54

Forschungsberichte aus dem Max-Planck-Institut
für Dynamik komplexer technischer Systeme

Philipp Schneider

Advances in Computational Strain Design with Minimal Cut Sets

Advances in Computational Strain Design with Minimal Cut Sets

Dissertation
zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

von Philipp Schneider
geb. am 22.10.1991 in Mainz

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik der
Otto-von-Guericke-Universität Magdeburg

Promotionskommission:

Dr.-Ing. Udo Reichl (Vorsitz)

Dr.-Ing. Steffen Klamt (Gutachter)

Dr. Jürgen Zanghellini (Gutachter)

Dr.-Ing. Andreas Kremling (Gutachter)

eingereicht am: 30.09.2021
Promotionskolloquium am: 02.12.2021

Forschungsberichte aus dem Max-Planck-Institut
für Dynamik komplexer technischer Systeme

Band 54

Philipp Schneider

Advances in Computational Strain Design with Minimal Cut Sets

Shaker Verlag
Düren 2022

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Magdeburg, Univ., Diss., 2021

Copyright Shaker Verlag 2022
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-8411-5
ISSN 1439-4804

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

Modern genetic engineering tools allow the manipulation of metabolic networks for different purposes, ranging from the effective bioproduction of value-added chemicals with microorganisms to medical applications. This practice, also known as metabolic engineering, is often aided by computational design methods that predict suitable genetic intervention targets. The minimal cut set (MCS) method presents a powerful tool for the computer-aided design of genome-scale metabolic networks, and is based on mixed integer linear programming (MILP). An MCS is a minimal set of interventions that reshapes a wild type metabolism according to a template of undesired and desired metabolic functions. This work extends and generalizes the framework of MCS and presents new algorithmic and theoretical developments which largely broaden its spectrum of applications.

Many metabolic engineering approaches aim to enforce or improve the product synthesis of bioproduction hosts. A common goal is to couple the microbial growth to the synthesis of a product and make the product of interest a mandatory byproduct of growth. Although a multitude of strain design approaches employ this general principle, various and partially contradicting notions of the actual growth coupling relationship prevail. To systematize these notions, this work proposes four degrees of growth-coupled product synthesis with gradually increasing rigidity. It will be shown that an extended MCS algorithm allows the computation of strain designs for all four coupling degrees within one framework, thereby closing the gap between bilevel and MCS-based strain design approaches.

Beyond growth-coupled strain design, the MCS approach is extended by a number of functional generalizations, including a new compression routine for gene-protein-reaction associations, the support of reaction and gene additions, the possibility to specify multiple undesired and desired flux regions and the use of individual cost factors for different interventions. These generalizations allow, for example, the efficient computation of gene-based MCS and of strain designs that involve the expression of heterologous pathways or substrate co-feeding.

The MCS algorithm can be used to generate large pools of strain design candidates. To select a specific candidate for experimental implementation, one must weigh its individual strengths and weaknesses. The screening procedure presented herein uses ten criteria to assess the performance, robustness and experimental effort of a candidate and is not limited to MCS strain designs. Different intervention strategies with identical metabolic outcomes, e.g., by blocking the same pathway with the knockout of different enzymes, are gathered in equivalence classes of which only one candidate needs to be assessed. A ranked list of the candidates can be obtained by weighting the benchmarks from each criterion with suitable ranking factors. The ranking approach can be used to characterize reaction- and gene-based MCS and strain designs with different growth-coupling degrees.

All algorithmic developments of this work have been integrated into the MATLAB-based *CellNetAnalyzer* toolbox and were verified with various computation examples.

Acronyms

Notation	Description
1,4-BDO	1,4-butanediol
2,3-BDO	2,3-butanediol
ACP	ATP-coupled production
API	application programming interface
ATP	adenosine triphosphate
CBM	constrained-based modeling
CDW	cell dry weight
CNA	*CellNetAnalyzer*
dACP	directionally ATP-coupled production
dGCP	directionally growth-coupled production
DNF	disjunctive normal form
ED	Entner-Doudoroff pathway
EFM	elementary flux mode
EFV	elementary flux vector
FBA	flux balance analysis
FVA	flux variability analysis
GCP	growth-coupled production
GPR	gene-protein-reaction
GRPY	growth-rate-product-yield space
GUI	graphical user interface
IS	intervention strategy
LFP	linear-fractional program
LP	linear program
MBLP	mixed binary linear program
MCS	minimal cut set
MDF	max-min driving force
MILP	mixed integer linear program
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NGAM	non-growth-associated ATP maintenance
optMDF	optimal max-min driving force
PE	production envelope
pGCP	potentially growth-coupled production
SL	synthetic lethal
SUCP	substrate-uptake-coupled production
wGCP	weakly growth-coupled production
YS	yield space

Contents

1 Introduction **1**

2 Foundations of metabolic modeling and computational strain design **7**

2.1 Mathematical optimization 7

2.1.1 Linear programming (LP) 7

2.1.2 Mixed integer linear programming (MILP) 10

2.2 Constraint-based modeling 14

2.2.1 Modeling metabolic networks with linear (in)equalities 14

2.2.2 Basic constraint-based methods 15

2.3 Computational strain design 17

2.3.1 Bilevel-based methods 17

2.3.2 Minimal cut sets 18

2.4 Models 26

2.4.1 *i*JO1366 and *EColiCore2* 26

2.4.2 *i*ML1515 and *i*ML1515core 26

2.5 Software for constraint-based analysis and design 27

3 Systematizing the different notions of growth-coupled product synthesis and computing corresponding MCS strain designs **28**

3.1 Definition of four growth-coupling degrees 30

3.2 Computing minimal cut sets for all four coupling degrees 36

3.2.1 MCS setup for substrate-uptake-coupled production (SUCP) and directionally growth-coupled production (dGCP) 37

3.2.2 MCS setup for weakly growth-coupled production (wGCP) 39

3.2.3 MCS setup for potentially growth-coupled production (pGCP) and OptKnock-like strain designs 42

3.3 Comparing strain designs for the growth-coupled production of ethanol at all coupling degrees 46

3.4 Genome-scale calculations for selected products 48

3.5 Generalization of coupled production: from growth-coupled to ATP-coupled production 52

3.6 Discussion 55

4 Extensions and generalizations of the MCS framework **57**

4.1 Multiple desired or undesired flux spaces 57

4.2 Weighting interventions with cost factors 60

4.3 Reaction additions by inverting the knockout-logic 61
4.4 Combining reaction deletions and additions to find substrate co-feeding strategies 61
4.5 Compressing GPR rules for the efficient computation of MCSs with genetic interventions 65
4.6 Example strain design for the production of 2,3-butanediol using substrate co-feeding 72
4.6.1 Effect of GPR rule compression on the MCS computation performance 72
4.6.2 Applying the new MCS features for strain design 75
4.6.3 Key principles of the found intervention strategies for 2,3-BDO synthesis 77
4.6.4 MCSs for growth-coupled 2,3-BDO synthesis in two other genome-scale models 78
4.7 Discussion 78

5 Characterizing and ranking computed metabolic engineering strategies 82
5.1 Characterization of metabolic engineering strategies by 10 different properties 82
5.2 Scoring and ranking 89
5.3 Examples for the characterization and ranking of computed metabolic engineering strategies 90
5.3.1 Example 1: production of L-methionine 91
5.3.2 Example 2: production of the heterologous product 1,4-butanediol . . 95
5.4 Discussion 96

6 Implementation in *CellNetAnalyzer* 99
6.1 *CellNetAnalyzer* API functions 99
6.1.1 Overview 99
6.1.2 *CNAMCSEnumerator2* 101
6.1.3 *CNAgeneMCSEnumerator2* and *CNAintegrateGPRrules* 102
6.1.4 *CNAMCSEnumerator3* and *CNAgeneMCSEnumerator3* 103
6.1.5 *CNAcharacterizeIS* 108
6.1.6 *CNAcharacterizeGeneMCS* 109
6.2 GUI-based MCS computation 110
6.3 Code availability and compatibility 111

7 Discussion and outlook 112

Appendix 119

Bibliography 122