA AAAC TTTAA ATGGT TGCAG GATCC AATCG C

3 Originalpublikationen

(XIV)

```
1   GCGAT TGGAT CCTGC AACCA TTATT TAGTT TTTTA TGTTG TATGT CAAGA TTTAA ACGAG  
61  CTTTT AACCC ACAAA ACAAA ATTAA ACTAG AACAA AACAC TAA6A GCAAG AGGGT ACT
```

Figure S7. DNA sequences of the labeled single stranded DNA oligos that were used to obtain double stranded DNA constructs for quenching experiments. Constructs with the 3 methylation states (0,1), (1,0) and (1,1) were obtained that show either fluorescence if the DNA exhibits a linear structure (LinearQuenchDNA), or if the DNA is present in the predicted cruciform conformation (CruciformQuenchDNA). All of them originate from four single stranded DNA oligos. LinearQuenchDNA(0,1): (V), (VI), (VII) and (IV); LinearQuenchDNA(1,0): (XII), (XIII), (XIV) and (XI); LinearQuenchDNA(1,1): (XII), (VI), (VII) and (XI); CruciformQuenchDNA(0,1): (I), (II), (III) and (IV); CruciformQuenchDNA(1,0): (VIII), (IX), (X) and (XI); CruciformQuenchDNA(1,1): (VIII), (II), (III) and (XI). The base labeled with 6-Fam is marked in green and the base labeled with BHQ1 is marked in red. The yellow mark indicates a phosphate group at the 5'-end that is necessary for ligation.

0.9 Comparison of 6-Fam quenching for methylation states (0,1) and (1,0)

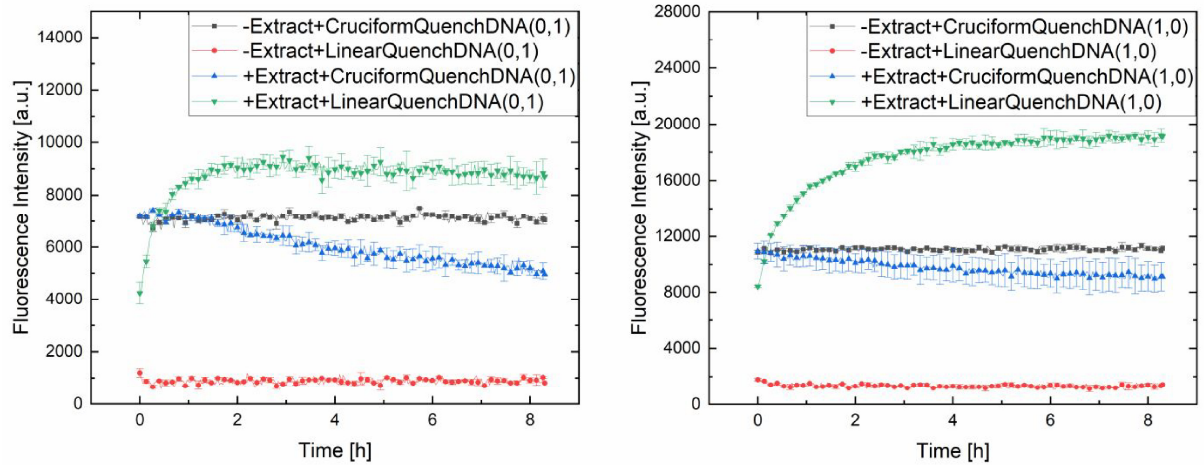


Figure S8. Time course of the fluorescence emission intensity of 6-Fam/BHQ1 labelled DNA constructs for the methylation states (0,1) (left) and (1,0) (right) after introduction into the expression system ($c = 120$ nM). The positions of 6-Fam and the corresponding quencher BHQ1 were chosen to show quenching either for the linear conformation (LinearQuenchDNA) or for the predicted DNA cruciform conformation (CruciformQuenchDNA). Experiments were performed in the presence of crude extract and in its absence. In the absence of crude extract the DNA remains linear. In its presence the linear conformation is lost to approach the cruciform conformation. For comparison, the fluorescence intensity of 6-Fam of the LinearQuenchDNA in the absence of BHQ1 is $12,111 \pm 487.9$ a.u. ($\text{dsDNA}_{\text{Ann,Linear}(0,1)}$, $c = 120$ nM) or $21,799.5 \pm 395.3$ a.u. ($\text{dsDNA}_{\text{Ann,Linear}(1,0)}$, $c = 120$ nM). The fluorescence intensity of 6-Fam of the CruciformQuenchDNA in the absence of BHQ1 is $6,831 \pm 446.9$ a.u. ($\text{dsDNA}_{\text{Ann,Cruciform}(0,1)}$, $c = 120$ nM) or $10,957 \pm 246.1$ a.u. ($\text{dsDNA}_{\text{Ann,Cruciform}(1,0)}$, $c = 120$ nM). The differences in the absolute fluorescence intensities for the different methylation states result from different fluorescence intensities of the oligos as received from the manufacturer.

0.10 A combination of our synthetic pap regulatory region and magnetic tweezers

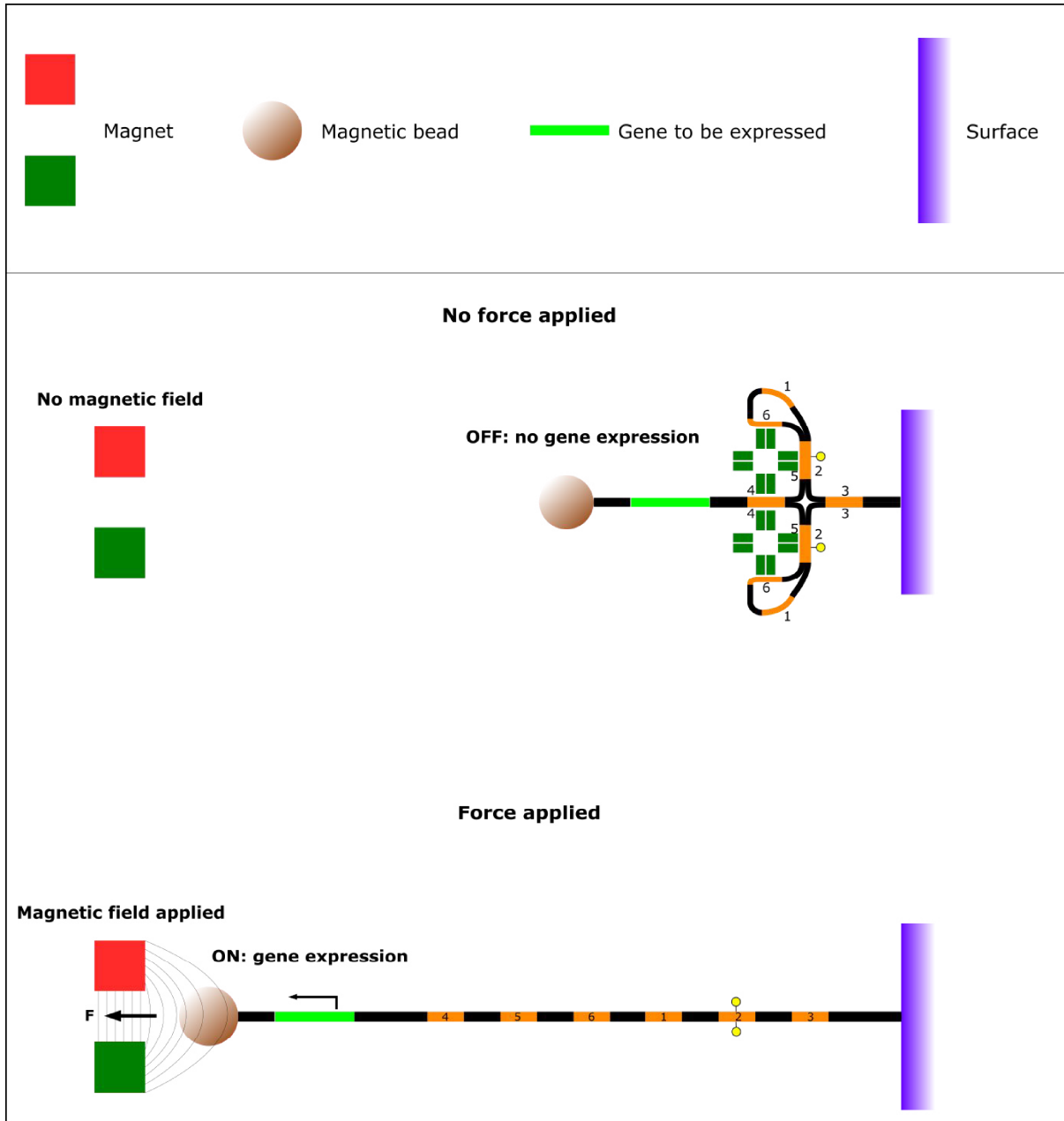


Figure S9. Force dependent gene expression at ultra-low forces. We exemplify the force application through magnetic tweezers. Lrp stabilizes a cruciform conformation of our synthetic pap regulatory DNA in methylation state (0,1). This results in suppressed gene expression. If one end of the DNA is bound to a surface and the other end is bound to a magnetic bead, the force of a magnetic tweezer will linearize the construct. The linear conformation enables gene expression.

REFERENCES

- (1) Hernday, A. D., Braaten, B. A., and Low, D. A. (2003) The Mechanism by Which DNA Adenine Methylase and PapI Activate the Pap Epigenetic Switch. *Mol. Cell* 12, 947–957. [https://doi.org/10.1016/S1097-2765\(03\)00383-6](https://doi.org/10.1016/S1097-2765(03)00383-6).
- (2) Braaten, B. A., Nou, X., Kaltenbach, L. S., and Low, D. A. (1994) Methylation Patterns in Pap Regulatory DNA Control Pyelonephritis-Associated Pili Phase Variation in E. Coli. *Cell* 76, 577–588. [https://doi.org/10.1016/0092-8674\(94\)90120-1](https://doi.org/10.1016/0092-8674(94)90120-1).
- (3) Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H., and Low, D. A. (1989) Phase-Variation of Pyelonephritis-Associated Pili in Escherichia Coli: Evidence for Transcriptional Regulation. *EMBO J.* 8, 613–620. <https://doi.org/10.1002/j.1460-2075.1989.tb03416.x>.
- (4) Zuker, M. (2003) Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* 31, 3406–3415. <https://doi.org/10.1093/nar/gkg595>.
- (5) De Rijk, P., Wuyts, J., and De Wachter, R. (2003) RnaViz 2: An Improved Representation of RNA Secondary Structure. *Bioinformatics* 19, 299–300. <https://doi.org/10.1093/bioinformatics/19.2.299>.

3.3 Finkler & Ott (2019)

Bead-based assay for spatiotemporal gene expression control in cell-free transcription–translation systems.

Finkler, M., and Ott, A.

BioTechniques (2019), 66, 29–33.

DOI: 10.2144/btn-2018-0097

<https://doi.org/10.2144/btn-2018-0097>

Reproduced from *BioTechniques*. (2019) 66, 29-33 with permission of Future Science Group.

Bead-based assay for spatiotemporal gene expression control in cell-free transcription–translation systems

Marc Finkler¹ & Albrecht Ott^{*1}

ABSTRACT

Cell-free gene expression has applications in synthetic biology, biotechnology and biomedicine. In this technique gene expression regulation plays an important role. Transcription factors do not completely suppress expression while other methods for expression control, for example, CRISPR/Cas, often require important biochemical modifications. Here we use an all *Escherichia coli*-based cell-free expression system and present a bead-based method to instantly start and, at a later stage, completely stop gene expression. Magnetic beads coated with DNA of the gene of interest trigger gene expression. The expression stops if we remove the bead-bound DNA as well as transcribed mRNA by hybridization to bead-bound ssDNA. Our method is a simple way to control expression duration very accurately in time and space.

METHOD SUMMARY

In our paper we present a new method that enables control of gene expression in an *Escherichia coli*-based cell-free expression system. Using paramagnetic, DNA-coated beads we introduce and remove not only the DNA of interest, but also endogenous mRNA to achieve, almost instantly, full activation or suppression of gene expression without affecting or modifying the expression system biochemically.

KEYWORDS:

cell-free expression • *Escherichia coli* cell extract • microbeads • transcription control • translation control

¹Universität des Saarlandes, Biologische Experimentalphysik, Naturwissenschaftlich-Technische Fakultät, B2 1, Campus, 66123 Saarbrücken, Germany; *Author for correspondence: albrecht.ott@physik.uni-saarland.de

BioTechniques 66: 29–33 (January 2019)
10.2144/btn-2018-0097

INTRODUCTION

The use of cell-free expression systems can avoid some of the difficulties linked to host organisms. For instance, toxic substances will kill the host [1] while, in most cases, a cell-free system will continue to perform [2]. The control of different biochemical parameters and processes is often simpler *in vitro* [3]. Many different applications of cell-free expression systems in synthetic biology, biotechnology and biomedicine require the expression of genes and their regulation or control [4–6].

A common method for transcription control is the use of inducible promoters. The presence of an inducer instantly starts transcription [7,8]. Examples of inducers include different metabolites, antibiotics and steroids [7]. Stopping transcription requires complete removal of the inducer, which is difficult without affecting the expression system. A different method to control transcription uses repressing transcription factors recognizing specific DNA sequences or specific DNA modifications such as methylation [9]. However, due to basal transcription, complete suppression cannot be achieved [9,10]. A light-dependent gene switch enables transcription as a response to visible light [7,11,12], but the problem with basal transcription remains [12]. CRISPR/Cas switches genes on and off much more efficiently than transcription factors. The usage of arbitrary crRNA mediates the direction of Cas9 to the DNA sequence of interest [8,13,14]. However, CRISPR/Cas requires biochemical modification of the expression system.

There are substances that stop transcription at the level of the RNA polymerase (Tagetin™, Rifampicin) or by mRNA cleavage (MazF) [15–18]. Suppression is efficient, but once included in the expression system, future expression is prevented.

Here we present a bead-based method that addresses transcription and translation control in an *E. coli* cell-free system. The method has no influence on the gene expression ability of the expression system. The method can be used to express genes locally, within a defined time window.

MATERIALS & METHODS

All experiments were performed using an *E. coli* cell-free system described previously [19,20]. For the preparation of DNA-coated beads, the plasmid pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 was derived from the commercially available plasmid pBest-Or2-Or1-Pr-UTR1-deGFP-T500 (Addgene Cat# 40019) by replacing the ColE1 ori with the p15a ori (to reduce background expression during plasmid amplification [20]) and adding a C-terminal His-tag DNA sequence to the *degfp* gene. The derived plasmid was used as a template for a PCR reaction using Q5-Polymerase (Q5® High-Fidelity DNA Polymerase, Neb) and the primers Biotin-TEG-Seq1 (TEG stands for TriEthylene Glycol) (5'-Biotin-TEG-CAC CAT CAG CCA GAA AAC-3', Metabion) and Seq4 (5'-GAG CTG ACT GGG TTG AAG G-3', Metabion) to amplify biotinylated Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500-Seq4 (DNA_{deGFP}). The PCR product was purified using a PureLink™ PCR Purification Kit (Invitrogen). Subsequently the product was bound to streptavidin-coated magnetic beads T1 (Dynabeads® MyOne™ Streptavidin T1, Invitrogen) following the bead manufacturer's protocol [21]. The T1-Biotin-TEG-Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500-Seq4 (bead-bound DNA_{deGFP}) stock solution was at 400 nM (DNA_{deGFP}-concentration) in water. Two neodymium cube magnets (WM-20x20x20-N, Magnet-shop, Germany) served for bead separation.

Streptavidin-coated magnetic beads T1 were coated with single-stranded biotinylated DNA complementary to ▶

Reports

► the 3'-end of the Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500-Seq4-mRNA (3') (5'-Biotin-TEG-CGG CGG GCT TTG CTC GAG TTA GTG GTG ATG GTG ATG-3', Metabion). Poly-T-coated beads (poly-T) (Dynabeads™ Oligo(dT)_{25'}, Invitrogen), complementary to the poly(A) tail of the transcribed mRNA, were used without further purification. The aqueous stock solution of the 3'-coated beads was prepared at the same bead concentration as the poly-T-coated beads (5 mg/ml).

DNA concentration was determined using a spectrophotometer (NanoDrop™ 2000c Spectrophotometer, Thermo Scientific). DeGFP was quantified by transferring 10 µl of sample to a microwell plate (Nunc™ 384-Well Optical Bottom Plates # 242764, Thermo Scientific) and determining the fluorescence intensity with a microplate reader (POLARstar OPTIMA, BMG LABTECH).

To 30 µl of crude *E. coli* cell-free extract we added: 3.6 µl magnesium glutamate (at 100 mM), 2.4 µl potassium glutamate (at 3 M), 6.43 µl PGA-buffer (3-Phosphoglyceric Acid) (14x), 4.5 µl 40% (v/v) PEG8000, 13.07 µl amino acids (mixture contains all canonical amino acids at 6 mM except Leu at 5 mM) and 3 µl GamS (Shortened lambda phage Gam protein) (at 99 µM) to reach a total of 63 µl, aliquoted in 8.4 µl (seven samples in batch). According to Shin *et al.* [20] the amount for magnesium glutamate, potassium glutamate and amino acids has to be adjusted for each stock of crude extract to optimize the cell-free protein synthesis. GamS was added to prevent degradation of linear DNA [15,22–24]. For expression we used a reaction volume of 12 µl by adding 3.6 µl of different DNA_{deGFP} solutions and double-distilled water to the 8.4 µl aliquots. Concentrations of bead-bound DNA_{deGFP} ranged from 5 nM to 120 nM. As a reference, unbound DNA_{deGFP} was added at 10 nM (this concentration leads to optimal protein yield).

All chemicals involved in sample preparation and the sample itself were put on ice until expression was started by raising the temperature. To study the optimal bead-bound DNA_{deGFP} concentration for protein yield, samples were incubated at 29°C for 16 h. Then beads were separated and the fluorescence

intensity was measured. Similar fluorescence intensities were measured when using 10 nM unbound DNA_{deGFP} (optimal for protein yield) (54132,6 ± 1680,87 a.u.) and 50 nM bead-bound DNA_{deGFP} (52551,4 ± 4329,48 a.u.). Therefore, 50 nM of bead-bound DNA_{deGFP} was used in all further expression reactions (addition of 1–5 µl of T1-Biotin-TEG-Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500-Seq4 stock solution and 2.1 µl double-distilled water to 8.4 µl aliquots).

To study transcription, expression reactions were incubated at 29°C for different time periods (T from 1 min to 16 h) followed by bead separation. The samples were again incubated at 29°C for T_{res} = 16 h-T to reach a summed incubation time of 16 h at 29°C. Subsequently the fluorescence intensity was measured (Figure 1).

To study translation, the expression reaction was incubated for 30 min at 29°C. Then the bead-bound DNA_{deGFP} was separated, and the supernatant was transferred to a vial containing freshly prepared ss-DNA-coated beads. The ss-DNA-coated beads consisted of 50 µl 3' and 50 µl poly-T-coated (complementary sequences to the deGFP-mRNA) beads at 5 mg/ml, prepared according to specifications from the manufacturer [25]. The mixture was incubated on a shaker for 15 min at room temperature at 230 rpm. Subsequently the ssDNA-coated beads were removed and the time course of fluorescence intensity was recorded by a microplate reader to visualize expression kinetics (Figure 2). To restart expression, 10 nM unbound DNA_{deGFP} was added to the sample immediately after ssDNA-coated bead separation, and the expression kinetics was visualized. As a reference a measurement of the kinetics without mRNA removal was performed (Figure 2).

RESULTS & DISCUSSION

Both experiments are illustrated in Figure 3.

In a first set of experiments, we focused on controlling transcription duration. The experiments were performed as illustrated in Figure 3A. Figure 1 shows the measured fluorescence intensity after

16 h of incubation for different periods T of initial exposure to bead-bound DNA_{deGFP}. During the incubation period the time course of fluorescence intensity corresponds to the black squares in Figure 2. Data in Figure 1 give the end points of this type of measurement.

Figure 1 shows that 30 min of DNA_{deGFP} in contact with a cell-free expression system leads to optimal protein yield. For only 1 min of exposure we can already achieve approximately a fifth of the optimal yield.

In a second set of experiments, we focused on controlling the duration of translation. The experiments were performed as illustrated in Figure 3B. Translation was stopped by adding and retrieving bead-bound ssDNA with a sequence that is complementary to the produced mRNA (Figure 2, red circles and black squares). The blue triangles in Figure 2 show the kinetics after adding fresh, unbound DNA_{deGFP} to the sample immediately after mRNA removal.

Figure 2 shows that adding beads coated with ss-DNA complementary to the transcribed mRNA (3' and poly-T) and their subsequent removal suppress further translation without affecting the ability of the cell-free system to perform further gene expressions.

We have shown that in our *E. coli* cell-free expression system, gene expression occurs if linear, bead-bound DNA_{deGFP} is added. For other expression systems the usage of bead-bound DNA as the DNA source has already been reported [26]. The necessity to use a fivefold concentration of bead-bound DNA_{deGFP} compared with unbound DNA_{deGFP} in bulk may be explained by lower accessibility of the bead-bound DNA_{deGFP}. The biotin-streptavidin bond is known to be very strong [27,28]. We do not expect DNA_{deGFP} to leak from beads since during a few weeks at 4°C no decay with bead storage in stock solution could be detected. Moreover, reusing beads with different exposure times to cell-free reactions as the DNA_{deGFP} source in new expression experiments produced the same deGFP expression level (Supplementary Figure S1). The fact that mRNA removal stops expression confirms that DNA_{deGFP} leaking from beads into the cell-free reaction is negligible. Only the

addition of fresh DNA again triggers protein expression (Figure 2).

We observed that the presence of DNA_{deGFP} during $T = 30$ min leads to optimal protein yield (Figure 1). Shorter exposure times result in lower yields, explained by smaller amounts of transcribed mRNA. A small decrease in protein yield for longer exposure times may well be due to sharing effects of translation and transcription processes. DeGFP fluorescence quenching was never observed with our experiments at comparable concentrations (Supplementary Figure S2). The relatively high protein production observed for $T = 1$ min and the fact that $T \geq 20$ min yields a comparable level of deGFP with $T = 16$ h, suggest that mRNA is not a limiting factor for protein production as long as $T \geq 20$ min.

The long lifetime of mRNA in the cell-free reaction is a likely cause for the minor increase of fluorescence intensity even after removal of DNA_{deGFP}. In principle, a delay in protein maturation could also cause an increase (Figure 2, black squares). However, maturation of deGFP requires 8 to 16.2 min [15,16], too short to explain the observed increase of fluorescence intensity over several hours. If mRNA lifetime was a limiting factor, the deGFP-mRNA mean lifetime could be calculated by fitting the time-dependent production of matured deGFP from the moment where transcription is stopped (Figure 2, black squares) with a mathematical expression that describes mRNA inactivation as exponentially decaying [15,16,29]. Doing so results in a mean mRNA lifetime of about 124 min. However, a DNA_{deGFP} exposure time of 30 min leads to a similar protein yield as exposure of over 16 h (Figure 1). We conclude that fatigue of our expression system, rather than mRNA degradation, is the main cause for saturation of deGFP production. Accordingly 124 min must be understood as a lower bound of the deGFP-mRNA lifetime. Previously the lifetime of mRNA was claimed to be short (*E. coli*: 6.8 min; *E. coli* cell-free system: about 13 min [3,16,29]). However, mRNA stability depends on the particular cell-free system as well as the mRNA itself [16,18,30]. A possible reason for the difference could lie in the presence

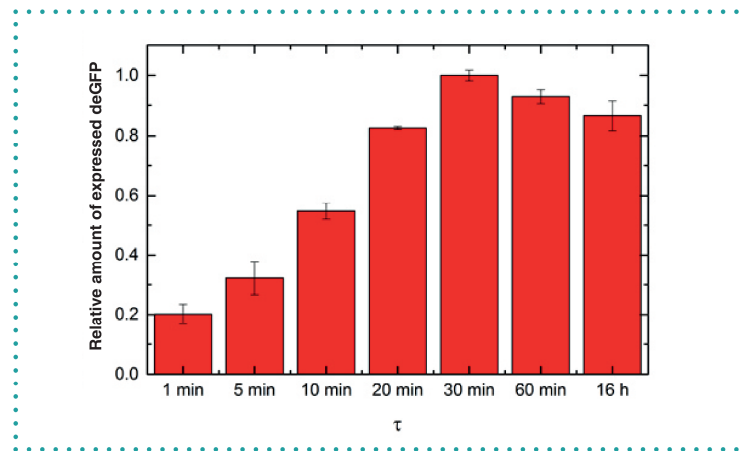


Figure 1. Relative amounts (normalized to the maximum) of expressed deGFP in a cell-free expression reaction after 16 h of incubation as a function of the time T , the time interval that the cell-free system was in contact with bead-bound DNA_{deGFP}. A DNA_{deGFP} exposure time of 30 min causes the highest level of deGFP production.

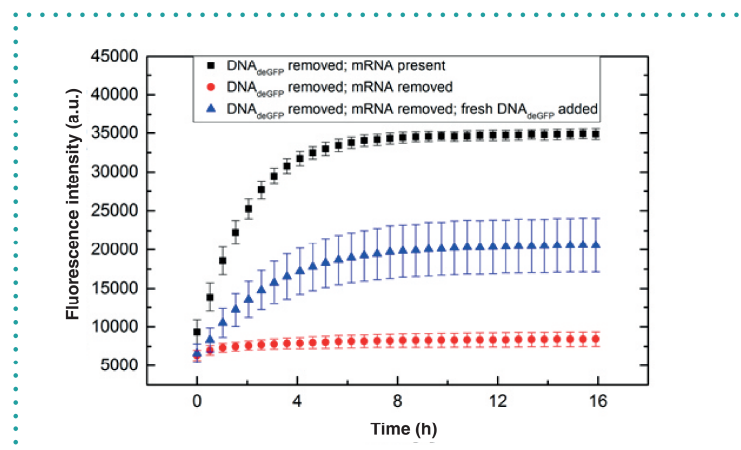


Figure 2. Time course of fluorescence intensity from deGFP as produced in a cell-free expression reaction. The samples were in contact with bead-bound DNA_{deGFP} (during $T = 30$ min, not shown). After DNA_{deGFP} removal the expression kinetics was monitored (i) without further modification of the expression system (black squares), (ii) after mRNA removal (red circles) and (iii) after mRNA removal and subsequent addition of fresh DNA (blue triangles). For mRNA removal ss-DNA-coated beads (3' and poly-T, each 50 μ l) were added and removed 15 min later (followed by the start of recording). We address the initial increase in (ii) to deGFP maturing and a small amount of residual mRNA. Adding fresh, unbound DNA_{deGFP} to the expression reaction restarted expression. To maintain overall concentration (iii) required pipetting of small volumes (about 0.1 μ l) of highly concentrated DNA_{deGFP} causing increased error.

of GamS in our case, which may affect mRNA degradation. Moreover, different methods to stop transcription will affect the expression system differently.

Figure 2 shows that treating mRNA containing cell-free systems (no DNA_{deGFP}) for 15 min with bead-bound ss-DNA complementary to the mRNA sequence

stops protein production and, by extension, translation. We suggest that the formation of a DNA-mRNA hybrid [31] removes free mRNA almost completely. The small increase in fluorescence intensity during the first minutes (Figure 2) can be explained by expressed deGFP that matured within this time frame ►

Reports

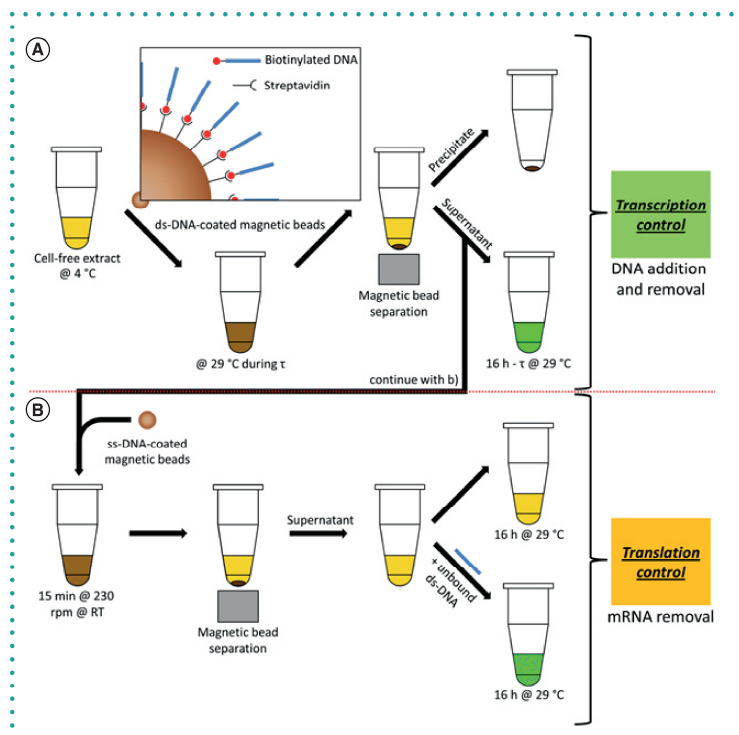


Figure 3. Scheme of experiments addressing control of transcription as well as translation. (A) For transcription control, DNA_{deGFP}-coated magnetic beads are added to a cell-free reaction followed by incubation at 29°C during the time interval *T*. Bead-bound DNA_{deGFP} is then removed by magnetic separation. The supernatant is incubated for 16 h-T at 29°C. During this time interval the amount of expressed protein is monitored (Figure 2 black squares, end points Figure 1). (B) To stop translation, magnetic beads coated with ss-DNA complementary to the transcribed mRNA are added immediately after DNA removal. Then the sample is mixed for 15 min on a shaker (230 rpm, room temperature) followed by magnetic separation of the beads. Expression is suppressed (Figure 2, red circles). Further expression reactions can be performed again by addition of fresh DNA (Figure 2, blue triangles).

► in conjunction with a small amount of remaining mRNA. The fact that gene expression restarts upon adding fresh DNA_{deGFP} shows that the expression system is not affected by bead treatment.

A possible application of the latter is the removal of residual canonical amino acids within the cell-free system for the incorporation of noncanonical amino acids. If the canonical amino acid is still present within the system, it will be preferably incorporated instead of the noncanonical one [32]. Using the method proposed here, a first expression can be performed to incorporate the residual canonical amino acid into a dummy protein, followed by addition of the corre-

sponding noncanonical amino acid in a second expression reaction.

The method enables spatially and temporally separated expression of several, different genes: beads coated with different DNA can be successively added and removed from the expression reaction. Simultaneous co-expression of several genes in cell-free systems results in competition for the expression machinery, making the control of individual expression rates difficult; a situation that our method can help to avoid. In principle, any cell-free system should remain unaltered following the protocol presented here. Our method can be used for any linear, biotinylated DNA-construct.

In this paper, we demonstrated a new and simple method for controlling the expression duration within a cell-free expression system. The method acts on two different levels, transcription and translation, without biochemically altering the expression system. DNA-coated beads are used to provide and remove the DNA of the gene of interest, or to remove the corresponding mRNA. We achieve good temporal control. Spatial control of gene expression through the arrangement of magnetic beads can also be useful, for example, in microfluidic experiments.

FUTURE PERSPECTIVE

The proposed method regulates gene expression by nothing but spatiotemporal separation of the nucleic acids without usage of any other biochemical substances. This way, the gene expression system remains chemically unaltered while controlling gene expression very accurately. Gene expression can be triggered in a defined location by applying magnetic fields on paramagnetic DNA-coated beads. This can be of use in microfluidics. Another field of application of the method proposed here is the incorporation of noncanonical amino acids into proteins by including them into the expression system. Proteins with noncanonical analogs may develop new functions in the organism. In cell-free extracts, however, there are residual amino acids that suppress the simple incorporation of their noncanonical analogs. With our technique we are able, in a first expression reaction, to incorporate the canonical amino acids into a dummy protein followed by a second expression reaction using noncanonical analogs. Furthermore, our method may be an interesting tool in the field of nanomaterials. Bead-bound DNA caged within a material that can release the DNA in response to an outer stimulus may be a way to enable protein synthesis as response to a stimulus.

AUTHOR CONTRIBUTIONS

AO and MF designed the research, MF performed the research and analyzed data with the help of AO. MF and AO wrote the paper.

FINANCIAL DISCLOSURE

This work was supported by the Collaborative Research Center SFB 1027 funded by Deutsche Forschungsgemeinschaft (DFG) and by the Human Frontier Science Program (HFSP, RGP0037/2015). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

ACKNOWLEDGMENTS

We thank Emanuel Worst and Vincent Noireaux for help and fruitful discussions. We thank Mina Mohammadi-Kambs for fitting our data to estimate the deGFP-mRNA mean lifetime.

OPEN ACCESS

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

SUPPLEMENTARY DATA

To view the supplementary data that accompany this article please visit the journal website at www.future-science.com/doi/suppl/10.2144/btn-2018-0097

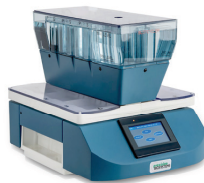
REFERENCES

- Schachtele CF, Rogers P. Canavanine death in *Escherichia coli*. *J. Mol. Biol.* 14, 474–489 (1965).
- Worst EG, Exner MP, DeSimone A *et al.* Cell-free expression with the toxic amino acid canavanine. *Bioorg. Med. Chem. Lett.* 25, 3658–3660 (2015).
- Shin J, Noireaux V. An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* 1, 29–41 (2012).
- Weber W, Fussenegger M. Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.* 13, 21–35 (2012).
- Michalodimitrakis K, Isalan M. Engineering prokaryotic gene circuits. *FEMS Microbiol. Rev.* 33, 27–37 (2009).
- Endy D. Foundations for engineering biology. *Nature* 438, 449–453 (2005).
- Polstein LR, Gersbach CA. Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. *J. Am. Chem. Soc.* 134, 16480–16483 (2012).
- Bikard D, Jiang W, Samai P *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429–7437 (2013).
- Schenkelberger M, Shanak S, Finkler M *et al.* Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system. *Phys. Biol.* 14(2), 026002 (2017).
- Kaludov NK, Wolffe AP. MeCP2 driven transcriptional repression *in vitro*: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *Nucleic Acids Res.* 28, 1921–1928 (2000).
- Pinheiro A, Baptista P, Lima JC. Light activation of transcription: photocaging of nucleotides for control over RNA polymerization. *Nucleic Acids Res.* 36, 1–7 (2008).
- Shimizu-Sato S, Huq E, Tepperman JM, Quail PH. A light-switchable gene promoter system. *Nat. Biotechnol.* 20, 1041–1044 (2002).
- Polstein LR, Gersbach CA. A light-inducible CRISPR/Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* 11, 198–200 (2015).
- Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355 (2014).
- Garamella J, Marshall R, Rustad M, Noireaux V. The all *E. coli* TX-TL toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth. Biol.* 5, 344–355 (2016).
- Shin J, Noireaux V. Study of messenger RNA inactivation and protein degradation in an *Escherichia coli* cell-free expression system. *J. Biol. Eng.* 4, 1–9 (2010).
- Sarkar N. Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* 66, 173–197 (1997).
- Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA* 99, 9697–9702 (2002).
- Sun ZZ, Hayes CA, Shin J *et al.* Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* 79, 1–14 (2013).
- Shin J, Noireaux V. Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4, 8 (2010).
- Invitrogen. *Manual for Dynabeads® MyOne™ Streptavidin T1.* (2017).
- Takahashi MK, Hayes CA, Chapell J *et al.* Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods* 86, 60–72 (2015).
- Sitaraman K, Esposito D, Klarmann G *et al.* A novel cell-free protein synthesis system. *J. Biotechnol.* 110, 257–263 (2004).
- Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* 3, 387–397 (2014).
- Ambion. *Manual for Dynabeads® Oligo (dT)25.* (2012).
- Nord O, Uhlén M, Nygren PÅ. Microbead display of proteins by cell-free expression of anchored DNA. *J. Biotechnol.* 106, 1–13 (2003).
- Weber P, Ohlendorf D, Wendoloski J, Salem F. Structural origins of high-affinity biotin binding to streptavidin. *Science* 243, 85–88 (1989).
- Chivers CE, Koner AL, Lowe ED, Howarth M. How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimeric tetramer. *Biochem. J.* 435, 55–63 (2011).
- Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* 13, 216–223 (2003).
- Belasco JG, Higgins CF. Mechanisms of mRNA decay in bacteria: a perspective. *Gene* 72, 15–23 (1989).
- Naiser T, Ehler O, Kaiser J, Mai T, Michel W, Ott A. Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges. *BMC Biotechnol.* 8, 1–23 (2008).
- Worst EG, Exner MP, DeSimone A, Schenkelberger M *et al.* Residue-specific incorporation of noncanonical amino acids into model proteins using an *Escherichia coli* cell-free transcription-translation system. *J. Vis. Exp.* 114, 1–14 (2016).



Blot-to-blot consistency. Time after time.

To achieve meaningful comparison of expressed proteins from blot-to-blot, trustworthy data is essential. **BlotCycler™** is designed to eliminate slight variations in fluid delivery, timing and washing that may occur when protocols are performed manually. With rigorous precision, **BlotCycler** delivers an affordable solution to busy labs demanding reproducible results.



BlotCycler™
by PRECISION BIOSYSTEMS

www.blotcycler.com

TRY OUR FREE 1-MONTH TRIAL | CALL 1-888-490-4443

Supplementary

Bead-based assay for spatiotemporal gene expression control in cell-free transcription–translation systems.

Finkler, M., and Ott, A.

BioTechniques (2019), 66, 29–33.

DOI: 10.2144/btn-2018-0097

<https://doi.org/10.2144/btn-2018-0097>

Reproduced from *BioTechniques*. (2019) 66, 29-33 with permission of Future Science Group.



Article Body Template

Supplementary

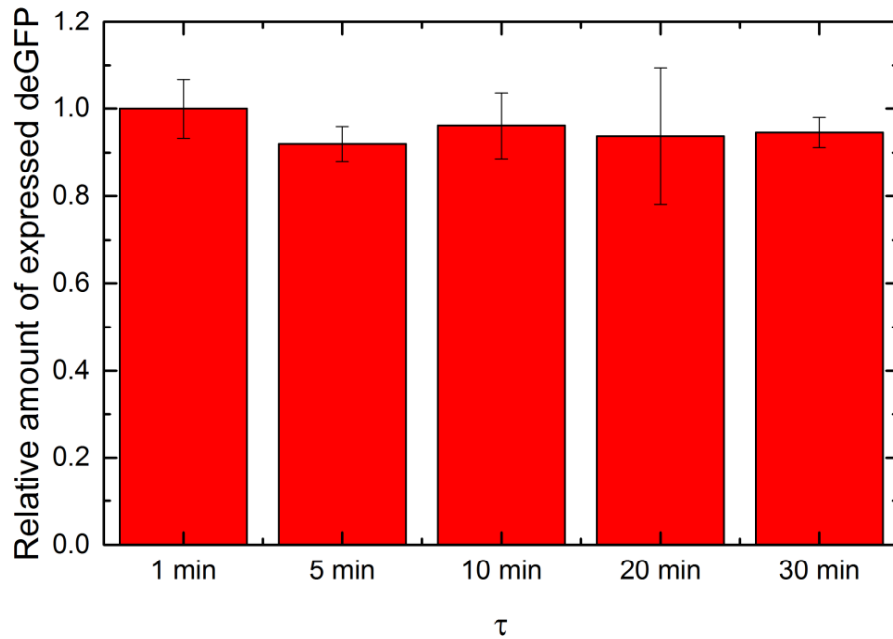


Figure S 1 Relative amounts (normalized to the observed maximum) of expressed deGFP if bead-bound DNA_{deGFP} is used for a second expression reaction. τ indicates the time the beads spent in the cell-free system during a first expression reaction. Immediately after bead separation, the beads were introduced into a second, fresh cell-free system (same volume as during the first expression reaction) and the sample was incubated for 16 hours at 29 °C. After incubation the beads were separated and the fluorescence-intensity was determined. Within the experimental error the same amount of fluorescent deGFP was recorded. DNA_{deGFP} lost during the first expression reaction would lead to decreasing deGFP production with increasing τ .

Article Body Template

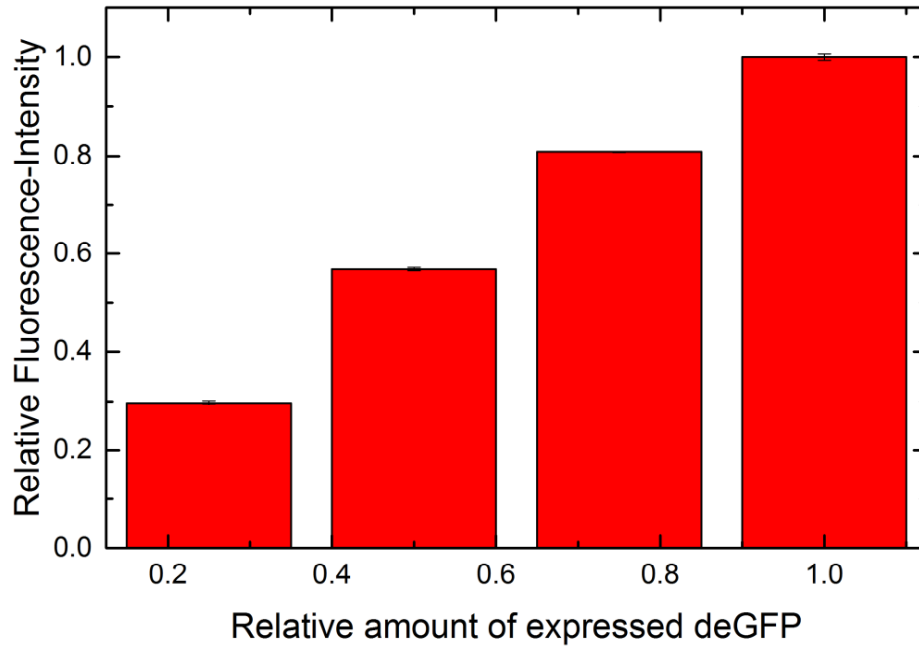


Figure S 2 Relative fluorescence emission-intensities determined with a plate reader for different relative concentrations of expressed deGFP. On the horizontal axis 1 corresponds to the typical, final concentration of GFP in an expression reaction. The other values were obtained by diluting our sample with the expression system several times and determining the amount of fluorescence emission intensity at each step. We find that the measured emission-intensities are linear in GFP concentration within the experimental error. We conclude that fluorescence quenching does not play a role in the concentration range relevant for our study.

3.4 Finkler *et al.* (2020)

A bead-based method for the removal of the amino acid lysine from cell-free transcription-translation systems.

Finkler, M., Kurt, Ö., Grimm, F., Hartz, P., and Ott, A.

Journal of Biotechnology (2020), 324S, 1–6.

DOI: 10.1016/j.btecx.2020.100024

<https://doi.org/10.1016/j.btecx.2020.100024>

This article was originally published in *Journal of Biotechnology: X*. *Journal of Biotechnology: X* is now discontinued and the article is republished in *Journal of Biotechnology* for the reader's convenience.

Reproduced from *Journal of Biotechnology*. (2020) 324S, 1–6 with permission of Elsevier.



Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

A bead-based method for the removal of the amino acid lysine from cell-free transcription-translation systems^{*,*}

Marc Finkler^a, Ömer Kurt^a, Florent Grimm^a, Philip Hartz^b, Albrecht Ott^{a,*}

^a Universität des Saarlandes, Biologische Experimentalphysik, Naturwissenschaftlich-Technische Fakultät, B2 1, Campus, 66123, Saarbrücken, Germany

^b Universität des Saarlandes, Institut für Biochemie, Naturwissenschaftlich-Technische Fakultät, B2 2, Campus, 66123, Saarbrücken, Germany

ARTICLE INFO

Keywords:

Escherichia coli cell extract
Cell-free expression
Microbeads
Lysine
Separation

ABSTRACT

Cell-free transcription-translation systems are a versatile tool to study gene expression, enzymatic reactions and biochemical regulation mechanisms. Because cell-free transcription-translation systems are often derived from cell lysates, many different substances, among them amino acids, are present. However, experiments concerning the incorporation of non-canonical amino acids into proteins require a system with negligible amounts of canonical analogs. Here we propose a two-step method for the removal of residual free lysine in an all *Escherichia coli*-based cell-free expression system. The first step consists of the expression of a high-lysine dummy protein. The second step consists of direct removal *via* binding between lysine and DNA. The presented method is an efficient, fast and simple way to remove residual lysine without altering the system ability to perform gene expression.

1. Introduction

In vitro transcription translation systems have numerous applications, among them the study of gene expression and regulation, biochemical reactions and metabolic pathways (Dudley et al., 2015; Hold and Panke, 2009; Rolf et al., 2019; Schenkelberger et al., 2017). While *in vivo* the concentrations of different biochemical compounds can often only be estimated, *in vitro* systems allow for adjustable concentrations by addition or removal (Hold and Panke, 2009; Rolf et al., 2019; Sullivan et al., 2016). Optimal concentrations of their ingredients, e.g. for maximal protein yield or maximal reaction rate, depend on the particular application (Noireaux et al., 2005; Shin and Noireaux, 2012).

The residue specific incorporation of non-canonical amino acids into proteins is a far reaching application of cell-free transcription-translation systems. Non-canonical amino acids as a surrogate at canonically prescribed locations can lead to new or improved properties of proteins with potential applications in biochemistry, biotechnology and biomedicine (Worst et al., 2015, 2016). However, for residue specific incorporation of noncanonical amino acids even small amounts of a corresponding canonical one will preferentially bind at the level of aminoacyl-tRNA-synthetases (Worst et al., 2016). Experiments by Hopfield have shown that a mechanism termed kinetic proofreading leads to a discrimination of up to 1/1000 against the incorrect amino acid (Hopfield, 1974; Hopfield et al., 1976).

Escherichia coli based expression systems are prepared “devoid” of amino acids, however, still there is a high level of remaining lysine,

compared to other amino acids (Worst et al., 2016). In the following we will discuss methods for the removal of canonical amino acids, in particular lysine.

(Pure) expression systems contain only the strictly necessary compounds for gene expression. These compounds are individually expressed, purified and mixed at a later stage (Gao et al., 2019; Singh-Blom et al., 2014). This avoids the presence of residual matter from the cell lysate (in particular amino acids such as lysine). However, the usage of such systems is expensive (in time or money). Moreover results cannot be transferred to *in vivo* situations because of the minimalistic constitution of the pure system (Gao et al., 2019; Singh-Blom et al., 2014). The system depends on the bacteriophage T7 RNA polymerase, which leads to high protein yield. At the same time the system remains limited in terms of transcription regulation mechanisms (Shin and Noireaux, 2010).

Our cell-free expression systems are derived from *E. coli* cell lysates. They include the entire endogenous transcription-translation machinery (Garenne et al., 2019; Schenkelberger et al., 2017; Worst et al., 2015). In principle lysine-free expression systems could be derived from lysine auxotrophic *E. coli* strains (Gao et al., 2019; Singh-Blom et al., 2014). However, these strains require addition of the depleted amino acid for growth, leading again to the presence of residual amino acids in the final extract. Here, our aim is the creation of a method that will not affect extract preparation and that could be used for any cell-free system.

* This article was originally published in *Journal of Biotechnology: X*. *Journal of Biotechnology: X* is now discontinued and the article is republished here for the reader's convenience. For citation purposes, please use the publication details of this article: *Journal of Biotechnology*, 324S.

* Corresponding author.

E-mail addresses: marc.finkler@uni-saarland.de (M. Finkler), oeurt@physik.uni-saarland.de (Ö. Kurt), s8flgrim@stud.uni-saarland.de (F. Grimm), philip.hartz@uni-saarland.de (P. Hartz), albrecht.ott@physik.uni-saarland.de (A. Ott).

<http://doi.org/10.1016/j.btecx.2020.100024>

Received 6 January 2020; Received in revised form 22 June 2020; Accepted 19 July 2020

Available online 22 July 2020

0168-1656/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

There are different methods to remove particular substances from an expression system. Filtering processes separate as a function of different properties, like hydrophobicity, size, charge, however, confounding lysine with other compounds with similar properties (Brödel et al., 2014; Gao et al., 2019). Antibodies or aptamers enable the separation of compounds in a much more specific way. To the best of our knowledge, however, there is no antibody against free lysine on the market. Related products will only recognize lysine in peptides and proteins (Steinbrecher et al., 1984; Xu et al., 2017). Furthermore antibodies are sensitive to chemical conditions (Jayasena, 1999; Song et al., 2008). A lysine aptamer can bind very specifically against lysine (Carter and Katakay, 2017; Fiegland et al., 2012; Sudarsan et al., 2003). However, depending on the type of aptamer its binding affinity is subject to change as a function of conditions. In our case the presence of K^+ or Mg^{2+} may interfere (Carter and Katakay, 2017; Fiegland et al., 2012; Shin and Noireaux, 2010, 2012).

In this paper we describe a bead-based method for the removal of residual lysine from an all *Escherichia coli*-based cell-free expression system. Our method contains two steps. A first removal is obtained by the incorporation of lysine into a dummy protein with high lysine content. The second step uses the electrostatic attraction between lysine and DNA. The proposed method maintains the extract's ability to perform gene expression.

2. Material and methods

2.1. Cell-free expression system preparation

Our cell-free expression system is prepared from *E. coli* as previously described (Shin and Noireaux, 2010; Sun et al., 2013).

2.2. Gene design

The reporter plasmid is pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500. This plasmid is optimized for our cell-free system. It can be derived from the commercially available plasmid pBest-Or2-Or1-Pr-UTR1-deGFP-T500 (Addgene Cat# 40019) (Finkler and Ott, 2019; Schenkelberger et al., 2017; Shin and Noireaux, 2010). The gene for the high-lysine dummy protein (HLDP), DNA-sequence in Supplementary Fig. S1, is from Geneart (Thermo Fisher Scientific), subcloned into the pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-T500 plasmid using the *Bam*HI and *Xho*I restriction sites to obtain pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-HLDP-T500. The deGFP sequence renders the expression of HLDP by production of the fluorescent fusion-protein deGFP-HLDP (GHLDP).

Adenine repeats impede gene expression (Arthur et al., 2015; Koutmou et al., 2015). Accordingly the sequence of the HLDP was chosen in such a way that lysine is at every second position, separated by any other canonical amino acid. The number of lysine in our HLDP was 53, 3 for all other amino acids (except histidine: six of them are used for His-tagging HLDP). The amino acid sequence can be written as $M(XK)_53H_6$, where X stands for all amino acids except lysine and histidine (3 of each amino acid, 2 methionines, see amino acid-sequence in Supplementary Fig. S1). While the production of this protein reduces the lysine content drastically, compared to lysine, the content of the other amino acids does not decrease significantly.

2.3. Gene preparation and bead coating

The gene coding for the fusion-protein GHLDP is linked to magnetic beads (Finkler and Ott, 2019; Invitrogen, 2017). Using the primers Biotin-TEG-Seq1 (TEG stands for TriEthyleneGlycol) (5'-Biotin-TEG-CAC CAT CAG CCA GAA AAC C-3', Metabion) and Seq4 (5'-GAG CTG ACT GGG TTG AAG G-3', Metabion) and streptavidin-coated magnetic beads T1 (Dynabeads™ MyOne™ Streptavidin T1, Invitrogen) we obtain the construct T1-Biotin-TEG-Seq1-Or2-Or1-Pr-UTR1-deGFP-HLDP-T500-Seq4 (T1-GHLDP) at a DNA-concentration of 400 nM.

2.4. SsDNA preparation and bead coating

As previously shown, mRNA is quite stable in our system. It needs to be removed before new genes can be added (Finkler and Ott, 2019). For removal, DNA complementary to the mRNA of GHLDP is used. It also removes residual lysine. We use biotinylated single-stranded DNA complementary to the 3'-end of the Seq1-Or2-Or1-Pr-UTR1-deGFP-HLDP-T500-Seq4-mRNA (5'-Biotin-TEG-CGG CGG GCT TTG CTC GAG TTA GTG GTG ATG GTG ATG-3', Metabion), linked to streptavidin-coated magnetic beads T1 (3'). Further we used poly-T-coated beads (poly-T) (Dynabeads® Oligo(dT)₂₅, Invitrogen), complementary to the poly(A) tail of the transcribed mRNA. Both ssDNA-bead stock solutions were at 5 mg/mL.

2.5. Experimental procedure

Experiments were performed on ice. The following ingredients were added to 30 μ L of crude *E. coli* cell-free extract: 3.6 μ L magnesium glutamate (at 100 mM), 2.7 μ L potassium glutamate (at 3 M), 6.43 μ L PGA-buffer (3-Phospho-Glyceric Acid) (14x), 4.5 μ L 40 % (v/v) PEG8000, 11.25 μ L amino acids - lysine (mixture contains all canonical amino acids but no lysine at 6 mM except Leu at 5 mM), 3 μ L GamS (Shortened lambda phage Gam protein, prevents linear DNA degradation (Garamella et al., 2016; Sun et al., 2014)) (at 99 μ M) and 1.52 μ L double-distilled water. The obtained 63 μ L were aliquoted in 8.4 μ L (seven samples in a batch).

For experiments 1.5 μ L of the T1-GHLDP stock solution and 2.1 μ L of double-distilled water were added to the 8.4 μ L aliquots. This concentration is optimal for bead-bound DNA (Finkler and Ott, 2019). Aliquots filled up with double-distilled water to 12 μ L were taken as a reference for comparison.

Samples were incubated at 29 °C for 30 min followed by magnetic separation of T1-GHLDP if present. During this time a mixture containing 50 μ L of the 3' and 50 μ L of the poly-T stock solution was freshly prepared according to poly-T manufacturer's protocol to obtain dried ssDNA coated beads (Ambion, 2012). After T1-GHLDP separation the supernatant was transferred to the prepared ssDNA coated beads and incubated for 15 min at room temperature at 230 rpm, followed by magnetic separation of the ssDNA coated beads. Subsequently 10 nM pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 was added to the supernatant to visualize the incorporation of residual lysine in deGFP. The deGFP contains 18 lysine residues that is about 10 % of the whole protein mass. To see if all these preparation steps affect the system's ability to perform further gene expression, we also added 1 mM L-lysine to the supernatant with pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 as reference. The increased error in this measurement (see Fig. 2) is due to small volume pipetting of highly concentrated lysine (about 0.3 μ L of a 40 mM solution) required for maintaining overall concentration constant. For relative quantification of deGFP, 10 μ L of samples were transferred to a microwell plate (Nunc™ 384-Well Optical Bottom Plates # 242764, Thermo Scientific) and the time course of fluorescence-intensity (see Fig. 2) was recorded with a microplate reader (POLARstar OPTIMA, BMG LABTECH) over 16 h. Measurements with an initial GHLDP expression were corrected to obtain the fluorescence-intensity of the reporter protein deGFP only. The fluorescence of GHLDP was estimated by performing the corresponding measurement without addition of 10 nM pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500. The software Origin (Origin 2019b, OriginLab) was used for data evaluation.

It should be mentioned that using magnetic beads for our purpose requires very accurate processing. Using less beads results in diminished lysine removal, increased bead concentrations can remove other compounds in critical amounts so that system efficiency is reduced. Further, incomplete removal of magnetic beads from our expression reaction may interfere with gene expression. The amount of residual

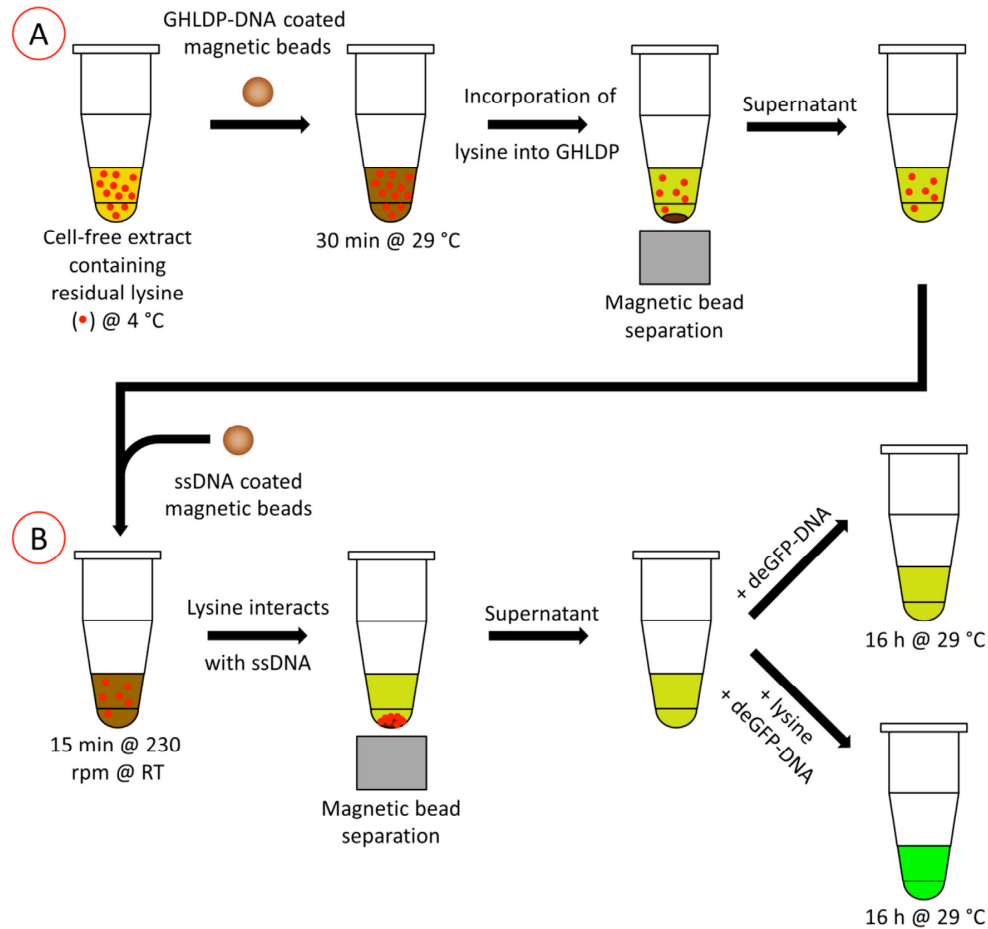


Fig. 1. Scheme of the two step removal of residual lysine in a cell-free expression reaction. The process consists of A) lysine incorporation into a deGFP-high-lysine dummy protein (GHLDP) and B) removal of lysine through binding of lysine to DNA-coated beads. After the two step procedure, addition of deGFP-DNA will not lead to expression because of missing lysine. Only if lysine is added, deGFP will be produced.

lysine differs between different stocks of crude extract. Therefore the amount of magnetic beads has to be adjusted for each stock. The amount of amino acids, magnesium glutamate and potassium glutamate must be adjusted for each stock for optimal cell-free protein synthesis (Finkler and Ott, 2019; Shin and Noireaux, 2010).

3. Results and discussion

As illustrated in Fig. 1 the first step towards lysine removal consists of its incorporation into a GHLDP (deGFP-high-lysine dummy protein, Fig. 1 A). For GHLDP expression all amino acids except lysine are added to the cell-free extract, resulting in lysine the limiting factor. GHLDP production can be observed by measuring its fluorescence-intensity. Because the expression time is short (30 min) not all of the residual lysine is incorporated into GHLDP. A prolonged GHLDP expression results in less residual lysine but also causes fatigue of the cell-free system. Fig. 1 B illustrates how we remove the remaining lysine through electrostatic attraction to DNA. Single-stranded DNA, complementary to the GHLDP-mRNA, is added to remove mRNA and stop GHLDP translation (and GHLDP expression) (Finkler and Ott, 2019; Naiser et al., 2008). Lysine is removed at the same time from the expression system. As a result the addition of DNA coding for deGFP does lead to significant production of deGFP only if lysine is added at the same time.

Fig. 2 shows the measured time course of fluorescence-intensity of expressed deGFP over 16 h for different samples.

To check for self-similarity of the curves for lysine deprived systems, we rescaled the time and fluorescence-intensity (characteristic time, τ , and characteristic amplitude, A). The scaling values are listed in Table 1.

By rescaling the sigmoid curves with the characteristic parameters in Table 1 we obtained Fig. 3.

According to Fig. 3, the curves are self-similar, suggesting a power law description $A \tau^\gamma$. Taking γ as a fitting parameter, we obtain $\gamma = -0.64 \pm 0.06$. A value of -1 corresponds to a simple diffusionlimited process. The result shows that other processes come into play. The exponent is larger than -1, compatible with the idea that proteolysis releases lysine, yielding increased amounts of deGFP. To investigate this idea, we performed experiments with protease inhibitors (Protease Inhibitor Cocktail P8849, Sigma-Aldrich). High concentrations resulted in a non-functioning expression system (possibly non-specific side effects), while lower concentrations had no effect. We conclude that proteolysis is unlikely. In experiments we observed that our expression reaction can be kept for at least 1 h without fatigue if no expression is mediated by DNA addition. Hence another possibility for the observed scaling exponent could be due to system fatigue caused by expression. This would lead to more efficient usage of lysine at low expression rates. The asymptotic endpoints of deGFP expression (the amplitude) then represent a lower bound for lysine

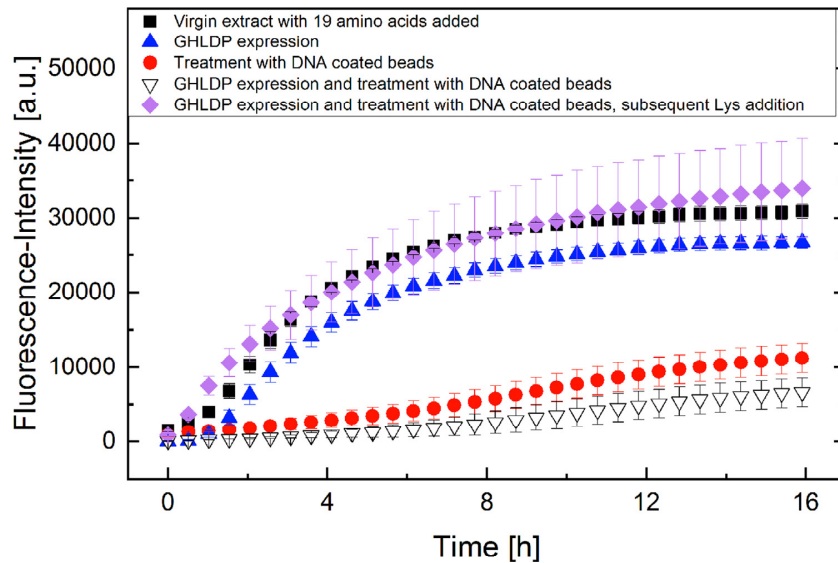


Fig. 2. Time course of deGFP fluorescence-intensity from different cell-free expression reactions. Initial addition of 20 amino acids to a virgin extract results in a final deGFP fluorescence-intensity of about $75,000 \pm 5000$ a.u. (data not shown). All reactions shown were performed with initial addition of only 19 amino acids (all except lysine), making the residual lysine from the initial preparation the limiting factor. Black squares represent the time course of deGFP fluorescence-intensity from a virgin 19 amino acid extract. The other studied samples differ by their treatment to remove residual lysine. The incorporation of residual lysine into GHLDP (blue triangles, following Fig. 1 A), or lysine separation by treatment with ssDNA coated beads (red circles, following Fig. 1 B), results in lower lysine content. The combination of both methods exhibits a cumulative effect (black open triangles). The subsequent addition of lysine mediates the production of deGFP (purple diamonds) comparable to the virgin extract with 19 amino acids added (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

removal. Accordingly we interpret $A \cdot [\text{lysine}]_{\text{Virgin extract}}$ as an upper bound of the remaining lysine and $\tau^{-1} \cdot [\text{lysine}]_{\text{Virgin extract}}$ as an estimation of the lysine concentration.

To show that lysine is the limiting reactant in these experiments we added different amounts of lysine to a lysine depleted cell-free reaction. Supplementary Fig. S2 shows that different deGFP expression levels are reached. We conclude that lysine indeed is the limiting reactant. Because of this we can interpret the deGFP fluorescence-intensity after 16 h of expression to reflect the initial content of lysine in the system (Shin and Noireaux, 2010). With this assumption, the incorporation of residual lysine into GHLDP decreases the lysine content by 14 % while lysine removal through DNA coated beads resulted in a 64 % reduction. The combination of both processes reduces the initial lysine content by almost 80 %. This is in excellent agreement with the values for τ .

According to literature, with our system and all amino acids provided, under optimal conditions a deGFP concentration of 0.65 mg/mL is reached (Shin and Noireaux, 2010). We deduce a concentration of 0.025 mg/mL of residual lysine in our virgin extract. Accordingly the 12 μL of our expression reaction contain at least about 300 ng of residual lysine. Our two step method decreased this to about 60 ng.

Fig. 2 shows that the difference in the final fluorescence-intensity between the black squares and the blue triangles (4636.24 ± 1925.96 a.u.) is similar to the difference between the red circles and the black open triangles (4250.33 ± 914.91 a.u.). In both cases the difference in the final fluorescence-intensity was obtained by initial GHLDP expression. Although additional treatment with DNA coated beads decreases the final fluorescence-intensity, the relative contribution of GHLDP to the decrease is the same as without the additional bead treatment. The two steps of our proposed method to remove residual lysine are cumulative.

Subsequent addition of lysine after initial two-step removal of residual lysine led to strong deGFP expression (Fig. 2, purple diamonds). This

shows that our proposed method is minimally invasive. The expression system is not affected in its ability to perform gene expression. The subsequent addition of lysine to an expression reaction whose lysine was removed with our two step method (Fig. 2, purple diamonds) led to only half the amount of deGFP, compared to adding 20 amino acids to a virgin extract. Beginning fatigue of the expression system by initial GHLDP expression is one possible explanation.

We achieved a five-fold terminal fluorescence-intensity by addition of lysine at $t = 0$ h to an extract that was initially treated with our two step method (Fig. 2, black open triangles and purple diamonds), which means that 80 % of the produced deGFP is from the added (not the residual) lysine.

4. Conclusion

We presented a new method, simple and fast, to remove a huge fraction (80 %) of residual lysine from an *E. coli* based cell-free expression system. Our minimally-invasive method does not alter the system in its ability to perform gene expression. After lysine removal the subsequent addition of deGFP-DNA and lysine resulted in strong deGFP production.

We found a relationship between the characteristic amplitude, A and the characteristic time, τ in experiments with limited lysine content, $A \cdot \tau^{-0.64 \pm 0.06}$. Because the exponent differs from -1 we suggest that other processes, like expression mediated system fatigue may play a role.

Although our method removes up to 80 % of the residual lysine, different experiments may require to even further decrease the amount of lysine. To achieve this, modifications in the crude extract preparation like prolonged dialysis (from 3 h (Shin and Noireaux, 2010) to up to 12 h) may be useful. Further the usage of a lysine auxotrophic bacterial strain for extract preparation in combination with our method may be an effective way to decrease the residual lysine to even lower levels.

Table 1

Characteristic time until inflection of the time-resolved fluorescence-intensity (Characteristic time, τ) and characteristic amplitude of fluorescence-intensity (Characteristic amplitude, A) for different samples. These values were used for rescaling the curves in Fig. 2 (Fig. 3).

Sample	Characteristic time, τ [a.u.]	Characteristic amplitude, A [a.u.]
Virgin extract with 19 amino acids added	1	1
GHLDP expression	1.1	0.91
Treatment with DNA coated beads	3.29	0.5
GHLDP expression and treatment with DNA coated beads	4.06	0.37

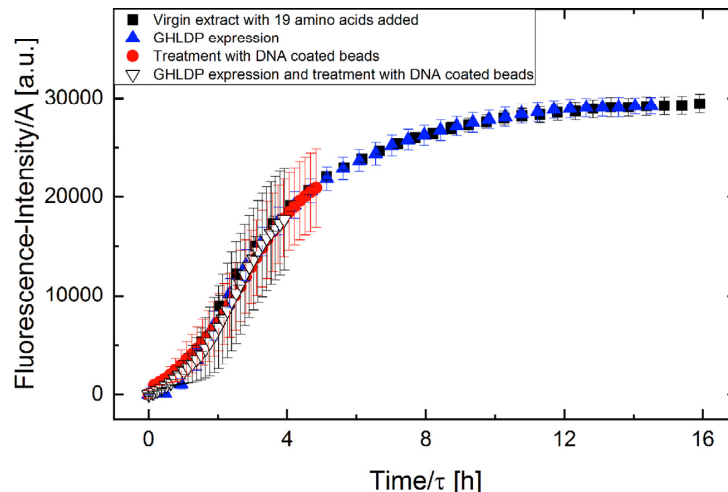


Fig. 3. Rescaled curves from Fig. 2, using the parameters in Table 1.

The indirect removal of lysine by incorporation into a dummy-protein can be transferred to other amino acids easily. In contrast, the removal of amino acids through binding to DNA is unlikely to work for negatively charged amino acids.

Author contributions

AO and MF designed the research. MF performed the research with the help of ÖK and FG. MF analyzed data with the help of AO, ÖK and PH. MF and AO wrote the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the collaborative research center SFB 1027 funded by Deutsche Forschungsgemeinschaft (DFG), and by the Human Frontier Science Program (HFSP, RGP0037/2015). We thank Emanuel Worst for help and fruitful discussions.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jbtecx.2020.100024>.

References

- Ambion, 2012. Manual for Dynabeads® Oligo (dT)₂₅. Life Technologies.
- Arthur, L.L., Pavlovic-Djuranovic, S., Koutmou, K.S., Green, R., 2015. Translational control by lysine-encoding A-rich sequences. *Sci. Adv.* 1, 1–11. <https://doi.org/10.1126/sciadv.1500154>.
- Brödel, A.K., Sonnabend, A., Kubick, S., 2014. Cell-free protein expression based on extracts from CHO cells. *Biotechnol. Bioeng.* 111, 25–36. <https://doi.org/10.1002/bit.25013>.
- Carter, Z.A., Katakly, R., 2017. A G-quadruplex aptamer based impedimetric sensor for free lysine and arginine. *Sensors Actuators B Chem.* 243, 904–909. <https://doi.org/10.1016/j.snb.2016.12.010>.
- Dudley, Q.M., Karim, A.S., Jewett, M.C., 2015. Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnol. J.* 10, 69–82. <https://doi.org/10.1002/biot.201400330>.
- Fiegand, L.R., Garst, A.D., Batey, R.T., Nesbitt, D.J., 2012. Single-molecule studies of the lysine riboswitch reveal effector dependent conformational dynamics of the aptamer domain. *Biochemistry* 51, 9223–9233. <https://doi.org/10.1021/bi3007753>.

- Finkler, M., Ott, A., 2019. Bead-based assay for spatiotemporal gene expression control in cell-free transcription-translation systems. *Biotechniques* 66, 29–33. <https://doi.org/10.2144/btn-2018-0097>.
- Gao, W., Cho, E., Liu, Y., Lu, Y., 2019. Advances and challenges in cell free incorporation of unnatural amino acids into proteins. *Front. Pharmacol.* 10, 1–8. <https://doi.org/10.3389/fphar.2019.00611>.
- Garamella, J., Marshall, R., Rustad, M., Noireaux, V., 2016. The all *E. coli* TX-TL toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth. Biol.* 5, 344–355. <https://doi.org/10.1021/acssynbio.5b00296>.
- Garenne, D., Beisel, C.L., Noireaux, V., 2019. Characterization of the all-*E. coli* transcription-translation system myTXTL by mass spectrometry. *Rapid Commun. Mass Spectrom.* 33, 1036–1048. <https://doi.org/10.1002/rcm.8438>.
- Hold, C., Panke, S., 2009. Towards the engineering of *in vitro* systems. *J. R. Soc. Interface* 6, 507–521. <https://doi.org/10.1098/rsif.2009.0110.focus>.
- Hopfield, J.J., 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U.S.A.* 71, 4135–4139. <https://doi.org/10.1073/pnas.71.10.4135>.
- Hopfield, J.J., Yamane, T., Yue, V., Coutts, S.M., 1976. Direct experimental evidence for kinetic proofreading in amino acylation of tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 73, 1164–1168. <https://doi.org/10.1073/pnas.73.4.1164>.
- Invitrogen, 2017. Manual for Dynabeads™ MyOne™ Streptavidin T1. Thermo Fisher Scientific.
- Jayasena, S.D., 1999. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.* 45, 1628–1650.
- Koutmou, K.S., Schuller, A.P., Brunelle, J.L., Radhakrishnan, A., Djuranovic, S., Green, R., 2015. Ribosomes slide on lysine-encoding homopolymeric A stretches. *Elife* 4, 1–18. <https://doi.org/10.7554/eLife.05534>.
- Naiser, T., Ehler, O., Kayser, J., Mai, T., Michel, W., Ott, A., 2008. Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges. *BMC Biotechnol.* 8, 1–23. <https://doi.org/10.1186/1472-6750-8-48>.
- Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H., Libchaber, A., 2005. Toward an artificial cell based on gene expression in vesicles. *Phys. Biol.* 2, 1–8. <https://doi.org/10.1088/1478-3975/2/3/P01>.
- Rolf, J., Rosenthal, K., Lütz, S., 2019. Application of cell-free protein synthesis for faster biocatalyst development. *Catalysts* 9, 1–18. <https://doi.org/10.3390/catal9020190>.
- Schenkelberger, M., Shanak, S., Finkler, M., Worst, E., Noireaux, V., Helms, V., Ott, A., 2017. Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system. *Phys. Biol.* 14, 1–11. <https://doi.org/10.1088/1478-3975/aa5d37>.
- Shin, J., Noireaux, V., 2010. Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4, 1–9. <https://doi.org/10.1186/1754-1611-4-8>.
- Shin, J., Noireaux, V., 2012. An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* 1, 29–41. <https://doi.org/10.1021/sb200016s>.
- Singh-Blom, A., Hughes, R.A., Ellington, A.D., 2014. An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J. Biotechnol.* 178, 12–22. <https://doi.org/10.1016/j.jbiotec.2014.02.009>.
- Song, S., Wang, L., Li, J., Zhao, J., Fan, C., 2008. Aptamer-based biosensors. *Trends Anal. Chem.* 27, 108–117. <https://doi.org/10.1016/j.trac.2007.12.004>.
- Steinbrecher, U.P., Fisher, M., Witztum, J.L., Curtiss, L.K., 1984. Immunogenicity of homologous low density lipoprotein after methylation, ethylation, acetylation, or carbamylation: generation of antibodies specific for derivatized lysine. *J. Lipid Res.* 25, 1109–1116.
- Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., Breaker, R.R., 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* 17, 2688–2697. <https://doi.org/10.1101/gad.1140003>.

M. Finkler et al.

Journal of Biotechnology 324S (2020) 100024

- Sullivan, C.J., Pendleton, E.D., Sasmor, H.H., Hicks, W.L., Farnum, J.B., Muto, M., Amendt, E.M., Schoborg, Ja., Martin, R.W., Clark, L.G., Anderson, M.J., Choudhury, A., Fior, R., Lo, Y.H., Griffey, R.H., Chappell, Sa., Jewett, M.C., Mauro, V.P., Dresios, J., 2016. A cell-free expression and purification process for rapid production of protein biologics. *Biotechnol. J.* 11, 238–248. <https://doi.org/10.1002/biot.201500214>.
- Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., Noireaux, V., 2013. Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* 79, 1–14. <https://doi.org/10.3791/50762>.
- Sun, Z.Z., Yeung, E., Hayes, C.A., Noireaux, V., Murray, R.M., 2014. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* 3, 387–397. <https://doi.org/10.1021/sb400131a>.
- Worst, E.G., Exner, M.P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., Ott, A., 2015. Cell-free expression with the toxic amino acid canavanine. *Bioorg. Med. Chem. Lett.* 25, 3658–3660. <https://doi.org/10.1016/j.bmcl.2015.06.045>.
- Worst, E.G., Exner, M.P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., Ott, A., 2016. Residue-specific incorporation of noncanonical amino acids into model proteins using an *Escherichia coli* cell-free transcription-translation system. *J. Vis. Exp.* 114, 1–14. <https://doi.org/10.3791/54273>.
- Xu, H., Zhou, J., Lin, S., Deng, W., Zhang, Y., Xue, Y., 2017. PLMD: an updated data resource of protein lysine modifications. *J. Genet. Genomics* 44, 243–250. <https://doi.org/10.1016/j.jgg.2017.03.007>.

Supplementary Material

A bead-based method for the removal of the amino acid lysine from cell-free transcription-translation systems.

Finkler, M., Kurt, Ö., Grimm, F., Hartz, P., and Ott, A.

Journal of Biotechnology (2020), 324S, 1–6.

DOI: 10.1016/j.btecx.2020.100024

<https://doi.org/10.1016/j.btecx.2020.100024>

This article was originally published in *Journal of Biotechnology: X*. *Journal of Biotechnology: X* is now discontinued and the article is republished in *Journal of Biotechnology* for the reader's convenience.

Reproduced from *Journal of Biotechnology*. (2020) 324S, 1–6 with permission of Elsevier.

Supplementary material



Fig. S1. DNA- and the deduced amino acid-sequence of HLDP. Because the gene expression of a polylysine gene was not successful if adenine repeats were present (Arthur et al., 2015; Koutmou et al., 2015) the sequence was adapted to a maximum of 3 adenines in a row. The codons coding for lysine are AAA and AAG, and the insertion of another amino acid between two lysines was necessary to prevent adenine stretches. Codons optimal for our system were chosen using the Kazusa codon usage database preventing 4 adenines in a row. Alanine was chosen to be a spacer between the initial methionine and the first lysine. Further a 6xHis-Tag was added to the last lysine. To keep the content of all amino acids except lysine and histidine constant, each amino acid is present at the same number of 3. This gives 53 lysine residues in our HLDP. *BamHI* and *XhoI* restriction sites are flanking our sequence.

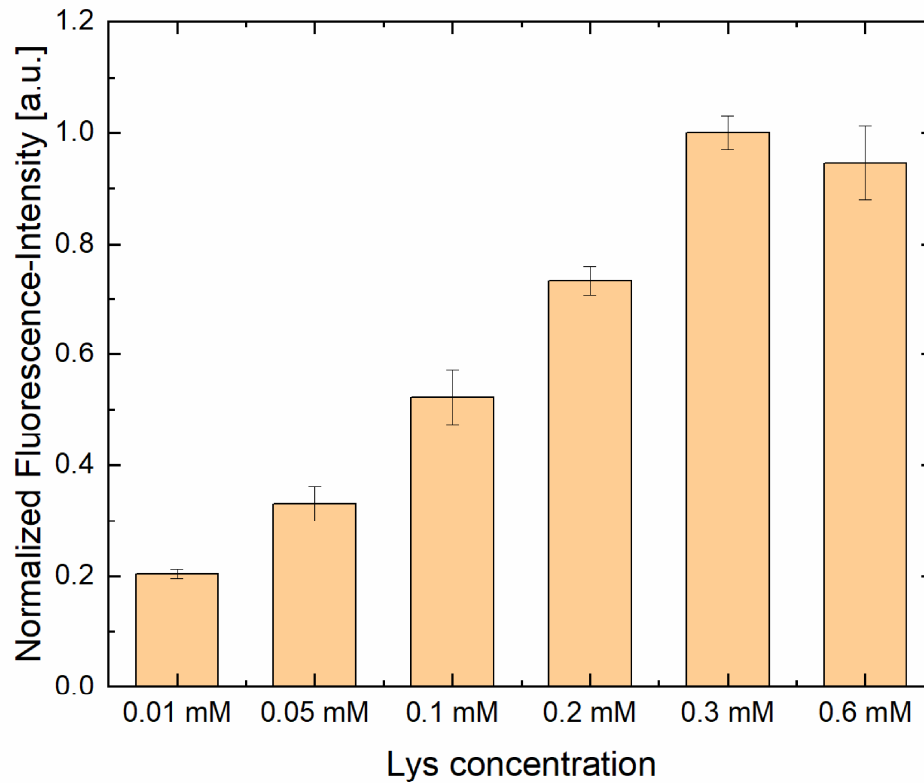


Fig. S2 Relative amounts of expressed deGFP for cell-free expression reactions that had residual lysine removed using our method, followed by addition of different amounts of lysine. The data is normalized to the observed maximum that occurs at the considered concentrations of lysine and different batches of crude extract. DeGFP production strongly depends on the amount of freshly added lysine. According to the data lysine becomes the limiting factor in the cell-free expression systems where residual lysine was removed with our method.

References

- Arthur, L.L., Pavlovic-Djuranovic, S., Koutmou, K.S., Green, R., 2015. Translational control by lysine-encoding A-rich sequences. *Sci. Adv.* 1, 1–11. <https://doi.org/10.1126/sciadv.1500154>
- Koutmou, K.S., Schuller, A.P., Brunelle, J.L., Radhakrishnan, A., Djuranovic, S., Green, R., 2015. Ribosomes slide on lysine-encoding homopolymeric A stretches. *Elife* 2015, 1–18. <https://doi.org/10.7554/eLife.05534>

3.5 Finkler *et al.* (2021)

Full incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in green fluorescent protein.

Finkler, M., Ravanbodshirazi, S., Grimm, F., Hartz, P., and Ott, A.

Bioorganic & Medicinal Chemistry (2021), *41*, 1–6.

DOI: 10.1016/j.bmc.2021.116207

<https://doi.org/10.1016/j.bmc.2021.116207>

Reproduced from *Bioorganic & Medicinal Chemistry*. (2021) 41, 1–6 with permission of Elsevier.



Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Full incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in green fluorescent protein

Marc Finkler^a, Sina Ravanbodshirazi^a, Florent Grimm^a, Philip Hartz^b, Albrecht Ott^{a,*}^a Universität des Saarlandes, Biologische Experimentalphysik, Naturwissenschaftlich-Technische Fakultät, B2 1, Campus, 66123 Saarbrücken, Germany^b Universität des Saarlandes, Institut für Biochemie, Naturwissenschaftlich-Technische Fakultät, B2 2, Campus, 66123 Saarbrücken, Germany

ARTICLE INFO

Keywords:

Hydroxylysine
Lysine
Escherichia coli cell extract
Cell-free expression
Noncanonical amino acid

ABSTRACT

The canonical set of amino acids leads to an exceptionally wide range of protein functionality, nevertheless, this set still exhibits limitations. The incorporation of noncanonical amino acids into proteins can enlarge its functional scope. Although proofreading will counteract the charging of tRNAs with other amino acids than the canonical ones, the translation machinery may still accept noncanonical amino acids as surrogates and incorporate them at the canonically prescribed locations within the protein sequence. Here, we use a cell-free expression system to demonstrate the full replacement of L-lysine by L-hydroxylysine at all lysine sites of recombinantly produced GFP. *In vivo*, as a main component of collagen, post-translational L-hydroxylysine generation enables the formation of cross-links. Our work represents a first step towards *in vitro* production of (modified) collagens, more generally of proteins that can easily be crosslinked.

1. Introduction

Production, optimization and modification of proteins and enzymes through protein engineering^{1–3} often aim at novel, improved or specialized protein functionalities.³ Proteins can be obtained by isolation from native organisms, and by chemical or recombinant production (*in vivo* or *in vitro*).^{4–7} Proteins can be post-translationally modified by enzymes, or more generally by chemical processes, to alter their amino acids or motifs.^{8,9} A different strategy consist of producing the protein directly with the envisioned alterations.^{10,11} During chemical protein synthesis the involved substances, in particular the amino acids can be varied, enabling the production of proteins with novel amino acid sequences or non-natural ingredients.^{6,12} At the same time, chemical synthesis can be difficult to control and perform, among others because of the production of by-products.¹³ Recombinant protein production by an organism or an *in vitro* system⁷ enables better control.¹³ An *in vitro* system requires DNA carrying the gene to be expressed as well as all regulatory units for expression that will act as a template for protein production.^{14–16} The amino acid sequence of a protein can simply be modified by changing the DNA-sequence of the related gene. Unnatural (noncanonical) amino acids can be introduced in a site-specific or residue-specific manner.^{17–19}

The site-specific incorporation of noncanonical amino acids into proteins requires an orthogonal pair of tRNA and aminoacyl-tRNA synthetase, and the selection of a desired recognition site for the tRNA (codon).^{17,18,20–22} Such an approach can overcome the necessity of amino-acid depletion for the incorporation of analogs into protein. *In vivo*, orthogonal systems were used for instance for site-specific incorporation of lysine/pyrrolysine derivatives.⁴¹ However, such orthogonal systems need to be optimized for each noncanonical amino acid and the number of incorporated noncanonical amino acids in one protein is limited to one or just a few.^{19,42}

Residue-specific incorporation of noncanonical amino acids into proteins is simpler.¹⁹ However, it is necessary that the noncanonical amino acid is recognized by the endogenous translation machinery of the expression system,^{17,18,23} making the technique suitable for analogs of canonical amino acids.^{24,25}

Each noncanonical amino acid presents different affinities to its corresponding aminoacyl-tRNA-synthetases.^{17,18,26} Due to proofreading^{27,28} performed by the translation machinery, even trace amounts of the canonical amino acids originating from the expression system will be preferred over the surrogate.^{17,18,29} The extent of proofreading activity depends on the particular noncanonical amino acid in question.¹⁷ While the successful substitution of tryptophan by a

* Corresponding author.

E-mail addresses: marc.finkler@uni-saarland.de (M. Finkler), sina.ravanbodshirazi@physik.uni-saarland.de (S. Ravanbodshirazi), sflgrim@stud.uni-saarland.de (F. Grimm), philip.hartz@uni-saarland.de (P. Hartz), albrecht.ott@physik.uni-saarland.de (A. Ott).

<https://doi.org/10.1016/j.bmc.2021.116207>

Received 20 February 2021; Received in revised form 22 April 2021; Accepted 26 April 2021

Available online 9 May 2021

0968-0896/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

tryptophan analog had to be performed in an expression system devoid of tryptophan, low quantities of arginine did not interfere with incorporation of the arginine analog canavanine.^{18,26} The affinity of the arginyl-tRNA-synthetase for canavanine is similar or identical to arginine.²⁶ However, in the expression system the yield of modified protein was lower than the yield of the wild type.²⁶ As a general rule, gene expression should be performed in the presence of the surrogate and the absence of the canonical amino acid to be replaced,^{18,19} a condition forcing the expression system to accept the noncanonical amino acid.^{17,24}

In this paper, we investigate the incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in GFP protein. Our work is a first step for the incorporation of hydroxylysine in more complex proteins. Hydroxylysine (L-5-Hydroxylysine) is an important component of collagen fibers.³⁰⁻³² Its absence can cause several disorders.³³⁻³⁵ In the organism hydroxylysine is not directly incorporated in the nascent polypeptide chains of collagen,^{31,36} rather hydroxylysine results from posttranslational hydroxylation of specific lysine residues by the enzyme lysyl hydroxylase.^{31,33,37} Since collagen can only be obtained by isolation from animal sources or cell-cultures originating from animal sources, direct incorporation leads the way to produce collagen and its derivative gelatin without harming animals.³⁸ Moreover, *in vitro* replacement enables simple, controlled modifications in collagen because the amino acid sequence can easily be changed in the expression system, opening possibilities to adjust the properties of collagen. Modified collagens are important for medical purposes like skin or bone tissue engineering.^{39,40} Another purpose of a direct hydroxylysine incorporation is the engineering of proteins with cross-linking activity.^{32,35}

We showed earlier¹⁷ that even trace quantities of lysine in the expression system will prevent the efficient incorporation of hydroxylysine as a surrogate. Lysine in the system originates from its preparation. The release of free lysine due to proteolysis is negligible.²⁹ We call 'residual lysine' the trace quantities of lysine in the virgin extract that can be incorporated into proteins.

Here, as a first step, we investigated the kinetics of green fluorescent protein production as a function of the amount of freshly added lysine. The comparison of the kinetics in a virgin and in a lysine depleted system helped us to see that freshly added and residual lysine are incorporated at different time intervals. In a second step we produced lysine depleted cell-free expression systems that minimize the inhibitory effect of residual lysine on the incorporation of hydroxylysine. For this purpose, we used ssDNA coated beads in order to remove residual lysine from the cell-free expression system.²⁹ With such a lysine depleted system we were able to produce proteins containing hydroxylysine at all lysine sites.

2. Material and methods

2.1. Expression system

The *E. coli* based cell-free expression system was prepared according to.^{7,14}

We used pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 as a reporter plasmid to visualize gene expression. This plasmid is derived from the plasmid pBest-Or2-Or1-Pr-UTR1-deGFP-T500 (Addgene Cat# 40019).¹⁴⁻¹⁶

2.2. Lysine depletion

For removing residual lysine from the expression system we used a previously described method with minor modifications.²⁹ We did not perform the step of producing a dummy protein with high lysine content because the non-specific lysine removal by treatment with ssDNA coated beads was sufficient for our purposes. For treatment, biotinylated ssDNA (5'-Biotin-TEG-CGG CGG GCT TTT GAG TTA GTG GTG ATG GTG

ATG-3', Metabion) was attached to streptavidin-coated magnetic beads T1 (Dynabeads® MyOne™ Streptavidin T1, Invitrogen).⁴³ Additionally, poly-T-coated beads (Dynabeads™ Oligo(dT)25, Invitrogen) were used. 400 µl of each ssDNA coated bead stock solution ($C_{\text{bead}} = 5 \text{ mg/ml}$) were mixed and prepared according to instructions.⁴⁴ 3 µl of GamS (Shortened lambda phage Gam protein, prevents linear DNA degradation^{16,45,46}) (at 99 µM) were added to 30 µl of crude *E. coli* cell-free extract. This mixture was then added to the dried beads and incubated for 20 min at 23 °C and 230 rpm. As a reference, 30 µl of crude *E. coli* cell-free extract with 3 µl of GamS were incubated with the same parameters but without ssDNA coated beads.

2.3. Gene expression

During incubation, the following master mix (for 30 µl of crude *E. coli* cell-free extract with 3 µl of GamS) was prepared: to 15 µl amino acids (mixture containing all canonical amino acids at 6 mM, except Leu at 5 mM, but no lysine) we added 3.6 µl magnesium glutamate (at 100 mM), 2.7 µl potassium glutamate (at 3 M), 6.43 µl PGA-buffer (3-Phosphoglyceric Acid) (14x) and 4.5 µl 40% (v/v) PEG8000. Further we added pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 to get a final concentration of 10 nM and filled up with ultrapure water to get a total volume of 39 µl. The final concentration of amino acids, magnesium glutamate and potassium glutamate had to be adjusted for each batch of crude extract to obtain optimal gene expression.

Experiments investigating the incorporation of freshly added and residual lysine into proteins were performed as followed. After 20 min of incubation with (or without) ssDNA coated beads the crude extract including GamS was aliquoted to volumes of 4.4 µl (seven aliquots in one batch). We added 5.2 µl of the master mix to each aliquot. Different quantities of lysine were added to the obtained volume of 9.6 µl. The total reaction volume was adjusted to 12 µl by adding ultrapure water. For investigation of the incorporation of L-5-hydroxylysine (Hyl) (DL-5-Hydroxylysine hydrochloride, Sigma-Aldrich) into proteins, we performed analogous experiments but without aliquoting. The concentrations as given in this paper represent the concentration of the L-enantiomer, that is, half of the concentration of the racemate obtained from the manufacturer.

The incorporation of the freshly added and/or residual amino acid into deGFP was visualized by transferring 10 µl of samples to a micro-well plate (Nunc™ 384-Well Optical Bottom Plates # 242764, Thermo Scientific), recording the time course of fluorescence-intensity with a microplate reader (POLARstar OPTIMA, BMG LABTECH) over 16 h at 29 °C.

2.4. Protein purification and mass spectrometry

Mass spectrometry was performed to confirm the incorporation of hydroxylysine into deGFP. For this purpose, we used immobilized metal ion affinity chromatography (His-Spin Protein Miniprep, Zymo Research) to purify the recombinantly produced protein. Protein concentrators (Pierce™ Protein Concentrators PES, 10 K MWCO, 0.5 mL, Thermo Scientific) were used to desalt and concentrate the purified protein solutions. For the latter we washed the purified protein two times with 500 µl ultrapure water followed by two wash steps with 500 µl 20% acetonitrile + 0.5% formic acid. The supernatant containing concentrated purified protein in 20% acetonitrile + 0.5% formic acid was injected into the mass spectrometer (Solarix FTICR-ESI, Bruker) for measurement. The obtained raw mass spectrums were deconvoluted using the Maximum Entropy algorithm of the Compass DataAnalysis software (Bruker Compass DataAnalysis Version 5.0, Bruker). Origin (Origin 2020b, OriginLab) was used for data evaluation.

GamS is a necessary ingredient for our method of lysine removal. It was extracted by His-Tag purification as well, producing strong peaks in the mass spectrum that interfered with deconvolution ($c(\text{GamS}) \gg c(\text{deGFP18Hyl})$). Consequently, the spectrum of GamS had to be

subtracted before deconvolution.

2.5. Sds-PAGE

We performed SDS-PAGE to strengthen the results obtained by mass spectrometry. For this purpose, we produced deGFP and purified, desalted and concentrated as described in section 2.4 except that washing exclusively consisted of four steps using 500 μ l ultrapure water. The concentrated and purified protein was denatured in sample buffer (TruPAGE™ LDS Sample Buffer, Sigma-Aldrich). For SDS-PAGE we used precast gels (TruPAGE™ Precast Gels 10%, 10 \times 10 cm, 12-well, Sigma-Aldrich) and MOPS buffer (TruPAGE™ Tris-MOPS SDS Express Running Buffer, Sigma-Aldrich).

3. Results and discussion

3.1. Kinetics of green fluorescent protein production as a function of supplemented lysine

We looked at the kinetics of protein production as a function of the amount of freshly added lysine in a virgin expression system and a lysine depleted one (Fig. 1). We were able to monitor real time protein production by choosing green fluorescent protein (deGFP) as a model protein and measuring the fluorescence-intensity of deGFP produced in our cell-free expression system.

Fig. 1 shows the time course of the fluorescence-intensity of the cell-free expression system resulting from deGFP production. We take the fluorescence-intensity as proportional to the amount of produced deGFP. Only for the highest concentrations of freshly added lysine we observed a single regime only. In all other cases we distinguished two different regimes: a first plateau of deGFP production was reached after a few hours. Then a conversion point emerged where deGFP production restarted to saturate again only later. The restart in production was more prominent in Fig. 1A than in Fig. 1B where an important fraction of the residual lysine was removed from the cell-free expression reaction ($c_{\text{residual Lys, 1A}} = 0.009 \mu\text{g}/\mu\text{l}$; $c_{\text{residual Lys, 1B}} = 0.002 \mu\text{g}/\mu\text{l}$, concentrations deduced from deGFP fluorescence-intensity as previously described²⁹). Compared to Fig. 1A, in Fig. 1B the conversion point appeared much later, and the second increase was substantially weaker. Moreover, in both cases the addition of fresh lysine had an immediate effect on the increase in fluorescence-intensity during the first hours of the experiment. The position of the first plateau differed, while the difference between the first and the second saturation level remained almost constant ($\Delta_{\text{S1S2, 1A}} = 5445.33 \pm 347.36$; $\Delta_{\text{S1S2, 1B}} = 1715.7 \pm 409.86$,

$c_{\text{added Lys}} = 0 \text{ mM}, \dots, 0.2 \text{ mM}$).

Note that the first saturation levels of fluorescence-intensity of the virgin cell-free expression system (Fig. 1A) and the final saturation levels of lysine depleted system (Fig. 1B) were comparable for different concentrations of lysine added (Fig. 2).

We interpret the observations above by the first plateau of the fluorescence-intensity representing the incorporation of the freshly added lysine into deGFP while the second plateau corresponds to the incorporation of residual lysine. The collapse from two plateaus to only one at high concentrations of freshly added lysine may well correspond to a suppression of the incorporation of residual lysine due to the high ratio of freshly added to residual lysine.

To verify our interpretation, we performed the same experiments using the stable lysine isotope $^{13}\text{C}_6 \text{ }^{15}\text{N}_2$ -lysine (LysIso) (L-Lysine-2HCl, 13C6, 15 N2 for SILAC, Thermo Scientific). This isotope has the same chemical properties as the non-isotopic isomer but differs in mass.⁴⁷⁻⁴⁹ The label enabled us to detect the incorporation of isotopic lysine within the deGFP (18 lysine sites) since $M_{\text{LysIso}} - M_{\text{Lys}} = 8 \text{ Da}$. For our purpose,

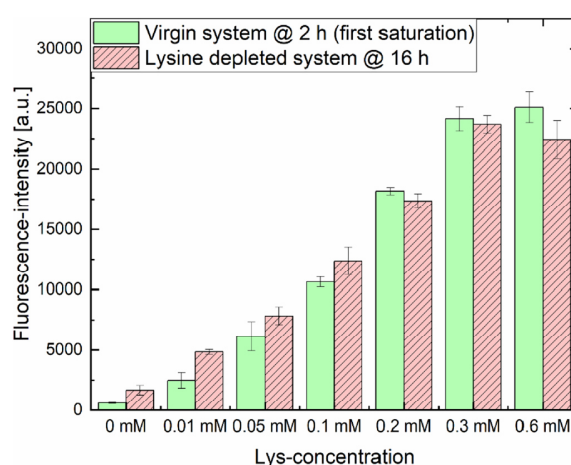


Fig. 2. Comparison of the first saturation level of deGFP production in a virgin system ($c_{\text{residual Lys}} = 0.009 \mu\text{g}/\mu\text{l}$) and the final saturation level of deGFP production in a lysine depleted system ($c_{\text{residual Lys}} = 0.002 \mu\text{g}/\mu\text{l}$). The graph shows the ratios of the values from Fig. 1A @ $t = 2 \text{ h}$ and Fig. 1B @ $t = 16 \text{ h}$ for different concentrations of freshly added lysine.

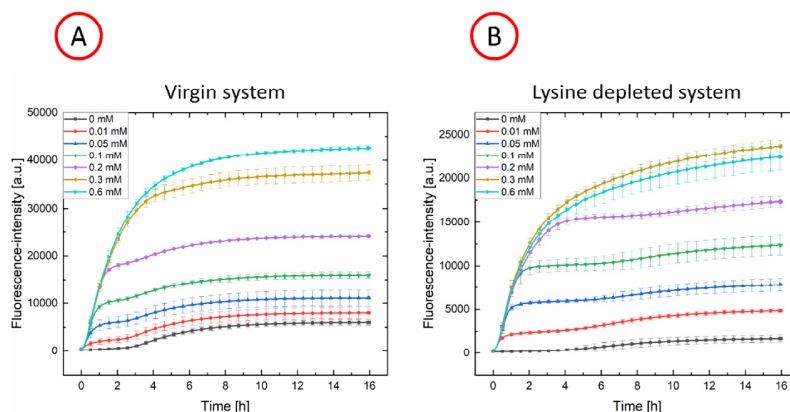


Fig. 1. Time course of the fluorescence-intensity of green fluorescent protein (deGFP) in our cell-free expression system for different concentrations of freshly added lysine (insert). The experiment was performed using a virgin system (A, residual lysine present) as well as a lysine depleted system (B, amount of residual lysine reduced by treatment with ssDNA coated beads).

we recorded a mass spectrum of deGFP from a virgin cell-free expression reaction with freshly added isotopic lysine ($c = 0.1$ mM) after 2 h of incubation (first plateau) and after 16 h (second plateau) at 29 °C. We also determined the mass spectrum after 16 h if no isotopic lysine ($c = 0$ mM) or a high amount of isotopic lysine ($c = 0.6$ mM) was added. The four mass spectra are shown in Fig. 3.

The results shown in Fig. 3 are in excellent agreement with our assumption regarding the kinetics of freshly added vs. residual lysine.

From the different kinetics discussed above, we conclude that there is a threshold in the concentration of freshly added lysine at which the incorporation of residual lysine is suppressed. In the following we show how to use this observation to achieve efficient hydroxylysine incorporation in green fluorescent protein.

3.2. Hydroxylysine incorporation

We investigated the incorporation of hydroxylysine into deGFP. Fig. 4 shows the chemical structure of L-hydroxylysine compared to L-lysine.

In a first step we tried to force our virgin cell-free expression system

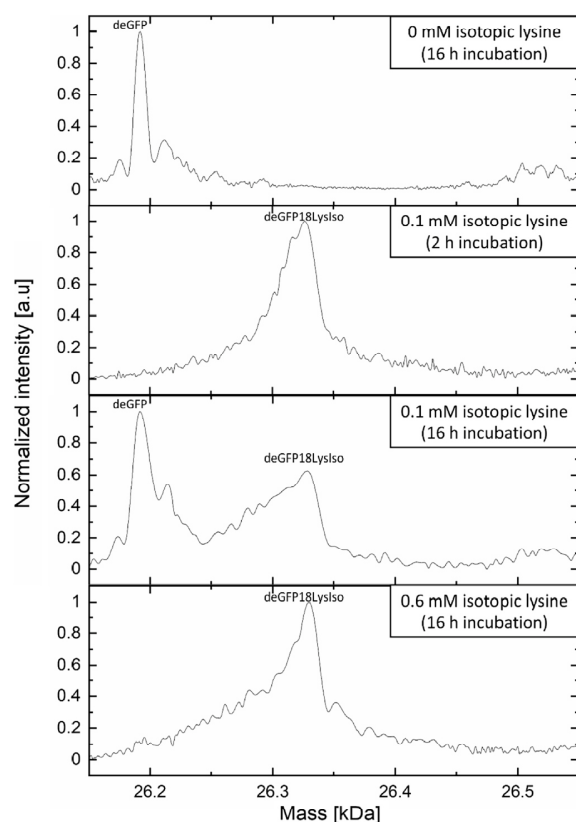


Fig. 3. Mass spectra of deGFP from different cell-free expression reactions, taken after 16 h of gene expression at 29 °C with 0 mM, 0.1 mM, or 0.6 mM of isotopic lysine added. Additionally, for 0.1 mM of isotopic lysine, a mass spectrum was recorded after 2 h of expression. Comparing the cases of no isotopic lysine added ($c = 0$ mM) and isotopic lysine at 0.6 mM, we understand that the incorporation of residual lysine ($M_{\text{deGFP}} = 26.193$ kDa) was suppressed by the presence of the isotope ($M_{\text{deGFP18LysIso}} = 26.337$ kDa). In the case of 0.1 mM of isotopic lysine and 2 h of incubation we detected solely the isotopic peak, $M = 26.333$ kDa (the difference between measured and literature value of 4 Da is due to non-isotopic impurities resulting in broadened peaks). After 16 h, however, both peaks (isotopic and non-isotopic) were present.

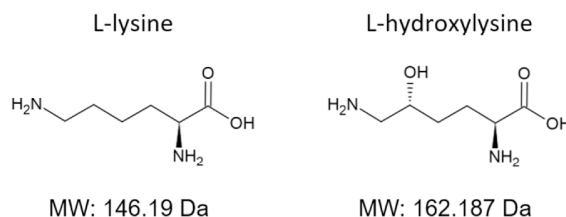


Fig. 4. Chemical structure of L-lysine (left) and L-hydroxylysine (right).

to accept hydroxylysine by increasing the amount of hydroxylysine. Fig. 5 shows the mass spectra for different, initial concentrations of hydroxylysine after 16 h of incubation. (Measurements after 2 h did not lead to any spectrum because of low protein content (data not shown).) We just observed peaks of correctly produced deGFP with different ratios of lysine to hydroxylysine. Peaks of ill-produced deGFP were not observed.

By increasing the concentration from 1 mM of hydroxylysine to 16.8 mM ($C_{\text{Hyl}}/C_{\text{residual Lys}} \approx 154$) we achieved comparable peak intensities for what we attribute to all possible combinations of lysine/hydroxylysine (deGFP, deGFP1Hyl, deGFP2Hyl, ..., deGFP18Hyl) (Fig. 5). Even if the ratio of incorporated lysine to hydroxylysine can be controlled statistically, there is a huge spread in the distribution, and we were not able to achieve the production of proteins with a defined ratio. The peak corresponding to full replacement of lysine exhibited very low intensity. Further increase of the concentration of hydroxylysine to 50 mM ($C_{\text{Hyl}}/C_{\text{residual Lys}} \approx 458$) led to a spectrum showing different lysine/hydroxylysine combinations but with lower signal-to-noise ratio (figure S 2). We understand that high concentrations of hydroxylysine diminish the probability of protein production. This was confirmed by SDS-PAGE

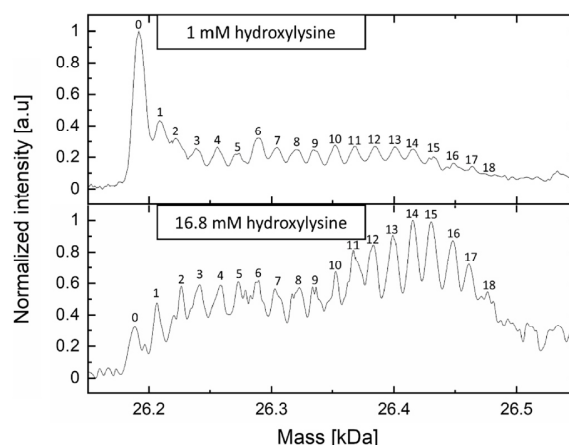


Fig. 5. Mass spectra from deGFP as obtained at two different concentrations of hydroxylysine (insert) supplemented to a virgin cell-free expression system. The spectra were recorded after 16 h of incubation at 29 °C. The numbers above the peaks indicate the number n of incorporated hydroxylysines. The number of incorporated lysines is $18-n$. The mass spectrum for a concentration of 1 mM of hydroxylysine (top) exhibits a most intense peak from deGFP containing 18 lysines ($M_{\text{deGFP}} = 26.193$ kDa). This lysine corresponds to residual lysine from the virgin extract. Lysine was present at much lower concentrations than hydroxylysine ($C_{\text{Hyl}}/C_{\text{residual Lys}} \approx 9$, where C_{Hyl} corresponds to the amount of added hydroxylysine, $C_{\text{residual Lys}}$ determined as previously described²⁹). All peaks corresponding to 1, 2, 3, ..., 17, (18) lysines replaced by hydroxylysine ($M_{\text{Hyl}} - M_{\text{Lys}} = 16$ Da; $M_{\text{deGFPnHyl}} = M_{\text{deGFP}} + n \cdot 16$ Da, $n = 1, \dots, 18$) were also present, albeit with clearly lower intensity. For higher concentrations of hydroxylysine (16.8 mM), the probability of hydroxylysine incorporation into deGFP is seen to increase.

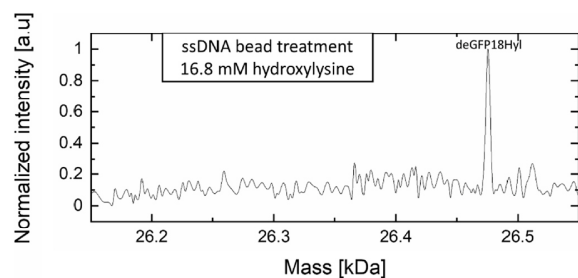


Fig. 6. Mass spectrum of deGFP produced from a cell-free expression system depleted in lysine, containing 16.8 mM of hydroxylysine. The spectrum was recorded after 16 h of incubation at 29 °C. We observed a peak at $M = 26.476$ kDa that corresponds to deGFP with 18 hydroxylysines instead of lysines ($M_{\text{deGFP18Hyl}} = 26.481$ kDa, the difference between the measured and the literature value of 5 Da for $M_{\text{deGFP18Hyl}}$ was likely due to deconvolution errors caused by the background noise).

(Fig. 7). Our observation agrees well with the fact that for each batch of the expression system, the concentrations of amino acids need to be adjusted to optimize protein production rates. Moreover, at higher concentrations of hydroxylysine ($c_{\text{Hyl}} > 50$ mM), deGFP related peaks vanished from the mass spectrum (data not shown). We conclude that elevated concentrations of hydroxylysine are inhibitory to gene expression. To achieve full incorporation, it is essential to decrease the amount of residual lysine within the cell-free expression system.

For this reason, we added hydroxylysine to a lysine depleted expression system. At a concentration of 16.8 mM ($c_{\text{Hyl}}/c_{\text{residual Lys}} \approx 665$), after 16 h of gene expression, we observed the presence of hydroxylysine only (Fig. 6).

Peaks representing deGFP species with different amounts of hydroxylysine and lysine were hardly visible, indicating the efficient incorporation of hydroxylysine into deGFP. Further increase of the hydroxylysine concentration entailed the absence of any related peaks (data not shown). We understand that high levels of hydroxylysine are, again, inhibitory.

The noise in the mass spectra was caused by GamS, present in the sample, interfering with deconvolution. Therefore, we performed SDS-PAGE to strengthen our results (Fig. 7).

Although the mass difference of 288 Da between native deGFP and deGFP18Hyl is difficult to detect by SDS-PAGE, we see that the deGFP indicating band in lane 2 (Dep Hyl) is positioned slightly higher than the band in lane 3 (Lys) strengthening the results presented in Fig. 6. Comparing the intensities of the corresponding bands, we observe that the amount of deGFP produced per cell-free expression reaction was much higher than the amount of deGFP18Hyl, produced under otherwise identical conditions.

Our method for lysine removal also depletes residual arginine. In the case of arginine, we observed similar kinetics as for lysine (see figures S 1A and S 1B). Our work represents a first step towards the incorporation of arginine surrogates as well.

4. Conclusion

In this paper we presented a method to fully replace lysine by hydroxylysine in green fluorescent protein. The expression system was designed to produce proteins just by adding DNA. It seems reasonable to assume that if our method works for GFP, it will also work for many other proteins. However, verification will be mandatory if one plans to realize a dedicated construct. To the best of our knowledge, the full replacement of lysine by hydroxylysine in proteins had not been achieved previously. Our method is based on the observation that freshly added lysine is incorporated earlier than residual lysine, which is incorporated only after a lag of a few hours (Fig. 1). Consequently,

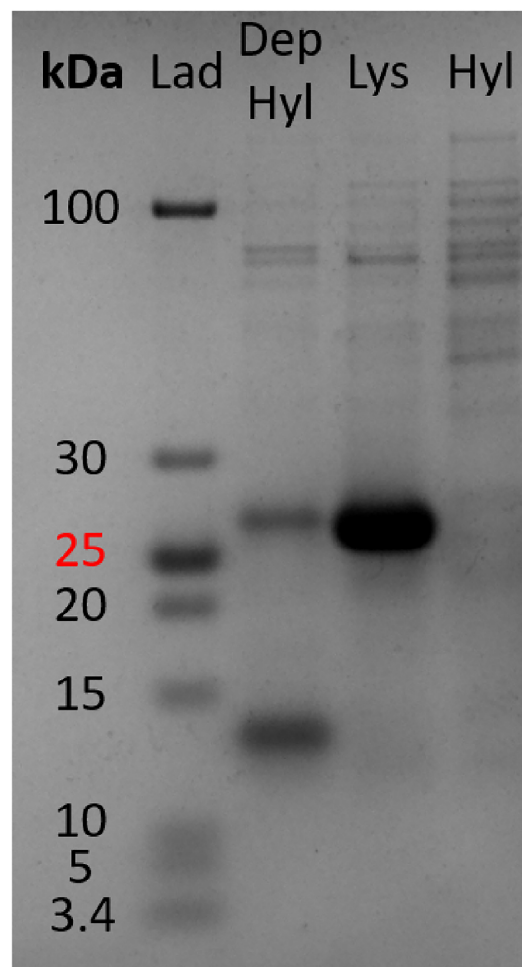


Fig. 7. SDS-PAGE of deGFP extracted from different preparations of the cell-free expression system after 16 h of incubation at 29 °C. The first lane (Lad) shows the ladder (PageRuler™ Unstained Low Range Protein Ladder, Thermo Scientific). The numbers left to this lane indicate the weight of the bands in kDa. The second lane (Dep Hyl) results from a lysine depleted cell-free expression system containing 16.8 mM of hydroxylysine. The third lane (Lys) is from a cell-free expression system containing 1 mM lysine and the fourth lane (Hyl) results from a cell-free expression system containing 50 mM hydroxylysine. In lane 2 (Dep Hyl) two bands indicate deGFP18Hyl ($M_{\text{deGFP18Hyl}} = 26.481$ kDa) and GamS ($M_{\text{GamS}} = 12.634$ kDa), while lane 3 (Lys) contains a band indicating deGFP ($M_{\text{deGFP}} = 26.193$ kDa). Lane 4 (Hyl) shows no bands indicating deGFP production (high molecular weight bands are residues from the cell-free expression system). This confirms our assumption that high concentrations of hydroxylysine diminish the probability of protein production.

protein synthesis can be stopped sufficiently early (after 2 h) to prevent the incorporation of residual lysine. We showed that further depletion of residual lysine avoids its incorporation even at a later stage of protein synthesis (16 h), increasing the yield. We believe that the here proposed method represents a powerful tool that can be useful for the incorporation of other lysine and arginine surrogates as well.

Author contributions

AO and MF designed the research. MF performed the research with the help of SR and FG. MF analyzed data with the help of AO, SR and PH.

MF and AO wrote the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the collaborative research center SFB 1027 funded by the Deutsche Forschungsgemeinschaft (DFG), and by the Human Frontier Science Program (HFSP, RGP0037/2015). We thank Emanuel Worst for help and fruitful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bmc.2021.116207>.

References

- Ulmer K. Protein engineering. *Science*. 1983;219:666–671. <https://doi.org/10.1126/science.6572017>.
- Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol*. 1992; 224:771–782. [https://doi.org/10.1016/0022-2836\(92\)90561-W](https://doi.org/10.1016/0022-2836(92)90561-W).
- Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol*. 1996;6: 178–182. [https://doi.org/10.1016/S0960-9822\(02\)00450-5](https://doi.org/10.1016/S0960-9822(02)00450-5).
- Tan SC, Yip BC. DNA, RNA, and protein extraction: the past and the present. *J Biomed Biotechnol*. 2009;2009:1–10. <https://doi.org/10.1155/2009/574398>.
- Chen R. Bacterial expression systems for recombinant protein production: E. coli and Beyond. *Biotechnol Adv*. 2012;30:1102–1107. <https://doi.org/10.1016/j.biotechadv.2011.09.013>.
- Fang G-M, Li Y-M, Shen F, et al. Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chemie*. 2011;50:7645–7649. <https://doi.org/10.1002/anie.201100996>.
- Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V. Protocols for implementing an Escherichia Coli based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp*. 2013;9:1–14. <https://doi.org/10.3791/50762>.
- Spicer CD, Davis BG. Selective chemical protein modification. *Nat Commun*. 2014;5: 1–14. <https://doi.org/10.1038/ncomms5740>.
- Rashidian M, Dozier JK, Distefano MD. Enzymatic labeling of proteins: techniques and approaches. *Bioconjug Chem*. 2013;24:1277–1294. <https://doi.org/10.1021/bc400102w>.
- Zoller MJ. New molecular biology methods for protein engineering. *Curr Opin Biotechnol*. 1991;2:526–531. [https://doi.org/10.1016/S0959-440X\(05\)80084-3](https://doi.org/10.1016/S0959-440X(05)80084-3).
- Lu Y. Cell-free synthetic biology: engineering in an open world. *Synth Syst Biotechnol*. 2017;2:23–27. <https://doi.org/10.1016/j.synbio.2017.02.003>.
- Muir TW, Sondhi D, Cole PA. Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA*. 1998;95:6705–6710. <https://doi.org/10.1073/pnas.95.12.6705>.
- Gavrilescu M, Chisti Y. Biotechnology - a sustainable alternative for chemical industry. *Biotechnol Adv*. 2005;23:471–499. <https://doi.org/10.1016/j.biotechadv.2005.03.004>.
- Shin J, Noireaux V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *J Biol Eng*. 2010;4:1–9. <https://doi.org/10.1186/1754-1611-4-8>.
- Schenkelberger M, Shanak S, Finkler M, et al. Expression regulation by a methyl-CpG binding domain in an E. coli based, cell-free TX-TL system. *Phys Biol*. 2017;14:1–11. <https://doi.org/10.1088/1478-3975/aa5d37>.
- Finkler M, Ott A. Bead-based assay for spatiotemporal gene expression control in cell-free transcription-translation systems. *Biotechniques*. 2019;66:29–33. <https://doi.org/10.2144/btn-2018-0097>.
- Worst EG, Exner MP, de Simone A, et al. Residue-specific incorporation of noncanonical amino acids into model proteins using an Escherichia coli cell-free transcription-translation system. *J Vis Exp*. 2016;114:1–14. <https://doi.org/10.3791/54273>.
- Singh-Blom A, Hughes RA, Ellington AD. An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J Biotechnol*. 2014;178:12–22. <https://doi.org/10.1016/j.jbiotec.2014.02.009>.
- Johnson JA, Lu YY, Van Deventer JA, Tirrell DA. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr Opin Chem Biol*. 2010;14:774–780. <https://doi.org/10.1016/j.cbpa.2010.09.013>.
- Sakamoto K, Hayashi A, Sakamoto A, et al. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res*. 2002;30: 4692–4699. <https://doi.org/10.1093/nar/gkf589>.
- Goerke AR, Swartz JR. High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnol Bioeng*. 2009;102:400–416. <https://doi.org/10.1002/bit.22070>.
- Wan W, Huang Y, Wang Z, et al. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia Coli. *Angew Chemie*. 2010;49:3211–3214. <https://doi.org/10.1002/anie.201000465>.
- Gao W, Cho E, Liu Y, Lu Y. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Front Pharmacol*. 2019;10:1–8. <https://doi.org/10.3389/fphar.2019.00611>.
- Yang B, Ayyadurai N, Yun H, et al. In vivo residue-specific dopa-incorporated engineered mussel bioglut with enhanced adhesion and water resistance. *Angew Chemie*. 2014;53:13360–13364. <https://doi.org/10.1002/anie.201406099>.
- Yoo TH, Tirrell DA. High-throughput screening for methionyl-TRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. *Angew Chemie*. 2007;46:5340–5343. <https://doi.org/10.1002/anie.200700779>.
- Worst EG, Exner MP, De Simone A, et al. Cell-free expression with the toxic amino acid canavanine. *Bioorg Med Chem Lett*. 2015;25:3658–3660. <https://doi.org/10.1016/j.bmcl.2015.06.045>.
- Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci*. 1974;71: 4135–4139. <https://doi.org/10.1073/pnas.71.10.4135>.
- Ninio J. Kinetic amplification of enzyme discrimination. *Biochimie*. 1975;57: 587–595. [https://doi.org/10.1016/S0300-9084\(75\)80139-8](https://doi.org/10.1016/S0300-9084(75)80139-8).
- Finkler M, Kurt Ö, Grimm F, Hartz P, Ott A. A bead-based method for the removal of the amino acid lysine from cell-free transcription-translation systems. *J Biotechnol X*. 2020;8:1–6. <https://doi.org/10.1016/j.btecx.2020.100024>.
- Dunn JA, McCance DR, Thorpe SR, Lyons TJ, Baynes JW. Age-Dependent Accumulation of Nε-(Carboxymethyl)Lysine and Nε-(Carboxymethyl)Hydroxyllysine in Human Skin Collagen. *Biochemistry*. 1991;30:1205–1210. <https://doi.org/10.1021/bi00219a007>.
- Cudic M, Lauer-Fields JL, Fields GB. Improved synthesis of 5-hydroxylysine (Hyl) derivatives. *J Pept Res*. 2005;65:272–283. <https://doi.org/10.1111/j.1399-3011.2005.00215.x>.
- Cudic M, Fields GB. Modulation of receptor binding to collagen by glycosylated 5-hydroxylysine: chemical biology approaches made feasible by Carpino's Fmoc Group. *Pept Sci*. 2020;112:1–12. <https://doi.org/10.1002/pep2.24156>.
- Krane SM, Pinnell SR, Erbe RW. Lysyl-protocollagen hydroxylase deficiency in fibroblasts from siblings with hydroxylysine-deficient collagen. *Proc Natl Acad Sci USA*. 1972;69:2899–2903. <https://doi.org/10.1073/pnas.69.10.2899>.
- Burkhardt A, Grupe G. Hydroxylysine deficiency, conspicuous skeletal lesions and a strange burial practice in a historical german island population. *Int J Osteoarchaeol*. 2018;28:227–236. <https://doi.org/10.1002/oa.2649>.
- Gjaltema RAF, Bank RA. Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. *Crit Rev Biochem Mol Biol*. 2017;52:74–95. <https://doi.org/10.1080/10409238.2016.1269716>.
- Popenoe EA, Aronson RB, Van Slyke DD. Hydroxylysine formation from lysine during collagen biosynthesis. *Proc Natl Acad Sci*. 1966;55:393–397. <https://doi.org/10.1073/pnas.55.2.393>.
- Makris EA, MacBarb RF, Responde DJ, Hu JC, Athanasiou KA. A copper sulfate and hydroxylysine treatment regimen for enhancing collagen cross-linking and biomechanical properties in engineered neocartilage. *FASEB J*. 2013;27:2421–2430. <https://doi.org/10.1096/fj.12.224030>.
- Deane CS, Bass JJ, Crossland H, Phillips BE, Atherton PJ. Animal, plant, collagen and blended dietary proteins: effects on musculoskeletal outcomes. *Nutrients*. 2020;12: 1–35. <https://doi.org/10.3390/nu12092670>.
- Maisani M, Ziane S, Ehret C, et al. A new composite hydrogel combining the biological properties of collagen with the mechanical properties of a supramolecular scaffold for bone tissue engineering. *J Tissue Eng Regen Med*. 2018;12:e1489–e1500. <https://doi.org/10.1002/term.2569>.
- Ma L, Gao C, Mao Z, et al. Collagen/chitosan porous scaffolds with improved biostability for skin tissue engineering. *Biomaterials*. 2003;24:4833–4841. [https://doi.org/10.1016/S0142-9612\(03\)00374-0](https://doi.org/10.1016/S0142-9612(03)00374-0).
- Gerrits M, Budisa N, Merk H. Site-specific chemoselective pyrrolysine analogues incorporation using the cell-free protein synthesis system. *ACS Synth Biol*. 2019;8: 381–390. <https://doi.org/10.1021/acssynbio.8b00421>.
- Agostini F, Völler JS, Koksich B, Acevedo-Rocha CG, Kubyshev V, Budisa N. Biocatalysis with unnatural amino acids: enzymology meets xenobiology. *Angew Chemie - Int Ed*. 2017;56:9680–9703. <https://doi.org/10.1002/anie.201610129>.
- Invitrogen. Manual for Dynabeads™ MyOne™ Streptavidin T1; 2017.
- Ambion. Manual for Dynabeads® Oligo (DT)25; 2012.
- Garamella J, Marshall R, Rustad M, Noireaux V. The all E. coli TX-TL Toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth Biol*. 2016;5:344–355. <https://doi.org/10.1021/acssynbio.5b00296>.
- Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM. Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system. *ACS Synth Biol*. 2014;3:387–397. <https://doi.org/10.1021/sb400131a>.
- Ong S-E, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002;1:376–386. <https://doi.org/10.1074/mcp.M200025-MCP200>.
- Ong S-E, Kratchmarova I, Mann M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res*. 2003;2: 173–181. <https://doi.org/10.1021/pr0255708>.
- Ong S-E. The expanding field of SILAC. *Anal Bioanal Chem*. 2012;404:967–976. <https://doi.org/10.1007/s00216-012-5998-3>.

Supplementary

Full incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in green fluorescent protein.

Finkler, M., Ravanbodshirazi, S., Grimm, F., Hartz, P., and Ott, A.

Bioorganic & Medicinal Chemistry (2021), *41*, 1–6.

DOI: 10.1016/j.bmc.2021.116207

<https://doi.org/10.1016/j.bmc.2021.116207>

Reproduced from *Bioorganic & Medicinal Chemistry*. (2021) 41, 1–6 with permission of Elsevier.

Supplementary

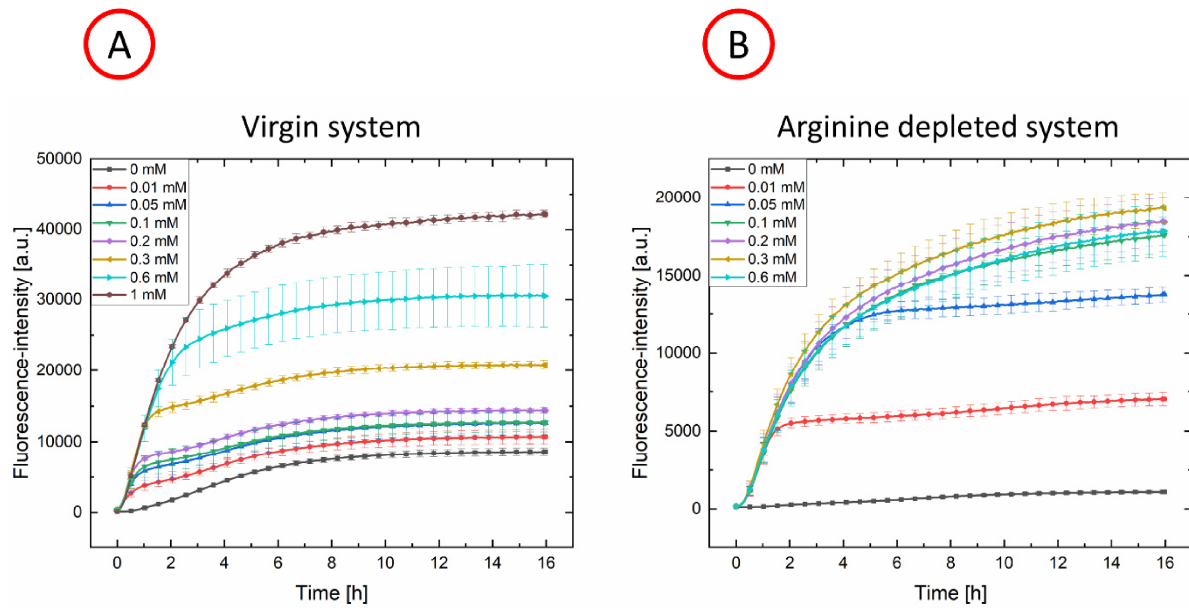


Figure S 1 Time course of the fluorescence-intensity of green fluorescent protein (deGFP) in our cell-free expression system for different concentrations of freshly added arginine (insert). The experiment was performed using a virgin system (A, residual arginine present) as well as an arginine depleted system (B, amount of residual arginine reduced by treatment with ssDNA coated beads).

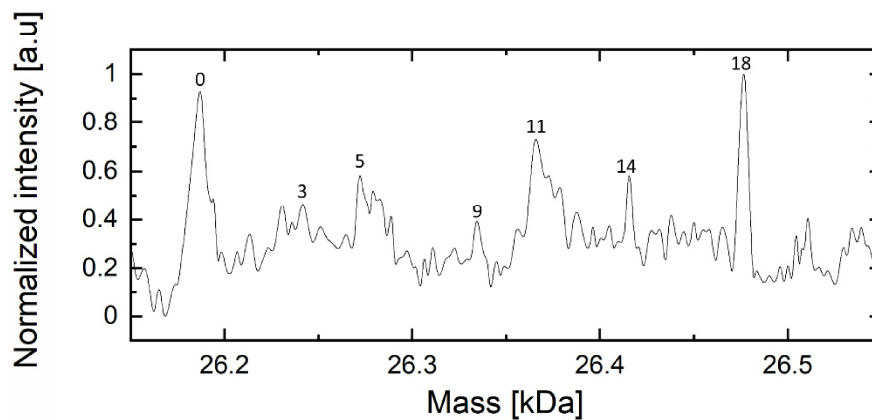


Figure S 2 Mass spectra of deGFP produced by a virgin cell-free expression system supplemented with 50 mM of hydroxylysine. The spectrum was recorded after 16 h of incubation at 29 °C. Even if the signal-to-noise ratio is low, we still distinguished a peak from deGFP with 18 hydroxylysines ($M_{\text{deGFP18Hyl}} = 26.481$ kDa) and a peak from wild type deGFP with 18 lysines. Because of the presence of many other species (deGFPnHyl, $n = 3, 5, 9, 11, 14$) with low intensities, it was difficult to draw more quantitative conclusions. The numbers above the peaks indicate the number of incorporated hydroxylysines.

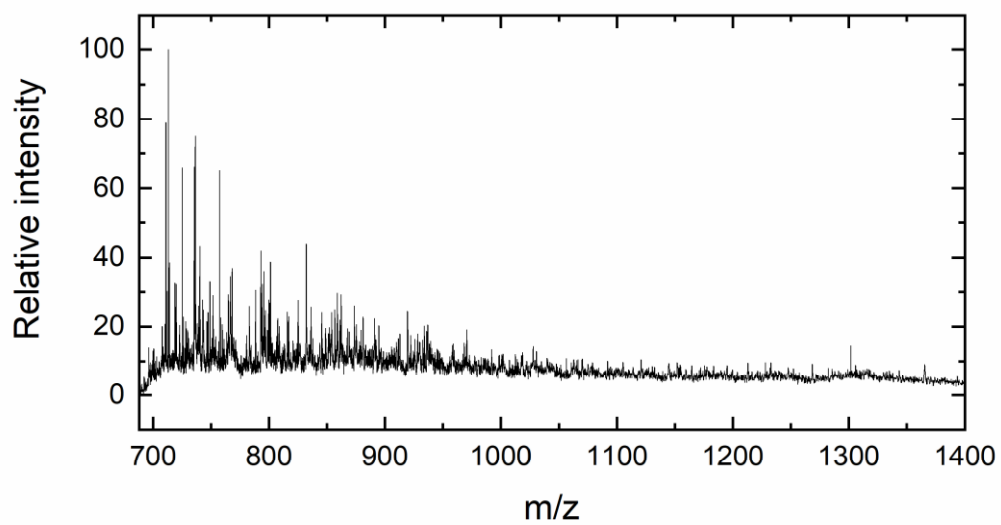


Figure S 3 Raw data leading to the deconvoluted spectrum shown in figure S 2 (50 mM hydroxylysine).

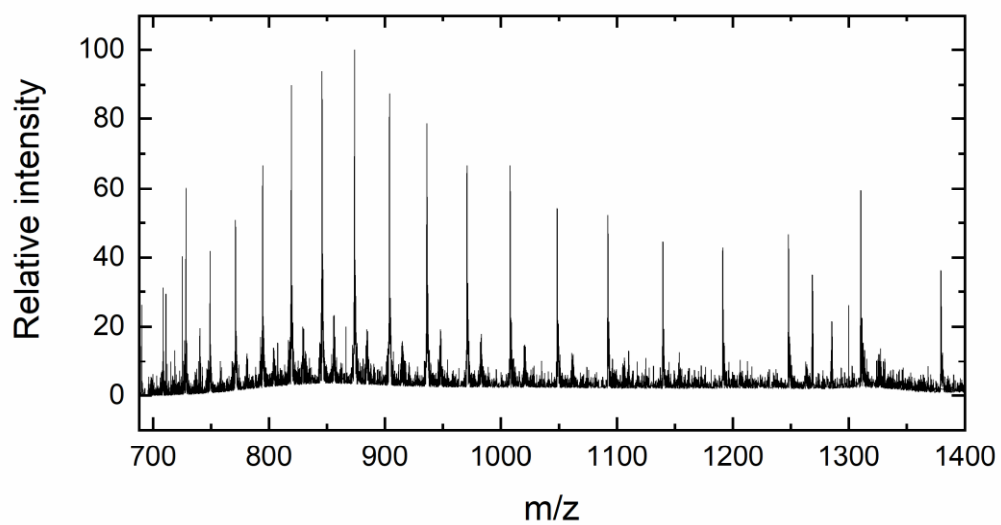


Figure S 4 Raw data leading to the deconvoluted spectrum shown in figure 3, top, "0 mM of isotopic lysine (16 h of incubation)".

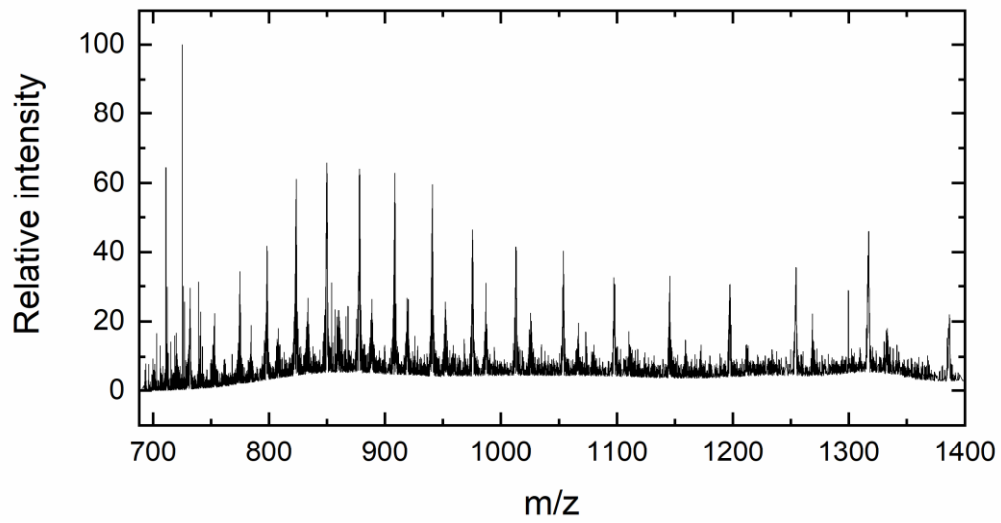


Figure S 5 Raw data leading to the deconvoluted spectrum shown in figure 3, second from top, "0.1 mM of isotopic lysine (2 h of incubation)".

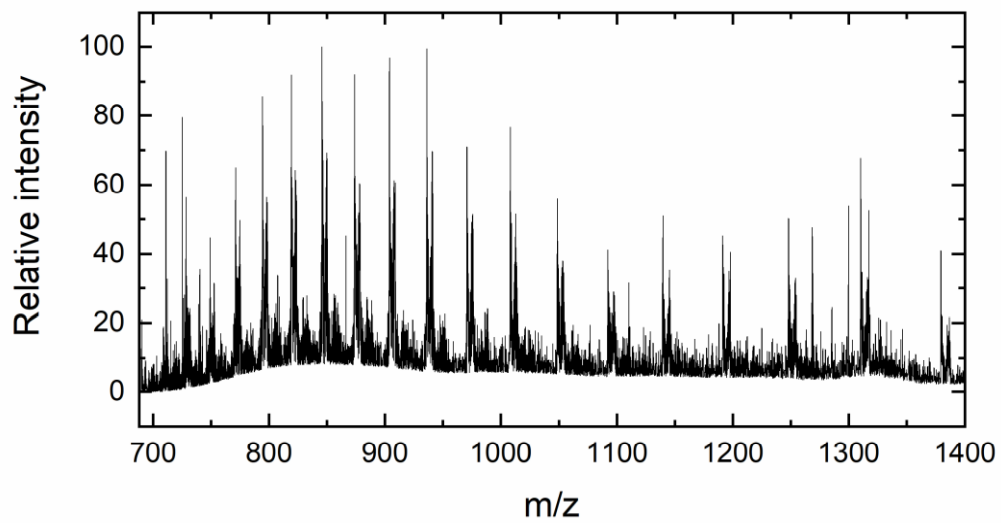


Figure S 6 Raw data leading to the deconvoluted spectrum shown in figure 3, third from top, "0.1 mM isotopic lysine (16 h incubation)".

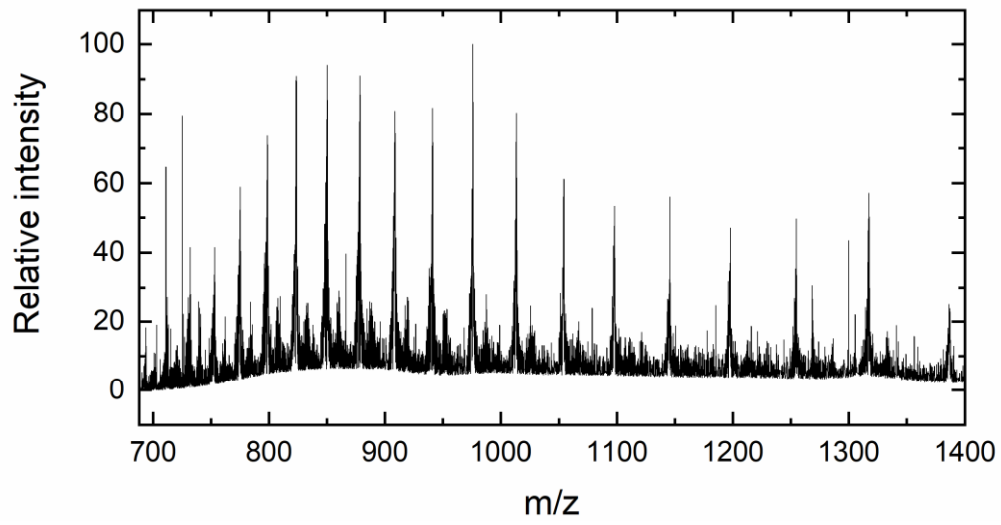


Figure S 7 Raw data leading to the deconvoluted spectrum shown in figure 3, bottom spectrum, "0.6 mM isotopic lysine (16 h of incubation)".

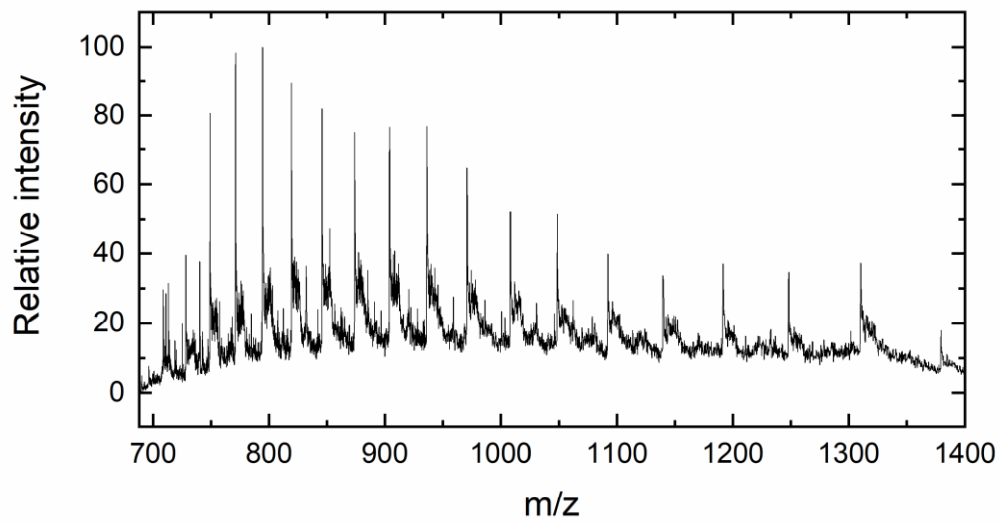


Figure S 8 Raw data leading to the deconvoluted spectrum shown in figure 5, top, "1 mM of hydroxylysine".

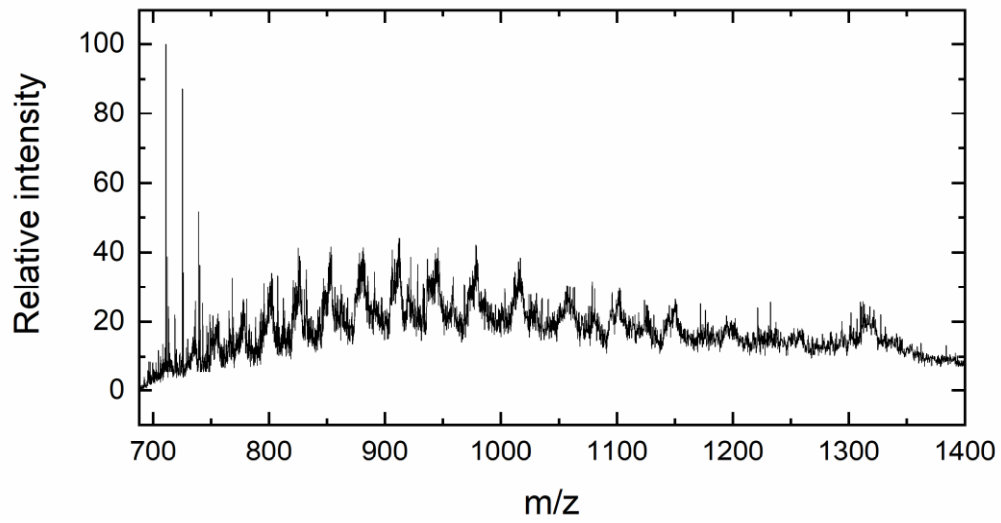


Figure S 9 Raw data leading to the deconvoluted spectrum shown in figure 5, bottom, "16.8 mM of hydroxylysine".

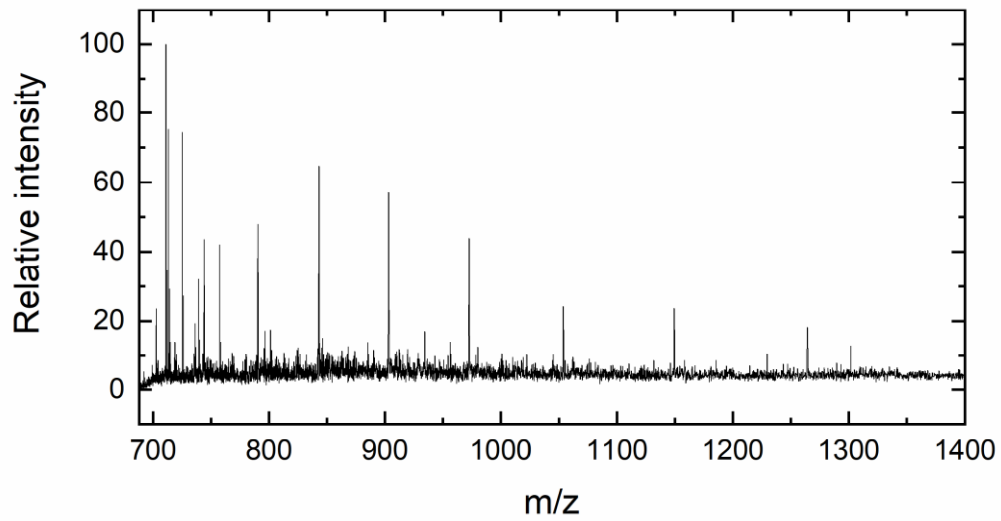


Figure S 10 Raw data leading to the deconvoluted spectrum shown in figure 6, "ssDNA bead treatment, 16.8 mM hydroxylysine".

4 Diskussion

In diesem Abschnitt werden die Originalpublikationen, auf denen die vorliegende Dissertation basiert, diskutiert. Hierbei wird zunächst der wissenschaftliche Hintergrund der jeweiligen Publikation erörtert, bevor die erhaltenen Ergebnisse betrachtet werden. Anschließend folgen eine kurze Zusammenfassung und ein Ausblick.

4.1 Schenkelberger *et al.* (2017)

4.1.1 Hintergrund

In Schenkelberger *et al.* (2017) wurde die Rolle der Cytosin-Methylierung in der eukaryotischen epigenetischen Genregulation näher untersucht. Diese kann in Säugetieren an Cytosinen in der DNA-Sequenz „CpG“ stattfinden (5-Methylcytosin) ^{33,90,360,361}. Die so genannte Methyl-CpG Bindedomäne (MBD) ist in diversen eukaryotischen Transkriptions-Regulatoren für deren Bindung an methylierte CpGs verantwortlich ^{200,362–365}. So verfügt auch der humane Transkriptions-Regulator MeCP2 über diese Bindedomäne ^{200,362–365}. Der Promotor III des „Brain-derived neurotrophic factor“ (BDNF), welcher fortan als BDNF-Promotor bezeichnet wird, verfügt über ein einzelnes zentrales CpG-Motiv und weist dadurch im methylierten Zustand von CpG eine hohe Affinität zu MeCP2 auf ^{43–46,366}. Die Bindung von MeCP2 an den BDNF-Promotor geschieht über eine Interaktion zwischen MBD und der Methylgruppe des CpG-Motivs, welche einen „Fußabdruck“ von 12 Basen (CpG in der Mitte) auf der DNA hinterlässt ³⁶⁷.

Da MBD eine maßgebliche Rolle bei der Interaktion zwischen MeCP2 und dem BDNF-Promotor spielt ^{44,362,365,367}, sollte MBD auf transkriptions-regulatorische Eigenschaften in Verbindung mit dem BDNF-Promotor untersucht werden. Um zu untersuchen, ob diese Regulation Organismus-spezifisch ist, wurden die entsprechenden Untersuchungen in einem auf *E. coli* basierenden, zellfreien Expressionssystem durchgeführt.

4.1.2 Ergebnisse

In den Experimenten wurde ein Plasmid verwendet, mit dem eine Expression von *degfp* unter der Kontrolle des BDNF-Promotors im zuvor erwähnten prokaryotischen, zellfreien Expressionssystem durchgeführt werden konnte. Dies wurde ermöglicht, indem der BDNF-Promotor im verwendeten Plasmid zwischen die Sigma-Faktor σ^{70} spezifischen Konsensussequenzen (–35- und –10-Element) von *E. coli* Promotoren, die mit Haushaltsgenen assoziiert sind, inseriert wurde ⁴⁹. In Verbindung mit der RNA-Polymerase initiiert σ^{70} als Holoenzym die Transkription ^{50,368,369}. Das verwendete Plasmid konnte

mit methyliertem oder mit unmethyliertem BDNF-Promotor dem Expressionssystem hinzugefügt werden. Durch eine zusätzliche Hinzugabe von MBD konnte dessen Einfluss auf die Expression von *degfp* durch Messung der Fluoreszenz-Intensität von deGFP untersucht werden (siehe Schenkelberger *et al.* (2017), Abbildung 1).

MBD wurde in den Experimenten nicht direkt in rekombinanter Form hinzugefügt, sondern im Rahmen einer Koexpression synthetisiert. Während die Konzentration des deGFP-codierenden Plasmids (mit methyliertem oder unmethyliertem BDNF-Promotor) im Expressionssystem konstant war (5 nM), wurde die Konzentration des hinzugegebenen MBD-codierenden Plasmids variiert (0 nM - 1 nM). Nach der Hinzugabe aller benötigten Komponenten wurde eine Genexpressionsreaktion über 16 h bei 29 °C durchgeführt und anschließend die Fluoreszenz-Intensität des synthetisierten deGFP gemessen.

Anhand der Messungen konnte gezeigt werden, dass lediglich im Falle eines methylierten BDNF-Promotors die Expression von *degfp* effektiv durch MBD reprimiert wurde. Erhöhte Konzentrationen an synthetisiertem MBD, welche durch erhöhte Konzentrationen am MBD-codierenden Plasmid erreicht werden konnten (²⁸ zeigt die Korrelation zwischen Plasmid-Konzentration und synthetisiertem Protein), hatten eine stärkere Repression zur Folge (siehe Schenkelberger *et al.* (2017), Abbildung 2). Somit konnte gezeigt werden, dass MBD in Verbindung mit dem BDNF-Promotor als Transkriptions-Regulator fungieren kann. Des Weiteren wurde gezeigt, dass diese Regulation nicht Organismus-spezifisch ist. Nach Gewinnung dieser neuen Erkenntnisse wurde der Genregulationsmechanismus weiter untersucht.

Da an der Interaktion zwischen MBD und dem BDNF-Promotor maßgeblich das zentrale, am Cytosin methylierte CpG-Motiv beteiligt ist, stellte sich die Frage, ob die benachbarten DNA-Sequenzen einen Einfluss auf diese Interaktion haben, wie dies für MeCP2 der Fall ist ^{44,45,362,365,367,370,371}. In *in vitro* DNase-Fußabdruck-Untersuchungen konnte bereits gezeigt werden, dass durch die Bindung von MBD an ein am Cytosin methyliertes CpG-Motiv die jeweils 5 benachbarten Basen „blockiert“ werden ³⁶⁷. Somit können auch diese Basen mit MBD interagieren. Um dies genauer zu untersuchen, wurden verschiedene Mutanten des nativen BDNF-Promotors generiert, welche mutierte Basen in unmittelbarer Nähe des zentralen CpGs besitzen (siehe Schenkelberger *et al.* (2017), Tabelle 1). Diese wurden als alternative Promotoren für die Expression von *degfp* in analogen Experimenten verwendet.

Dabei konnte festgestellt werden, dass die Mutationen die absolute Expression von *degfp* beeinflussten. Dennoch konnte bei allen Mutanten, in Analogie zum nativen BDNF-Promotor, eine effektive Repression der *degfp*-Expression nur im Falle einer Methylierung des Promotors und erhöhten MBD-Konzentrationen erreicht werden (siehe Schenkelberger *et al.* (2017), Abbildung 3). In diesem Zusammenhang wurde die Fluoreszenz-Intensität von synthetisiertem deGFP bei verschiedenen MBD-Konzentrationen auf die Fluoreszenz-Intensität von synthetisiertem deGFP in Abwesenheit von MBD normiert. Anhand der normierten Werte konnte für alle Mutanten und den nativen BDNF-Promotor gezeigt werden, dass die durch MBD verursachte Repression der *degfp*-Expression in allen Fällen

vergleichbar war (siehe Schenkelberger *et al.* (2017), Abbildung 4). Somit konnte durch Mutationen der von CpG benachbarten Basen die absolute Genexpression beeinflusst werden, nicht jedoch die relative Repression in Abhängigkeit der MBD-Konzentration.

Im Anschluss sollte gezeigt werden, ob das beobachtete Verhalten durch ein einfaches kinetisches Modell für das Binden von MBD und/oder den Sigma-Faktor σ^{70} an den methylierten BDNF-Promotor beschrieben werden konnte.

Diesbezüglich wurde ein Modell entwickelt, welches davon ausgeht, dass zunächst MBD und/oder σ^{70} mit unterschiedlichen Raten an den BDNF-Promotor binden können. Hierbei bindet MBD mit der Rate k_M^{on} an den BDNF-Promotor. Diese ist für alle Mutanten gleich, da die Bindung maßgeblich zwischen dem zentralen methylierten CpG und MBD erfolgt^{44,362,365,367}. Im Gegensatz dazu ist die Rate k_σ^{on} , mit der σ^{70} an den BDNF-Promotor bindet, von der gesamten Sequenz des Promotors abhängig und somit für unterschiedliche absolute Genexpressionslevels der verschiedenen Mutanten verantwortlich. Der Einfluss der Promotor-Sequenz auf die Affinität zwischen Promotor und RNA-Polymerase wurde bereits in der Literatur beschrieben und lässt eine analoge Annahme bzgl. σ^{70} zu^{197–199,368}. Des Weiteren hängt k_σ^{on} davon ab, ob am BDNF-Promotor bereits MBD gebunden wurde. Ist dies der Fall, so verringert sich k_σ^{on} durch sterische Hinderungen ($k_{\sigma,-MBD}^{on} \gg k_{\sigma,+MBD}^{on}$). Die Ergebnisse der zuvor erwähnten *in vitro* DNase-Fußabdruck-Untersuchungen³⁶⁷ unterstützen diese Vermutung. Dadurch, dass MBD und/oder σ^{70} an den BDNF-Promotor binden können, existieren für diesen vier mögliche Zustände: BDNF ohne Bindeelemente (1), BDNF mit MBD (2), BDNF mit σ^{70} (3) und BDNF mit MBD und σ^{70} (4) (siehe Schenkelberger *et al.* (2017), Abbildung 5).

Eine Expression von *degfp* erfolgt lediglich in den Zuständen (3) und (4), in denen σ^{70} gebunden ist. Hierbei kann die Gesamtexpression E als Superposition der Expression aus Zustand (3) (kein MBD gebunden) E_- und der Expression aus Zustand (4) (MBD gebunden) E_+ geschrieben werden, wobei $E_- > E_+$ (siehe Gleichung (7)).

$$E = \frac{[(3)]}{[(3)] + [(4)]} E_- + \frac{[(4)]}{[(3)] + [(4)]} E_+ \quad (7)$$

In der Gleichung repräsentiert $[(x)]$ die Konzentration des Promotors im Zustand x .

Wird berücksichtigt, dass ein Übergang zwischen den Zuständen (3) und (4) durch Assoziation bzw. Dissoziation von MBD erfolgen kann, lässt sich die Dissoziationskonstante von MBD $K_{D,M}$ als weitere Größe einführen (siehe Gleichung (8)).

$$K_{D,M} = \frac{[(3)][MBD]}{[(4)]} \quad (8)$$

Mit Gleichung (7) und Gleichung (8) kann somit ein neuer Ausdruck für die Gesamtexpression gefunden werden, bei dem die Konzentration des Promotors im Zustand x nicht benötigt wird und der eine Abhängigkeit von der MBD-Konzentration $[MBD]$ aufweist (siehe Gleichung (9)).

$$E = E_+ + \frac{(E_- - E_+) \cdot K_{D,M}}{K_{D,M} + [MBD]} \quad (9)$$

Die erhaltene Gleichung (9) wurde für die Anpassung der experimentellen Daten verwendet. Hierbei konnte festgestellt werden, dass die experimentellen Daten für alle untersuchten BDNF-Varianten (nativ und mutiert) mit Gleichung (9) angepasst werden konnten (siehe Schenkelberger *et al.* (2017), Abbildung 6 und Abbildung S8). Somit konnte das vorgeschlagene einfache kinetische Modell bestätigt werden.

Für eine weitere Untersuchung des Genregulationsmechanismus von MBD in Verbindung mit dem BDNF-Promotor wurden Molekuldynamik-Simulationen durchgeführt (siehe Schenkelberger *et al.* (2017), Abbildung 7 - 10). Hierbei konnte gezeigt werden, dass durch die Bindung von MBD an die methylierte BDNF-Promotor-DNA deren große Furche im Bereich der Bindestelle von ca. 15 Å auf ca. 20 Å vergrößert wird. Liegt der BDNF-Promotor in der unmethylierten Form vor, so fällt diese Vergrößerung geringer aus. Somit konnte gezeigt werden, dass Konformationsänderungen der DNA in den Genregulationsmechanismus von MBD (und vermutlich auch MeCP2) in Verbindung mit dem BDNF-Promotor involviert sind. Da die maßgebliche Interaktion zwischen MBD und dem BDNF-Promotor lediglich am zentralen CpG-Motiv stattfindet, die Konformationsänderungen sich aber über mehrere Basenpaare erstrecken, ist dies ein Indiz für kooperative Effekte. Eine solche Kooperativität kann eine mögliche Erklärung für die hohe Spezifität von MBD gegenüber dem BDNF-Promotor darstellen.

4.1.3 Zusammenfassung und Ausblick

In Schenkelberger *et al.* (2017) wurde eine durch MBD verursachte Genregulation in Verbindung mit dem BDNF-Promotor in einem auf *E. coli* basierenden, zellfreien Expressionssystem beschrieben. Daraus resultierte der experimentelle Nachweis, dass MBD, welches Bestandteil diverser eukaryotischer Transkriptions-Regulatoren ist^{200,362–365}, auch alleine als Transkriptions-Regulator fungieren kann. Des Weiteren konnte dadurch, dass ein prokaryotisches, zellfreies Expressionssystem für die durchgeführten Experimente verwendet wurde, erstmals ein eukaryotischer epigenetischer Genregulationsmechanismus in einem solchen System realisiert werden. Dies zeigte, dass der zugrunde liegende Regulationsmechanismus nicht Organismus-spezifisch ist. Des Weiteren konnten Hinweise gefunden werden, dass die durch MBD verursachte Genregulation auf DNA-Konformationsänderungen des BDNF-Promotors zurückzuführen ist.

4 Diskussion

Diese Erkenntnisse könnten in zukünftigen Experimenten eine wichtige Rolle spielen, da hiermit die Generierung neuer universeller Genregulationsmechanismen ermöglicht wird, die nicht an einen Organismus gebunden sind ^{29,372–375}. Der Nachweis, dass MBD transkriptions-regulatorische Eigenschaften aufweist, könnte eine neue Betrachtungsweise der Genregulationsmechanismen von MBD-assoziierten Transkriptions-Regulatoren, wie z. B. MeCP2, zur Folge haben.

Die Transkriptionsregulation durch MeCP2, welches über MBD verfügt, steht in Verbindung mit vielen neuropsychiatrischen Erkrankungen, wie z. B. dem Rett-Syndrom ^{46,48,376}. In Untersuchungen dieser Krankheit wurde lediglich MeCP2 als verantwortlicher Transkriptions-Regulator betrachtet ^{46,48,376,377}, wobei 20 % aller Rett-Syndrom-Fälle durch Mutationen in der MBD-Region von MeCP2 verursacht werden ³⁷⁶. Aus diesem Grund können die in Schenkelberger *et al.* (2017) präsentierten Erkenntnisse der MBD vermittelten Transkriptionsregulation eine neue Betrachtung ermöglichen. Da Punktmutationen in MBD dessen Funktionalität beeinflussen und folglich das Rett-Syndrom auslösen können ³⁷⁶, kann angenommen werden, dass diese Anfälligkeit von MBD gegenüber Punktmutationen aus kooperativen Effekten bei der Bindung von MBD an die DNA resultiert.

4.2 Worst *et al.* (2021)

4.2.1 Hintergrund

In Worst *et al.* (2021) wurde der methylierungsabhängige Regulationsmechanismus der Gene für *Pyelonephritis*-assoziierte Pili (Pap), welche in uropathogenen *E. coli* (UPEC) als Virulenzfaktoren fungieren^{50–53}, näher untersucht. Das zugrunde liegende native DNA-Konstrukt ist der so genannte Pap epigenetische Schalter, welcher aus der *pap*-regulatorischen Region und den *pap*-Genen *papI*, *papB* und *papA* besteht (siehe Abbildung 25)^{40,52,92,378–380}. Da es sich bei *papA*, im Gegensatz zu den Regulatorgenen *papI* und *papB*, um ein Strukturgen handelt, ist dieses nicht direkt in den Regulationsmechanismus involviert^{50,378,379}. Aus diesem Grund wurde *papA* in der vorliegenden Studie nicht betrachtet.

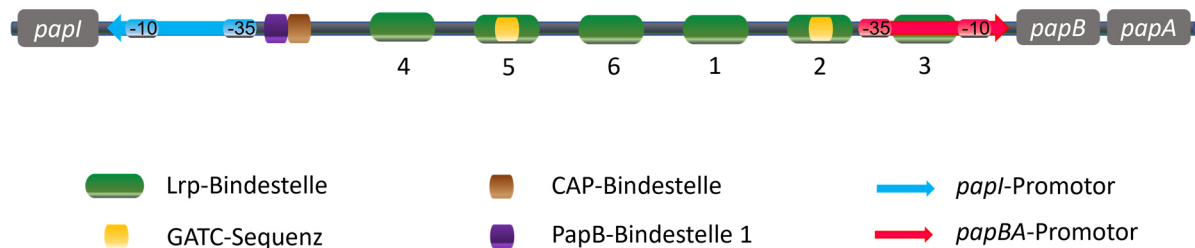


Abbildung 25 Schematische Darstellung des Pap epigenetischen Schalters.

Die *pap*-regulatorische Region beinhaltet 6 Bindestellen (BS) für den Transkriptions-Regulator „Leucine-responsive regulatory protein“ (Lrp), wobei in den Bindestellen 5 und 2 das DNA-Motiv „GATC“ vorhanden ist^{52,380}. Dieses ermöglicht eine durch die DNA-Adenin-Methyltransferase (Dam) vermittelte DNA-Methylierung des Adenins zu N⁶-Methyladenin^{52,380–382}.

Je nachdem, ob die GATC-Sequenz in Lrp-BS 5 (GATC^{dist}) oder in Lrp-BS 2 (GATC^{prox}) methyliert vorliegt, wird von der OFF-Phase (GATC^{dist} methyliert) oder der ON-Phase (GATC^{prox} methyliert) des Pap epigenetischen Schalters gesprochen^{40,53}. Während in der OFF-Phase die Expression der *pap*-Gene reprimiert wird, werden in der ON-Phase die *pap*-Gene exprimiert^{40,53} (siehe Worst *et al.* (2021), Abbildung 1).

Bei diesem Regulationsmechanismus spielt der Transkriptions-Regulator Lrp eine entscheidende Rolle^{40,52,383}. Liegt der Pap epigenetische Schalter in der OFF-Phase vor, so wird die Bindung von Lrp an die Lrp-Bindestellen 4–6 durch die Methylierung von Lrp-BS 5 unterbunden und Lrp bindet kooperativ an die Lrp-Bindestellen 1–3^{40,52,383}. Dadurch wird der *papBA*-Promotor sterisch blockiert und folglich kein PapB produziert, welches wiederum für die *papI*-Expression als Aktivator benötigt wird^{52,379,384,385}. Somit findet auch keine *papI*-Expression statt^{379,384}. Im Gegensatz dazu wird in der ON-Phase des Pap epigenetischen Schalters die Bindung von Lrp an die Bindestellen 1–3 durch die Methylierung von Lrp-

4 Diskussion

BS 2 unterbunden und Lrp bindet kooperativ an die Lrp-Bindestellen 4-6^{40,52,383}. Somit sind der *papBA*- sowie der *papI*-Promotor für die RNA-Polymerase zugänglich und es wird zunächst *papB* exprimiert⁵². Das gebildete PapB kann anschließend durch Bindung an PapB-BS 1 die Expression von *papI* aktivieren^{40,50,379,384}. PapI erhöht wiederum die Bindeaffinität von Lrp gegenüber den Lrp-BS 2 und 5, wobei ein kooperatives Binden an Lrp-BS 2 durch die Methylgruppe verhindert wird^{40,52,383}. Daraus resultiert eine Stabilisierung der ON-Phase⁴⁰. Des Weiteren ist die Bindung des Transkriptions-Regulators „Catabolite activator protein“ (CAP) an die CAP-BS essenziell für die Expression der *pap*-Gene^{384,386}.

Bezüglich der kooperativen Bindung von Lrp an je drei Lrp-Bindestellen wurde vorgeschlagen, dass sich durch einen Zusammenschluss von vier Lrp-Dimeren ringförmige Lrp-Oktamere ausbilden, bei denen drei Dimere mit drei Lrp-Bindestellen wechselwirken können³⁸⁷⁻³⁸⁹. Dies kann die Konformation der DNA beeinflussen, was wiederum essenziell für den methylierungsabhängigen Regulationsmechanismus sein kann (siehe Abbildung 26)^{52,383}.

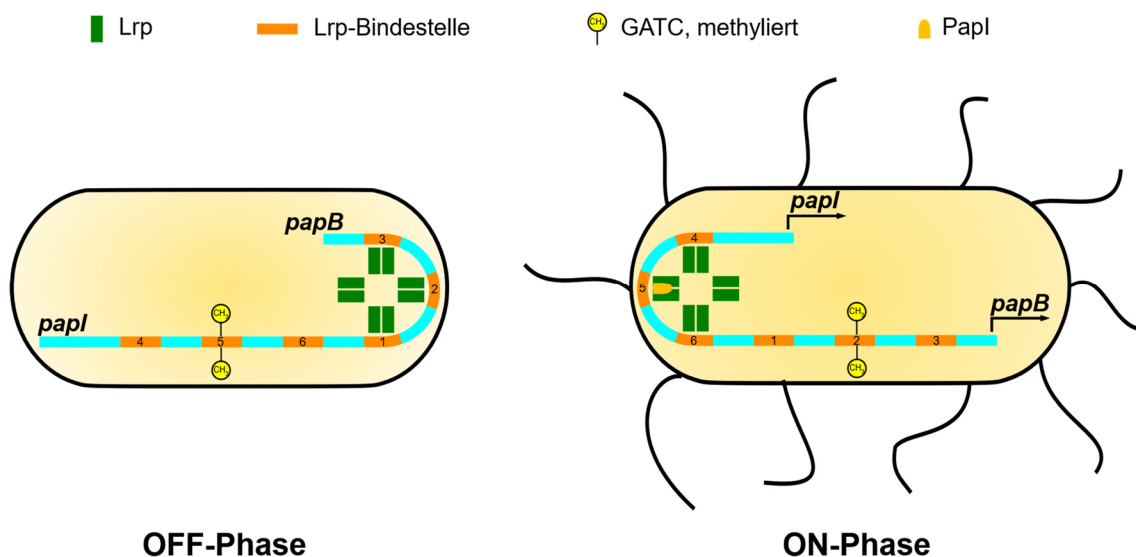


Abbildung 26 Mögliche Lrp vermittelte Konformationsänderungen der *pap*-regulatorischen Region[†]. Vorgeschlagen von Peterson & Reich³⁸³.

Der Pap epigenetische Schalter kann durch einen Phasenvariationsmechanismus zwischen den beiden Phasen „schalten“^{40,50,383,390}. Hierbei müssen für einen Übergang zwei DNA-Replikationszyklen durchlaufen werden, damit sich der Methylierungszustand ändert^{40,50,383,390}.

Für eine bessere Untersuchung des mit dem Phasenvariationsmechanismus des Pap epigenetischen Schalters assoziierten methylierungsabhängigen Genregulationsmechanismus wurde ein modifizierter, synthetischer Pap epigenetischer Schalter designt⁵⁰ (siehe Worst *et al.* (2021), Abbildung 2) und in anschließenden Experimenten charakterisiert.

[†] Die dargestellte Abbildung basiert auf einer Grafik, welche von Emanuel Worst im Rahmen seiner Dissertation am Lehrstuhl für Biologische Experimentalphysik an der Universität des Saarlandes angefertigt wurde.

4.2.2 Ergebnisse

Im synthetischen Pap epigenetischen Schalter wurden die *pap*-Gene durch die codierenden Gene für die Fluoreszenzproteine mCherry (Ersatz für *papI*) und deGFP (Ersatz für *papB* und *papA*) ersetzt. Dadurch kann die Genexpression stromauf- und stromabwärts der *pap*-regulatorischen Region visualisiert werden. In der *pap*-regulatorischen Region (native Sequenz: GenBank: X14471.1, basierend auf ⁵¹; DNA-Sequenzen der 6 Lrp-Bindestellen basierend auf ⁴⁰) wurden sowohl das –35- als auch das –10-Element des *papBA*-Promotors, welche die Lrp-BS 3 umschließen, gegen die entsprechenden Elemente des Or2-Or1-Pr-Promotors (Sequenzen der Elemente basierend auf ^{28,49,50}) ausgetauscht. Des Weiteren wurde der *papI*-Promotor entfernt und stattdessen das –35- und das –10-Element des Or2-Or1-Pr-Promotors so eingefügt, dass diese Elemente die Lrp-BS 4 flankieren. Das –35- und das –10-Element des Or2-Or1-Pr-Promotors wurden verwendet, um eine effektive Transkription im verwendeten auf *E. coli* basierenden, zellfreien Expressionssystem zu erreichen. Zwischen der modifizierten *pap*-regulatorischen Region und den beiden endständigen Genen *mcherry* und *degfp* wurde jeweils die Sequenz der untranslatierten Region UTR1 (Sequenz basierend auf ²⁸) eingefügt.

Im Gegensatz zum nativen Pap epigenetischen Schalter soll beim synthetischen Pap epigenetischen Schalter in beiden Phasen eine Genexpression stattfinden. Durch unterschiedliche Reportergene kann somit das Vorliegen der jeweiligen Phase visualisiert werden. In der OFF-Phase soll *mcherry* exprimiert werden, in der ON-Phase *degfp*. Um die verschiedenen Methylierungszustände der beiden Phasen zu erreichen, wurde entweder GATC^{dist} oder GATC^{prox} zur Sequenz „GCTC“ mutiert, wobei diese Mutation die Bindeaffinität von Lrp gegenüber der Lrp-BS nicht verändert ⁵³. Anschließend wurde eine Methylierung mit Dam durchgeführt. Während die verbleibende GATC-Sequenz dabei methyliert wurde, blieb die GCTC-Sequenz unmethyliert. Zusätzlich wurde ein weiterer Methylierungszustand generiert, bei dem sowohl GATC^{dist} als auch GATC^{prox} methyliert wurde. In den durchgeführten Experimenten wurde der designte synthetische Pap epigenetische Schalter in den Methylierungszuständen (0,1), (1,0) und (1,1) verwendet. Hierbei steht die erste Zahl des 2-Tupels für den Methylierungsstatus (1 = methyliert; 0 = unmethyliert) der Lrp-BS 5 und die zweite Zahl für den Methylierungsstatus der Lrp-BS 2. Die korrekte und effiziente Methylierung der generierten Konstrukte wurde mittels DpnI-Restriktionsverdau (schneidet nur an methylierten GATC-Sequenzen ^{41,391,392}) und anschließender Agarose-Gelelektrophorese nachgewiesen (siehe Worst *et al.* (2021), Abbildung 10 i. V. m. Tabelle 2).

Die generierten Konstrukte wurden in ein auf *E. coli* basierendes, zellfreies Expressionssystem gegeben und die Expression von *mcherry* bzw. *degfp* durch Messung der Fluoreszenz-Intensität der korrespondierenden Fluoreszenzproteine visualisiert. Hierbei konnte die höchste deGFP Fluoreszenz-Intensität im Methylierungszustand (0,1) und die höchste mCherry Fluoreszenz-Intensität im Methylierungszustand (1,0) gemessen werden. Die geringste Fluoreszenz-Intensität beider

4 Diskussion

Fluoreszenzproteine wurde im Methylierungszustand (1,1) gemessen (siehe Worst *et al.* (2021), Abbildung 3).

Da die Transkriptions-Regulatoren Lrp und PapI einen entscheidenden Einfluss auf den Phasenvariationsmechanismus des nativen Pap epigenetischen Schalters haben ^{40,50,52,380,383,393}, sollte gezeigt werden, ob dies auch für den synthetischen Pap epigenetischen Schalter zutrifft. Zu diesem Zweck wurden Genexpressionsreaktionen mit dem synthetischen Pap epigenetischen Schalter in den verschiedenen Methylierungszuständen durchgeführt, bei denen eine Koexpression von *lrp* bzw. *papI* stattfand. Hierbei wurden Plasmide mit den codierenden Genen dem Expressionssystem in verschiedenen Konzentrationen hinzugefügt, um unterschiedliche Mengen des jeweiligen Transkriptions-Regulators zu synthetisieren (²⁸ zeigt die Korrelation zwischen Plasmid-Konzentration und synthetisiertem Protein) (siehe Worst *et al.* (2021), Abbildung 4 und S5).

In den Experimenten konnte gezeigt werden, dass eine Erhöhung der Lrp-Konzentration in allen Methylierungszuständen und unterschiedlichen PapI-Konzentrationen zu einer Verminderung der Genexpression führte. Dennoch war für alle Lrp-Konzentrationen die deGFP Fluoreszenz-Intensität im Methylierungszustand (0,1) und die mCherry Fluoreszenz-Intensität im Methylierungszustand (1,0) am höchsten. Dies zeigte, dass durch die Erhöhung der Lrp-Konzentration die Wahrscheinlichkeit der Bindung von Lrp an alle Lrp-Bindestellen erhöht wurde, was eine gleichzeitige Repression beider Reportergene zur Folge hatte. Auch ohne Koexpression von *papI* konnte dieses Verhalten beobachtet werden. Da für die Präparation des Expressionssystems ein Bakterienstamm verwendet wurde, der kein PapI synthetisiert, ist im Expressionssystem generell kein PapI vorhanden (das Genom von *E. coli* Stamm BL21 enthält kein Gen für PapI; Sequenz: GenBank: CP010816.1 ³⁹⁴). Erst durch die rekombinante Expression von *papI* wird PapI synthetisiert. Somit konnte Lrp in unseren Experimenten auch in Abwesenheit von PapI eine methylierungsabhängige Genregulation bewirken, während für die methylierungsabhängige Genregulation des nativen Pap epigenetischen Schalters beide Transkriptions-Regulatoren benötigt wurden ^{40,52,383}.

Eine Erhöhung der PapI-Konzentration konnte nur im Methylierungszustand (0,1) ohne Koexpression von *lrp* die deGFP Fluoreszenz-Intensität erhöhen, während die mCherry Fluoreszenz-Intensität verringert wurde. Da PapI für die Interaktion mit der *pap*-regulatorischen Region Lrp benötigt ³⁹⁵, zeigte dies, dass im verwendeten Expressionssystem geringe Mengen von Lrp aufgrund dessen Präparation vorhanden sein mussten, was auch bereits in der Literatur bestätigt wurde (siehe Anhang von ¹⁸³). Die erhaltenen Messergebnisse können dadurch erklärt werden, dass die Erhöhung der PapI-Konzentration zu einer erhöhten Bindeaffinität von Lrp gegenüber der Lrp-BS 5 führte, da Vergleichbares bereits für das native System gezeigt wurde ^{40,52,53,383}. Hierbei gilt anzumerken, dass auch im Methylierungszustand (1,1) ohne Koexpression von *lrp*, die mCherry Fluoreszenz-Intensität durch Erhöhung der PapI-Konzentration verringert wurde, wobei dieses Verhalten weniger ausgeprägt war als im Methylierungszustand (0,1).

In Analogie zum nativen Pap epigenetischen Schalter ^{40,52,383,396} wurde angenommen, dass die Methylierung von Lrp-BS 5 oder Lrp-BS 2 des synthetischen Pap epigenetischen Schalters eine kooperative Bindung von Lrp an diese Lrp-BS verhindert. Vielmehr kann davon ausgegangen werden, dass Lrp kooperativ an die verbleibende, unmethylierte Lrp-BS bindet, dadurch eine Bindung der RNA-Polymerase verhindert und somit die Expression des benachbarten Gens reprimiert. Daraus resultiert lediglich die Expression des Gens, welches sich in der Nähe der methylierten Lrp-BS befindet. Wird Lrp-BS 5 methyliert (1,0), findet die Expression von *mcherry* statt. Wird Lrp-BS 2 methyliert (0,1), findet die Expression von *degfp* statt. Somit sollte im Methylierungszustand (1,1) die Expression beider Reportergene stattfinden, da kein Lrp kooperativ binden kann (siehe Worst *et al.* (2021), Abbildung 5). In Analogie zum nativen System ^{52,383} wurde auch für den synthetischen Pap epigenetischen Schalter angenommen, dass ringförmige Lrp-Oktamere die Konformation der DNA verändern können (vergleiche Abbildung 26 mit Worst *et al.* (2021), Abbildung 6).

Während die Methylierungszustände (1,0) und (0,1) in den durchgeführten Experimenten die erwarteten Ergebnisse lieferten, konnte festgestellt werden, dass im Methylierungszustand (1,1) eine Repression beider Reportergene stattfand (siehe Worst *et al.* (2021), Abbildung 3, 4 und S5). Somit konnten die erhaltenen Ergebnisse nicht durch den von Peterson & Reich vorgeschlagenen Regulationsmechanismus ³⁸³ erklärt werden.

Im Rahmen des Designs des synthetischen Pap epigenetischen Schalters konnte bereits festgestellt werden, dass sich dessen Synthese als schwierig gestaltete ⁵⁰. Diesbezüglich wurde vermutet, dass selbstkomplementäre Bereiche innerhalb der DNA-Sequenz der *pap*-regulatorischen Region Sekundärstrukturen zur Folge haben, welche die Synthese erschweren. Da Sekundärstrukturen einen Einfluss auf die Genexpression haben können ⁹⁵, wurde folglich angenommen, dass diese Sekundärstrukturen in den methylierungsabhängigen Genregulationsmechanismus des synthetischen Pap epigenetischen Schalters involviert sind.

In diesem Zusammenhang wurden die DNA-Einzelstränge der synthetischen *pap*-regulatorischen Region mit dem „mfold web server“ ³⁹⁷ auf mögliche Sekundärstrukturen untersucht. Hierbei konnte festgestellt werden, dass diese durch selbstkomplementäre Sequenzen Sekundärstrukturen ausbilden können, bei denen Lrp-BS 5 und 2 fast vollständig und Lrp-BS 1 und 6 sowie Lrp-BS 3 und 4 zumindest teilweise gepaart vorliegen. Somit kann die doppelsträngige synthetische *pap*-regulatorische Region neben der linearen DNA-Konformation auch kreuzförmige DNA-Konformationen ausbilden (siehe Worst *et al.* (2021), Abbildung 7).

Um die Ausbildung solcher kreuzförmiger DNA-Konformationen nachzuweisen, wurden für die drei Methylierungszustände je zwei unterschiedliche, doppelt-gelabelte lineare DNA-Konstrukte mit der DNA-Sequenz der synthetischen *pap*-regulatorischen Region generiert. Als Labels wurden der Fluoreszenzfarbstoff 6-Fam sowie dessen Quencher BHQ1 ²⁹³ verwendet. Da eine direkte Synthese der DNA-Konstrukte nicht möglich war, wurden zunächst einzelsträngige DNA-Oligos synthetisiert

(Metabion international AG, Planegg/Steinkirchen, Deutschland; DNA-Sequenzen, siehe Worst *et al.* (2021), Abbildung S7) und anschließend hybridisiert und ligiert, um die beiden Konstrukte in den drei möglichen Methylierungszuständen zu erhalten. Bei einem Konstrukt („LinearQuenchDNA“) wurden die beiden Labels so angeordnet, dass ein Quenching der Fluoreszenz-Intensität von 6-Fam bei Vorliegen einer linearen DNA-Konformation stattfindet. Bei dem anderen Konstrukt („CruciformQuenchDNA“) erfolgte die Anordnung so, dass bei Vorliegen einer vermuteten kreuzförmigen DNA-Konformation Quenching auftritt (siehe Worst *et al.* (2021), Abbildung 11).

Unter Verwendung der generierten Konstrukte wurde kontinuierlich die 6-Fam Fluoreszenz-Intensität in einer gepufferten sowie einer bakteriellen („all *E. coli* cell-free expression system“^{28,29}; Lrp vorhanden, siehe Anhang von¹⁸³) Umgebung gemessen (siehe Worst *et al.* (2021), Abbildung 8 und S8). Hierbei konnte festgestellt werden, dass in der gepufferten Umgebung das Konstrukt „LinearQuenchDNA“ eine konstante, stark gequenchte 6-Fam Fluoreszenz-Intensität aufwies, während für das Konstrukt „CruciformQuenchDNA“ eine konstante, ungequenchte 6-Fam Fluoreszenz-Intensität erhalten wurde. Folglich wurde die lineare DNA-Konformation der synthetischen *pap*-regulatorischen Region in der gepufferten Umgebung beibehalten.

In der bakteriellen Umgebung, in der auch Lrp vorhanden war, wurde hingegen ein anderes Verhalten beobachtet. Wurde das Konstrukt „LinearQuenchDNA“ in die bakterielle Umgebung eingeführt, konnte festgestellt werden, dass die anfänglich gequenchte 6-Fam Fluoreszenz-Intensität innerhalb kürzester Zeit stark anstieg und anschließend einen konstanten Wert erreichte. Dies kann mit der „Öffnung“ der linearen DNA-Konformation erklärt werden. Bei Einführung des Konstruktes „CruciformQuenchDNA“ erfolgte über die gesamte Messdauer eine Abnahme der anfänglich ungequenchten 6-Fam Fluoreszenz-Intensität. Dies zeigt die Ausbildung der vermuteten kreuzförmigen DNA-Konformation. Die Ergebnisse beider Konstrukte lieferten den Nachweis, dass die anfänglich lineare DNA-Konformation der synthetischen *pap*-regulatorischen Region in der bakteriellen Umgebung aufgelöst und tatsächlich die vermutete kreuzförmige DNA-Konformation ausgebildet wurde.

Während die „Öffnung“ der DNA schnell über ein kurzes Zeitintervall von mehreren Minuten erfolgte, geschah die Ausbildung der vermuteten kreuzförmigen DNA-Konformation verhältnismäßig langsam über ein langes Zeitintervall von mehreren Stunden. Des Weiteren war die finale gemessene 6-Fam Fluoreszenz-Intensität des Konstruktes „LinearQuenchDNA“ geringer als die maximal mögliche, welche in Abwesenheit des Quenchers bestimmt wurde. Zusätzlich war die finale gemessene 6-Fam Fluoreszenz-Intensität des Konstruktes „CruciformQuenchDNA“ zu hoch, um von einem kompletten Quenching auszugehen. Zusammenfassend lassen sich diese Beobachtungen dadurch erklären, dass die „Öffnung“ der DNA den Übergang von einer definierten Konformation in eine Vielzahl anderer möglicher Konformationen darstellt und die vermutete kreuzförmige DNA-Konformation nur eine von diesen ist. Da auch andere Konformationen ausgebildet werden können, geschieht die „Öffnung“ schneller als die Ausbildung einer neuen definierten Konformation. Des Weiteren erfolgte keine

komplette „Öffnung“ der linearen DNA-Konformation, was die Vermutung nahelegt, dass ein geringer Anteil des Konstruktes in der linearen oder einer vergleichbaren Konformation verweilt. Somit kann die synthetische *pap*-regulatorische Region gleichzeitig in einer Vielzahl von verschiedenen Konformationen vorliegen, wobei der Anteil der nicht linearen Konformationen stark überwiegt. Diese Annahme wird dadurch verstärkt, dass die simultane Präsenz verschiedener DNA-Konformationen für gleiche DNA-Konstrukte bereits in der Literatur beschrieben wurde^{95,96,98}.

Mit dem Nachweis der vermuteten kreuzförmigen DNA-Konformation der synthetischen *pap*-regulatorischen Region konnte ein neuer Ansatz bezüglich des methylierungsabhängigen Genregulationsmechanismus des synthetischen Pap epigenetischen Schalters gemacht werden (siehe Worst *et al.* (2021), Abbildung 9). Dieser besagt, dass in den drei möglichen Methylierungszuständen eine kreuzförmige DNA-Konformation ausgebildet werden kann. Hierbei stellt die vorgeschlagene kreuzförmige DNA-Konformation nur eine unter vielen möglichen dar, welche sich u. a. durch die Länge der Haarnadelstrukturen unterscheiden. Je nach Länge der Haarnadelstrukturen kann eine Genexpression der Reportergene erfolgen oder nicht. Vergleichbare Beispiele aus der Literatur lassen eine solche Vermutung zu^{95,98}.

Somit kann für den Methylierungszustand (1,1) angenommen werden, dass gleichzeitig kreuzförmige DNA-Konformationen vorliegen, in denen eine Expression von *mcherry* und *degfp* stattfindet, sowie kreuzförmige DNA-Konformationen, in denen eine Genexpression beider Reportergene reprimiert wird.

Für die Methylierungszustände (1,0) und (0,1) gestaltet sich eine andere Situation. Dadurch, dass Lrp kooperativ an die unmethylierte Lrp-BS 2 bzw. 5 bindet, wird die Expression von *degfp* bzw. *mcherry* unabhängig von der DNA-Konformation reprimiert, während *mcherry* bzw. *degfp* exprimiert wird. Prinzipiell kann aber auch für diese Methylierungszustände davon ausgegangen werden, dass verschiedene kreuzförmige DNA-Konformationen möglich sind, die die Expression des zu exprimierenden Gens fördern oder unterbinden. Hierbei gilt anzumerken, dass die Methylierungszustände (1,0) und (0,1) die Besonderheit aufweisen, dass bei der Ausbildung kreuzförmiger DNA-Konformationen hemimethylierte GATC-Sequenzen entstehen. In der Literatur wurde bereits beschrieben, dass solche Hemimethylierungen eine veränderte DNA-Struktur mit einer verengten großen Furche bewirken, was das Bindungsverhalten diverser Proteine an diese Stellen beeinflussen kann^{398,399}. Diesbezüglich konnte für die Transkriptions-Regulatoren Lrp und PapI nachgewiesen werden, dass deren Bindungsverhalten gegenüber dem Pap epigenetischen Schalter durch das Vorliegen von hemimethylierten GATC-Sequenzen innerhalb der *pap*-regulatorischen Region und zusätzlich durch die Art der Hemimethylierung beeinflusst wird⁴⁰. In Analogie zur Literatur (Betrachtung von vollständig methylierten GATC-Sequenzen)³⁸³ kann auch für hemimethylierte GATC-Sequenzen angenommen werden, dass die kooperative Bindung von Lrp spezifische DNA-Konformationen stabilisieren kann. Aufgrund der in Worst *et al.* (2021) präsentierten experimentellen Ergebnisse für die Methylierungszustände (1,0) und (0,1) wird vermutet, dass durch die kooperative

4 Diskussion

Bindung von Lrp die Expressions-fördernden kreuzförmigen DNA-Konformationen in diesen Methylierungszuständen stabilisiert werden.

Abbildung 27 zeigt, wie die aus den Fluoreszenz-Quenching-Experimenten gewonnenen Ergebnisse mit dem neuen Ansatz bezüglich des methylierungsabhängigen Genregulationsmechanismus des synthetischen Pap epigenetischen Schalters in Einklang gebracht werden können.

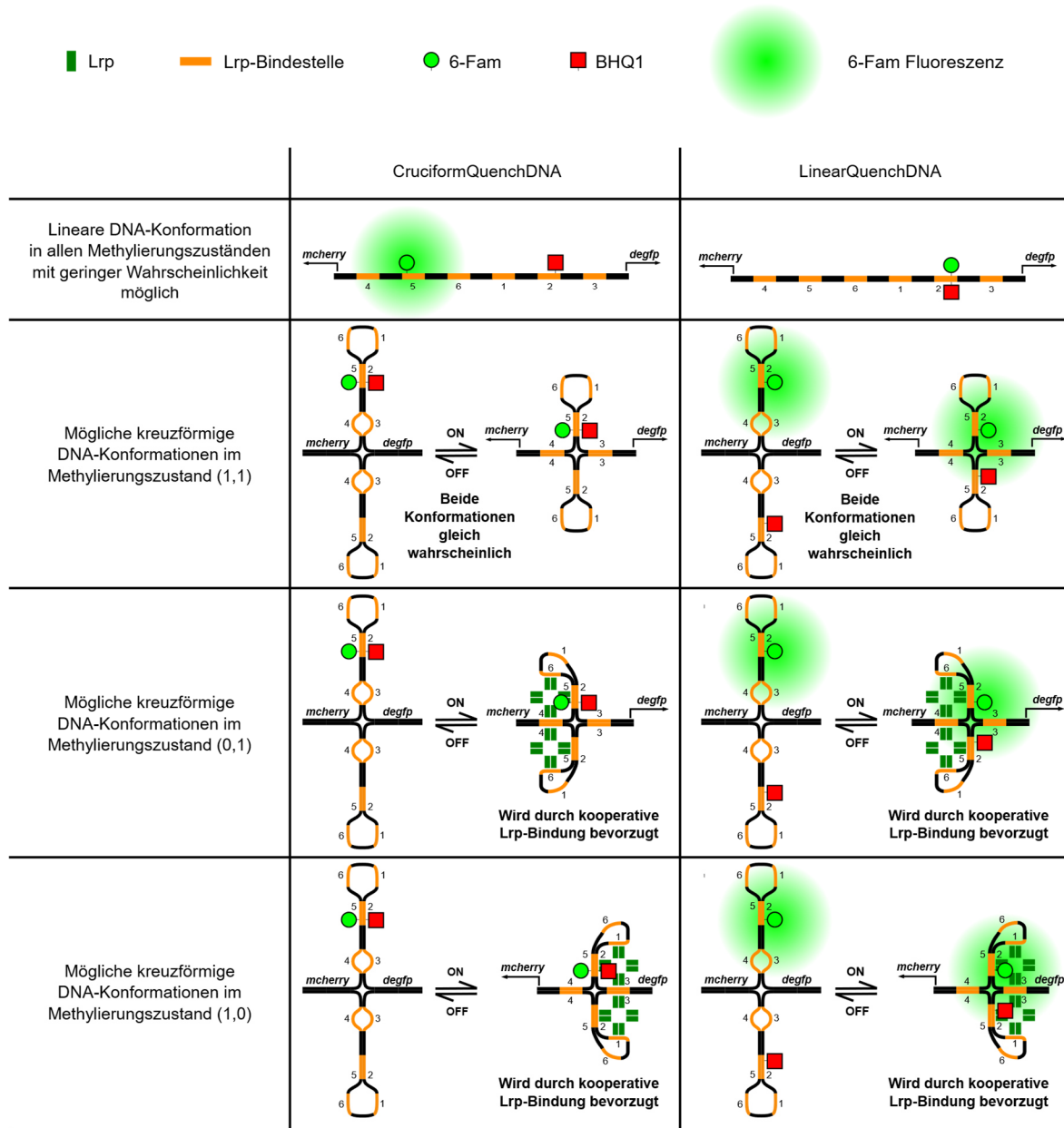


Abbildung 27 Mögliche DNA-Konformationen von zwei unterschiedlich gelabelten Versionen des synthetischen Pap epigenetischen Schalters in Abhängigkeit des Methylierungszustandes[‡]. Diese können eine gequenchete (6-Fam und BHQ1 in unmittelbarer Nähe) oder eine ungequenchete (6-Fam und BHQ1 nicht in unmittelbarer Nähe) 6-Fam Fluoreszenz aufweisen, was einen indirekten Nachweis der vorliegenden DNA-Konformation ermöglicht. Zur besseren Überschaubarkeit sind die Methylgruppen in der Abbildung nicht abgebildet.

[‡] Die dargestellte Abbildung beinhaltet Grafiken von DNA-Konformationen, welche von Emanuel Worst im Rahmen seiner Dissertation am Lehrstuhl für Biologische Experimentalphysik an der Universität des Saarlandes angefertigt wurden.

4.2.3 Zusammenfassung und Ausblick

Worst *et al.* (2021) beschreibt das Design eines synthetischen Pap epigenetischen Schalters, bei dem in den beiden möglichen Phasen unterschiedliche Fluoreszenzprotein-Gene exprimiert werden. Mit diesem Konstrukt konnte in einem auf *E. coli* basierenden, zellfreien Expressionssystem eine genauere Analyse des zugrundeliegenden Regulationsmechanismus erfolgen. Hierbei konnte gezeigt werden, dass neben Lrp auch die Sekundärstruktur der *pap*-regulatorischen Region einen entscheidenden Einfluss auf die betrachtete methylierungsabhängige Genregulation hat.

Obwohl die Ausbildung von kreuzförmigen DNA-Konformationen lediglich für den synthetischen Pap epigenetischen Schalter experimentell nachgewiesen werden konnte, ist dies auch für den nativen Pap epigenetischen Schalter wahrscheinlich (siehe Abbildung 28).

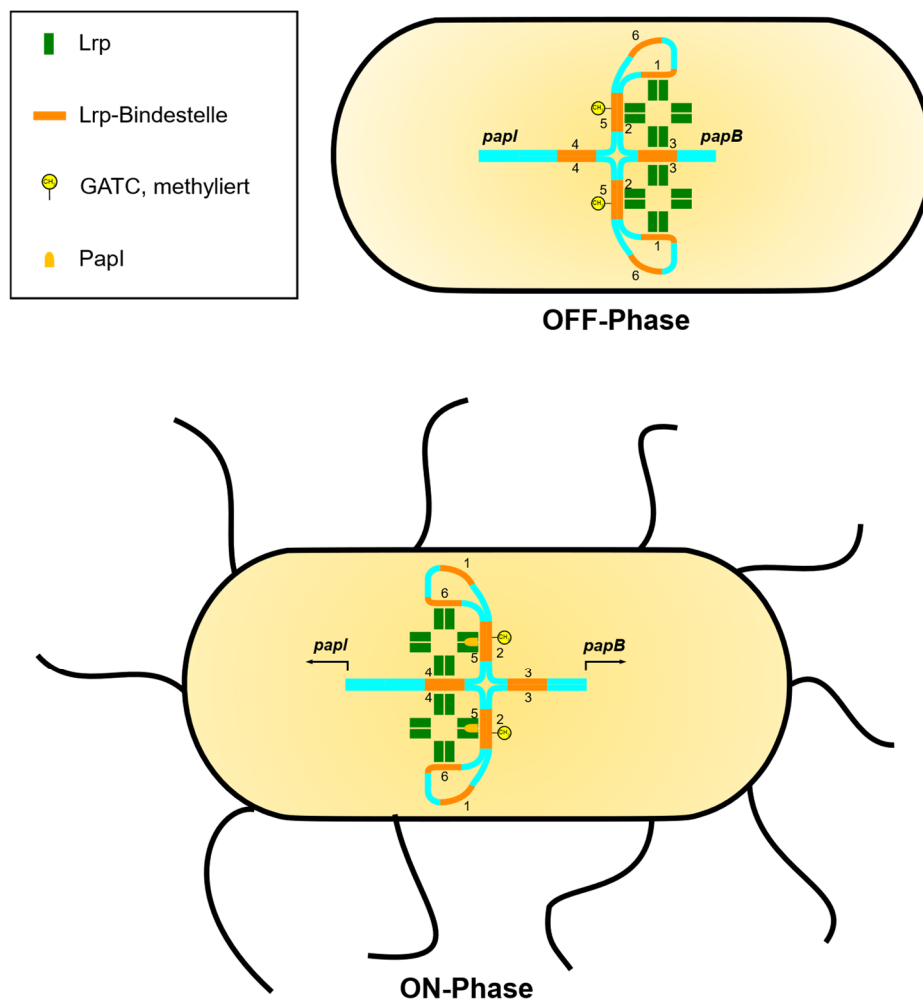


Abbildung 28 Aus Worst *et al.* (2021) abgeleitete mögliche Lrp vermittelte Konformationsänderungen der *pap*-regulatorischen Region⁵. Die Abbildung zeigt eine neue Betrachtung der in Abbildung 26 dargestellten, von Peterson & Reich vorgeschlagenen³⁸³, Lrp vermittelten Konformationsänderungen.

⁵ Die dargestellte Abbildung basiert auf einer Grafik, welche von Emanuel Worst im Rahmen seiner Dissertation am Lehrstuhl für Biologische Experimentalphysik an der Universität des Saarlandes angefertigt wurde.

4 Diskussion

Somit kann angenommen werden, dass die Sekundärstruktur der *pap*-regulatorischen Region auch einen entscheidenden Einfluss auf den Phasenvariationsmechanismus des nativen *Pap* epigenetischen Schalters hat. Mit dieser neuen Erkenntnis konnte ein besseres Verständnis des molekularen Mechanismus der methylierungsabhängigen Genregulation des *Pap* epigenetischen Schalters gewonnen werden.

Der neu designte synthetische *Pap* epigenetische Schalter eignet sich zur genaueren Untersuchung des Phasenvariationsmechanismus des *Pap* epigenetischen Schalters. Des Weiteren stellt er ein DNA-Konstrukt dar, welches für verschiedene Anwendungen verwendet werden könnte. Neben der möglichen Verwendung als Sensor für Lrp-Präsenz oder Methylase-Aktivität kann der synthetische *Pap* epigenetische Schalter als mögliches Grundkonstrukt für die Generierung von genetischen Schaltern fungieren, welche durch Kraft-induzierte Konformationsänderungen die Genexpression regulieren können^{95,400,401} (siehe Abbildung 29).

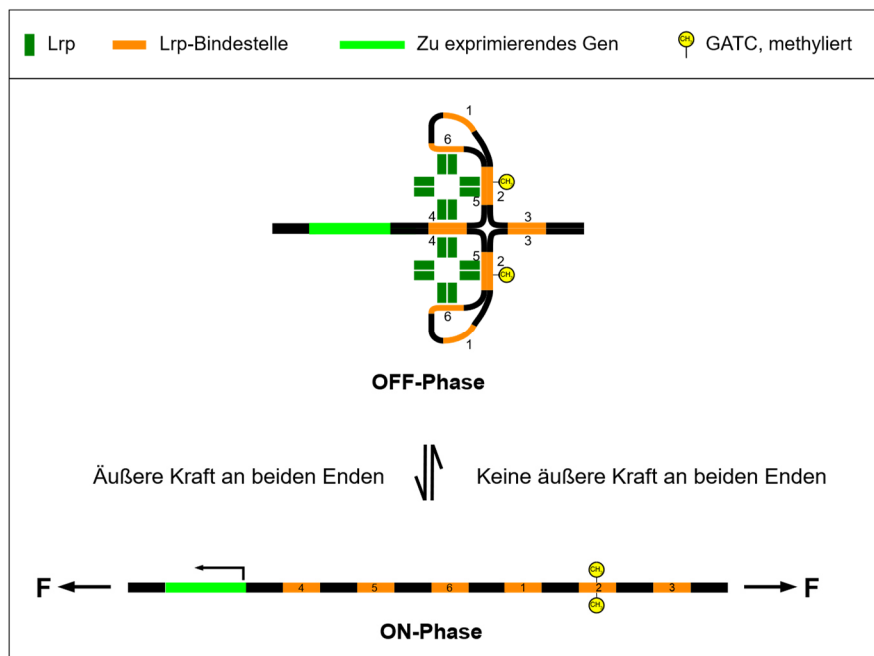


Abbildung 29 Kraft-induzierte Genexpression unter Verwendung der synthetischen *pap*-regulatorischen Region**. Das dargestellte Konstrukt (*pap*-regulatorische Region im Methylierungszustand (0,1); nur 1 Gen, stromaufwärts der *pap*-regulatorischen Region) bildet in Anwesenheit von Lrp eine kreuzförmige DNA-Konformation aus, welche die Genexpression unterbindet. Durch Anlegen einer äußeren Kraft F an beide Enden des Konstrukts kann eine lineare DNA-Konformation ausgebildet werden, die eine Genexpression ermöglicht. Wird die äußere Kraft entfernt, geht das Konstrukt wieder in die kreuzförmige DNA-Konformation über und die Genexpression wird gestoppt. Somit kann das Konstrukt als genetischer Schalter betrachtet werden.

Worst *et al.* (2021) zeigt eine mögliche Versuchsanordnung unter Verwendung einer magnetischen Pinzette, die ein solches Experiment realisieren kann (siehe Worst *et al.* (2021), Abbildung S9).

** Die dargestellte Abbildung beinhaltet Grafiken von DNA-Konformationen, welche von Emanuel Worst im Rahmen seiner Dissertation am Lehrstuhl für Biologische Experimentalphysik an der Universität des Saarlandes angefertigt wurden.

4.3 Finkler & Ott (2019)

4.3.1 Hintergrund

Für zahlreiche *in vitro* Expressionsexperimente wird eine Regulation der Genexpression benötigt. Eine solche erfolgt prinzipiell durch klassische biologische Methoden, wie z. B. durch die Verwendung von induzierbaren Promotoren oder von Aktivatoren bzw. Repressoren (siehe Abschnitt 2.5)^{18,24,29,113,130,206,226}. Entsprechende Regulationsmechanismen benötigen DNA-Modifikationen in Form von Mutationen, Klonierungen oder Methylierungen^{160,206,209,228}. Des Weiteren ist es mit den aktuellen Methoden schwierig, die Genexpression vollständig in einem kurzen Zeitraum an- oder abzuschalten^{30,160,206,209,228}. Da dies jedoch in diversen Experimenten von Nutzen sein kann, sollte in Finkler & Ott (2019) eine neue Methode entwickelt werden, welche eine räumlich und zeitlich begrenzte Genexpression ohne größere DNA-Modifikationen ermöglicht.

4.3.2 Ergebnisse

Zunächst erforderte die neue Methode ein geeignetes DNA-Template. Zu diesem Zweck wurde ausgehend von dem in Abschnitt 2.4 beschriebenen Plasmid pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-Tag-T500 (siehe Abbildung 10) das linearisierte, am 5'-Ende biotinylierte DNA-Konstrukt „Biotin-Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-Tag-T500-Seq4“ (DNA_{deGFP}) mittels einer Polymerase-Kettenreaktion (PCR) hergestellt. Hierbei bezeichnet Biotin-Seq1 den für die PCR verwendeten, am 5'-Ende biotinylierten, vorwärts gerichteten Primer und Seq4 den für die PCR verwendeten rückwärts gerichteten Primer. Essenziell für die präsentierte neue Methode zur räumlich und zeitlich begrenzten Genexpression war die Biotinylierung der DNA, da hierdurch eine Bindung der DNA an Streptavidin-beschichtete Oberflächen ermöglicht wurde^{116,118,119}.

In diesem Zusammenhang wurde DNA_{deGFP} an Streptavidin-beschichtete magnetische Beads (T1) (Dynabeads™ MyOne™ Streptavidin T1, Invitrogen, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) gebunden (Bindung gemäß Hersteller-Protokoll⁴⁰²). Die so erhaltenen DNA-beschichteten Beads („Bead-gebundene DNA_{deGFP}“) wurden dem Expressionssystem als DNA-Template für die Expression von *degfp* hinzugegeben. Des Weiteren musste dem System gleichzeitig der Desoxyribonuklease-Inhibitor GamS hinzugegeben werden, da dieser die Degradation linearer DNA im Expressionssystem durch die darin enthaltene Desoxyribonuklease RecBCD verhindert^{177,178,183,185}. Experimentell konnte die Funktionalität von Bead-gebundener DNA_{deGFP} als DNA-Template für eine Genexpressionsreaktion im verwendeten Expressionssystem bestätigt werden, auch wenn die 5-fache DNA-Konzentration von Bead-gebundener DNA_{deGFP} gegenüber ungebundener DNA_{deGFP} für eine vergleichbare Genexpressionsrate benötigt wurde (siehe Abbildung 30). Die Verwendung von Bead-

4 Diskussion

gebundener DNA als DNA-Template wurde bereits in der Literatur für ein anderes Expressionssystem beschrieben ⁴⁰³.

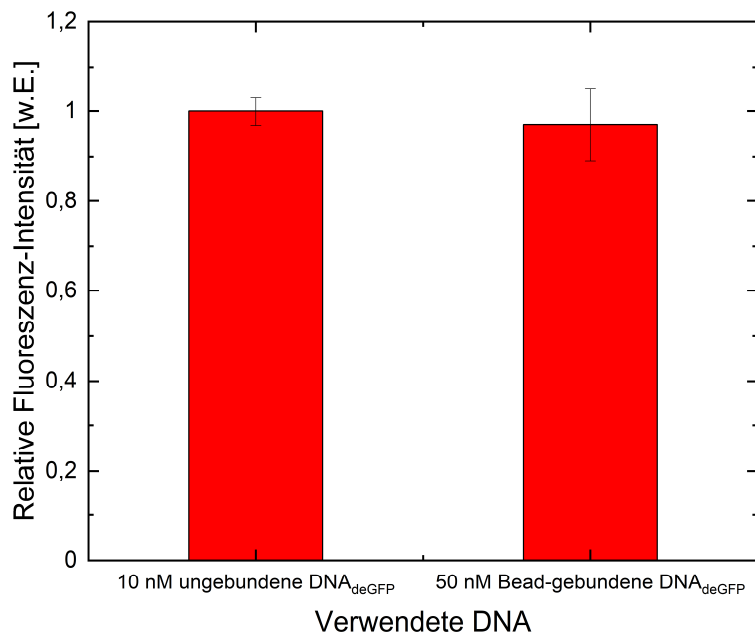


Abbildung 30 Vergleich der gemessenen Fluoreszenz-Intensitäten bei Verwendung von 10 nM ungebundener DNA_{deGFP} oder von 50 nM Bead-gebundener DNA_{deGFP} als DNA-Template in einer zellfreien Genexpressionsreaktion.

Eine Erklärung für die höhere Konzentration an benötigter DNA kann darin liegen, dass durch die Bindung der DNA an die Beads deren Zugänglichkeit für die Polymerase beeinträchtigt wird und sich dies nachteilig auf die Genexpressionsrate auswirkt.

Des Weiteren konnte festgestellt werden, dass die Art der Beads, an die DNA_{deGFP} gebunden wird, einen entscheidenden Einfluss auf die Genexpressionsrate hat. Während analoge Experimente mit vier verschiedenen Arten von Streptavidin-beschichteten Beads (M-270, M-280, C1 und T1) (Dynabeads™ Streptavidin Trial Kit, Invitrogen, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) (Charakteristika in ⁴⁰⁴) durchgeführt wurden, so lieferten nur die Beads T1 eine optimale Genexpressionsrate (siehe Abbildung 31).

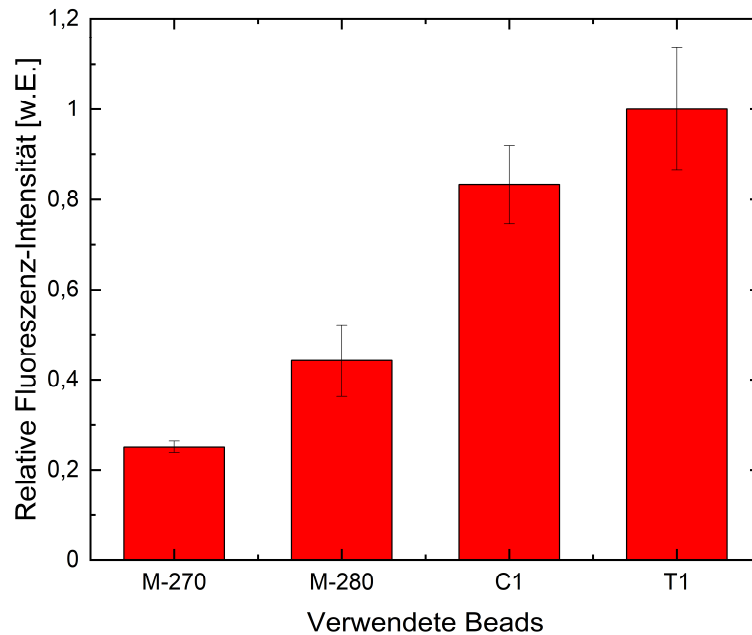


Abbildung 31 Vergleich der gemessenen Fluoreszenz-Intensitäten bei Verwendung verschiedener Arten von Streptavidin-beschichteten Beads (M-270, M-280, C1 und T1), an die DNA_{deGFP} gebunden wurde, als DNA-Template ($c = 50 \text{ nM}$) in einer zellfreien Genexpressionsreaktion.

Diese Beobachtung kann durch unterschiedliche Beschaffenheiten der verschiedenen Bead-Arten erklärt werden. So weisen diese unterschiedliche Bindungskapazitäten, Größen, Sedimentationsraten etc. auf

404

Da im ersten Schritt der Publikation die Funktionalität von Bead-gebundener DNA als DNA-Template für eine Genexpressionsreaktion bestätigt wurde, ermöglichte dies eine räumlich und zeitlich kontrollierbare DNA-Exposition im Expressionssystem. Für die DNA-Entfernung musste lediglich ein Magnet verwendet werden. Hierbei konnte festgestellt werden, dass in einer Genexpressionsreaktion mit einer Dauer von 16 h eine anfängliche DNA-Exposition im Expressionssystem über 30 Minuten die gleiche deGFP-Produktionsrate mit sich führte wie eine DNA-Exposition über die gesamte Reaktionsdauer von 16 h (siehe Finkler & Ott (2019), Abbildung 1).

Durch kontinuierliche Messung der Fluoreszenz-Intensität, welche mit der Menge des im Expressionssystem synthetisiertem deGFP korreliert, konnte gezeigt werden, dass selbst nach Entfernung der deGFP-codierenden DNA weiteres deGFP gebildet wird. Die Freisetzung von Bead-gebundener DNA ist durch die starke Interaktion zwischen Biotin und Streptavidin unwahrscheinlich^{118,119} und konnte experimentell widerlegt werden (siehe Finkler & Ott (2019), Abbildung S 1). Dies zeigte die Präsenz einer im Expressionssystem stabilen, von *degfp* transkribierten mRNA (Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-Tag-T500-Seq4-mRNA), welche eine kontinuierliche Translation ermöglichte. Während eine Entfernung von Bead-gebundenem *degfp* nur die Transkription beenden konnte, benötigte ein vollständiger Genexpressionsstopp die zusätzliche Entfernung der gebildeten mRNA und die daraus resultierende Beendigung der Translation.

Eine mRNA-Entfernung sollte über deren Hybridisierung mit einzelsträngiger Bead-gebundener DNA^{21,239,405} erfolgen. Zu diesem Zweck wurde einzelsträngige DNA (ssDNA) verwendet, welche eine komplementäre Sequenz zu Bereichen der Seq1-Or2-Or1-Pr-UTR1-deGFP-6xHis-Tag-T500-Seq4-mRNA aufwies. Neben einzelsträngiger DNA komplementär zum 3'-Ende der transkribierten mRNA (3'-ssDNA) wurde auch einzelsträngige DNA komplementär zum darauffolgenden poly-A-Schwanz²³⁹ (poly-T-ssDNA) verwendet. Erstere wurde als biotinyliertes DNA-Oligo erworben (Metabion international AG, Planegg/Steinkirchen, Deutschland) und an magnetische Beads mit Streptavidin-Beschichtung (T1) (Dynabeads™ MyOne™ Streptavidin T1, Invitrogen, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) gebunden (Bindung gemäß Hersteller-Protokoll⁴⁰²). Letztere konnte bereits in Bead-gebundener Form kommerziell erworben werden (Dynabeads™ Oligo (dT)₂₅, Invitrogen, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA).

Experimentell konnte gezeigt werden, dass eine Mischung aus 250 µg Beads mit 3'-ssDNA-Oberfläche und 250 µg Beads mit poly-T-ssDNA-Oberfläche in einem Reaktionsvolumen von 12 µl einer zellfreien Genexpressionsreaktion zuvor transkribierte mRNA optimal binden konnte (Bindung gemäß Dynabeads® Oligo (dT)₂₅-Hersteller-Protokoll⁴⁰⁶). Nach magnetischer Separation der einzelsträngigen Bead-gebundenen DNA und somit auch der hybridisierten mRNA konnte die Translation der mRNA und folglich auch die deGFP-Produktion beendet werden. Die Hinzugabe neuer codierender DNA zu einem so behandeltem Expressionssystem führte zu einer neuen Genexpression, womit gezeigt werden konnte, dass bei Verwendung der präsentierten Methode zur mRNA-Entfernung die Funktionalität des Expressionssystems erhalten bleibt (siehe Finkler & Ott (2019), Abbildung 2). Dennoch war die Expressionsrate der zweiten Genexpression verringert. Dies kann daran liegen, dass die vorangegangene *degfp*-Expression bereits Ressourcen verbrauchte. Zusätzlich besteht prinzipiell die Möglichkeit, dass durch die Entfernung der mRNA auch daran gebundene Ribosomen entfernt werden und daraus eine verringerte Expressionsrate resultiert.

Insgesamt kann die präsentierte Methode zur Genregulation somit in zwei Einzelschritte untergliedert werden kann (siehe Finkler & Ott (2019), Abbildung 3). Hierbei reguliert der erste Schritt die Transkription und der zweite Schritt die darauffolgende Translation.

4.3.3 Zusammenfassung und Ausblick

In Finkler & Ott (2019) wurde eine Methode erarbeitet, die eine räumlich und zeitlich kontrollierbare Genexpression in einem auf *E. coli* basierenden, zellfreien Expressionssystem ermöglicht. Zunächst wurde codierende DNA an magnetische Beads gebunden und dieses Konstrukt als DNA-Template für die zellfreie Genexpression verwendet. Hierbei wurde festgestellt, dass die Genexpression durch magnetische Separation der Bead-gebundenen DNA nicht beendet werden konnte, da die transkribierte mRNA im Expressionssystem stabil war und somit eine kontinuierliche Translation veranlasste. Die

Beendigung der Genexpression erforderte somit die zusätzliche Entfernung der mRNA. In diesem Zusammenhang wurden magnetische Beads mit gebundener einzelsträngiger DNA verwendet, welche unter der Ausbildung von DNA/mRNA-Hybriden die transkribierte mRNA zunächst binden konnten. Durch eine anschließende magnetische Separation der Beads konnte die mRNA aus dem Expressionssystem entfernt und folglich die Genexpression beendet werden. Somit konnte die Genexpression durch magnetische Separation der DNA und mRNA räumlich und zeitlich reguliert werden.

Die neu entwickelte Methode könnte in zukünftigen Experimenten verwendet werden, in denen mehrere Gene nacheinander im selben Expressionssystem exprimiert werden sollen. Dadurch wird eine Koexpression der verschiedenen Gene vermieden und Ressourcen werden lediglich für die erwünschte Genexpression verbraucht. Dies könnte z. B. bei der Realisierung von Reaktionskaskaden in zellfreien Expressionssystemen von Nutzen sein.

4.4 Finkler *et al.* (2020)

4.4.1 Hintergrund

Lysin ist eine Aminosäure, die in zellfreien Expressionssystemen aufgrund deren Präparation in geringen Konzentrationen vorhanden sein und somit die Genexpression beeinflussen kann ⁵⁷. In vorangegangenen Experimenten konnte bereits gezeigt werden, dass diese geringen Konzentrationen z. B. dem seitenkettenspezifischen Einbau von nicht-kanonischen Lysin-Analoga in Proteine entgegenwirken können ⁵⁷. Folglich erfordern diverse Experimente die Entfernung des „restlichen“ Lysins vor Durchführung einer Genexpression.

Selbst die Verwendung von Lysin-auxotrophen Bakterienstämmen für die Präparation von Expressionssystemen gewährt keine komplette Abwesenheit von Lysin, da dieses für das vorangegangene Wachstum der Bakterien hinzugegeben werden muss ^{22,109,189,250}. Eine verlängerte Dialyse bei der Präparation des Expressionssystems zur Verringerung der Konzentration des „restlichen“ Lysins führte zu Expressionssystemen mit verminderten Genexpressionsraten. Da klassische Trennverfahren keine Separation einzelner spezifischer Substanzen ermöglichen ^{27,189} und somit auch für das Expressionssystem essenzielle Substanzen entfernen können, sind solche Methoden nicht für die Entfernung des „restlichen“ Lysins geeignet. Die Verwendung von Antikörpern sowie von Aptameren zur Lysin-Entfernung gestaltet sich als schwierig, da deren Bindungseigenschaften stark durch die chemische Umgebung beeinflusst werden können ⁴⁰⁷⁻⁴¹⁰. In diesem Zusammenhang sollte in Finkler *et al.* (2020) eine neue Methode entwickelt werden, die eine effektive Lysin-Entfernung aus einem zellfreien Expressionssystem ermöglicht, ohne dessen Genexpressionseigenschaften nachteilig zu beeinflussen.

4.4.2 Ergebnisse

Die ursprüngliche Idee war, „restliches“ Lysin in ein Protein mit einem hohen Lysin-Gehalt (High-Lysine-Dummy-Protein, (HLDP)) im Rahmen einer Genexpressionsreaktion einzubauen. Tabelle 2 zeigt die Aminosäuresequenzen verschiedener designer HLDPs, deren „High-Lysin“-codierende Genabschnitte synthetisiert (GeneArt, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA; DNA-Sequenzen im Anhang, Abbildung A1) und jeweils in pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-Tag-T500 (siehe Abbildung 10) inseriert wurden, um pBest-p15a-Or2-Or1-Pr-UTR1-HLDP-T500 zu erhalten. Dieses Plasmid wurde dem Expressionssystem als DNA-Template für eine Genexpressionsreaktion hinzugegeben. Nach Beendigung der Reaktion wurde mittels verschiedener Methoden überprüft, ob eine HLDP-Synthese stattfand.

Tabelle 2 Designte Proteine mit hohem Lysin (K)-Gehalt. Dargestellt sind die theoretisch gebildeten Proteine (Gelb: Purifikations-Tags; grün: deGFP-Tag; X: beliebige Aminosäure, außer Lysin und Histidin; X*: beliebige Aminosäure, außer Lysin und Methionin), deren codierende DNA synthetisiert und in einer Genexpressionsreaktion verwendet wurde. Anschließend wurde die Bildung des Proteins mittels verschiedener Methoden überprüft (-: nicht nachweisbar; +: nachweisbar).

Nummer	Theoretisch gebildetes Protein	Nachweisverfahren	Nachweis
1	M(K) ₁₈	SDS-PAGE	-
2	ME(K) ₄₀ -His-Tag	SDS-PAGE	-
3	MA-Strep-Tag II-SA(K) ₃₀ GS	SDS-PAGE	-
4	MA-Strep-Tag II-SA(K) ₃₀ GS-deGFP	Fluoreszenzmessung	-
5	deGFP-LY(K) ₃₀ -His-Tag	Fluoreszenzmessung	-
6	deGFP-LYTM(X*K) ₁₈	Fluoreszenzmessung	+
7	deGFP-LYTM(XK) ₅₃ -His-Tag	Fluoreszenzmessung	+

Die designten HLDPs, welche über Polylysin-Abschnitte (K)_n verfügten, konnten nicht im Expressionssystem synthetisiert werden. Dies kann dadurch erklärt werden, dass Lysin lediglich durch die Condons AAA und AAG codiert wird (siehe Abbildung 8). Folglich besitzt ein Polylysin-codierendes Gen Wiederholungen von zumindest 5 Adeninen. In der Literatur wurde bereits gezeigt, dass solche Adenin-Wiederholungen zu einer verminderten Translation führen können, da durch diese Sequenzen ein „Rutschen“ der Ribosomen auftreten kann^{411,412}.

Dieses Problem konnte dadurch umgangen werden, ein HLDP zu designen, welches lediglich an jeder zweiten Position ein Lysin aufweist. Alle anderen Aminosäuren wurden gleichmäßig über die verbleibenden Positionen verteilt. Diesbezüglich wurden zwei Konstrukte M(X*K)₁₈ und M(XK)₅₃H₆ designt. Das Konstrukt M(X*K)₁₈ enthält 18 Lysine, während alle anderen Aminosäuren nur jeweils einmal vertreten sind. X* entspricht allen Aminosäuren außer Lysin und Methionin. Da das Konstrukt M(XK)₅₃H₆ über einen His-Tag verfügt, enthält dieses Protein 53 Lysine und 6 Histidine, während alle anderen Aminosäuren dreimal vertreten sind. Somit steht X für alle Aminosäuren außer Lysin und Histidin, wobei aufgrund des Startmethionins nur zwei Methionine an die Positionen für X eingebaut werden. Beide Konstrukte konnten als Fusionsprotein mit deGFP (LYT fungiert als „Linker“) in einer zellfreien Genexpressionsreaktion synthetisiert werden, was anhand einer Fluoreszenzmessung nachgewiesen werden konnte. Hierbei ließ sich feststellen, dass die bei beiden Konstrukten gemessene Fluoreszenz-Intensität und die damit korrelierende Proteinmenge vergleichbar war (siehe Abbildung 32).

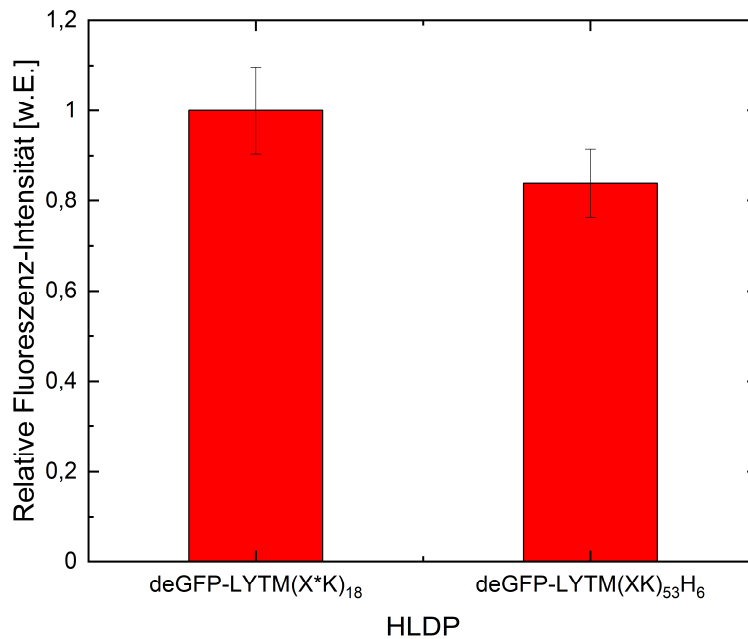


Abbildung 32 Vergleich der gemessenen Fluoreszenz-Intensitäten bei Verwendung der codierenden DNA von deGFP-LYTM(X*K)₁₈ oder von deGFP-LYTM(XK)₅₃H₆ als DNA-Template ($c = 10 \text{ nM}$) in einer zellfreien Genexpressionsreaktion.

Da in das Konstrukt M(XK)₅₃H₆ mehr Lysine eingebaut werden können als in M(X*K)₁₈ und beide Protein-Produktionsraten vergleichbar sind, ermöglicht M(XK)₅₃H₆ eine effektivere Entfernung „restlichen“ Lysins durch dessen Einbau in ein HLDP. Folglich wurde M(XK)₅₃H₆ für die durchzuführenden Experimente ausgewählt. In diesen wurde stets M(XK)₅₃H₆ als Fusionsprotein mit deGFP (GHLDP) synthetisiert, um einen visuellen Nachweis der Synthese zu ermöglichen. Das zugrundeliegende Plasmid für die rekombinante Produktion von GHLDP war pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-LYTM(XK)₅₃H₆-T500 bzw. pBest-p15a-Or2-Or1-Pr-UTR1-GHLDP-T500.

Um ein Expressionssystem mit vernachlässigbarem Lysin-Gehalt zu erhalten, mit dem weitere Genexpressionen möglich sind, war es für optimale Expressionsbedingungen notwendig, die rekombinante Produktion von GHLDP zeitlich zu begrenzen.

Für die zeitlich begrenzte Expression von *ghldp* wurde die in Finkler & Ott (2019) beschriebene Methode verwendet. Diese Methode hatte den positiven Nebeneffekt, dass die verwendete, Bead-gebundene DNA eine zusätzliche Lysin-Entfernung ermöglichte. Der zugrunde liegende Mechanismus sind elektrostatische Interaktionen zwischen Lysin und DNA^{63–65}. Neben dem Einbau in GHLDP konnte „restliches“ Lysin somit auch durch Interaktion mit Bead-gebundener DNA und deren anschließender magnetischer Separation entfernt werden. Nach Beendigung der *ghldp*-Expression wurde dem Expressionssystem mit stark verringertem Lysin-Gehalt deGFP-codierende DNA hinzugegeben. Dessen Expression konnte nur durch die zusätzliche Hinzugabe von Lysin veranlasst werden (siehe Finkler *et al.* (2020), Abbildung 1).

In einem weiteren Experiment wurde untersucht, wie viel des „restlichen“ Lysins indirekt durch den Einbau in GHLDP und direkt durch Interaktion mit Bead-gebundener DNA entfernt wird. Hierzu wurde kontinuierlich die Fluoreszenz-Intensität in unterschiedlich präparierten zellfreien Genexpressionsreaktionen bei Expression von *degfp* gemessen (siehe Finkler *et al.* (2020), Abbildung 2). In allen Fällen wurden dem Expressionssystem alle Aminosäuren außer Lysin hinzugegeben, sodass lediglich „restliches“ Lysin als Lysin-Quelle fungieren konnte. Da Lysin der limitierende Faktor der Proteinproduktion war (durch Sättigung der kontinuierlich gemessenen Fluoreszenz-Intensität erkennbar), konnten durch Bestimmung der gebildeten Proteinmenge Rückschlüsse über die Menge an „restlichem“ Lysin gewonnen werden. In den durchgeführten Experimenten wurde die Konzentration an synthetisiertem deGFP aus der Messung der Fluoreszenz-Intensität abgeleitet. Der Umrechnungsfaktor wurde aus der Fluoreszenz-Intensität der maximalen deGFP-Produktion erhalten ($c_{\max} = 0,65 \text{ mg/ml}^{28}$). Wenn davon ausgegangen wird, dass annähernd das komplette „restliche“ Lysin in deGFP eingebaut wurde, lässt sich aus der finalen deGFP-Konzentration die Konzentration an „restlichem“ Lysin bestimmen, welche zu Beginn der *degfp*-Expression im Expressionssystem vorhanden war.

Zunächst wurde die Konzentration an „restlichem“ Lysin in einem unbehandelten Expressionssystem als Referenz bestimmt. Hierbei konnte eine Konzentration von 0,025 mg/ml ermittelt werden. Während der Einbau von „restlichem“ Lysin in GHLDP dessen Konzentration um 14 % reduzierte, verursachte die Interaktion von „restlichem“ Lysin mit Bead-gebundener DNA eine Reduktion um 64 %. Wurden beide Möglichkeiten zur Entfernung von „restlichem“ Lysin verwendet, verringerte sich dessen Konzentration um 80 %.

Um zu zeigen, dass die vorgestellte Methode zur Entfernung von „restlichem“ Lysin durch Einbau in GHLDP und durch Interaktion mit Bead-gebundener DNA keinen nachteiligen Effekt auf die Genexpressionseigenschaften des Expressionssystems hat, wurde ein mit dieser Methode behandeltes Expressionssystem verwendet. Während in einem solchem System nur eine vernachlässigbare Menge an deGFP nach Zugabe der codierenden DNA synthetisiert werden konnte, bewirkte die zusätzliche Hinzugabe von „frischem“ Lysin eine hohe deGFP-Produktion. Somit kann die vorgestellte Methode als effektive Methode zur Entfernung von „restlichem“ Lysin aus einem zellfreien Expressionssystem betrachtet werden.

Zur genaueren Untersuchung des Einbaus von „restlichem“ Lysin musste die kontinuierliche Messung der Fluoreszenz-Intensität in allen Genexpressionsreaktionen, in denen Lysin der limitierende Faktor der deGFP-Produktion war, betrachtet werden. Hierbei wurde festgestellt, dass die erhaltenen Kurven einen vergleichbaren Verlauf hatten, aber unterschiedlich skalierten und somit einen selbstähnlichen Charakter aufwiesen.

Diese Selbstähnlichkeit konnte durch Reskalierung der Zeit mit einer charakteristischen Zeit τ sowie der Fluoreszenz-Intensität mit einer charakteristischen Amplitude A gezeigt werden (siehe Finkler *et al.*

(2020), Tabelle 1 und Abbildung 3). Folglich wurde ein Potenzgesetz der Form $A \sim \tau^\gamma$ vorgeschlagen. Durch Anpassen der für A und τ bestimmten Werte (siehe Finkler *et al.* (2020), Tabelle 1) mit diesem Gesetz konnte für den Fitparameter γ ein Wert von $\gamma = -0,64 \pm 0,06$ erhalten werden. Während ein Wert von -1 einen einfachen diffusionslimitierten Prozess widerspiegelt⁴¹³⁻⁴¹⁶, lieferte der erhaltene Wert Hinweise für weitere Prozesse, welche in den Einbau von „restlichem“ Lysin in deGFP involviert sind. Da eine Freisetzung von Lysin durch Proteolyse widerlegt werden konnte, wurde angenommen, dass eine mit fortschreitender Genexpressionsdauer zunehmende Erschöpfung des Expressionssystems²⁴ den von -1 abweichenden Exponenten bewirkte.

4.4.3 Zusammenfassung und Ausblick

In Finkler *et al.* (2020) wurde eine neue Methode präsentiert, mit der es möglich ist, in einem auf *E. coli* basierenden, zellfreien Expressionssystem vorhandenes „restliches“ Lysin effektiv zu entfernen, ohne dessen Genexpressionseigenschaften nachteilig zu verändern. Diesbezüglich entstand zunächst die Idee, „restliches“ Lysin in ein Dummy-Protein mit einem hohen Lysin-Gehalt (HLDP) im Rahmen einer Genexpression einzubauen. Hieraus sollte ein Expressionssystem mit annähernd vernachlässigbarem Lysin-Gehalt resultieren, welches für eine weitere Genexpression verwendet werden kann. Zu diesem Zweck musste die HLDP-codierende DNA jedoch vor Durchführung der zweiten Genexpression entfernt werden, um eine verminderte Genexpressionsrate aufgrund von Koexpression zu vermeiden. In diesem Zusammenhang wurde die in Finkler & Ott (2019) beschriebene Methode verwendet, um eine räumlich und zeitlich begrenzte Expression von *hldp* durchzuführen. Dadurch konnte die Menge des „restlichen“ Lysins im zellfreien Expressionssystem um 80 % reduziert wurde. Hierbei wurde festgestellt, dass die Produktion von HLDP nur einen geringen Anteil an der Lysin-Entfernung ausmachte. Der größte Anteil an „restlichem“ Lysin wurde durch dessen Interaktion mit an den Beads gebundener DNA und anschließender magnetischer Separation entfernt. Durch Hinzugabe einer neuen codierenden DNA sowie von Lysin konnte eine weitere Genexpression veranlasst werden. Somit wurden die Genexpressionseigenschaften des Expressionssystems nicht nachteilig verändert.

Die neu entwickelte Methode ermöglicht die Produktion zellfreier Expressionssysteme mit vernachlässigbarem Lysin-Gehalt und könnte somit in Experimenten, in denen die Lysin-Konzentration ein wichtiger Faktor ist, eine entscheidende Rolle spielen. Ein Beispiel ist der vollständige seitenkettenspezifische Einbau von Lysin-Analoga, welcher in Präsenz von Lysin unter Umständen nicht durchgeführt werden kann⁵⁷.

4.5 Finkler *et al.* (2021)

4.5.1 Hintergrund

Das Lysin-Derivat Hydroxylysin (L-5-Hydroxylysin) ist eine wichtige Komponente von Kollagenfasern und fungiert in diesen als Ausgangspunkt für die Ausbildung von Quervernetzungen⁴¹⁷⁻⁴¹⁹. Da in der Natur kein direkter Einbau von Hydroxylysin in die naszierenden Polypeptidketten von Kollagen erfolgt^{418,420}, resultiert Hydroxylysin aus der posttranslationalen Modifikation von Lysin^{418,421,422}. Dennoch ist der direkte Einbau von besonderem Interesse, da dadurch die rekombinante Produktion von modifizierten Kollagenen oder von quervernetzenden Proteinen ermöglicht wird^{417-419,423-426}.

In der Literatur konnte bereits gezeigt werden, dass der vollständige seitenkettenspezifische Einbau von Hydroxylysin aufgrund geringer Mengen an Lysin, welche im Allgemeinen in Expressionssystemen vorhanden sind, nicht ohne Weiteres durchgeführt werden konnte⁵⁷. Somit wurde in Finkler *et al.* (2021) nach einer Möglichkeit gesucht, diesen Einbau dennoch durchzuführen. In diesem Zusammenhang wurde auch das Einbauverhalten von im Expressionssystem vorhandenen („restlichem“) und dem Expressionssystem hinzugefügten („frischem“) Lysin näher betrachtet.

4.5.2 Ergebnisse

In einem ersten Schritt wurde zunächst untersucht, ob „restliches“ und „frisches“ Lysin im zellfreien Expressionssystem unterschiedliche Charakteristika beim Einbau in Proteine aufwiesen. Als Modellprotein wurde deGFP verwendet, da dessen Konzentration und somit die Menge an eingebautem Lysin, in Analogie zu Finkler *et al.* (2020), durch Messung der Fluoreszenz-Intensität bestimmt werden kann. In den durchgeführten Experimenten wurden zu einem unbehandelten Expressionssystem sowie zu einem Expressionssystem, dessen „restliches“ Lysin durch Interaktion mit Bead-gebundener DNA größtenteils entfernt wurde (siehe Finkler *et al.* (2020)), unterschiedliche Mengen an „frischem“ Lysin hinzugegeben. Im Gegensatz zu der in Finkler *et al.* (2020) beschriebenen Methode wurde auf eine Entfernung von „restlichem“ Lysin durch den Einbau in GHLDLP verzichtet, da die Entfernung durch Interaktion mit Bead-gebundener DNA ausreichte und somit keine für die Genexpression benötigten Ressourcen zur Lysin-Entfernung verbraucht wurden. Alle anderen Aminosäuren wurden den Expressionssystemen in ausreichender Menge ($c = 1 \text{ mM}$ bzw. $c = 0,83 \text{ mM}$ für Leucin) hinzugegeben, sodass Lysin der limitierende Faktor der Genexpression war. Anschließend wurde die Fluoreszenz-Intensität kontinuierlich über die gesamte Dauer der Genexpressionsreaktion gemessen (siehe Finkler *et al.* (2021), Abbildung 1).

Hierbei konnte festgestellt werden, dass sich die aus der kontinuierlichen Messung der Fluoreszenz-Intensität erhaltenen Kurven bei geringen Konzentrationen an „frischem“ Lysin aus 2 Regimen

zusammensetzten. Während eine anfängliche deGFP-Produktion nach wenigen Stunden bereits eine Sättigung erreichte, startete später eine erneute deGFP-Produktion, welche nach mehreren Stunden auch gesättigt war. Für die Höhe der ersten Sättigung konnte eine Abhängigkeit zur Konzentration an „frischem“ Lysin festgestellt werden. Der Zeitpunkt, an dem die deGFP-Produktion wieder aufgenommen wurde, sowie die Menge des ab dann gebildeten deGFP, unterschied sich je nach Menge an „restlichem“ Lysin im Expressionssystem. Eine Abhängigkeit zur Konzentration an „frischem“ Lysin konnte jedoch nicht festgestellt werden. Lediglich für hohe Konzentrationen an „frischem“ Lysin verschwand das 2. Regime, da keine anfängliche Sättigung erreicht wurde.

Diese Beobachtungen legten es nahe, dass im Expressionssystem zunächst „frisches“ Lysin in deGFP eingebaut wird. Nach dessen Erschöpfung erreicht die deGFP-Produktion eine Sättigung. Die Wiederaufnahme der deGFP-Produktion wird durch den Einbau des „restlichen“ Lysins bewirkt. Ist „frisches“ Lysin im Überschuss vorhanden, so kann der Einbau von „restlichem“ Lysin supprimiert werden.

In einem Expressionssystem mit vernachlässigbarem Gehalt an „restlichem“ Lysin sollte somit die finale Sättigung der deGFP-Produktion mit der ersten Sättigung der deGFP-Produktion in einem unbehandelten Expressionssystem bei gleicher Konzentration an „frischem“ Lysin übereinstimmen. Dies konnte durch Betrachtung der kontinuierlichen Messung der Fluoreszenz-Intensität bestätigt werden (siehe Finkler *et al.* (2021), Abbildung 2).

Während die zuvor erwähnten Experimente Hinweise auf ein unterschiedliches Einbauverhalten von „restlichem“ und „frischem“ Lysin lieferten, musste der Einbau der verschiedenen Spezies zu verschiedenen Zeitpunkten auch experimentell bestätigt werden. Zu diesem Zweck wurde einem unbehandelten Expressionssystem isopenmarkiertes Lysin hinzugegeben. Diese Isopenmarkierung hat keine Auswirkungen auf den Einbau von Lysin in Proteine^{427–429} und ermöglicht die Unterscheidung von deGFP mit eingebautem „restlichem“ und deGFP mit eingebautem „frischem“ Lysin mittels Massenspektrometrie.

Zu diesem Zweck musste zunächst eine massenspektrometrische Methode entwickelt werden, welche eine möglichst sensitive Detektion von deGFP ermöglichte. Neben den Einstellungen des für die Messung verwendeten Massenspektrometers (SolariX FTICR-ESI, Bruker Corporation, Billerica, Massachusetts, USA) hat auch die Wahl des Lösungsmittels einen entscheidenden Einfluss auf die Messung^{430–434} (siehe Abbildung 33 am Beispiel für EGFP). Die optimierten, bei den Messungen verwendeten Einstellungen können Tabelle A1 im Anhang entnommen werden.

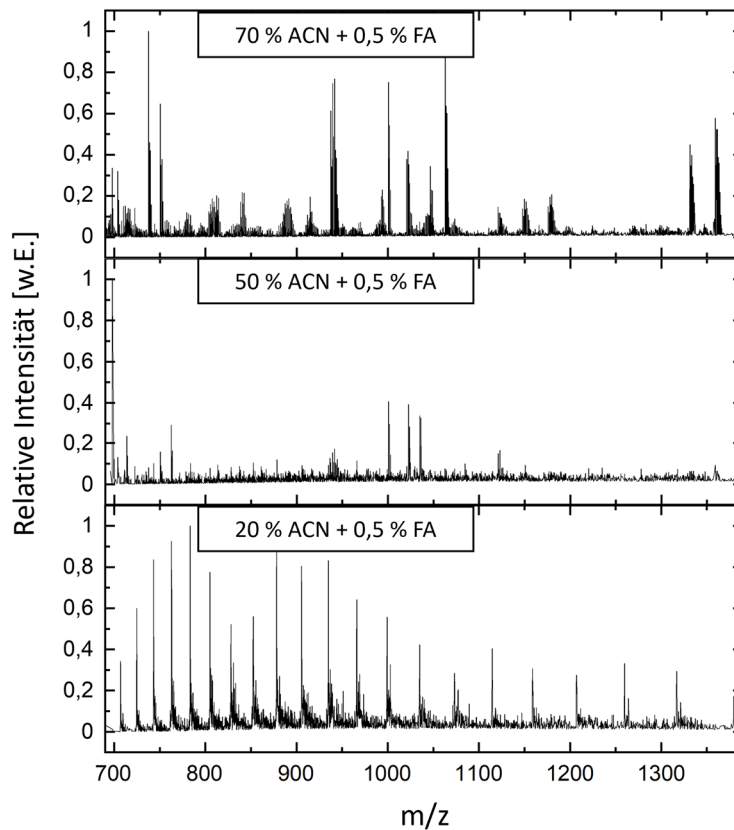


Abbildung 33 Massenspektrum von EGFP bei Verwendung von verschiedenen Lösungsmitteln. Während bei 50 % ACN (Acetonitril) + 0,5 % FA (Ameisensäure) fast keine Signale detektierbar sind, können die Signale bei 70 % ACN + 0,5 % FA keinem Protein und somit auch nicht EGFP zugeordnet werden. Lediglich die Verwendung von 20 % ACN + 0,5 % FA führt zu einem Massenspektrum, in dem verschiedene Ladungszustände von EGFP zu erkennen sind.

Wie aus Abbildung 33 ersichtlich, eignet sich 20%iges Acetonitril, versetzt mit 0,5 % Ameisensäure, sehr gut als Lösungsmittel für EGFP. Somit wurde dieses Lösungsmittel auch für deGFP verwendet.

Für alle in dieser Arbeit durchgeführten massenspektrometrischen Messungen von deGFP wurden die folgenden Schritte durchgeführt. Zunächst wurde zellfrei synthetisiertes deGFP, welches über einen His-Tag verfügt, mittels immobilisierter-Metallionen-Affinitätschromatographie (IMAC) (His-Spin Protein Miniprep, Zymo Research Europe GmbH, Freiburg im Breisgau, Deutschland) purifiziert (Purifikation gemäß Hersteller-Protokoll ⁴³⁵). Anschließend wurde das purifizierte deGFP entsalzt und aufkonzentriert. Zu diesem Zweck wurden Mikrozentrifugierröhrchen verwendet, die eine Separation und Aufkonzentrierung von Molekülen mit einer Masse größer als 10 kDa ermöglichen (Pierce™ Protein Concentrators PES, 10 K MWCO, 0.5 mL, Thermo Scientific, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Mit diesen wurde das purifizierte deGFP zweimal mit Reinstwasser und anschließend zweimal mit 20%igem Acetonitril, versetzt mit 0,5 % Ameisensäure, gewaschen (Durchführung gemäß Hersteller-Protokoll ⁴³⁶). Der Überstand bestand somit aus aufgereinigtem, hochkonzentriertem deGFP in 20%igem Acetonitril, versetzt mit 0,5 % Ameisensäure. Aufgrund des geringen Volumens dieser Probe (ca. 20 µl) war eine direkte Injektion in das Massenspektrometer

4 Diskussion

schwierig durchzuführen, weshalb für die Injektion ein konstanter Fluss (0,08 ml/min) von 50 % Methanol verwendet wurde.

Ein Schema der entwickelten massenspektrometrischen Methode kann Abbildung 34 entnommen werden.

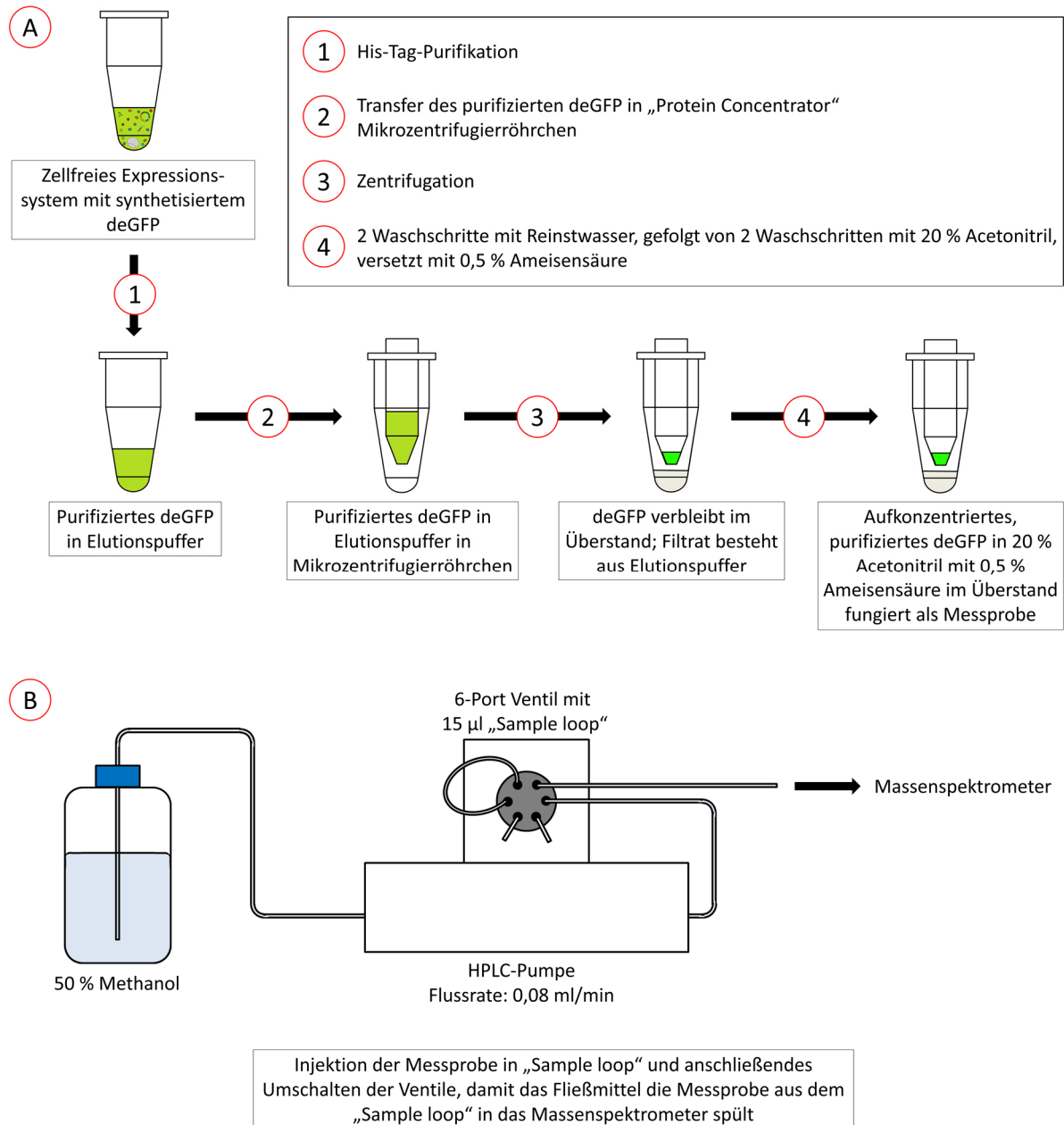


Abbildung 34 Schematische Darstellung der entwickelten massenspektrometrischen Methode zum Nachweis von deGFP. Die Abbildung zeigt die Purifikation und Aufkonzentrierung von in einer zellfreien Genexpressionsreaktion synthetisiertem deGFP (A). Die purifizierte und aufkonzentrierte deGFP-Lösung wird aufgrund des geringen Volumens indirekt über einen Flüssigkeitsstrom in das Massenspektrometer injiziert (B).

Anschließend wurde die Probe mittels des Massenspektrometers analysiert. In diesem Zusammenhang wurde mit Hilfe der Software „Compass DataAnalysis“ (Bruker Compass DataAnalysis Version 5.0,

Bruker Corporation, Billerica, Massachusetts, USA) zunächst ein Massenspektrum generiert. Die zugehörigen Daten wurden daraufhin mit dem in Compass DataAnalysis integrierten „Maximum entropy“-Algorithmus prozessiert, um ein dekonvolutioniertes Massenspektrum zu erhalten.

Für die durchzuführenden Experimente bzgl. des Einbaus von isotoopenmarkiertem Lysin wurden einem unbehandelten Expressionssystem verschiedene Mengen an isotoopenmarkiertem Lysin hinzugegeben und die Genexpression nach verschiedenen Zeitpunkten gestoppt. Im Anschluss wurde das zellfrei synthetisierte deGFP aufgereinigt und massenspektrometrisch untersucht (siehe Finkler *et al.* (2021), Abbildung 3). Wurde kein isotoopenmarkiertes Lysin hinzugefügt, so wurde lediglich „restliches“ Lysin in deGFP eingebaut. Bei einer Konzentration von 0,1 mM an isotoopenmarkiertem Lysin konnte festgestellt werden, dass nach einer Genexpressionsdauer von 2 h (Zeitpunkt der ersten Sättigung der deGFP-Produktion in einem unbehandelten Expressionssystem bei geringen Konzentrationen an „frischem“ Lysin) kein Einbau von „restlichem“ Lysin nachweisbar war, sondern nur deGFP mit isotoopenmarkiertem Lysin gebildet wurde. Bei einer Verlängerung der Genexpressionsdauer auf 16 h (Zeitpunkt der finalen Sättigung der deGFP-Produktion) wurde neben deGFP mit eingebautem isotoopenmarkiertem Lysin auch deGFP mit eingebautem „restlichem“ Lysin nachgewiesen. Wurde dem Expressionssystem eine erhöhte Konzentration an isotoopenmarkiertem Lysin (0,6 mM) zur Verfügung gestellt, so wurde lediglich dieses in deGFP eingebaut und der Einbau von „restlichem“ Lysin unterbunden. Somit konnte die Vermutung, dass „restliches“ und „frisches“ Lysin zu verschiedenen Zeitpunkten vom zellfreien Expressionssystem in Proteine eingebaut wird, experimentell verifiziert werden. Zunächst verwendet das System zu Beginn einer Genexpression „frisches“ Lysin für die Proteinbiosynthese. Erst wenn dieses erschöpft ist, wird zu einem späteren Zeitpunkt „restliches“ Lysin eingebaut.

Aufgrund dieser neuen Erkenntnis wurde zunächst versucht, den Einbau von Hydroxylysin (Strukturformel-Vergleich mit Lysin, siehe Finkler *et al.* (2021), Abbildung 4) in Proteine innerhalb eines Expressionssystems mit „restlichem“ Lysin durch einen Überschuss an Hydroxylysin zu forcieren. Hierfür wurden einem unbehandelten Expressionssystem verschiedene Mengen an Hydroxylysin hinzugegeben. Alle kanonischen Aminosäuren außer Lysin waren in einer Konzentration von 1 mM (bzw. 0,83 mM für Leucin) vorhanden. Nach einer Genexpressionsdauer von 16 h wurde das synthetisierte deGFP purifiziert und massenspektrometrisch untersucht (siehe Finkler *et al.* (2021), Abbildung 5).

Hierbei konnte festgestellt werden, dass selbst bei einer erhöhten Hydroxylysin-Konzentration kein vollständiger Einbau von Hydroxylysin in deGFP erreicht werden konnte. Vielmehr wurden alle möglichen deGFP-Mischspezies mit verschiedenen Verhältnissen an eingebautem „restlichem“ Lysin und Hydroxylysin gebildet. Eine Erhöhung der Hydroxylysin-Konzentration hatte lediglich eine verstärkte Bildung der Hydroxylysin-reichen Spezies zur Folge. Des Weiteren konnte festgestellt

werden, dass hohe Konzentrationen an Hydroxylysin ($c > 50 \text{ mM}$) die Proteinproduktion inhibierten (siehe Finker *et al.* (2021), Abbildung 7 und Abbildung S 2).

Da somit die Konzentration an Hydroxylysin nicht beliebig erhöht werden konnte, aber ein großes Verhältnis zwischen Hydroxylysin und „restlichem“ Lysin ($c(\text{Hydroxylysin})/c(\text{„restliches“ Lysin}) > 458$) elementar für den effektiven vollständigen Einbau von Hydroxylysin in deGFP ist, konnte eine Reduktion des „restlichen“ Lysins eine Vergrößerung dieses Verhältnisses bewirken.

Zu diesem Zweck wurde „restliches“ Lysin durch Interaktion mit Bead-gebundener DNA, wie in Finkler *et al.* (2020) beschrieben, aus einem zellfreien Expressionssystem entfernt. Auch hier wurde im Gegensatz zu der in Finkler *et al.* (2020) beschriebenen Methode auf eine Entfernung von „restlichem“ Lysin durch den Einbau in GHLDLP wegen der zuvor genannten Gründe verzichtet.

Mit einem solchen System wurde eine Genexpressionsreaktion bei einem Verhältnis zwischen Hydroxylysin und „restlichem“ Lysin von $c(\text{Hydroxylysin})/c(\text{„restliches“ Lysin}) \approx 665$ durchgeführt. In dieser Reaktion konnte der vollständige seitenkettenspezifische Einbau von Hydroxylysin in deGFP realisiert werden, während der Einbau von „restlichem“ Lysin supprimiert wurde. Dies konnte durch Massenspektrometrie und SDS-PAGE verifiziert werden (siehe Finker *et al.* (2021), Abbildung 6 und Abbildung 7).

4.5.3 Zusammenfassung und Ausblick

In Finkler *et al.* (2021) konnte der vollständige seitenkettenspezifische Einbau von Hydroxylysin in Proteine in einem auf *E. coli* basierenden, zellfreien Expressionssystem erreicht werden. Da dieser in bisherigen Experimenten durch „restliches“ Lysin verhindert wurde⁵⁷, erfolgte zunächst eine Untersuchung von dessen Einbauverhalten. Hierbei konnte gezeigt werden, dass „restliches“ und „frisches“ Lysin im zellfreien Expressionssystem zu unterschiedlichen Zeitpunkten in Proteine eingebaut werden. Des Weiteren hatte eine Erhöhung der Konzentration an „frischem“ Lysin die Suppression des Einbaus von „restlichem“ Lysin zur Folge. Diese Erkenntnisse sollten anschließend dafür verwendet werden, den Einbau von Hydroxylysin zu realisieren.

Bei der Verwendung von Hydroxylysin als „frische“ Aminosäure konnte festgestellt werden, dass das Verhältnis zwischen Hydroxylysin und „restlichem“ Lysin das entscheidende Kriterium für den vollständigen seitenkettenspezifischen Einbau von Hydroxylysin ist. Dennoch konnte die Hydroxylysin-Konzentration nicht beliebig erhöht werden, da dadurch die Genexpression supprimiert wurde. Dies erforderte die Entfernung von „restlichem“ Lysin aus dem Expressionssystem. Zu diesem Zweck wurde eine gering modifizierte Version der in Finkler *et al.* (2020) beschriebenen Methode verwendet, bei der „restliches“ Lysin lediglich durch Interaktion mit Bead-gebundener DNA entfernt wurde. Durch die stark reduzierte Lysin-Konzentration und eine hohe Hydroxylysin-Konzentration konnte das

Expressionssystem zum vollständigen seitenkettenspezifischen Einbau von Hydroxylysin als Lysin-Ersatz forciert werden. Hierbei konnten Proteine produziert werden, welche Hydroxylysin an allen Lysin-Positionen aufwiesen.

Mit dieser neuen Methode konnte somit der direkte Einbau von Hydroxylysin im Rahmen einer Genexpressionsreaktion realisiert werden. Dies könnte in Zukunft eine Vielzahl an möglichen Anwendungen mit sich ziehen, welche im Folgenden näher erläutert werden.

Der direkte Einbau von Hydroxylysin erlaubt die rekombinante Produktion von Proteinen, welche an spezifischen Positionen Hydroxylysin besitzen und durch Ausbildung von Quervernetzungen spezielle Strukturen ausbilden können^{419,425}. Zusätzlich können rekombinant produzierte, Hydroxylysin enthaltende Proteine für neue Untersuchungen bzgl. Hydroxylysin-assoziiierter Krankheiten verwendet werden^{421,425,437}.

Des Weiteren ermöglicht der direkte Einbau von Hydroxylysin die Proteinbiosynthese von modifizierten Kollagenen, welche z. B. in der Medizin für die Entwicklung neuer Haut- oder Knochengewebs-Transplantate verwendet werden können^{423,424}.

Die rekombinante Produktion von Kollagen ermöglicht auch die rekombinante Produktion von Gelatine, einem Kollagen-Derivat⁴³⁸. Da beides aktuell nur aus tierischen Quellen oder aus Zellkulturen mit tierischem Ursprung gewonnen werden kann^{426,438}, ermöglicht deren rekombinante Produktion eine Alternative, bei der keine Tiere geschädigt werden. Dies könnte in Zukunft eine wichtige Rolle in der Industrie spielen.

Die präsentierte Methode zum vollständigen seitenkettenspezifischen Einbau von Hydroxylysin in Proteine kann auch für andere Lysin-Analoga benutzt werden. Die einzige Voraussetzung ist, dass diese von der endogenen Translations-Maschinerie akzeptiert werden^{22,57,189}.

In Finkler *et al.* (2021) wurde bereits gezeigt, dass mit der verwendeten Methode zur Entfernung von „restlichem“ Lysin auch „restliches“ Arginin aus einem zellfreien Expressionssystem entfernt werden kann (siehe Finkler *et al.* (2021), Abbildung S 1). Auch wenn das „restliche“ Arginin den vollständigen Einbau von beispielsweise Canavanin in Proteine nicht negativ beeinflusste^{57,174}, lässt sich dies nicht auf andere Arginin-Analoga übertragen, da jede nicht-kanonische Aminosäure individuelle Eigenschaften bzgl. des Einbaus aufweist^{22,57,106,189,245}. Somit könnte die in Finkler *et al.* (2021) präsentierte Methode auch für den vollständigen seitenkettenspezifischen Einbau von Arginin-Analoga in Proteine hilfreich sein.

Zwei nicht-kanonische Aminosäuren, deren vollständiger Einbau in Proteine von besonderem Interesse ist, sind das Arginin-Derivat Argpyrimidin und das Lysin-Derivat Carboxy-Methyl-Lysin (CML) (siehe Abbildung 35).

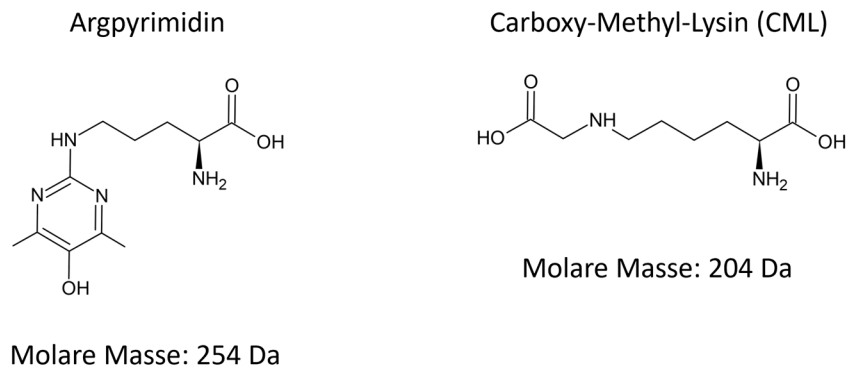


Abbildung 35 Strukturformeln von L-Argpyrimidin (links) und L-Carboxy-Methyl-Lysin (CML) (rechts).

Da es sich bei diesen beiden nicht-kanonischen Aminosäuren um „Advanced glycation end products“ (AGEs) handelt, könnte sich deren vollständiger Einbau in rekombinante Proteine als ein wichtiger Bestandteil in der Untersuchung degenerativer Erkrankungen erweisen^{439–445}. Um diesen Einbau zu bewerkstelligen, kann die in Finkler *et al.* (2021) präsentierte Methode ein essenzieller Faktor sein. Zumindest für CML konnte bereits experimentell bestätigt werden, dass „restliches“ Lysin den vollständigen Einbau von CML in Proteine verhindert (siehe Abbildung 36).

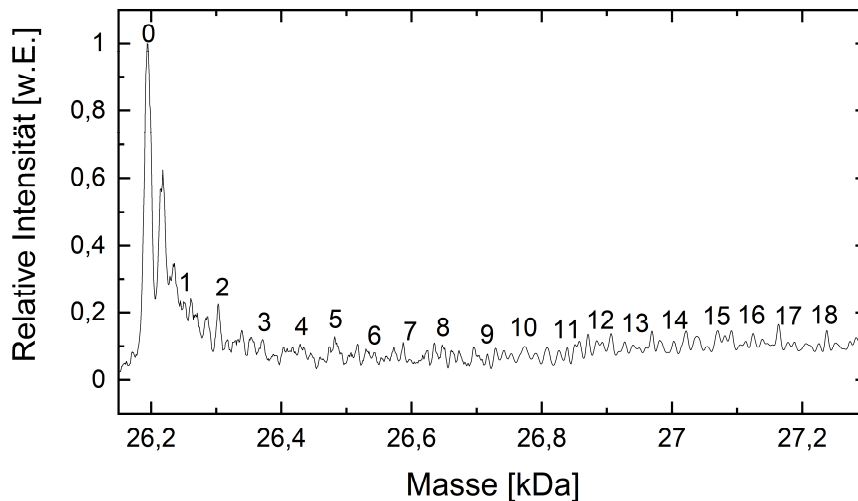


Abbildung 36 Dekonvolutionsmassenspektrum von deGFP, welches in einem unbehandelten zellfreien Expressionssystem bei Präsenz von 1 mM CML ohne Lysin-Hinzugabe synthetisiert wurde. Dieses zeigt, dass das Expressionssystem „restliches“ Lysin stark präferierte (hoher Peak von nativem deGFP; $M_{\text{deGFP}} = 26,193$ kDa) und somit ein vollständiger Einbau von CML verhindert wurde. Peaks, welche deGFP (18 Lysin-Positionen) mit mindestens einem eingebauten CML repräsentieren ($M_{\text{CML}} - M_{\text{Lys}} = 58$ Da; $M_{\text{deGFPnCML}} = M_{\text{deGFP}} + n \cdot 58$ Da, $n = 1, \dots, 18$) sind nur sehr schwach bis gar nicht vertreten (Die Nummern über den Peaks stehen für die Anzahl an CML im deGFP). Die Daten wurden auf das Maximum des Massenspektrums normiert. Für die Generierung des Massenspektrums wurde ein ESI-TOF-Massenspektrometer verwendet^{††}.

^{††} Die präsentierten Daten wurden im Rahmen einer Kooperation mit Prof. Dr. Nediljko Budisa am Institut für Chemie an der technischen Universität Berlin generiert.

Die in Finkler *et al.* (2021) beobachtete Diskriminierung von Hydroxylysin wurde u. a. durch den „Kinetic Proofreading“-Prozess^{54,58,61,62} (siehe Abschnitt 2.7) verursacht. In der klassischen Betrachtung dieses Prozesses wurden jedoch keine kompetitiven Bedingungen berücksichtigt⁵⁸, welche in den Experimenten vorlagen. Somit könnten die erhaltenen Ergebnisse als Ausgangspunkt für eine neue Betrachtung des „Kinetic Proofreading“-Prozesses angesehen werden. Zukünftige Experimente erfordern jedoch die Kenntnis, welche Mengen der konkurrierenden Aminosäuren in Proteine eingebaut werden. In diesem Zusammenhang wurden bereits weitere Experimente durchgeführt, welche in Abschnitt 5 näher beschrieben werden.

5 Weitere Ergebnisse

Herkömmliche Methoden zur Bestimmung der Konzentration eines spezifischen Proteins unterliegen diversen Einschränkungen⁴⁴⁶⁻⁴⁴⁸. Somit eignen sich diese nur bedingt für die quantitative Bestimmung von Proteinen mit eingebauten kanonischen und Proteinen mit eingebauten nicht-kanonischen Aminosäuren. Aus diesem Grund ist die Entwicklung neuer quantitativer Methoden erforderlich, die eine solche Bestimmung erlauben. In den Abschnitten 5.1 und 5.2 werden zwei mögliche Methoden vorgestellt.

5.1 Spektroskopische Methode zur Untersuchung des Aminosäure-Einbaus

In Analogie zu Finkler *et al.* (2020) und Finkler *et al.* (2021) kann aus der gemessenen Fluoreszenz-Intensität die Konzentration von deGFP mit eingebauten nicht-kanonischen Aminosäuren abgeleitet werden. Hierbei muss vorausgesetzt sein, dass der Einbau von diesen nicht-kanonischen Aminosäuren in deGFP dessen Fluoreszenzeigenschaften nicht beeinflusst. Zu diesem Zweck muss zunächst überprüft werden, ob die zu ersetzende kanonische Aminosäure im nativen deGFP eine Interaktion mit dessen Fluorophor aufweist.

In der Literatur wurde bereits gezeigt, dass deGFP mit eingebautem Canavanin andere Fluoreszenzeigenschaften aufweist als natives deGFP¹⁷⁴. Die Interaktion eines Arginins mit dem Fluorophor wird als Grund für diese veränderten Fluoreszenzeigenschaften vermutet¹⁷⁴.

Für Hydroxylysin gestaltet sich die Situation anders, da keines der 18 Lysine in deGFP einen Kontakt mit dem Fluorophor aufweist (siehe Abbildung 37, Bändermodell von EGFP, einer verlängerten Form von deGFP^{303,309,310}).

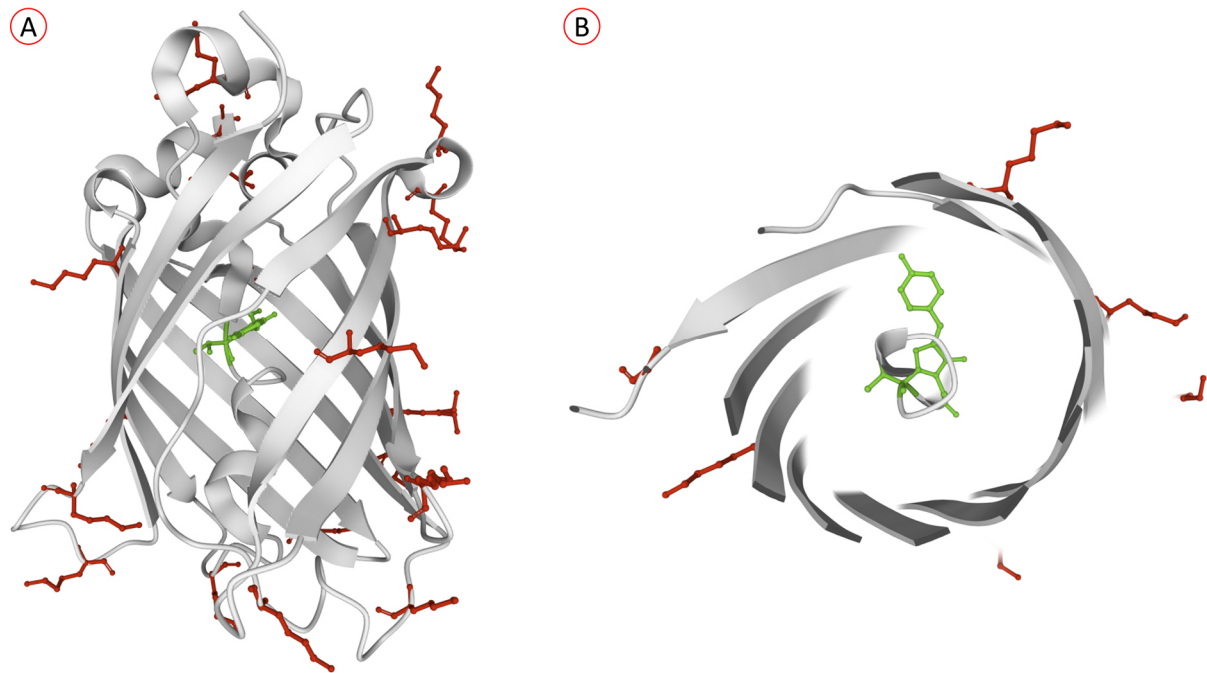


Abbildung 37 Bändermodell von EGFP (PDB ID 2Y0G ^{303,309}, bereitgestellt von RCSB PDB. Für die Generierung der Abbildung wurde der Mol* Viewer ³¹⁰ verwendet) (A: Seitenansicht, B: Querschnitt auf Ebene des Fluorophors). Anhand beider Ansichten kann gezeigt werden, dass keine Interaktionen zwischen dem Fluorophor (grün) und den im EGFP vorhandenen Lysinen (rot) auftreten.

Aus diesem Grund kann angenommen werden, dass deGFP mit eingebautem Hydroxylysin und natives deGFP gleiche Fluoreszenzeigenschaften aufweisen. Dies muss jedoch in zukünftigen Experimenten validiert werden.

Unter der Annahme gleicher Fluoreszenzeigenschaften kann durch Messung der Fluoreszenz-Intensität, in Analogie zu Finkler *et al.* (2020) und Finkler *et al.* (2021), die Konzentration von deGFP mit eingebautem Hydroxylysin bestimmt werden. Eine quantitative Bestimmung von Proteinen mit eingebautem „restlichen“ Lysin und Proteinen mit eingebautem Hydroxylysin ist jedoch nicht direkt möglich. Diese erfordert zusätzliche Betrachtungen.

In Finkler *et al.* (2021) konnte gezeigt werden, dass „frisches“ und „restliches“ Lysin in zellfreien Expressionssystemen zu unterschiedlichen Zeitpunkten in Proteine eingebaut wird. Ausgehend von dieser Beobachtung kann angenommen werden, dass gleiches für „frisches“ Hydroxylysin und „restliches“ Lysin gilt. Aus diesem Grund wurde in Analogie zu Finkler *et al.* (2021) folgendes Experiment durchgeführt: Einem zellfreien Expressionssystem, dessen „restliches“ Lysin durch Interaktion mit Bead-gebundener DNA größtenteils entfernt wurde, wurden unterschiedliche Mengen an „frischem“ Hydroxylysin hinzugegeben. Anschließend wurde *degfp* exprimiert und kontinuierlich die Fluoreszenz-Intensität gemessen. Die Entfernung des „restlichen“ Lysins war notwendig, da ein zu hoher Gehalt von „restlichem“ Lysin keinen vollständigen Einbau von Hydroxylysin erlaubte.

5 Weitere Ergebnisse

Die Ergebnisse des durchgeführten Experimentes können Abbildung 38 entnommen werden (siehe Abbildung 38 B). Als Referenz enthält die Abbildung die Ergebnisse des ursprünglichen Experimentes (siehe Finkler *et al.* (2021), Abbildung 1 A), bei dem einem unbehandelten Expressionssystem unterschiedliche Mengen an „frischem“ Lysin hinzugegeben wurden (siehe Abbildung 38 A).

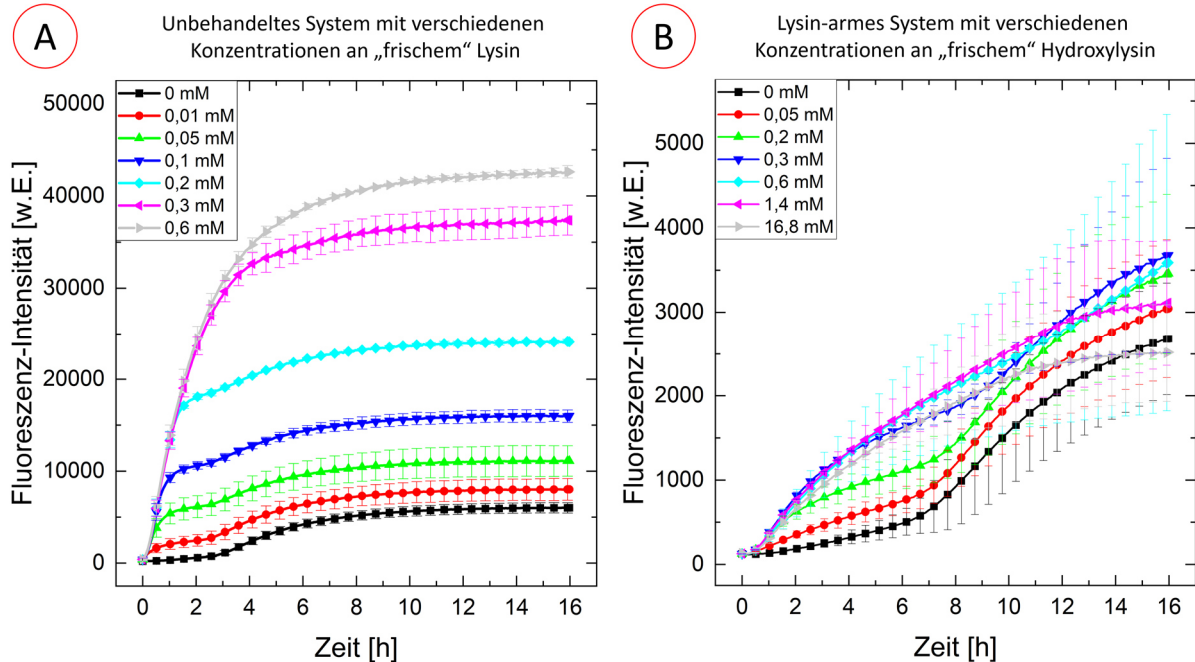


Abbildung 38 Kontinuierliche Messung der Fluoreszenz-Intensität für verschiedene Konzentrationen an Lysin bzw. Hydroxylysin, welches zu unterschiedlich behandelten zellfreien Genexpressionsreaktionen hinzugefügt wurde. Die Abbildung zeigt die Messungen für ein unbehandeltes Expressionssystem, in dem verschiedene Konzentrationen an hinzugefügtem Lysin vorhanden waren (A), sowie für ein Lysin-armes Expressionssystem, in dem verschiedene Konzentrationen an hinzugefügtem Hydroxylysin vorhanden waren (B).

Abbildung 38 zeigt analoge Ergebnisse für die beiden präsentierten Experimente. Bei geringen Konzentrationen der „frischen“ Aminosäure setzten sich die aus der kontinuierlichen Messung der Fluoreszenz-Intensität erhaltenen Kurven aus zwei unterschiedlichen Regimen zusammen. Der Übergang vom ersten in das zweite Regime ist durch einen Wendepunkt gekennzeichnet. Im Gegensatz dazu verschwand das zweite Regime für hohe Konzentrationen der „frischen“ Aminosäure.

Somit kann in Analogie zu Finkler *et al.* (2021) angenommen werden, dass zunächst Hydroxylysin in Proteine eingebaut wird (erstes Regime), bevor ein Einbau von „restlichem“ Lysin zu einem späteren Zeitpunkt stattfindet (zweites Regime). Für hohe Konzentrationen an „frischem“ Hydroxylysin war lediglich ein einzelnes Regime erkennbar, was eine Suppression des Einbaus von „restlichem“ Lysin nahelegt. Dies konnte experimentell für eine Konzentration von 16,8 mM an „frischem“ Hydroxylysin in Kombination mit einem Verhältnis zwischen Hydroxylysin und „restlichem“ Lysin von $c(\text{Hydroxylysin})/c(\text{„restliches“ Lysin}) \approx 665$ bestätigt werden (siehe Finkler *et al.* (2021), Abbildung 6 und Abbildung 7).

Anhand der in Abbildung 38 dargestellten, durch kontinuierliche Messung der Fluoreszenz-Intensität erhaltenen Kurven kann der Zeitpunkt bestimmt werden, an dem der Einbau von „restlichem“ Lysin begann (durch einen Wendepunkt repräsentiert). Folglich kann bis zu diesem Zeitpunkt die Konzentration von deGFP mit eingebauten „frischen“ Aminosäuren (Lysin oder Hydroxylysin) aus der Fluoreszenz-Intensität abgeleitet werden (Berechnung gemäß Finkler *et al.* (2020) und Finkler *et al.* (2021)). In diesem Zusammenhang kann auch für verschiedene Konzentrationen der „frischen“ Aminosäure gezeigt werden, wie viel von dieser in deGFP eingebaut wird. Dies erlaubt die Bestimmung der Effektivität des Einbaus in Abhängigkeit von der Konzentration (siehe Tabelle 3).

Tabelle 3 Vergleich der Konzentration an „frischem“ Lysin bzw. Hydroxylysin, das in deGFP eingebaut wurde (nach 2, 6 oder 16 h: Zeitpunkt vor dem Einbau von „restlichem“ Lysin, repräsentiert durch einen Wendepunkt in Abbildung 38) und der Konzentration an „frischem“ Lysin bzw. Hydroxylysin, welches zu Beginn der zellfreien Genexpressionsreaktion im Expressionssystem vorhanden war. Das Verhältnis zwischen eingebautem und vorhandenem „frischem“ Lysin bzw. Hydroxylysin spiegelt die Effektivität des Einbaus wider.

Einbau von „frischem“ Lysin		
c(eingebautes Lys) [mg/ml]	c(„frisches“ Lys) [mg/ml]	Verhältnis
0,0029 (2 h)	0,0015	1,933
0,0085 (2 h)	0,0073	1,164
0,0154 (2 h)	0,0146	1,055
0,0269 (2 h)	0,0292	0,921
0,0572 (16 h)	0,0438	1,306
0,0653 (16 h)	0,0877	0,745

Einbau von „frischem“ Hydroxylysin (Hyl)		
c(eingebautes Hyl) [mg/ml]	c(„frisches“ Hyl) [mg/ml]	Verhältnis
0,0004 (6 h)	0,0081	0,049
0,001 (6 h)	0,0324	0,031
0,0019 (6 h)	0,0487	0,039
0,0021 (6 h)	0,0973	0,022
0,0052 (16 h)	0,2271	0,023
0,0042 (16 h)	2,725	0,0015

Tabelle 3 zeigt, dass in den zellfreien Genexpressionsreaktionen, in denen zunächst die „frische“ Aminosäure der limitierende Faktor der deGFP-Produktion war (2 h und 6 h Werte), die Effektivität des Einbaus unabhängig von der Konzentration der „frischen“ Aminosäure war (repräsentiert durch vergleichbare Verhältnisse). Im Falle von Lysin wurden ca. 100 % des im Expressionssystem vorhandenen „frischen“ Lysins innerhalb von 2 h in deGFP eingebaut (repräsentiert durch ein Verhältnis von ca. 1). Im Gegensatz dazu wurden im Falle von Hydroxylysin lediglich ca. 3 % des vorhanden

5 Weitere Ergebnisse

„frischen“ Hydroxylysins innerhalb von 6 h in deGFP eingebaut (repräsentiert durch ein Verhältnis von ca. 0,03). Hieraus ergibt sich eine 100-fach reduzierte Einbaurrate von Hydroxylysin verglichen mit Lysin.

Dadurch, dass im Falle von Hydroxylysin 97 % des vorhanden „frischen“ Hydroxylysins dem Expressionssystem auch nach Start des Einbaus von „restlichem“ Lysin zur Verfügung standen, wird angenommen, dass nach diesem Zeitpunkt „restliches“ Lysin und „frisches“ Hydroxylysin simultan in deGFP eingebaut wurde. Diese Vermutung wird dadurch verstärkt, dass in Abbildung 38 B keine vollständige Sättigung vor dem Wendepunkt erkennbar ist. Folglich kann angenommen werden, dass daraus die in Finkler *et al.* (2021), Abbildung 5, beobachteten Mischspezies von deGFP mit verschiedenen Verhältnissen an eingebautem Lysin und Hydroxylysin resultierten.

Des Weiteren konnte in Finkler *et al.* (2021) gezeigt werden, dass bei einer Hydroxylysin-Konzentration von 16,8 mM (2,725 mg/ml) in Kombination mit einem Verhältnis zwischen Hydroxylysin und „restlichem“ Lysin von $c(\text{Hydroxylysin})/c(\text{„restliches“ Lysin}) \approx 665$ ein vollständiger Einbau von Hydroxylysin in Proteine möglich war. Da diese Bedingungen auch bei Generierung der in Tabelle 3 aufgelisteten Werte für eine Konzentration von 2,725 mg/ml an „frischem“ Hydroxylysin vorlagen, sind diese Werte von besonderem Interesse. Bei dieser Konzentration an „frischem“ Hydroxylysin wurden lediglich 0,15 % (1/667) des vorhandenen Hydroxylysins in deGFP eingebaut. Zusammengefasst werden 665 Hydroxylysine pro einem „restlichem“ Lysin benötigt, um den Einbau von letzterem zu supprimieren während von den vorhandenen 665 Hydroxylysinen lediglich ca. eines in deGFP eingebaut wird.

Die in diesem Abschnitt beschriebene Methode zur quantitativen Analyse des Einbaus von Hydroxylysin in deGFP liefert erste Hinweise für einen nicht linearen Zusammenhang zwischen dem Prozess der Fehlerreduktion beim Beladen der tRNA mit der „falschen“ Aminosäure und dem Verhältnis der Konzentrationen von „falscher“ und „richtiger“ Aminosäure. Diese Nichtlinearität ist ein Indiz für das Vorhandensein kooperativer Effekte ⁴⁴⁹. Da bei der klassischen Betrachtung des „Kinetic Proofreading“-Prozesses jedoch ein linearer Zusammenhang vermutet wurde ^{58,62}, können die in diesem Abschnitt präsentierten Ergebnisse zu einer Erweiterung der klassischen Betrachtung des „Kinetic Proofreading“-Prozesses führen. Dennoch müssen weitere Experimente dieser Art durchgeführt werden, um die aufgestellte Hypothese zu bestätigen.

5.2 Massenspektrometrische Methode zur Untersuchung des Aminosäure-Einbaus

Eine quantitative Analyse von Proteinen ist auch mittels Massenspektrometrie möglich^{447,448,450,451}. Diesbezüglich existieren verschiedene Methoden für eine relative oder absolute Quantifizierung^{448,450}. Welche Methode sich für ein Experiment am besten eignet, hängt von der jeweiligen Problemstellung ab⁴⁵⁰.

Eine mögliche massenspektrometrische Methode für eine quantitative Bestimmung von Proteinen mit eingebauten kanonischen und Proteinen mit eingebauten nicht-kanonischen Aminosäuren ist die so genannte „Absolute quantification“ (AQUA)-Methode^{452–455}.

Im Folgenden soll die AQUA-Methode für ein besseres Verständnis kurz zusammengefasst werden, wobei eine genauere Beschreibung in der Literatur eingesehen werden kann^{452–455}.

Bei der AQUA-Methode muss zunächst ein geeignetes Peptid ausgesucht werden, welches durch einen proteolytischen Verdau des zu untersuchenden Proteins erhalten wird. Anschließend muss das ausgesuchte Peptid als isotoopenmarkierte Variante synthetisiert werden. Da die Menge des synthetisierten Peptides nach dessen Produktion bekannt ist, kann dieses als interner Standard verwendet werden. Mit diesem internen Standard kann eine Eichgerade angefertigt werden.

Für die Quantifizierung des zu untersuchenden Proteins muss dieses zunächst proteolytisch verdaut werden. Diesem Ansatz wird das synthetische isotoopenmarkierte Peptid in einer bekannten Konzentration hinzugegeben. Im Anschluss wird eine massenspektrometrische Messung durchgeführt. Durch die Isotoopenmarkierung (*) haben das synthetische und das aus dem Verdau erhaltene Peptid unterschiedliche Massen, was zu unterschiedlichen Peaks im Massenspektrum führt. Durch den Vergleich der Flächen beider Peaks kann in Verbindung mit der Eichgerade die Konzentration des durch den Verdau erhaltenen Peptides und somit des zu untersuchenden Proteins erhalten werden.

Abbildung 39 liefert ein Schema der zuvor beschriebenen AQUA-Methode.

5 Weitere Ergebnisse

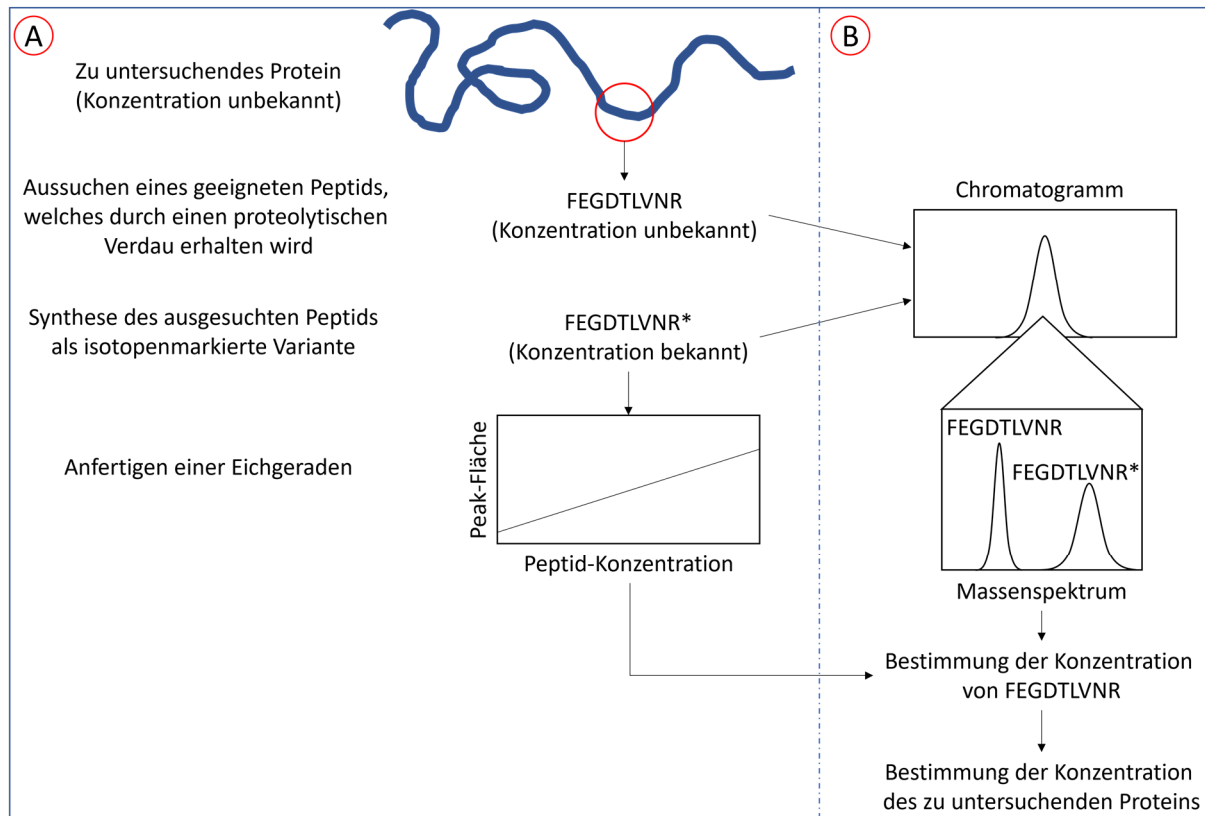


Abbildung 39 Schema zur AQUA-Methode. Die Abbildung zeigt die Präparation und Verwendung geeigneter Peptid-Proben (A) sowie die eigentliche analytische Methode (B).

Soll die AQUA-Methode zur quantitativen Bestimmung von Proteinen mit eingebauten kanonischen und Proteinen mit eingebauten nicht-kanonischen Aminosäuren verwendet werden, muss zunächst ein Peptid ausgewählt werden, welches die zu ersetzende kanonische Aminosäure enthält. Die AQUA-Methode benötigt in diesem Falle zwei synthetische Peptide: ein isotopenmarkiertes Peptid mit der kanonischen Aminosäure und ein isotopenmarkiertes Peptid mit der nicht-kanonischen Aminosäure. Mit diesen beiden internen Standards werden zwei Eichgeraden angefertigt. Wird ein Gemisch des zu untersuchenden Proteins mit eingebauten kanonischen und mit eingebauten nicht-kanonischen Aminosäuren proteolytisch verdaut, so können mit der AQUA-Methode die Konzentrationen beider Spezies bestimmt werden. Des Weiteren ist es mit dieser Methode möglich, alle Positionen im Protein, an denen die kanonische Aminosäure durch die nicht-kanonische ersetzt werden kann, einzeln zu untersuchen (sofern lediglich eine zu ersetzende Aminosäure zwischen zwei Schnittstellen vorhanden ist).

Im Falle des Einbaus von Canavanin (Can) in Proteine kann eine modifizierte Version der klassischen AQUA-Methode verwendet werden. Hierbei muss das zu untersuchende Protein zunächst auf das Aminosäure-Motiv „...KX...R...K...“ überprüft werden, wobei zwischen den beiden Lysinen kein weiteres Lysin und lediglich ein Arginin vorhanden sein darf (X entspricht einer beliebigen Aminosäure außer Lysin oder Arginin). Ist ein solches Motiv im zu untersuchenden Protein vorhanden, so müssen

die isopenmarkierten Peptide „X...R*“ und „X...Can*...K“ synthetisiert werden, welche als interne Standards fungieren. Wird ein Gemisch des zu untersuchenden Proteins mit eingebautem Arginin und eingebautem Canavanin mit Trypsin behandelt, so ergeben sich die Peptide „X...R“ und „X...Can...K“, da Trypsin an Lysin und Arginin, nicht aber an Canavanin schneidet^{251,456}. Mit diesem Ansatz kann die modifizierte AQUA-Methode durchgeführt werden (siehe Abbildung 40).

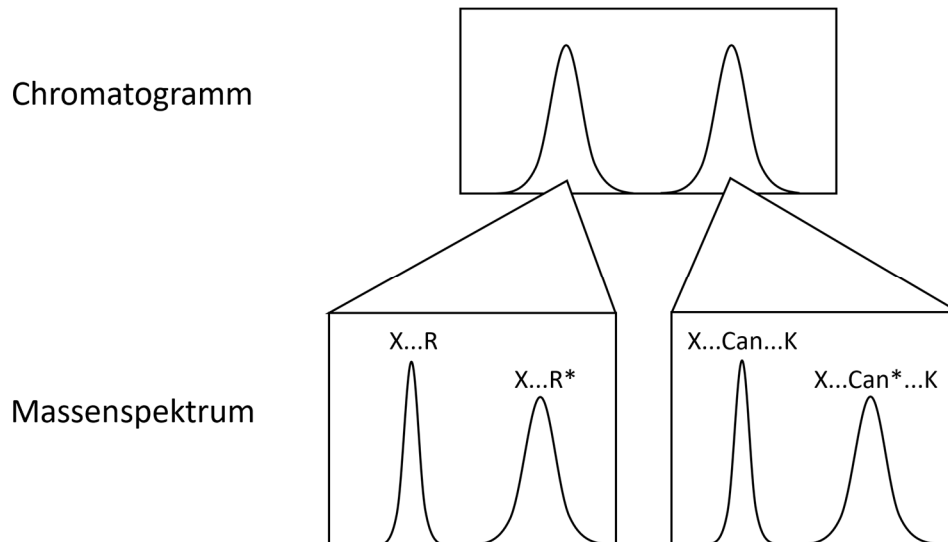


Abbildung 40 Modifizierte Version der AQUA-Methode zur gleichzeitigen Quantifizierung von eingebautem Canavanin und eingebautem Arginin in Proteinen.

In ersten Experimenten zur Etablierung der modifizierten AQUA-Methode^{**} konnten auch hier Hinweise für einen nicht linearen Zusammenhang zwischen dem Prozess der Fehlerreduktion beim Beladen der tRNA mit der „falschen“ Aminosäure und dem Verhältnis der Konzentrationen von „falscher“ und „richtiger“ Aminosäure gefunden werden (siehe Abbildung 41).

^{**} Die Experimente wurden von Pirkko Wenzel im Rahmen ihrer Master-Thesis am Lehrstuhl für Biologische Experimentalphysik an der Universität des Saarlandes durchgeführt.

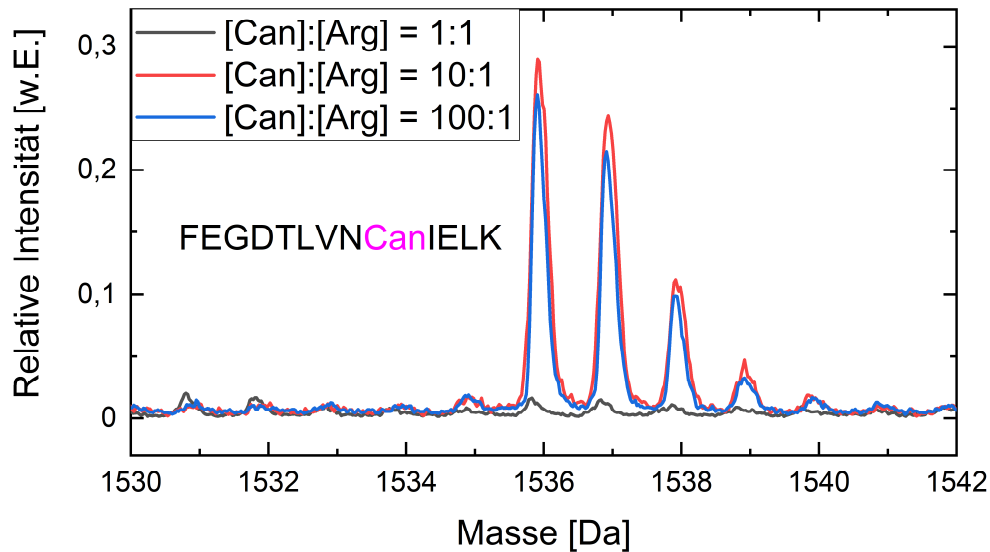


Abbildung 41 Massenspektrum eines tryptischen Verdau von deGFP, welches in einer zellfreien Genexpressionsreaktion bei verschiedenen Verhältnissen von Canavanin und Arginin synthetisiert wurde^{§§}. Dargestellt ist der Massenbereich für das Peptid „FEGDTLVNXIELK“, wobei X entweder Arg ($M_{\text{Peptid}} = 1533,8$ Da) oder Can ($M_{\text{Peptid}} = 1535,8$ Da) repräsentiert. Falls Arginin in deGFP eingebaut wurde, ist dieses Peptid nicht vorhanden, da durch den tryptischen Verdau am Arginin geschnitten wird. Wurde anstatt Arginin Canavanin eingebaut, ist dieses Peptid vorhanden, da am Canavanin nicht geschnitten wird^{251,456}. Aus diesem Grund sind im Massenspektrum nur die Peaks von „FEGDTLVNCanIELK“ vorhanden (Molekül- und Isotopenpeaks). Während bei einem Verhältnis von 1:1 (Can:Arg) kein Canavanin in deGFP eingebaut wurde, resultierten aus den Verhältnissen 10:1 und 100:1 vergleichbare Mengen an eingebautem Canavanin. Die Daten wurden auf das Maximum des Massenspektrums normiert. Für die Generierung des Massenspektrums wurde ein MALDI-TOF-Massenspektrometer verwendet.

Dennoch befindet sich die modifizierte AQUA-Methode in einem Anfangsstadium, weshalb weitere Experimente zur Validierung der aufgestellten Vermutung benötigt werden.

^{§§} Die präsentierten Daten wurden von Pirkko Wenzel im Rahmen ihrer Master-Thesis am Lehrstuhl für Biologische Experimentalphysik an der Universität des Saarlandes generiert.

6 Zusammenfassung und Ausblick

Die Publikationen, auf denen diese Dissertation basiert, lieferten neue Erkenntnisse über die Regulation der zellfreien Genexpression durch DNA-Methylierung sowie über die Konkurrenz zwischen kanonischen und nicht-kanonischen Aminosäuren beim seitenkettenspezifischen Einbau in Proteine.

Basierend auf den Ergebnissen aus Schenkelberger *et al.* (2017) und Worst *et al.* (2021) konnten DNA-Konformationsänderungen zumindest in den betrachteten Fällen als ein wichtiger Faktor für die durch DNA-Methylierung verursachte Genregulation identifiziert werden. Somit war es möglich, einen Organismus-übergreifenden Genregulationsmechanismus zu realisieren. Des Weiteren konnte ein Genschalter synthetisiert werden, dessen Gene durch unterschiedliche Methylierungszustände und daraus resultierende unterschiedliche DNA-Konformationen reguliert werden konnten. Solche Regulationsmechanismen könnten in naher Zukunft als werkvolle Werkzeuge in der synthetischen Biologie angesehen werden, da sie die Konstruktion synthetischer DNA-Konstrukte ermöglichen, deren Genexpression Organismus-unabhängig^{29,372–375} bzw. durch eine variable DNA-Konformation steuerbar ist^{95,400,401}. In diesem Zusammenhang könnten auch kraftabhängige Genschalter generiert werden.

Mit den Ergebnissen aus Finkler & Ott (2019), Finkler *et al.* (2020) und Finkler *et al.* (2021) konnte der vollständige seitenkettenspezifische Einbau der nicht-kanonischen Aminosäure Hydroxylysin in Proteine realisiert werden, welcher bisher durch die Konkurrenz zwischen Hydroxylysin und im Expressionssystem befindlichem „restlichem“ Lysin verhindert wurde⁵⁷. Solche rekombinant produzierten, Hydroxylysin enthaltenden Proteine könnten in Zukunft u. a. in der Medizin eine wichtige Rolle spielen^{421,423–425,437}. Eine genauere Analyse des Einbauverhaltens von „restlichem“ und „frischem“ Lysin war ein essenzieller Bestandteil für die Realisierung des vollständigen seitenkettenspezifischen Einbaus von Hydroxylysin. Des Weiteren konnten neue Erkenntnisse über den Einfluss des Verhältnisses zwischen Hydroxylysin und „restlichem“ Lysin auf den Einbau von Hydroxylysin gewonnen werden. Diese Ergebnisse können als der Startpunkt für neue Untersuchungen von Nichtlinearitäten durch kompetitives Binden bei der Fehlerreduktion durch den „Kinetic Proofreading“-Prozesses gesehen werden. Des Weiteren kann die vorgestellte Methode zum seitenkettenspezifischen Einbau von Hydroxylysin in Proteine unter kompetitiven Bedingungen prinzipiell auch für den Einbau anderer nicht-kanonischer Aminosäuren verwendet werden.

$$M(X^*K)_{18}$$

ORF	Protected sites	Protected areas	Motifs to avoid
9-122 [ATG...TAA]	1-6 BsrGI [TGTACA] 7-12 NcoI [CCATGG] 141-146 XhoI [CTCGAG]		BsrGI [TGTACA] NcoI [CCATGG] XhoI [CTCGAG]

1.	M A K S K G K L K P K T K V K C K E K I K
	TGTACACCATGGCGAAGAGCAAAGGCAAGCTGAAACCGAAGACCAAAGTGAAGTGCAAAGAGAAGATCA
	F K N K D K Y K R K Q K W K H K *
70.	AATTCAGAATAAAGATAAGTATAAACGCAAGCAGAAATGGAAGCATAAATAACATCACCATCACCCT
139.	AACTCGAG

$$M(XK)_{53}\text{-His-Tag}$$

ORF	Protected sites	Protected areas	Motifs to avoid
9-350 [ATG...TAA]	1-6 BsrGI [TGTACA] 7-12 NcoI [CCATGG] 351-356 XhoI [CTCGAG]		BsrGI [TGTACA] NcoI [CCATGG] XhoI [CTCGAG]

1.	M A K S K G K L K P K T K V K C K E K I K
	TGTACACCATGGCGAAGAGCAAAGGCAAGCTGAAACCGAAGACCAAAGTGAAGTGCAAAGAGAAGATCA
	F K N K D K Y K R K Q K W K M K A K S K G K L
70.	AATTCAGAATAAAGATAAGTATAAACGCAAGCAGAAATGGAAGATGAAAGCGAAGAGCAAAGGCAAGC
	K P K T K V K C K E K I K F K N K D K Y K R K
139.	TGAAACCGAAGACCAAAGTGAAGTGCAAAGAGAAGATCAAATTCAGAATAAAGATAAGTATAAACGCA
	Q K W K M K A K S K G K L K P K T K V K C K E
208.	AGCAGAAATGGAAGATGAAAGCGAAGAGCAAAGGCAAGCTGAAACCGAAGACCAAAGTGAAGTGCAAAG
	K I K F K N K D K Y K R K Q K W K H H H H H
277.	AGAAGATCAAATTCAGAATAAAGATAAGTATAAACGCAAGCAGAAATGGAAGCATCATCACCATCACC
	*
346.	ACTAACTCGAG

Abbildung A1 DNA-Sequenzen der synthetisierten „High-Lysin“-codierenden Gene und die daraus abgeleiteten Aminosäuresequenzen^{***}. Ein weiteres „High-Lysin“-codierendes Gen (codiert für LY(K)₃₀-His-Tag) wurde mittels PCR aus dem Gen von MA-Strep-Tag II-SA(K)₃₀GS generiert.

^{***} Die einzelnen Abbildungen stammen aus den vom Hersteller (GeneArt, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) bereitgestellten Analyse-Berichten.

Massenspektrometer-Einstellungen

Tabelle A1 Optimierte Einstellungen des verwendeten Massenspektrometers (SolariX FTICR-ESI, Bruker Corporation, Billerica, Massachusetts, USA) zum Nachweis von deGFP.

Kategorie	Parameter	Wert/Einstellung
General Settings	Size	1M
	Low m/z	688,10
	High m/z	1400,00
	Avg Scans	1
	Accum (s)	0,25
General Settings – Polarity	Polarity	Pos
	API High Voltage	✓
	Source Quench	✓
General Settings – Chromatography	Serial Mode	✓
API Source – API Source	Capillary	4500 V
	End Plate Offset	-500 V
API Source – Source Gas Tune	Nebulizer	2,5 bar
	Dry Gas	6,0 L/min
	Dry Temp	240 °C
Ion Transfer – Source Optics	Capillary Exit	200,0 V
	Deflector Plate	220,0 V
	Funnel 1	150,0 V
	Skimmer 1	15,0 V
	Funnel RF Amplitude	150,0 Vpp
Ion Transfer – Octopole	Frequency	5 MHz
	RF Amplitude	300,0 Vpp
Ion Transfer – Quadrupole	Q1 Mass	600,00 m/z
Ion Transfer – Collision Cell	Collision Voltage	-2.0 V
	DC Extract Bias	0,5 V
	RF Frequency	2 MHz
	Collision RF Amplitude	1800,0 Vpp
Ion Transfer – Transfer Optics	Time of Flight	1,200 ms
	Frequency	2 MHz
	RF Amplitude	350,0 Vpp
Ion Transfer – Gas Control	Flow	32,0 %
	Enable	✓
Analyzer – Infinity Cell	Transfer Exit Lens	-14,0 V

	Analyzer Entrance	-7,0 V
	Side Kick	0,0 V
	Side Kick Offset	-1,5 V
	Front Trap Plate	0,650 V
	Back Trap Plate	0,550 V
	Sweep Excitation Power	14,0 %
Analyzer – Multiple Cell Accumulations	ICR Cell Fills	1

Literaturverzeichnis

- (1) Weissbach, H., and Pestka, S. (1977) *Molecular Mechanisms of Protein Biosynthesis*. Academic Press. <https://doi.org/10.1016/B978-0-12-744250-1.X5001-X>
- (2) Ibba, M., and Söll, D. (1999) Quality Control Mechanisms During Translation. *Science* 286, 1893–1897. <https://doi.org/10.1126/science.286.5446.1893>
- (3) Lengyel, P., and Söll, D. (1969) Mechanism of Protein Biosynthesis. *Bacteriol. Rev.* 33, 264–301. <https://doi.org/10.1128/br.33.2.264-301.1969>
- (4) Mierke, C. T. (2020) Translation and Post-Translational Modifications in Protein Biosynthesis. In *Cellular Mechanics and Biophysics*. pp 595–665, Springer, Cham. https://doi.org/10.1007/978-3-030-58532-7_14
- (5) Loettgers, A. (2007) Model Organisms and Mathematical and Synthetic Models to Explore Gene Regulation Mechanisms. *Biol. Theory* 2, 134–142. <https://doi.org/10.1162/biot.2007.2.2.134>
- (6) Mirzaei, M., Pushpitha, K., Deng, L., Chitranshi, N., Gupta, V., Rajput, R., Mangani, A. B., Dheer, Y., Godinez, A., McKay, M. J., Kamath, K., Pascovici, D., Wu, J. X., Hosseini Salekdeh, G., Karl, T., Haynes, P. A., Graham, S. L., and Gupta, V. K. (2019) Upregulation of Proteolytic Pathways and Altered Protein Biosynthesis Underlie Retinal Pathology in a Mouse Model of Alzheimer’s Disease. *Mol. Neurobiol.* 56, 6017–6034. <https://doi.org/10.1007/s12035-019-1479-4>
- (7) Buckel, P. (1996) Recombinant proteins for therapy. *Trends Pharmacol. Sci.* 17, 450–456. [https://doi.org/10.1016/S0165-6147\(96\)01011-5](https://doi.org/10.1016/S0165-6147(96)01011-5)
- (8) Jang, Y., and Champion, J. A. (2016) Self-Assembled Materials Made from Functional Recombinant Proteins. *Acc. Chem. Res.* 49, 2188–2198. <https://doi.org/10.1021/acs.accounts.6b00337>
- (9) Chappell, J., Watters, K. E., Takahashi, M. K., and Lucks, J. B. (2015) A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Curr. Opin. Chem. Biol.* 28, 47–56. <https://doi.org/10.1016/j.cbpa.2015.05.018>
- (10) Fernandez, J. M., and Hoeffler, J. P. (1999) *Gene Expression Systems: Using Nature for the Art of Expression*. Academic Press. <https://doi.org/10.1016/B978-0-12-253840-7.X5000-7>
- (11) Stiege, W., and Erdmann, V. A. (1995) The potentials of the in vitro protein biosynthesis system. *J. Biotechnol.* 41, 81–90. [https://doi.org/10.1016/0168-1656\(95\)00005-B](https://doi.org/10.1016/0168-1656(95)00005-B)
- (12) Sitaraman, K., Esposito, D., Klarmann, G., Le Grice, S. F., Hartley, J. L., and Chatterjee, D. K. (2004) A novel cell-free protein synthesis system. *J. Biotechnol.* 110, 257–263.

- <https://doi.org/10.1016/j.jbiotec.2004.02.014>
- (13) Sullivan, C. J., Pendleton, E. D., Sasmor, H. H., Hicks, W. L., Farnum, J. B., Muto, M., Amendt, E. M., Schoborg, J. A., Martin, R. W., Clark, L. G., Anderson, M. J., Choudhury, A., Fior, R., Lo, Y.-H., Griffey, R. H., Chappell, S. A., Jewett, M. C., Mauro, V. P., and Dresios, J. (2016) A cell-free expression and purification process for rapid production of protein biologics. *Biotechnol. J.* *11*, 238–248. <https://doi.org/10.1002/biot.201500214>
- (14) Serrano-Heras, G., Salas, M., and Bravo, A. (2005) A new plasmid vector for regulated gene expression in *Bacillus subtilis*. *Plasmid* *54*, 278–282. <https://doi.org/10.1016/j.plasmid.2005.04.002>
- (15) Inouye, M. (1983) *Experimental Manipulation of Gene Expression*. Academic Press. <https://doi.org/10.1016/C2013-0-10902-6>
- (16) Casali, N., and Preston, A. (2003) *E. coli Plasmid Vectors: Methods and Applications*. Humana Press. <https://doi.org/10.1385/1592594093>
- (17) Biswas, I., Jha, J. K., and Fromm, N. (2008) Shuttle expression plasmids for genetic studies in *Streptococcus mutans*. *Microbiology* *154*, 2275–2282. <https://doi.org/10.1099/mic.0.2008/019265-0>
- (18) Weinberger, C., and Bradley, D. J. (1990) Gene Regulation by Receptors Binding Lipid-Soluble Substances. *Annu. Rev. Physiol.* *52*, 823–840. <https://doi.org/10.1146/annurev.ph.52.030190.004135>
- (19) Schachtele, C. F., and Rogers, P. (1965) Canavanine Death in *Escherichia coli*. *J. Mol. Biol.* *14*, 474–489. [https://doi.org/10.1016/S0022-2836\(65\)80197-8](https://doi.org/10.1016/S0022-2836(65)80197-8)
- (20) Dudley, Q. M., Karim, A. S., and Jewett, M. C. (2015) Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnol. J.* *10*, 69–82. <https://doi.org/10.1002/biot.201400330>
- (21) Rolf, J., Rosenthal, K., and Lütz, S. (2019) Application of Cell-Free Protein Synthesis for Faster Biocatalyst Development. *Catalysts* *9*, 1–18. <https://doi.org/10.3390/catal9020190>
- (22) Singh-Blom, A., Hughes, R. A., and Ellington, A. D. (2014) An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J. Biotechnol.* *178*, 12–22. <https://doi.org/10.1016/j.jbiotec.2014.02.009>
- (23) Lu, Y. (2017) Cell-free synthetic biology: Engineering in an open world. *Synth. Syst. Biotechnol.* *2*, 23–27. <https://doi.org/10.1016/j.synbio.2017.02.003>
- (24) Shin, J., and Noireaux, V. (2012) An *E. coli* Cell-Free Expression Toolbox: Application to

- Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* 1, 29–41. <https://doi.org/10.1021/sb200016s>
- (25) Hold, C., and Panke, S. (2009) Towards the engineering of *in vitro* systems. *J. R. Soc. Interface* 6, 507–521. <https://doi.org/10.1098/rsif.2009.0110.focus>
- (26) He, M. (2008) Cell-free protein synthesis: applications in proteomics and biotechnology. *N. Biotechnol.* 25, 126–132. <https://doi.org/10.1016/j.nbt.2008.08.004>
- (27) Brödel, A. K., Sonnabend, A., and Kubick, S. (2014) Cell-Free Protein Expression Based on Extracts From CHO Cells. *Biotechnol. Bioeng.* 111, 25–36. <https://doi.org/10.1002/bit.25013>
- (28) Shin, J., and Noireaux, V. (2010) Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4, 1–9. <https://doi.org/10.1186/1754-1611-4-8>
- (29) Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013) Protocols for Implementing an *Escherichia coli* Based TX-TL Cell-Free Expression System for Synthetic Biology. *J. Vis. Exp.* 79, 1–14. <https://doi.org/10.3791/50762>
- (30) McClure, W. R. (1985) Mechanism and Control of Transcription Initiation in Prokaryotes. *Annu. Rev. Biochem.* 54, 171–204. <https://doi.org/10.1146/annurev.bi.54.070185.001131>
- (31) Ryazanov, A. G., Rudkin, B. B., and Spirin, A. S. (1991) Regulation of protein synthesis at the elongation stage: New insights into the control of gene expression in eukaryotes. *FEBS Lett.* 285, 170–175. [https://doi.org/10.1016/0014-5793\(91\)80798-8](https://doi.org/10.1016/0014-5793(91)80798-8)
- (32) Baumberg, S. (1999) *Prokaryotic Gene Expression*. OUP Oxford.
- (33) Jaenisch, R., and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33, 245–254. <https://doi.org/10.1038/ng1089>
- (34) Park, M., Patel, N., Keung, A. J., and Khalil, A. S. (2019) Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* 176, 227–238. <https://doi.org/10.1016/j.cell.2018.11.002>
- (35) Lo, A., and Qi, L. (2017) Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Research* 6, 1–16. <https://doi.org/10.12688/f1000research.11113.1>
- (36) Casadesús, J., and Low, D. (2006) Epigenetic Gene Regulation in the Bacterial World. *Microbiol. Mol. Biol. Rev.* 70, 830–856. <https://doi.org/10.1128/MMBR.00016-06>
- (37) Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. <https://doi.org/10.1101/GAD.947102>
- (38) Hmadcha, A., Bedoya, F. J., Sobrino, F., and Pintado, E. (1999) Methylation-Dependent Gene

- Silencing Induced by Interleukin 1 β via Nitric Oxide Production. *J. Exp. Med.* 190, 1595–1603. <https://doi.org/10.1084/jem.190.11.1595>
- (39) Neidhart, M. (2015) *DNA Methylation and Complex Human Disease*. Academic Press. <https://doi.org/10.1016/C2013-0-13028-0>
- (40) Hernday, A. D., Braaten, B. A., and Low, D. A. (2003) The Mechanism by Which DNA Adenine Methylase and PapI Activate the Pap Epigenetic Switch. *Mol. Cell* 12, 947–957. [https://doi.org/10.1016/S1097-2765\(03\)00383-6](https://doi.org/10.1016/S1097-2765(03)00383-6)
- (41) Chang, T. Z., Kuo, J., and Silver, P. A. (2019) Beyond the Four Bases: A Home Run for Synthetic Epigenetic Control? *Mol. Cell* 74, 5–7. <https://doi.org/10.1016/j.molcel.2019.03.016>
- (42) Kahramanoglou, C., Prieto, A. I., Khedkar, S., Haase, B., Gupta, A., Benes, V., Fraser, G. M., Luscombe, N. M., and Seshasayee, A. S. N. (2012) Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat. Commun.* 3, 1–9. <https://doi.org/10.1038/ncomms1878>
- (43) Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R., and Greenberg, M. E. (2003) Derepression of BDNF Transcription Involves Calcium-Dependent Phosphorylation of MeCP2. *Science* 302, 885–889. <https://doi.org/10.1126/science.1086446>
- (44) Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., and Walkinshaw, M. D. (2008) MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol. Cell* 29, 525–531. <https://doi.org/10.1016/j.molcel.2007.12.028>
- (45) Klose, R. J., Sarraf, S. A., Schmiedeberg, L., McDermott, S. M., Stancheva, I., and Bird, A. P. (2005) DNA Binding Selectivity of MeCP2 Due to a Requirement for A/T Sequences Adjacent to Methyl-CpG. *Mol. Cell* 19, 667–678. <https://doi.org/10.1016/j.molcel.2005.07.021>
- (46) Li, W., and Pozzo-Miller, L. (2014) BDNF deregulation in Rett syndrome. *Neuropharmacology* 76, 737–746. <https://doi.org/10.1016/j.neuropharm.2013.03.024>
- (47) Guy, J., Cheval, H., Selfridge, J., and Bird, A. (2011) The Role of MeCP2 in the Brain. *Annu. Rev. Cell Dev. Biol.* 27, 631–652. <https://doi.org/10.1146/annurev-cellbio-092910-154121>
- (48) Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T. C., Qin, J., and Zoghbi, H. Y. (2008) MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. *Science* 320, 1224–1229. <https://doi.org/10.1126/science.1153252>
- (49) Schenkelberger, M. (2014) Cooperative Biomolecular Binding: High specificity of competitive single stranded DNA hybridization, influence of DNA methylation on the duplex stability, and epigenetic regulation of in vitro gene switches. Thesis, Universität des Saarlandes. <https://doi.org/10.22028/D291-22946>

- (50) Worst, E. G. (2018) Evolution wachsender, sich reproduzierender Polymere. Zellfreie Genexpression zum Einbau nicht-kanonischer Aminosäuren in Proteine sowie zur Analyse des epigenetischen *Escherichia-coli*-Pili-Phasenvariationsmechanismus. Thesis, Universität des Saarlandes. <https://doi.org/10.22028/D291-27165>
- (51) Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H., and Low, D. A. (1989) Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. *EMBO J.* 8, 613–620. <https://doi.org/10.1002/j.1460-2075.1989.tb03416.x>
- (52) Hernday, A., Krabbe, M., Braaten, B., and Low, D. (2002) Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl. Acad. Sci. USA.* 99, 16470–16476. <https://doi.org/10.1073/pnas.182427199>
- (53) Braaten, B. A., Nou, X., Kaltenbach, L. S., and Low, D. A. (1994) Methylation Patterns in *pap* Regulatory DNA Control Pyelonephritis-Associated Pili Phase Variation in *E. coli*. *Cell* 76, 577–588. [https://doi.org/10.1016/0092-8674\(94\)90120-1](https://doi.org/10.1016/0092-8674(94)90120-1)
- (54) Hopfield, J. J. (1974) Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proc. Natl. Acad. Sci. USA.* 71, 4135–4139. <https://doi.org/10.1073/pnas.71.10.4135>
- (55) Murugan, A., Huse, D. A., and Leibler, S. (2012) Speed, dissipation, and error in kinetic proofreading. *Proc. Natl. Acad. Sci. USA.* 109, 12034–12039. <https://doi.org/10.1073/pnas.1119911109>
- (56) Pak, D., Kim, Y., and Burton, Z. F. (2018) Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code. *Transcription* 9, 205–224. <https://doi.org/10.1080/21541264.2018.1467718>
- (57) Worst, E. G., Exner, M. P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., and Ott, A. (2016) Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an *Escherichia coli* Cell-free Transcription-translation System. *J. Vis. Exp.* 114, 1–14. <https://doi.org/10.3791/54273>
- (58) Hopfield, J. J., Yamane, T., Yue, V., and Coutts, S. M. (1976) Direct experimental evidence for kinetic proofreading in amino acylation of tRNA^{Ile}. *Proc. Natl. Acad. Sci. USA.* 73, 1164–1168. <https://doi.org/10.1073/pnas.73.4.1164>
- (59) Budisa, N. (2004) Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire. *Angew. Chemie - Int. Ed.* 43, 6426–6463. <https://doi.org/10.1002/anie.200300646>
- (60) DeLey Cox, V. E., Cole, M. F., and Gaucher, E. A. (2019) Incorporation of Modified Amino Acids by Engineered Elongation Factors with Expanded Substrate Capabilities. *ACS Synth. Biol.*

- 8, 287–296. <https://doi.org/10.1021/acssynbio.8b00305>
- (61) Ninio, J. (1975) Kinetic amplification of enzyme discrimination. *Biochimie* 57, 587–595. [https://doi.org/10.1016/S0300-9084\(75\)80139-8](https://doi.org/10.1016/S0300-9084(75)80139-8)
- (62) Hopfield, J. J. (1980) The energy relay: A proofreading scheme based on dynamic cooperativity and lacking all characteristic symptoms of kinetic proofreading in DNA replication and protein synthesis. *Proc. Natl. Acad. Sci. USA*. 77, 5248–5252. <https://doi.org/10.1073/pnas.77.9.5248>
- (63) Solovyev, A. Y., Tarnovskaya, S. I., Chernova, I. A., Shataeva, L. K., and Skorik, Y. A. (2015) The interaction of amino acids, peptides, and proteins with DNA. *Int. J. Biol. Macromol.* 78, 39–45. <https://doi.org/10.1016/j.ijbiomac.2015.03.054>
- (64) Cherstvy, A. G. (2011) Electrostatic interactions in biological DNA-related systems. *Phys. Chem. Chem. Phys.* 13, 9942–9968. <https://doi.org/10.1039/C0CP02796K>
- (65) Liu, L., Guo, Z., Huang, Z., Zhuang, J., and Yang, W. (2016) Size-selective separation of DNA fragments by using lysine-functionalized silica particles. *Sci. Rep.* 6, 1–8. <https://doi.org/10.1038/srep22029>
- (66) Davidson, J. N. (1972) *The Biochemistry of the Nucleic Acids*. Academic Press. <https://doi.org/10.1016/B978-0-12-205350-4.X5002-1>
- (67) Munk, K. (2017) *Taschenlehrbuch Biologie: Genetik*, 2. Auflage. Georg Thieme Verlag.
- (68) Cosgrove, M. (1998) Nucleotides. *Nutrition* 14, 748–751. [https://doi.org/10.1016/S0899-9007\(98\)00075-6](https://doi.org/10.1016/S0899-9007(98)00075-6)
- (69) Lopez, A., and Liu, J. (2017) Self-Assembly of Nucleobase, Nucleoside and Nucleotide Coordination Polymers: From Synthesis to Applications. *ChemNanoMat.* 3, 670–684. <https://doi.org/10.1002/cnma.201700154>
- (70) Koecke, H.-U., Emschermann, P., and Härle, E. (2000) *Biologie: Lehrbuch der allgemeinen Biologie für Mediziner und Naturwissenschaftler ; mit 22 Tabellen und 53 Exkursen*, 4. Auflage. Schattauer Verlag.
- (71) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) The Structure and Function of DNA. In *Molecular Biology of the Cell*, 4. Auflage. Garland Science.
- (72) Loeb, L. A., Dube, D. K., Beckman, R. A., Koplitz, M., and Gopinathan, K. P. (1981) On the Fidelity of DNA Replication: Nucleoside Monophosphate Generation during Polymerization. *J. Biol. Chem.* 256, 3978–3987. [https://doi.org/10.1016/s0021-9258\(19\)69555-4](https://doi.org/10.1016/s0021-9258(19)69555-4)
- (73) Da, L.-T., Wang, D., and Huang, X. (2012) Dynamics of Pyrophosphate Ion Release and Its Coupled Trigger Loop Motion from Closed to Open State in RNA Polymerase II. *J. Am. Chem.*

- Soc.* 134, 2399–2406. <https://doi.org/10.1021/ja210656k>
- (74) Tan, S. C., and Yiap, B. C. (2009) DNA, RNA, and Protein Extraction: The Past and The Present. *J. Biomed. Biotechnol.* 2009, 1–10. <https://doi.org/10.1155/2009/574398>
- (75) Watson, J. D., and Crick, F. H. C. (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738. <https://doi.org/10.1038/171737a0>
- (76) Cooper, V. R., Thonhauser, T., Puzder, A., Schröder, E., Lundqvist, B. I., and Langreth, D. C. (2008) Stacking Interactions and the Twist of DNA. *J. Am. Chem. Soc.* 130, 1304–1308. <https://doi.org/10.1021/ja0761941>
- (77) Calladine, C. R. (1982) Mechanics of Sequence-dependent Stacking of Bases in B-DNA. *J. Mol. Biol.* 161, 343–352. [https://doi.org/10.1016/0022-2836\(82\)90157-7](https://doi.org/10.1016/0022-2836(82)90157-7)
- (78) Dias, R., and Lindman, B. (2008) *DNA Interactions with Polymers and Surfactants*. John Wiley & Sons.
- (79) Takeda, Y., Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1983) DNA-Binding Proteins. *Science* 221, 1020–1026. <https://doi.org/10.1126/science.6308768>
- (80) Jen-Jacobson, L., Engler, L. E., and Jacobson, L. A. (2000) Structural and Thermodynamic Strategies for Site-Specific DNA Binding Proteins. *Structure* 8, 1015–1023. [https://doi.org/10.1016/S0969-2126\(00\)00501-3](https://doi.org/10.1016/S0969-2126(00)00501-3)
- (81) Pabo, C. O., and Sauer, R. T. (1992) Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem.* 61, 1053–1095. <https://doi.org/10.1146/annurev.bi.61.070192.005201>
- (82) Schreiter, E. R., and Drennan, C. L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* 5, 710–720. <https://doi.org/10.1038/nrmicro1717>
- (83) Nelson, H. C. (1995) Structure and function of DNA-binding proteins. *Curr. Opin. Genet. Dev.* 5, 180–189. [https://doi.org/10.1016/0959-437X\(95\)80006-9](https://doi.org/10.1016/0959-437X(95)80006-9)
- (84) Shen, C.-H. (2019) Chapter 4 - Gene Expression: Translation of the Genetic Code. In *Diagnostic Molecular Biology*. pp 87–116, Academic Press. <https://doi.org/10.1016/B978-0-12-802823-0.00004-3>
- (85) Kumari, S., Taginik, G., Varadaraj, S., and Varadaraj, K. (2019) Positively charged amino acid residues in the extracellular loops A and C of lens aquaporin 0 interact with the negative charges in the plasma membrane to facilitate cell-to-cell adhesion. *Exp. Eye Res.* 185. <https://doi.org/10.1016/j.exer.2019.05.022>
- (86) Thomas, J. O., and Kornberg, R. D. (1975) An octamer of histones in chromatin and free in

- solution. *Proc. Natl. Acad. Sci. USA.* 72, 2626–2630. <https://doi.org/10.1073/pnas.72.7.2626>
- (87) Palmer, B. R., and Marinus, M. G. (1994) The *dam* and *dcm* strains of *Escherichia coli* - a review. *Gene* 143, 1–12. [https://doi.org/10.1016/0378-1119\(94\)90597-5](https://doi.org/10.1016/0378-1119(94)90597-5)
- (88) He, W., Jia, C., and Zou, Q. (2019) 4mCPred: machine learning methods for DNA N⁴-methylcytosine sites prediction. *Bioinformatics* 35, 593–601. <https://doi.org/10.1093/bioinformatics/bty668>
- (89) Zhao, Z., Zhang, X., Chen, F., Fang, L., and Li, J. (2020) Accurate prediction of DNA N⁴-methylcytosine sites via boost-learning various types of sequence features. *BMC Genomics* 21, 1–11. <https://doi.org/10.1186/s12864-020-07033-8>
- (90) Heyn, H., and Esteller, M. (2015) An Adenine Code for DNA: A Second Life for N⁶-Methyladenine. *Cell* 161, 710–713. <https://doi.org/10.1016/j.cell.2015.04.021>
- (91) Zaccara, S., Ries, R. J., and Jaffrey, S. R. (2019) Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* 20, 608–624. <https://doi.org/10.1038/s41580-019-0168-5>
- (92) Marinus, M. G., and Løbner-Olesen, A. (2009) DNA Methylation. *EcoSal Plus* 3. <https://doi.org/10.1128/ecosalplus.4.4.5>
- (93) Singh, J., Hanson, J., Paliwal, K., and Zhou, Y. (2019) RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.* 10, 1–13. <https://doi.org/10.1038/s41467-019-13395-9>
- (94) Zhang, H., Zhang, C., Li, Z., Li, C., Wei, X., Zhang, B., and Liu, Y. (2019) A New Method of RNA Secondary Structure Prediction Based on Convolutional Neural Network and Dynamic Programming. *Front. Genet.* 10, 1–12. <https://doi.org/10.3389/fgene.2019.00467>
- (95) Brázda, V., Laister, R. C., Jagelská, E. B., and Arrowsmith, C. (2011) Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* 12, 1–16. <https://doi.org/10.1186/1471-2199-12-33>
- (96) Swadling, J. B., Ishii, K., Tahara, T., and Kitao, A. (2018) Origins of biological function in DNA and RNA hairpin loop motifs from replica exchange molecular dynamics simulation. *Phys. Chem. Chem. Phys.* 20, 2990–3001. <https://doi.org/10.1039/c7cp06355e>
- (97) Hald, M., Pedersen, J. B., Stein, P. C., Kirpekar, F., and Jacobsen, J. P. (1995) A comparison of the hairpin stability of the palindromic d(CGCG(A/T)₄CGCG) oligonucleotides. *Nucleic Acids Res.* 23, 4576–4582. <https://doi.org/10.1093/nar/23.22.4576>
- (98) Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* 13, 770–780.

- <https://doi.org/10.1038/nrg3296>
- (99) Gillespie, J. J. (2004) Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. *Mol. Phylogenet. Evol.* *33*, 936–943. <https://doi.org/10.1016/j.ympev.2004.08.004>
- (100) Tamura, K. (2015) Origins and Early Evolution of the tRNA Molecule. *Life* *5*, 1687–1699. <https://doi.org/10.3390/life5041687>
- (101) Dedkova, L. M., Fahmi, N. E., Golovine, S. Y., and Hecht, S. M. (2006) Construction of Modified Ribosomes for Incorporation of D-Amino Acids into Proteins. *Biochemistry* *45*, 15541–15551. <https://doi.org/10.1021/bi060986a>
- (102) Genchi, G. (2017) An overview on D-amino acids. *Amino Acids* *49*, 1521–1533. <https://doi.org/10.1007/s00726-017-2459-5>
- (103) Bodanszky, M. (1993) *Peptide Chemistry: A Practical Textbook*, 2. Auflage. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-78206-0>
- (104) Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* *35*, 225–249. <https://doi.org/10.1146/annurev.biophys.35.101105.121507>
- (105) Branden, C. I., and Tooze, J. (2012) *Introduction to Protein Structure*, 2. Auflage. Garland Science.
- (106) Johnson, J. A., Lu, Y. Y., Van Deventer, J. A., and Tirrell, D. A. (2010) Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* *14*, 774–780. <https://doi.org/10.1016/j.cbpa.2010.09.013>
- (107) Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991) Selenocysteine: the 21st amino acid. *Mol. Microbiol.* *5*, 515–520. <https://doi.org/10.1111/j.1365-2958.1991.tb00722.x>
- (108) Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science* *296*, 1459–1462. <https://doi.org/10.1126/science.1069588>
- (109) Agostini, F., Völler, J.-S., Kokschi, B., Acevedo-Rocha, C. G., Kubyskin, V., and Budisa, N. (2017) Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chemie - Int. Ed.* *56*, 9680–9703. <https://doi.org/10.1002/anie.201610129>
- (110) Walsh, C. (2006) *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. Roberts and Company Publishers.
- (111) Schröder, E., and Lübke, K. (1965) *The Peptides: Methods of Peptide Synthesis*. Academic Press.

- <https://doi.org/10.1016/C2013-0-12066-1>
- (112) Damodaran, S., Parkin, K. L., and Fennema, O. R. (2007) *Fennema's Food Chemistry*, 4. Auflage. CRC Press.
- (113) Epstein, W., and Beckwith, J. R. (1968) Regulation of Gene Expression. *Annu. Rev. Biochem.* 37, 411–436. <https://doi.org/10.1146/annurev.bi.37.070168.002211>
- (114) Baltes, W., and Matissek, R. (2011) *Lebensmittelchemie*, 7. Auflage. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-16539-9>
- (115) Yugandhar, K., and Gromiha, M. M. (2014) Protein-protein binding affinity prediction from amino acid sequence. *Bioinformatics* 30, 3583–3589. <https://doi.org/10.1093/bioinformatics/btu580>
- (116) Kimple, M. E., Brill, A. L., and Pasker, R. L. (2013) Overview of Affinity Tags for Protein Purification. *Curr. Protoc. Protein Sci.* 73, 1–23. <https://doi.org/10.1002/0471140864.ps0909s73>
- (117) Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., and Stüber, D. (1988) Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Nat. Biotechnol.* 6, 1321–1325. <https://doi.org/10.1038/nbt1188-1321>
- (118) Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., and Salemme, F. R. (1989) Structural Origins of High-Affinity Biotin Binding to Streptavidin. *Science* 243, 85–88. <https://doi.org/10.1126/science.2911722>
- (119) Chivers, C. E., Koner, A. L., Lowe, E. D., and Howarth, M. (2011) How the biotin–streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochem. J.* 435, 55–63. <https://doi.org/10.1042/bj20101593>
- (120) Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) Disulfide Bonds and Protein Folding. *Biochemistry* 39, 4207–4216. <https://doi.org/10.1021/bi992922o>
- (121) Bohr, H., Bohr, J., Brunak, S., Cotterill, R. M. J., Lautrup, B., Nørskov, L., Olsen, O. H., and Petersen, S. B. (1988) Protein secondary structure and homology by neural networks: The α -helices in rhodopsin. *FEBS Lett.* 241, 223–228. [https://doi.org/10.1016/0014-5793\(88\)81066-4](https://doi.org/10.1016/0014-5793(88)81066-4)
- (122) Browning, D. F., and Busby, S. J. W. (2016) Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* 14, 638–650. <https://doi.org/10.1038/nrmicro.2016.103>
- (123) Levine, M., and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature* 424, 147–151. <https://doi.org/10.1038/nature01763>

- (124) Karlson, P., and Doenecke, D. (2005) *Karlsons Biochemie und Pathobiochemie*. Georg Thieme Verlag.
- (125) Browning, D. F., and Busby, S. J. W. (2004) The Regulation of Bacterial Transcription Initiation. *Nat. Rev. Microbiol.* 2, 57–65. <https://doi.org/10.1038/nrmicro787>
- (126) Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E., and Ebright, R. H. (2012) Structural Basis of Transcription Initiation. *Science* 338, 1076–1080. <https://doi.org/10.1126/science.1227786>
- (127) Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., and Werner, T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21, 2933–2942. <https://doi.org/10.1093/bioinformatics/bti473>
- (128) Soutourina, J. (2018) Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* 19, 262–274. <https://doi.org/10.1038/nrm.2017.115>
- (129) Hermsen, R., Tans, S., and ten Wolde, P. R. (2006) Transcriptional Regulation by Competing Transcription Factor Modules. *PLoS Comput. Biol.* 2, 1552–1560. <https://doi.org/10.1371/journal.pcbi.0020164>
- (130) Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987) Regulation of Inducible and Tissue-Specific Gene Expression. *Science* 236, 1237–1245. <https://doi.org/10.1126/science.3296191>
- (131) Beljanski, M. (2013) *The Regulation of DNA Replication and Transcription*. Demos Medical Publishing.
- (132) Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., and Salzberg, S. L. (2000) Prediction of Transcription Terminators in Bacterial Genomes. *J. Mol. Biol.* 301, 27–33. <https://doi.org/10.1006/jmbi.2000.3836>
- (133) Peters, J. M., Vangeloff, A. D., and Landick, R. (2011) Bacterial Transcription Terminators: The RNA 3'-End Chronicles. *J. Mol. Biol.* 412, 793–813. <https://doi.org/10.1016/j.jmb.2011.03.036>
- (134) Green, M. R. (1986) Pre-mRNA Splicing. *Annu. Rev. Genet.* 20, 671–708. <https://doi.org/10.1146/annurev.ge.20.120186.003323>
- (135) Proudfoot, N. J., Furger, A., and Dye, M. J. (2002) Integrating mRNA Processing with Transcription. *Cell* 108, 501–512. [https://doi.org/10.1016/S0092-8674\(02\)00617-7](https://doi.org/10.1016/S0092-8674(02)00617-7)
- (136) Darnell, J. E. (1978) Implications of RNA-RNA Splicing in Evolution of Eukaryotic Cells. *Science* 202, 1257–1260. <https://doi.org/10.1126/science.364651>
- (137) Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187–208.

- [https://doi.org/10.1016/S0378-1119\(99\)00210-3](https://doi.org/10.1016/S0378-1119(99)00210-3)
- (138) Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361, 13–37. <https://doi.org/10.1016/j.gene.2005.06.037>
- (139) Nakagawa, S., Niimura, Y., Miura, K., and Gojobori, T. (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc. Natl. Acad. Sci. USA.* 107, 6382–6387. <https://doi.org/10.1073/pnas.1002036107>
- (140) Wachter, A. (2014) Gene regulation by structured mRNA elements. *Trends Genet.* 30, 172–181. <https://doi.org/10.1016/j.tig.2014.03.001>
- (141) Mayer, C., Köhrer, C., Kenny, E., Prusko, C., and RajBhandary, U. L. (2003) Anticodon Sequence Mutants of *Escherichia coli* Initiator tRNA: Effects of Overproduction of Aminoacyl-tRNA Synthetases, Methionyl-tRNA Formyltransferase, and Initiation Factor 2 on Activity in Initiation. *Biochemistry* 42, 4787–4799. <https://doi.org/10.1021/bi034011r>
- (142) Jin, H., Zhao, Q., Gonzalez de Valdivia, E. I., Ardell, D. H., Stenström, M., and Isaksson, L. A. (2006) Influences on gene expression *in vivo* by a Shine-Dalgarno sequence. *Mol. Microbiol.* 60, 480–492. <https://doi.org/10.1111/j.1365-2958.2006.05110.x>
- (143) Hansen, J. L., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2002) Structural insights into peptide bond formation. *Proc. Natl. Acad. Sci. USA.* 99, 11670–11675. <https://doi.org/10.1073/pnas.172404099>
- (144) Zhang, B., and Cech, T. R. (1997) Peptide bond formation by *in vitro* selected ribozymes. *Nature* 390, 96–100. <https://doi.org/10.1038/36375>
- (145) Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., and Noller, H. F. (2001) Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science* 292, 883–896. <https://doi.org/10.1126/science.1060089>
- (146) Zhou, J., Lancaster, L., Donohue, J. P., and Noller, H. F. (2014) How the ribosome hands the A-site tRNA to the P site during EF-G-catalyzed translocation. *Science* 345, 1188–1191. <https://doi.org/10.1126/science.1255030>
- (147) Agrawal, R. K., Penczek, P., Grassucci, R. A., Li, Y., Leith, A., Nierhaus, K. H., and Frank, J. (1996) Direct Visualization of A-, P-, and E-Site Transfer RNAs in the *Escherichia coli* Ribosome. *Science* 271, 1000–1002. <https://doi.org/10.1126/science.271.5251.1000>
- (148) Kapoor, S., Das, G., and Varshney, U. (2011) Crucial contribution of the multiple copies of the initiator tRNA genes in the fidelity of tRNA^{Met} selection on the ribosomal P-site in *Escherichia coli*. *Nucleic Acids Res.* 39, 202–212. <https://doi.org/10.1093/nar/gkq760>

- (149) Liljas, A., and Ehrenberg, M. (2013) *Structural Aspects of Protein Synthesis*, 2. Auflage. World Scientific. <https://doi.org/10.1142/7793>
- (150) Buttlar, J., Klein, C., Bruch, A., Fachinger, A., Funk, J., Hawer, H., and Kuijpers, A. (2020) *Tutorium Genetik - eine (ausführliche) Einführung*. Springer Spektrum, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-56067-9>
- (151) Bresch, C., and Hausmann, R. (1972) *Klassische und molekulare Genetik*. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-87168-9>
- (152) Stephan, W., and Hörger, A. C. (2019) *Molekulare Populationsgenetik: Theoretische Konzepte und empirische Evidenz*. Springer Spektrum, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-59428-5>
- (153) Wu, Y., Wang, Z., Qiao, X., Li, J., Shu, X., and Qi, H. (2020) Emerging Methods for Efficient and Extensive Incorporation of Non-canonical Amino Acids Using Cell-Free Systems. *Front. Bioeng. Biotechnol.* 8, 1–14. <https://doi.org/10.3389/fbioe.2020.00863>
- (154) Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008) A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature* 452, 108–111. <https://doi.org/10.1038/nature06683>
- (155) Livingston, D. M., and Leder, P. (1969) Deformylation and Protein Biosynthesis. *Biochemistry* 8, 435–443. <https://doi.org/10.1021/bi00829a059>
- (156) Johnson, D. B. F., Wang, C., Xu, J., Schultz, M. D., Schmitz, R. J., Ecker, J. R., and Wang, L. (2012) Release Factor One Is Nonessential in *Escherichia coli*. *ACS Chem. Biol.* 7, 1337–1344. <https://doi.org/10.1021/cb300229q>
- (157) Fang, G.-M., Li, Y.-M., Shen, F., Huang, Y.-C., Li, J.-B., Lin, Y., Cui, H.-K., and Liu, L. (2011) Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chemie - Int. Ed.* 50, 7645–7649. <https://doi.org/10.1002/anie.201100996>
- (158) Gavrilescu, M., and Chisti, Y. (2005) Biotechnology - a sustainable alternative for chemical industry. *Biotechnol. Adv.* 23, 471–499. <https://doi.org/10.1016/j.biotechadv.2005.03.004>
- (159) Kochendoerfer, G. G., and Kent, S. B. (1999) Chemical protein synthesis. *Curr. Opin. Chem. Biol.* 3, 665–671. [https://doi.org/10.1016/S1367-5931\(99\)00024-1](https://doi.org/10.1016/S1367-5931(99)00024-1)
- (160) Overton, T. W. (2014) Recombinant protein production in bacterial hosts. *Drug Discov. Today* 19, 590–601. <https://doi.org/10.1016/j.drudis.2013.11.008>
- (161) Heinrich, P. C., Müller, M., and Graeve, L. (2014) *Löffler/Petrides Biochemie und*

- Pathobiochemie*, 9. Auflage. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-17972-3>
- (162) Chen, R. (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol. Adv.* 30, 1102–1107. <https://doi.org/10.1016/j.biotechadv.2011.09.013>
- (163) Çelik, E., and Çalık, P. (2012) Production of recombinant proteins by yeast cells. *Biotechnol. Adv.* 30, 1108–1118. <https://doi.org/10.1016/j.biotechadv.2011.09.011>
- (164) Ehrlich, S. D. (1977) Replication and expression of plasmids from *Staphylococcus aureus* in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA.* 74, 1680–1682. <https://doi.org/10.1073/pnas.74.4.1680>
- (165) Bailey, J. E., Da Silva, N. A., Peretti, S. W., Seo, J. -H., and Srienc, F. (1986) Studies of Host-Plasmid Interactions in Recombinant Microorganisms. *Ann. N. Y. Acad. Sci.* 469, 194–211. <https://doi.org/10.1111/j.1749-6632.1986.tb26498.x>
- (166) Schmidt, O. (2017) *Genetik und Molekularbiologie*. Springer Spektrum, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-50274-7>
- (167) Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* 10, 411–421. [https://doi.org/10.1016/S0958-1669\(99\)00003-8](https://doi.org/10.1016/S0958-1669(99)00003-8)
- (168) Yoshida, N., and Sato, M. (2009) Plasmid uptake by bacteria: a comparison of methods and efficiencies. *Appl. Microbiol. Biotechnol.* 83, 791–798. <https://doi.org/10.1007/s00253-009-2042-4>
- (169) Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000) *Molecular Cell Biology*, 4. Auflage. W. H. Freeman and Company.
- (170) Xu, Z.-L., Mizuguchi, H., Ishii-Watabe, A., Uchida, E., Mayumi, T., and Hayakawa, T. (2001) Optimization of transcriptional regulatory elements for constructing plasmid vectors. *Gene* 272, 149–156. [https://doi.org/10.1016/S0378-1119\(01\)00550-9](https://doi.org/10.1016/S0378-1119(01)00550-9)
- (171) Liu, Q., Li, M. Z., Leibham, D., Cortez, D., and Elledge, S. J. (1998) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* 8, 1300–1309. [https://doi.org/10.1016/s0960-9822\(07\)00560-x](https://doi.org/10.1016/s0960-9822(07)00560-x)
- (172) Tolmachov, O. (2009) Designing Plasmid Vectors. In *Gene Therapy of Cancer. Methods in Molecular Biology™ (Methods and Protocols)*. Vol. 542, pp 117–129, Humana Press. https://doi.org/10.1007/978-1-59745-561-9_6
- (173) Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H., and Libchaber, A. (2005) Toward an artificial cell based on gene expression in vesicles. *Phys. Biol.* 2, 1–8.

- <https://doi.org/10.1088/1478-3975/2/3/p01>
- (174) Worst, E. G., Exner, M. P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., and Ott, A. (2015) Cell-free expression with the toxic amino acid canavanine. *Bioorg. Med. Chem. Lett.* *25*, 3658–3660. <https://doi.org/10.1016/j.bmcl.2015.06.045>
- (175) Izri, Z., Garenne, D., Noireaux, V., and Maeda, Y. T. (2019) Gene Expression in on-Chip Membrane-Bound Artificial Cells. *ACS Synth. Biol.* *8*, 1705–1712. <https://doi.org/10.1021/acssynbio.9b00247>
- (176) Takahashi, M. K., Hayes, C. A., Chappell, J., Sun, Z. Z., Murray, R. M., Noireaux, V., and Lucks, J. B. (2015) Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods* *86*, 60–72. <https://doi.org/10.1016/j.ymeth.2015.05.020>
- (177) Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016) The All *E. coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synth. Biol.* *5*, 344–355. <https://doi.org/10.1021/acssynbio.5b00296>
- (178) Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014) Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an *Escherichia coli* Based TX-TL Cell-Free System. *ACS Synth. Biol.* *3*, 387–397. <https://doi.org/10.1021/sb400131a>
- (179) Zhu, B., Gan, R., Cabezas, M. D., Kojima, T., Nicol, R., Jewett, M. C., and Nakano, H. (2020) Increasing cell-free gene expression yields from linear templates in *Escherichia coli* and *Vibrio natriegens* extracts by using DNA-binding proteins. *Biotechnol. Bioeng.* *117*, 3849–3857. <https://doi.org/10.1002/bit.27538>
- (180) Kuzminov, A., and Stahl, F. W. (1997) Stability of Linear DNA in *recA* Mutant *Escherichia coli* Cells Reflects Ongoing Chromosomal DNA Degradation. *J. Bacteriol.* *179*, 880–888. <https://doi.org/10.1128/jb.179.3.880-888.1997>
- (181) Kuzminov, A., Schabtach, E., and Stahl, F. W. (1994) χ sites in combination with RecA protein increase the survival of linear DNA in *Escherichia coli* by inactivating *exoV* activity of RecBCD nuclease. *EMBO J.* *13*, 2764–2776. <https://doi.org/10.1002/j.1460-2075.1994.tb06570.x>
- (182) Dabert, P., Ehrlich, S. D., and Gruss, A. (1992) χ sequence protects against RecBCD degradation of DNA *in vivo*. *Proc. Natl. Acad. Sci. USA.* *89*, 12073–12077. <https://doi.org/10.1073/pnas.89.24.12073>
- (183) Garenne, D., Beisel, C. L., and Noireaux, V. (2019) Characterization of the all-*E. coli* transcription-translation system myTXTL by mass spectrometry. *Rapid Commun. Mass Spectrom.* *33*, 1036–1048. <https://doi.org/10.1002/rcm.8438>
- (184) Seki, E., Matsuda, N., Yokoyama, S., and Kigawa, T. (2008) Cell-free protein synthesis system

- from *Escherichia coli* cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation. *Anal. Biochem.* 377, 156–161. <https://doi.org/10.1016/j.ab.2008.03.001>
- (185) Yim, S. S., Johns, N. I., Noireaux, V., and Wang, H. H. (2020) Protecting Linear DNA Templates in Cell-Free Expression Systems from Diverse Bacteria. *ACS Synth. Biol.* 9, 2851–2855. <https://doi.org/10.1021/acssynbio.0c00277>
- (186) Hong, S. H., Kwon, Y.-C., and Jewett, M. C. (2014) Non-standard amino acid incorporation into proteins using *Escherichia coli* cell-free protein synthesis. *Front. Chem.* 2, 1–7. <https://doi.org/10.3389/fchem.2014.00034>
- (187) Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755. <https://doi.org/10.1038/90802>
- (188) Shimizu, Y., Kanamori, T., and Ueda, T. (2005) Protein synthesis by pure translation systems. *Methods* 36, 299–304. <https://doi.org/10.1016/j.ymeth.2005.04.006>
- (189) Gao, W., Cho, E., Liu, Y., and Lu, Y. (2019) Advances and Challenges in Cell-Free Incorporation of Unnatural Amino Acids Into Proteins. *Front. Pharmacol.* 10, 1–8. <https://doi.org/10.3389/fphar.2019.00611>
- (190) Sibley, D. R., Benovic, J. L., Caron, M. G., and Lefkowitz, R. J. (1987) Regulation of Transmembrane Signaling by Receptor Phosphorylation. *Cell* 48, 913–922. [https://doi.org/10.1016/0092-8674\(87\)90700-8](https://doi.org/10.1016/0092-8674(87)90700-8)
- (191) Krauss, G. (2006) *Biochemistry of Signal Transduction and Regulation*, 3. Auflage. John Wiley & Sons. <https://doi.org/10.1002/3527601864>
- (192) Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014) A brief history of synthetic biology. *Nat. Rev. Microbiol.* 12, 381–390. <https://doi.org/10.1038/nrmicro3239>
- (193) Latchman, D. S. (2002) *Gene Regulation: A Eukaryotic Perspective*. Garland Science.
- (194) Dorman, C. J. (2020) *Structure and Function of the Bacterial Genome*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119309697>
- (195) Young, B. A., Gruber, T. M., and Gross, C. A. (2002) Views of Transcription Initiation. *Cell* 109, 417–420. [https://doi.org/10.1016/S0092-8674\(02\)00752-3](https://doi.org/10.1016/S0092-8674(02)00752-3)
- (196) Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z., and Landick, R. (2009) Regulator Trafficking on Bacterial Transcription Units In Vivo. *Mol. Cell* 33, 97–108. <https://doi.org/10.1016/j.molcel.2008.12.021>

- (197) Li, J., and Zhang, Y. (2014) Relationship between promoter sequence and its strength in gene expression. *Eur. Phys. J. E* 37, 1–6. <https://doi.org/10.1140/epje/i2014-14086-1>
- (198) Werel, W., Schickor, P., and Heumann, H. (1991) Flexibility of the DNA enhances promoter affinity of *Escherichia coli* RNA polymerase. *EMBO J.* 10, 2589–2594. <https://doi.org/10.1002/j.1460-2075.1991.tb07800.x>
- (199) Van Brempt, M., Clauwaert, J., Mey, F., Stock, M., Maertens, J., Waegeman, W., and De Mey, M. (2020) Predictive design of sigma factor-specific promoters. *Nat. Commun.* 11, 1–13. <https://doi.org/10.1038/s41467-020-19446-w>
- (200) Lopez-Serra, L., Ballestar, E., Fraga, M. F., Alaminos, M., Setien, F., and Esteller, M. (2006) A Profile of Methyl-CpG Binding Domain Protein Occupancy of Hypermethylated Promoter CpG Islands of Tumor Suppressor Genes in Human Cancer. *Cancer Res.* 66, 8342–8346. <https://doi.org/10.1158/0008-5472.can-06-1932>
- (201) Latchman, D. S. (1997) Transcription factors: An overview. *Int. J. Biochem. Cell Biol.* 29, 1305–1312. [https://doi.org/10.1016/S1357-2725\(97\)00085-X](https://doi.org/10.1016/S1357-2725(97)00085-X)
- (202) Pabo, C. O., and Sauer, R. T. (1984) Protein-DNA Recognition. *Annu. Rev. Biochem.* 53, 293–321. <https://doi.org/10.1146/annurev.bi.53.070184.001453>
- (203) Romanel, A., Jensen, L. J., Cardelli, L., and Csikász-Nagy, A. (2012) Transcriptional Regulation Is a Major Controller of Cell Cycle Transition Dynamics. *PLoS One* 7, 1–9. <https://doi.org/10.1371/journal.pone.0029716>
- (204) Deb, A., Davis, B. H., Guo, J., Ni, A., Huang, J., Zhang, Z., Mu, H., and Dzau, V. J. (2008) SFRP2 Regulates Cardiomyogenic Differentiation by Inhibiting a Positive Transcriptional Autofeedback Loop of Wnt3a. *Stem Cells* 26, 35–44. <https://doi.org/10.1634/stemcells.2007-0475>
- (205) Pan, G., Li, J., Zhou, Y., Zheng, H., and Pei, D. (2006) A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *FASEB J.* 20, 1730–1732. <https://doi.org/10.1096/fj.05-5543fje>
- (206) Siegele, D. A., and Hu, J. C. (1997) Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. USA.* 94, 8168–8172. <https://doi.org/10.1073/pnas.94.15.8168>
- (207) Polstein, L. R., and Gersbach, C. A. (2012) Light-Inducible Spatiotemporal Control of Gene Activation by Customizable Zinc Finger Transcription Factors. *J. Am. Chem. Soc.* 134, 16480–16483. <https://doi.org/10.1021/ja3065667>
- (208) Pinheiro, A. V., Baptista, P., and Lima, J. C. (2008) Light activation of transcription: photocaging

- of nucleotides for control over RNA polymerization. *Nucleic Acids Res.* 36, 1–7. <https://doi.org/10.1093/nar/gkn415>
- (209) Shimizu-Sato, S., Huq, E., Tepperman, J. M., and Quail, P. H. (2002) A light-switchable gene promoter system. *Nat. Biotechnol.* 20, 1041–1044. <https://doi.org/10.1038/nbt734>
- (210) Polstein, L. R., and Gersbach, C. A. (2015) A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* 11, 198–200. <https://doi.org/10.1038/nchembio.1753>
- (211) Satyanarayana, U. (2014) *Regulation of gene expression*, 4. Auflage. Elsevier Health Sciences.
- (212) Struhl, K. (1986) Constitutive and Inducible *Saccharomyces cerevisiae* Promoters: Evidence for Two Distinct Molecular Mechanisms. *Mol. Cell. Biol.* 6, 3847–3853. <https://doi.org/10.1128/mcb.6.11.3847-3853.1986>
- (213) Yansura, D. G., and Henner, D. J. (1984) Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA.* 81, 439–443. <https://doi.org/10.1073/pnas.81.2.439>
- (214) Dickson, R. C., Abelson, J., Barnes, W. M., and Reznikoff, W. S. (1975) Genetic Regulation: The Lac Control Region: The nucleotide sequence of the lac control region containing the promoter and operator is presented. *Science* 187, 27–35. <https://doi.org/10.1126/science.1088926>
- (215) Fitzgerald-Hayes, M., and Reichsman, F. (2009) *DNA and Biotechnology*, 3. Auflage. Academic Press.
- (216) Becker, N. A., Peters, J. P., Lionberger, T. A., and Maher III, L. J. (2013) Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Res.* 41, 156–166. <https://doi.org/10.1093/nar/gks1011>
- (217) Atwater, J. A., Wisdom, R., and Verma, I. M. (1990) Regulated mRNA Stability. *Annu. Rev. Genet.* 24, 519–541. <https://doi.org/10.1146/annurev.ge.24.120190.002511>
- (218) Belasco, J. G., and Higgins, C. F. (1988) Mechanisms of mRNA decay in bacteria: a perspective. *Gene* 72, 15–23. [https://doi.org/10.1016/0378-1119\(88\)90123-0](https://doi.org/10.1016/0378-1119(88)90123-0)
- (219) Alifano, P., Bruni, C. B., and Carlomagno, M. S. (1994) Control of mRNA processing and decay in prokaryotes. *Genetica* 94, 157–172. <https://doi.org/10.1007/bf01443430>
- (220) Cheng, Z.-F., and Deutscher, M. P. (2005) An Important Role for RNase R in mRNA Decay. *Mol. Cell* 17, 313–318. <https://doi.org/10.1016/j.molcel.2004.11.048>
- (221) Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S., and Cohen, S. N. (2002) Global

- analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA.* 99, 9697–9702. <https://doi.org/10.1073/pnas.112318199>
- (222) Mauger, D. M., Cabral, B. J., Presnyak, V., Su, S. V., Reid, D. W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M. J., and McFadyen, I. J. (2019) mRNA structure regulates protein expression through changes in functional half-life. *Proc. Natl. Acad. Sci. USA.* 116, 24075–24083. <https://doi.org/10.1073/pnas.1908052116>
- (223) Seo, S. W., Yang, J., and Jung, G. Y. (2009) Quantitative Correlation Between mRNA Secondary Structure Around the Region Downstream of the Initiation Codon and Translational Efficiency in *Escherichia coli*. *Biotechnol. Bioeng.* 104, 611–616. <https://doi.org/10.1002/bit.22431>
- (224) Kortmann, J., Sczodrok, S., Rinnenthal, J., Schwalbe, H., and Narberhaus, F. (2011) Translation on demand by a simple RNA-based thermosensor. *Nucleic Acids Res.* 39, 2855–2868. <https://doi.org/10.1093/nar/gkq1252>
- (225) Huang, H.-Y., Chien, C.-H., Jen, K.-H., and Huang, H.-D. (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res.* 34, 429–434. <https://doi.org/10.1093/nar/gkl333>
- (226) Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020) Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* 21, 151–170. <https://doi.org/10.1038/s41576-019-0186-3>
- (227) Hansen, L. H., Knudsen, S., and Sørensen, S. J. (1998) The Effect of the *lacY* Gene on the Induction of IPTG Inducible Promoters, Studied in *Escherichia coli* and *Pseudomonas fluorescens*. *Curr. Microbiol.* 36, 341–347. <https://doi.org/10.1007/s002849900320>
- (228) Kaludov, N. K., and Wolffe, A. P. (2000) MeCP2 driven transcriptional repression *in vitro*: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *Nucleic Acids Res.* 28, 1921–1928. <https://doi.org/10.1093/nar/28.9.1921>
- (229) Fineran, P. C., and Dy, R. L. (2014) Gene regulation by engineered CRISPR-Cas systems. *Curr. Opin. Microbiol.* 18, 83–89. <https://doi.org/10.1016/j.mib.2014.02.007>
- (230) Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. a. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429–7437. <https://doi.org/10.1093/nar/gkt520>
- (231) Bikard, D., and Marraffini, L. A. (2013) Control of gene expression by CRISPR-Cas systems. *Fl1000Prime Rep.* 5, 1–8. <https://doi.org/10.12703/p5-47>
- (232) Sander, J. D., and Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting

- genomes. *Nat. Biotechnol.* 32, 347–355. <https://doi.org/10.1038/nbt.2842>
- (233) Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., and Church, G. M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10, 1116–1121. <https://doi.org/10.1038/nmeth.2681>
- (234) Marshall, R., Beisel, C. L., and Noireaux, V. (2020) Rapid Testing of CRISPR Nucleases and Guide RNAs in an *E. coli* Cell-Free Transcription-Translation System. *STAR Protoc.* 1, 1–11. <https://doi.org/10.1016/j.xpro.2019.100003>
- (235) Dastjerdi, A. H., Newman, A., and Burgio, G. (2019) The Expanding Class 2 CRISPR Toolbox: Diversity, Applicability, and Targeting Drawbacks. *BioDrugs* 33, 503–513. <https://doi.org/10.1007/s40259-019-00369-y>
- (236) Makarova, K. S., and Koonin, E. V. (2015) Annotation and Classification of CRISPR-Cas Systems. In *CRISPR. Methods in Molecular Biology*. Vol. 1311, pp 47–75, Humana Press. https://doi.org/10.1007/978-1-4939-2687-9_4
- (237) McSweeney, M. A., and Styczynski, M. P. (2021) Effective Use of Linear DNA in Cell-Free Expression Systems. *Front. Bioeng. Biotechnol.* 9, 1–12. <https://doi.org/10.3389/fbioe.2021.715328>
- (238) Shin, J., and Noireaux, V. (2010) Study of messenger RNA inactivation and protein degradation in an *Escherichia coli* cell-free expression system. *J. Biol. Eng.* 4, 1–9. <https://doi.org/10.1186/1754-1611-4-9>
- (239) Sarkar, N. (1997) Polyadenylation of mRNA in Prokaryotes. *Annu. Rev. Biochem.* 66, 173–197. <https://doi.org/10.1146/annurev.biochem.66.1.173>
- (240) Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y., and Alakhov, Y. B. (1988) A Continuous Cell-Free Translation System Capable of Producing Polypeptides in High Yield. *Science* 242, 1162–1164. <https://doi.org/10.1126/science.3055301>
- (241) Shimizu, Y., Kuruma, Y., Ying, B.-W., Umekage, S., and Ueda, T. (2006) Cell-free translation systems for protein engineering. *FEBS J.* 273, 4133–4140. <https://doi.org/10.1111/j.1742-4658.2006.05431.x>
- (242) Almhjell, P. J., Boville, C. E., and Arnold, F. H. (2018) Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* 47, 8980–8997. <https://doi.org/10.1039/c8cs00665b>
- (243) Gerrits, M., Budisa, N., and Merk, H. (2019) Site-Specific Chemoselective Pyrrolysine Analogues Incorporation Using the Cell-Free Protein Synthesis System. *ACS Synth. Biol.* 8, 381–390. <https://doi.org/10.1021/acssynbio.8b00421>

- (244) Cusack, S. (1997) Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* 7, 881–889. [https://doi.org/10.1016/S0959-440X\(97\)80161-3](https://doi.org/10.1016/S0959-440X(97)80161-3)
- (245) Yang, B., Ayyadurai, N., Yun, H., Choi, Y. S., Hwang, B. H., Huang, J., Lu, Q., Zeng, H., and Cha, H. J. (2014) In Vivo Residue-Specific Dopa-Incorporated Engineered Mussel Bioglue with Enhanced Adhesion and Water Resistance. *Angew. Chemie - Int. Ed.* 53, 13360–13364. <https://doi.org/10.1002/anie.201406099>
- (246) Yoo, T. H., and Tirrell, D. A. (2007) High-Throughput Screening for Methionyl-tRNA Synthetases That Enable Residue-Specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells. *Angew. Chemie - Int. Ed.* 46, 5340–5343. <https://doi.org/10.1002/anie.200700779>
- (247) Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I., and Yokoyama, S. (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* 30, 4692–4699. <https://doi.org/10.1093/nar/gkf589>
- (248) Goerke, A. R., and Swartz, J. R. (2009) High-Level Cell-Free Synthesis Yields of Proteins Containing Site-Specific Non-Natural Amino Acids. *Biotechnol. Bioeng.* 102, 400–416. <https://doi.org/10.1002/bit.22070>
- (249) Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P.-J., Russell, D. H., and Liu, W. R. (2010) A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in *Escherichia coli*. *Angew. Chemie - Int. Ed.* 49, 3211–3214. <https://doi.org/10.1002/anie.201000465>
- (250) Acevedo-Rocha, C. G., and Budisa, N. (2016) Xenomicrobiology: a roadmap for genetic code engineering. *Microb. Biotechnol.* 9, 666–676. <https://doi.org/10.1111/1751-7915.12398>
- (251) Bence, A. K., and Crooks, P. A. (2003) The Mechanism of l-Canavanine Cytotoxicity: Arginyl tRNA Synthetase as a Novel Target for Anticancer Drug Discovery. *J. Enzyme Inhib. Med. Chem.* 18, 383–394. <https://doi.org/10.1080/1475636031000152277>
- (252) Hopfield, J. J. (1978) Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proc. Natl. Acad. Sci. USA.* 75, 4334–4338. <https://doi.org/10.1073/pnas.75.9.4334>
- (253) Nam, H., Lewis, N. E., Lerman, J. A., Lee, D.-H., Chang, R. L., Kim, D., and Palsson, B. O. (2012) Network Context and Selection in the Evolution to Enzyme Specificity. *Science* 337, 1101–1104. <https://doi.org/10.1126/science.1216861>
- (254) Hedstrom, L. (2010) Enzyme Specificity and Selectivity. In *eLS*. John Wiley & Sons.

- <https://doi.org/10.1002/9780470015902.a0000716.pub2>
- (255) Peracchi, A. (2018) The Limits of Enzyme Specificity and the Evolution of Metabolism. *Trends Biochem. Sci.* 43, 984–996. <https://doi.org/10.1016/j.tibs.2018.09.015>
- (256) Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S., and Puglisi, J. D. (2004) tRNA selection and kinetic proofreading in translation. *Nat. Struct. Mol. Biol.* 11, 1008–1014. <https://doi.org/10.1038/nsmb831>
- (257) Kessler, K. J., Kaufmann, W. K., Reardon, J. T., Elston, T. C., and Sancar, A. (2007) A mathematical model for human nucleotide excision repair: Damage recognition by random order assembly and kinetic proofreading. *J. Theor. Biol.* 249, 361–375. <https://doi.org/10.1016/j.jtbi.2007.07.025>
- (258) Bartlett, D. W., and Gilbert, A. M. (2021) A kinetic proofreading model for bispecific protein degraders. *J. Pharmacokinet. Pharmacodyn.* 48, 149–163. <https://doi.org/10.1007/s10928-020-09722-z>
- (259) McKeithan, T. W. (1995) Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl. Acad. Sci. USA.* 92, 5042–5046. <https://doi.org/10.1073/pnas.92.11.5042>
- (260) Gowans, G. J., Hawley, S. A., Ross, F. A., and Hardie, D. G. (2013) AMP Is a True Physiological Regulator of AMP-Activated Protein Kinase by Both Allosteric Activation and Enhancing Net Phosphorylation. *Cell Metab.* 18, 556–566. <https://doi.org/10.1016/j.cmet.2013.08.019>
- (261) Lasorne, B., Worth, G. A., and Robb, M. A. (2011) Excited-state dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 460–475. <https://doi.org/10.1002/wcms.26>
- (262) Sobel'man, I. I., Vainshtein, L. A., and Yukov, E. A. (1995) *Excitation of Atoms and Broadening of Spectral Lines*, 2. Auflage. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-57825-0>
- (263) Fano, U. (1974) Excitation of atoms to states of high orbital momentum. *J. Phys. B: Atom. Mol. Phys.* 7, 401–404. <https://doi.org/10.1088/0022-3700/7/14/003>
- (264) Moiseiwitsch, B. L., and Smith, S. J. (1968) Electron Impact Excitation of Atoms. *Rev. Mod. Phys.* 40, 238–353. <https://doi.org/10.1103/RevModPhys.40.238>
- (265) Otter, G., and Honecker, R. (1993) *Atome - Moleküle - Kerne: Band I Atomphysik*, 2. Auflage. Vieweg+Teubner Verlag, Wiesbaden. <https://doi.org/10.1007/978-3-663-11966-1>
- (266) Lakowicz, J. R., Ray, K., Chowdhury, M., Szymanski, H., Fu, Y., Zhang, J., and Nowaczyk, K. (2008) Plasmon-controlled fluorescence: a new paradigm in fluorescence spectroscopy. *Analyst* 133, 1308–1346. <https://doi.org/10.1039/b802918k>

- (267) McGervey, J. D. (1995) *Quantum Mechanics: Concepts and Applications*. Academic Press. <https://doi.org/10.1016/C2009-1-28546-7>
- (268) Sathyanarayana, D. N. (2001) *Electronic Absorption Spectroscopy and Related Techniques*. Universities Press.
- (269) Latscha, H. P., Kazmaier, U., and Klein, H. A. (2002) *Organische Chemie: Chemie - Basiswissen II*, 5. Auflage. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-09138-8>
- (270) Jørgensen, C. K. (1962) *Absorption Spectra and Chemical Bonding in Complexes*. Pergamon. <https://doi.org/10.1016/C2013-0-05287-5>
- (271) Walsh, A. (1955) The application of atomic absorption spectra to chemical analysis. *Spectrochim. Acta* 7, 108–117. [https://doi.org/10.1016/0371-1951\(55\)80013-6](https://doi.org/10.1016/0371-1951(55)80013-6)
- (272) Kasha, M. (1987) Fifty Years of the Jabłoński Diagram. *Acta Phys. Pol. A* 71, 661–670.
- (273) Zimmermann, J., Zeug, A., and Röder, B. (2003) A generalization of the Jablonski diagram to account for polarization and anisotropy effects in time-resolved experiments. *Phys. Chem. Chem. Phys.* 5, 2964–2969. <https://doi.org/10.1039/b303138a>
- (274) Jabłoński, A. (1935) Über den Mechanismus der Photolumineszenz von Farbstoffphosphoren. *Zeitschrift für Phys.* 94, 38–46. <https://doi.org/10.1007/bf01330795>
- (275) Jabłoński, A. (1933) Efficiency of Anti-Stokes Fluorescence in Dyes. *Nature* 131, 839–840. <https://doi.org/10.1038/131839b0>
- (276) Feng, G., Zhang, G.-Q., and Ding, D. (2020) Design of superior phototheranostic agents guided by Jablonski diagrams. *Chem. Soc. Rev.* 49, 8179–8234. <https://doi.org/10.1039/d0cs00671h>
- (277) Zhang, T., Ma, X., Wu, H., Zhu, L., Zhao, Y., and Tian, H. (2020) Molecular Engineering for Metal-Free Amorphous Materials with Room-Temperature Phosphorescence. *Angew. Chemie - Int. Ed.* 59, 11206–11216. <https://doi.org/10.1002/anie.201915433>
- (278) Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3. Auflage. Springer, Boston, MA. <https://doi.org/10.1007/978-0-387-46312-4>
- (279) Scheibe, G., Daltrozzo, E., and Wörz, O. (1969) Das Franck-Condon-Prinzip und die Lichtabsorption von Merocyaninen. *Zeitschrift für Phys. Chemie* 64, 97–114. https://doi.org/10.1524/zpch.1969.64.1_4.097
- (280) Condon, E. (1926) A Theory of Intensity Distribution in Band Systems. *Phys. Rev.* 28, 1182–1201. <https://doi.org/10.1103/PhysRev.28.1182>
- (281) Franck, J., and Dymond, E. G. (1926) Elementary Processes of Photochemical Reactions. *Trans. Faraday Soc.* 21, 536–542. <https://doi.org/10.1039/tf9262100536>

- (282) Callis, P. R. (2014) Binding phenomena and fluorescence quenching. I: Descriptive quantum principles of fluorescence quenching using a supermolecule approach. *J. Mol. Struct.* 1077, 14–21. <https://doi.org/10.1016/j.molstruc.2014.04.050>
- (283) Harris, D. C. (2014) *Lehrbuch der Quantitativen Analyse*, 8. Auflage. Springer Spektrum, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-37788-4>
- (284) Niehues, I., Marauhn, P., Deilmann, T., Wigger, D., Schmidt, R., Arora, A., de Vasconcellos, S. M., Rohlfig, M., and Bratschitsch, R. (2020) Strain tuning of the Stokes shift in atomically thin semiconductors. *Nanoscale* 12, 20786–20796. <https://doi.org/10.1039/d0nr04557h>
- (285) Wöhrle, D., Tausch, M. W., and Stohrer, W.-D. (2012) *Photochemie: Konzepte, Methoden, Experimente*. John Wiley & Sons.
- (286) Gustavsson, T., Baldacchino, G., Mialocq, J.-C., and Pommeret, S. (1995) A femtosecond fluorescence up-conversion study of the dynamic Stokes shift of the DCM dye molecule in polar and non-polar solvents. *Chem. Phys. Lett.* 236, 587–594. [https://doi.org/10.1016/0009-2614\(95\)00276-a](https://doi.org/10.1016/0009-2614(95)00276-a)
- (287) Ren, T.-B., Xu, W., Zhang, W., Zhang, X.-X., Wang, Z.-Y., Xiang, Z., Yuan, L., and Zhang, X.-B. (2018) A General Method To Increase Stokes Shift by Introducing Alternating Vibronic Structures. *J. Am. Chem. Soc.* 140, 7716–7722. <https://doi.org/10.1021/jacs.8b04404>
- (288) Nagasawa, Y., Oishi, A., Itoh, T., Yasuda, M., Muramatsu, M., Ishibashi, Y., Ito, S., and Miyasaka, H. (2009) Dynamic Stokes Shift of 9,9'-Bianthryl in Ionic Liquids: A Temperature Dependence Study. *J. Phys. Chem. C* 113, 11868–11876. <https://doi.org/10.1021/jp9020454>
- (289) Hötzer, B., Ivanov, R., Altmeier, S., Kappl, R., and Jung, G. (2011) Determination of Copper(II) Ion Concentration by Lifetime Measurements of Green Fluorescent Protein. *J. Fluoresc.* 21, 2143–2153. <https://doi.org/10.1007/s10895-011-0916-1>
- (290) Fraiji, L. K., Hayes, D. M., and Werner, T. C. (1992) Static and Dynamic Fluorescence Quenching Experiments for the Physical Chemistry Laboratory. *J. Chem. Educ.* 69, 424–428. <https://doi.org/10.1021/ed069p424>
- (291) Barrio, J. R., Tolman, G. L., Leonard, N. J., Spencer, R. D., and Weber, G. (1973) Flavin 1, N⁶-Ethenoadenine Dinucleotide: Dynamic and Static Quenching of Fluorescence. *Proc. Natl. Acad. Sci. USA.* 70, 941–943. <https://doi.org/10.1073/pnas.70.3.941>
- (292) Duncan, R. R. (2006) Fluorescence lifetime imaging microscopy (FLIM) to quantify protein-protein interactions inside cells. *Biochem. Soc. Trans.* 34, 679–682. <https://doi.org/10.1042/bst0340679>
- (293) Fan, Z., Zhou, H., Li, P. Y., Speer, J. E., and Cheng, H. (2014) Structural elucidation of cell

- membrane-derived nanoparticles using molecular probes. *J. Mater. Chem. B* 2, 8231–8238. <https://doi.org/10.1039/c4tb00980k>
- (294) Tsien, R. Y. (1998) The Green Fluorescent Protein. *Annu. Rev. Biochem.* 67, 509–544. <https://doi.org/10.1146/annurev.biochem.67.1.509>
- (295) Tsien, R. Y. (2010) Fluorescence Readouts of Biochemistry in Live Cells and Organisms. In *Molecular Imaging: Principles and Practice*. pp 808–828, People’s Medical Publishing House.
- (296) Hussain, S. A. (2009) An Introduction to Fluorescence Resonance Energy Transfer (FRET). *arXiv preprint arXiv:0908.1815*, 1–4.
- (297) Saini, S., Singh, H., and Bagchi, B. (2006) Fluorescence resonance energy transfer (FRET) in chemistry and biology: Non-Förster distance dependence of the FRET rate. *J. Chem. Sci.* 118, 23–35. <https://doi.org/10.1007/BF02708762>
- (298) Zimmermann, T., Rietdorf, J., Girod, A., Georget, V., and Pepperkok, R. (2002) Spectral imaging and linear un-mixing enables improved FRET efficiency with a novel GFP2-YFP FRET pair. *FEBS Lett.* 531, 245–249. [https://doi.org/10.1016/S0014-5793\(02\)03508-1](https://doi.org/10.1016/S0014-5793(02)03508-1)
- (299) Wallace, B., and Atzberger, P. J. (2017) Förster resonance energy transfer: Role of diffusion of fluorophore orientation and separation in observed shifts of FRET efficiency. *PLoS One* 12, 1–22. <https://doi.org/10.1371/journal.pone.0177122>
- (300) Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* 20, 448–455. [https://doi.org/10.1016/S0968-0004\(00\)89099-4](https://doi.org/10.1016/S0968-0004(00)89099-4)
- (301) Yang, T.-T., Cheng, L., and Kain, S. R. (1996) Optimized Codon Usage and Chromophore Mutations Provide Enhanced Sensitivity with the Green Fluorescent Protein. *Nucleic Acids Res.* 24, 4592–4593. <https://doi.org/10.1093/nar/24.22.4592>
- (302) Chudakov, D. M., Matz, M. V., Lukyanov, S., and Lukyanov, K. A. (2010) Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiol. Rev.* 90, 1103–1163. <https://doi.org/10.1152/physrev.00038.2009>
- (303) Royant, A., and Noirclerc-Savoye, M. (2011) Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *J. Struct. Biol.* 174, 385–390. <https://doi.org/10.1016/j.jsb.2011.02.004>
- (304) Heim, R., Cubitt, A., and Tsien, R. (1995) Improved green fluorescence. *Nature* 373, 663–664. <https://doi.org/10.1038/373663b0>
- (305) Cheng, L., Fu, J., Tsukamoto, A., and Hawley, R. G. (1996) Use of green fluorescent protein

- variants to monitor gene transfer and expression in mammalian cells. *Nat. Biotechnol.* 14, 606–609. <https://doi.org/10.1038/nbt0596-606>
- (306) Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R., and Piston, D. W. (1997) Use of the Green Fluorescent Protein and Its Mutants in Quantitative Fluorescence Microscopy. *Biophys. J.* 73, 2782–2790. [https://doi.org/10.1016/S0006-3495\(97\)78307-3](https://doi.org/10.1016/S0006-3495(97)78307-3)
- (307) Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173, 33–38. [https://doi.org/10.1016/0378-1119\(95\)00685-0](https://doi.org/10.1016/0378-1119(95)00685-0)
- (308) Li, B., Shahid, R., Peshkepija, P., and Zimmer, M. (2012) Water diffusion in and out of the β -barrel of GFP and the fast maturing fluorescent protein, TurboGFP. *Chem. Phys.* 392, 143–148. <https://doi.org/10.1016/j.chemphys.2011.11.001>
- (309) Royant, A., and Noirclerc-Savoie, M. (2011) X-ray structure of Enhanced Green Fluorescent Protein (EGFP). *PDB Entry - 2Y0G*. <https://doi.org/10.2210/pdb2y0g/pdb>
- (310) Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., and Rose, A. S. (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* 49, 431–437. <https://doi.org/10.1093/nar/gkab314>
- (311) Ayrton, J., Dear, G. J., Leavens, W. J., Mallett, D. N., and Plumb, R. S. (1998) Use of generic fast gradient liquid chromatography-tandem mass spectroscopy in quantitative bioanalysis. *J. Chromatogr. B: Biomed. Sci. Appl.* 709, 243–254. [https://doi.org/10.1016/S0378-4347\(98\)00074-7](https://doi.org/10.1016/S0378-4347(98)00074-7)
- (312) Holtz, R. B. (1971) Qualitative and Quantitative Analyses of Free Neutral Carbohydrates in Mushroom Tissue by Gas-Liquid Chromatography and Mass Spectrometry. *J. Agric. Food Chem.* 19, 1272–1273. <https://doi.org/10.1021/jf60178a052>
- (313) Ho, C., Lam, C., Chan, M., Cheung, R., Law, L., Lit, L., Ng, K., Suen, M., and Tai, H. (2003) Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* 24, 3–12.
- (314) Gross, J. H. (2013) *Massenspektrometrie: Ein Lehrbuch*. Springer Spektrum, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-8274-2981-0>
- (315) Dawson, P. H. (1976) *Quadrupole Mass Spectrometry and its Applications*. Elsevier. <https://doi.org/10.1016/C2013-0-04436-2>
- (316) Downard, K., and Downard, K. M. (2004) *Mass Spectrometry: A Foundation Course*. Royal Society of Chemistry.

- (317) Wollrab, E. (2014) Emergente Strukturierung durch getriebene, chemische Reaktionen in komplexen, selbstorganisierten Gemischen. Thesis, Universität des Saarlandes. <https://doi.org/10.22028/d291-22987>
- (318) Rabinovitz, M., Cohen, Y., and Halpern, M. (1986) Hydroxide Ion Initiated Reactions Under Phase Transfer Catalysis Conditions: Mechanism and Implications. *Angew. Chemie - Int. Ed.* 25, 960–970. <https://doi.org/10.1002/anie.198609601>
- (319) Banerjee, S., and Mazumdar, S. (2012) Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* 2012, 1–40. <https://doi.org/10.1155/2012/282574>
- (320) International Equipment Trading Ltd. MICROMASS Q-ToF micro MASS SPECTROMETER. *Data Sheet*.
- (321) Cleaves, H. J., Chalmers, J. H., Lazcano, A., Miller, S. L., and Bada, J. L. (2008) A Reassessment of Prebiotic Organic Synthesis in Neutral Planetary Atmospheres. *Orig. Life Evol. Biosph.* 38, 105–115. <https://doi.org/10.1007/s11084-007-9120-3>
- (322) Danielson, J. W., Snell, R. P., and Oxborrow, G. S. (1990) Detection and Quantitation of Ethylene Oxide, 2-Chloroethanol, and Ethylene Glycol with Capillary Gas Chromatography. *J. Chromatogr. Sci.* 28, 97–101. <https://doi.org/10.1093/chromsci/28.3.97>
- (323) Heftmann, E. (2004) *Chromatography: Fundamentals and applications of chromatography and related differential migration methods - Part A: Fundamentals and techniques*, 6. Auflage. Elsevier.
- (324) Gritter, R. J., Bobbitt, J. M., and Schwarting, A. E. (1987) *Einführung in die Chromatographie*. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-72789-4>
- (325) PerkinElmer. Clarus 500 Mass Spectrometer. *Data Sheet*.
- (326) Smith, C. (2005) Striving for purity: advances in protein purification. *Nat. Methods* 2, 71–77. <https://doi.org/10.1038/nmeth0105-71>
- (327) Poulsen, A. K., Arleth, L., Almdal, K., and Scharff-Poulsen, A. M. (2007) Unusually large acrylamide induced effect on the droplet size in AOT/Brij30 water-in-oil microemulsions. *J. Colloid Interface Sci.* 306, 143–153. <https://doi.org/10.1016/j.jcis.2006.10.048>
- (328) Dumetz, A. C., Lewus, R. A., Lenhoff, A. M., and Kaler, E. W. (2008) Effects of Ammonium Sulfate and Sodium Chloride Concentration on PEG/Protein Liquid-Liquid Phase Separation. *Langmuir* 24, 10345–10351. <https://doi.org/10.1021/la801180n>
- (329) Toupance, G., Bossard, A., and Raulin, F. (1977) Far UV irradiation of model prebiotic

- atmospheres. *Orig. Life Evol. Biosph.* 8, 259–266. <https://doi.org/10.1007/bf00930687>
- (330) Campana, J. E. (1980) Elementary theory of the quadrupole mass filter. *Int. J. Mass Spectrom. Ion Phys.* 33, 101–117. [https://doi.org/10.1016/0020-7381\(80\)80042-8](https://doi.org/10.1016/0020-7381(80)80042-8)
- (331) Portolés, T., Pitarch, E., López, F. J., Hernández, F., and Niessen, W. M. A. (2011) Use of soft and hard ionization techniques for elucidation of unknown compounds by gas chromatography/time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 25, 1589–1599. <https://doi.org/10.1002/rcm.5028>
- (332) Hejazi, L., Guilhaus, M., Hibbert, D. B., and Ebrahimi, D. (2015) Gas chromatography with parallel hard and soft ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 29, 91–99. <https://doi.org/10.1002/rcm.7091>
- (333) Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* 246, 64–71. <https://doi.org/10.1126/science.2675315>
- (334) Nordhoff, E., Krogsdam, A.-M., Jørgensen, H. F., Kallipolitis, B. H., Clark, B. F. C., Roepstorff, P., and Kristiansen, K. (1999) Rapid identification of DNA-binding proteins by mass spectrometry. *Nat. Biotechnol.* 17, 884–888. <https://doi.org/10.1038/12873>
- (335) Wilm, M. (2011) Principles of Electrospray Ionization. *Mol. Cell. Proteomics* 10, 1–8. <https://doi.org/10.1074/mcp.M111.009407>
- (336) Alon, T., and Amirav, A. (2015) How enhanced molecular ions in Cold EI improve compound identification by the NIST library. *Rapid Commun. Mass Spectrom.* 29, 2287–2292. <https://doi.org/10.1002/rcm.7392>
- (337) Nguyen, D. N., Becker, G. W., and Riggin, R. M. (1995) Protein mass spectrometry: applications to analytical biotechnology. *J. Chromatogr. A* 705, 21–45. [https://doi.org/10.1016/0021-9673\(94\)01256-e](https://doi.org/10.1016/0021-9673(94)01256-e)
- (338) Loo, J. A., Udseth, H. R., and Smith, R. D. (1989) Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Anal. Biochem.* 179, 404–412. [https://doi.org/10.1016/0003-2697\(89\)90153-x](https://doi.org/10.1016/0003-2697(89)90153-x)
- (339) Görlach, E., and Richmond, R. (1999) Discovery of Quasi-Molecular Ions in Electrospray Spectra by Automated Searching for Simultaneous Adduct Mass Differences. *Anal. Chem.* 71, 5557–5562. <https://doi.org/10.1021/ac9904011>
- (340) Xu, G., Stupak, J., Yang, L., Hu, L., Guo, B., and Li, J. (2018) Deconvolution in mass spectrometry based proteomics. *Rapid Commun. Mass Spectrom.* 32, 763–774. <https://doi.org/10.1002/rcm.8103>

- (341) Strobel, H. A., and Heineman, W. R. (1989) *Chemical Instrumentation: A Systematic Approach*, 3. Auflage. Wiley.
- (342) Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207. <https://doi.org/10.1038/nature01511>
- (343) Scrivens, R. (2013) Requirements for Ion Sources. In *Proceedings of the CAS–CERN Accelerator School: Ion Sources, Senec, Slovakia, 29 May – 8 June 2012*, edited by R. Bailey, CERN-2013-007. pp 1–8, CERN. <https://doi.org/10.5170/CERN-2013-007>
- (344) Wieser, M. E., and Schwieters, J. B. (2005) The development of multiple collector mass spectrometry for isotope ratio measurements. *Int. J. Mass Spectrom.* 242, 97–115. <https://doi.org/10.1016/j.ijms.2004.11.029>
- (345) Rehm, H., and Letzel, T. (2010) *Der Experimentator: Proteinbiochemie/Proteomics*, 6. Auflage. Spektrum Akademischer Verlag. <https://doi.org/10.1007/978-3-8274-2313-9>
- (346) Mamyrin, B. A. (2001) Time-of-flight mass spectrometry (concepts, achievements, and prospects). *Int. J. Mass Spectrom.* 206, 251–266. [https://doi.org/10.1016/S1387-3806\(00\)00392-4](https://doi.org/10.1016/S1387-3806(00)00392-4)
- (347) Wick, C. H. (2013) *Identifying Microbes by Mass Spectrometry Proteomics*. CRC Press.
- (348) Cornish, T. J., and Cotter, R. J. (1993) A curved-field reflectron for improved energy focusing of product ions in time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 7, 1037–1040. <https://doi.org/10.1002/rcm.1290071114>
- (349) Schermann, J.-P. (2008) 3 - Experimental Methods. In *Spectroscopy and Modeling of Biomolecular Building Blocks*. pp 129–207, Elsevier. <https://doi.org/10.1016/B978-044452708-0.50004-6>
- (350) Nikolaev, E. N., Kostyukevich, Y. I., and Vladimirov, G. (2019) Chapter 4 - Fundamentals and simulations in FT-ICR-MS. In *Fundamentals and Applications of Fourier Transform Mass Spectrometry*. pp 89–111, Elsevier. <https://doi.org/10.1016/B978-0-12-814013-0.00004-1>
- (351) Mellon, F. A. (2003) MASS SPECTROMETRY | Principles and Instrumentation. In *Encyclopedia of Food Sciences and Nutrition*, 2. Auflage. pp 3739–3749, Academic Press. <https://doi.org/10.1016/B0-12-227055-X/00746-X>
- (352) Heeren, R. M. A. (2017) MALDI Techniques in Mass Spectrometry Imaging. In *Encyclopedia of Spectroscopy and Spectrometry*, 3. Auflage. pp 711–718, Academic Press. <https://doi.org/10.1016/B978-0-12-803224-4.00008-X>
- (353) Arentz, G., Mittal, P., Zhang, C., Ho, Y.-Y., Briggs, M., Winderbaum, L., Hoffmann, M. K., and

- Hoffmann, P. (2017) Chapter Two - Applications of Mass Spectrometry Imaging to Cancer. In *Advances in Cancer Research*. Vol. 134, pp 27–66, Academic Press. <https://doi.org/10.1016/bs.acr.2016.11.002>
- (354) Kellogg, M. D. (2017) Chapter 8 - Measurement of Biological Materials. In *Clinical and Translational Science: Principles of Human Research*, 2. Auflage. pp 137–155, Academic Press. <https://doi.org/10.1016/B978-0-12-802101-9.00008-9>
- (355) Heyns, K., and Grützmaier, H.-F. (1963) Massenspektrometrische Untersuchungen, IV. Massenspektren von *N*-Acetyl-peptiden einfacher Monoaminocarbonsäuren. *Justus Liebigs Ann. Chem.* 669, 189–201. <https://doi.org/10.1002/jlac.19636690121>
- (356) Siuzdak, G. (1996) *Mass Spectrometry for Biotechnology*. Academic Press. <https://doi.org/10.1016/B978-0-12-647471-8.X5000-5>
- (357) Suzuki, T., Hayashi, S., Ishiuchi, S., Saeki, M., and Fujii, M. (2005) A New, Highly Sensitive Time-of-Flight Mass Spectrometer Consisting of a Flangeon-type Conical Ion Lens System and a Proto-type Daly Detector for Exhaust Gas Analysis Based on the Jet-REMPI Technique. *Anal. Sci.* 21, 991–996. <https://doi.org/10.2116/analsci.21.991>
- (358) Brehm, B., Grosser, J., Ruschinski, T., and Zimmer, M. (1995) Absolute detection efficiencies of a microchannel plate detector for ions. *Meas. Sci. Technol.* 6, 953–958. <https://doi.org/10.1088/0957-0233/6/7/015>
- (359) Holbrow, C. H., Lloyd, J. N., Amato, J. C., Galvez, E., and Parks, M. E. (2010) *Modern Introductory Physics*, 2. Auflage. Springer, New York, NY. <https://doi.org/10.1007/978-0-387-79080-0>
- (360) Moore, L. D., Le, T., and Fan, G. (2013) DNA Methylation and Its Basic Function. *Neuropsychopharmacology* 38, 23–38. <https://doi.org/10.1038/npp.2012.112>
- (361) Attwood, J. T., Yung, R. L., and Richardson, B. C. (2002) DNA methylation and the regulation of gene transcription. *Cell. Mol. Life Sci.* 59, 241–257. <https://doi.org/10.1007/s00018-002-8420-z>
- (362) Nan, X., Campoy, F. J., and Bird, A. (1997) MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell* 88, 471–481. [https://doi.org/10.1016/S0092-8674\(00\)81887-5](https://doi.org/10.1016/S0092-8674(00)81887-5)
- (363) Buck-Koehntop, B. A., and Defossez, P.-A. (2013) On how mammalian transcription factors recognize methylated DNA. *Epigenetics* 8, 131–137. <https://doi.org/10.4161/epi.23632>
- (364) Warren, A. J. (2002) Eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* 12, 107–114. [https://doi.org/10.1016/s0959-440x\(02\)00296-8](https://doi.org/10.1016/s0959-440x(02)00296-8)

- (365) Khrapunov, S., Warren, C., Cheng, H., Berko, E. R., Grealley, J. M., and Brenowitz, M. (2014) Unusual Characteristics of the DNA Binding Domain of Epigenetic Regulatory Protein MeCP2 Determine Its Binding Specificity. *Biochemistry* 53, 3379–3391. <https://doi.org/10.1021/bi500424z>
- (366) Ghosh, R. P., Nikitina, T., Horowitz-Scherer, R. A., Gierasch, L. M., Uversky, V. N., Hite, K., Hansen, J. C., and Woodcock, C. L. (2010) Unique Physical Properties and Interactions of the Domains of Methylated DNA Binding Protein 2. *Biochemistry* 49, 4395–4410. <https://doi.org/10.1021/bi9019753>
- (367) Nan, X., Meehan, R. R., and Bird, A. (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* 21, 4886–4892. <https://doi.org/10.1093/nar/21.21.4886>
- (368) Typas, A., Becker, G., and Hengge, R. (2007) The molecular basis of selective promoter activation by the σ^S subunit of RNA polymerase. *Mol. Microbiol.* 63, 1296–1306. <https://doi.org/10.1111/j.1365-2958.2007.05601.x>
- (369) Marr, M. T., and Roberts, J. W. (1997) Promoter Recognition As Measured by Binding of Polymerase to Nontemplate Strand Oligonucleotide. *Science* 276, 1258–1260. <https://doi.org/10.1126/science.276.5316.1258>
- (370) Ho, K. L. (2009) Structural studies of MeCP2 in complex with methylated DNA. Thesis, University of Edinburgh.
- (371) Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. *Cell* 105, 487–497. [https://doi.org/10.1016/S0092-8674\(01\)00324-5](https://doi.org/10.1016/S0092-8674(01)00324-5)
- (372) Yen, L., Svendsen, J., Lee, J.-S., Gray, J. T., Magnier, M., Baba, T., D'Amato, R. J., and Mulligan, R. C. (2004) Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* 431, 471–476. <https://doi.org/10.1038/nature02844>
- (373) Mapp, A. K., and Ansari, A. Z. (2007) A TAD Further: Exogenous Control of Gene Activation. *ACS Chem. Biol.* 2, 62–75. <https://doi.org/10.1021/cb600463w>
- (374) Dubrovina, A. S., and Kiselev, K. V. (2019) Exogenous RNAs for Gene Regulation and Plant Resistance. *Int. J. Mol. Sci.* 20, 1–21. <https://doi.org/10.3390/ijms20092282>
- (375) Liang, H., Huang, L., Cao, J., Zen, K., Chen, X., and Zhang, C.-Y. (2012) Regulation of mammalian gene expression by exogenous microRNAs. *Wiley Interdiscip. Rev. RNA* 3, 733–742. <https://doi.org/10.1002/wrna.1127>
- (376) Kucukkal, T. G., Yang, Y., Uvarov, O., Cao, W., and Alexov, E. (2015) Impact of Rett Syndrome

- Mutations on MeCP2 MBD Stability. *Biochemistry* 54, 6357–6368. <https://doi.org/10.1021/acs.biochem.5b00790>
- (377) Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007) Reversal of Neurological Defects in a Mouse Model of Rett Syndrome. *Science* 315, 1143–1147. <https://doi.org/10.1126/science.1138389>
- (378) Båga, M., Göransson, M., Normark, S., and Uhlin, B. E. (1985) Transcriptional activation of a Pap pilus virulence operon from uropathogenic *Escherichia coli*. *EMBO J.* 4, 3887–3893. <https://doi.org/10.1002/j.1460-2075.1985.tb04162.x>
- (379) Forsman, K., Göransson, M., and Uhlin, B. E. (1989) Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. *EMBO J.* 8, 1271–1277. <https://doi.org/10.1002/j.1460-2075.1989.tb03501.x>
- (380) Nou, X., Braaten, B., Kaltenbach, L., and Low, D. A. (1995) Differential binding of Lrp to two sets of *pap* DNA binding sites mediated by PapI regulates Pap phase variation in *Escherichia coli*. *EMBO J.* 14, 5785–5797. <https://doi.org/10.1002/j.1460-2075.1995.tb00267.x>
- (381) Sánchez-Romero, M. A., Cota, I., and Casadesús, J. (2015) DNA methylation in bacteria: from the methyl group to the methylome. *Curr. Opin. Microbiol.* 25, 9–16. <https://doi.org/10.1016/j.mib.2015.03.004>
- (382) van der Woude, M., Braaten, B., and Low, D. (1996) Epigenetic phase variation of the *pap* operon in *Escherichia coli*. *Trends Microbiol.* 4, 5–9. [https://doi.org/10.1016/0966-842X\(96\)81498-3](https://doi.org/10.1016/0966-842X(96)81498-3)
- (383) Peterson, S. N., and Reich, N. O. (2008) Competitive Lrp and Dam Assembly at the *pap* Regulatory Region: Implications for Mechanisms of Epigenetic Regulation. *J. Mol. Biol.* 383, 92–105. <https://doi.org/10.1016/j.jmb.2008.07.086>
- (384) Göransson, M., Forsman, K., Nilsson, P., and Uhlin, B. E. (1989) Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus-adhesin in *Escherichia coli*. *Mol. Microbiol.* 3, 1557–1565. <https://doi.org/10.1111/j.1365-2958.1989.tb00141.x>
- (385) Weyand, N. J., and Low, D. A. (2000) Regulation of Pap Phase Variation: Lrp IS SUFFICIENT FOR THE ESTABLISHMENT OF THE PHASE OFF *pap* DNA METHYLATION PATTERN AND REPRESSION OF *pap* TRANSCRIPTION IN VITRO. *J. Biol. Chem.* 275, 3192–3200. <https://doi.org/10.1074/jbc.275.5.3192>
- (386) Weyand, N. J., Braaten, B. A., van der Woude, M., Tucker, J., and Low, D. A. (2001) The essential role of the promoter-proximal subunit of CAP in *pap* phase variation: Lrp- and helical

- phase-dependent activation of *papBA* transcription by CAP from -215. *Mol. Microbiol.* 39, 1504–1522. <https://doi.org/10.1046/j.1365-2958.2001.02338.x>
- (387) Leonard, P. M., Smits, S. H. J., Sedelnikova, S. E., Brinkman, A. B., de Vos, W. M., van der Oost, J., Rice, D. W., and Rafferty, J. B. (2001) Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO J.* 20, 990–997. <https://doi.org/10.1093/emboj/20.5.990>
- (388) de los Rios, S., and Perona, J. J. (2007) Structure of the *Escherichia coli* Leucine-responsive Regulatory Protein Lrp Reveals a Novel Octameric Assembly. *J. Mol. Biol.* 366, 1589–1602. <https://doi.org/10.1016/j.jmb.2006.12.032>
- (389) Kawamura, T., Vartanian, A. S., Zhou, H., and Dahlquist, F. W. (2011) The Design Involved in PapI and Lrp Regulation of the *pap* Operon. *J. Mol. Biol.* 409, 311–332. <https://doi.org/10.1016/j.jmb.2011.01.058>
- (390) Peterson, S. N., and Reich, N. O. (2006) GATC Flanking Sequences Regulate Dam Activity: Evidence for how Dam Specificity may Influence *pap* Expression. *J. Mol. Biol.* 355, 459–472. <https://doi.org/10.1016/j.jmb.2005.11.003>
- (391) Mierzejewska, K., Siwek, W., Czapinska, H., Kaus-Drobek, M., Radlinska, M., Skowronek, K., Bujnicki, J. M., Dadlez, M., and Bochtler, M. (2014) Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Res.* 42, 8745–8754. <https://doi.org/10.1093/nar/gku546>
- (392) Li, F., Liu, S.-L., and Mullins, J. I. (1999) Site-Directed Mutagenesis Using Uracil-Containing Double-Stranded DNA Templates and DpnI Digestion. *BioTechniques* 27, 734–738. <https://doi.org/10.2144/99274st03>
- (393) Cui, Y., Midkiff, M. A., Wang, Q., and Calvo, J. M. (1996) The Leucine-responsive Regulatory Protein (Lrp) from *Escherichia coli*: STOICHIOMETRY AND MINIMAL REQUIREMENTS FOR BINDING TO DNA. *J. Biol. Chem.* 271, 6611–6617. <https://doi.org/10.1074/jbc.271.12.6611>
- (394) Jeong, H., Kim, H. J., and Lee, S. J. (2015) Complete Genome Sequence of *Escherichia coli* Strain BL21. *Genome Announc.* 3, 1. <https://doi.org/10.1128/genomeA.00134-15>
- (395) Kaltenbach, L. S., Braaten, B. A., and Low, D. A. (1995) Specific Binding of PapI to Lrp-*pap* DNA Complexes. *J. Bacteriol.* 177, 6449–6455. <https://doi.org/10.1128/JB.177.22.6449-6455.1995>
- (396) Calvo, J. M., and Matthews, R. G. (1994) The Leucine-Responsive Regulatory Protein, a Global Regulator of Metabolism in *Escherichia coli*. *Microbiol. Rev.* 58, 466–490. <https://doi.org/10.1128/mr.58.3.466-490.1994>

- (397) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. <https://doi.org/10.1093/nar/gkg595>
- (398) Bae, S.-H., Cheong, H.-K., Cheong, C., Kang, S., Hwang, D. S., and Choi, B.-S. (2003) Structure and Dynamics of Hemimethylated GATC Sites: IMPLICATIONS FOR DNA-SeqA RECOGNITION. *J. Biol. Chem.* 278, 45987–45993. <https://doi.org/10.1074/jbc.M306038200>
- (399) Han, J. S., Kang, S., Lee, H., Kim, H. K., and Hwang, D. S. (2003) Sequential Binding of SeqA to Paired Hemi-methylated GATC Sequences Mediates Formation of Higher Order Complexes. *J. Biol. Chem.* 278, 34983–34989. <https://doi.org/10.1074/jbc.M304923200>
- (400) Hohng, S., Zhou, R., Nahas, M. K., Yu, J., Schulten, K., Lilley, D. M. J., and Ha, T. (2007) Fluorescence-Force Spectroscopy Maps Two-Dimensional Reaction Landscape of the Holliday Junction. *Science* 318, 279–283. <https://doi.org/10.1126/science.1146113>
- (401) Swoboda, M., Grieb, M. S., Hahn, S., and Schlierf, M. (2014) Measuring Two at the Same Time: Combining Magnetic Tweezers with Single-Molecule FRET. In *Fluorescent Methods for Molecular Motors. Experientia Supplementum*. Vol. 105, pp 253–276, Springer, Basel. https://doi.org/10.1007/978-3-0348-0856-9_12
- (402) Invitrogen. (2018) Dynabeads™ MyOne™ Streptavidin T1. *Manual*.
- (403) Nord, O., Uhlén, M., and Nygren, P.-Å. (2003) Microbead display of proteins by cell-free expression of anchored DNA. *J. Biotechnol.* 106, 1–13. <https://doi.org/10.1016/j.jbiotec.2003.09.002>
- (404) Invitrogen. (2018) Dynabeads™ Streptavidin Trial Kit. *Manual*.
- (405) Naiser, T., Ehler, O., Kayser, J., Mai, T., Michel, W., and Ott, A. (2008) Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: Comparison of single base mismatches and base bulges. *BMC Biotechnol.* 8, 1–23. <https://doi.org/10.1186/1472-6750-8-48>
- (406) Invitrogen. (2020) Dynabeads™ Oligo (dT)₂₅. *Manual*.
- (407) Jayasena, S. D. (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.* 45, 1628–1650.
- (408) Song, S., Wang, L., Li, J., Zhao, J., and Fan, C. (2008) Aptamer-based biosensors. *TrAC Trends Anal. Chem.* 27, 108–117. <https://doi.org/10.1016/j.trac.2007.12.004>
- (409) Carter, Z. A., and Katakly, R. (2017) A G-quadruplex aptamer based impedimetric sensor for free lysine and arginine. *Sensors Actuators B: Chem.* 243, 904–909. <https://doi.org/10.1016/j.snb.2016.12.010>

- (410) Fiegand, L. R., Garst, A. D., Batey, R. T., and Nesbitt, D. J. (2012) Single-Molecule Studies of the Lysine Riboswitch Reveal Effector-Dependent Conformational Dynamics of the Aptamer Domain. *Biochemistry* 51, 9223–9233. <https://doi.org/10.1021/bi3007753>
- (411) Arthur, L. L., Pavlovic-Djuranovic, S., Koutmou, K. S., Green, R., Szczesny, P., and Djuranovic, S. (2015) Translational control by lysine-encoding A-rich sequences. *Sci. Adv.* 1, 1–11. <https://doi.org/10.1126/sciadv.1500154>
- (412) Koutmou, K. S., Schuller, A. P., Brunelle, J. L., Radhakrishnan, A., Djuranovic, S., and Green, R. (2015) Ribosomes slide on lysine-encoding homopolymeric A stretches. *eLife* 4, 1–18. <https://doi.org/10.7554/eLife.05534>
- (413) Eliazar, I., and Klafter, J. (2010) Ultra diffusions. *J. Phys. A: Math. Theor.* 43, 1–8. <https://doi.org/10.1088/1751-8113/43/13/132002>
- (414) Jeon, J.-H., Leijnse, N., Oddershede, L. B., and Metzler, R. (2013) Anomalous diffusion and power-law relaxation of the time averaged mean squared displacement in worm-like micellar solutions. *New J. Phys.* 15, 1–16. <https://doi.org/10.1088/1367-2630/15/4/045011>
- (415) Malacarne, L. C., Mendes, R. S., Pedron, I. T., and Lenzi, E. K. (2001) Nonlinear equation for anomalous diffusion: Unified power-law and stretched exponential exact solution. *Phys. Rev. E* 63, 1–4. <https://doi.org/10.1103/PhysRevE.63.030101>
- (416) Barthelemy, P., Bertolotti, J., and Wiersma, D. S. (2008) A Lévy flight for light. *Nature* 453, 495–498. <https://doi.org/10.1038/nature06948>
- (417) Dunn, J. A., McCance, D. R., Thorpe, S. R., Lyons, T. J., and Baynes, J. W. (1991) Age-Dependent Accumulation of N^{ϵ} -(Carboxymethyl)lysine and N^{ϵ} -(Carboxymethyl)hydroxylysine in Human Skin Collagen. *Biochemistry* 30, 1205–1210. <https://doi.org/10.1021/bi00219a007>
- (418) Cudic, M., Lauer-Fields, J. L., and Fields, G. B. (2005) Improved synthesis of 5-hydroxylysine (Hyl) derivatives. *J. Pept. Res.* 65, 272–283. <https://doi.org/10.1111/j.1399-3011.2005.00215.x>
- (419) Cudic, M., and Fields, G. B. (2020) Modulation of receptor binding to collagen by glycosylated 5-hydroxylysine: Chemical biology approaches made feasible by Carpino’s Fmoc group. *Pept. Sci.* 112, 1–12. <https://doi.org/10.1002/pep2.24156>
- (420) Popenoe, E. A., Aronson, R. B., and Van Slyke, D. D. (1966) Hydroxylysine Formation from Lysine during Collagen Biosynthesis. *Proc. Natl. Acad. Sci. USA.* 55, 393–397. <https://doi.org/10.1073/pnas.55.2.393>
- (421) Krane, S. M., Pinnell, S. R., and Erbe, R. W. (1972) Lysyl-Protocollagen Hydroxylase Deficiency in Fibroblasts from Siblings with Hydroxylysine-Deficient Collagen. *Proc. Natl. Acad. Sci. USA.* 69, 2899–2903. <https://doi.org/10.1073/pnas.69.10.2899>

- (422) Makris, E. A., MacBarb, R. F., Responde, D. J., Hu, J. C., and Athanasiou, K. A. (2013) A copper sulfate and hydroxylysine treatment regimen for enhancing collagen cross-linking and biomechanical properties in engineered neocartilage. *FASEB J.* 27, 2421–2430. <https://doi.org/10.1096/fj.12-224030>
- (423) Maisani, M., Ziane, S., Ehret, C., Levesque, L., Siadous, R., Le Meins, J.-F., Chevallier, P., Barthélémy, P., De Oliveira, H., Amédée, J., Mantovani, D., and Chassande, O. (2018) A new composite hydrogel combining the biological properties of collagen with the mechanical properties of a supramolecular scaffold for bone tissue engineering. *J. Tissue Eng. Regen. Med.* 12, 1489–1500. <https://doi.org/10.1002/term.2569>
- (424) Ma, L., Gao, C., Mao, Z., Zhou, J., Shen, J., Hu, X., and Han, C. (2003) Collagen/chitosan porous scaffolds with improved biostability for skin tissue engineering. *Biomaterials* 24, 4833–4841. [https://doi.org/10.1016/S0142-9612\(03\)00374-0](https://doi.org/10.1016/S0142-9612(03)00374-0)
- (425) Gjaltema, R. A. F., and Bank, R. A. (2017) Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. *Crit. Rev. Biochem. Mol. Biol.* 52, 74–95. <https://doi.org/10.1080/10409238.2016.1269716>
- (426) Deane, C. S., Bass, J. J., Crossland, H., Phillips, B. E., and Atherton, P. J. (2020) Animal, Plant, Collagen and Blended Dietary Proteins: Effects on Musculoskeletal Outcomes. *Nutrients* 12, 1–35. <https://doi.org/10.3390/nu12092670>
- (427) Ong, S.-E. (2012) The expanding field of SILAC. *Anal. Bioanal. Chem.* 404, 967–976. <https://doi.org/10.1007/s00216-012-5998-3>
- (428) Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* 1, 376–386. <https://doi.org/10.1074/mcp.M200025-MCP200>
- (429) Ong, S.-E., Kratchmarova, I., and Mann, M. (2003) Properties of ¹³C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *J. Proteome Res.* 2, 173–181. <https://doi.org/10.1021/pr0255708>
- (430) van den Broek, I., Niessen, W. M. A., and van Dongen, W. D. (2013) Bioanalytical LC–MS/MS of protein-based biopharmaceuticals. *J. Chromatogr. B* 929, 161–179. <https://doi.org/10.1016/j.jchromb.2013.04.030>
- (431) Kaddis, C. S., and Loo, J. A. (2007) Native Protein MS and Ion Mobility: Large Flying Proteins with ESI. *Anal. Chem.* 79, 1778–1784. <https://doi.org/10.1021/ac071878c>
- (432) Loo, J. A., Loo, R. R. O., Udseth, H. R., Edmonds, C. G., and Smith, R. D. (1991) Solvent-

- induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 5, 101–105. <https://doi.org/10.1002/rcm.1290050303>
- (433) Lomeli, S. H., Yin, S., Loo, R. R. O., and Loo, J. A. (2009) Increasing charge while preserving noncovalent protein complexes for ESI-MS. *J. Am. Soc. Mass Spectrom.* 20, 593–596. <https://doi.org/10.1016/j.jasms.2008.11.013>
- (434) Javanshad, R., Maser, T. L., Honarvar, E., and Venter, A. R. (2019) The Addition of Polar Organic Solvent Vapors During the Analysis of Proteins by DESI-MS. *J. Am. Soc. Mass Spectrom.* 30, 2571–2575. <https://doi.org/10.1007/s13361-019-02345-w>
- (435) Zymo Research Corp. His-Spin Protein Miniprep™. *Manual*.
- (436) Thermo Fisher Scientific. (2019) Pierce™ Protein Concentrators, PES. *Manual*.
- (437) Burkhardt, A., and Grupe, G. (2018) Hydroxylysine deficiency, conspicuous skeletal lesions and a strange burial practice in a historical German island population. *Int. J. Osteoarchaeol.* 28, 227–236. <https://doi.org/10.1002/oa.2649>
- (438) Liu, D., Nikoo, M., Boran, G., Zhou, P., and Regenstein, J. M. (2015) Collagen and Gelatin. *Annu. Rev. Food Sci. Technol.* 6, 527–557. <https://doi.org/10.1146/annurev-food-031414-111800>
- (439) van Heijst, J. W. J., Niessen, H. W. M., Hoekman, K., and Schalkwijk, C. G. (2005) Advanced Glycation End Products in Human Cancer Tissues: Detection of N^ε-(Carboxymethyl)lysine and Argpyrimidine. *Ann. N. Y. Acad. Sci.* 1043, 725–733. <https://doi.org/10.1196/annals.1333.084>
- (440) van Heijst, J. W. J., Niessen, H. W. M., Musters, R. J., van Hinsbergh, V. W. M., Hoekman, K., and Schalkwijk, C. G. (2006) Argpyrimidine-modified Heat Shock Protein 27 in human non-small cell lung cancer: A possible mechanism for evasion of apoptosis. *Cancer Lett.* 241, 309–319. <https://doi.org/10.1016/j.canlet.2005.10.042>
- (441) Castellani, R. J., Harris, P. L. ., Sayre, L. M., Fujii, J., Taniguchi, N., Vitek, M. P., Founds, H., Atwood, C. S., Perry, G., and Smith, M. A. (2001) Active glycation in neurofibrillary pathology of Alzheimer disease: N^ε-(Carboxymethyl) lysine and hexitol-lysine. *Free Radic. Biol. Med.* 31, 175–180. [https://doi.org/10.1016/s0891-5849\(01\)00570-6](https://doi.org/10.1016/s0891-5849(01)00570-6)
- (442) Pashikanti, S., Boissonneault, G. A., and Cervantes-Laurean, D. (2011) Ex vivo detection of histone H1 modified with advanced glycation end products. *Free Radic. Biol. Med.* 50, 1410–1416. <https://doi.org/10.1016/j.freeradbiomed.2011.01.034>
- (443) Wilker, S. C., Chellan, P., Arnold, B. M., and Nagaraj, R. H. (2001) Chromatographic Quantification of Argpyrimidine, a Methylglyoxal-Derived Product in Tissue Proteins:

- Comparison with Pentosidine. *Anal. Biochem.* 290, 353–358. <https://doi.org/10.1006/abio.2001.4992>
- (444) Nass, N., Ignatov, A., Andreas, L., Weißenborn, C., Kalinski, T., and Sel, S. (2017) Accumulation of the advanced glycation end product carboxymethyl lysine in breast cancer is positively associated with estrogen receptor expression and unfavorable prognosis in estrogen receptor-negative cases. *Histochem. Cell Biol.* 147, 625–634. <https://doi.org/10.1007/s00418-016-1534-4>
- (445) Schleicher, E. D., Wagner, E., and Nerlich, A. G. (1997) Increased Accumulation of the Glycooxidation Product N^ε-(carboxymethyl)lysine in Human Tissues in Diabetes and Aging. *J. Clin. Invest.* 99, 457–468. <https://doi.org/10.1172/JCI119180>
- (446) Zhu, W., Smith, J. W., and Huang, C.-M. (2010) Mass Spectrometry-Based Label-Free Quantitative Proteomics. *J. Biomed. Biotechnol.* 2010, 1–6. <https://doi.org/10.1155/2010/840518>
- (447) Ong, S.-E., Foster, L. J. ., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* 29, 124–130. [https://doi.org/10.1016/s1046-2023\(02\)00303-1](https://doi.org/10.1016/s1046-2023(02)00303-1)
- (448) Domon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* 28, 710–721. <https://doi.org/10.1038/nbt.1661>
- (449) Qian, H. (2008) Cooperativity and Specificity in Enzyme Kinetics: A Single-Molecule Time-Based Perspective. *Biophys. J.* 95, 10–17. <https://doi.org/10.1529/biophysj.108.131771>
- (450) Elliott, M. H., Smith, D. S., Parker, C. E., and Borchers, C. (2009) Current trends in quantitative proteomics. *J. Mass Spectrom.* 44, 1637–1660. <https://doi.org/10.1002/jms.1692>
- (451) Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., and Aebersold, R. (2017) Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat. Protoc.* 12, 1289–1294. <https://doi.org/10.1038/nprot.2017.040>
- (452) Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 35, 265–273. <https://doi.org/10.1016/j.ymeth.2004.08.018>
- (453) Kettenbach, A. N., Rush, J., and Gerber, S. A. (2011) Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides. *Nat. Protoc.* 6, 175–186. <https://doi.org/10.1038/nprot.2010.196>
- (454) Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA.* 100, 6940–6945. <https://doi.org/10.1073/pnas.0832254100>

Literaturverzeichnis

- (455) Brun, V., Masselon, C., Garin, J., and Dupuis, A. (2009) Isotope dilution strategies for absolute quantitative proteomics. *J. Proteomics* 72, 740–749. <https://doi.org/10.1016/j.jprot.2009.03.007>
- (456) Crine, P., and Lemieux, E. (1982) Incorporation of Canavanine into Rat Pars Intermedia Proteins Inhibits the Maturation of Pro-opiomelanocortin, the Common Precursor to Adrenocorticotropin and β -Lipotropin. *J. Biol. Chem.* 257, 832–838. [https://doi.org/10.1016/s0021-9258\(19\)68272-4](https://doi.org/10.1016/s0021-9258(19)68272-4)