

Zusammenfassung von dem Artikel “Optimality and evolutionary tuning of the expression level of a protein“ [1]

Sandro Andreotti

24. Februar 2006

1 Hintergrund

In dem Experiment wurde ein Kosten-Nutzen Modell für die Expression des Lac-Operon in E.coli erstellt. Die Kosten- und Nutzen-Funktionen wurden an experimentell ermittelten Daten gefittet, mit dem Ziel, die optimale Expression für das Lac-Protein β -Galactosidase, welches für die Metabolisierung von Lactose nötig ist, unter verschiedenen Lactosekonzentrationen des Nährmediums vorauszusagen.

2 Motivation

Es ist aus vielen Versuchen schon lange bekannt, dass Organismen wie E.coli die Fähigkeit besitzen ihren Phänotyp über mehrere Generationen zu verändern, um sich so an äußere Bedingungen anzupassen und die Wachstumsrate zu erhöhen. Dieses Phänomen fällt besonders negativ auf bei der Bildung von Resistenzen bei krankheitserregenden Mikroorganismen und Viren. Für die Veränderung der Bakterien über mehrere Generationen kommen zwei Prozesse in Frage

- neutraler evolutionärer Gendrift
- Selektion von Mutanten mit erhöhter Fitness

Der Gendrift stellt dabei eine zufällige Änderung des Genpools dar, welche durch zufällige Änderung der Genfrequenzen in einer Population bedingt ist. Diese Änderung ist dabei unabhängig von der Fitness, des durch Gendrift entstandenen Phänotyps. Der zweite Punkt stellt den typischen Prozess von Mutation und Selektion dar, wobei sich Mutanten mit erhöhter Fitness durchsetzen. Dabei wird häufig von einem Optimierungsprozess ausgegangen. Es wurden bereits viele Fälle von Optimierung auf der Ebene des ganzen Organismus Phänotyps experimentell demonstriert. In diesem Versuch wird hingegen die Optimalität auf der Proteinebene untersucht. Die konkrete Fragestellung lautet dabei, ob man das (optimale) Expressionslevel für ein bestimmtes Protein unter bestimmten äußeren Faktoren voraussagen kann.

3 Das untersuchte System

Das untersuchte System in diesem Versuch ist das Lac-Operon in E.coli. Das Lac-Operon enthält drei Gene LacA, LacY, LacZ, welche Proteine codieren die für die Metabolisierung von Lactose notwendig sind. Das untersuchte Protein ist, die durch LacZ codierte, β -Galactosidase. β -Galactosidase katalysiert die hydrolytische Spaltung von Lactose in Galactose und Glucose, und damit den ersten Schritt zur Energiegewinnung aus Lactose. In Abbildung 1 ist eine schematische Versuchsanordnung zusammen mit den wichtigsten Prozessen am Lac-Operon dargestellt. Eine wichtige Rolle in den durchgeführten Versuchen spielt der Induktor IPTG (Isopropyl- β -Thiogalaktosid). IPTG bewirkt ebenso wie Lactose die Dissoziation des Lac Repressorproteins vom Operator, wird jedoch nicht metabolisiert und kann daher nicht in Form von Energie nutzbar gemacht werden.

4 Kosten - Nutzen Modell

Um von Optimierung sprechen zu können und sie zu modellieren, ist eine möglichst realistische Fitness-Funktion zu definieren, welche für jeden Phänotyp seine Fitness ermittelt. In diesem Experiment wird die jeweilige Expression des LacZ Gens als Phänotyp aufgefasst. Dementsprechend muss die Fitnessfunktion eine Funktion in Abhängigkeit von der LacZ Expression, bei einer gegebenen Lactose Konzentration, sein. Die Fitnessfunktion besteht aus den zwei Komponenten Kosten und Nutzen, welche für diesen Versuch wie folgt definiert sind.

- Kosten: relative Reduktion der Wachstumsrate aufgrund der Belastung durch die Produktion und Anwesenheit von Lac-Proteinen (relativ zu Zellen ohne Lac Expression)

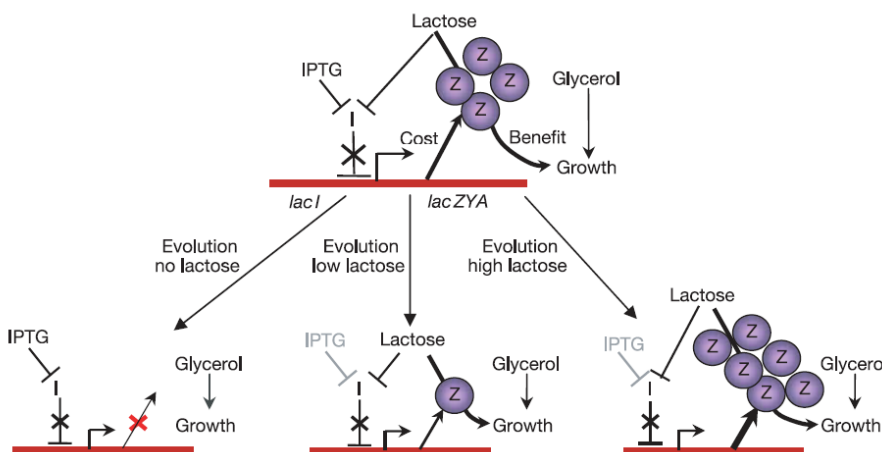


Abbildung 1: exemplarische Übersicht des Lac Operon und der Versuchsanordnung. E.coli Stämme werden in einer Verdünnungsreihe bei verschiedenen Lactosekonzentrationen gehalten und ändern evolutionär das Expressionslevel des LacZ Proteins. Glycerin ist in allen Testreihen als (relativ schwacher) Energielieferant vorhanden.

- Nutzen: relative Zunahme der Wachstumsrate bedingt durch den Energiegewinn aus der Lactose Metabolisierung (relativ zu Zellen ohne Lactose)

Die Fitness-Funktion ergibt sich somit als Nutzen - Kosten. In diesem Modell ist schließlich diese Expression optimal, für welche die Fitness-Funktion ihr Maximum annimmt. Nachdem Kosten und Nutzen definiert sind, müssen beide genauer bestimmt werden. In diesem Versuch werden sie direkt gemessen. Um die Kosten zu bestimmen, misst man die Wachstumsrate von E.coli auf einem Medium ohne Lactose und bei unterschiedlichen Konzentrationen von IPTG. Diese Versuchsanordnung gestattet die isolierte Bestimmung der Kosten, da der Nutzen, aufgrund der Abwesenheit von Lactose, stets konstant bei 0 liegt. In Abbildung 2 sind die gemessenen Werte einiger Proben abgetragen. Der Parameter Z_{WT} kennzeichnet dabei die Expression von LacZ für das voll induzierte Lac Operon. Es ist deutlich zu erkennen, dass die Kosten nicht linear ansteigen, also die Kosten für die Produktion und Anwesenheit eines LacZ Moleküls abhängig sind von der herrschenden Konzentration an LacZ. Die biologische Erklärung für diese Beobachtung könnte sein, dass die Zelle beschränkte Ressourcen hat, und eine hohe Produktionsrate von Lac Proteinen die Synthese von anderen essentiellen Proteinen einschränkt. Zur Beschreibung der gemessenen Daten werden zunächst zwei Kandidaten für die Kostenfunktion vorgestellt.

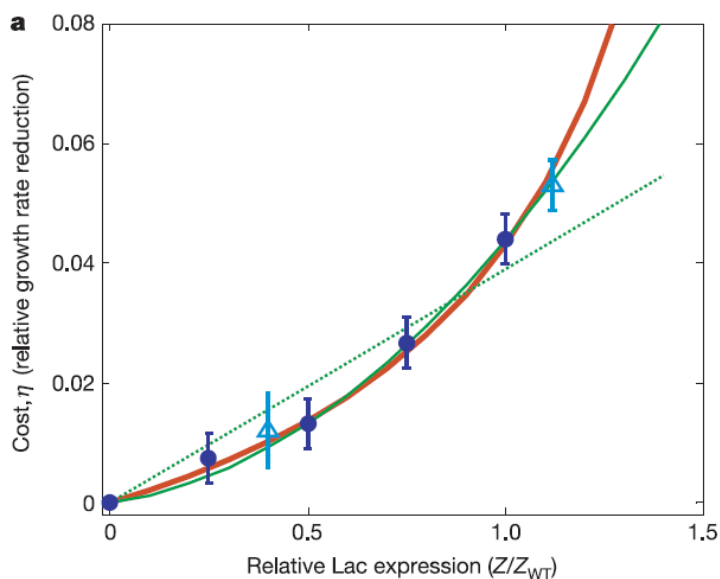


Abbildung 2: gemessene Kosten und Verlauf der gefitteten Kostenfunktionen. Rot kennzeichnet Verlauf von Funktion 2, Grün ist Verlauf von Funktion 1. Gestrichelt dargestellt ist der Verlauf von linearer Kostenfunktion. Messwerte als Punkt stammen von Wildtyp, Dreieck bei 0.4 auf x-Achse von Stamm nach 530 Generationen bei 0.2mM Lactose, Dreieck bei 1.12 auf x-Achse von Stamm nach 400 Generationen bei 5mM Lactose

$$\eta_1(Z) = \eta_0 Z + \eta'_0 Z^2 \quad (1)$$

$$\eta_2(Z) = \frac{\eta_0 Z}{1 - Z/M} \quad (2)$$

In Funktion 1 geht das Expressionslevel von LacZ (Z) einmal linear und einmal quadriert ein. Durch Fitten der Funktion an die ermittelten Daten ergaben sich für die Parameter η_0 und η'_0 die Werte $\eta_0 Z_{WT} = 0.09 \pm 0.01$ und $\eta'_0 Z_{WT}^2 = 3 \pm 0.1$. Der Verlauf von Funktion 1 ist in Abbildung 2 gezeigt (grüne Linie). Funktion 2, die, wie sich im weiteren Verlauf der Untersuchungen zeigte, eine bessere Modellierung darstellt, verwendet einen Parameter M . Der Parameter M stellt dabei die maximale Kapazität für die Produktion nicht essentieller Proteine dar, womit die oben angegebene biologische Erklärung für die gemessenen Daten in die Modellierung einfließt. Durch Fitten an die experimentellen Daten ergaben sich für die beiden Parameter η_0 und M die Werte $\eta_0 Z_{WT} = 0.02 \pm 0.003$ und $M = (1.8 \pm 0.3) Z_{WT}$. Der Verlauf der gefitteten Funktion 2 ist in Abbildung 2 in Rot dargestellt. Im Folgenden ist eine kurze Herleitung von Funktion 2 gegeben. Die Ausgangsfunktion stellt dabei die Funktion für die Wachstumsrate mit interner Ressource R dar.

$$g = \beta R / (K + R)$$

Dabei definiert β die maximale Wachstumsrate und K stellt das Ressourcenlevel für ein genau halbmaximales Wachstum dar. Wird nun ein Protein Z produziert, so reduziert dies die Ressourcen um ϵ . Also gilt:

$$g(Z) = \beta(R - \epsilon Z) / (K + R - \epsilon Z)$$

Somit ergibt sich die Kostenfunktion

$$\eta(Z) = (g(0) - g(Z)) / g(0) = \eta_0 Z / (1 - Z/M).$$

Der letzte Term stellt dabei genau die verwendete Funktion 2 dar, mit $\eta_0 = \epsilon K / R(K + R)$ und $M = (K + R / \epsilon)$

Zur Bestimmung des Nutzen wird die Wachstumsrate bei verschiedenen Konzentrationen von Lactose gemessen, wobei das Lac System stets mittels IPTG voll induziert bleibt. Analog zur Kostenbestimmung erlaubt dieser Versuchsaufbau die isolierte Bestimmung des Nutzen, da die Kosten durch die Produktion der Lac Proteine mittels IPTG stets maximal und somit konstant gehalten werden. Die Änderung der Wachstumsrate in Abhängigkeit der Lactosekonzentration ist in Abbildung 3 dargestellt. Die Meßwerte beschreiben relativ deutlich einen sigmoidalen Verlauf, und man kann zeigen, dass die Nutzenfunktion sehr gut durch die bekannte Transport- und Katabolismuskinetik von LacZ und LacY beschrieben sind. Die sich daraus ergebende Funktion für den Nutzen ist:

$$B(Z) = \delta [Z L_{in}] \quad (3)$$

Der Parameter L_{in} kennzeichnet die intrazelluläre Lactosekonzentration und δ stellt den Wachstumsvorteil je LacZ Molekül bei Lactosesättigung dar, welcher aus Abbildung 3 direkt entnommen werden kann ($\delta = 0.17 / Z_{WT}$). Funktion 3

berechnet den Nutzen jedoch in Abhängigkeit der intrazellulären Lactosekonzentration L_{in} , wobei das Ziel aber ist, den Nutzen in Abhängigkeit von der Lactosekonzentration des umgebenden Nährmediums zu quantifizieren. Man umgeht dieses Problem indem man sich auf das Ergebnis von Wong, Gladney&Keasing (1997) stützt, welches besagt, dass die Umsatzrate von LacZ proportional zum Einstrom von Lactose in die Zelle, und damit zur Permease Aktivität von LacY, ist. Somit gilt:

$$V_Z[ZL_{in}] \approx V_Y[YL] \approx V_Y \frac{YL}{K_Y + L}$$

Dabei sind V_Z bzw. V_Y die Umsatzgeschwindigkeiten von LacZ und LacY, K_Y die Michaelis Konstante von LacY und Y bezeichnet die Konzentration von LacY. Wenn nun der Quotient von Z und Y , sowie der Umsatzgeschwindigkeiten V_Y und V_Z mit in den Proportionalitätsfaktor δ eingehen, so erhält man als Kostenfunktion in Abhängigkeit von der äußeren Lactose Konzentration die Funktion 4

$$B(Z) = \delta \frac{ZL}{K_Y + L}. \quad (4)$$

Die Fitnessfunktion ergibt sich wie gewohnt aus der Differenz von Nutzen und Kosten, also

$$g(Z) = \delta \frac{ZL}{K_Y + L} - \frac{\eta_0 Z}{1 - Z/M}. \quad (5)$$

In Abbildung 4 sind die Verläufe der Fitnessfunktion g in Abhängigkeit von Z für drei verschiedene Lactosekonzentrationen gezeigt. Der Verlauf der Funktionen zeigt, dass es für die verschiedenen Lactosekonzentrationen einen optimalen Wert für die Expression von LacZ gibt. Diesen Wert bestimmt man analytisch über die Ableitung der Fitnessfunktion nach Z und Nullsetzen $dg/dz = 0$. Aus

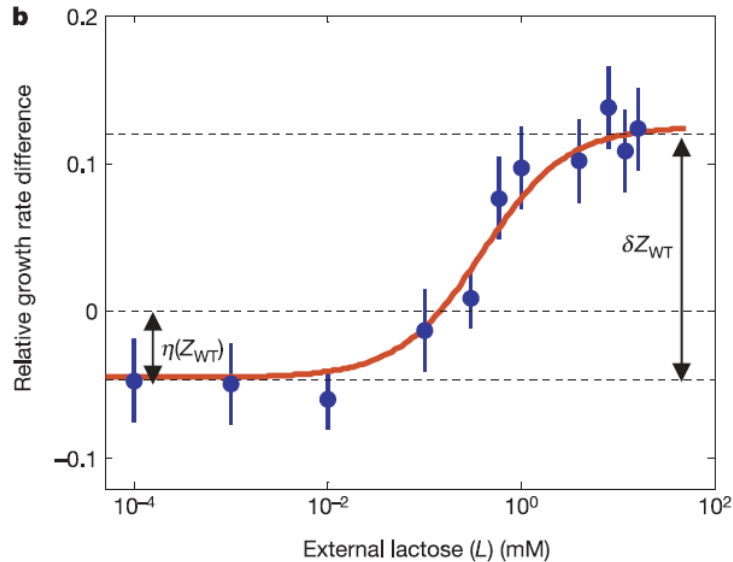


Abbildung 3: gemessener relativer Wachstumsvorteil bei voll induzierten Lac Operon, unter verschiedenen externen Lactose Konzentrationen. Rote Linie kennzeichnet Verlauf der gefitteten Nutzenfunktion (Funktion 3)

der daraus resultierenden Formel berechnet sich die optimale Expression für LacZ (Z_{OPT}) wie folgt:

$$Z_{opt} = M \left(1 - \sqrt{\frac{\eta_0 L + K_Y}{\delta L}} \right) \quad (6)$$

5 Vergleich zwischen theoretischem Optimum und biologischer Evolution

Die Fragestellung, welche über der gesamten Arbeit steht, ist, ob E.coli seine LacZ Expression, in einem Nährmedium mit bestimmter Lactose Konzentration, auf das durch die Fitnessfunktion bestimmte Optimum einstellt.

5.1 Versuchsdurchführung

In dem entsprechenden Versuch wurden sieben Testreihen durchgeführt wobei E.coli in sieben verschiedenen Lactosekonzentrationen (0mM, 0.1mM, 0.2mM, 0.5mM, 1mM, 2mM, 5mM) gehalten wurde. Damit in allen Fällen das Lac Operon stets induziert wird, wurde den Testreihen mit 0mM und 0.1mM IPTG zugefügt. In allen anderen Testreihen ist die Lactosekonzentration hoch genug, um das Lac Operon zu induzieren. Alle 24 Stunden wurde bei jeder Testreihe eine Verdünnung im Verhältnis 1:100 durchgeführt. Alle 72 Stunden wurden aus jeder Testreihe Samples tiefgefroren. Die Samples wurden anschließend auf

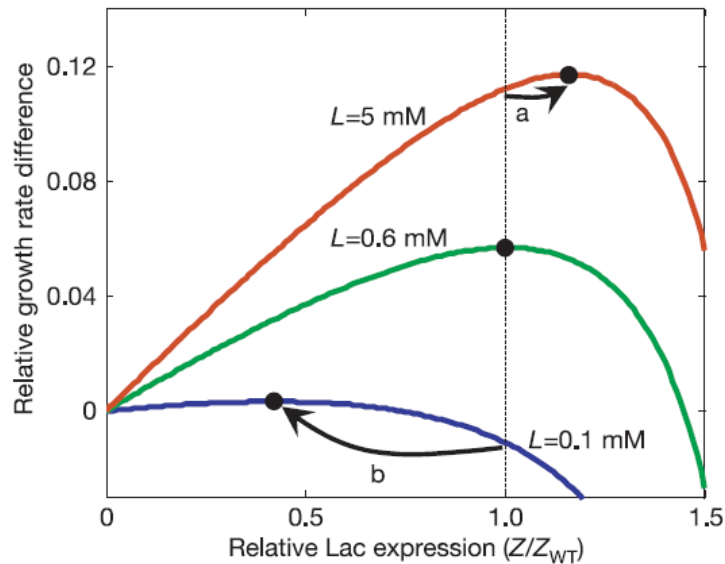


Abbildung 4: Verlauf der Fitnessfunktion (Funktion 5) in Abhängigkeit von der relativen Expression von LacZ (Z) für drei verschiedene externe Lactosekonzentrationen. Der Punkt kennzeichnet jeweils das Optimum. Bei höherer Lactosekonzentration erhöht sich auch der optimale Wert für Z . Die LacZ Expression beim Wildtyp (Z_{WT}) ist optimal für eine externe Lactosekonzentration von 0.6mM

die LacZ Enzymaktivität und die LacZ Konzentration untersucht. Die LacZ Enzymaktivität wurde mittels ONPG-Enzymassay bestimmt. Dabei macht man sich die Tatsache zu nutze, dass LacZ nicht spezifisch Lactose spaltet, sondern jegliche β -1-4-Verknüpfungen mit Galactose. Man nutzt deshalb Nitrophenyl- β -D-Galactosid (ONPG), welches an seiner β -1-4-Verknüpfung durch LacZ in D-Galactose und o-Nitrophenol gespalten wird. Die Konzentration an o-nitrophenol kann anschließend mittels optischer Verfahren bestimmt werden. Die Konzentration von LacZ wurde mittels quantitativer Gelelektrophorese bestimmt.

5.2 Ergebnisse

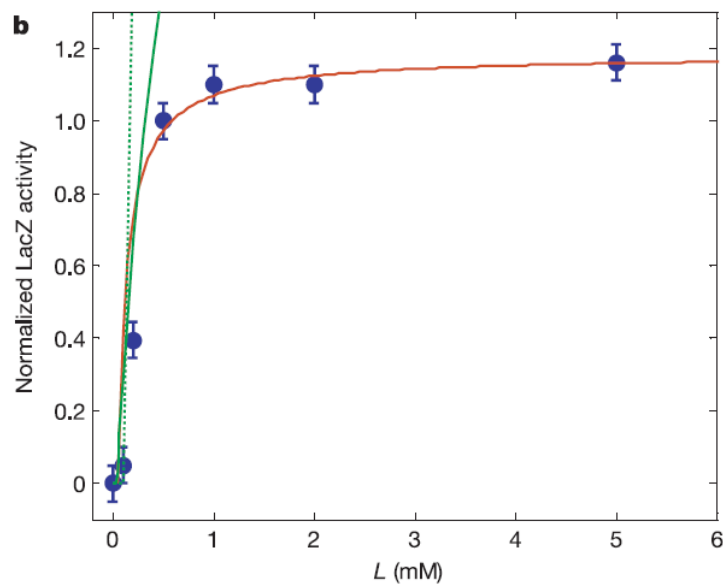


Abbildung 5: Vergleich, der nach Verdünnungsreihe gemessenen, LacZ-Aktivität für die sieben Testreihen mit der jeweiligen externen Lactosekonzentration (da alle LacZ Moleküle die gleiche Aktivität aufweisen, ist anzunehmen, dass die Änderung der Aktivität gleichbedeutend der Änderung der Expression ist). Die rote Linie kennzeichnet den Verlauf der Fitnessfunktion (Funktion 5) für die sieben Testreihen mit der jeweiligen externen Lactosekonzentration. Die grüne Linie kennzeichnet den Verlauf der Fitnessfunktion unter Verwendung der quadratischen Kostenfunktion (Funktion 1). Gestrichelte Linie zeigt Verlauf von Fitnessfunktion unter Verwendung einer linearen Kostenfunktion. Es ist gut erkennbar, dass die quadratische und die lineare Kostenfunktion die realen Kosten bei höheren Lactosekonzentrationen unterschätzen.

Mittels der Tests zur Aktivität und Konzentration von LacZ konnte festgestellt werden, dass die Aktivität jedes einzelnen LacZ Proteins unverändert geblieben ist. Änderungen der gesamten Enzymaktivität sind somit auf die Änderung der LacZ Expression zurückzuführen. Wie Abbildung 5 zeigt, stimmen die experimentell erfassten LacZ Aktivitäten und damit die LacZ Expression relativ genau mit den vorher mittels Fitnessfunktion bestimmten, optimalen Expressionslevel (rote Linie) überein. Demnach kann man für dieses kleine biologische System mit der gemessenen Kosten-Nutzen-Funktion, das durch Evo-

lution erreichte, LacZ Expressionslevel vorhersagen. In einem Zusatzexperiment wurde eine Populationsdynamik-Simulation durchgeführt. Bei dieser Simulation ist der einzige freie Parameter die Mutationswahrscheinlichkeit (p), mit der je Zelle und Generation die gewünschte Mutation auftritt. Wie in Abbildung 6 zu entnehmen ist, entspricht der Verlauf der Simulation wiederum relativ genau den gemessenen Werten. Aus dem Simulationsexperiment ergeben sich zwei Haupterkenntnisse. Zum einen bewegt sich der Wert für p im Bereich von 10^{-7} und ist damit um den Faktor 10^2 kleiner als die normale Mutationsrate in *E.coli*. Daher läßt sich vermuten, dass jeweils im Bereich 10^2 Basen existieren, welche bei einer Mutation zum entsprechenden Phänotyp, also zum entsprechenden LacZ Expressionslevel führen. Eine weitere Erkenntnis ist, dass die jeweiligen veränderten LacZ Expressionslevel mit hoher Wahrscheinlichkeit auf eine einzelne Mutation zurückzuführen ist, und nicht auf eine Kombination von Mutationen.

6 Zusammenfassung und Ausblick

In ihrem Versuch konnten Alon und Dekel für ein kleines Beispielsystem zeigen, dass Optimalsteuerung auf der Ebene von Proteinen existiert und in diesem kleinen Beispielsystem auch voraussagbar ist. Mit diesem Ergebnis legten sie eine Basis für weitere Untersuchungen auf Optimalitätssteuerung an größeren und komplexeren Systemen. Die Ergebnisse dieser Experimente können dann verwendet werden, um eine Theorie über optimales Genregulationsdesign auf-

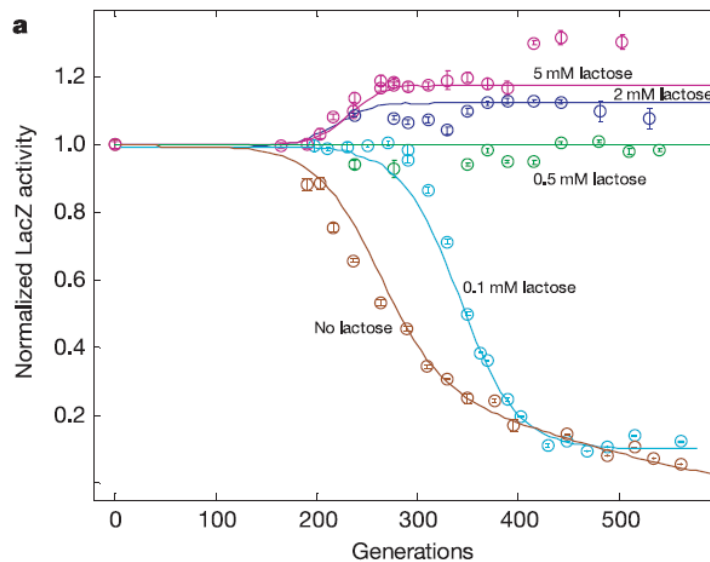


Abbildung 6: Vergleich der gemessenen LacZ Aktivität zu bestimmten Zeitpunkten während der Verdünnungsreihe mit dem Verlauf einer Populationsdynamik-Simulation. Einziger freier Parameter bei Simulation ist Mutationswahrscheinlichkeit (p). Ein erkennbarer Sprung bei der 5mM Testreihe nach etwa 400 Generationen wird auf Änderung in einem anderen metabolischen System zurückgeführt und läßt sich mathematisch mit einer Erhöhung der Kapazität (M) beschreiben.

zustellen und zu testen. Auf diesem Wege kann man auf eine Lösung, der noch immer sehr kontrovers geführten Diskussion, über den Begriff von Optimalität in Bezug auf Evolution und Genregulation hoffen.

Literatur

- [1] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436:588–592, 2005.