Optimal experimental design for grey-box models

Ph.D. Thesis

Florin Paul Davidescu
CAPEC
Department of Chemical Engineering
Technical University of Denmark

January 14, 2009

Copyright © Florin Paul Davidescu
CAPEC
Department of Chemical Engineering
Technical University of Denmark, 2009
ISBN 87-91435-93-5
Printed by Book Partner, Nørhaven Digital, Copenhagen, Denmark

Preface

This thesis is submitted as partial fulfillment of the requirements for the Ph.D. degree at Technical University of Denmark. The work has been carried out at the Computer Aided Process Engineering Center research center (CAPEC), at Department of Chemical Engineering at Technical University of Denmark between January 1st 2005 and 31st of December 2007. The project has been funded by the European Union by the EUROBIOSYN project under the 6th Research Framework Program.

I will take the opportunity to acknowledge the persons that directly or indirectly have contributed to the work presented in this thesis.

First of all I would like to express my gratitude to the main supervisor, Professor Sten Bay Jørgensen from the Department of Chemical Engineering for the many discussions that we have had during this project. I would like to thank my co-supervisor, Professor Henrik Madsen from the Department of Informatics and Mathematical Modelling for his inputs all long the way.

Secondly, I would like to express my gratitude to the European Union for funding this project through the EUROBIOSYN project under the 6th Research Framework Program. I would like to give thanks to all the colleagues from all the four research centers involved in this project and especially Mr. Michael Schümperli for carrying out the necessary laboratory experiments.

Thirdly, I would like to thank all the CAPEC co-workers and especially to Loic, Mauricio, Vipasha, Piotr, Jacob, Jan and Steen for creating a good social environment at the CAPEC and during the conferences. I'm sorry if I have not included all of you here.

Fourthly, I would like to thank the Department of Chemical Engineering at the Technical University of Denmark for giving me the opportunity to do the study related to this PhD project. Furthermore I would like to acknowledge the computing services kindly offered by Danish High Performance Computing Center at the Department of Informatics and Mathematical Modelling. Moreover, I would like to thank the workshop for their help and assistance even though their input did not enter directly into the work presented here but in my teaching assistance activities around the pilot plant of the department.

Finally I would like to thank my parents and my brothers for their love and support during this project while being away from them and for trying to understand my work.

København,

Florin Paul Davidescu

Abstract

In the recent years, chemical and biochemical process modelling has dominated the research and development. There are various targets of the developed models such as steady-state and dynamic simulation in order to better investigate and understand the process, process design, process control, process optimization and even more recently fault monitoring and diagnosis.

In this thesis the main reason to develop process models is to first better understand the behavior of the enzymatic reaction network and to possibly identify the bottle-necks and limitations. The secondary objective and equally important was and still is to optimize the operating conditions of the system of biotransformations (SBT) in order to maximize the yield. Moreover in order to move to the production on a larger scale the reactor design can become an important target.

There is a need however for different tools that the end-user needs in order to advance faster in this very time consuming task of building dynamic process models.

The contribution of this thesis is therefore twofold at least. The first way is toward the model development for the enzymatic reaction network in form of stochastic differential models. The second way is toward the necessary systematic methodology and tools needed at various steps in this process.

One important step is assessing the possibility of estimating the model parameters from real life data once the a model has been formulated. In addressing this issues a systematic methodology has been set up. Given a model structure, a set of measured states and perturbed inputs this methodology provides the parameters which can be at least theoretically estimated.

Another important step is represented by qualitative experimental design which aims at determining the optimal set of measured states and perturbed inputs. By optimal we mean either the case of minimum number of measured states and perturbed inputs rendering all model parameters estimable or at least as many parameters as possible.

A related step is designing the experimental conditions, sometimes called quantitative experimental design, for the optimally selected set of measured states and perturbed inputs. Typically, the initial values of measured states, the optimal sampling points, the input profiles are searched for, within a dynamic optimization problem framework formulation.

In terms of software, a computer program to determine the optimal quantitative experimental design for models described by stochastic differential equations has been developed. The program called Continuous Time Stochastic Modelling (CTSM) Kristensen et al. (2004b) previously developed, has been

further extended to be more flexible in handling experimental data obtained under different conditions and in estimating each of the the parameters either globally for all the experiments or locally for each experiment, allowing varying parameters.

Resumé på dansk

I de seneste år har kemisk og biokemisk procesmodellering domineret forskning og udvikling. Der er forskellige formål med de udviklede modeller såsom steady-state og dynamisk simulering af en given proces for bedre at kunne undersøge og forstå processen, procesdesign, procesregulering, procesoptimering samt overvågning og diagnose af procesfejl.

I dette arbejde har det primære formål med udviklingen af procesmodeller været at øge forståelsen af det enzymatiske reaktionsnetværk og om muligt identificere flaskehalse og begrænsninger. Det sekundære formål af lige så stor vigtighed har været og er stadig at optimere procesbetingelserne for systemet af biotransformationer (SBT) for at optimere udbyttet. Derudover, kan reaktordesign blive en vigtig applikation af modellen ved opskalering af processen.

Der er et behov for forskellige værktøjer, som kan hjælpe brugeren til hurtigt at nå fremskridt under den uhyre tidskrævende opgave det er at opbygge dynamiske procesmodeller.

Bidraget i dette arbejde er derfor mindst tofoldigt. Dels i retning af model udvikling for det enzymatiske reaktionsnetværk i form af stokastiske differentialligninger. Dels i retning af en nødvendig systematisk metodologi og værktøjer, der er behov for under de forskellige trin i processen.

Et vigtigt trin er at vurdere muligheden af at estimere model parametre fra virkelige forsøg når en model er blevet formuleret. For at håndtere dette er en systematisk metodologi blevet sat op. Givet en modelstruktur, et sæt af målte tilstande og pertuberede inputs giver denne metodologi de parametre, som i det mindste teoretisk kan blive estimeret.

Et andet vigtigt trin og intet andet end et udvidet identificerbarhedsproblem er repræsenteret ved kvalitative eksperimentelle designs, som søger at bestemme det optimale sæt af målte tilstande og pertuberede inputs. Med optimale menes enten det tilfælde hvor alle eller så mange parametre som muligt kan blive estimeret ud fra det mindste mulige antal målte tilstande og pertuberede inputs.

Et beslægtet trin omhandler design af de eksperimentelle betingelser, ofte benævnt kvantitativt eksperimentelt design, for det valgte sæt af målte tilstande og pertuberede inputs. Typisk søges der efter begyndelsesværdierne af de målte tilstande, de optimale tidspunkter for prøveudtagning samt inputprofilerne indenfor en dynamisk optimeringsproblem framework formulation.

Vedrørende software er et computer program til bestemmelse af det optimale kvantitative eksperimentelle design for modeller beskrevet ved stokastiske differentialligninger blevet udviklet. Dette tidligere udviklede (Kristensen et al. 2004b) program, Continuous Time Stocastic Modelling (CTSM) er blevet

videre udviklet til at være mere fleksibelt ved håndteringen af eksperimentelle data erhvervet under forskellige betingelser og i estimeringen af parametre enten globalt for alle eksperimenter eller lokalt hvilket tillader varierende parametre.

Contents

Preface **iii**

Abstract **v**

Resumé på dansk **vii**

1 Introduction **9**
1.1 Modelling issues 9
1.2 Motivation 10
1.2.1 Modelling purpose 11
1.2.2 Grey-box stochastic modelling of SBT 11
1.2.3 Identifiability analysis and qualitative experimental design 12
1.2.4 Quantitative experimental design 13
1.3 Hypotheses of the thesis 15
1.4 Structure of the thesis 15
1.5 Publications 16
1.5.1 Reviewed journal papers 16
1.5.2 Reviewed conference papers 16
1.5.3 Oral presentations and posters 17

2 Reaction Network Modelling within Grey Box Framework **19**
2.1 Introduction 19
2.1.1 Why using stochastic differential equations? 20
2.1.2 Review of the existing grey-box stochastic modelling framework 21
2.2 Methods and Software 22
2.2.1 Model structure of SDE 22
2.2.2 Parameter estimation 23
2.2.3 Statistical tests 24
2.2.4 Software 25
2.3 Experimental data 26
2.4 Model I for an enzymatic reaction network 27
2.4.1 Model I equations and validation 29
2.4.2 Discussion of Model I 37
2.5 Model II, Improved kinetics 37
2.5.1 Validation of Model II 45
2.5.2 Multiple sets of experimental data 47
2.6 Discussion on Model II 49
2.7 Conclusions 50

3 Structural parameter identifiability analysis for dynamic reaction networks 51
3.1 Introduction . . . 51
3.2 Methodology . . . 54
3.2.1 Reaction rate identifiability . . . 54
3.2.2 Structural Parameter Identifiability Analysis . . . 55
3.2.3 Structural identifiability algorithm . . . 57
3.3 Example Process . . . 59
3.3.1 Model formulation . . . 60
3.3.2 Structural identifiability analysis for the enzymatic reaction network (SBT) . . . 62
3.4 Validation of the algorithm . . . 67
3.4.1 Combination r_1, r_3, r_6, r_7 . . . 69
3.4.2 Combination r_1, r_2, r_5, r_7 . . . 69
3.4.3 Combination r_1, r_2, r_3, r_7 . . . 70
3.5 Discussion . . . 71
3.6 Conclusions . . . 72

4 Qualitative experimental design for nonlinear dynamic reaction networks 73
4.1 Introduction . . . 73
4.2 Methodology . . . 74
4.3 Workflow of structural experimental design . . . 75
4.4 Example Process . . . 81
4.4.1 Model formulation . . . 81
4.5 Qualitative experimental design for the enzymatic reaction network (SBT) . . . 84
4.5.1 Combination r_1, r_5, r_6, r_7 . . . 85
4.5.2 Combination r_1, r_3, r_6, r_7 . . . 85
4.5.3 Combination r_1, r_2, r_5, r_7 . . . 86
4.5.4 Combination r_1, r_2, r_3, r_7 . . . 86
4.5.5 One extra measured state . . . 86
4.5.6 Glucose feed input and five measured states . . . 87
4.6 Considering a sixth measured state . . . 88
4.6.1 Lactate as sixth measured state . . . 88
4.6.2 Six measured states and the Glucose input . . . 89
4.6.3 G3P as sixth measured state . . . 89
4.6.4 Six measured states and the Glucose input . . . 89
4.7 Seven measured states . . . 90
4.7.1 PO4 included as measured state . . . 90
4.7.2 PO4 included and the Glucose input . . . 90
4.7.3 F16B instead of PO4 . . . 90
4.7.4 F16B instead of PO4 and Glucose input . . . 91
4.8 Discussion . . . 91
4.9 Conclusions and Future work . . . 92

5 Quantitative experimental design for parameter estimation 95
5.1 Introduction . . . 95
5.2 Methodology . . . 97
5.2.1 Optimization . . . 99
5.2.2 Evaluating the design objective . . . 101
5.2.3 Model structure of SDE . . . 103
5.2.4 Conditional density distribution function . . . 104
5.3 The algorithm . . . 105
5.4 Benchmark case . . . 109
5.4.1 Results . . . 109
5.5 Discussion . . . 114
5.6 Conclusion . . . 114

6 An improved grey-box stochastic modelling framework 117
6.1 Introduction . . . 117
6.2 An improved framework for grey-box model development . . . 118
6.3 Model III for a simplified SBT . . . 119
6.4 Decoupled qualitative experimental design . . . 122
6.4.1 Identifiable rates . . . 123
6.4.2 Identifiable parameters . . . 124
6.5 Parameter estimation using data from Exp 30 A . . . 127
6.6 Quantitative experimental design for Model III of SBT . . . 131
6.6.1 Evaluating the initial experimental data . . . 131
6.6.2 Optimal experimental design . . . 132
6.7 Conclusion . . . 135

7 Software issues 137
7.1 CTSM software . . . 137
7.2 EXPDSGN . . . 137
7.3 Maple code . . . 138

8 Conclusions and Future work 139
8.1 General conclusions on hypotheses . . . 139
8.2 New contributions . . . 140
8.3 Suggestions for future work . . . 141
8.3.1 Further modelling . . . 141
8.3.2 Parameter independent experimental design . . . 141
8.3.3 Experimental design for model discrimination . . . 142

A A very detailed kinetic model 143

B Experimental data 147
B.1 Experimental procedure . . . 148
B.2 Experiment 1 . . . 148
B.3 Experiment 2 . . . 151
B.4 Experiment 3 . . . 154

B.5 Experiment 4 . . . 157
B.6 Experiment 5 . . . 160
B.7 Experiment 6 . . . 163
B.8 Experiment 7 . . . 164
B.9 Experiment 8 . . . 167
B.10 Experiment 9 . . . 168
B.11 Experiment 10 . . . 171
B.12 Experiment 11 . . . 174
B.13 Experiment 12 . . . 177
B.14 Experiment 17 . . . 184
B.15 Experiment 20 . . . 191
B.16 Experiment 23 . . . 194
B.17 Experiment 24 . . . 196
B.18 Experiment 26 . . . 202
B.19 Experiment 27 . . . 205
B.20 Experiment 28 . . . 210
B.21 Experiment 30 . . . 215
B.22 Data analysis . . . 222
B.22.1 Duplicate experiments . . . 222
B.22.2 Hexokinase influence . . . 222
B.22.3 Temperature influence . . . 222
B.22.4 Protein concetration . . . 223
B.22.5 ATP influence . . . 223
B.22.6 NAD+ influence . . . 223
B.22.7 DHAP-aldolase reaction . . . 223
B.22.8 Glucose influence . . . 224

References **229**

List of Figures

1.1 Existing grey-box modelling framerok Kristensen et al. (2004b) 10

2.1 Reaction network used for Model I development with the measured species indicated by a $*$ 28
2.2 Simulation of Model I against Experiment 17A, set 1; pure simulation-red line, one step ahead prediction-blue line 36
2.3 Residuals of Model I simulation for Experiment 17A, set 1 . . . 36
2.4 The simplified reaction network for Model II including the discussed additions 38
2.5 Simulation of Model II against Experiment 17A set 1; pure simulation-red line, one step ahead prediction-blue line 45
2.6 Residuals for Model II simulation and Experiment 17A set 1 . . 46
2.7 Validation of Model II using data from Experiment 17B set 1; pure simulation-red line, one step ahead prediction-blue line . . 46
2.8 Simulation of Model II against data from Experiment 17A set 1; pure simulation-red line, one step ahead prediction-blue line . . 48
2.9 Simulation of Model II against data from Experiment 17B set 1; pure simulation-red line, one step ahead prediction-blue line . . 49

3.1 Work-flow of the structural identifiability analysis 58
3.2 The simplified reaction network investigated in this analysis . . 61

4.1 Workflow of qualitative experimental design methodology I . . 77
4.2 Workflow of qualitative experimental design methodology II . . 78
4.3 Workflow of qualitative experimental design methodology III . 79
4.4 The simplified reaction network used in this analysis 82

5.1 Proposed work-flow for the optimization algorithm used for quantitative experimental design for systems modelled by stochastic differential equations 106
5.2 Proposed work-flow for evaluation of the experimental design objective function in *step 2* of quantitative experimental design (Fig. 5.1) for SDE based models 107
5.3 Experimental design - initial design 111
5.4 Initial experimental design mean values 111
5.5 Progression of the convergence using $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$ 112
5.6 Experimental design - final input design using $G = 50000$, $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$ 113

5.7 Experimental design - final output design using $G = 50000$, $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$ 113

6.1 Grey-box stochastic modelling framework extended with identifiability analysis 119
6.2 Grey-box stochastic modelling framework extended with qualitative and quantitative experimental design 119
6.3 The simplified reaction network corresponding to Model III . . 120
6.4 Simulation of Model III against data from Experiment 30A, reactor II, set 1; pure simulation-red, one step prediction-blue. Data from Experiment 30A, reactor I, set 1 and 2 used for estimation 129
6.5 Progression of the convergence using $NP = 5$, $N = 100$ and $F = 0.79$, $CR = 0.8$ 130
6.6 Data collected in Experiment 30A, set 1 and 2, reactor I and II 132
6.7 Simulated profiles for optimal experimental design for Model III, obj func =-9.66E+01 after 1900 generations 134
6.8 Progression of the convergence using $NP = 20$, $N = 100$ and $F = 0.9$, $CR = 0.9$ 135

A.1 The simplified reaction network used in this analysis 144

List of Tables

2.1 Conditions defining the experiment 17A 27
2.2 Log-likehood function value for the estimation results for Model I when estimating different number of parameters, using data from Experiment 17A, set 1 and 2, reactor I 32
2.3 Loglikelihood ratio test applied for estimation results for Model I using data from Experiment 17A, set 1 and 2, reactor I 33
2.4 Estimation results for Model I, *set 2* from Table 2.2 33
2.5 Fixed parameters values for Model I 33
2.6 Estimation results for Model I using data from Experiment 17A, set 1 and 2, reactor I 34
2.7 Correlation matrix of the estimates for Model I using data from Experiment 17A, set 1 and 2, reactor I 35
2.8 Log-likelihood function value for the estimation results for Model II when estimating different number of parameters, using data from Experiment 17A, data set 1 and 2, reactor I 42
2.9 Loglikelihood ratio test applied for estimation results of Model II using data from Experiment 17A, set 1 and 2, reactor I . . . 42
2.10 Estimation Results for Model II, using data from Experiment 17A, data set 1 and 2, reactor I 42
2.11 Estimation results for *set 7* from Table 2.9 part 1 43
2.12 Fixed parameters value for Model II 43
2.13 Correlation matrix for *set 7* from Table 2.9 44
2.14 Estimation results for Model II using data from Experiment 17A set 1 and 2, and Experiment 17B set 1 and 2 47
2.15 Estimation results for Model II using data from Experiment 17A set 1 and 2, and Experiment 17B set 1 and 2 48

3.1 Full stoichiometric matrix N for the reaction network 63
3.2 Stoichiometric matrix N_m for the measured species 63
3.3 Stoichiometric matrix N_u, for the unmeasured species 63
3.4 Identifiable combinations of reaction rates for the enzymatic reaction network 64
3.5 Reduced stoichiometric matrix — first set 64
3.6 Reduced stoichiometric matrix — second set 64
3.7 Reduced stoichiometric matrix — third set 65
3.8 Reduced stoichiometric matrix — fourth set 65
3.9 One combination of Lie derivatives for which a unique solution was found 66

3.10 One combination of first and second order Lie derivatives for which a unique solution was found 67
3.11 One combination of first, second, and third Lie derivatives for which a unique solution was found 67
3.12 Combination of Lie derivatives for which a unique solution was found 68
3.13 Initial conditions for simulation 68
3.14 Estimation results 69
3.15 Correlation matrix 69
3.16 Estimation results 69
3.17 Correlation matrix 70
3.18 Estimation results 70
3.19 Estimation results 70
3.20 Estimation results 70
3.21 Estimation results 71

4.1 Full stoichiometric matrix N for the reaction network 84
4.2 Stoichiometric matrix N_m for the four originally measured species 84
4.3 Stoichiometric matrix N_u, for the unmeasured species 85
4.4 Identifiable combinations of reaction rates for the enzymatic reaction network with the four original measurements 85
4.5 One combination of Lie derivatives for which a unique solution was found with the four original measurements 85
4.6 One combination of Lie derivative for which a unique solution was found 86
4.7 One combinations of Lie derivatives for which a unique solution was found 86
4.8 One combination of Lie derivatives for which a unique solution was found 86
4.9 Combinations of species rendering the five reaction rates identifiable 87
4.10 One combination of parameters for which a unique solution was found 87
4.11 One combination of Lie derivative for which a unique solution was found for ten parameters when perturbing Glucose feed with five measurements 88
4.12 Combinations of species rendering six reaction rates identifiable with six measurements 88
4.13 One combination of ten parameters for which a unique solution was found with six measurements 88
4.14 A combination of eleven parameters for which a unique solution was found with six measured states 89
4.15 Parameter sets for which a unique solution was found with six measured states and an input perturbation 89

4.16 Combinations of species rendering the seven reaction rates identifiable 90

5.1 Initial experimental conditions used by Asprey and Macchietto (2002) 110
5.2 Experimental conditions obtained 112

6.1 Combinations of measured species rendering all the nine reaction rates identifiable 124
6.2 Correspondence between the reaction rates and the measured species 124
6.3 Analysis for subsets of two parameters taken simultaneously . . 124
6.4 Analysis for subsets of two parameters 125
6.5 Analysis for subsets of one parameter 125
6.6 Analysis for subsets of two parameters taken simulateously . . 125
6.7 Analysis for subsets of one parameter 125
6.8 Analysis for subsets of one parameter 126
6.9 Analysis for subsets of two parameters 126
6.10 Analysis for subsets of two parameters taken simultaneously . . 126
6.11 Analysis for subsets of one parameter 126
6.12 One parameter vector set considered qualitatively identifiable from the measurement set GL, G6P, DHAP, ATP, ADP, AMP, PYR 126
6.13 Fixed parameters value for Model III 127
6.14 Estimation results for Model III, using data from Experiment 30A, reactor I set 1 and 2 128
6.15 Experimental conditions used in Experiment 30A 131
6.16 Computed mean variance for data collected in Experiment 30A, set 1 and 2, reactor I and II 133
6.17 Optimized experimental conditions obtained after 1900 generations 133

B.1 Experiment 1 description 149
B.2 Experiment 1A 149
B.3 Experiment 1B 149
B.4 Experiment 2 description 151
B.5 Experiment 2A, 20 [mg/ml] 151
B.6 Experiment 2B, 10 [mg/ml] 152
B.7 Experiment 2C, [2 mg/ml] 152
B.8 Experiment 3 description 154
B.9 Experiment 3A, 37 C 154
B.10 Experiment 3B, 30 C 155
B.11 Experiment 3C, 25 C 155
B.12 Experiment 4 description 157
B.13 Experiment 4A, 11.5 mM ATP 157

B.14 Experiment 4B, 5.75 mM ATP 158
B.15 Experiment 4C, 1.15 ATP 158
B.16 Experiment 5 description 160
B.17 Experiment 5A, 11.5 mM NAD 160
B.18 Experiment 5B, 5.75 mM NAD 161
B.19 Experiment 5C, 1.15 mM NAD 161
B.20 Experiment 6 description 163
B.21 Experiment 6A, 44 mM GL 163
B.22 Experiment 7 description 164
B.23 Experiment 7A, nonwashed extract 164
B.24 Experiment 7B, ultrafiltered extract 165
B.25 Experiment 7C, washed extract 165
B.26 Experiment 8 description 167
B.27 Experiment 8A, double mutant 167
B.28 Experiment 9 description 168
B.29 Experiment 9A, double mutant 168
B.30 Experiment 9B, double mutant 169
B.31 Experiment 10 description 171
B.32 Experiment 10A, 11.5 mM ATP, Reactor I 171
B.33 Experiment 10B, 5.75 mM ATP, Reactor II 173
B.34 Experiment 11 description 174
B.35 Experiment 11A, 11.5 mM ATP, reactor I 174
B.36 Experiment 11B, 5.75 mM ATP, reactor II 175
B.37 Experiment 12 description 177
B.38 Experiment 12A, unwashed extract, Reactor I 178
B.39 Experiment 12B, washed extract I, Reactor II 178
B.40 Experiment 12A, washed extract II ,Reactor III 179
B.41 Experiment 12C, filtered extract I, Reactor IV 179
B.42 Experiment 12C, filtered extract II, Reactor V 180
B.43 Experiment 17 description 184
B.44 Experiment 17A, Stirred double-walled beakers. Reactor I, 11.5 mM ATP 184
B.45 Experiment 17B. Stirred double-walled beakers. Reactor II, 5.75 mM ATP 185
B.46 Experiment 17B. Stirred double-walled beakers. Reactor III, 5.75 mM ATP 185
B.47 Experiment 17C. Stirred double-walled beakers.Reactor IV, 1.15 mM ATP 186
B.48 Experiment 17C. Stirred double-walled beakers. Reactor V, 1.15 mM ATP 186
B.49 Experiment 20 description 191
B.50 Experiment 20A, reactor I 191
B.51 Experiment 20A, reactor II 193
B.52 Experiment 23 description 194
B.53 Experiment 23A, reactor I 194

B.54 Experiment 23A, reactor II 195
B.55 Experiment 24 description 196
B.56 Experiment 24A, 11.5 mM ATP, reactor I 196
B.57 Experiment 24A, 11.5 mM ATP, reactor II 197
B.58 Experiment 24B, 5.75 mM, reactor III 197
B.59 Experiment 24B, 5.75 mM, reactor IV 197
B.60 Experiment 24C, 1.15 mM, reactor V 201
B.61 Experiment 26 description 202
B.62 Experiment 26A, 11,5 mM, reactor I 202
B.63 Experiment 26A, 11,5 mM, reactor II 204
B.64 Experiment 27 description 205
B.65 Experiment 27A, 11,5 mM ATP, reactor I 206
B.66 Experiment 27A, 11,5 mM ATP, reactor II 207
B.67 Experiment 28 description 210
B.68 Experiment 28A, 11,5 mM ATP, reactor I 211
B.69 Experiment 28A, 11,5 mM ATP, reactor II 212
B.70 Experiment 30 description 215
B.71 Experiment 30A, GLC pulse, reactor I 216
B.72 Experiment 30A, GLC pulse, reactor II 217
B.73 Experiment 30B, 1.15 mM ATP, reactor I 220
B.74 Experiment 30B, 1.15 mM ATP, reactor II 220

1

Introduction

The purpose of this chapter is to motivate, to define the hypotheses and the objectives of this thesis. Moreover, a brief overview of the state of the art for each of the main areas dealt with in the following chapters is given. Each chapter contains a section, where the state of the art, is review and presented in more detail. A section describing the organization of the thesis closes this chapter.

1.1 Modelling issues

Over the last years, a significant effort has been put into modelling the various process that both the industry and academia has to deal with. The effort meaning both time and money. The model application drives usually the model development, thus any methods and tools that can facilitate and ease the model development is highly relevant.

As far as the applications are concerned there are various reasons to build and develop process models. One basic and fundamental application is understanding the process under investigation, the developed model helping the investigator by process simulation. A different application is process design and optimization. In many cases e.g. for reactor or bio-reactor design a good kinetic model is required. Mathematical optimization uses the model as its workhorse besides the optimization routine. Yet another application is process control where a model is used to compute the optimal input trajectories.

In the description above the focus was on general dynamic process models but in this thesis the focus is on dynamic process models described by stochastic differential equations or grey-box models. The scope of this thesis is thus on improving grey-box stochastic models by introducing parameter identifiability analysis, qualitative and quantitative experimental design. The main application for these tools is represented by an enzymatic reaction network denoted in this thesis a system of bio-transformation (SBT). The starting point in this work is represented by an existing grey-box modelling framework proposed by (Kristensen et al. 2004b). A graphical description of this modelling framework is given below in Figure 1.1.

This methodology focused upon utilizing existing experimental data. Moreover the assumption is that the data are sufficient from the statistical point of

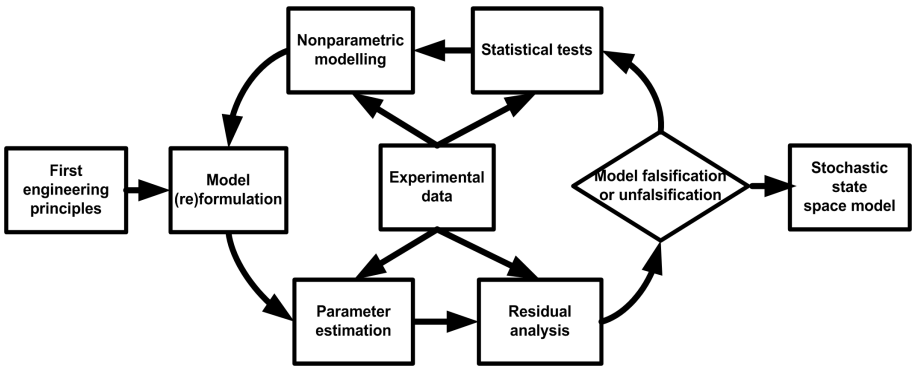


Figure 1.1: Existing grey-box modelling framerok Kristensen et al. (2004b)

view such that it is possible to recover the states and model parameters appropriately. In this situation it is possible to identify the model deficiencies by assessing the significance of the diffusion terms which represent a key feature of the available methodology Kristensen et al. (2004b).

Another key feature of the existing grey-box stochastic modelling framework is that by applying state estimation and nonparametric modelling it is possible to reveal how to model the deficient parts of the model, i.e. the kinetic rates for the enzymatic reaction network.

However if the data are insufficient then it is difficult to estimate the model parameters and thus it is not really clear which of the diffusion terms are significant. Secondly, when performing state estimation and nonparametric modelling if the data series are not sufficient, then the obtained dependencies between various process variables may be biased and thus the model correction may not go in the right direction.

In the present thesis the focus is therefore on designing experiments to better estimate the model parameters and the diffusion terms and thereby ease the identification of model deficiencies. Moreover the state estimation used for nonparametric modeling should be more accurate thus facilitating model reformulation and correction.

1.2 Motivation

This section gives an overview of the state of the art for the various fields covered in this thesis and gives the motivation for the work performed during this PhD study. There are several different points to be discussed. The first point to be discussed is modelling of the SBT described above. The second point is that of using grey-box models to model this system. The third point regards performing identifiability analysis and qualitative experimental design. Finally, quantitative experimental design for grey-box models will be discussed.

1.2.1 Modelling purpose

This enzymatic reaction network renders a very complex system, partly known through measurements with a series of feed-back and feed-forward mechanisms. In order to be able to run this SBT in an optimal way with respect to productivity of the desired compound, DHAP, it is necessary to optimize the design and the operation of network. Several means are used in practice to achieve this goal. One way regards the identification of the possible bottlenecks and their removal when possible. This can be done using reaction network analysis, knowledge about the enzymes and their manipulation, i.e. saturation constants which can be modified by manipulating the proteins and their corresponding genes and finally the operation may be optimized based on simulation studies of a model with good prediction properties developed for this enzymatic reaction network.

Another way used, is by performing a schedule of experiments organized based on a trial and error and heuristics approach. These experiments may be designed based again on a model of the process. There are at least two types of model based experimental design. The first type regards collecting information about estimation of the model parameter. The second type regards collecting data in order to differentiate between few potential candidates models. These cases all requires a fairly good model of SBT.

1.2.2 Grey-box stochastic modelling of SBT

Modelling chemical and biochemical processes the most common way is the first principle engineering modelling, sometimes called white box modelling. The main idea is to start with the general principles of mass, energy and momentum conservation and derive the model equations. However, more knowledge is required in terms of transfer phenomena parameters i.e. heat, mass transfer coefficients, diffusion coefficients or when it comes to reaction engineering various kinetic equations and parameters. Most often this information is partly known or only an educated guess of the reaction kinetics in terms of structure and parameters values is available. Even so, when trying to validate the developed model for the particular process under investigation against experimental data, it is assumed that the model structure is correct. Thus, by estimating the model parameters, the total error will be absorbed depending on the estimation scheme and on the way the measurement errors is considered.

Another situation when developing process models is the case when nearly nothing is known about the whole process or just part of it. In this situation, black box modelling may be used. In black box modelling various polynomial models linking the input and the output variables are proposed and the model polynomials are estimated using a standard estimation scheme as output error or prediction error method. The performance of these models is best for short time horizon but depends upon estimation method and weightings used.

In order to utilize both the knowledge available and the experimental data

for model development, grey-box modelling has been proposed. The workhorse of these models is represented by stochastic differential equations. The aim is to use all the knowledge available from first principle engineering and simultaneously account for model uncertainties and deficiencies, input uncertainties, by assuming stochastic evolution of the states. Thus the models are developed as first engineering principles models incorporating stochastic terms denoted as diffusion terms. Grey box modelling may be viewed as founded on fundamental behavior of molecules but here applied on macroscopic systems.

The particular feature, of combining the available knowledge with stochastic variables motivates the use of stochastic differential equation for modelling. The framework developed by Kristensen et al. (2004a), describes a systematic methodology for improving this kind of models. This framework will be employed and further developed here.

1.2.3 Identifiability analysis and qualitative experimental design

In order to evaluate the model performance it is necessary to estimate the model parameters. However estimation of model parameters may be difficult if data is not informative and impossible if the parameters are not identifiable. Consequently, analysis of the properties of the parameter estimation problem may be highly beneficial prior to both experimental design and to undertaking actual estimation of model parameters in practice. Hence it is desirable to evaluate whether the parameters of a given model are identifiable and to perform experimental design aimed at developing most beneficial conditions for identifying unknown model parameters.

Experimental design may be decomposed into two steps: qualitative experimental design and quantitative experimental design. Qualitative experimental design concerns first selection of the operation mode, secondly, which input variables should be varied, and thirdly, which outputs should be measured in the experiments in order to render the unknown model parameters identifiable. The first author to define the concept of qualitative and quantitative experimental design and their relationship with identifiability analysis, mainly working with linear or linearized models, seems to be Walter and Pronzato (1987), Vajda et al. (1989) and Walter and Pronzato (1990).

One approach to address the structural identifiability problem is based on state isomorphisms. If indistinguishable state space systems exist then one can try to parameterize the equivalence classes by the admissible state space isomorphisms. And if one can show that there is an identity, then global identifiability follows under certain conditions as outlined by Peeters and Hanzon (2005).

A second approach to structural identifiability analysis is based on differential algebra proposed by Ljung and Glad (1994), which has been further developed by Audoly et al. (2001), Saccomani et al. (2003) and Saccomani (2004). Here a set of algebraic equations denoted the exhaustive summary is obtained and

this set is solved by algebraic methods e.g. the Buchberger algorithm.

The third approach to structural identifiability analysis reviewed here is based on power series expansion of the model outputs as a function of inputs and time (Pohjanpalo (1978), Fliess (1980), Fliess and Lagarrigue (1980) and Walter and Pronzato (1996)). Two types of expansions may be used, one is based on Taylor series and the other based on generating series. In a more recent paper, (Walter et al. 2004), an optimization based method is presented. The classical definitions of identifiability are slightly modified by defining a validity domain for the model parameters.

In the case of practical, or *a-posteriori* identifiability analysis, several methods are available. The first method is the local or global (multi local) sensitivity analysis (Sarmiento Ferrero et al. (2006), Kontoravdi et al. (2005)), which is a widely used method for large models.

The second method for establishing practical identifiability, is an optimization-based approach, proposed by Asprey and Macchietto (2000). In principle, the idea is to maximize the distance between two parameter vectors that essentially give the same model output. If the maximized distance is smaller than some threshold then the model is deemed identifiable. In a more recent paper, the optimization based approach is modified and combined with the multi-local sensitivity analysis into what is called the perturbation algorithm (Sidoli et al. 2005).

A systematic approach to identification of complex reaction networks was developed by Brendel et al. (2006). Their model development is decomposed into steps, where each subsequent step is related to only one part of the model but all possible model candidates are considered. On the other hand, methods based on sensitivity analysis or perturbation study tend to be more applicable to real life applications, but they require a substantial computational effort and are only black box methods which do not directly provide insight into the nonlinearities of the dynamic model.

Even though the two problems mentioned above for nonlinear dynamic systems have attracted attention for a long time it seems that there is still a need to develop methods which can address development of systematic methodologies for structural identifiability of parameters and qualitative experimental design at different stages in model development more efficiently.

It is the focus of chapters 3 and 4 to provide an attempt toward development of a systematic methodology for qualitative experimental design based upon the notion of structural identifiability. The presented methodology focuses upon reaction networks since the determination of approximate kinetics can be very beneficial for subsequent reactor design and development of operational strategies.

1.2.4 Quantitative experimental design

The parameter estimation problem mentioned above in section 1.2.3 can be greatly improved by using quantitative experimental design for parameter pre-

cision improvement.

A review on the topic is provided by Walter and Pronzato (1990) where the focus is mainly on linear models. When performing quantitative experimental design, however, for linear dynamic systems another way is to convert the model from time domain into the frequency domain Sadegh et al. (1994), Sadegh et al. (1995) and then determine the optimal set of frequencies for inputs. This study focuses on stochastic differential equation models. For nonlinear dynamic models which are nearly linear, a common approach is to linearize the model around the operational state of the process and then apply the same method as for dynamic linear models. Körkel et al. (1999) and Bauer et al. (2000), were among the first to develop the concepts and algorithms for quantitative experimental design for dynamic models of systems of ordinary differential equations or differential-algebraic equations. Asprey and Machietto, in a series of papers: Asprey and Macchietto (2000), Asprey and Macchietto (2002) address the problem of robust quantitative experimental design for nonlinear dynamic systems.

Two approaches are usually taken when considering the prior available knowledge about the model parameters.

The first approach takes into account the *a-priori* uncertainty in the model parameters θ. The parameters are assumed to belong to a population with a known distribution $p(\theta)$. The experiments designed in this way are good on average but can be poor for some values of the parameters (Walter and Pronzato (1987), Asprey and Macchietto (2000)).

The second approach aims to determine experimental designs Φ_{WC}, that optimize the worst possible performance for any value of $\theta \in \Theta$ the only prior information about θ is the admissible domain Θ. In this way the design attempts to ensure acceptable performance for all possible values of θ. The WC approach had a limited usage due to burdensome computation. In order to circumvent this problem Asprey and Macchietto (2000), Asprey and Macchietto (2002) proposed a sequential algorithm derived from the worst case approach and they denote this as R-optimal experimental design.

Benabbas et al. (2005) included a criterion containing information about the curvature of the response surface. In order to further account for non-linearity of the system a criterion based on global sensitivity analysis has been introduced in recent paper by Rodriguez-Fernandez et al. (2007).

Thus it seems like, a methodology addressing the problem of quantitative experimental design for stochastic differential equations is still open. The reason to develop such a methodology, concerns the estimation of model parameters in models described by stochastic differential equations, especially because using different terms for state noise (diffusion terms) and measurement variance it is possible to separate the influences of the model error and measurements error.

The development of an algorithm (procedure) and its implementation as software for performing quantitative experimental design for parameters estimation for processes described by stochastic differential equations is the focus of chapter 5. The classical D-optimal design criterion from the linear models theory

is employed here Atkinson and Donev (1996).

1.3 Hypotheses of the thesis

After introducing the motivation and the state of the art it is appropriate to state the hypotheses of this thesis:

1. In order to facilitate efficient optimization of a production process it is most beneficial to develop a reasonably accurate process model where information about model uncertainty is available
2. A structural identifiability analysis step, before the parameter estimation step will guide the model parameters estimation and will reduce both the time and the number of estimations by focusing only on the theoretically identifiable process parameters
3. A qualitative experimental design step will guide the experimentalist to focus on measuring only the relevant states and on perturbing only the relevant inputs in order to identify structurally identifiable process parameters
4. A quantitative experimental design step for grey box stochastic models will improve the parameter estimation step by providing the experimental data containing the maximum amount of information

1.4 Structure of the thesis

The introduction discussed the motivation of this thesis, then it briefly reviewed the state of the art for the main aspects dealt with and then established the hypothesis of the work. Following the hypothesis, the grey box stochastic modelling of SBT is presented in two chapters.

In chapter 2 the modelling concepts and tools are introduced together with two developed models, Model I and Model II. The chapter describes the model development process for an enzymatic reaction network, the so-called system of bio-transformation. Various grey box stochastic models for SBT, usually increasing in complexity are presented discussed and validated against laboratory data. The performance of the models gradually improves from version to version.

Chapter 3 introduces and illustrates the identifiability analysis algorithm developed for assessing the possibility of estimating the model parameters from the available data for reaction network models. The identifiability analysis is applied for the simplest developed model in chapter 2. A validation of the correctiveness of the results is provided as well.

Chapter 4 continues and expands the ideas presented in chapter 3 into the problem of qualitative experimental design. The analyzed model is Model I introduced in chapter 2 and provides a completion of the analysis.

Chapter 5 develops quantitative experimental design for grey box stochastic models. After introducing the work-flow and the methodology, the algorithm is illustrated on a simple model mainly for testing and comparison.

In chapter 6 an improved version of the grey-box stochastic modelling framework is introduced. The methodologies presented in chapters 3–5 have been incorporated in the proposed framework. One more model, Model III is developed and discussed.

A second approach for qualitative experimental design is proposed and illustrated on Model III. Parts of the results obtained from this analysis were used to establish which parameters can be estimated during model development.

An optimal quantitative experimental design for parameter estimation for Model III is performed as well.

Chapter 7 describes the software contributions related to the work presented here.

Finally, chapter 8 closes the thesis by briefly reviewing the hypothesis formulated in this introductory section and by reviewing the contributions to the grey box stochastic modelling framework. This closing chapter includes a short discussion about the software implementations and provides suggestions for future work.

1.5 Publications

The work presented in this thesis has been presented in several publications as reviewed journal papers, reviewed conference papers, various oral presentations and posters. Below, the most important are listed, and the connection with the various parts of the thesis is described.

1.5.1 Reviewed journal papers

The content of chapter 3 has been submitted for publications as:

1. Davidescu Florin Paul, Jørgensen Sten Bay, "Structural parameter identifiability analysis for dynamic reaction networks", Chemical Engineering Science, volume 28, pp 4754-4762, Elsevier 2008.

1.5.2 Reviewed conference papers

Small parts of the work presented in chapter 6 has been already presented as contribution for the ESCAPE 16 proceedings volume as:

1. Florin Paul Davidescu, Madsen Henrik, Schümperli Michael, Heinemann Matthias, Panke Sven, Jørgensen Sten Bay, "Stochastic Grey Box Modeling of the enzymatic biochemical reaction network E. coli mutants", Computer-Aided chemical engineering, volume 24, Elsevier 2006.

The work in chapters 3 and 4 contains some of the work presented at ESCAPE 17 in proceedings volume:

1. "Systematic qualitative experimental design based upon identifiability analysis", Computer-Aided chemical engineering, volume 25, Elsevier 2007.

2. Florin Paul Davidescu, Henrik Madsen, Sten Bay Jørgensen, "Using Lie algebra to study the parameters identifiability and to perform experimental design", ECCE 7 abstract

1.5.3 Oral presentations and posters

Regarding oral presentations and posters the content of chapters 3 and 4 has been presented in the following contributions also:

Different aspects of chapter 6 have been presented in the following contributions also:

1. Florin Paul Davidescu, Henrik Madsen, Sten Bay Jørgensen, "Production of complex fine chemicals using a system of bio-transformation", CAPEC consortium annual meeting, held at Kob æk Strand, Denmark, June 2006

2. Florin Paul Davidescu, Sten Bay Jørgensen, "Stochastic grey-box modeling of a system of biotransformations", DK2, held at Technical University of Denmark Kgs. Lyngby June 2006, published in the proceedings conference volume

3. Florin Paul Davidescu, Henrik Madsen, Sten Bay Jørgensen, "Systematic improvement of grey-box stochastic modelling of sequential continuous cultivations", NPCW 13, at Technical University of Denmark Kgs. Lyngby January 2006, included in the workshop proceedings

4. Florin Paul Davidescu, Henrik Madsen, Sten Bay Jørgensen, "Grey-box stochastic modelling of a an enzymatic reactions network for biotransformation", presented in session 656b Applied Mathematics in Bioengineering II, at AICHE annual meeting, San Francisco, November 2006

5. Florin Paul Davidescu, Sten Bay Jørgensen, "Qualitative And Quantitative Experimental Design For An Enzymatic Reaction Network", presented in session 10B01 Process Monitoring and Identification - I, at AICHE annual meeting, Salt Lake city, November 2007

2

Reaction Network Modelling within Grey Box Framework

Abstract

Key elements of a grey-box stochastic modelling framework are (reviewed and) applied for preliminary model development of an enzymatic reaction network. The main purposes are to illustrate and pinpoint the properties and limitations of the grey box stochastic modelling methodology identifying essential targets in the reaction network for maximizing productivity of the network. Experimental data for four measured compounds of the enzymatic reaction network are available as time series data collected during batch experiments. Two models for the system of bio-transformation are developed. The first model assumes mass action type kinetics for the main pathway. The second model incorporates additional reactions and more complex kinetics for the main pathways. The models are developed by using the grey-box modelling framework for combining information embedded in experimental data with first principle engineering knowledge. Various statistical tests are used to assess statistical quality of the model and parameter estimates. Limitations of the present grey box framework are pinpointed in an effort to identify areas for improving the grey box modelling framework.

2.1 Introduction

The main objective of this work is to develop an improved modelling methodology for dynamic systems which can provide models with good long term prediction properties of relatively large complexity. The starting point is represented by an existing grey-box modelling framework (Kristensen et al. 2004a). Along with the methodology development, a model for an enzymatic reaction network constitutes the secondary objective of this work. The considered case is of large complexity showing multiple recirculations.

2.1.1 Why using stochastic differential equations?

The motivation to use stochastic differential equations (SDEs) is primarily discussed based on the main application represented by the enzymatic reaction network. Enzymatic reaction networks in general represents nonlinear dynamic systems of large complexity. First engineering principles to develop dynamic models represents the preferred choice by several authors Chassagnole et al. (2002), Bali and Thomas (2001) for modelling enzymatic reaction networks in *E-coli* microorganisms. However the kinetic mechanism and the subsequent kinetic equation for many enzymatic reactions have been investigated for purified enzymes only. The influence of the other enzymes and cofactors on the particular enzymatic kinetic mechanism is not clarified for many enzymes, thus model uncertainty is present. Moreover, the enzymatic reaction network (SBT) considered in this study is an *in-vitro* realization since the cells are destroyed after the cultivations and the (most relevant) enzymes used for synthesis of the desired compound in batch of fed-batch operation. The first principles engineering models assume the kinetic equations to be correct, thus model error may be absorbed in the parameter estimates. By using SDEs with a pure diffusion term or random walk, a possibility to account for model deficiency is provided.

The experimental points are collected only at discrete points in time and only for a subset of the model states. Thus for ODE models in general and for enzymatic networks described by ODE models the model errors are most often serially correlated - i.e. the difference between the data and the model predictions are correlated in time. This correlation has to be accounted for by any estimation procedure, and by using a maximum likelihood procedure the likelihood function is well known to be constructed as a product of conditional densities. In this (one-step) conditional densities the autocorrelation must be taken into account, and this is most adequately done by considering SDEs instead of ODEs. When working with enzymatic networks, some random effects occur in reality which the first principles engineering models can not account for. While simulating the enzymatic reaction network it is not possible to simulate most real life systems in a realistic manner without using SDEs.

If inputs are perturbed, then even these inputs contain some uncertainty. SDEs allow for measurement errors for the input variables - and this is then accounted for by the diffusion part. For both ODE and SDE models it is well known that errors of the dependent (output) variable is described by the observation error in the measurement equation of the state space model, but only SDEs allow for measurement errors in the input variables.

If the scope of the current work is broaden, there are a few other advantages of using SDEs. If the developed models are to be used for developing model based controllers the following differences need to be mentioned. For ODEs the future of the states is assumed to be known exactly - and this is rarely the situation. For SDEs the uncertainty in general increases as the prediction horizon increases, and a knowledge of the uncertainty about future values of

the states is very important. For forecasting applications, the SDEs allow for a simpler (often more 'operational') model than ODEs since the model approximation is implicitly described by the diffusion part of the model.

Finally the SDEs are important in order to enable statistical methods for model estimation, model identification, model validation and falsification.

2.1.2 Review of the existing grey-box stochastic modelling framework

The grey-box stochastic modelling framework is based upon stochastic differential equation models. The purpose of the framework is to enable a number of powerful mathematical and statistical tools to assist the model development in a systematic way. Using first principle engineering the model equations are derived in a form of a set of ordinary differential equations, and then complemented with diffusion terms to obtain the set of stochastic differential equations. The diffusion terms account for model errors and/or for the un-modeled effects of the system plus the noisy input.

By considering the model as a system of stochastic differential equations and by introducing a pure diffusion process or random walk enables a possibility to identify variation not recognized at a certain point in the model building thereby establishing possibilities for pinpointing the model error in several steps (Kristensen et al. 2004a). The measurement equations include measurements errors as well, thus in this approach it is possible to distinguish between measurement and process errors.

In the next step the set of unknown parameters together with the diffusion terms and the measurements variances are estimated from experimental data using a maximum likelihood or a maximum a-posteriori method in a prediction error setting. More details about the estimation scheme are given in Section 2.2.

The model is (un-)falsified using different statistic tests. Then, the model is reformulated and the iterations continued until the model is un-falsified given available data i.e. the information contained in the data is extracted. In the original framework the statistical test for significance applied to the diffusion terms together with some residual analysis are used to pinpoint the model deficiencies and to formulate hypotheses about the deficiencies. Once the deficiencies have been identified non-parametric tools are used to improve the deficient part of the model.

The remaining part of the chapter is organized as follows: Section 2.2 describes in some detail the parameter estimation scheme, Section 2.3 describes the experimental data, Section 2.4 describes the development of Model I for an enzymatic reaction network using the grey box stochastic modelling methodology. Section 2.5 presents a second model, Model II. A validation of Model II is performed in section 2.5.1. The chapter is closed by a discussion of the results and modelling methodological deficiencies and the conclusions are drawn.

2.2 Methods and Software

In this section, a brief review of theoretical aspects related to the parameter estimation scheme based on maximum likelihood estimation for stochastic differential equations is provided for convenience. A short description of the software tool used is given before the grey box stochastic modelling methodology is summarized.

2.2.1 Model structure of SDE

A continuous-discrete stochastic state space model consists in the general case of a set of non-linear discretely, partially observed (Jazwinski 1970), *Itô* SDE's with measurement noise, i.e.:

$$dx_t = f(x_t, u_t, t, \theta)\, dt + \sigma(u_t, t, \theta)\, d\omega_t \tag{2.1}$$

$$y_k = h(x_k, u_k, t_k, \theta)\, dt + e_k \tag{2.2}$$

In the above formulation:

- $t \in \mathbf{R}$ is the time variable;
- $x_t \in X \subset \mathbf{R}^n$ is a vector of state variables;
- $u_t \in U \subset \mathbf{R}^m$ is a vector of input variables;
- $y_t \in Y \subset \mathbf{R}^l$ is a vector of output variables;
- $\theta \in \Theta \subset \mathbf{R}^p$ is a vector of (possibly unknown) parameters;
- $f(\cdot) \in \mathbf{R}^n$ represents the deterministic (or drift) functions of the SDE model;
- $\sigma(\cdot) \in \mathbf{R}^{n \times n}$ represents the stochastic (diffusion) functions of the SDE model;
- $h(\cdot) \in \mathbf{R}^l$ are known non-linear functions representing the measurement equations;
- $\{\omega_t\}$ is an n-dimensional standard Wiener process (Jazwinski, 1970);
- $\{e_k\}$ is an l-dimensional white noise process with $e_k \in N(0, S(u_k, t_k, \theta))$.

The diffusion term of Eq. 2.1 is assumed to be independent of the process states and in the grey-box modelling framework (Kristensen et al. 2004a) the Σ matrix of the equation set is formulated in a diagonal form, one term for each state. The Σ matrix is a matrix whose diagonal elements are the $\sigma(\cdot)$ parameters while the off-diagonal elements are zero. The Σ matrix can have any parametrization, however the diagonal form enables the possibility of pinpointing the model deficiencies as the cross coupling to the other states is neglegted.

2.2.2 Parameter estimation

The solution to Eq. set 2.1, is a Markov process with an estimation method of the unknown parameters of the model in Eqs. 2.1–2.2, e.g. *maximum likelihood* (ML) or *maximum a posteriori* (MAP).

The latter method can be applied when prior information about the parameters is available. For a stochastically independent sequence set of S consecutive measurements (Kristensen et al. 2004b):

$$\mathbf{Y} = \left[Y^1_{N_1}, Y^2_{N_2}, \ldots, Y^i_{N_i}, \ldots Y^S_{N_S} \right] \tag{2.3}$$

each sequence is:

$$Y^i_{N_i} = \left[y^i_{N_i}, \ldots, y^i_k, \ldots, y^i_1, y^i_0 \right] \tag{2.4}$$

Denote $p(\theta)$ as a prior probability density function for the parameters. In the general case, estimates of the parameters in Eqs. 2.1–2.2 can then be found as the parameters θ that maximize the joint posterior probability density function given in the general form:

$$p\left(Y|\theta\right) = \left(\prod_{i=1}^{S} p\left(Y^i_{N_i}|\theta\right) \right) p\left(y^i_0|\theta\right) p\left(\theta\right) \tag{2.5}$$

or equivalently:

$$p\left(\mathbf{Y}|\theta\right) = \prod_{i=1}^{S} \left(\prod_{k=1}^{N_i} p\left(y^i_k|Y^i_{k-1},\theta\right) \right) p\left(y^i_0|\theta\right) p\left(\theta\right) \tag{2.6}$$

where the conditional probability rule $P(A \cap B) = P(A|B) \cdot P(B)$ was applied successively in order to form a product of conditional probability functions. This formulation allows MAP estimation on multiple data sets. If the $p\left(\theta\right)$ is uniform, then the equation 2.6 becomes:

$$p\left(\mathbf{Y}|\theta\right) = \prod_{i=1}^{S} \left(\prod_{k=1}^{N_i} p\left(y^i_k|Y^i_{k-1},\theta\right) \right) p\left(y^i_0|\theta\right) \tag{2.7}$$

which allows ML estimation.
In order to obtain an exact evaluation of the likelihood function in Eq. 2.5, a general nonlinear filtering problem has to be solved (Jazwinski 1970), but this is computationally infeasible in practice.

However, since the increments of the standard Wiener process $\{\omega_t\}$ driving the SDE in Eqs. 2.1 are Gaussian, it is reasonable to assume that the conditional probability densities in Eq. 2.5 can be well approximated by Gaussian densities, which means that a method based on the extended Kalman filter (EKF) can be applied (Kristensen et al. 2004b). The Gaussian density is completely characterized by its mean and covariance. Introducing the following notation:

$$\hat{y}^i_{k|k-1} = E\{y^i_k | Y^i_{k-1}, \theta\} \tag{2.8}$$

$$R^i_{k|k-1} = V\{y^i_k | Y^i_{k-1}, \theta\} \tag{2.9}$$

$$\varepsilon^i_k = y^i_k - \hat{y}^i_{k|k-1} \tag{2.10}$$

The ϵ^i_k is denoted the *innovations* since this represents the new information which is not predicted based on previous information. By replacing Eqs. 2.8–2.10 in Eq. 2.6 and assuming a Gaussian pdf the likelihood function becomes:

$$p\left(\mathbf{Y}|\theta\right) = \prod_{i=1}^{S}\left(\prod_{k=1}^{N_i}\frac{\exp\left(-\frac{1}{2}\left(\epsilon^i_k\right)^T\left(R^i_{k|k-1}\right)^{-1}\left(\epsilon^i_k\right)\right)}{\sqrt{det\left(R^i_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right)p\left(y^i_0|\theta\right)p\left(\theta\right) \tag{2.11}$$

Then further conditioning on:

$$y_0 = \left[y^1_0, y^2_0, \ldots, y^i_0, \ldots y^S_0\right] \tag{2.12}$$

and applying nonlinear optimization and finding the minimum of the negative logarithm of the function in Eq. 2.11 the set of parameters θ are obtained.

$$\hat{\theta} = argmin_{\theta\in\Theta}\left\{-ln\left(p\left(\theta|\mathbf{Y}, y_0\right)\right)\right\} \tag{2.13}$$

For each set of the parameters θ in the optimization operation, the *innovations* ϵ_k and their covariances $R_{k|k-1}$ are computed recursively by means of the EKF (Kristensen et al. 2004b).

2.2.3 Statistical tests

An estimate of the uncertainty of the parameter estimates is obtained by using the fact that by the central limit theorem the estimator in Eq. 2.13 is asymptotically Gaussian with mean θ and covariance matrix $\Sigma_{\hat{\theta}} = H^{-1}$ where the matrix H is given by Eq. 2.14:

$$h_{ij} = -E\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}ln\left(p\left(\theta|Y; y_0\right)\right)\right\}; i, j = 1, \ldots, p \tag{2.14}$$

An approximation to H can be obtained from :

$$h_{ij} = -\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}ln\left(p\left(\theta|Y; y_0\right)\right)\right)\Bigg|_{\theta-\hat{\theta}}; i, j = 1, \ldots, p \tag{2.15}$$

which represents the Hessian matrix evaluated at a minimum of the objective function.

The asymptotic Gaussianity of the estimator in Eq. 2.15 also allows t-tests to be performed to test the hypothesis:

$$H_0 : \theta_j = 0 \tag{2.16}$$

against the corresponding alternative:

$$H_1 : \theta_j \neq 0 \tag{2.17}$$

i.e. to test whether a given parameter θ_j is marginally insignificant or not. The test quantity is the value of the parameter estimate divided by the standard deviation of the estimate, and under H_0 this quantity is asymptotically t-distributed with a number of degrees of freedom DF that equals the total number of observations minus the number of estimated parameters (Kristensen et al. 2004b).

The Hessian matrix in Eq. 2.15, represents the observed Fisher Information Matrix while the definition in Eq. 2.14 represents the definition of the Fisher Information Matrix. In this way, a direct connection between the parameter estimate uncertainty and quantitative experimental design is given by the Kramer-Rao inequality.

When it comes to discriminating a model between two or more models, e.g. one model being a simplification of another model, a statistical test can be used. The test is called the log-likelihood ratio test (the Wilk test) and is described in the following. The asymptotic behavior of the quantity $-2log\lambda\,(Y)$ converges to a χ^2 distribution with nd degrees of freedom DF and a quantile value of $1-\alpha$, where α is to be specified by the user. The degrees of freedom are represented by the difference between the number of parameters for the two compared models. The λ is determined by computing the ratio of the log-likelihood function values for the two models. The numerator is considered to be the value of the log-likelihood function for the model containing a smaller parameter set. Thus the general hypothesis is formulated as follows:

$$H_0 : \theta \in M(I) \tag{2.18}$$

and the alternative hypothesis:

$$H_1 : \theta \in M(II) \tag{2.19}$$

The hypothesis H_0 is that the θ vector is the first model parameter vector and the alternative hypothesis H_0^a is that the θ vector is the second model parameter vector.

2.2.4 Software

The methods described above for parameter estimation in stochastic differential equations are implemented in a software toolbox for continuous time stochastic modeling by previous authors (Madsen and Melgaard (1991), Kristensen

et al. (2004a)). The software has been used in several applications already, for example in the case of modeling of heat dynamics of a building (Madsen and Holst 1995), modeling of a green-house dynamics (Nielsen and Madsen 1998), modeling of pharmaceutical processes, (Tornøe et al. 2004) In the current work the software has been modified and complemented to allow for the possibility of having multiple data sets starting with different initial conditions considered simultaneously. From the algorithm point of view, when computing the maximum likelihood function given in Eq. 2.11 an extra outer loop was added to include all the different experimental data sets. That is, when parameter estimation is performed, in the final estimate of the parameters, information from different experiments is included simultaneously. Clearly some parameters may belong just to one data set, while other parameters may be common for all data. The performance of this extended version was tested and later used when multiple different experiments were used to provide complementary information for the parameter estimation.

2.3 Experimental data

The experiments have been conducted at ETH Zurich (Schümperli et al. 2007). In phase I, fed-batch (semi-batch) fermentations of *E. coli* LJ110 *tpi* are conducted until the optical density (OD) in the bioreactor reaches a preset value of 600. The broth is centrifuged and the cells are resuspended in SBT-buffer: 100 mM HEPES, 0.84 mM KCl, 1 mM ZnSO4 and at pH = 7. The cells are disrupted by high-pressure homogenization. The remaining solids are eliminated by centrifugation/filtration and the liquid extract is recovered. The total protein concentration is determined by Bradford analysis and adjusted to the desired concentration by dilution with the SBT buffer. The liquid extract contains the enzymes and compounds present in the cell at the time when the fermentation was stopped. In phase II, a volume of 5 ml of SBT extract is used for each experiment. Defined amounts of *hexokinase* and Lactate-DH as well as ATP and NAD^+, are added. The reactions are initiated by adding Glucose. Samples are collected according to a previously defined time plan. The experiments are terminated after 300 or 360 minutes. First, the proteins are removed by precipitation with HCl (1M) followed by centrifugation. The samples are analyzed by enzymatic assays. First Glucose-6-Phosphate is determined, and then Glucose and Glucose-6-Phosphate are determined together by addition of *Glucose-6-Phosphate-de-hydrogenase* and *hexokinase* to form NADPH which is determined spectro-photometrically. Di-Hydroxy-Acetone-Phosphate is determined by addition of Glycerol-3-Phosphate-dehydrogenase and measuring the NADH consumption spectro-photometrically. ATP is measured as well. Each experiment is defined by the SBT extract design, by the initial concentrations of Glucose, cofactors, Phosphate and enzymes, by the sampling times and by the end time. The samples were collected frequently in the beginning (between 1-15 minutes) and less frequent towards the end of the batch and the total

time was 360 minutes. The initial concentration of the enzymes, cofactors, phosphate and glucose is also known, with some uncertainty. As experimental data, time series data are initially available for four measured compounds. The first measurement consists of the sum of Glucose and Glucose-6-Phosphate the second represents Glucose-6-Phosphate, the third represents DHAP and the fourth ATP.

In the Table 2.1 a summary of the conditions defining each experiment is given. For each experiment two reactors with a volume of 5.5 ml were used and for each of the reactors, two samples were taken for each measurements, which means four duplicates for each of the samples. Each time series consists of 12 samples. Thus the two reactors constitute different realizations.

Experiment ID	**Exp.17 A**
Hexokinase U/ml	0.01
Lactate-DH U/ml	1
Glc mM	11.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
Prot.tot. mg/ml	10
T	37
c-AMP mM	10
PPi mM	10

Table 2.1: Conditions defining the experiment 17A

2.4 Model I for an enzymatic reaction network

Since it is not clearly known which reactions occur in the cell free extract, a simplified reaction network of the reactions occurring during the kinetic experiments with the enzymatic extract is depicted in Figure 2.1. The starting point in establishing the network, represents the central carbon metabolism reaction network of the *E-coli* mutants. The reversible reaction between DHAP and Glyceraldehide-3-Phosphate does not take place since the *tpi* gene which is responsible for the expression of the enzyme catalyzing this reaction has been knocked-out. Since it is desired to have all the reactions regenerating the cofactors required in the first reaction it is necessary to include the genes catalyzing the reactions after Glyceraldehide-3-Phosphate that are responsible for ATP regeneration. NAD^+ is consumed and then produced during ATP regeneration. A simple generic reaction for spontaneous degradation of ATP was included to account for the ATP degradation due to the numerous enzymes present in the extract. Efforts are made to isolate further only the necessary enzymatic reactions in order to reduce the fast consumption of ATP toward undesired reactions. The compounds marked with a star, Glucose, Glucose-6-Phosphate,

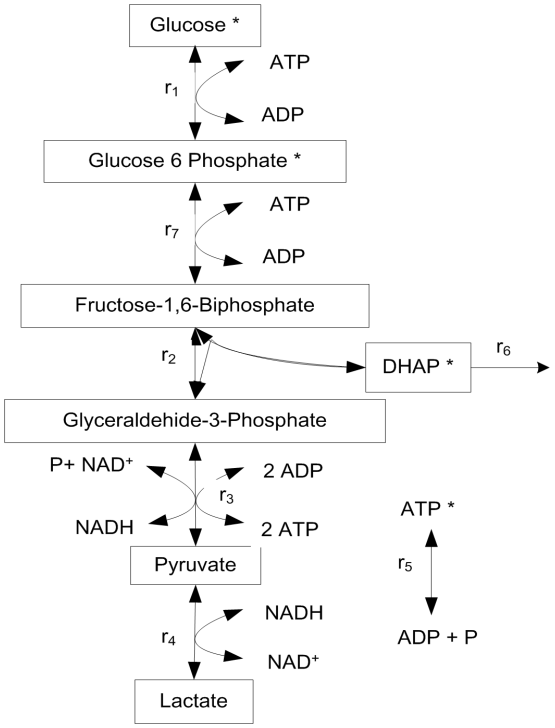


Figure 2.1: Reaction network used for Model I development with the measured species indicated by a $*$

DHAP and ATP are measured during the SBT batch experiments. Thus four full measurement sequences are available along with the initial conditions for NAD^+, PO4.

The first reaction considered in Model I, r_1 is the enzymatic transformation from Glucose to Glucose-6-Phosphate. The second reaction considered is the reaction from Fructose-1,6-Biphosphate to Glyceraldehide-3-Phosphate and Di-Hydroxy-Acetyl-Phosphate, r_2 which is catalyzed by *aldolase*. The reactions consuming the Glyceraldehyde-3-Phosphate towards Pyruvate were all lumped in the model into one single reaction denoted r_3. The reaction producing Lactate from Pyruvate was included as reaction r_4. The reason to include these two reactions was already mentioned. As a decrease for DHAP has been observed in the experimental data, a degradation of DHAP was considered in the enzymatic reaction by r_6. The reaction consuming the ATP cofactor is reaction r_5. Two enzymatic transformations from Glucose-6-Phosphate to Fructose-1,6-Biphosphate were considered as a single reaction in the network denoted r_7.

2.4.1 Model I equations and validation

Based on the above simplified enzymatic reaction network a first model of the enzymatic reaction network is formulated as follows in terms of an ordinary differential model. The time evolution of each individual chemical species is described by a continuous time dynamic mass balance as given in Eqs. 2.20.

$$
\begin{aligned}
\frac{dc_{GL}}{dt} &= -r_1 \\
\frac{dc_{F16B}}{dt} &= r_7 - r_2 \\
\frac{dc_{DHAP}}{dt} &= r_2 - r_6 \\
\frac{dc_{G3P}}{dt} &= r_2 - r_3 \\
\frac{dc_{PYR}}{dt} &= r_3 - r_4 \\
\frac{dc_{LAc}}{dt} &= r_4 \\
\frac{dc_{ATP}}{dt} &= -r_1 - r_7 + 2r_3 - r_5 \\
\frac{dc_{NAD}}{dt} &= -r_3 + r_4 \\
\frac{dc_{PO4}}{dt} &= -r_3 + r_5 \\
\frac{dc_{G6P}}{dt} &= r_1 - r_7 \\
\frac{dc_{ADP}}{dt} &= r_1 + r_7 - 2r_3 + r_5 \\
\frac{dc_{NADH}}{dt} &= r_3 - r_4
\end{aligned} \tag{2.20}
$$

Since many of the reactions described above are in fact sequences of enzymatic reactions and considering the present limited availability of information about the system in terms of experimental data, simple mass action kinetic models are assumed, as given in Eqs. 2.21. Improvement of both the kinetic model for each of the enzymatic reactions considered and the general model structure represents the the focus of this work.

$$
\begin{aligned}
r_1 &= r_{1maxf} c_{GL} c_{ATP} - r_{1maxb} c_{G6P} c_{ADP} \\
r_2 &= r_{2maxf} c_{F16B} - r_{2maxb} c_{DHAP} c_{G3P} \\
r_3 &= r_{3maxf} c_{G3P} c_{NAD} c_{PO4} c_{ADP} - r_{3maxb} c_{PYR} c_{NADH} c_{ATP} \\
r_4 &= r_{4maxf} c_{PYR} c_{NADH} - r_{4maxb} c_{LAC} c_{NAD} \\
r_5 &= r_{5maxf} c_{ATP} - r_{5maxb} c_{ADP} \\
r_6 &= r_{6maxf} c_{DHAP} \\
r_7 &= r_{7maxf} c_{G6P} c_{ATP} - r_{7maxb} c_{ADP} c_{F16B}
\end{aligned} \tag{2.21}
$$

The model equations, Eqs. 2.20–2.21 are converted into a system of stochastic differential equations (Kristensen et al. 2004a) by adding the diffusion terms

multiplied by a standard Wiener process. The set of stochastic differential equations appears as in Eqs. 2.22.

$$\begin{aligned}
dc_{GL} &= (-r_1)\,dt + \sigma_{11}d\omega \\
dc_{F16B} &= (r_7 - r_2)\,dt + \sigma_{22}d\omega \\
dc_{DHAP} &= (r_2 - r_6)\,dt + \sigma_{33}d\omega \\
dc_{G3P} &= (r_2 - r_3)\,dt + \sigma_{44}d\omega \\
dc_{PYR} &= (r_3 - r_4)\,dt + \sigma_{55}d\omega \\
dc_{LAc} &= (r_4)\,dt + \sigma_{66}d\omega \\
dc_{ATP} &= (-r_1 - r_7 + 2r_3 - r_5)\,dt + \sigma_{77}d\omega \\
dc_{NAD} &= (-r_3 + r_4)\,dt + \sigma_{88}d\omega \\
dc_{PO4} &= (-r_3 + r_5)\,dt + \sigma_{99}d\omega \\
dc_{G6P} &= (r_1 - r_7)\,dt + \sigma_{1010}d\omega \\
dc_{ADP} &= (r_1 + r_7 - 2r_3 + r_5)\,dt + \sigma_{1111}d\omega \\
dc_{NADH} &= (r_3 - r_4)\,dt + \sigma_{1212}d\omega
\end{aligned} \tag{2.22}$$

To complete the model it is necessary to include the equations for discrete measurements. Four equations for each measurement variable will be considered: Glucose, Glucose-6-Phosphate, DHAP and ATP. The measurement equations are given in set 2.23, where the $S_{11} - S_{44}$ parameters represent the variances of the measured experimental data.

$$\begin{aligned}
y_{GLC} &= c_{GLC} + e, \quad e \in N(0, S_{11}) \\
y_{G6P} &= c_{G6P} + e, \quad e \in N(0, S_{22}) \\
y_{DHAP} &= c_{DHAP} + e, \quad e \in N(0, S_{33}) \\
y_{ATP} &= c_{ATP} + e, \quad e \in N(0, S_{44})
\end{aligned} \tag{2.23}$$

The second step in the grey-box stochastic methodology framework represents estimation of the parameters given Model I discussed above and the available experimental data. Using Continuous Time Stochastic Modelling (CTSM) software, (Kristensen et al. 2004b), selected sets of parameters have been estimated and statistical tests for significance of individual parameters have been applied in order to determine which parameters are significant in a preliminary step. Along with this information, CTSM provides an estimate of the covariance matrix of the parameters that is used to check the parameter correlation. A few parameter estimation attempts showed that many parameters were correlated and insignificant while for other runs it was not even possible to achieve convergence.

A preliminary local sensitivity analysis was used to check which of the model parameters are impossible to estimate using available experimental batch data, where the initial amount of ATP was varied. From the identifiable parameters

several subsets of parameters have been considered for estimation. Moreover, as a measure of simultaneous significance of parameters a statistic test based on the ratio of the log-likelihood function value is used as discussed in Section 2.2.3.

The experimental data obtained in Experiment 17A, see conditions in table 2.1, have been used for parameter estimation. One reactor has been used in this experiment and two time series data sets have been collected. Tables 2.2 contains the value of the log likelihood function and the number of estimated parameters.

Set	$LogL$	$Npar$
1	-1.155707605130632E+01	14
2	-1.174498134588278E+01	15
3	-1.189327118730689E+01	16
4	-1.189328815830789E+01	17
5	-1.222756558170784E+01	17

Table 2.2: Log-likehood function value for the estimation results for Model I when estimating different number of parameters, using data from Experiment 17A, set 1 and 2, reactor I

For these results, the log-likelihood ratio test presented in section 2.2.3 has been applied in order to statistically determine the most significant number of estimated parameters. Table 2.3 presents the result of the log-likelihood ratio test for some of the pairings of estimated parameters considered from the results presented in Table 2.2.

Each row represents the results of the application of the log-likelihood ratio test for two sets of parameters. For example, in the first row of the Table 2.3, the hypothesis is H_0 : the parameter vector θ is represented by the parameter set 5 from Table 2.2, while the alternative hypothesis is H_1 : the parameter vector θ is represented by the parameter set 1 from the Table 2.2. First column contains the log-likelihood ratio $\lambda(Y)$, used to compute the statistical quantity used for the test $-2log\lambda(Y)$. The degrees of freedom DF is given by the column $\#\theta_0 - \#\theta_1$. The quantile of the χ^2 distribution for an $alpha = 0.05$ is given as well in the next column. The last column contains the selected true hypothesis, for this comparison the true hypothesis being the alternative hypothesis since the value of the statistical quantity $-2log\lambda$ is much lower than the quantile.

It looks like the case when less parameters have been estimated is favored by the log-likelihood ratio test, while the t-test statistics shows that some parameters are insignificant when more parameters are estimated. Based on the results presented in Tables 2.2-2.3, a good compromise between the uncertainty of the parameters estimates and the value of the log-likelihood function seems to be *set 2*. Strictly following the results of the log-likelihood ratio test, the *set 1* should be selected, however there were insignificant changes in model performance (results not shown).

The estimation results corresponding to set 1 are given in the following Tables

H_1/H_0	$\lambda(Y)$	$-2log\lambda(Y)$	$\#\theta_0 - \#\theta_1$	$\chi^2(\#\theta_0 - \#\theta_1, 0.95)$	*True hypothesis*
1/5	0.945	0.0489	3	7.8147	H_1, 1
3/5	0.972	0.024	1	3.8415	H_1, 3
1/4	0.945	0.0489	3	7.8147	H_1, 1
3/4	0.999	7.3E-7	1	3.8415	H_1, 3
1/2	0.9839	0.014	1	3.8415	H_1, 1

Table 2.3: Loglikelihood ratio test applied for estimation results for Model I using data from Experiment 17A, set 1 and 2, reactor I

2.4–2.7.

Value of objective function	-1.174498134588278E+01
Value of prior function	-1.843101136264748E+01
Value of penalty function	5.897797922664172E-04
Negative logarithm of determinant of Hessian	-1.312117454303914E+02
Number of iterations	35
Number of objective function evaluations	60

Table 2.4: Estimation results for Model I, *set 2* from Table 2.2

In Table 2.4 the objective function represents the log-likelihood function corresponding to the MAP estimation scheme, the prior function represents the term related to the prior probability density function for parameters θ, $p(\theta)$. In the CTSM software program, the bounds on the parameters to be estimated are converted into an open interval by a mathematical transformation and the corresponding optimization problem becomes an unconstrained optimization problem. A penalty function is therefore necessary to force the solution to remain in the feasible interval. The program gives as results the value of the negative logarithm of the determinant of Hessian also, which is a measure of the parameter uncertainty. The last two rows in Table 2.4 shows the number of iterations and the number of log likelihood function evaluations needed to achieve convergence.

Name	Value	Name	Value	Name	Value
c_{GL0}	11.1	c_{G6P0}	1.0E-5	r_{7maxf}	9.0
c_{F16B0}	1.0E-5	c_{ADP0}	1.0E-5	$ln(\sigma_{44})$	-8.0
c_{DHAP0}	1.0E-5	c_{NADH0}	1.0E-5	$ln(\sigma_{55})$	-8.0
c_{G3P0}	1.0E-5	r_{1maxb}	9.5355E-9	$ln(\sigma_{66})$	-8.0
c_{PYR0}	1.0E-5	r_{2maxb}	1.8581	$ln(\sigma_{88})$	-8.0
c_{LAC0}	1.0E-5	r_{3maxb}	0.01	$ln(\sigma_{99})$	-8.0
c_{ATP0}	11.5	r_{4maxb}	0.00855	$ln(\sigma_{1111})$	-8.0
c_{NAD0}	5.75	r_{5maxb}	0.15	$ln(\sigma_{1212})$	-8.0
c_{PO40}	11.1	r_{6maxf}	1.0E-12		

Table 2.5: Fixed parameters values for Model I

The model parameter which have been fixed during the estimations are given

in Table 2.5. The values of the initial states are given as well. For the compounds which are added to the reactor the values from the actual experiments have been assigned. For the rest of the compounds it was assumed that initially they are not present and only will be formed later after the reactions are initiated. A value of $1.0E-5$ represents zero value. Since not much information about the diffusion terms was available for the unmeasured states they were assumed to be very small. For the results shown here, in fact, instead of the normal values of the diffusion terms σ due to robustness issues their natural logarithm was used/estimated. As for the fixed kinetic parameters educated guess have been used. Generally, the range of values for similar enzymatic reactions have been used, or alternatively the values obtained from previous optimizations.

Name		Estimate	Std. dev.	t-score	$p(>\|t\|)$	dF/dPar	dPen/dPar
r_{1maxf}	ML	1.1606E-01	2.6920E-02	4.3111	0.0000	0.0000	0.0000
r_{2maxf}	ML	6.3043E-02	1.1785E-02	5.3494	0.0000	-0.0000	0.0000
r_{3maxf}	ML	5.5181E-03	1.6963E-03	3.2531	0.0016	-0.0000	0.0000
r_{4maxf}	ML	9.2214E-02	3.4610E-02	2.6644	0.0091	-0.0000	0.0000
r_{5maxf}	ML	2.1798E+00	5.5699E-01	3.9136	0.0002	-0.0000	0.0000
r_{7maxb}	ML	9.4748E-02	2.6147E-02	3.6237	0.0005	0.0000	0.0000
$ln(\sigma_{11})$	MAP	-1.0000E+01	9.9928E-03	-1000.7194	0.0000	0.0000	0.0001
$ln(\sigma_{22})$	MAP	-6.0000E+00	1.0018E-02	-598.9330	0.0000	0.0000	0.0000
$ln(\sigma_{33})$	MAP	-6.0000E+00	9.9367E-03	-603.8229	0.0000	-0.0000	0.0000
$ln(\sigma_{77})$	MAP	-5.0002E+00	9.9931E-03	-500.3612	0.0000	-0.0000	0.0000
$ln(\sigma_{1010})$	MAP	-6.0000E+00	1.0022E-02	-598.7078	0.0000	0.0000	0.0000
S_{11}	ML	1.2233E-02	4.1473E-03	2.9497	0.0040	0.0000	0.0000
S_{22}	ML	3.9002E-02	1.1130E-02	3.5043	0.0007	0.0000	0.0000
S_{33}	ML	5.3089E-02	1.4798E-02	3.5875	0.0005	-0.0000	0.0000
S_{44}	ML	2.2945E+00	6.9438E-01	3.3045	0.0014	-0.0000	0.0000

Table 2.6: Estimation results for Model I using data from Experiment 17A, set 1 and 2, reactor I

Table 2.6 contains the parameter estimates along with the evaluated standard deviations and the values of the t-test quantity. The probability of a parameter to be insignificant is given in the fourth column. The last two columns contain the partial derivative of the log-likelihood function with respect to each parameter and the partial derivative of the penalty function with respect to each parameter. The first derivative is an indication for convergence and the second derivative indicates wether the estimate is very close to one of the bounds on the parameters.
The pure simulation residuals have been plotted in Figure 2.3.

Generally, for these estimates, the correlation matrix shows limited correlations between parameters, except one pairing showing a correlation higher than 0.99, e.g. r_{1maxf} and r_{5maxf}. Only a few more pairs have a correlation higher than 0.5.

The performance of these estimates is explored by plotting the one step ahead and the pure simulation data. Figure 2.2 shows these for experimental data from Experiment 17A set 1.

	r_{1maxf}	r_{2maxf}	r_{3maxf}	r_{4maxf}	r_{5maxf}	r_{7maxb}	$ln(\sigma_{11})$	$ln(\sigma_{22})$	$ln(\sigma_{33})$	$ln(\sigma_{77})$	$ln(\sigma_1 010)$	S_{1111}	S_{22}	S_{33}	S_{44}
r_{1maxf}	1														
r_{2maxf}	-0.0553	1													
r_{3maxf}	-0.0042	-0.9630	1												
r_{4maxf}	-0.7843	0.4645	-0.5153	1											
r_{5maxf}	0.9973	-0.0174	-0.0383	-0.7727	1										
r_{7maxb}	-0.9598	0.1046	-0.0363	0.7683	-0.9542	1									
$ln(\sigma_{11})$	-0.0056	0.0058	-0.0038	0.0083	-0.0067	0.0082	1								
$ln(\sigma_{22})$	0.0044	0.0040	-0.0016	-0.0061	0.0046	-0.0033	0.0005	1							
$ln(\sigma_{33})$	0.0078	-0.0055	0.0033	-0.0110	0.0093	-0.0109	0.0045	-0.0000	1						
$ln(\sigma_{77})$	-0.0004	0.0057	-0.0056	0.0020	0.0003	0.0025	-0.0042	0.0039	0.0035	1					
$ln(\sigma_{1010})$	0.0005	0.0164	-0.0109	0.0020	-0.0002	0.0053	-0.0008	0.0046	0.0034	0.0008	1				
S_{11}	-0.3472	0.2093	-0.1755	0.3362	-0.3320	0.3361	-0.0012	0.0047	0.0010	0.0010	0.0062	1			
S_{22}	0.1959	-0.1698	0.1807	-0.2729	0.1909	-0.1743	-0.0036	0.0028	0.0058	-0.0008	-0.0025	-0.1199	1		
S_{33}	-0.0859	-0.0196	-0.0054	0.1134	-0.0864	0.0618	-0.0167	0.0053	0.0181	-0.0054	-0.0150	0.0610	-0.0490	1	
S_{44}	0.1879	0.0134	-0.0403	-0.1065	0.1861	-0.1761	-0.0145	0.0022	0.0216	0.0008	-0.0140	0.0178	0.0566	-0.0030	1

Table 2.7: Correlation matrix of the estimates for Model I using data from Experiment 17A, set 1 and 2, reactor I

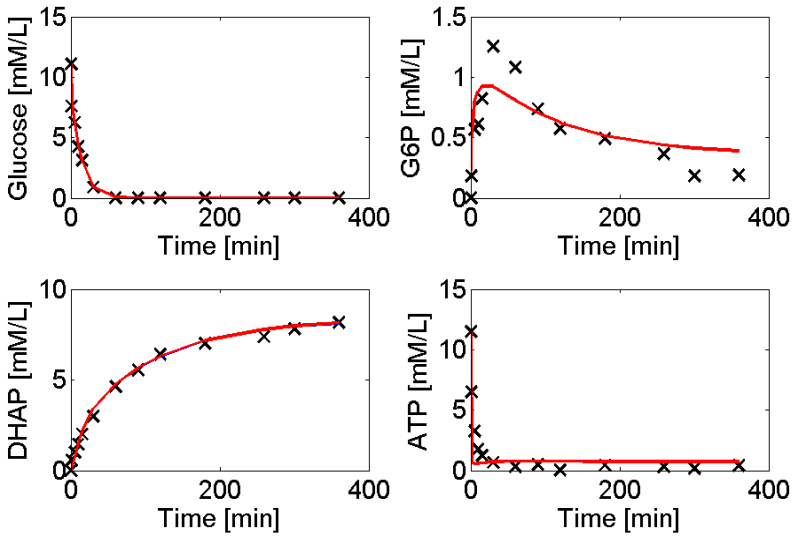


Figure 2.2: Simulation of Model I against Experiment 17A, set 1; pure simulation-red line, one step ahead prediction-blue line

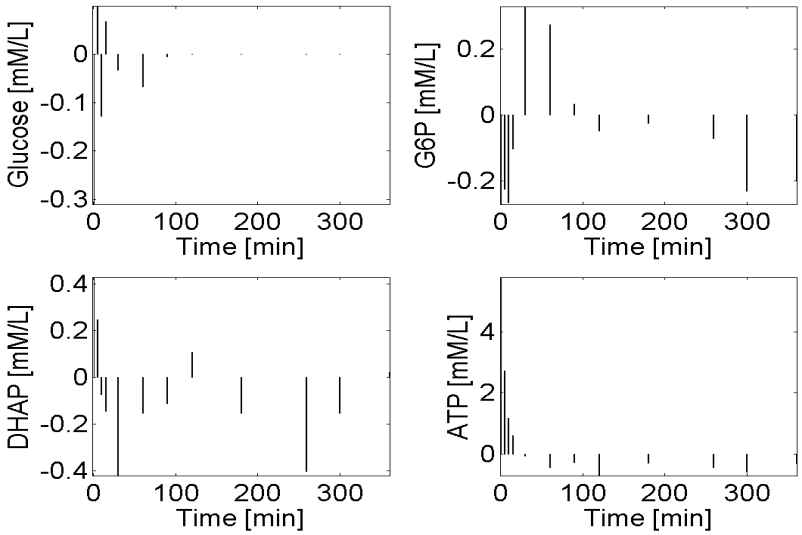


Figure 2.3: Residuals of Model I simulation for Experiment 17A, set 1

2.4.2 Discussion of Model I

Inspecting the estimates for the diffusion terms, σ, in Table 2.6 it can be noticed that are very small and based upon the t-test values are significant. Obtaining small values for the diffusion terms renders nearly no difference between one step ahead prediction and pure simulation data, as seen in the Figure 2.2. Having only a relatively low number of data points in the time series could hinder the correct evaluation of the diffusion terms and the t-test in fact can lead to a wrong conclusion about the significance. This is an indication that is necessary to have longer time series data to obtain a good estimate of the diffusion term. The measurement variances seems to be important as well, since their values are significant. Within the existing grey box stochastic modelling framework if a diffusion term is significant, that indicates that a model deficiency is present in the corresponding drift term of the stochastic differential equation. Thus the original methodology for grey-box stochastic modelling framework developed by Kristensen et al. (2004a) cannot be applied directly.

The main limitation of Model I, as can be seen in Fig. 2.2, concerns the ability to reproduce the dynamic trajectory for Glucose-6-Phosphate and ATP concentrations when performing pure simulations. Analyzing the residuals shown in Figure 2.3 is likely that there is at least a model deficiency in the balances for Glucose-6-Phosphate and for ATP. Since it is not very clear what should be the starting point in investigating these deficiencies in the way described in the original framework, a more general attempt to improve the model will be pursued in the following.

The following reasoning is used to motivate the model reformulation. It is known from the literature, (Chassagnole et al. 2002) that there are many more reactions in which ATP is not only cofactor but a direct reactant or product therefore in the reformulated model additional ATP reactions will be considered. By inspecting the mass balance for Glucose-6-Phosphate r_7 and r_1 seems to be deficient as well, therefore an improved kinetic expression for r_7 and r_1 will be investigated.

r_2 is a reaction catalyzed by *aldolase* and throughout the literature there has been a long debate about the mechanism of this enzymatic reaction thus all the available knowledge about this enzyme kinetics will be incorporated as well. Since the fourth reaction is a single reaction and kinetic model is available in the literature this knowledge is going to be incorporated as well in the model. There is no direct connection, however, between the obtained results and the fact that an improved kinetic rate shall be used for r_4.

2.5 Model II, Improved kinetics

Based upon the discussion given in section 2.4.2 the model is reformulated as described in the following. The reaction network considered for Model II is given in Figure 2.4.

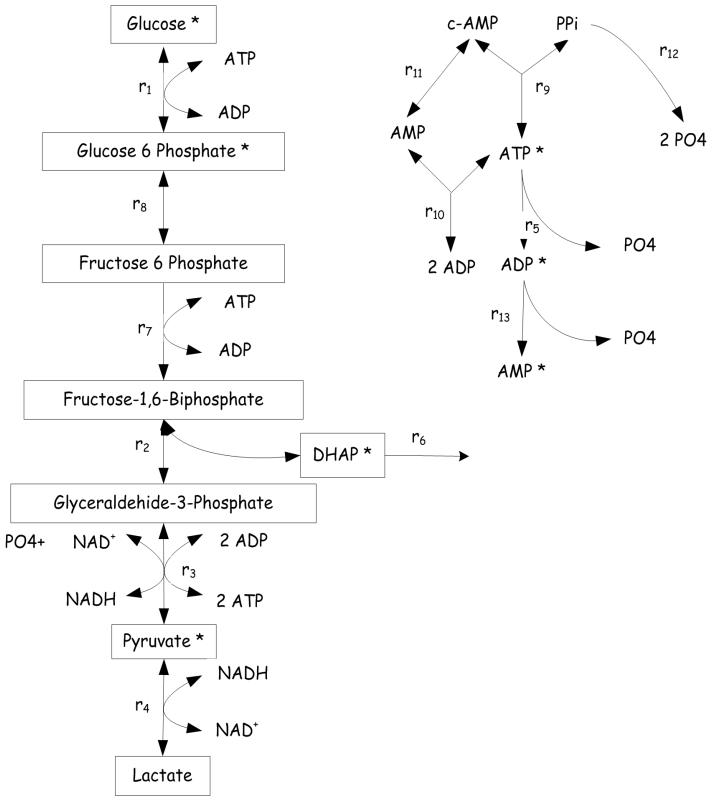


Figure 2.4: The simplified reaction network for Model II including the discussed additions

- In order to address the lack of fit for Glucose-6-Phosphate an extra reaction r_8, is included. This new reaction represents the isomerization of Glucose-6-Phosphate to Fructose-6-Phosphate which is catalyzed by *phosphoglucoisomerase*. A kinetic equation based on the kinetics presented by Chassagnole et al. (2002) has been included.

- More reactions related to the ATP chemistry have been incorporated in the reaction network as shown in the new reaction scheme in Figure 2.4. A reversible reaction where ATP is involved as reactant is catalyzed by the *adenylate cyclase* enzyme to form cyclic-AMP and Poly-phosphate. This reaction is reaction r_9 in Figure 2.4. *Adenylate kinase* is the enzyme catalyzing the reversible reaction of ATP production along with an AMP molecule from 2 molecules of ADP; in the network this is reaction r_{10}. A reaction degrading ADP to AMP and PO4 is included and denoted reaction r_{13}. Poly-phosphate is an unstable compound and therefore it is assumed that a decomposition reaction to PO4 occurs as reaction r_{12}. An extra reaction r_{11} converting cyclic-AMP into AMP is included, although there is only limited evidence in the literature for its presence.

- The reaction r_1 is catalyzed by *hexokinase*. Several authors, Bali and Thomas (2001), Teusink et al. (2000), Viola et al. (1982), Ning et al. (1969), Hanson and Fromm (1967), have investigated the kinetic mechanism for *hexokinase* at physiological conditions. It seems plausible that the reaction kinetics is described by a rapid equilibrium random mechanism. Similar values for the saturation constants K_{ADP}, K_{ATP}, and K_{GL} as well as K_{G6P} have been found for different types of *hexokinases*. Since *hexokinase* is not naturally present in *E-coli*, is added in a small amount to the SBT. As for the kinetic parameters, the values from the literature have been used to define the search space and the mean value of the interval as initial value for parameter estimation.

- For the reaction producing DHAP, which is catalyzed by *aldolase* a more complex kinetic model considering saturation for Glyceraldehide-3-Phosphate is considered. Since all literature sources found referred either to isolated purified enzyme Zgiby et al. (2000), (2002) or the in-vivo behavior (Teusink et al. (2000), Rizzi et al. (1996)), only an educated guess for the parameters e.g. saturation constants, inhibition constants, maximal velocities, could be established. The values found in the literature have been used to establish the search interval for estimation where a mean value is used as initial guess.

- Now, r_7 concerns transformation from Fructose-6-Phosphate to Fructose-1,6-Bisphosphate. This reaction is catalyzed by *phosphofructokinase kinase (PFK)*. It is known that the enzyme acts in a complex way by being inhibited by Phospho-enol-Pyruvate, ADP and AMP when operating under *in-vivo* conditions. Since these compounds are not measured and the

enzyme is cell free, only a simplified kinetic equation is considered. The starting point for the simplification was the kinetic equation from Chassagnole et al. (2002), and the results of simplifications are given in the kinetic Eq. 2.24.

- Reaction four r_4, corresponds to a single transformation catalyzed by *lactate-dehydrogenase (LDH)* it is possible to find kinetics (Zewe and Fromm 1965) which have been included in Model II.

The saturation constant for a compound is assumed to be the same for all the enzymes involved. The main reason for this simplification is that for most of the saturation constants the numerical values found in the literature are fairly similar.

After implementing the modifications and additions discussed above the modified kinetic equations are given in Eqs. 2.24, while completed stochastic differential equations are given in Eqs. 2.25.

$$
\begin{aligned}
r_1 &= \frac{(r_{1maxf}\frac{c_{GL}c_{ATP}}{K_{GL}K_{ATP}} - r_{1maxb}\frac{c_{G6P}c_{ADP}}{K_{GL}K_{ATP}K_{eqHK}})}{(1+\frac{c_{GL}}{K_{GL}}+\frac{c_{G6P}}{K_{G6P}})(1+\frac{c_{ATP}}{K_{ATP}}+\frac{c_{ADP}}{K_{ADP}})} \\
r_2 &= \frac{r_{2maxf}\frac{c_{F16B}}{K_{F16B}} - r_{2maxb}\frac{c_{DHAP}c_{G3P}}{K_{F16B}K_{eqALD}}}{1+\frac{c_{F16B}}{K_{F16B}}+\frac{c_{DHAP}}{K_{DHAP}}+\frac{c_{G3P}}{K_{G3P}}+\frac{c_{F16B}c_{G3P}}{K_{F16B}K_{iG3P}}+\frac{c_{DHAP}c_{G3P}}{K_{DHAP}K_{G3P}}} \\
r_3 &= r_{3maxf}c_{G3P}c_{NAD}c_{PO4}c_{ADP} - r_{3maxb}c_{PYR}c_{NADH}c_{ATP} \\
r_4 &= \frac{r_{4maxf}\frac{c_{PYR}c_{NADH}}{K_{PYR}K_{NADH}} - r_{4maxb}\frac{c_{LAC}c_{NAD}}{K_{PYR}K_{NADH}K_{eqLDH}}}{(1+\frac{c_{NAD}}{K_{NAD}}+\frac{c_{NADH}}{K_{NADH}})(1+\frac{c_{LAC}}{K_{LAC}}+\frac{c_{PYR}}{K_{PYR}})} \\
r_5 &= r_{5maxf}c_{ATP} \\
r_6 &= r_{6maxf}c_{DHAP} \\
r_7 &= \frac{r_{7maxf}c_{F6P}c_{ATP}}{((c_{ATP}+K_{ATP}(1+\frac{c_{ADP}}{K_{ADP}}))(c_{F6P}+K_{F6P}c_{AMP}))} \\
r_8 &= \frac{r_{8maxf}\frac{c_{G6P}}{K_{G6P}} - r_{8maxb}\frac{c_{F6P}}{K_{G6P}K_{eqPGI}}}{1+\frac{c_{G6P}}{K_{G6P}\left(1+\frac{c_{G6P}}{K_{F6PG6P}}\right)}+\frac{c_{F6P}}{K_{F6P}}} \\
r_9 &= r_{9maxf}c_{ATP} - r_{9maxb}c_{PPi}c_{cAMP} \\
r_{10} &= r_{10maxf}c_{ADP}^2 - r_{10maxb}\frac{c_{ATP}c_{AMP}}{K_{eqAK}} \\
r_{11} &= r_{11maxf}c_{cAMP} - r_{11maxb}c_{AMP} \\
r_{12} &= r_{12maxf}c_{PPi} \\
r_{13} &= r_{13maxf}c_{ADP}
\end{aligned}
\tag{2.24}
$$

$$
\begin{aligned}
dc_{GL} &= (-r_1)\,dt + \sigma_{11}d\omega \\
dc_{F16B} &= (r_7 - r_2)\,dt + \sigma_{22}d\omega \\
dc_{DHAP} &= (r_2 - r_6)\,dt + \sigma_{33}d\omega \\
dc_{G3P} &= (r_2 - r_3)\,dt + \sigma_{44}d\omega \\
dc_{PYR} &= (r_3 - r_4)\,dt + \sigma_{55}d\omega \\
dc_{LAC} &= (r_4)\,dt + \sigma_{66}d\omega \\
dc_{ATP} &= (-r_1 - r_7 + 2r_3 - r_5 - r_9 + r_{10})\,dt + \sigma_{77}d\omega \\
dc_{NAD} &= (-r_3 + r_4)\,dt + \sigma_{88}d\omega \\
dc_{PO4} &= (-r_3 + r_5 + 2r_{12} + r_{13})\,dt + \sigma_{99}d\omega \\
dc_{G6P} &= (r_1 - r_8)\,dt + \sigma_{1010}d\omega \\
dc_{ADP} &= (r_1 + r_7 - 2r_3 + r_5 - 2r_{10})\,dt + \sigma_{1111}d\omega \\
dc_{NADH} &= (r_3 - r_4)\,dt + \sigma_{1212}d\omega \\
dc_{F6P} &= (r_8 - r_7)\,dt + \sigma_{1313}d\omega \\
dc_{AMP} &= (r_{10} + r_{11} + r_{13})\,dt + \sigma_{1414}d\omega \\
dc_{cAMP} &= (r_9 - r_{11})\,dt + \sigma_{1515}d\omega \\
dc_{PPi} &= (r_9 - r_{12})\,dt + \sigma_{1616}d\omega
\end{aligned}
\tag{2.25}
$$

Given the model as described above, a preliminary identifiability analysis based on sensitivity analysis has been performed in order to determine which parameters can be identified. Once a set of potentially identifiable parameters has been established the parameters have been estimated using the CTSM program (Kristensen et al. 2004a). Several subsets of parameters have been considered for estimation. In some cases the estimation was repeated starting from a different initial solution. Experimental data from Experiment 17A set 1 and 2 have been used for these estimations. The statistical tests described in Section 2.2.3 have been applied again.

Table 2.8 contains the value of the log-likelihood function after estimating the parameters in dependence of the number of estimated parameters in each case. The log-likelihood ratio test has been applied in order to determine how many parameters should be estimated form a statistical point of view.

First the hypothesis H_0 was: the θ parameter vector is represented by *set 3*, while the alternative hypothesis H_1 was: the θ parameter vector is represented by *set 7*. The alternative hypothesis H_1 was true. After testing *set 2* against *set 7* it turns out that the alternative hypothesis *set 7* with a smaller number of parameters is true. Finally when testing *set 7* against *set 6* the hypothesis having less parameters *set 7* is valid. Thus the results obtained when estimating *set 7* will be discussed in the following. The estimation for *set 7* is given in tables 2.10 – 2.13.

The estimation statistics when estimating *set 7* is given in Table 2.10. For this run the ML estimation scheme has been used thus no prior term was added

Set	$LogL$	$Npar$
1	-6.483216320707882E+01	16
2	-6.483224308069573E+01	17
3	-6.483209530254184E+01	18
4	-6.475481101023928E+01	14
5	-6.476421561113025E+01	15
6	-6.575493616769320E+01	15
7	-6.481667750801653E+01	13
8	-6.481958029720100E+01	14
9	-6.483215873045927E+01	15
10	-6.483216320707882E+01	16

Table 2.8: Log-likelihood function value for the estimation results for Model II when estimating different number of parameters, using data from Experiment 17A, data set 1 and 2, reactor I

H_1/H_0	$\lambda(Y)$	$-2log\lambda(Y)$	$\#\theta_0 - \#\theta_1$	$\chi^2(\#\theta_0 - \#\theta_1, 0.95)$	$True\ hypothesis$
7/3	0.99976	2.0848e-4	5	11.0705	H_1, 7
7/2	0.99976	2.0848e-4	4	9.4877	H_1, 7
7/6	0.98573	0.01248	2	5.9915	H_1, 7

Table 2.9: Loglikelihood ratio test applied for estimation results of Model II using data from Experiment 17A, set 1 and 2, reactor I

Value of objective function	-6.481667750801653E+01
Value of penalty function	1.797273227015682E-04
Negative logarithm of determinant of Hessian	-1.150425812314553E+02
Number of iterations	46
Number of objective function evaluations	70

Table 2.10: Estimation Results for Model II, using data from Experiment 17A, data set 1 and 2, reactor I

in the log-likelihood function. The cost function at the optimal point is much lower than for Model I, see Table 2.4.

The fixed parameters of Model II are listed in Table 2.12. The initial values of the compounds which are added to the reactor are known and were assigned to the model states. As for Model I, for the species which are formed during the reactions the initial concentration was set to zero, or a very small number. The diffusion terms for the unmeasured compounds were set to a small value. The saturation constants for the enzymes have been taken from the literature as mention above. For the maximum velocities, the values used were either taken from the previous runs and then fixed or educated guesses have been used.

Analyzing the results of the estimates and the t- test statistics in Table 2.11 it is clear that the estimated parameters are significant and the uncertainty of the estimates is small. Analyzing the parameter correlations in Table 2.13

Name		Estimate	Std. dev.	t-score	$p(>\|t\|)$	dF/dPar	dPen/dPar
r_{1maxf}	ML	1.0549E+01	2.5235E-01	41.8049	0.0000	0.0007	0.0000
r_{2maxf}	ML	1.1136E+00	2.5257E-02	44.0902	0.0000	0.0009	0.0001
r_{5maxf}	ML	2.3145E-01	5.9657E-03	38.7960	0.0000	0.0004	0.0000
r_{8maxb}	ML	1.0294E+00	5.4740E-02	18.8056	0.0000	0.0002	0.0000
K_{eqPGI}	ML	3.1166E-01	1.0314E-02	30.2173	0.0000	-0.0008	0.0001
σ_{11}	ML	2.7117E-06	8.0005E-07	3.3894	0.0010	0.0000	0.0000
σ_{22}	ML	7.0537E-05	2.1971E-05	3.2105	0.0018	-0.0000	0.0000
σ_{33}	ML	3.4870E-02	7.3341E-03	4.7544	0.0000	-0.0000	0.0000
σ_{1010}	ML	6.9409E-05	1.9851E-05	3.4965	0.0007	0.0000	0.0000
S_{11}	ML	5.5295E-02	1.1962E-02	4.6224	0.0000	0.0000	0.0000
S_{22}	ML	8.8041E-03	1.5465E-03	5.6929	0.0000	0.0001	0.0000
S_{33}	ML	9.8563E-03	3.2709E-03	3.0133	0.0033	0.0001	0.0000
S_{44}	ML	3.2140E-02	6.2343E-03	5.1555	0.0000	0.0000	0.0000

Table 2.11: Estimation results for *set 7* from Table 2.9 part 1

Name	Parameter	Name	Parameter	Name	Parameter
c_{GL0}	11.1	K_{ADP}	0.14524	r_{9maxf}	0.0020
c_{F16B0}	1.0E-5	K_{F16B}	0.88	r_{9maxb}	0.0010983
c_{DHAP0}	1.0E-5	r_{2maxb}	0.050572	r_{10maxf}	0.019706
c_{G3P0}	1.0E-5	K_{eqALD}	0.4395	r_{10maxb}	0.0010
c_{PYR0}	1.0E-5	K_{DHAP}	1.22	K_{eqAK}	0.35
c_{LAC0}	1.0E-5	K_{G3P}	1.0	r_{11maxf}	1.0E-8
c_{ATP0}	11.5	K_{iG3P}	0.325	r_{11maxb}	1.0E-8
c_{NAD0}	5.75	r_{3maxf}	3.1115E-4	r_{12maxf}	1.0E-5
c_{PO40}	11.1	r_{3maxb}	0.0010	r_{13maxf}	1.0E-5
c_{G6P0}	1.0E-5	r_{4maxf}	0.0044597	σ_{44}	1.0E-5
c_{ADP0}	1.0E-5	K_{PYR}	0.854	σ_{55}	1.0E-5
c_{NADH0}	1.0E-5	K_{NADH}	0.044	σ_{66}	1.0E-5
c_{F6P0}	1.0E-5	r_{4maxb}	0.012525	σ_{77}	1.0E-5
c_{AMP0}	1.0E-5	K_{eqLDH}	10.0	σ_{88}	1.0E-5
c_{cAMP0}	10.0	K_{NAD}	1.285	σ_{99}	1.0E-5
c_{PPi0}	10.0	K_{LAC}	50.0	σ_{1111}	1.0E-5
K_{GL}	0.1	r_{6maxf}	1.0E-8	σ_{1212}	1.0E-5
K_{ATP}	1.5733	r_{7maxf}	3.9651	σ_{1313}	1.0E-5
r_{1maxb}	0.1	K_{F6P}	0.28047	σ_{1414}	1.0E-5
K_{eqHK}	7250.0	r_{8maxf}	15.309	σ_{1515}	1.0E-5
K_{G6P}	0.090873	K_{F6PG6P}	0.2	σ_{1616}	1.0E-5

Table 2.12: Fixed parameters value for Model II

it can be noticed that there are still four pairs having correlation higher than 50 %, that is the pair r_{1maxf} –r_{5maxf}, r_{8maxb}–K_{eqPGI}, r_{5maxf}–r_{8maxb} and r_{8maxb}–S_{33}. But these are all below 70 %. The performance of the model is illustrated in the following by plotting both the one step ahead prediction (blue line) and pure simulation data (red line) in Figure 2.5 and comparing these to experimental data from Experiment 17A set 1. It can be seen that there is a

	r_{1maxf}	r_{2maxf}	r_{5maxf}	r_{8maxb}	Keq_{PGI}	σ_{11}	σ_{22}	σ_{33}	σ_{1010}	S_{11}	S_{22}	S_{33}	S_{44}
r_{1maxf}	1												
r_{2maxf}	0.1576	1											
r_{5maxf}	0.5736	0.2425	1										
r_{8maxb}	-0.2814	-0.4101	-0.5171	1									
Keq_{PGI}	0.0068	-0.4743	-0.0999	0.6992	1								
σ_{11}	0.1167	-0.2215	0.0610	-0.0716	0.1055	1							
σ_{22}	-0.1180	0.1514	-0.0775	0.1318	0.0269	-0.0399	1						
σ_{33}	0.1137	0.3724	0.2888	-0.3655	-0.2758	-0.0016	-0.0686	1					
σ_{1010}	-0.0211	-0.1985	-0.0146	-0.0556	-0.1073	-0.0316	-0.0798	0.0257	1				
S_{11}	-0.2474	0.0318	-0.1156	0.2908	0.2765	-0.1380	0.1702	-0.1443	0.0827	1			
S_{22}	0.1061	-0.3576	0.2112	0.1102	0.1439	-0.0511	0.0972	-0.1715	0.0904	0.1342	1		
S_{33}	-0.3506	-0.4042	-0.4088	0.6041	0.4841	0.1500	0.0793	-0.3027	0.0051	0.2815	0.2675	1	
S_{44}	0.3450	0.0056	0.2632	-0.1827	0.1775	0.2235	-0.0376	0.2175	0.1334	-0.2639	-0.1243	-0.0441	1

Table 2.13: Correlation matrix for *set 7* from Table 2.9

slight difference between one step ahead and pure simulation data for DHAP. The diffusion term, σ_{33} for this equation is much larger than any other values. This indicates a possible model deficiency in the drift term of this equation, however the difference is rather small.

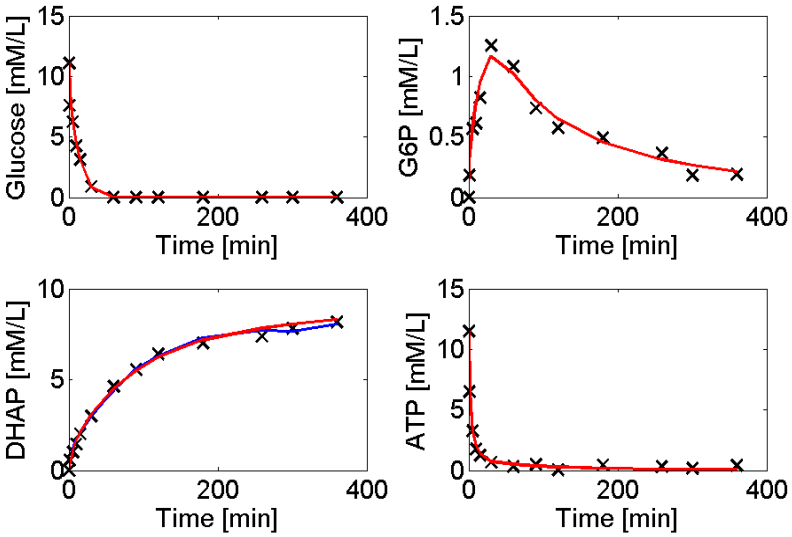


Figure 2.5: Simulation of Model II against Experiment 17A set 1; pure simulation-red line, one step ahead prediction-blue line

The residuals of this model for pure simulation data (red line) have been plotted in Figure 2.6. For this fit the sequence of residuals looks more like white noise compared to the residuals for obtained for pure simulation data (red line) of Model I plotted in Figure 2.3. Thus the estimation has extracted the available information in the data.

2.5.1 Validation of Model II

In order to assess the validity of Model II described above, a validation test is performed using the estimated parameters listed in Table 2.11. The results are shown in Figure 2.7 as plots of the pure simulation (red line) and one step ahead prediction (blue line).

Experimental data obtained during Experiment 17B set 1 have been used for validation while Experiment 17A set 1 and 2 were used for estimation. Thus, the validation has been performed on data obtained under experimental conditions where 50% of the intial ATP was used. Analyzing the graphical results it seems as if there still is a small systematic error either in the model structure or in the model parameters for Glucose-6-Phosphate.

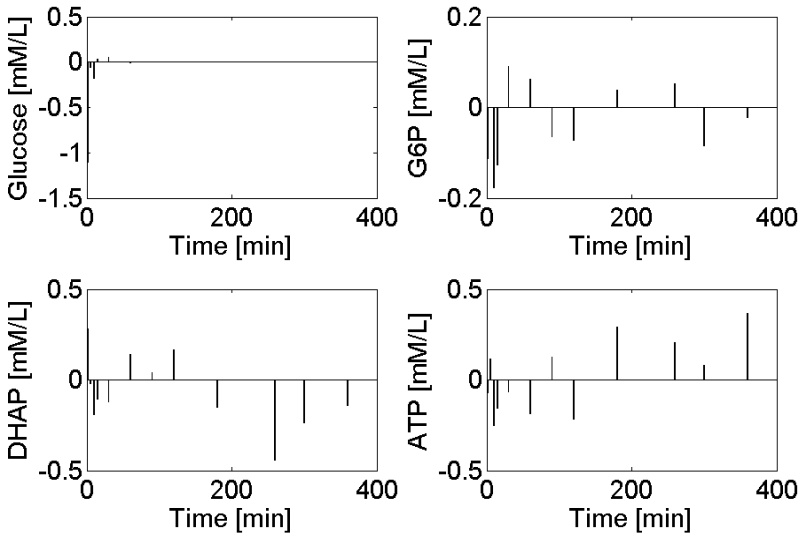


Figure 2.6: Residuals for Model II simulation and Experiment 17A set 1

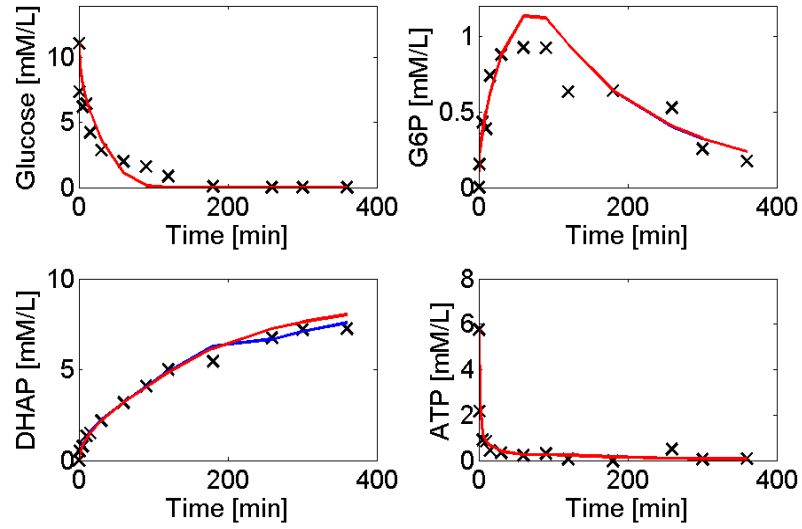


Figure 2.7: Validation of Model II using data from Experiment 17B set 1; pure simulation-red line, one step ahead prediction-blue line

2.5.2 Multiple sets of experimental data

In order to improve the prediction performance of Model II experimental data obtained from different experimental data sets are used for estimating the parameters of the model. It was assumed that the model structure as well as the fixed parameters are correct, which means that the diffusion term and the measurements variations will be estimated as having a single global value for the two different experiments. Experimental data obtained during Experiment 17A and B, four time series of data each with four measured variables have been used for estimation.

The second assumption is that some of the reactions occur at different maximal velocities depending on the initial amount of ATP present in the system, which means that the corresponding parameters will have different values and they will be estimated locally for each experiment.

Value of objective function	-76.32745746
Value of penalty function	0.00038834
Negative logarithm of determinant of Hessian	-184.38995603
Number of iterations	128
Number of objective function evaluations	185

Table 2.14: Estimation results for Model II using data from Experiment 17A set 1 and 2, and Experiment 17B set 1 and 2

Table 2.14 shows the statistic of the estimation for the case where both Experiments 17A and B have been used for estimation. Comparing the value of the minimized negative logarithm of the log-likelihood function with the case where only data from Experiment 17 A was used, see Table 2.10, then a smaller value was obtained when using more experimental data, thus it makes sense to use experimental data obtained under different experimental conditions. The estimation results obtained are given in Table 2.15.

The pure simulation prediction together with the one step prediction data are plotted for both experiments below in Figures 2.8–2.9.

The first reaction influenced by ATP is reaction r_1. Analyzing the parameter estimates from Table 2.15, it can be seen that the estimate for r_{1maxf} for Experiment 17A is nearly half of r_{1maxf} for Experimental 17B. Another reaction where ATP is involved is reaction r_5. The parameter estimates related to this reaction rate are fairly similar. The parameter estimates for r_{2maxf}, in Experiment 17B is approximately half of the value for Experiment 17A. It seems that the reactions where ATP is involved are influenced by the initial concentration. The reaction producing DHAP is clearly influenced as well.

All the estimated diffusion terms when estimated using only the experimental data stemming from a single experiment, Experiment 17A, see Table 2.11, significant. However after estimating them using experimental data obtained from two different experiments only σ_{33} remained significant while the rest became insignificant, see Table 2.15. This observation, together with the graphical

Name		Estimate	Std. dev.	t-score	$p(>\|t\|)$	dF/dPar	dPen/dPar
r_{1maxf}	ML	0.627865E+01	0.238603E+00	26.3142	0.0000	0.59E-06	0.11E-03
r_{2maxf}	ML	0.121157E+02	0.643749E+01	1.8821	0.0613	0.80E-07	0.24E-04
r_{5maxf}	ML	0.321098E+00	0.759980E-02	42.2508	0.0000	0.44E-06	0.34E-05
r_{8maxf}	ML	0.180458E+02	0.737216E+00	24.4783	0.0000	-0.11E-05	0.56E-04
σ_{11}	ML	0.402606E-23	0.454404E-19	0.0001	0.9999	-0.24E-10	-0.25E-10
σ_{22}	ML	0.448364E-06	0.319173E-03	0.0014	0.9989	0.33E-09	0.90E-11
σ_{33}	ML	0.590991E-01	0.599582E-02	9.8567	0.0000	0.50E-06	0.12E-05
σ_{1010}	ML	0.443131E-08	0.508128E-05	0.0009	0.9993	-0.30E-09	0.89E-13
S_{11}	ML	0.664381E+00	0.139934E+00	4.7478	0.0000	-0.53E-07	0.76E-05
S_{22}	ML	0.106647E-01	0.208694E-02	5.1102	0.0000	-0.40E-06	0.11E-06
S_{33}	ML	0.724092E-21	0.204366E-17	0.0004	0.9997	0.00E+00	-0.14E-12
S_{44}	ML	0.293980E-01	0.630407E-02	4.6633	0.0000	-0.45E-06	0.30E-06
r_{1maxf}	ML	0.916814E+01	0.607950E+00	15.0804	0.0000	-0.14E-05	0.31E-03
r_{2maxf}	ML	0.685951E+01	0.216262E+01	3.1718	0.0018	-0.34E-06	-0.45E-05
r_{5maxf}	ML	0.385214E+00	0.195990E-01	19.6548	0.0000	0.74E-05	0.42E-05
r_{8maxf}	ML	0.189876E+02	0.117089E+01	16.2164	0.0000	-0.25E-05	0.63E-04

Table 2.15: Estimation results for Model II using data from Experiment 17A set 1 and 2, and Experiment 17B set 1 and 2

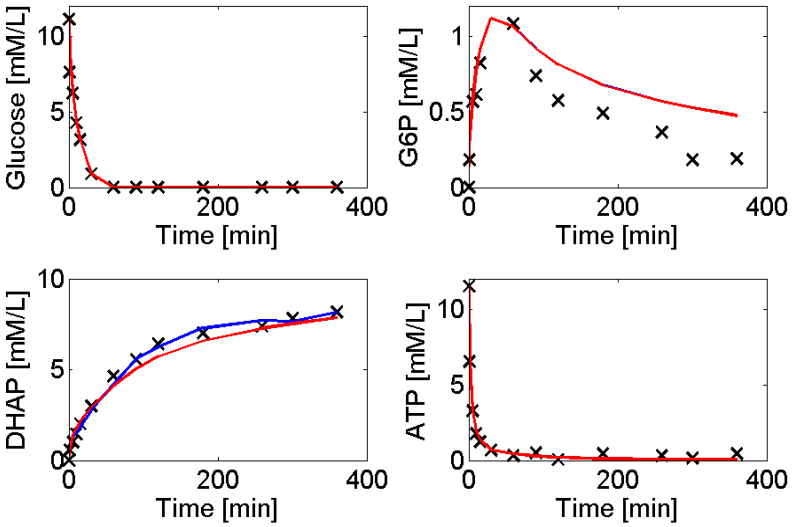


Figure 2.8: Simulation of Model II against data from Experiment 17A set 1; pure simulation-red line, one step ahead prediction-blue line

results from Figures 2.8–2.9 points to some model deficiency present in the balance for DHAP.

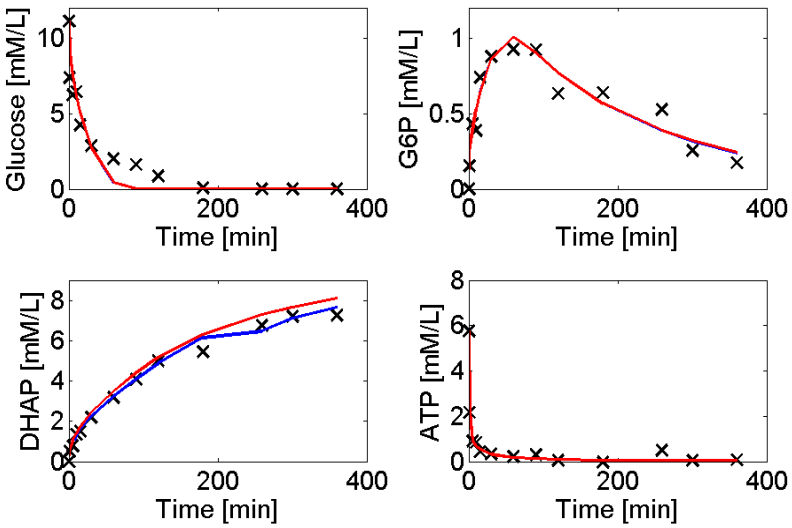


Figure 2.9: Simulation of Model II against data from Experiment 17B set 1; pure simulation-red line, one step ahead prediction-blue line

2.6 Discussion on Model II

Generally, the modifications proposed in section 2.5 improved the pure simulation prediction properties of the model. Thus Model II clearly represents an improvement over Model I. The validation performed in section 2.5.1 showed improved performances since the pure simulation obtained for a slightly different experiment (Experiment 17 B) are fairly good. Analyzing the plots in Figure 2.7, there is hardly any difference between one step ahead and pure simulation prediction. The Glucose and Glucose-6-Phosphate outputs indicate a systematic deviation from the experimental data indicating that a model deficiency may still be present. This can be noticed when comparing the plots in Figure 2.7 with Figures 2.8–2.9. Similar systematic deviations are observed for the Glucose and Glucose-6-Phosphate outputs.

When estimating the parameters using experimental data from two different experiments, Experiment 17 A and B and considering some of the model parameters to have different values for each of the experimental conditions showed a slight improvement in the fit. Interestingly, most of the diffusion terms which have been estimated globally for all experiments became insignificant, while σ_{33} still remained significant. This fact reduces the doubts about model deficiency of the corresponding mass balance. Moreover this was reflected in the plots for the two experiments where there is a difference between one step ahead prediction and pure simulation, Figures 2.8–2.9.

Note that the parameter estimates for r_{1maxf}, and r_{2maxf} obtained by estimating them individually for each experiment are indeed different.

Based upon these remarks, it is obvious that there is a need for performing experimental design to improve the data available for estimation of kinetic parameters. Such an experimental design may be base upon obtaining qualitative information on structural parameter identifiability prior to carrying out quantitative experimental design thereby obtaining a sound basis for decision making.

2.7 Conclusions

Two grey-box stochastic models have been developed using the grey-box stochastic modelling methodology including the available experimental data and knowledge. The grey-box approach used facilitates the use of statistical procedures for selecting the most reasonable model while considering the experimental data.

Regarding the grey-box stochastic modelling methodology, some limitations have been encountered. The first limitation observed during its application is that it requires some minimum number of data points available in the measured time series. This minimum number depends on the case. However it is clearly related to the number of measured compounds and the number of unknown parameters in the kinetic model. This was noticed particulary to be related to the difficulty of estimating the diffusion terms.

A second limitation concerns obtaining parameter estimates which are both significant and uncorrelated as a prerequisite to apply the tools available in the grey-box stochastic modelling methodology. It takes sometimes many iterations just to get a set of estimated uncorrelated parameters and only then it is possible to meaningfully apply the significance test to the diffusion terms in order to check for model deficiencies.

Clearly some desirable and useful features have been identified. The key feature is related to the possibility of determining *a-priori* which parameters can at least be potentially estimated from the available experimental data. Such a tool for qualitative identifiability analysis may prove very useful. This leads directly to a second desirable feature which concerns the possibility of subdividing experimental design into a qualitative and a quantitative phases. Thereby facilitating clarifying some structural aspects before executing the computationally more burdensome quantitative experimental design in order to facilitate reliable estimation of the model parameters.

3

Structural parameter identifiability analysis for dynamic reaction networks

Abstract

A fundamental problem in model identification is to investigate whether unknown parameters in a given model structure potentially can be uniquely recovered from experimental data. This issue of global or structural identifiability is essential during nonlinear first principles model development where for a given set of measured variables it is desirable to investigate which parameters may be estimated prior to spending computational effort on the actual estimation. This contribution addresses the structural parameter identifiability problem for the typical case of reaction network models. The proposed analysis is performed in two phases. The first phase determines the structurally identifiable reaction rates based on reaction network stoichiometry. The second phase assesses the structural parameter identifiability of the specific kinetic rate expressions using a generating series expansion method based on Lie derivatives. The proposed systematic two phase methodology is illustrated on a mass action based model for an enzymatically catalyzed reaction pathway network where only a limited set of variables is measured. The proposed two phase methodology clearly pinpoints the structurally identifiable parameters in dependence of the given measurements and input perturbations.

3.1 Introduction

Process models are applied increasingly for design, control, optimization and risk analysis within bio- and chemical engineering. For many such applications, which include batch and fed-batch operation, it is highly desirable to develop quantitative dynamic process models with good long term prediction properties over a wide operating region. During process model development, an important issue is to determine a suitable model structure, e.g. by selection

from a set of possible structures for a given application purpose. To facilitate model selection it is necessary to evaluate the model performance and for this purpose it is necessary to estimate the model parameters. However, estimation of model parameters may be quite time consuming. Consequently analysis of the properties of the parameter estimation problem may be highly beneficial prior to both experimental design and to undertaking actual estimation of model parameters in practice. Therefore it is desirable to evaluate whether the parameters of a given model are structurally identifiable.

A property is said to be *structural* or *generic* if it is true for almost any parameter value excluding some atypical hyper surface and possibly false on a parametric subspace of zero measure. Such a property may also be labeled *qualitative* as opposed to quantitative. When performing *a priori* or structural identifiability analysis, following Walter and Pronzato (1997), it is possible to distinguish between the following types of identifiability:

- Global identifiability: The structural identifiability analysis problem yields a unique solution
- Local identifiability: The structural identifiability analysis problem yields multiple solutions
- Unidentifiable: The structural identifiability analysis problem does not yield any feasible solution

In contrast, during quantitative parameter estimation, an *a posteriori* analysis may be performed, based on collected experimental information, where however only local identifiability results may be obtained. Thus practical identifiability analysis is defined as *a posteriori* quantitative analysis based upon numerical parameter estimation.

While structural parameter identifiability only has received some attention in the literature, there has been a substantial interest in *practical* identifiability analysis. Here the main literature on structural identifiability of nonlinear dynamics systems is reviewed while only related approaches based upon practical identifiability are covered. Subsequently, identification of reaction networks is reviewed before the purpose of this article is stated. One approach to address the structural identifiability problem is based on state isomorphism's, which uses the fact that under certain conditions, indistinguishable state space systems have locally isomorphic state spaces. If indistinguishable state space systems exist then one can parameterize the equivalence classes by the admissible state space isomorphisms. If an identity can be shown, then global identifiability follows under certain conditions (Peeters and Hanzon 2005).

A second approach to structural identifiability analysis is based on differential algebra (Ljung and Glad 1994), which has been further developed by Audoly et al. (2001), Saccomani et al. (2003) and Saccomani (2004). A set of algebraic equations denoted the exhaustive summary is obtained, and solved by algebraic methods e.g. the Buchberger algorithm.

The third approach to structural identifiability analysis reviewed here is based on power series expansion of the model outputs as a function of inputs and time (Pohjanpalo 1978), (Fliess 1980), (Fliess and Lagarrigue 1980) and (Walter and Pronzato 1996). This approach has been applied by e.g. Dochain et al. (1995) and Petersen (2000). Two types of expansions may be used, one is based on Taylor series and the other based on generating series. In the case of Taylor series, the output vector and its time derivatives are typically developed around the initial time. Successive time derivatives starting with the $zero^{th}$ order term and going up to the n^{th} order (where n is the number of parameters) are used to form an algebraic equation system. Then identifiability is assessed by investigating whether the algebraic equation system is solvable symbolically, by determining the number of solutions for the parameter set under investigation (Walter and Pronzato 1996). In the generating series case, the model output is expanded into a series whose coefficients are computed using Lie derivatives. A set of algebraic equations corresponding to the number of unknown parameters is formed to investigate structural identifiability. In a more recent paper, (Walter et al. 2004), a guaranteed optimization based method is presented. The classical definitions of identifiability are slightly modified by defining a valid domain for the model parameters.

In the case of practical, or *a-posteriori* identifiability, several methods are available. The first method is the local or global (multi local) sensitivity analysis (Brun et al. 2002), (Sarmiento Ferrero et al. 2006) and (Kontoravdi et al. 2005), which is a widely used method for large models. The second method for establishing practical identifiability, is an optimization-based approach, proposed by Asprey and Macchietto (2000). In principle, the idea is to maximize the distance between two parameter vectors that essentially give the same model output. If the maximized distance is smaller than some threshold then the model is deemed identifiable. In a more recent paper, the optimization based approach is modified and combined with the multi-local sensitivity analysis into what is called the perturbation algorithm (Sidoli et al. 2005).

For identification of complex reaction networks Brendel et al. (2006) developed a systematic approach, where the model development is decomposed into steps. Each subsequent step is related to only one part of the model. The steps included are: stoichiometry, reaction rates, reaction kinetics and kinetic parameters. However the analysis of structural identifiability of reaction kinetic parameters is not considered. Instead, all possible model candidates are considered. After estimating the parameters for each model candidate the statistically insignificant model parameters or model candidates are eliminated. Another potential drawback is that each individual reaction rate is considered independently, thus interaction effects between reactions are neglected.

The main advantage of the structural identifiability analysis methods based on the differential algebra or power series expansions is that these methods may provide a global identifiability test. Another advantage is that they provide insight concerning the nonlinearities of the dynamic model. The applicability

of methods based on differential geometry for real life applications such as enzymatic reaction networks is limited by the fact that the complexity of the derived equations grows exponentially, when the number of model states and parameters grows. On the other hand, methods based on sensitivity analysis or a perturbation study tend to be more applicable to real life applications, but they require a very substantial computational effort. The present article aims at combining an extension of structural parameter identifiability analysis with a structural rate identifiability analysis based upon stoichiometry analysis (Bonvin and Rippin 1990) to obtain a methodology which provides information on qualitative parameter identifiability for specific models.

The article is organized with a Section 3.2 presenting the methodology and the developed algorithm. Section 3.3.1 describes a model for an enzymatic reaction network on which the application of the methodology is illustrated in Section 3.3.2. In Section 3.5 the results and their limitations are discussed and the conclusions are drawn in Section 3.6.

3.2 Methodology

The methodology implemented in this work combines identifiability analysis of reaction rates in dependence of the available measurements with a structural parameter identifiability analysis for the parameters in the proposed kinetic expressions for the identifiable reaction rates. The background for the two types of analysis is outlined first, and the work flow of the global identifiability algorithm is presented second.

3.2.1 Reaction rate identifiability

Following the notation of Brendel et al. (2006), the stoichiometric matrix for the reaction network is denoted $N \in \mathbf{R}^{n_R \times n_S}$, the part corresponding to chemical species being measured is denoted $N_m \in \mathbf{R}^{n_R \times n_{Sm}}$, and for the unmeasured species $N_u \in \mathbf{R}^{n_R \times n_{Su}}$. The matrix containing the true reaction rates is denoted $R \in \mathbf{R}^{n_Q \times n_R}$. The matrix containing the measured reaction fluxes is denoted $F_m^r \in \mathbf{R}^{n_Q \times n_{Sm}}$.

V_o is a diagonal matrix containing the reactor volume measurements ($n_Q \times n_Q$), n_Q being the number of samples.

For the case where all reaction rates are identifiable, (Rank(N_m)=n_R) then Bonvin and Rippin (1990) estimates the reaction rates from the measured flux rates:

$$\hat{R} = V_o^{-1} \cdot \hat{F}_m^r \cdot N_m^+ \tag{3.1}$$

Here, N_m^+ is the Moore Penrose inverse defined by $N_m^+ = N_m^T \left(N_m N_m^T\right)^{-1}$. For the case where not all the reaction rates are identifiable (i.e. Rank(N_m) < n_R), Brendel et al. (2006) introduces an identifiability criterion based on the difference between $N_m \cdot N_m^+$ and I:

$$\Delta^r = \left[N_m N_m^+ - I\right] \tag{3.2}$$

The identifiability criterion Δ^r, computed using Eq. 3.2, states that a reaction rate r_j is structurally identifiable if the corresponding column in Δ^r is represented by the null vector.

Thus, in order to determine whether a vector of reaction rates is identifiable or not, it is necessary to form the corresponding stoichiometric matrix from the full N_m matrix and then to compute the identifiability criterion vector Δ^r using Eq. 3.2.

3.2.2 Structural Parameter Identifiability Analysis

In a given input affine model structure $M(\theta)$ (Eq. 3.3), the f functions represent the part of the right hand-side of the equations which are unaffected by the inputs; while the parts affected by the affine inputs, are represented by the functions g_i (i=1,..m).

$$M(\theta): \begin{cases} \frac{dx(t)}{dt} = f\left(x\left(t\right),\theta\right) + \sum\limits_{i=1}^{m} u_i\left(t\right) g_i\left(x\left(t\right),\theta\right),\ x\left(0\right) = x_0\left(\theta\right) \\ y\left(t,\theta\right) = h\left(x\left(t\right),\theta\right) \end{cases} \tag{3.3}$$

For such a given input-output behavior $M\left(\theta\right)$, it is of interest to investigate if identical behavior for the plant parameters θ^* and for the model parameters $\hat{\theta}$, imply that the parameters are identical. Thus three structural identifiability situations can occur:

1. A model parameter θ_i will be structurally globally identifiable, if, for almost any θ^*, $M(\hat{\theta}) = M(\theta^*) \ \rightarrow \ \hat{\theta}_i = \theta_i^*$. The structure M will be structurally globally identifiable if all its parameters are structurally globally identifiable.

2. A model parameter θ_i will be structurally locally identifiable, if, for almost any θ^*, there exists a neighborhood $V(\theta^*)$ such that $\hat{\theta} \in V(\theta^*)$ and $M(\hat{\theta}) = M(\theta^*) \ \rightarrow \ \hat{\theta}_i = \theta_i^*$. The structure M will be structurally locally identifiable if all its parameters are structurally locally identifiable.

3. The parameter θ_i is structurally unidentifiable if for almost any θ^* there is no neighborhood $V(\theta^*)$ such that $\hat{\theta} \in V(\theta^*)$ and $M(\hat{\theta}) = M(\theta^*) \ \rightarrow \ \hat{\theta}_i = \theta_i^*$. The structure M will be structurally unidentifiable if at least one of its parameters is structurally unidentifiable.

A generating series expansion can be written based on the Lie derivatives where the model output is expanded in a series with respect to inputs and time (Fliess and Lagarrigue 1980), (Fliess et al. 1983), (Pohjanpalo 1978), (Walter and Pronzato 1996) around an initial time denoted by subscript $_0$. If the initial

time does not provide sufficient information, that is, all the derivatives are zero Pohjanpalo (1978), then alternatively, a different time for which the values of the states are known can be chosen.

Index m represents the number of inputs, and d is the maximum order of derivation. The Lie derivative of h, $L_f h$ along the vector field f is:

$$L_f h\left(x\left(t\right),\theta\right) = \sum_{j=1}^{n} f_j\left(x\left(t\right),\theta\right) \frac{\partial}{\partial x_j} h\left(x\left(t\right),\theta\right) \tag{3.4}$$

For a mixed term, where the inputs are involved, the equation can be written as:

$$L_{g_i} L_f h\left(x\left(t\right),\theta\right) = \sum_{j=1}^{n} g_{i,j}\left(x\left(t\right),\theta\right) \frac{\partial}{\partial x_j} L_f h\left(x\left(t\right),\theta\right) \tag{3.5}$$

where n is the number of states and index j is the j^{th} element of the f or g functions vector, while i represents the i^{th} input. The above determined Lie derivatives represents the vector of the coefficients of the corresponding generating expansion. These are loaded into vector $s(\theta) = \{s_k(\theta)\}$ with $k \in (0, m \times n \times d)$. Then the identifiability analysis follows from (Walter and Pronzato 1996):

Proposition 1 *For two arbitrary parameterizations $\hat{\theta}$ and $\tilde{\theta}$, the identity $M(\hat{\theta}) \equiv M(\tilde{\theta})$ translates into $s(\hat{\theta}) \equiv s(\tilde{\theta})$.*

Consequently one can test the structural identifiability of the model parameters $M(\theta)$ by analytically, e.g. symbolically determining the number of solutions for $\hat{\theta}$ of the set of equations $s(\hat{\theta}) \equiv s(\tilde{\theta})$. The determination of the number of analytical solutions to the above equation set can be further facilitated by the following corollary:

Corollary 1 *Determining the number of analytical solutions to the system of equations $s(\hat{\theta}) \equiv s(\tilde{\theta})$ is equivalent to determining the number of analytical solutions for $s(\theta) = k$ where k is an arbitrary constant vector.*

Proof 1 *The set of equations $s(\hat{\theta}) \equiv s(\tilde{\theta})$ is split into two systems of equations $s(\hat{\theta}) \equiv k$ and $s(\tilde{\theta}) \equiv k$, where k is an arbitrary constant. By solving these analytically, both equation systems will provide the same solution(-s). Hence it is sufficient to solve only one set symbolically.*

If the above underlying system of equations has only one solution, then the parameter set is structurally globally identifiable; if there are more solutions, then the parameters at best may be structurally locally identifiable and if the system has no solution the parameters are structurally unidentifiable. Once the Lie derivatives have been computed i.e. the vector $s(\hat{\theta})$ is determined, then a system of algebraic equations is obtained. The elements in the parameter

vector θ are the model parameters possibly including the initial state values. It should be mentioned, that if two parameters e.g. a and b are unidentifiable, then an arbitrary combination of the two, e.g. the sum or the ratio may be structurally identifiable. For batch models, applying the generating series the obtained set of equations is similar to the set obtained by applying Taylor series, but for fed-batch models, the set obtained by applying generating series is less complex than that obtained by applying Taylor series. Therefore the former is applied in this work.

Although structural identifiability analysis thus seems most appealing, it is limited by the ability to solve the above algebraic equation set. Obtaining this solution analytically is in general not feasible. While symbolic methods provide a help, they cannot overcome this fundamental limitation. Hence alternative approaches should be investigated. This fundamental limitation implies that the number of solutions are not easily determined in general. However for polynomial kinetic expressions the algebraic equations are in principle solvable, and in practice also by computational symbolic methods for low order cases.

Next the work flow of the structural identifiability algorithm is described.

3.2.3 Structural identifiability algorithm

The work-flow of the developed algorithm is illustrated in Figure 3.1. The steps are detailed in the following.

The algorithm is composed of two main phases where each phase encompasses several steps. The first phase concerns structural identifiability of the reaction rates given the reaction network and the measured species. The second phase, investigates parameter identifiability of the kinetic models for the identifiable reaction rate set.

- *Step 1* and *2* of the algorithm presented in the Figure 3.1 concerns model formulation for a reaction network for the process under investigation based upon physical knowledge including the reaction stoichiometry.
- *Step 3* concerns determination of the structurally identifiable reaction rates. In order to determine which reaction rates are identifiable a criterion introduced by Brendel et al. (2006) Eq. 3.2 is used. The criterion utilizes only the information related to the reaction stoichiometry.
- After the possible combinations of identifiable reaction rates have been determined, the kinetic expressions together with the associated parameters are introduced in *step 4*, where the parameter vector to be investigated for identifiability is formulated for each combination of identifiable reaction rates.
- The first order Lie derivatives which represents the first term of a power series expansion are derived in *Step 5*. If the model comes from a fed-batch or continuous reactor where perturbed inputs to the process may be present, then,

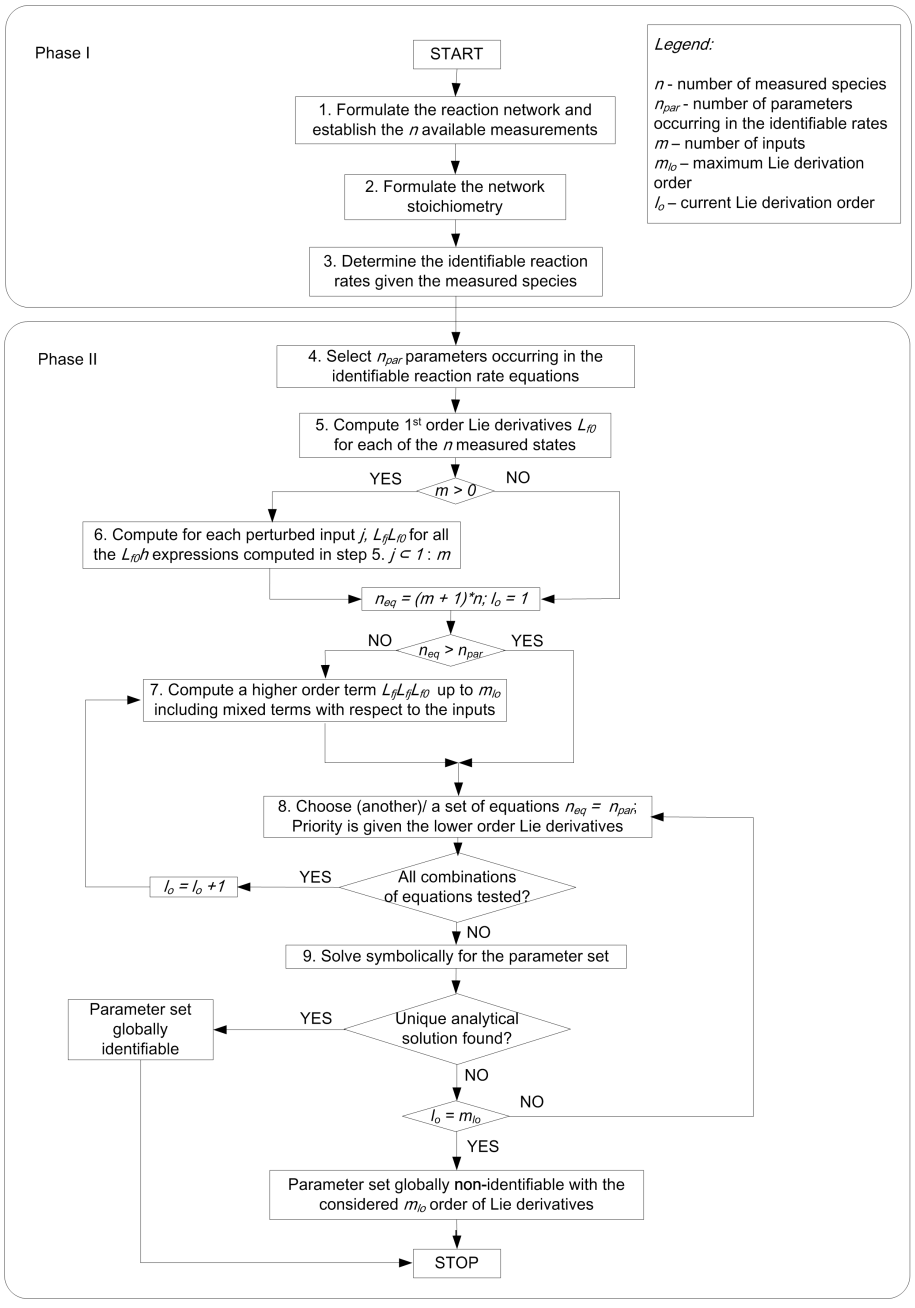


Figure 3.1: Work-flow of the structural identifiability analysis

- *step 6* concerns calculation of mixed first order Lie derivatives terms with respect to such inputs.
- *step 6* is skipped if no perturbed inputs are present.
- *Step 7* deals with computing higher order Lie derivatives if necessary in order to form a fully determined system of equations.
- *step 8* concerns selection of the equations to be included in the algebraic system. When a selection is being performed, the lower order Lie derivatives, being the simplest are given priority. If inputs are present, priority is given to the derivatives containing inputs. If no possible unique solution has been found, the system of equations will be expanded by considering including higher order terms as well.
- In *step 9* the equation system considered in the previous step is solved for the parameter set containing n_{par} parameters. If a unique solution is found, then the parameter set is globally identifiable. If the system of equations solved in *step 9* did not provided a unique solution, then returning to *step 7*, equations obtained from higher order Lie derivatives are included in the equation vector.

The procedure is iterated until all possible combinations including highest order Lie derivation m_{lo} have been considered. For practical reasons m_{lo} is set to 3, and the main reason is that the obtained solutions (when obtained) become complex.

The loop of computing additional derivatives and solving for the various parameter combinations, is terminated if a unique solution has been found for the parameters considered in *step 4* or when all the possible equation combinations have been analyzed.

3.3 Example Process

There is an increasing interest in producing complex intermediates and expensive fine chemicals in the pharmaceutical industry using biochemical synthesis. Presently, only a few bio-transformation steps are involved in complex synthesis in industry, although enzymes are widely known as being specific, fast and working under mild conditions. To develop a purely enzymatic synthesis for complex molecules from simple substrates, often relatively large reaction networks are necessary. One way to achieve such a functional network is by using a so-called System of Bio-Transformations (SBT). An SBT is based on parts of the metabolic network of a microorganism containing the synthesis pathway including cofactor regeneration reactions down to the desired product, which most often is an intermediate in the metabolic network. Suitably genetically modified mutants of *E-coli* are here used to produce the metabolic network for an SBT (Schümperli et al. 2007). The mutants are grown up to a certain

level. Deletion of the genes for enzymes catalyzing reactions which compete with the desired pathway is triggered a short while before the cultivation is stopped. Subsequently the cells are isolated by centrifugation. The cells are then resuspended in a buffer. The cell walls are destroyed by high pressure homogenization and the resulting particles are then separated by filtration. The cell free liquid extract is recovered and then the desired protein concentration is obtained by dilution with a buffer solution. The bio-transformations are performed with the cell free extract in the production phase, combining the easy handling of a viable culture for producing the desired enzymes with the advantages of in vitro bio-transformations. For this example, the key product is Di-hydroxy-acetone phosphate (DHAP), which represents an important precursor for the production of phosphorylated, non-natural carbohydrates. The DHAP-producing SBT contains all the enzymes for the glycolysis reactions except the ones corresponding to the genes which were deleted prior to cessation of cultivation, leading to a system of high complexity. Below, the developed structural parameter identifiability analysis is illustrated on a relatively simple mass action model for the SBT.

3.3.1 Model formulation

Since the system described above is a system of high order dynamics and complexity which is intended to operate at a substantially higher concentration level than in the microbial cell, it is not considered to be realistic to develop a "perfect model" using purely first principle engineering methods. Thus, in this work a grey-box stochastic model development framework (Kristensen et al. 2004b) is used. The idea behind this methodological chain is to be able to combine the limited amount of data for measured states with available first principle engineering knowledge. A very large model considering all the possible intermediates of glycolysis, will inherently contain a very large number of parameters which in practice would be impossible to validate against the limited available data. Thus for many parameters only an educated guess can be used. For practical reasons this article only describes the deterministic part of the grey-box model. A network representing the most important reactions occurring in the SBT is given in Figure 3.1. The network will form the basis for formulating the mass balance model equations which constitute the model.

Based upon the pathway shown in Figure 3.2, there are 7 enzymatic reactions included in this network. It is further assumed that except degradation of DHAP the remaining reactions are reversible. It is assumed that the kinetics can be described by simple mass action terms as follows:

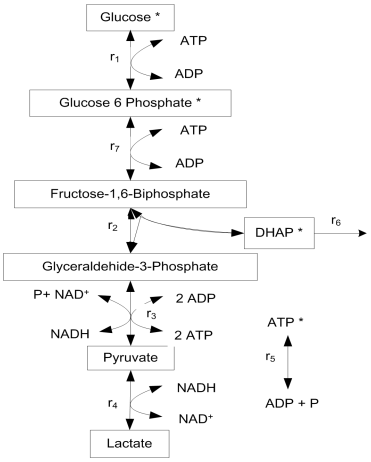


Figure 3.2: The simplified reaction network investigated in this analysis

$$
\begin{aligned}
r_1 &= r_{1maxf} c_{GL} c_{ATP} - r_{1maxb} c_{G6P} c_{ADP} \\
r_2 &= r_{2maxf} c_{F16B} - r_{2maxb} c_{DHAP} c_{G3P} \\
r_3 &= r_{3maxf} c_{G3P} c_{NAD} c_{PO4} c_{ADP} - r_{3maxb} c_{PYR} c_{NADH} c_{ATP} \\
r_4 &= r_{4maxf} c_{PYR} c_{NADH} - r_{4maxb} c_{LAC} c_{NAD} \\
r_5 &= r_{5maxf} c_{ATP} - r_{5maxb} c_{ADP} \\
r_6 &= r_{6maxf} c_{DHAP} \\
r_7 &= r_{7maxf} c_{G6P} c_{ATP} - r_{7maxb} c_{ADP} c_{F16B}
\end{aligned} \tag{3.6}
$$

For each chemical compound a dynamic mass balance is formulated as shown below. Since the bio-transformation reactor is a batch reactor, it does not include any perturbed input flow-rate. The corresponding terms therefore will not appear in the equations and thus will not be used in the further analysis.

$$
\begin{aligned}
\frac{dc_{GL}}{dt} &= -r_1 \\
\frac{dc_{F16B}}{dt} &= r_7 - r_2 \\
\frac{dc_{DHAP}}{dt} &= r_2 - r_6 \\
\frac{dc_{G3P}}{dt} &= r_2 - r_3 \\
\frac{dc_{PYR}}{dt} &= r_3 - r_4 \\
\frac{dc_{LAC}}{dt} &= r_4 \\
\frac{dc_{ATP}}{dt} &= -r_1 - r_7 + 2r_3 - r_5 \\
\frac{dc_{NAD}}{dt} &= -r_3 + r_4 \\
\frac{dc_{PO4}}{dt} &= -r_3 + r_5 \\
\frac{dc_{G6P}}{dt} &= r_1 - r_7 \\
\frac{dc_{ADP}}{dt} &= r_1 + r_7 - 2r_3 + r_5 \\
\frac{dc_{NADH}}{dt} &= r_3 - r_4
\end{aligned} \tag{3.7}
$$

In the analysis, it is assumed that only four states are measured. Using the notation introduced above in Eq. 3.7 the n_{Sm} measurement equations for this case are simply the states which are being measured:

$$
\begin{aligned}
y_1(c) &= c_{GL} \\
y_2(c) &= c_{G6P} \\
y_3(c) &= c_{DHAP} \\
y_4(c) &= c_{ATP}
\end{aligned} \tag{3.8}
$$

3.3.2 Structural identifiability analysis for the enzymatic reaction network (SBT)

This section presents an application of the algorithm presented in Section 3.2.3 for the enzymatic reaction network model described above. The analysis is practically started from *step 2* of the algorithm since the reaction network and the associated model were already formulated in the previous section. The full stoichiometric matrix N associated with the network, in Figure 3.2 is given in Table 1.

	GL	ATP	G6P	ADP	F16B	DHAP	G3P	PO4	NAD	NADH	PYR	LAC
r_1	-1	-1	1	1	0	0	0	0	0	0	0	0
r_2	0	0	0	0	-1	1	1	0	0	0	0	0
r_3	0	2	0	-2	0	0	-1	-1	-1	1	1	0
r_4	0	0	0	0	0	0	0	0	1	-1	-1	1
r_5	0	-1	0	1	0	0	0	1	0	0	0	0
r_6	0	0	0	0	0	-1	0	0	0	0	0	0
r_7	0	-1	-1	1	1	0	0	0	0	0	0	0

Table 3.1: Full stoichiometric matrix N for the reaction network

The seven reactions considered in the network are independent since the rank of the stoichiometric matrix is 7. The matrix is split into two matrices corresponding to the species measured and unmeasured respectively. The matrix containing the measured species is denoted N_m (Table 3.2), and the one containing the unmeasured species N_u, (Table 3.3).

	GL	ATP	G6P	DHAP
r_1	-1	-1	1	0
r_2	0	0	0	1
r_3	0	2	0	0
r_4	0	0	0	0
r_5	0	-1	0	0
r_6	0	0	0	-1
r_7	0	-1	-1	0

Table 3.2: Stoichiometric matrix N_m for the measured species

	ADP	F16B	G3P	PO4	NAD	NADH	PYR	LAC
r_1	1	0	0	0	0	0	0	0
r_2	0	-1	1	0	0	0	0	0
r_3	-2	0	-1	-1	-1	1	1	0
r_4	0	0	0	0	1	-1	-1	1
r_5	1	0	0	1	0	0	0	0
r_6	0	0	0	0	0	0	0	0
r_7	1	1	0	0	0	0	0	0

Table 3.3: Stoichiometric matrix N_u, for the unmeasured species

If N_m is not full rank, then only a subset of the rates will be identifiable. The rank of the N_m matrix is 4. Thus it is possible to identify a maximum of four independent reaction rates. In order to apply Eq. 3.2 the $N_m \cdot N_m^T$ matrix has to be invertible, thus it is necessary to have a full matrix with four rows and it will be obtained using four rows from the original matrix N_m matrix; this matrix will be denoted N_{mr} and replaces N_m in Eq. 3.2. All the possible

combinations of four reaction rates obtained from the full set composed of seven reaction rates are determined first. The identifiable combinations of parameters for each of the identifiable reaction rate sets are obtained, secondly.

The structural rate identifiability criterion presented in Eq. 3.2 was applied for all the combinations described above. The criterion was fulfilled for a series of combinations as shown in Table 3.4.

For other combinations it was not possible to invert the matrix N_{mr}, which is required to calculate the pseudo-inverse N_{mr}^{+} and thus the corresponding combinations have been rejected from the analysis.

Combination	Identifiable Rates			
1	r_1	r_5	r_6	r_7
2	r_1	r_3	r_6	r_7
3	r_1	r_2	r_5	r_7
4	r_1	r_2	r_3	r_7

Table 3.4: Identifiable combinations of reaction rates for the enzymatic reaction network

A simple check for the validity of the results is that the matrix N_{mr} is full rank. The N_{mr} matrix for the four combinations are given below in Tables 3.5 –3.8.

Rates	GL	ATP	G6P	DHAP	Δ^r
r_1	-1	-1	1	0	0
r_5	0	-1	0	0	0
r_6	0	0	0	-1	0
r_7	0	-1	-1	0	0

Table 3.5: Reduced stoichiometric matrix — first set

Rates	GL	ATP	G6P	DHAP	Δ^r
r_1	-1	-1	1	0	0
r_3	0	2	0	0	0
r_6	0	0	0	-1	0
r_7	0	-1	-1	0	0

Table 3.6: Reduced stoichiometric matrix — second set

The rank of all four matrices is 4 and the determinant different from zero, i.e. $det(M) = -1, 2, -1$ and 2 respectively. Thus the selected combinations of reaction rates are indeed identifiable.

For each of the combinations in Table 3.4,*steps 4 – 9*, have been performed in order to asses the global identifiability of the parameters involved in the

Rates	GL	ATP	G6P	DHAP	Δ^r
r_1	-1	-1	1	0	0
r_2	0	0	0	1	0
r_5	0	-1	0	0	0
r_7	0	-1	-1	0	0

Table 3.7: Reduced stoichiometric matrix — third set

Rates	GL	ATP	G6P	DHAP	Δ^r
r_1	-1	-1	1	0	0
r_2	0	0	0	1	0
r_3	0	2	0	0	0
r_7	0	-1	-1	0	0

Table 3.8: Reduced stoichiometric matrix — fourth set

corresponding kinetic expressions. For illustration purpose the first identifiable combination of reaction rates will be detailed below. For the remaining combinations the results will only be given in tabular form.

3.3.2.1 Combination r_1, r_5, r_6, r_7

The parameters included in the kinetic expressions for this case include r_{1maxb}, r_{1maxf}, r_{5maxb}, r_{5maxf}, r_{6maxb}, r_{6maxf}, r_{7maxb}, r_{7maxf}. In *step 5*, the first order Lie derivatives are computed for the four measured states of the model as follows:

$$
\begin{aligned}
L_f h_1 &= f_1 \frac{\partial h_1}{\partial c_{GL}}\bigg|_0 \\
&= -r_{1maxf} c_{GL} c_{ATP} + r_{1maxb} c_{G6P} c_{ADP} \\
L_f h_2 &= f_{10} \frac{\partial h_2}{\partial c_{G6P}}\bigg|_0 \\
&= r_{1maxf} c_{GL} c_{ATP} - r_{1maxb} c_{G6P} c_{ADP} - r_{7maxf} c_{G6P} c_{ATP} \\
&\quad + r_{7maxb} c_{ADP} c_{F16B} \\
L_f h_3 &= f_3 \frac{\partial h_3}{\partial c_{DHAP}}\bigg|_0 \\
&= r_{2maxf} c_{F16B} - r_{2maxb} c_{DHAP} c_{G3P} - r_{6maxf} c_{DHAP} \\
L_f h_4 &= f_4 \frac{\partial h_4}{\partial c_{ATP}}\bigg|_0 \\
&= -r_{1maxf} c_{GL} c_{ATP} + r_{1maxb} c_{G6P} c_{ADP} - r_{7maxf} c_{G6P} c_{ATP} \\
&\quad + r_{7maxb} c_{ADP} c_{F16B} + 2 r_{3maxf} c_{G3P} c_{NAD} c_{PO4} c_{ADP} \\
&\quad - 2 r_{3maxb} c_{PYR} c_{NADH} c_{ATP} - r_{5maxf} c_{ATP} + r_{5maxb} c_{ADP}
\end{aligned}
\tag{3.9}
$$

Since there is no perturbed input present in the process operation, *step 6* is skipped. Following the algorithm presented in Figure 3.1 it is necessary to include additional equations since $n_{eq} < n_{par}$. Thus, the second order Lie derivatives are computed. For the higher order terms only the compact form of the equations are shown in equations 3.10.

$$\begin{aligned}
L_f L_f h_1 &= f_1 \frac{\partial L_f h_1}{\partial c_{GL}}\bigg|_0 + f_7 \frac{\partial L_f h_1}{\partial c_{ATP}}\bigg|_0 + f_{10} \frac{\partial L_f h_1}{\partial c_{G6P}}\bigg|_0 \\
&+ f_{11} \frac{\partial L_f h_1}{\partial c_{ADP}}\bigg|_0 \\
L_f L_f h_2 &= f_1 \frac{\partial L_f h_2}{\partial c_{GL}}\bigg|_0 + f_7 \frac{\partial L_f h_2}{\partial c_{ATP}}\bigg|_0 + f_{10} \frac{\partial L_f h_2}{\partial c_{G6P}}\bigg|_0 \\
&+ f_{11} \frac{\partial L_f h_2}{\partial c_{ADP}}\bigg|_0 + f_2 \frac{\partial L_f h_2}{\partial c_{F16B}}\bigg|_0 \\
L_f L_f h_3 &= f_2 \frac{\partial L_f h_3}{\partial c_{F16B}}\bigg|_0 + f_3 \frac{\partial L_f h_3}{\partial c_{DHAP}}\bigg|_0 + f_4 \frac{\partial L_f h_3}{\partial c_{G3P}}\bigg|_0 \\
L_f L_f h_4 &= f_1 \frac{\partial L_f h_4}{\partial c_{GL}}\bigg|_0 + f_7 \frac{\partial L_f h_4}{\partial c_{ATP}}\bigg|_0 + f_{10} \frac{\partial L_f h_4}{\partial c_{G6P}}\bigg|_0 \\
&+ f_{11} \frac{\partial L_f h_4}{\partial c_{ADP}}\bigg|_0 + f_2 \frac{\partial L_f h_4}{\partial c_{F16B}}\bigg|_0 + f_4 \frac{\partial L_f h_4}{\partial c_{G3P}}\bigg|_0 \\
&+ f_8 \frac{\partial L_f h_4}{\partial c_{NAD}}\bigg|_0 + f_9 \frac{\partial L_f h_4}{\partial c_{PO4}}\bigg|_0 + f_5 \frac{\partial L_f h_4}{\partial c_{PYR}}\bigg|_0 \\
&+ f_{12} \frac{\partial L_f h_4}{\partial c_{NADH}}\bigg|_0
\end{aligned} \tag{3.10}$$

In *step 8*, several systems of seven equations may be formed using the first and second order Lie derivatives. No combination was found to render a unique solution using equations formed with first and second order Lie derivatives. Lie derivation has been applied once more and various combinations of equations have been solved for the parameters. The only combination of 7 equations that rendered a unique solution for the parameter set is listed in Table 3.9.

Comb.	r_{1maxb}	r_{1maxf}	r_{5maxb}	r_{5maxf}	r_{6maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_2$	$L_f h_1$	$L_f h_3$	$L_f h_4$	$L_f L_f h_2$	$L_f L_f L_f h_1$	$L_f L_f L_f h_4$

Table 3.9: One combination of Lie derivatives for which a unique solution was found

For the combination of reaction rates considered, it was possible to find a unique solution for various system of equations formed with Lie derivatives up to the third order originating from the four measurement functions. Thus, the investigated parameter set is deemed to be structurally globally identifiable.

The analysis is repeated for all the other vectors of identifiable rates given in Table 3.4.

3.3.2.2 Combination r_1, r_3, r_6, r_7

For this combination of parameters, steps *4–9* were performed again. The combinations for which a unique solution was found, is given below in Table 3.10.

A unique solution could be found for several combinations of the Lie derivatives, one combination is shown in Table 3.10. This time it was possible to obtain a unique solution by using only the first and the second order Lie derivatives. Thus the parameter set is deemed to be globally identifiable.

Comb.	r_{1maxf}	r_{1maxf}	r_{3maxf}	r_{3maxb}	r_{6maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_1$	$L_f h_3$	$L_f L_f h_1$	$L_f h_4$	$L_f L_f h_2$	$L_f L_f h_4$	$L_f L_f h_3$

Table 3.10: One combination of first and second order Lie derivatives for which a unique solution was found

3.3.2.3 Combination r_1, r_2, r_5, r_7

For the third combination of reaction rates a combination of Lie derivatives for which a unique solution was found when solved for the full parameter set is given in Table 3.11. Here, however third order Lie derivatives were required. Since a unique solution has been found, the parameter set is deemed to be globally identifiable.

Comb.	r_{1maxb}	r_{1maxf}	r_{2maxb}	r_{2maxf}	r_{5maxb}	r_{5maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_1$	$L_f h_3$	$L_f L_f h_2$	$L_f L_f h_3$	$L_f L_f h_1$	$L_f L_f L_f h_4$	$L_f L_f L_f h_3$	$L_f h_4$

Table 3.11: One combination of first, second, and third Lie derivatives for which a unique solution was found

3.3.2.4 Combination r_1, r_2, r_3, r_7

One combination of Lie derivatives for which a unique solution was found when considering the full parameter vector of eight parameters is given in Table 3.12. The parameter set for the last combination of identifiable reaction rates is deemed to be structurally globally identifiable.

3.4 Validation of the algorithm

The validation is undertaken with the specific understanding that structural identifiability as defined in this work is concerned with the possibility of de-

Comb.	r_{1maxb}	r_{1maxf}	r_{2maxb}	r_{2maxf}	r_{3maxb}	r_{3maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_1$	$L_f h_3$	$L_f h_2$	$L_f L_f h_3$	$L_f L_f h_1$	$L_f L_f h_4$	$L_f L_f L_f h_1$	$L_f L_f L_f h_4$

Table 3.12: Combination of Lie derivatives for which a unique solution was found

termining the model parameters. To quantitatively determine the parameter values requires that suitable parameter sensitivity information is available in the experimental design, which in this article thus far has been assumed given. For cases where such parameter sensitivity information is not available then a minimization may be performed but high correlations may result. To illustrate the qualitative identifiability result an arbitrary example based upon the reaction network modelled in this article is used.

The idea is to first simulate the noise free model with initial conditions which do provide information from different parts of the network with a given set of parameters and then to use these data to estimate the parameters.

A set of 100 data points was obtained by simulating the model given in Eqs. 3.6–3.8. The same value of 0.1 was used for all the model parameters. The initial concentration used for this simulation are given in Table 3.13. As measured compounds, the four measurements in Eqs. 3.8 have been considered.

Compound	Concentration	Compound	Concentration
c^0_{GL}	11.1 [mM/L]	c^0_{ATP}	11.5 [mM/L]
c^0_{F16B}	1e-5 [mM/L]	c^0_{NAD}	5.75 [mM/L]
c^0_{DHAP}	1e-5 [mM/L]	c^0_{PO4}	11.1 [mM/L]
c^0_{G3P}	1e-5 [mM/L]	c^0_{G6P}	1e-5 [mM/L]
c^0_{PYR}	1e-5 [mM/L]	c^0_{ADP}	1e-5 [mM/L]
c^0_{LAC}	1e-5 [mM/L]	c^0_{NADH}	1e-5 [mM/L]

Table 3.13: Initial conditions for simulation

The parameter sets given in Tables 3.10–3.12 were estimated from the simulated data. A Levenberg-Marquardt method was used for the estimation of the parameters using the *COPASI* software program Hoops et al. (2006). The estimation has been performed using the interval $[0, 100]$. A set of three different initial guesses of parameters 1, 5 and 10 have been used. However very similar results were obtained in all the cases.

The Table 3.14 gives the results obtained for the set r_1, r_5, r_6, r_7.

The objective function used, was the standard least squares estimator and the value for it was $Obj = 3.1839e - 010$.

The computed correlation matrix for the estimates is given in Table 3.15. However the correlation matrix reveals that a quantitative experimental design is desirable to reduce the three high parameter correlations between the forward and backward rate constants.

Parameters	LowerB	Initial	UpperB	Estimate	σ_D
r_{1maxf}	0	10	100	0.1	1.84456e-007
r_{1maxb}	0	10	100	0.1	1.98921e-007
r_{5maxf}	0	10	100	0.1	6.27943e-008
r_{5maxb}	0	10	100	0.1	6.39445e-008
r_{6maxf}	0	10	100	0.1	1.27507e-008
r_{7maxf}	0	10	100	0.0999997	1.76717e-007
r_{7maxb}	0	10	100	0.0999997	1.9734e-007

Table 3.14: Estimation results

	r_{1maxf}	r_{1maxb}	r_{5maxf}	r_{5maxb}	r_{6maxf}	r_{7maxf}	r_{7maxb}
r_{1maxf}	1	0.99872	-0.517642	-0.519229	-0.0893252	-0.538386	-0.536762
r_{1maxb}	0.99872	1	-0.509737	-0.510993	-0.0866214	-0.538004	-0.537169
r_{5maxf}	-0.517642	-0.509737	1	0.999352	0.0437292	-0.107299	-0.101635
r_{5maxb}	-0.519229	-0.510993	0.999352	1	0.0440764	-0.107957	-0.10211
r_{6maxf}	-0.0893252	-0.0866214	0.0437292	0.0440764	1	0.191113	0.210542
r_{7maxf}	-0.538386	-0.538004	-0.107299	-0.107957	0.191113	1	0.99867
r_{7maxb}	-0.536762	-0.537169	-0.101635	-0.10211	0.210542	0.99867	1

Table 3.15: Correlation matrix

3.4.1 Combination r_1, r_3, r_6, r_7

The estimation results for this combination of parameters is given in Table 3.16.

Parameters	LowerB	Initial	UpperB	Estimate	σ_D
r_{1maxb}	0	10	100	0.1	1.9942e-007
r_{1maxf}	0	10	100	0.1	2.13465e-007
r_{3maxb}	0	10	100	0.0999986	2.84563e-006
r_{3maxf}	0	10	100	0.0999988	2.84651e-006
r_{6maxb}	0	10	100	0.1	1.32055e-008
r_{7maxb}	0	10	100	0.0999992	2.46808e-007
r_{7maxf}	0	10	100	0.099999	2.89585e-007

Table 3.16: Estimation results

The objective function used, was the standard least square estimator and its value was $Obj = 3.1824e-010$. The computed correlation matrix for the estimates is given in Table 3.17.

3.4.2 Combination r_1, r_2, r_5, r_7

The estimation results for this combination of parameters is given in Table 3.18.

The objective function used, was the standard least square estimator and its value was $Obj = 3.17734e-010$. The computed correlation matrix for the estimates is given in Table 3.19.

	r_{1maxf}	r_{1maxb}	r_{3maxf}	r_{3maxb}	r_{6maxf}	r_{7maxf}	r_{7maxb}
r_{1maxf}	1	0.998777	-0.534569	-0.529313	-0.1671	0.693731	-0.702641
r_{1maxb}	0.998777	1	-0.526384	-0.521379	-0.164697	-0.684436	-0.693112
r_{3maxf}	-0.534569	-0.526384	1	0.999852	0.250377	0.178101	0.224431
r_{3maxb}	-0.529313	-0.521379	0.999852	1	0.251812	0.16639	0.212359
r_{6maxf}	-0.1671	-0.164697	0.250377	0.251812	1	0.121849	0.137729
r_{7maxf}	-0.693731	-0.684436	0.178101	0.16639	0.121849	1	0.997507
r_{7maxb}	-0.702641	-0.693112	0.224431	0.212359	0.137729	0.997507	1

Table 3.17: Correlation matrix

Parameters	LowerB	Initial	UpperB	Estimate	σ_D
r_{1maxf}	0	10	100	0.1	2.02008e-007
r_{1maxb}	0	10	100	0.1	2.16843e-007
r_{2maxf}	0	10	100	0.1	4.2731e-008
r_{2maxb}	0	10	100	0.1	8.07845e-008
r_{5maxf}	0	10	100	0.1	7.08752e-008
r_{5maxb}	0	10	100	0.1	7.20201e-008
r_{7maxf}	0	10	100	0.0999993	2.88396e-007
r_{7maxb}	0	10	100	0.0999992	3.35906e-007

Table 3.18: Estimation results

	r_{1maxf}	r_{1maxb}	r_{2maxf}	r_{2maxb}	r_{5maxf}	r_{5maxb}	r_{7maxf}	r_{7maxb}
r_{1maxf}	1	0.998863	-0.277813	0.0618174	-0.250194	-0.253889	-0.510127	-0.52624
r_{1maxb}	0.998863	1	-0.27424	0.0562994	-0.248757	-0.252145	-0.503898	-0.519858
r_{2maxf}	-0.277813	-0.27424	1	0.63551	-0.444459	-0.44141	-0.0955138	-0.0158406
r_{2maxb}	0.0618174	0.0562994	0.63551	1	-0.177082	-0.176897	-0.673023	-0.641941
r_{5maxf}	-0.250194	-0.248757	-0.444459	-0.177082	1	0.999481	-0.128604	-0.162374
r_{5maxb}	-0.253889	-0.252145	-0.44141	-0.176897	0.999481	1	-0.127623	-0.160973
r_{7maxf}	-0.510127	-0.503898	-0.0955138	-0.673023	-0.128604	-0.127623	1	0.995729
r_{7maxb}	-0.52624	-0.519858	-0.0158406	-0.641941	-0.162374	-0.160973	0.995729	1

Table 3.19: Estimation results

3.4.3 Combination r_1, r_2, r_3, r_7

The estimation results for this combination of parameters is given in Table 3.20.

Parameters	LowerB	Initial	UpperB	Estimate	σ_D
r_{1maxf}	0	10	100	0.1	2.005e-007
r_{1maxb}	0	10	100	0.1	2.15011e-007
r_{2maxf}	0	10	100	0.1	5.33221e-008
r_{2maxb}	0	10	100	0.0999993	1.51572e-007
r_{3maxf}	0	10	100	0.0999938	3.79976e-006
r_{3maxb}	0	10	100	0.0999947	3.80691e-006
r_{7maxf}	0	10	100	0.0999994	3.02979e-007
r_{7maxb}	0	10	100	0.0999992	3.48642e-007

Table 3.20: Estimation results

The objective function used, was the standard least square estimator and its

value was $Obj = 3.14981e-010$. The computed correlation matrix for the estimates is given in Table 3.21.

	r_{1maxf}	r_{1maxb}	r_{2maxf}	r_{2maxb}	r_{3maxf}	r_{3maxb}	r_{7maxf}	r_{7maxb}
r_{1maxf}	1	0.998777	-0.129228	0.0632667	-0.262375	-0.26111	-0.579635	-0.605234
r_{1maxb}	0.998777	1	-0.129683	0.0689804	-0.25525	-0.254352	-0.574832	-0.600227
r_{2maxf}	-0.129228	-0.129683	1	0.583089	-0.687561	-0.691127	-0.342496	-0.284492
r_{2maxb}	0.0632667	0.0689804	0.583089	1	-0.339403	-0.35967	-0.592717	-0.570033
r_{3maxf}	-0.262375	-0.25525	-0.687561	-0.339403	1	0.999661	0.288198	0.275173
r_{3maxb}	-0.26111	-0.254352	-0.691127	-0.35967	0.999661	1	0.292927	0.279581
r_{7maxf}	-0.579635	-0.574832	-0.342496	-0.592717	0.288198	0.292927	1	0.996537
r_{7maxb}	-0.605234	-0.600227	-0.284492	-0.570033	0.275173	0.279581	0.996537	1

Table 3.21: Estimation results

Analogous results were obtained when estimating parameters for the other three reaction rate combinations as can be seen in the Tables 3.16–3.20. Very similar solutions were obtained within the shown standard deviations, thus enforcing the fact that indeed unique solutions are obtained in each case.

3.5 Discussion

For the example illustrated above it was possible to find combinations of Lie derivatives rendering a unique analytical solution for the parameter set under investigation for each combination of identifiable rates. When analyzing each individual combination of identifiable reactions, since the model is nonlinear, the number of required Lie derivatives to achieve a system of equations rendering the parameters structurally identifiable was different from combination to combination. For one of the combinations of identifiable rates it was possible to conclude structural identifiability using only first and second order Lie derivatives.

Since the available number of measured states in the presented example is low compared to the number of independent reactions a fairly large number of combinations of identifiable reactions was determined as can be seen in Table 3.4. The higher the number of measured states, the fewer combinations of structurally identifiable reaction rates will be obtained, each containing a larger number of reactions.

An important drawback is that there is no upper bound for the order of the Lie derivation required to achieve a system of algebraic equations which would render the parameter set structurally identifiable. For the case where inputs may be perturbed the number of equations available is much larger even when using only first and possibly second order Lie derivatives to form the system of algebraic equations. The complexity of the equations is thus reduced significantly and the need for the higher order derivative information is reduced.

A limitation of the methodology is related to the complexity of the model equations. Due to high complexity of the subsequent algebraic equations it may not be possible to solve the system of equations analytically.

The attempted validation presented in section 3.4 exhibited large quantitative correlations for some parameter pairs. However to provide a proper validation of the qualitative identifiability requires design of the experimental (or simulation) conditions, in order to ensure reliable parameter estimates.

Moreover changing these variables represents in fact the objective for quantitative experimental design, which is the scope of chapter 5.

3.6 Conclusions

This article introduces a methodology and an application of structural identifiability analysis as a step towards developing more efficient tools for establishing parameter identifiability for quantitative mathematical models. The article illustrates potential benefits of such a methodology, even though there is a fundamental limitation in the suggested method. The proposed approach complements the approach of Brendel et al. (2006), by introducing a structural identifiability test for the parameters of proposed kinetic expressions of identifiable reaction rates. The key feature of the combined methodology for assessing structural parameter identifiability for reaction network models is the partitioning of the main problem into two sub problems. The first sub problem aims at finding the structurally identifiable reaction rates based on stoichiometric analysis. The second sub problem aims at finding the structurally identifiable kinetic parameters involved in the identifiable reaction rates based on the generating series method. The methodology thus combines knowledge about the reaction network in terms of stoichiometry and reaction kinetics with a generating series expansion in order to facilitate global identifiability assessment of the model parameters in a systematic manner. The methodology is limited by the ability to determine the number of solutions of an algebraic equation set. Even though computer algebra is used there also are fundamental limitations to this problem. Hence a search for alternative formulations of structural identifiability is relevant.

The presented systematic methodology is illustrated through a case of an enzymatic reaction network where a simple model is derived and global identifiability investigated for a few measurements.

Having established the influence of measurements and perturbed inputs upon structural identifiability renders it possible, in the cases where the algebraic problem is solvable, to perform qualitative experimental design prior to experimental work. This aspect significantly adds to the application potential of the structural identifiability methods.

4

Qualitative experimental design for nonlinear dynamic reaction networks

Abstract

For design, control and optimization it is desirable to develop quantitative process models which exhibit good long term prediction properties over the intended operating region. During nonlinear process model development an important step is to design experiments for testing model validity for the intended application and for enabling reliable estimation of model parameters. Experimental design may be subdivided into two main steps where the first qualitative step is to select the measured states and to decide which inputs to perturb. Thus this first step is labeled qualitative experimental design. The second step aims at quantitative determination of initial state values, the input profiles and the measurement times. This contribution aims at development of a systematic methodology for qualitative experimental design for dynamic reaction networks. Qualitative experimental design is particularly appealing and important since this may be carried out a-priori. The core of the developed methodology is based upon ensuring global parameter identifiability through appropriate selection of measurement set and input perturbations by ensuring that a certain set of algebraic equations can be solve symbolically. The developed methodology is illustrated through application to qualitative experimental design of an enzymatic reaction network developed for the production of Di-Hydroxy-Acetone-Phosphate (DHAP) where the enzymes are extracted from genetically modified E-coli cells.

4.1 Introduction

For design, control and optimization, within bio- and chemical engineering it is desirable to develop quantitative process models which exhibit good long term prediction properties over the intended operating region. During model development, an important step, is to ensure identifiability of model parameters.

Experimental design aims at developing most beneficial conditions for identifying unknown model parameters and for enabling discrimination between candidate models. Experimental design may be performed in two steps: qualitative experimental design and quantitative experimental design (Walter and Pronzato 1990). Qualitative experimental design concerns selection of first the operation mode, secondly which input variables should be varied (and how) and thirdly, which outputs should be measured in the experiments in order to be able to render the unknown model parameters identifiable.

Quantitative experimental design concerns determining the actual input profiles, the initial values of the states and the optimal sampling instants for the measurements to provide maximum information for quantitative determination of the unknown parameters.

The concepts of qualitative and quantitative experimental design and their relationship with identifiability analysis appear first to be treated by Vajda et al. (1989). Walter and Pronzato (1990) introduced qualitative experimental design for linear or linearized models. Subsequently very few attempts have been published dealing with input affine nonlinear models .

Based upon a recently developed structural identifiability analysis methodology presented in chapter 3 which establishes the influence of measurements and perturbed inputs on structural identifiability, the purpose of the present chapter is to present a systematic methodology for performing qualitative or structural experimental design.

This chapter provides an attempt towards development of a systematic methodology for qualitative experimental design based upon the notion of qualitative identifiability. The presented methodology focuses upon reaction networks where determination of approximate kinetics can be very beneficial for subsequent reactor design and control.

The chapter is organized as follows. Section 4.2 describes the proposed methodology and the work-flow of the systematic algorithm is given in Section 4.3. Section 4.4 describes the enzymatic reaction network model, Section 4.5 illustrates the application of the proposed algorithm. Finally, Section 4.9 concludes the paper.

4.2 Methodology

The methodology developed in this chapter is based on ensuring structural parameter identifiability. The aim is to define the necessary measurement sets and desirable inputs perturbations in order to render a maximum number of parameter structurally identifiable in a selected model.

In principle, the methodology, formulates and solves iteratively, a series of smaller problems then it builds around the results by considering more complex subproblems until the aim is reached or the possible measurements and perturbation sets are exhausted. Each subproblem represents a qualitative experimental design problem where a subset of measured states and/or perturbed

inputs are considered simultaneously. Each qualitative experimental design subproblem is solved by a two phase approach. The first phase, determines which reaction rates are identifiable based upon a stoichiometric analysis. In the second phase, for the identifiable rates, a method based on Lie derivatives is employed to determine which parameters are structurally globally identifiable and which are not. Both tools have been introduced in chapter 3. In this way for each proposed set of measured states and perturbed inputs a set of structurally identifiable parameters is determined. Once the subproblem is solved then the vectors of measured states and perturbed inputs are augmented with new candidates and a new problem is formulated and subsequently solved.

There are two cases considered for qualitative experimental design. In the first case, the model has been formulated but there is neither specification concerning which states are measured nor which inputs are to be perturbed. In this case, the algorithm starts simply by considering a single measured state and successively the number of measured states is increased. For perturbing inputs a similar strategy is applied e.g. no input is varied at the beginning. Then, one extra input is considered etc.

The second case deals with retrofitting an experimental design where it is assumed that, some settings are given already where a series of states are already being measured and the analysis starts from there and then builds around this set of measurements.

4.3 Workflow of structural experimental design

In Figure 4.1, the work-flow of the structural experimental design methodology is outlined. The steps are detailed for each of the two phases.

Phase 1:

- *Step 1* and *step 2* of the algorithm presented in Figure 4.1 concerns formulation of the reaction network under investigation based upon physical knowledge together with the associated stoichiometry.

- *Step 3* establishes the number of the species measured. It can be only one in the first case, as discussed above, or larger if there are already some states which are being measured in the retrofit case. This step is user dependent, which means the user can decide how many states are to be considered.

- *Step4.* In this step, given the number of measured species and based on the stoichiometry, all the possible combinations of species rendering a number of identifiable rates equal to the number of measured species are computed and listed.

- In *Step 5* from the complete list of possible combinations of measured species, the preferred species will be selected by the user. For instance,

some experimental setup may already exist and can provide measurements for the specific states considered.

- *Step 6* Determines the identifiable reaction rates for the selected measured species. A criterion introduced by Brendel et al. (2006) utilizing only the information related to the reaction stoichiometry is used.

- In *step 7* The kinetic expressions together with the associated parameters are now considered. The parameter vectors to be investigated for identifiability are formulated for each combination of identifiable reaction rates. The parameter vector corresponding to one selected set of identifiable reaction rates is investigated in the subsequent steps.

- *Step 8* computes the first order Lie derivative which represents the first term of a power series expansion.

- In *step 9* terms related to the perturbed inputs may be included. The number of parameters included in the current parameter subset is set equal to the number of measured states n.

- In *step 10* a subset of parameters ip_{set}, equal to the number of equations, is chosen from the full parameters set to be investigated.

- In *step 11* a system of equations is formed by stacking together the first order Lie derivatives, during the first iteration. When following the loop the second or third time the system of equations will be formed by considering higher order terms as well.

- In *step 12* the equation system formed during the previous step is solved analytically for all possible combinations of parameter subsets containing ip_{set} parameters taken from the full parameter set. If a unique solution is found, then the parameter set is structurally globally identifiable. After solving for all the possible combinations, the number of parameters included in the analysis is increased by one and correspondingly an extra Lie derivative term is included in *step 11*.

- *Step 13* includes higher order Lie derivatives if necessary in order to form a fully determined system of equations.

- If the system of equations solved in *step 12* did not provided a unique solution, in *step 14*, equations obtained from higher order Lie derivatives are considered in the equation vector.

- *Step 15* After all measured states are investigated, then potential additional measured states are considered by reiteration to *step 6*.

The procedure is iterated until all possible combinations of higher ip_{set} are considered. The whole loop of computing additional derivatives and solving for the various parameter combinations, is terminated if the number of parameters

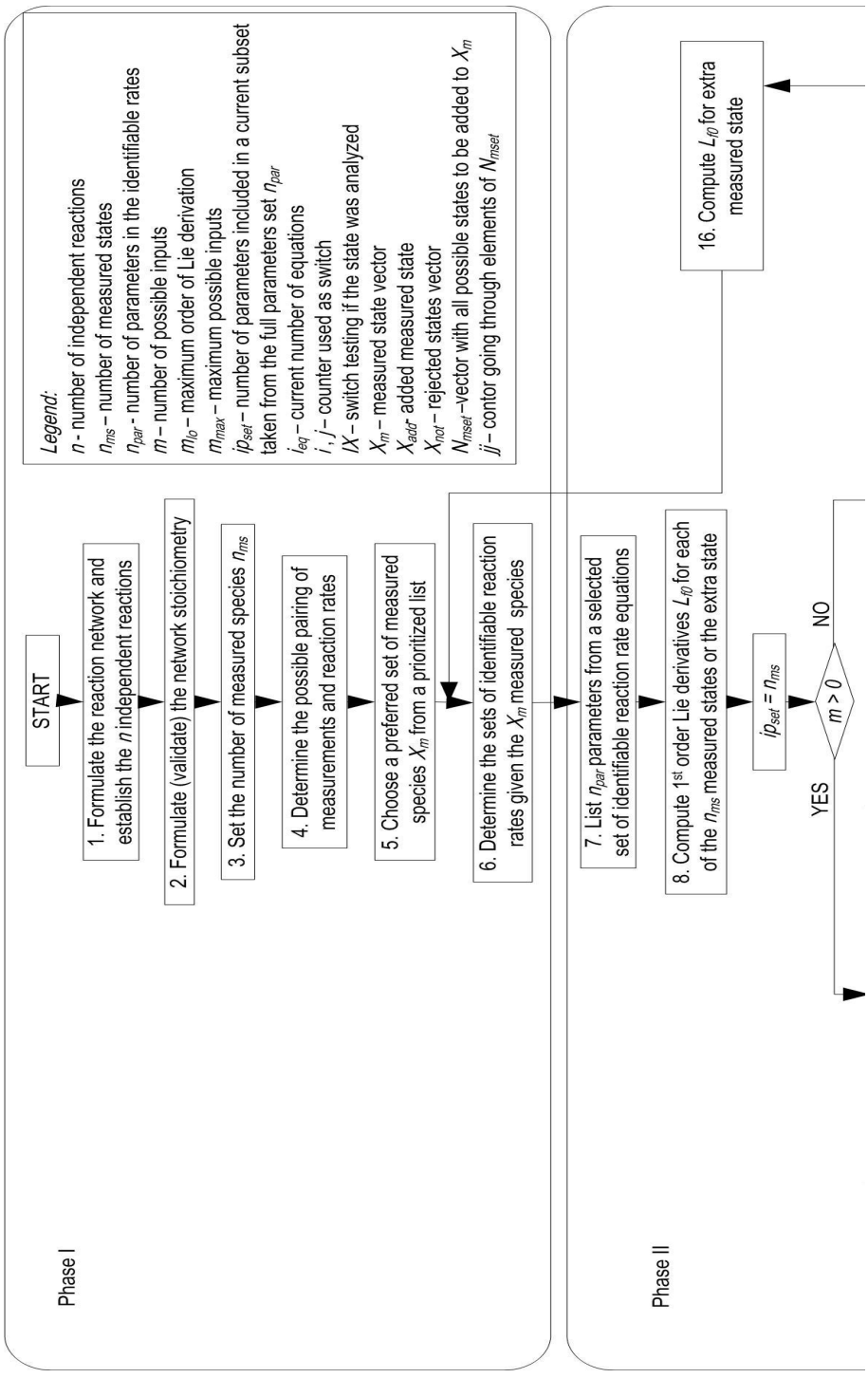


Figure 4.1: Workflow of qualitative experimental design methodology I

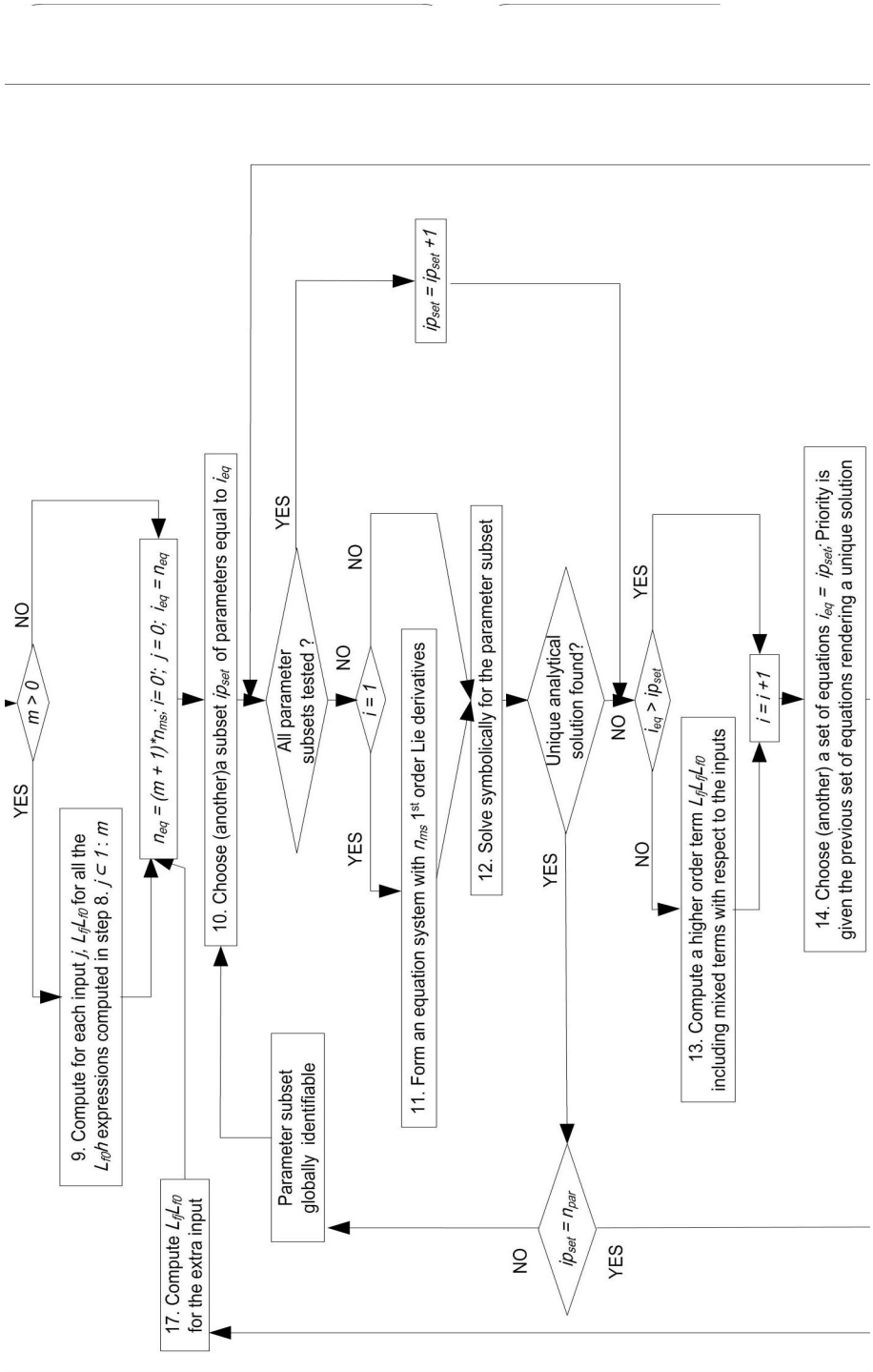


Figure 4.2: Workflow of qualitative experimental design methodology II

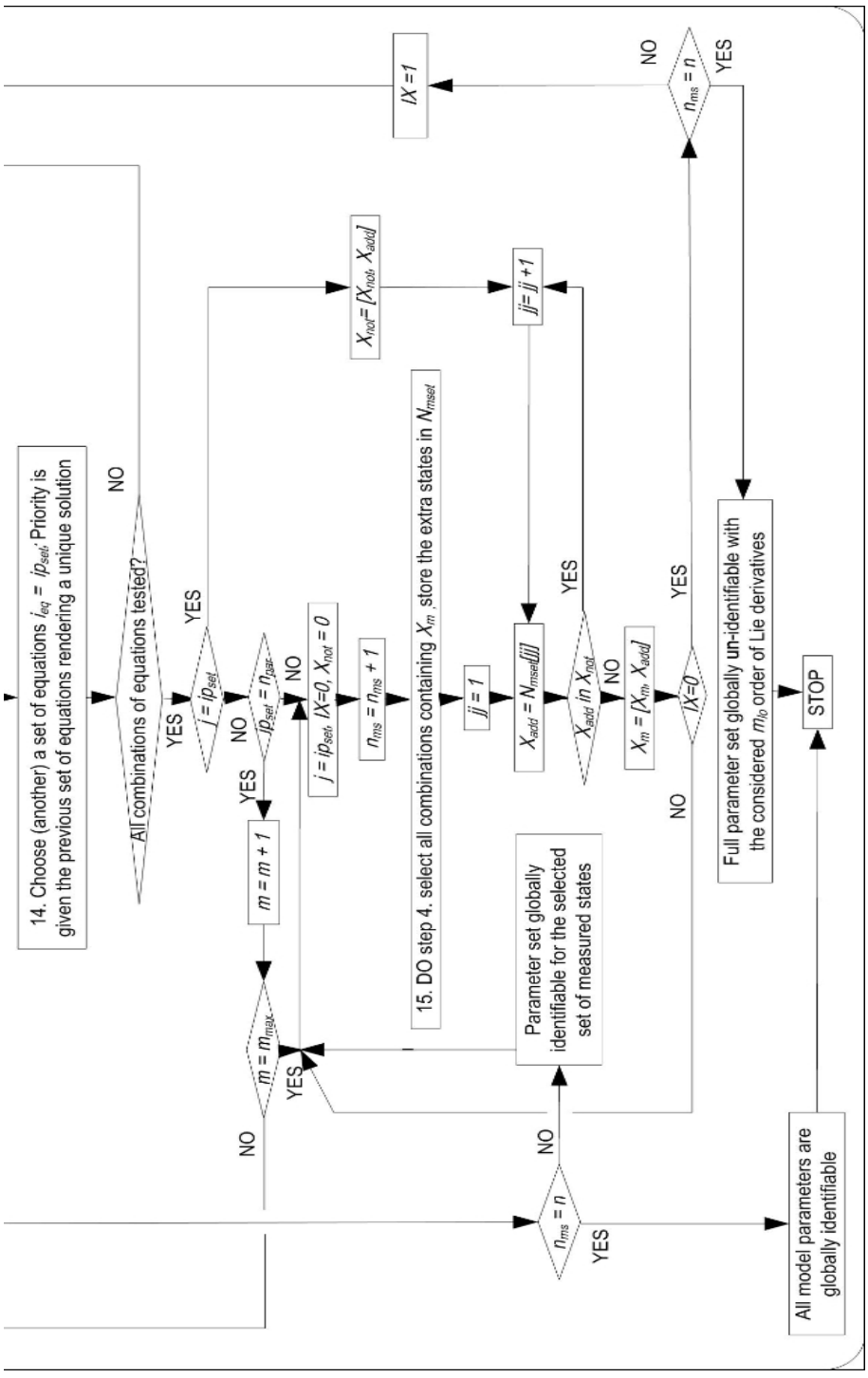


Figure 4.3: Workflow of qualitative experimental design methodology III

considered is equal to the number of parameters included in the full parameter set (n_{par}) defined in *step 7*, or if the maximum order of the Lie derivation has been reached.

If perturbed inputs are present initially, the work-flow is almost similar except during *steps 9,10 and 14* as discussed below.

Step 9 concerns the calculation of mixed first order Lie derivatives with respect to the inputs.

In *step 10* and *step 14* where the different equations are selected to form the system of equations, the mixed Lie derivatives containing the influence of the input(s) are considered with priority to lower order Lie derivatives without inputs if they are less complex.

At this point the analysis is completed for the given inputs and the selected measured states selected in *step 5*. Thus, it is necessary to choose whether the analysis should proceed by considering an extra input and/or an extra measured state.

The way to choose the extra state is as follows. In *step 15* all the possible combinations of $n_{ms} + 1$ measured states are determined. In a screening step, only combinations containing the n_{ms} previously measured states in the X_m vector species will be selected. The extra states are stored in the N_{mset} vector. This screening step reduces drastically the number of possible extra species to be measured. In an extra loop the extra species are tested individually one at the time. First, a state X_{add} is selected if it was not previously rejected and then added to the vector of measured states X_m. In *step 16* the first order Lie derivative is computed for this state, and then all the inner loops after *step 9* are followed again but with the new settings.

If an extra input is considered, in *step 17* the extra mixed first order Lie derivative is computed and then the inner loops are followed accordingly.

Once the new set of measured states is analyzed, some possible results can occur.

- The first case occurs when the extra state does not render any extra parameter identifiable. Then the state X_{add} will be included in the vector X_{not} which stores the rejected measured state candidates.

- The second case occurs when extra parameters could be rendered identifiable but there are still some potentially identifiable parameters, $ip_{set} < n_{par}$. Then the newly added state is kept in X_m. Subsequently an input is considered to render these parameters identifiable.

- The third case occurs when all remaining parameters are rendered identifiable, then the newly added state is kept in X_m.

The procedure continues until all the measured states and all possible inputs are included or until all the model parameters are found to be structurally identifiable. The value of n_{par} will increase with every new state accepted to be measured.

4.4 Example Process

There is an increasing interest in producing complex intermediates and expensive fine chemicals in the pharmaceutical industry using biochemical synthesis. Presently, only a few bio-transformation steps are involved in complex synthesis in industry, although enzymes are widely known as being specific, fast and working under mild conditions. To develop a purely enzymatic synthesis for complex molecules from simple substrates, often relatively large reaction networks are necessary. One way to achieve such a functional network is by using a so-called System of Bio-Transformations (SBT). An SBT is based on parts of the metabolic network of a microorganism containing the synthesis pathway including cofactor regeneration reactions down to the desired product, which most often is an intermediate in the metabolic network. Suitably genetically modified mutants of *E-coli* are here used to produce the metabolic network for an SBT (Schümperli et al. 2007). The mutants are grown up to a certain level. Deletion of the genes for enzymes catalyzing reactions which compete with the desired pathway is triggered a short while before the cultivation is terminated. Subsequently the cells are isolated by centrifugation. The cells are then resuspended in a buffer. The cell walls are destroyed by high pressure homogenization and the resulting particles are removed by filtration. The cell free liquid extract is recovered and then the desired protein concentration is obtained by dilution with a buffer solution. The bio-transformations are performed with the cell free extract in the production phase, combining the easy handling of a viable culture for producing the desired enzymes with the advantages of in vitro bio-transformations. For this example, the key product is Di-hydroxy-acetone phosphate (DHAP), which represents an important precursor for the production of phosphorylated, non-natural carbohydrates. The DHAP-producing SBT contains all the enzymes for the glycolysis reactions except the ones corresponding to the genes which were deleted prior to cessation of cultivation, leading to a system of high complexity. Below, the developed structural parameter identifiability analysis is illustrated on a relatively simple mass action model for the SBT.

4.4.1 Model formulation

Since the system described above is a complex dynamic system, it is not realistic to develop a "perfect model" using purely first principle engineering methods. Thus, in this work a grey-box stochastic model development framework (Kristensen et al. 2004b) is used to develop a grey-box stochastic state space model for the purpose of reaction network de-bottlenecking. The idea behind developing such a model is to have the possibility to combine the limited amount of data e.g. measured states with available first principle engineering knowledge. A large model considering all the possible intermediates will inherently contain a large number of parameters which in practice would be impossible to validate against limited available data. In such cases only an educated guess

can be used for many parameters and estimating them will lead to uncertain estimates. For practical reasons this section only describes the deterministic part of the grey-box model. A sketch representing the most important reactions occurring in the SBT is given in Figure 4.4. The sketch will form the basis for formulating the model equations.

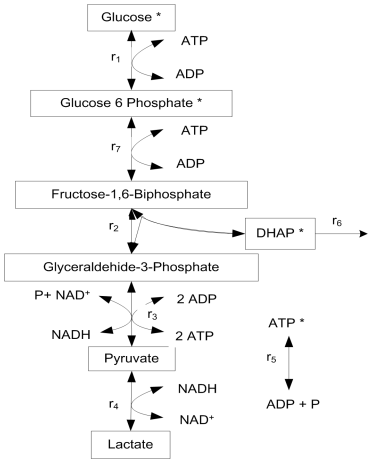


Figure 4.4: The simplified reaction network used in this analysis

There are 7 enzymatic reactions included in this network in Figure 4.4. It is assumed that except degradation of DHAP the remaining reactions are reversible. It is assumed that the reversible reaction kinetics can be described by simple mass action terms as follows:

$$
\begin{aligned}
r_1 &= r_{1maxf}c_{GL}c_{ATP} - r_{1maxb}c_{G6P}c_{ADP} \\
r_2 &= r_{2maxf}c_{F16B} - r_{2maxb}c_{DHAP}c_{G3P} \\
r_3 &= r_{3maxf}c_{G3P}c_{NAD}c_{PO4}c_{ADP} - r_{3maxb}c_{PYR}c_{NADH}c_{ATP} \\
r_4 &= r_{4maxf}c_{PYR}c_{NADH} - r_{4maxb}c_{LAC}c_{NAD} \\
r_5 &= r_{5maxf}c_{ATP} - r_{5maxb}c_{ADP} \\
r_6 &= r_{6maxf}c_{DHAP} \\
r_7 &= r_{7maxf}c_{G6P}c_{ATP} - r_{7maxb}c_{ADP}c_{F16B}
\end{aligned} \tag{4.1}
$$

For each chemical compound a dynamic mass balance is formulated as shown below. Since the reactor is a (fed-)batch reactor the model does include an input flow-rate.

$$
\begin{aligned}
\frac{dc_{GL}}{dt} &= -r_1 + \frac{F}{V}\left(c_{GLfeed} - c_{GL}\right) \\
\frac{dc_{F16B}}{dt} &= r_7 - r_2 - \frac{F}{V}c_{F16B} \\
\frac{dc_{DHAP}}{dt} &= r_2 - r_6 - \frac{F}{V}c_{DHAP} \\
\frac{dc_{G3P}}{dt} &= r_2 - r_3 - \frac{F}{V}c_{G3P} \\
\frac{dc_{PYR}}{dt} &= r_3 - r_4 - \frac{F}{V}c_{PYR} \\
\frac{dc_{LAC}}{dt} &= r_4 - \frac{F}{V}c_{LAC} \\
\frac{dc_{ATP}}{dt} &= -r_1 - r_7 + 2r_3 - r_5 - \frac{F}{V}c_{ATP} \\
\frac{dc_{NAD}}{dt} &= -r_3 + r_4 - \frac{F}{V}c_{NAD} \\
\frac{dc_{PO4}}{dt} &= -r_3 + r_5 - \frac{F}{V}c_{PO4} \\
\frac{dc_{G6P}}{dt} &= r_1 - r_7 - \frac{F}{V}c_{G6P} \\
\frac{dc_{ADP}}{dt} &= r_1 + r_7 - 2r_3 + r_5 - \frac{F}{V}c_{ADP} \\
\frac{dc_{NADH}}{dt} &= r_3 - r_4 - \frac{F}{V}c_{NADH} \\
\frac{dV}{dt} &= F
\end{aligned} \tag{4.2}
$$

Measurements equations are included as well, in practice one individual state or some simple combinations of more states are measured. Four species are currently measured during the batch experiments while for the last two measurement equations only the initial conditions are known. In the analysis, it is thus considered that only the four states are measured. Using the notation introduced above the y equations for this case are simply the states which are being measured:

$$
\begin{aligned}
y_1(c_{GL}) &= c_{GL} \\
y_2(c_{G6P}) &= c_{G6P} \\
y_3(c_{DHAP}) &= c_{DHAP} \\
y_4(c_{ATP}) &= c_{ATP}
\end{aligned} \tag{4.3}
$$

4.5 Qualitative experimental design for the enzymatic reaction network (SBT)

In the following, the methodology described in section 4.2 is illustrated on the enzymatic reaction network model described above. The analysis uses the case where a set of species are already being measured in a given setting i.e. the second case according to Section 4.2. Initially, no new (i.e besides the initially supplied amounts of substrates and cofactors) perturbing input is considered. The full stoichiometric matrix for the network described above is given below.

	1	2	3	4	5	6	7	8	9	10	11	12
	GL	ATP	G6P	ADP	F16B	DHAP	G3P	PO4	NAD	NADH	PYR	LAC
r_1	-1	-1	1	1	0	0	0	0	0	0	0	0
r_2	0	0	0	0	-1	1	1	0	0	0	0	0
r_3	0	2	0	-2	0	0	-1	-1	-1	1	1	0
r_4	0	0	0	0	0	0	0	0	1	-1	-1	1
r_5	0	-1	0	1	0	0	0	1	0	0	0	0
r_6	0	0	0	0	0	-1	0	0	0	0	0	0
r_7	0	-1	-1	1	1	0	0	0	0	0	0	0

Table 4.1: Full stoichiometric matrix N for the reaction network

The seven reactions considered in the network are independent since the rank of the stoichiometric matrix is 7. The matrix is split into two matrices corresponding to the species measured and unmeasured respectively. The matrix containing the measured species is denoted N_m, and the one containing the unmeasured species N_u.

	GL	ATP	G6P	DHAP
r_1	-1	-1	1	0
r_2	0	0	0	1
r_3	0	2	0	0
r_4	0	0	0	0
r_5	0	-1	0	0
r_6	0	0	0	-1
r_7	0	-1	-1	0

Table 4.2: Stoichiometric matrix N_m for the four originally measured species

First, all the possible combinations of four reaction rates obtained from the full set composed of seven reaction rates. The identifiable combinations of rates, for which the identifiability criterion, Eq. 3.2, was fulfilled are given below in Table 4.4:

For each of the combinations present in Table 4.4, the analysis has been performed. The combinations of parameter with maximum of parameters rendered identifiable are given below in tabular form.

	ADP	F16B	G3P	PO4	NAD	NADH	PYR	LAC
r_1	1	0	0	0	0	0	0	0
r_2	0	-1	1	0	0	0	0	0
r_3	-2	0	-1	-1	-1	1	1	0
r_4	0	0	0	0	1	-1	-1	1
r_5	1	0	0	1	0	0	0	0
r_6	0	0	0	0	0	0	0	0
r_7	1	1	0	0	0	0	0	0

Table 4.3: Stoichiometric matrix N_u, for the unmeasured species

Combination	Identifiable Rates			
1	r_1	r_5	r_6	r_7
2	r_1	r_3	r_6	r_7
3	r_1	r_2	r_5	r_7
4	r_1	r_2	r_3	r_7

Table 4.4: Identifiable combinations of reaction rates for the enzymatic reaction network with the four original measurements

4.5.1 Combination r_1, r_5, r_6, r_7

The Lie derivatives as introduced in chapter 3, Eqs: 3.4-3.5, are formulated in order to obtain the algebraic equations required for the structural identifiability analysis. The algebraic equations are grouped and solved analytically in order to assess the identifiability property of the parameters.

Table 4.5 contains all the combinations of equations formed with Lie derivatives for which a unique solution was found for the first combination of identifiable rates.

Comb.	r_{1maxb}	r_{1maxf}	r_{5maxb}	r_{5maxf}	r_{6maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_2$	$L_f h_1$	$L_f h_3$	$L_f h_4$	$L_f L_f h_2$	$L_f L_f L_f h_1$	$L_f L_f L_f h_4$

Table 4.5: One combination of Lie derivatives for which a unique solution was found with the four original measurements

4.5.2 Combination r_1, r_3, r_6, r_7

When selecting all the seven parameters appearing in the selected set of reaction rates kinetic equations a unique solution could be found for multiple combinations of the Lie derivatives. This time it was possible to obtain a unique solution by using only the first and the second order Lie derivatives. Table 4.6 contains one combination of equations formed with Lie derivatives for which a unique solution was found for the second combination of identifiable rates.

Comb.	r_{1maxf}	r_{1maxf}	r_{3maxf}	r_{3maxb}	r_{6maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_1$	$L_f h_3$	$L_f L_f h_1$	$L_f h_4$	$L_f L_f h_2$	$L_f L_f h_4$	$L_f L_f h_3$

Table 4.6: One combination of Lie derivative for which a unique solution was found

4.5.3 Combination r_1, r_2, r_5, r_7

Table 4.7 contains one of the combinations of equations formed with Lie derivatives for which a unique solution was found for the third combination of identifiable rates. It can be noticed that all the eight parameters were rendered identifiable.

Comb.	r_{1maxb}	r_{1maxf}	r_{2maxb}	r_{2maxf}	r_{5maxb}	r_{5maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_1$	$L_f h_3$	$L_f L_f h_2$	$L_f L_f h_3$	$L_f L_f h_1$	$L_f L_f L_f h_4$	$L_f L_f L_f h_3$	$L_f h_4$

Table 4.7: One combinations of Lie derivatives for which a unique solution was found

4.5.4 Combination r_1, r_2, r_3, r_7

Table 4.8 contains one of the combinations of equations formed with Lie derivatives for which a unique solution was found for the last combination of identifiable rates. It can be noticed that for this combination all the eight parameters were rendered identifiable.

Comb.	r_{1maxb}	r_{1maxf}	r_{2maxb}	r_{2maxf}	r_{3maxb}	r_{3maxf}	r_{7maxb}	r_{7maxf}
1	$L_f h_1$	$L_f h_3$	$L_f h_2$	$L_f L_f h_3$	$L_f L_f h_1$	$L_f L_f h_4$	$L_f L_f L_f h_1$	$L_f L_f L_f h_4$

Table 4.8: One combination of Lie derivatives for which a unique solution was found

Having analyzed all four identifiable reaction rate combinations, and determined that all the parameters could be structurally identifiable for the third and fourth combinations. Thus the question is which additional state to measure?

4.5.5 One extra measured state

The method will be illustrated using the fourth combination r_1, r_2, r_3, r_7 above. It is desired first to investigate r_4 to be included included in the set of identifiable reaction rates. Considering the five reaction rates, all possible combinations of 5 measured species including the original ones have been determined and are listed in Table 4.9. In the following, number 12 is for Lactate, 11 for Pyruvate, 10 for NADH and 9 for NAD. according to the stoichiometric matrix, Table 4.1.

No	Species				
1	GL	ATP	G6P	DHAP	LAC
2	GL	ATP	G6P	DHAP	PYR
3	GL	ATP	G6P	DHAP	NADH
4	GL	ATP	G6P	DHAP	NAD

Table 4.9: Combinations of species rendering the five reaction rates identifiable

For practical reasons, i.e. ease of measurement, Pyruvate is preferred to other compounds that would render the five reaction rates identifiable. Analyzing the results in Tables: 4.5 and 4.8, it seems that L_fh_1 , L_fh_2, L_fh_3 occurs in all the combinations rendering a unique solution. L_fh_4 occurs in almost all the combinations. Thus all the first order terms will be considered. As for the second order terms $L_fL_fh_2$, $L_fL_fh_3$, $L_fL_fh_4$ occurs quite frequently. $L_fL_fh_1$ occurs alternating with L_fh_4. Thus these eight equations will be used with priority. First and second order Lie derivative are computed for Pyruvate L_fh_5. Having four measured species rendered 8 parameters globally identifiable. Thus is expected to have at least 9 parameters globally identifiable after including Pyruvate as a new measurement. When considering combinations of 9 parameters simultaneously it was indeed possible to find a unique solution for many parameter sets. Table 4.10 contains one set of globally identifiable parameters.

Eq	L_fh_5	L_fh_2	L_fh_1	L_fh_3	$L_fL_fh_1$	$L_fL_fh_2$	$L_fL_fh_3$	$L_fL_fh_4$	$L_fL_fh_5$
1	r_{1maxb}	r_{1maxf}	r_{2maxb}	r_{2maxf}	r_{3maxb}	r_{3maxf}	r_{4maxb}	r_{4maxf}	r_{7maxb}

Table 4.10: One combination of parameters for which a unique solution was found

In the next step an unfruitful attempt was performed to solve for all 10 parameters occuring in the corresponding kinetic equations. Including higher order Lie derivatives in the equation set did not render all the parameters to be globally identifiable either. Thus one extra parameter was rendered globally identifiable and Pyruvate is included in the measured states vector.

4.5.6 Glucose feed input and five measured states

An input perturbation to the process is considered at this moment. The considered perturbed input is the Glucose feed. By considering the input to the process, the parameter vector to search for remains the same, but it is expected that at least one extra parameter becomes globally identifiable. Besides the first order mixed terms, first and second order Lie derivatives are used when forming the system of equations. One of the combinations of equations rendering all 10 parameters globally identifiable is given in the Table 4.11. In Table 4.11, the mixed Lie derivative is denoted with a g corresponding to the input function.

Based upon the information in Table 4.11 only one mixed term was necessary

No	r_{1maxf}	r_{1maxb}	r_{2maxb}	r_{2maxf}	r_{3maxb}	r_{3maxf}	r_{4maxb}	r_{4maxf}	r_{7maxb}	r_{7maxf}
2	L_fh_1	L_fh_3	L_fh_4	L_fh_5	$L_fL_fh_1$	$L_fL_fh_2$	$L_fL_fh_3$	$L_fL_fh_4$	$L_fL_fh_5$	$L_gL_fh_3$

Table 4.11: One combination of Lie derivative for which a unique solution was found for ten parameters when perturbing Glucose feed with five measurements

to render all 10 parameters identifiable. In fact the search could have been stopped as soon as one combination of equations rendering all the parameters globally identifiable was found.

4.6 Considering a sixth measured state

The aim is now to render six reaction rates identifiable. In order to do so one extra state is included in the set of measured states. A few options are available after determining again all the possibilities in terms of extra states to be measured. First of all the aim is to render r_5 identifiable. Table 4.12 gives all the possible combinations of six measured species, including the aforementioned five.

No	Species					
1	GL	ATP	G6P	DHAP	PYR	LAC
2	GL	ATP	G6P	DHAP	PYR	PO4
3	GL	ATP	G6P	DHAP	PYR	G3P

Table 4.12: Combinations of species rendering six reaction rates identifiable with six measurements

Analyzing the results it seems like there are three possibilities. In the following Table 4.1 component number 12 is Lactate, 8 is Phosphate (PO4) and 7 is G3P.

4.6.1 Lactate as sixth measured state

In a first attempt, Lactate (LAC) is considered as additional measured state. The procedure is followed first without input and then considering a perturbed input. For the case with no input perturbation, the first two order Lie derivatives have been computed L_fh_6 and $L_fL_fh_6$. Having no input, at least ten parameters should be globally identifiable, thus the search starts with 10 parameters. Table 4.13 lists the equations rendering a unique solution and the combination of parameters founded to be globally identifiable.

No	L_fh_1	L_fh_2	L_fh_3	L_fh_5	$L_fL_fh_1$	$L_fL_fh_2$	$L_fL_fh_3$	$L_fL_fh_4$	$L_fL_fh_5$	$L_fL_fh_6$
1	r_{1maxf}	r_{1maxb}	r_{2maxf}	r_{2maxb}	r_{3maxf}	r_{3maxb}	r_{4maxf}	r_{4maxb}	r_{7maxf}	r_{5maxf}

Table 4.13: One combination of ten parameters for which a unique solution was found with six measurements

An attempt to find 11 parameters globally identifiable related to the considered six reaction rates was unfruitful. Instead, only multiple solutions have been found, thus rendering the parameter set to be only locally identifiable.

4.6.2 Six measured states and the Glucose input

Considering Lactate as the sixth measured state a perturbed input is investigated. However including this input, no more than 10 parameters could be rendered globally identifiable, thus Lactate is discarded as a candidate for measured states.

4.6.3 G3P as sixth measured state

When including another sixth measured state, according to Table 4.12 a second possibility is G3P. Again, the first and second order Lie derivatives have been computed and included in the selected equations. This time the search was initiated with a parameter vector containing 11 parameters related to the six reaction rates. The search for a globally identifiable set of parameters was fruitful and the list of parameters/equations used are included in Table 4.14.

No.	L_fh_1	L_fh_2	L_fh_3	L_fh_5	$L_fL_fh_1$	$L_fL_fh_2$	$L_fL_fh_3$	$L_fL_fh_4$	$L_fL_fh_5$	$L_fL_fh_6$	L_fh_6
1	r_{1maxf}	r_{1maxb}	r_{2maxf}	r_{2maxb}	r_{3maxf}	r_{3maxb}	r_{4maxf}	r_{4maxb}	r_{5maxf}	r_{7maxf}	r_{5maxb}

Table 4.14: A combination of eleven parameters for which a unique solution was found with six measured states

The search for a set of 12 parameters globally identifiable was un-fruitfull. As results only multiple solutions have been found, thus the parameter set could be deemed to be only locally identifiable.

4.6.4 Six measured states and the Glucose input

The Glucose input perturbation is included again in the analysis. The first order mixed term of the Lie derivatives is included and a set of equations containing one mixed term is found to render a unique solution for a parameter set of 12 parameters.

L_fh_1	L_fh_2	L_fh_3	L_fh_5	$L_fL_fh_1$	$L_fL_fh_2$	$L_fL_fh_3$	$L_fL_fh_4$	$L_fL_fh_5$	$L_fL_fh_6$	L_fh_6	$L_gL_fh_5$
r_{1maxf}	r_{1maxb}	r_{2maxf}	r_{2maxb}	r_{3maxf}	r_{3maxb}	r_{4maxf}	r_{4maxb}	r_{5maxf}	r_{5maxb}	r_{7maxf}	r_{7maxb}

Table 4.15: Parameter sets for which a unique solution was found with six measured states and an input perturbation

Thus including G3P and considering an input perturbation rendered all the parameters related to the six reaction rates to be globally identifiable.

4.7 Seven measured states

There is only one reaction rate and one parameter left thus it is desirable to include an extra measured state. The procedure is reiterated and the combinations of seven measured species containing the previous six, have been determined and listed in Table 4.16.

No	Species						
1	GL	ATP	G6P	DHAP	G3P	PYR	LAC
2	GL	ATP	G6P	DHAP	G3P	PYR	PO4
3	GL	ATP	G6P	DHAP	G3P	PYR	F16B

Table 4.16: Combinations of species rendering the seven reaction rates identifiable

When analyzing for all the seven reaction rates then, three combinations were found to theoretically render all the seven reaction rates identifiable. All of them contains the previous six measured species. Number 8 corresponds to Phosphate, number 5 corresponds to F16B and finally number 12 to Lactate. Since Lactate did not improved the identifiability previously it will not be considered again. Thus two possibilities are left.

4.7.1 PO4 included as measured state

According to the stoichiometric matrix for the measured species, including PO4 in the measured states should render all 7 reaction rates identifiable. In other words it should be possible to find a globally identifiable parameter set containing parameters from all the seven rates, i.e. at least a parameter set containing r_{6maxf} should be globally identifiable. A set of 12 parameters was found to be globally identifiable but the parameters set did not included any parameter related to r_6. An attempt to find a reduced set of 11 parameters to be globally identifiable, containing r_{6maxf} was performed but it was unfruitful. Only multiple solutions were obtained and therefore the parameter set, including the r_{6maxf} is locally identifiable.

4.7.2 PO4 included and the Glucose input

In an attempt to render all the parameters globally identifiable an input as Glucose feed is considered. Several combinations of equations have been used and solving them, only resulting in multiple or undetermined solutions. However by no means it was possible to render a combination of at least 12 parameters containing r_{6maxf} to be globally identifiable.

4.7.3 F16B instead of PO4

The last available possibility is the inclusion of F16B in the measured states. F16B was considered as measured state and the first and second order Lie

derivatives have been computed. The search was initiated with combinations of 11 parameters including r_{6maxf} but no combination was found to be globally identifiable. Again only multiple solutions were found thus the parameter sets are only locally identifiable.

4.7.4 F16B instead of PO4 and Glucose input

The last attempt to render a set of parameters at least containing r_{6maxf} was to include an input perturbation. Considering the Glucose input did not help either in finding a unique solution for the parameter set. At least with the considered combinations of equations. By considering a seventh measured species it was not possible to render r_{6maxf} globally identifiable together with the rest of the parameters. Neither including PO4 as a measured species, nor including F16B could render the parameter set including r_{6maxf} globally identifiable. However including PO4 as a measured state helped identifiability of the rest of parameters. By using the Lie derivatives obtained from this state it was possible to render all the remaining 12 parameters globally identifiable without including a Glucose feed input perturbation.

4.8 Discussion

First, the four measured states were considered in the analysis. Pyruvate was included in the measured states and it improved the identifiability by rendering a maximum of nine parameters globally identifiable. The ninth parameter compared with the previous step was a parameter related to r_4 which was in accordance with the conclusions of the stoichiometry analysis. In the next step the Glucose input has been considered additionally. Including the input perturbation improved again the identifiability for 10 parameters corresponding to the reaction rates: r_1, r_2, r_3, r_4, r_7. In order to investigate the identifiability of the parameters of r_5, the stoichiometric analysis showed that LAC or G3P could be included in the measured states. G3P was considered in the set of measured species and rendered one extra parameter globally identifiable. An input perturbation has been considered again and this helped the identifiability by rendering all parameters related to reaction rates: r_1, r_2, r_3, r_4, r_5, r_7 globally identifiable. According to stoichiometric analysis applied for 7 measured species PO4 or F16B could be included in the measured states. Including PO4 improved the identifiability, by rendering 12 parameters globally identifiable without input perturbation but the parameter set did not contained r_{6maxf}. Considering the input again did not bring new qualitative information either. The fact that having $PO4$ measured rendered extra parameters globally identifiable shows interactions that the stoichiometric analysis could not detect.

The last step was repeated for F16B as an additional measured state. This choice did not improve the identifiability since not even a combination of 11 parameters became globally identifiable using Lie derivatives derived from this

state. Considering a perturbed input did not help either. The optimal set of outputs-inputs that can render maximum of 12 parameters identifiable is five measured states: GL, G6P, ATP, DHAP, G3P and the glucose input or the six measured states GL, G6P, ATP, DHAP, G3P and PO4.

The fact that the parameter set containing r_{6maf} can not be rendered globally identifiable can be related to the fact that for the reaction degrading DHAP no product is defined. The analysis could be initiated using a different set of states or maybe considering perturbation in other inputs. The stoichiometric analysis shows all the combinations of measured species and reaction rates once the stoichiometric matrix and the number of measured states is provided. The way to choose an extra measurement or an input perturbation ultimately depends on what is more expensive or difficult to perform in practice. Qualitative experimental design and qualitative identifiability analysis are in fact just two related issues of model development. Identifiability analysis starts out from a given setting in terms of measured states and perturbed inputs to determine the qualitatively identifiable parameters. Qualitative experimental design tries to determine the minimum number of measured species and(or) perturbed inputs to render a maximum number of parameters identifiable. Furthermore the states to be considered for measurement and the inputs to be perturbed are determined.

The proposed approach for qualitative experimental design has been most useful for pinpointing which potential model structures and measurement sets could render the model parameters globally identifiable. However, when working with more realistic enzyme kinetic expressions the parameter identifiability analysis becomes more complex in that a rather nonlinear set of equations should be solved symbolically. Hence for such cases a simpler methodology can be desirable.

4.9 Conclusions and Future work

A methodology for *a-priori* assessing structural identifiability of the model parameters and performing qualitative experimental design for reaction networks models is introduced. This methodology, is illustrated through a real life case study on a model derived from an enzymatic reaction network. The methodology is based on incorporating knowledge about the reaction network in terms of reaction stoichiometry and kinetics to ensure global model parameters identifiability in a systematic manner.

One drawback is that there is no upper bound for the order of the Lie derivation required to achieve a system of algebraic equations which render the parameter set globally identifiable. For the case where inputs are perturbed, the number of equations available is much larger even when using only first or second order Lie derivatives of the measured states to form the system of algebraic equations. The complexity of the equations is reduced significantly and the need for the higher order derivative information is thereby reduced.

The results of this analysis are qualitative, thus only deterministic aspects are considered.

However, even with all limitations the results of this procedure provides a very valuable *a-priori* basis for quantitative experimental design for parameter estimation and subsequent model (in-)validation.

5

Quantitative experimental design for parameter estimation

Abstract

An important step during dynamic model development is to design informative experiments. Design of experiments for dynamic models may be decomposed into qualitative experimental design and quantitative experimental design. Qualitative experimental design determines which states to be measured and which inputs to be perturbed. While the aim of quantitative experimental design is to maximize the information for parameter estimation to reduce the uncertainty of the parameter estimates by experimental design variables such as the initial values of the system states, the input profiles and the sampling times. The connection between quantitative design of experiments and parameter estimation is given by the Kramer-Rao inequality which provides a lower bound for the parameter uncertainty for unbiased estimates. This contribution addresses the problem of quantitative experimental design for models formulated using stochastic differential equations. The proposed method uses an iterated extended Kalman Filter for integrating the system of stochastic differential equations which is an integral part in computing the maximum likelihood function used further in the evaluation of the Fisher Information matrix. The method is demonstrated to provide a sound design for a benchmark fed-batch reactor experimental design.

5.1 Introduction

For many applications such as design, control, optimization, in chemical and biochemical engineering it is desirable to develop quantitative mechanistic dynamic process models used to describe and understand the process under investigation. In principle physical modeling generates complex nonlinear dynamic models in the form of systems of ordinary differential algebraic equations or stochastic differential equations containing many parameters which needs to be

estimated. For some parameters, like saturation constants if enzymatic catalysis is employed, or diffusion or heat transfer coefficients, numerical values can be available in the literature or from previous experiments, for others only limited knowledge is available. It is therefore necessary to perform parameter estimation and validation of the model performance against experimental data. Generating and collecting experimental data is an expensive operation and in order to reduce the amount of data required, experimental design is a prerequisite step. Performing model based experimental design implies maximizing the information content for parameter estimation while minimizing the number of experiments. Experimental design consists in principle of two phases: qualitative experimental design and quantitative experimental design as discussed in Chapter 4. Qualitative experimental design concerns determining which input variables should be varied and which outputs should be measured in the experiment in order to be able to render the unknown model parameters identifiable assuming a known model structure. Quantitative experimental design, based upon the selected inputs and outputs, concerns determining the input profiles, the initial values of the states and the optimal measurement sampling times.

Quantitative experimental design for dynamic models has been addressed by several authors. A key review on the topic is provided by Walter and Pronzato (1990) where the focus is mainly on linear models. When performing quantitative experimental design, however, for linear dynamic systems it is possible to convert the model from time domain into the frequency domain (Sadegh et al. 1995), and then determine the optimal set of frequencies for inputs. For nonlinear dynamic models which are nearly linear, a common approach is to linearize the model around the operational state of the process and then to apply the same strategy as for linear models. Körkel et al. (1999), Bauer et al. (2000), developed the concepts and algorithms for quantitative experimental design for nonlinear dynamic models in forms of systems of ordinary differential or differential-algebraic equations. Asprey and coworkers, (Asprey and Macchietto 2000), (Asprey and Macchietto 2002) further address the problem of quantitative experimental design for nonlinear dynamic systems.

If the models structure is deficient or more model structure candidates are available, then it is possible to design experiments to discriminate in an optimal way between two or more possible candidates (Asprey and Macchietto 2000, 2002, Chen and Asprey 2003, Chen et al. 2004).

The focus of this work is however the quantitative experimental design for parameter estimation. Depending on the level of prior knowledge available about the model parameters two approaches are considered. In the first approach some information about the parameter distribution considered in the optimization criterion Walter and Pronzato (1990) obtaining either an expected value design or a classical D design. In this case the choice of experimental decision variables maximizes the expected value over the population of possible parameter values (Asprey and Macchietto 2002). The experiments designed in this way are good on average but can be poor for some parameter values (Walter and Pronzato 1987, Asprey and Macchietto 2000).

In the second approach the idea is to compute a design which tries to maximize the informational content for any value of the parameters. Using this approach the only prior information about θ is the admissible domain Θ. In this way the design attempts to ensure acceptable performance for all possible values of θ. The max-min approach have had a limited use due to the burdensome computations associated with. In order to circumvent this problem Asprey and Macchietto (2000), Asprey and Macchietto (2002) proposed a sequential algorithm derived from the worst case approach and they denote this as robust R-optimal experimental design.

In a subsequent paper, Benabbas et al. (2005) investigated a criterion including information about the curvature of the response surface. In order to further account for non-liniarity of the system, a criterion based on global sensitivity analysis has been introduced more recently by Rodriguez-Fernandez et al. (2007).

However for stochastic differential equation models the topic of quantitative experimental design constitutes a subject of limited investigations so far. An approached exploiting the frequency domain has been employed by Sadegh et al. (1994) for linear systems.

In the present contribution quantitative experimental design is developed for models described by input affine stochastic differential equations. *The purpose is to address the issue of including prior knowledge about the parameters and possibly also their distribution function.*

The main contribution of this chapter is the development of a procedure for performing quantitative experimental design for parameters estimation and model (in-)validation for processes described by stochastic differential equations. The classical local D-design criterion, i.e. criteria given in Eqs. 5.9–5.10, is used here based upon on an estimate of the Fisher Information Matrix (FIM) for SDEs models. The criterion is denoted (SD)-optimal stochastic D design.

The developed software implementation is applied to a simple engineering example represented by a bio-reactor model as a benchmarking case.

The remainder of the chapter is organized as follows. Section 5.2 describes the systematic methodology, to perform quantitative experimental design for stochastic differential equations models, then Section 5.3 presents the workflow of the implemented algorithm. Section 5.4 introduces the benchmarking case study. Section 5.5 discusses the results and concludes the chapter.

5.2 Methodology

The quantitative experimental design for stochastic differential equations is solved as a dynamic optimization problem where the design variables are the initial values of the stochastic states, the input profiles, the sampling times with the objective to maximize the information content in the measurements, desired for parameter estimation. The information content for a specific model is stored in the Fisher Information Matrix. The Fisher Information Matrix

(FIM) is computed as the negative expectation of the second derivative of the log-likelihood function. The objective function for quantitative experimental design case is a standard scalar measure of the FIM e.g. the determinant of the matrix (D criterion), or the ratio of the maximum and the minimum eigenvalues i.e. the (A criterion) or another metric. Since the experimental data are to be collected for parameter estimation, initially the focus will be on all parameters, thus the D-optimal experimental design will be considered in this contribution. The work-flow used for calculating the design objective is illustrated later in Figure 5.2 and described in section 5.3 .

In this work the inputs are assumed piecewise constant. The notation used here:

1. the initial value of the state x_{0i}, $i \in (0, N)$
2. the sampling time t_{sp_i}, $i \in (0, n_{sp})$
3. the switching time for the inputs $t_{sw\,i,j}$, $i \in (1, n_u)$, $j \in (1, n_{sw})$
4. the values of the input $u_{i,j}$, $i \in (1, n_u)$, $j \in (1, n_{sw})$

In order to obtain a problem of a manageable size, the sampling time will be considered the same for all measured states and the switching time will be considered the same for the inputs. In an extended formulation the sampling time could be allowed to be different for each of the measured states. The same can be applied to the switching times. Various constraints can be imposed on the design variables:

1. bounds on ${x_0}_i^L \leq x_{0i} \leq {x_0}_i^U$, $i = 1, \ldots, M$
2. bounds on $u_{i,j}^L \leq u_{i,j} \leq u_{i,j}^U$, $i = 1, \ldots, n_u$, $j = 1 \ldots, n_{sw}$
3. inequality constraints $\Delta {t_{sp}}_l^{min} \leq t_{sp_l} - t_{sp_{l-1}} \leq \Delta {t_{sp}}_l^{max}$, $l = 1, \ldots, n_{sp}$
4. inequality constraints $\Delta {t_{sw}}_{i,j}^{min} \leq t_{sw\,i,j} - t_{sw-1\,i,j} \leq \Delta {t_{sw}}_{i,j}^{max}$, $i = 1, \ldots, n_u$, $i = 1, \ldots, n_{sw}$
5. $t_{sw\,i,nsw} \leq t_{sp_{nsp}}$, $i = 1, \ldots, n_u$

In this framework the difficulty arises from two points. The first is related to the nature of the stochastic differential equations and subsequently the evaluation of the log-likelihood function and the second from the evaluation of the expectation of the second derivative of the log-likelihood function. The optimization problem to be solved is thus a constrained nonlinear and stochastic optimization problem.

5.2.1 Optimization

For solving this optimization problem a differential-evolution (DE) based algorithm is employed. The motivation for this choice stems from the fact that it is a global optimization algorithm, which can handle constraints and the objective function can be noisy. Moreover it is reported to be one of the fastest algorithm for a global search in terms of number of function evaluations needed to achieve acceptable convergence (Storn and Price 1996).

The original version of differential evolution was introduced by Storn and Price (1995), Storn (1996), Storn (1997). Important extensions for handling multiple optimization criteria as well as constraints were introduced by Lampinen (2002). Particularly, Kukkonen and Lampinen (2005) has introduced the so called generalized differential evolution strategy (GDE), where an arbitrary number of objectives and constraints can be handled.

Differential evolution works with a population P_G, of parameter vectors to be optimized. Within the population, each member or candidate represents a vector of the design variables φ to optimize. These candidate solutions evolves over a number of generations G_{max} which is specified in advance.

A constant population, consisting of NP individuals characterized by real-valued vectors $\varphi_{i,G}$, is used. The i indexes the population and the subscript G designates the population generation, i.e:

$$P_G = (\varphi_{1,G}, \ldots, \varphi_{NP,G}), \; G = 0, \ldots, G_{max} \tag{5.1}$$

Furthermore, each vector contains N_φ real parameters:

$$\varphi_{i,G} = \left(\varphi_{1,i,G}, \ldots, \varphi_{N_\varphi,i,G}\right), \; i = 1, \ldots, NP, \; G = 0, \ldots, G_{max} \tag{5.2}$$

The population $P_{G=0}$ is initialized with random values chosen from within the given boundaries:

$$\varphi_{j,i,0} = rand_j\,[0,1] \cdot \left(\varphi_j^{(U)} - \varphi_j^{(L)}\right) + \varphi_j^{(L)} \; i = 1, \ldots, NP, \; j = 1, \ldots, N_\varphi \tag{5.3}$$

where $rand_j\,[0,1]$ denotes a uniformly distributed random value within the range: $[0,1]$. The random value is generated for each parameter (j). The vectors in the current population, P_G, are randomly sampled and combined to create candidate vectors for the next generation population, P_{G+1}. First, a population of candidates of trial vectors P'_{G+1} containing $\varphi'_{i,G+1}$ vectors $(i = 1, \ldots, NP)$ and each vector $\varphi'_{i,G+1}$ with the elements $= \varphi'_{j,i,G+1}$, $(j = 1, \ldots, N_\varphi)$, following the notation of Price (1999), $DE/rand/1/bin$, is generated as follows:

$$\varphi'_{j,i,G+1} = \begin{cases} \varphi_{j,r_3,G} + F \cdot (\varphi_{j,r_1,G} - \varphi_{j,r_2,G}), \; if \; rand_j\,[0,1) \leq CR \; \vee j = k \\ \varphi_{j,i,G}, \; otherwise \end{cases} \tag{5.4}$$

where, $i = 1, \ldots, NP$, $j = 1, \ldots, N_{\varphi}$, $k_i \in \{1, \ldots, NP\}$, $r_1, r_2, r_3 \in \{1, \ldots, NP\}$, and $r_1 \neq r_2 \neq r_3 \neq i$, and $CR \in [0, 1]$, $F \in (0, 1]$.

The randomly chosen indexes, r_1, r_2 and r_3, are different and also different from the running index, i. New, random, integer values for r_1, r_2 and r_3, are chosen for each value of the index i, i.e., for each individual trial vector. The index, k, refers to a randomly chosen chromosome (a chromosome is an optimization parameter) which is used to ensure that each individual trial vector, $\varphi'_{i,G+1}$, differs from its counterpart in the previous generation, $\varphi_{i,G}$ by at least one parameter. A new, random, integer value is assigned to k prior to the construction of each trial vector, i.e., for each value of the index i.

F and CR are the control parameters of the differential evolution (DE) algorithm. Like NP, both values remain constant during the search. F is a real-valued factor in the range $(0, 1]$ that scales the differential variations. The upper limit on F has been empirically determined as suggested by Storn (1997) and is not a strict limit. CR is a real-valued crossover factor in range $[0, 1]$ that controls the probability that a trial vector parameter will come from the randomly chosen, mutated vector, $\varphi'_{i,G+1}$ (with elements $\varphi'_{j,i,G+1}$), instead of from the current population vector, $\varphi_{i,G}$. Generally, both F and CR affect the convergence velocity and robustness of the search process (Lampinen 2002). Their optimal values depend both on the objective function characteristics and on the population size, NP. Suitable values for F, CR and NP are reported to be found by trial-and-error after a few trials by Storn and Price (1995).

Once a trial generation φ'_{G+1} has been obtained, a selection operation is performed in order to decide which population members will get into the next generation population φ_{G+1}. The comparison of a member from the trial population, $\varphi'_{i,G+1}$, with a corresponding member of the current population $\varphi_{i,G}$ is based on the following rules.

1. If $\varphi_{i,G}$ and $\varphi'_{i,G+1}$ are infeasible, then the sum of the constraint violation is computed for both parameter vectors. The vector rendering a smaller total violation is selected as member in the next generation population $\varphi_{i,G+1}$.

2. If $\varphi_{i,G}$ is not feasible while $\varphi'_{i,G+1}$ is feasible, the feasible vector is selected as member in the next generation population $\varphi_{i,G+1}$. The inverse is valid as well.

3. If $\varphi_{i,G}$ and $\varphi'_{i,G+1}$ are both feasible, then the vector which improves the cost function is selected as member in the next generation population $\varphi_{i,G+1}$.

These selection rules are actually analogous with the concepts of Pareto-optimality (Kukkonen and Lampinen 2005), (Kukkonen and Lampinen 2006). It should be mentioned that the cost function is evaluated just for the case where both parameter vectors are feasible. In case of infeasible solutions, the selection rule does not compare the objective function values but only the

evaluation of the constraints is required. In the original generalized differential evolution algorithm presented in Kukkonen and Lampinen (2005), the way to discriminate between two infeasible individuals was to compare the violation for just one of the constraints, however in practice it was noticed that sometimes for problems with a large number of constraints it is more beneficial to consider the sum of the violations over all constraints. Therefore for finding the first feasible solution an effective selection pressure is applied by only accepting feasible candidates in the next population generation. This results in a fast convergence to feasible regions of the search space (Kukkonen and Lampinen 2005).

At least six different strategy schemes based on the one reproduced in Eq. 5.4 have been proposed. Various strategies mixing the best candidate obtained in the current generation with randomly generated members have been implemented and investigated by(Price et al. 2005) . Here the first strategy, $DE/rand/1/bin$ was used.

5.2.2 Evaluating the design objective

The information content for a specific model is represented by the Fisher Information Matrix. The Fisher Information Matrix (FIM) is computed as the expectation of the second derivative of the log-likelihood function as given in Eq. 5.5.

$$FIM\left(\theta,\varphi\right)=E_{Y|\theta}\left\{\left(\frac{\partial log\ p\left(\mathbf{Y}|\theta,\mathbf{u}\right)}{\partial\theta}\right)^{T}\left(\frac{\partial log\ p\left(\mathbf{Y}|\theta,\mathbf{u}\right)}{\partial\theta}\right)\right\} \tag{5.5}$$

where $\mathbf{Y}$ represents the outputs, u the process inputs, $p\left(\cdot\right)$ the density function, and finally, θ denotes the parameter vector. The vector of φ design variables contains the inputs u as well as the switching time, the initial values of the states and the samples $\mathbf{Y}$.

The main reason to consider a scalar function of the information matrix is the well known Kramer-Rao inequality, which under some certain regularity conditions, states that:

A lower limit of the covariance of every unbiased efficient estimator of θ is asymptotically given by the inverse of the information matrix.
More precisely (Rao 1973):

$$V\left(\hat{\theta}\right)\geq FIM\left(\hat{\theta}\right)^{-1} \tag{5.6}$$

Depending on the level of prior knowledge available about the model parameters, two approaches are considered in defining the optimization criterion. In the first approach some information about the parameter distribution is considered in the criterion (Walter and Pronzato 1990). In the second approach the idea is to compute a design which maximizes the informational content for any value of the parameters.

The first approach takes into account the a-priori uncertainty in the model parameters θ. The parameters are assumed to belong to a population with a known distribution $p(\theta)$. Depending on the parameter estimation scheme, two criteria can be formulated. One for the maximum likelihood (ML) estimation scheme and another for the maximum a-posteriori (MAP) estimation scheme. The criterion for ML can be written as follows (Walter and Pronzato 1987) :

$$\varphi_{ED_{ML}} = arg\, min_{\varphi\in\Phi}\left[-E_\theta\left\{log\, detFIM\left(\theta,\varphi\right)\right\}\right] \tag{5.7}$$

where φ is the design variables vector, $E\{\cdot\}$ denotes the expected value, $FIM(\theta,\varphi)$ represents the Fisher information matrix, θ is within the Θ domain and φ is within the Φ domain. The equivalent criterion for MAP is written as follows (Melgaard and Madsen 1993):

$$\varphi_{ED_{MAP}} = arg\, min_{\varphi\in\Phi}\left[-E_\theta\left\{log\, detFIM\left(\theta,\varphi\right)/N + \Sigma_{pre}^{-1}\right\}\right] \tag{5.8}$$

The Σ_{pre}^{-1} matrix represents the prior covariance matrix of the parameters. Hence (ED)-optimal experiment is one for which the choice of experimental decision variables maximizes the expected value over the population of possible parameter values (Asprey and Macchietto 2002). The experiments designed in this way are good on average but can be poor for some values of the parameters (Walter and Pronzato 1987, Asprey and Macchietto 2000). With respect to the parameters space this approach is global .

When no prior information about the parameter distribution is available, it is possible to use only some nominal values of the parameters which in fact represent the prior mean values. In this case, the considered approach is local and the two criteria for ML and MAP mentioned above are further simplified as follows:

$$\varphi_{D\,ML} = arg\, min_{\varphi\in\Phi}\left[-log\, detFIM\left(\theta,\varphi\right)\right]_{\theta=E\{\theta\}} \tag{5.9}$$

$$\varphi_{D\,MAP} = arg\, min_{\varphi\in\Phi}\left[-log\, detFIM\left(\theta,\varphi\right)/N + \Sigma_{pre}^{-1}\right]_{\theta=E\{\theta\}} \tag{5.10}$$

In Eq. 5.10, N represents the number of samples.

A second approach aims to determine experimental designs φ_{WC}, that optimize the worst possible performance for any value of $\theta \in \Theta$. The criterion is given in Eq. 5.11.

$$\varphi_{WC} = arg\, min_{\varphi\in\Phi} max_{\theta\in\Theta}\left[-log\, detFIM\left(\theta,\varphi\right)\right] \tag{5.11}$$

Using this approach the only prior information about θ is the admissible domain Θ. In this way the design attempts to ensure acceptable performance for all possible values of θ. The max-min approach have had limited use due to the associated burdensome computations. In order to circumvent this problem

Asprey and Macchietto (2000), Asprey and Macchietto (2002) proposed a sequential algorithm derived from the worst case approach which they denote as robust R-optimal experimental design.

The scope of this chapter is limited to the simple local design where the selected optimization criterion is represented by the negative of the determinant of the FIM as given in equation 5.9. Furthermore, the data collected are intended for parameter estimation using a maximum likelihood (ML) estimation scheme. In performing experimental design for stochastic differential equations, one of the difficulties arises from the evaluation of the FIM.

Since it is nearly impossible to compute the FIM analytically for a system of stochastic differential equations, a method based on Monte Carlo simulations as proposed by Spall (2004, 2005) will be employed. The main idea is to generate a large number of Hessian matrices using Monte Carlo simulation. The efficient simultaneous perturbation stochastic approximation (SPSA) Spall (2004) algorithm which uses only four values of the log-likelihood function is used to obtain an estimate of the Hessian. Finally, averaging these Hessian matrices produces an estimate of FIM. The idea comes from the definition of FIM which is the expectation of the Hessian matrix of the log-likelihood function with respect to the model parameter vector θ. As stated by (Spall 2005) each Hessian estimate can be qualitatively very poor, even singular, but the interest is not in each Hessian estimate but the average obtained using many of these Hessian estimates. Thus, clearly, there is a difference from the previous approaches where only the deterministic case was considered in the evaluation of the FIM for quantitative experimental design.

5.2.3 Model structure of SDE

In the general case, a continuous-discrete stochastic state space model is a model that consists of a set of non-linear discretely, partially observed *Itô*, SDE's with measurement noise, (Jazwinski 1970) i.e.:

$$dx_t = f\left(x_t, u_t, t, \theta\right) dt + \sigma\left(u_t, t, \theta\right) d\omega_t \tag{5.12}$$

$$y_k = h\left(x_k, u_k, t_k, \theta\right) dt + e_k \tag{5.13}$$

In the above formulation $t \in \mathbf{R}$ is the time variable; $x_t \in X \subset \mathbf{R}^n$ is a vector of state variables; $u_t \in U \subset \mathbf{R^m}$ is a vector of input variables; $y_t \in Y \subset \mathbf{R}^l$ is a vector of output variables; $\theta \in \Theta \subset \mathbf{R}^p$ is a vector of possibly unknown parameters; $f\left(\cdot\right) \in \mathbf{R^n}$, $\sigma\left(\cdot\right) \in \mathbf{R^{n \times n}}$ and $h\left(\cdot\right) \in \mathbf{R^l}$ are known non-linear functions; $\{\omega_t\}$ is an n-dimensional standard Wiener process (Jazwinski 1970) and $\{e_k\}$ is an l-dimensional white noise process with $e_k \in N\left(0, S\left(u_k, t_k, \theta\right)\right)$. The diffusion term in Eq. 5.12 is assumed to be independent of the process states. The Σ matrix containing the diffusion terms σ is assumed to be diagonal.

5.2.4 Conditional density distribution function

Consider a set S of independent sequence of consecutive random measurements (Kristensen et al. 2004b):

$$\mathbf{Y} = \left[Y^1_{N_1}, Y^2_{N_2}, \ldots, Y^i_{N_i}, \ldots Y^S_{N_S} \right] \tag{5.14}$$

where each sequence is:

$$Y^i_{N_i} = \left[y^i_{N_i}, \ldots, y^i_k, \ldots, y^i_1, y^i_0 \right] \tag{5.15}$$

In the general case, if the parameters will be estimated using a ML estimation scheme and no prior information about the parameters density function is available, then the probability density distribution function is generated as follows:

$$p\left(\mathbf{Y}|\theta\right) = \left(\prod_{i=1}^{S} p\left(Y^i_{N_i}|\theta\right) \right) p\left(y^i_0|\theta\right) \tag{5.16}$$

or equivalently:

$$p\left(\mathbf{Y}|\theta\right) = \prod_{i=1}^{S} \left(\prod_{k=1}^{N_i} p\left(y^i_k|Y^i_{k-1},\theta\right) \right) p\left(y^i_0|\theta\right) \tag{5.17}$$

where the conditional probability rule $P(A \cap B) = P(A|B) \cdot P(B)$ is applied successively in order to form a product of conditional probability densities. In order to obtain an exact evaluation of the likelihood function in Eq. 5.17, a general nonlinear filtering problem has to be solved (Jazwinski 1970), but this is computationally infeasible in practice. However, since increments of the standard Wiener process $\{\omega_t\}$ driving the SDE in Eqs. 5.12 are Gaussian, it is reasonable to assume that the conditional probability densities in Eq. 5.17 can be well approximated by Gaussian densities which means that a method based on the extended Kalman filter (EKF) can be applied (Kristensen et al. 2004b). The Gaussian density is completely characterized by its mean and covariance, thus introducing the following notation:

$$\hat{y}^i_{k|k-1} = E\{y^i_k|Y^i_{k-1},\theta\} \tag{5.18}$$

$$R^i_{k|k-1} = V\{y^i_k|Y^i_{k-1},\theta\} \tag{5.19}$$

$$\varepsilon^i_k = y^i_k - \hat{y}^i_{k|k-1} \tag{5.20}$$

Where Eq. 5.20 is denoted as the *innovation* equation then by substituting Eqs. 5.18–5.20 into Eq. 5.17 the likelihood function becomes:

$$p\left(\mathbf{Y}|\theta\right)=\prod_{i=1}^{S}\left(\prod_{k=1}^{N_i}\frac{\exp\left(-\frac{1}{2}\left(\epsilon_k^i\right)^T\left(R_{k|k-1}^i\right)^{-1}\left(\epsilon_k^i\right)\right)}{\sqrt{det\left(R_{k|k-1}^i\right)}\left(\sqrt{2\pi}\right)^l}\right)p\left(y_0^i|\theta\right) \tag{5.21}$$

The *innovations* ϵ_k and their covariances $R_{k|k-1}$ are computed recursively by means of the EKF (Kristensen et al. 2004b). Given the initial probability density function $p\left(y_0|\theta\right)$ all subsequent conditional densities and hence the likelihood function can be determined by solving a continuous discrete nonlinear filtering problem. The work-flow of quantitative experimental design for stochastic differential equation models is presented in Figure 5.2 and the various steps are detailed in Section 5.3. Having introduced the optimization methodology, the stochastic model formulation and solution method the procedure for quantitative experimental design for SDE models is introduced next.

5.3 The algorithm

This section describes the developed algorithm for quantitative experimental design for stochastic differential equation models. In Figure 5.1, a work-flow diagram of the quantitative experimental design is depicted. The work-flow steps are detailed in this section.

- *Step 1* of the main algorithm simply generates a population of vector members as described in Section 5.2.1.
- In *Step 2* of the algorithm the objective function is computed and this step will be detailed in the sub-steps as detailed in Figure 5.2. The number of Hessians used in the evaluation of FIM is denoted N.
  - In *step 2.1* the unknown parameters of interest for experiment design are selected. In the case of D optimal design, the θ vector is composed of all unknown model parameters.
  - In *step 2.2* integration of the system of SDE is performed using a Kloeden-Platen explicit strong scheme of order 1 (Kloeden and Platen 1999) with a very small integration step. A set of pseudo experimental data $\bar{y}_k$ is obtained.

    Next steps, *2.3* - *2.7* concerns obtaining one of the Hessian matrix estimates used in the FIM calculation.
  - In *Step 2.3* two different vectors, Δ and Δ_T of random numbers of the same size as the θ vector are generated. These vectors are drawn from a uniform distribution. The more common distributions such as Gaussian distribution are excluded (Spall 2004). In principle these vectors are vectors giving the sign for the perturbations.

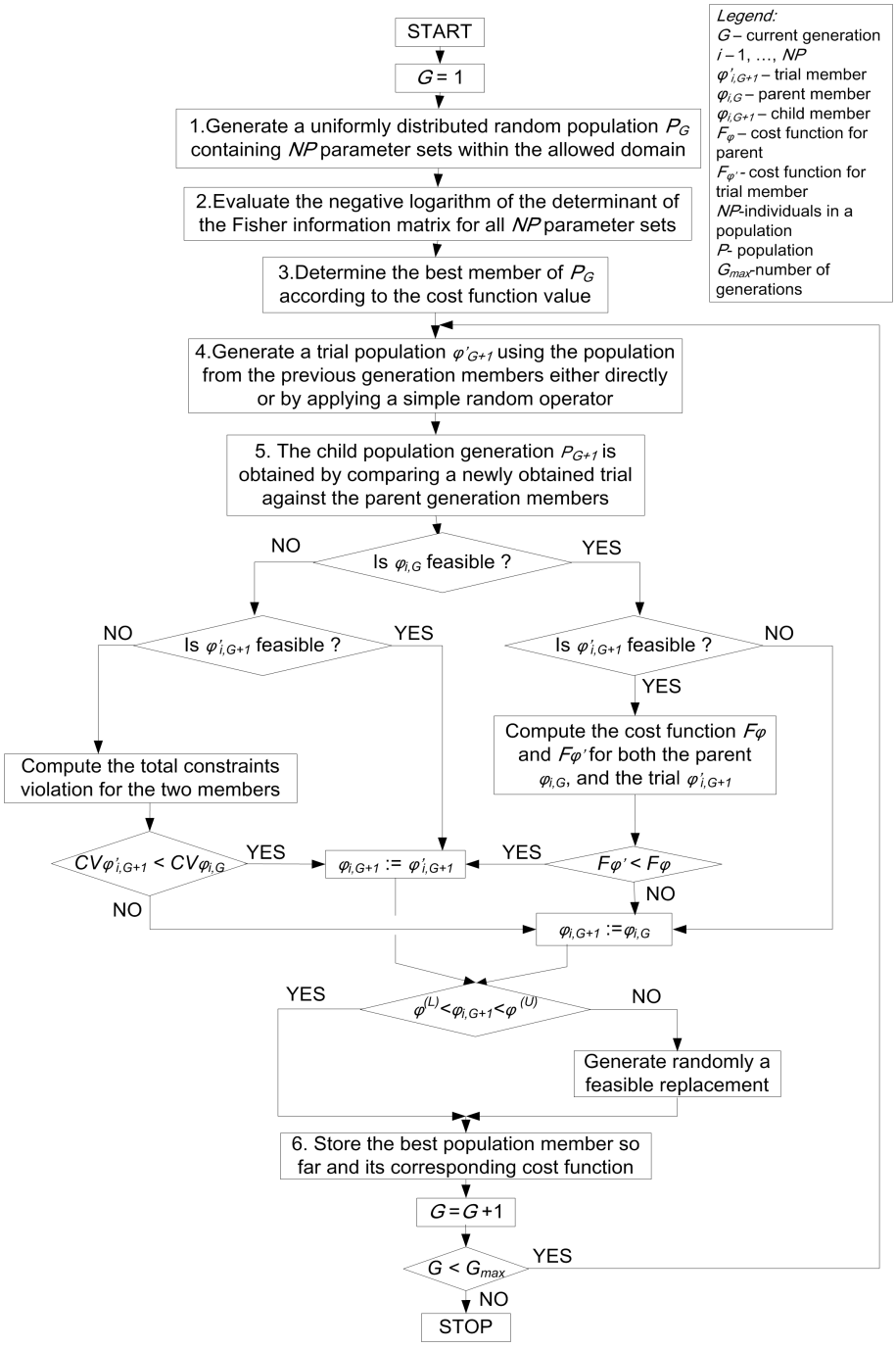


Figure 5.1: Proposed work-flow for the optimization algorithm used for quantitative experimental design for systems modelled by stochastic differential equations

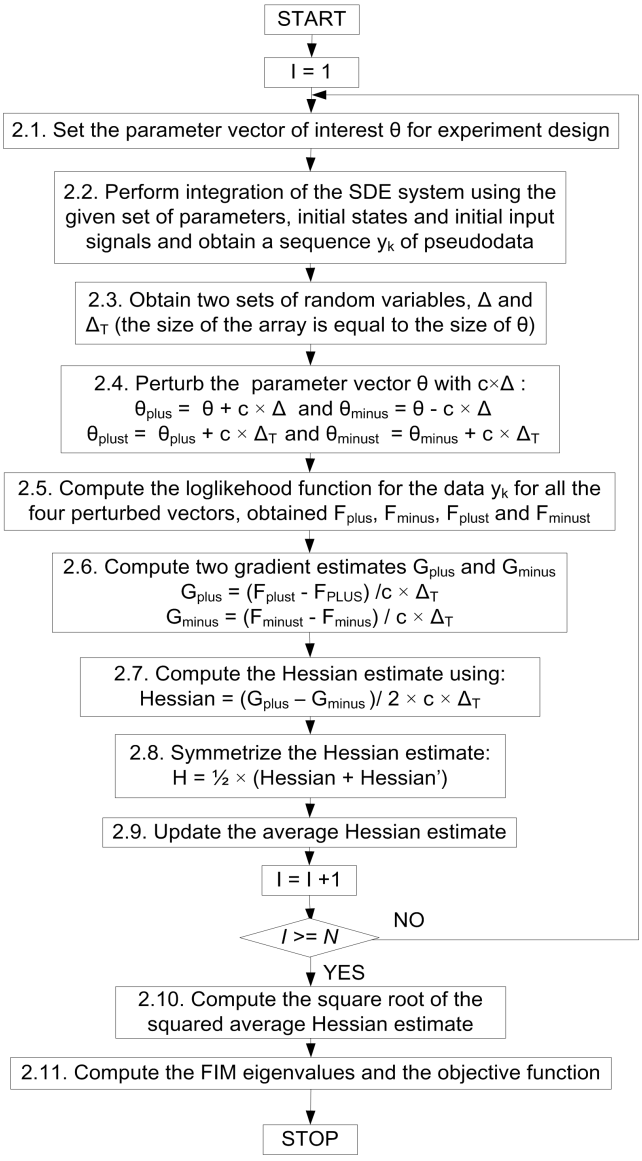


Figure 5.2: Proposed work-flow for evaluation of the experimental design objective function in *step 2* of quantitative experimental design (Fig. 5.1) for SDE based models

- In *Step 2.4* the parameter vector θ is perturbed with $\pm c \times \Delta$ and $\pm c \times \Delta_T$ respectively in order to obtain four new parameter vectors: θ_{plus}, θ_{minus}, θ_{plust} and θ_{minust}. In the original algorithm for FIM calculation by Spall (2004), Spall (2005) the c quantity is fixed to a small value i.e. $c = 0.001$ and similarly for all the parameters. In this contribution c was taken as 0.1 % of the nominal value of the parameters. As discussed by Spall (2004), c is a measure of accuracy of the estimate.
- In *Step 2.5* the log-likelihood function is computed for the sequence of pseudo experimental data y_k. The log-likelihood function is evaluated for all four parameter sets independently, and thus four values of the log-likelihood function: F_{plus}, F_{minus}, F_{plust}, F_{minust} are obtained. An extended Kalman Filter is used to recursively compute the log-likelihood function as presented in Section 5.2.4.
- *Step 2.6* concerns estimation of the gradient of the log-likelihood function for two points using the four values of the log-likelihood function using a central difference. Two values of the gradient G_{plus} and G_{minus} are obtained.
- *Step 2.7* evaluates the Hessian matrix using Eq. (3.1) from Spall (2004) and the two gradient values obtained in *Step 2.6*.
- *Step 2.8* simply makes the estimate symmetric by averaging the Hessian matrix with its transpose.
- In *step 2.9* the updated average Hessian matrix is computed. *Steps 2.1 - Steps 2.9* are repeated until the maximum number N of desired pseudo experimental data sets and the subsequent Hessian matrices are generated.
- *Step 2.10* concludes the FIM computation by computing the square root of the squared average Hessian matrix (Spall 2004). This operation is performed to assure a positive definite matrix.
- In *step 2.11* the FIM eigenvalues are computed using a singular value decomposition and then the optimization criterion can be computed immediately. In this case the determinant of the matrix is computed.

- Returning to the main algorithm (Fig. 5.1), in *Step 3*, the best member of the current generation is selected based on the cost function value.
- In *step 4* the trial randomly generated population, $\varphi'_{i,G+1}$ is obtained as discussed in Section 5.2.1, on page 5.2.1.
- In *step 5* using the obtained trial population and the selection criterion mentioned above in Section 5.2.1 the child generation $\varphi_{i,G+1}$ is obtained.

The optimization continues until convergence has been achieved or another stopping criterion is met e.g. the maximum number of generations, G_{max}.

5.4 Benchmark case

This section presents the application of the described algorithm for a model described by two coupled stochastic differential equations. The model represents a fed-batch bio-reactor also used by Asprey and Macchietto (2000). The system of ordinary differential equations is extended to a system of stochastic differential equations. The diffusion terms are considered to be states independent. The model equations are shown in Eqs. 5.22–5.23 including the stochastic terms.

$$\begin{aligned} dx_1 &= (r - u_1 - \theta_4) \cdot x_1 \cdot dt + \sigma_1 \cdot d\omega \\ dx_2 &= -\frac{r \cdot x_1}{\theta_3} + u_1 \cdot (u_2 - x_2) + \sigma_2 \cdot d\omega \end{aligned} \tag{5.22}$$

$$r = \frac{\theta_1 \cdot x_2}{\theta_2 + x_2} \tag{5.23}$$

In Eqs., 5.22-5.23, x_1 represents the biomass concentration $[g/l]$, x_2 represents the substrate concentration $[g/l]$, u_1 is the dilution rate $\left[h^{-1}\right]$ and u_2 is the feed substrate concentration $[g/l]$. The experimental conditions are:

1. the initial biomass concentration $x_{01} \in [1; 10]$ $[g/l]$
2. the dilution rate, $u_1 \in [0.05; 0.2]$ $\left[h^{-1}\right]$
3. the feed substrate concentration, $u_2 \in [5; 35]$ $[g/l]$

The mean initial concentration, x_{02}, is always 0.1 $[g/l]$ and cannot be manipulated for experimental design purposes. Both x_1 and x_2 can be measured during the experiments. The objective is to design an experiment that will provide the maximum amount of information for estimation of the four model parameters θ_1 *to* θ_4.

5.4.1 Results

The experimental design performed and reported here is equivalent to the first step of the robust optimal design proposed by Asprey and Macchietto (2002) . In the original paper the authors are considering also a known interval for each of the parameters. However the goal of this comparison is merely to test the algorithm and the developed program rather than a full reproduction of results and performance. The optimization algorithm used here does not start with a given initial parameter vector as a local search method but instead with a uniformly distributed random population as discussed above in section 5.3, *Step 2*.

The initial design conditions are given in Table 5.1.

In their paper Asprey and Macchietto (2002), do not mention any value of the measurement variance. The value of the experimental design criterion for

Variable	Values
x_1^0	5.5
t_{spl}, $l = 1, \ldots, n_{sp}$	2.0, 4.0, 6.0, 8.0, 10.0
	12.0, 14.0, 16.0, 18.0, 20.0
$t_{sw1,i}$, $i = 1, \ldots, n_{sw1}$	4.0, 8.0, 12.0, 16.0, 20.0
$t_{sw2,i}$, $i = 1, \ldots, n_{sw2}$	4.0, 8.0, 12.0, 16.0, 20.0
$u_{1,i}$, $i = 1, \ldots, n_{sw1}$	0.12, 0.12, 0.12, 0.12, 0.12
$u_{2,i}$, $i = 1, \ldots, n_{sw2}$	15.0, 15.0, 15.0, 15.0, 15.0

Table 5.1: Initial experimental conditions used by Asprey and Macchietto (2002)

the initial design (Table 5.1) reported by Asprey and Macchietto (2002) was $det(FIM) = 2.41E + 8$.

The initial design given in Table 5.1 was evaluated using the approach described in section 5.3, *steps 2.1-2.11.* All the diffusion terms, see Eq.5.12, were set to $\sigma(i, i) = 1.0E - 12$, and the variance of the measurements, see Eq. 5.13, to $S = 1.0E - 12$. The number of stochastic realization, was set to $N = 100$. The estimated FIM using the parameters described above was singular.

The selection of $N = 100$ independent realizations for evaluating the FIM was made after a series of trials where it was noticed that by increasing this number significantly, the variance in the FIM estimates remained nearly constant.

The simulation data using the initial design together with the parameters for diffusion terms and the measurement variances are plotted in Figures 5.3 - 5.4.

The diffusion terms and the measurement variances were set at the values mentioned above, $\Sigma = 1.0E - 12$, $S = 1.0E - 12$. The optimization algorithm described in section 5.2.1 with the following parameters $G = 50000$, $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$ has been run in order to determine the optimal design. In order to assess the performance of the algorithm the evolution of the cost function versus the number of iterations is plotted in Figure 5.5. It can be seen that the value of the objective function is constant for the last many generations. It can be concluded based on the pattern in Figure 5.5, that the results are very close to a minimum since an asymptotic evolution is noticed.

The resulting optimal design obtained after running the optimization is given in Table 5.2.

The design objective function criterion has been improved significantly to a value of $-log(det(FIM)) = -54.2352$ or $det(FIM) = 3.58E + 23$. The value reported by Asprey and Macchietto (2002) for the D-optimal design is $det(FIM) = 1.278E + 17$ and the ED-optimal design $det(FIM) = 5.346E + 1$.

The simulated profile of the obtained optimal experimental design as well as the values of the inputs and the switching time are plotted in Figures 5.6-5.7.

When comparing with the results obtained by Asprey and Macchietto (2002) for D optimal design both the minimum value, the optimal input profiles and subsequently the outputs are different. There are various reason for the differ-

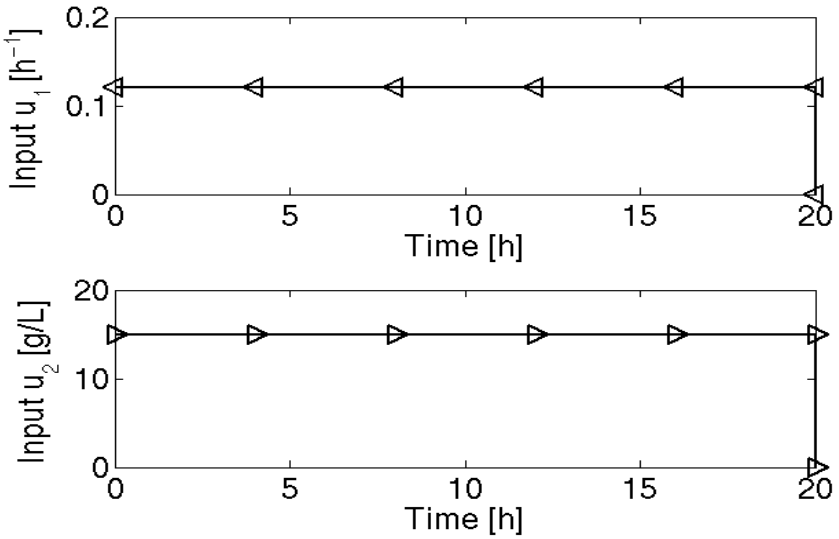


Figure 5.3: Experimental design - initial design

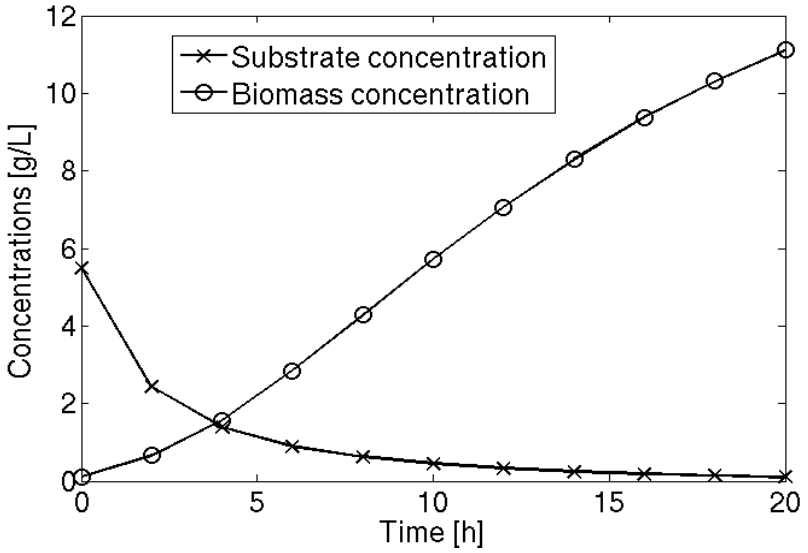


Figure 5.4: Initial experimental design mean values

ences and apparently one possible cause is the differences in the model formulation, i.e. the inclusion of the standard Wiener processes in the structure of

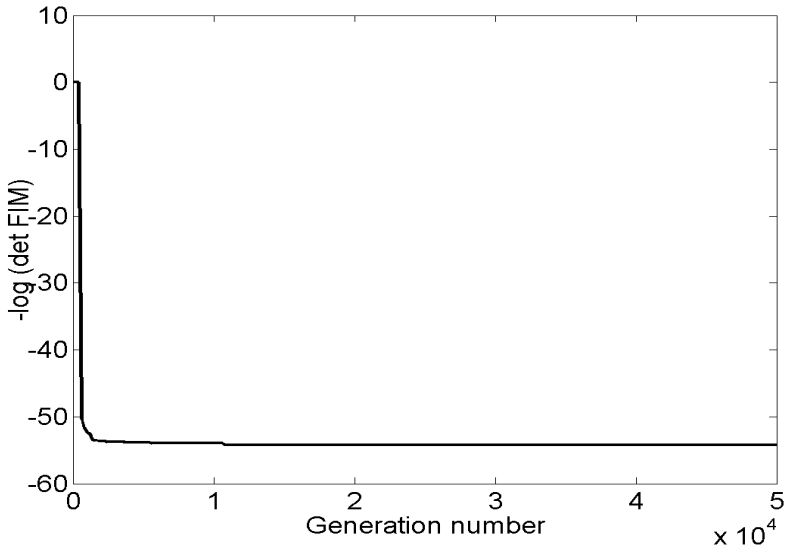


Figure 5.5: Progression of the convergence using $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$

Variable	Values
x_1^0	9.29
t_{sp_l}; $l = 1, \ldots, n_{sp}$	0.16, 9.66, 10.84, 13.68, 23.86, 25.89, 26.97, 35.1, 37.22, 39.98
$t_{sw1,i}$; $i = 1, \ldots, n_{sw1}$	3.072, 14.132, 20.164, 31.014, 36.238
$t_{sw2,i}$; $i = 1, \ldots, n_{sw2}$	0.151, 13.475, 19.91, 25.993, 31.432
$u_{1,i}$; $i = 1, \ldots, n_{sw1}$	0.104, 0.111, 0.126, 0.132, 0.138
$u_{2,i}$; $i = 1, \ldots, n_{sw2}$	7.23, 33.158, 24.853, 22.239, 18.646

Table 5.2: Experimental conditions obtained

the model equations; another possible causes are the algorithm used to compute the likelihood function, the optimization algorithm and also due to the minimum of the likelihood function used in the SD design.

However, when comparing the results obtained using the proposed SD design algorithm with the results obtained for ED design by Asprey and Macchietto (2002) the results seems to be more similar. The measurement times are grouped into three series. First sample is placed at the beginning of the experiment, then a short series composed of three samples is placed arount $t = 10[h]$ where the slope of the substrate concentration is large, then the second series is located toward the maximum of the substrate concentration and the last series of three samples is placed towards end of the experiment. The input profiles

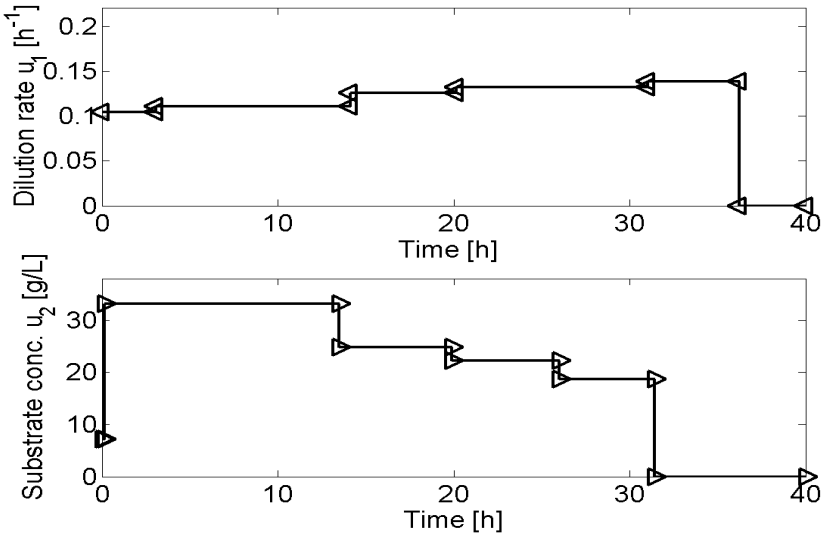


Figure 5.6: Experimental design - final input design using $G = 50000$, $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$

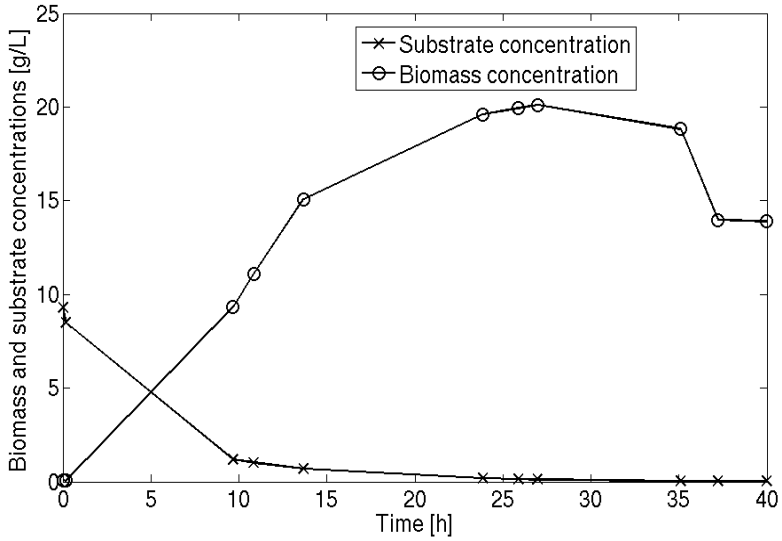


Figure 5.7: Experimental design - final output design using $G = 50000$, $NP = 31$, $N = 100$ and $F = 0.9$, $CR = 0.9$

seems to be different. The dillution rate seems to be nearly constant. In the optimization it was assumed that after the last switching time both inputs are zero.

The scheme used to evaluate the FIM is based on the SPSA algorithm which implies stochastic perturbation of the model parameters using an uniform distribution. Hence, this explains at least partly the similarity with the results obtained for the DAE version of the model using the expected value design criterion which accounts for model parameter uncertainty assuming a Gaussian distribution in Asprey and Macchietto (2002).

5.5 Discussion

In this contribution, a new method for model based quantitative experimental design for stochastic differential equation models is presented. The method is based upon minimizing the negative logarithm of the determinant of the Fisher Information matrix called SD design.

The core of the algorithm is represented by the efficient calculation of the Fisher information matrix using bootstrapping (Spall 2004) based on the simultaneous perturbation stochastic approximation (SPSA). This method is efficient because it is only using 4 values of the log-likelihood function.

The integration of the system of stochastic differential equations uses an iterated Kalman Filter approach. Within the integration routine a package for automatic differentiation (Kedem 1980) is used to evaluate the Jacobians of the system.

The optimization algorithm is based on a differential evolution strategy for global optimization with a small modification. The introduced modification concerns the discrimination between two infeasible individuals when selecting the new generation population.

The algorithm is applied to a benchmark model for a bio-reactor process. By applying experimental design it was possible to obtained pseudo experimental data with a much richer informational content for parameter estimation.

5.6 Conclusion

A new method for quantitative experimental design called SD-design has been developed. The method is based upon selecting the experimental design variables which maximizes the information content in the average FIM. The underlying models are stochastic differential equation models. Where the average FIM is evaluated based upon the probability density distribution function of the measurements. That includes prior information about parameter distributions. Thus this method bears similarity with the ED-design method of Asprey and Macchietto (2002). One of the advantages of the SD design is represented by the possibility to better evaluate the information content i.e. the Fisher Information Matrix instead of using just a sort of mean value in the form of

the Hessian matrix, thus distribution data are incorporated in a natural manner. The main drawback of this approach is represented by the additional computation burden.

6

An improved grey-box stochastic modelling framework

Abstract

Modelling complex systems such as enzymatic reaction networks is a challenge for current methodologies. This chapter is concerned with model development for an enzymatic reaction network under high uncertainty with respect to the reactions occurring and their kinetic mechanisms. An improved framework for development of mathematical models is presented and applied to an enzymatic reaction network. The improvements focus upon ensuring structural identifiability of the reaction kinetic parameters through qualitative experimental design followed by a quantitative experimental design step to exploit the remaining degrees of freedom. A stochastic differential equation model for an enzymatic reaction network is developed, analyzed and statistically validated. Various statistical tests are used to asses the model and parameters validity.

6.1 Introduction

Current methodologies for identification and model development for dynamic systems are still unable to handle both partial knowledge about the system and partial measurements systematically, despite the great efforts that have been made in this direction.

Among the attempts to systematize the methodology and the available tools it can be mentioned here one approach for general model development and identification by Kremling et al. (2004), Gadkar et al. (2005), and by Chen et al. (2004). However there is a main drawback of their approach, that is, the assumption that the model structure is correct such that all the error is absorbed in the measurement errors and in the model parameter estimates. Thus, the methodology neither pinpoints model deficiencies nor provides information on how to improve them.

A more systematic methodology for model improvement for nonlinear dy-

namic processes was developed recently by Kristensen et al. (2004b). By considering stochastic differential equations this methodology has built in features that accounts for model errors by the diffusion terms.

However several shortcomings where identified related to the quality of parameter identifiability. These observations led to subsequent methodological developments related to experimental design for ensuring qualitative parameter identifiability (chapters 3 & 4) and to quantitative experimental design for ensuring quantitative parameter identifiability through quantitative experimental design (chapter 5). The developed methods are used in this chapter to facilitate model development for reaction networks described by stochastic differential equations.

The purpose of the present chapter is to present a modified framework for development of grey-box stochastic models and to demonstrate its application on development of a few different stochastic differential equation models for an enzymatic reaction network for producing Di-Hydroxy-Acetone-Phosphate (DHAP) from Glucose. The modelling purpose is to improve understanding for optimizing production from a (fed-)batch production unit.

The remaining of the chapter is organized as follows: Section 6.2 introduces an improved version of the grey-box stochastic modelling framework. Section 6.3 presents an improved version of the SBT model. Section 6.4 introduces a method for qualitative experimental design which is applied on the model for an enzymatic reaction network presented in section 6.3. Section 6.6 presents a quantitative experimental design for parameter estimation performed on the model described in section 6.3. Finally, Section 6.7 closes this chapter with discussion and conclusions.

6.2 An improved framework for grey-box model development

Two improved grey-box modelling frameworks are proposed. The new frameworks include the developments presented in chapters 3–5.

The first improvement is represented by the situation where some experimental data and a first possible model are available. In this case the first step is to perform an identifiability analysis and determine which parameters can be estimated. Subsequently follows parameter estimation and residual analysis. If the model is still falsified the framework simply follows the original framework proposed by Kristensen et al. (2004b). Figure 6.1 illustrates this modification. Note that the experimental data used for estimation differs from the data used for residual analysis etc. as indicated by the hatched line.

The second improvement concerns only an initial version of the model is available. In this case the improved grey-box stochastic modelling framework includes two new steps. With a first model at hand a qualitative experimental design step is used to determine which variables to measure and which input to perturb to ensure parameter identifiability. The next step is to perform a

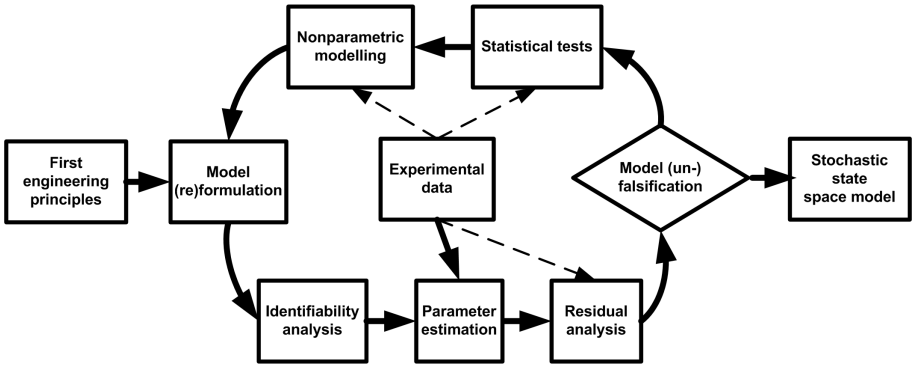


Figure 6.1: Grey-box stochastic modelling framework extended with identifiability analysis

quantitative experimental design in order to increase the information content of the experimental data required for parameter estimation. The remaining steps follow the idea of the original framework as proposed by Kristensen et al. (2004b). A sketch of this proposed algorithm is illustrated in Figure 6.2.

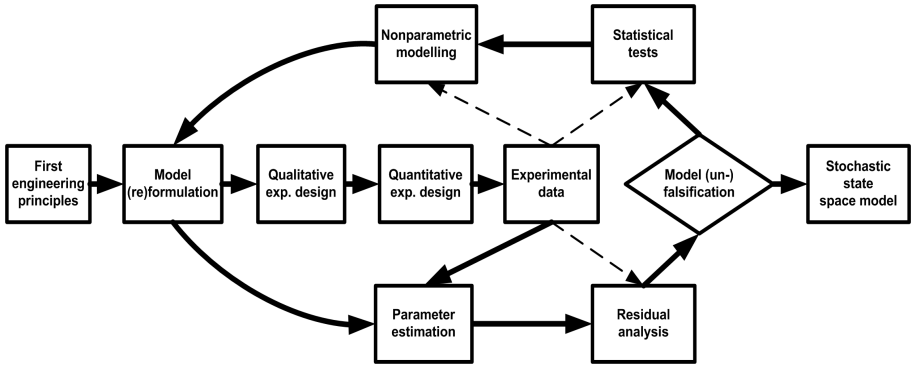


Figure 6.2: Grey-box stochastic modelling framework extended with qualitative and quantitative experimental design

The proposed framework will be used in this chapter to improve the SBT models proposed in chapter 2.

6.3 Model III for a simplified SBT

In this section a third model for an enzymatic reaction network is presented. The sketch of the new reaction network considered is given in Figure 6.3.

This version is motivated by the modifications occurred in the experimental

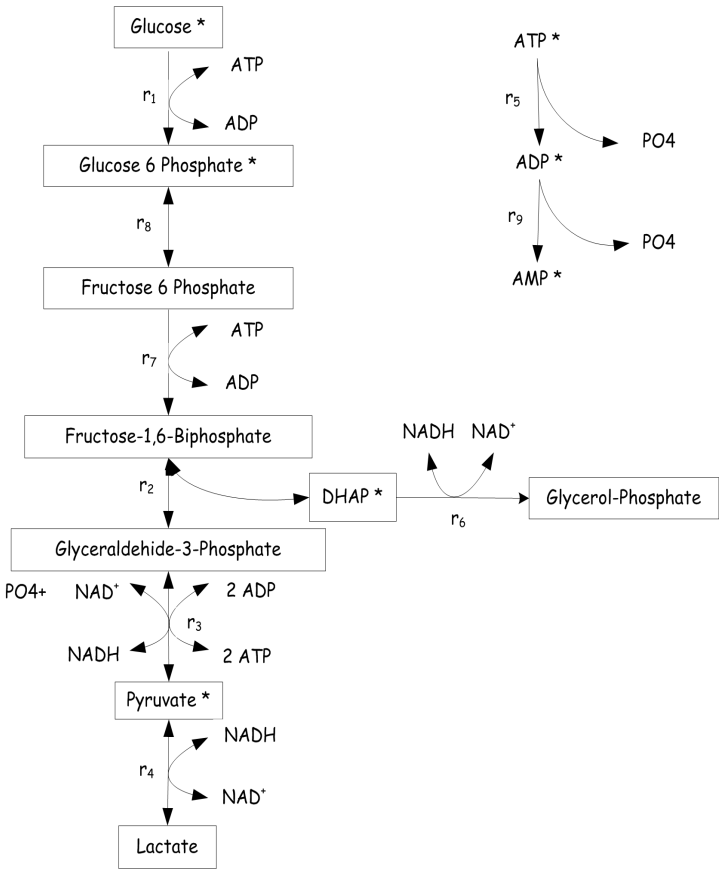


Figure 6.3: The simplified reaction network corresponding to Model III

settings. The genes responsible for the expression of enzymes catalyzing the reaction r_9, r_{11} and r_{12} were deleted, partly as a consequence of our modelling attempts. The new model takes these simplifications into account. The reaction r_{10} is eliminated since the gene for this enzyme is not expressed as well. The reaction r_{13} becomes r_9 in the new reaction network. The main reaction consuming DHAP in the SBT has been identified as being the conversion to Glycerol-Phosphate catalyzed by NADH. The model equations are given in Eqs. 6.1–6.3.

$$
\begin{aligned}
r_1 &= \frac{(r_{1maxf}\frac{c_{GL}c_{ATP}}{K_{GL}K_{ATP}} - r_{1maxb}\frac{c_{G6P}c_{ADP}}{K_{GL}K_{ATP}K_{eqHK}})}{(1+\frac{c_{GL}}{K_{GL}}+\frac{c_{G6P}}{K_{G6P}})(1+\frac{c_{ATP}}{K_{ATP}}+\frac{c_{ADP}}{K_{ADP}})} \\
r_2 &= \frac{r_{2maxf}\frac{c_{F16B}}{K_{F16B}} - r_{2maxb}\frac{c_{DHAP}c_{G3P}}{K_{F16B}K_{eqALD}}}{1+\frac{c_{F16B}}{K_{F16B}}+\frac{c_{DHAP}}{K_{DHAP}}+\frac{c_{G3P}}{K_{G3P}}+\frac{c_{F16B}c_{G3P}}{K_{F16B}K_{iG3P}}+\frac{c_{DHAP}c_{G3P}}{K_{DHAP}K_{G3P}}} \\
r_3 &= r_{3maxf}c_{G3P}c_{NAD}c_{PO4}c_{ADP} \\
r_4 &= \frac{r_{4maxf}\frac{c_{PYR}c_{NADH}}{K_{PYR}K_{NADH}} - r_{4maxb}\frac{c_{LAC}c_{NAD}}{K_{PYR}K_{NADH}K_{eqLDH}}}{(1+\frac{c_{NAD}}{K_{NAD}}+\frac{c_{NADH}}{K_{NADH}})(1+\frac{c_{LAC}}{K_{LAC}}+\frac{c_{PYR}}{K_{PYR}})} \\
r_5 &= r_{5maxf}\frac{c_{ATP}}{c_{ATP}+K_{ATP}} \\
r_6 &= r_{6maxf}c_{DHAP}c_{NADH} \\
r_7 &= \frac{r_{7maxf}c_{F6P}c_{ATP}}{((c_{ATP}+K_{ATP}(1+\frac{c_{ADP}}{K_{ADP}}))(c_{F6P}+K_{F6P}c_{AMP}))} \\
r_8 &= \frac{r_{8maxf}\frac{c_{G6P}}{K_{G6P}} - r_{8maxb}\frac{c_{F6P}}{K_{G6P}K_{eqPGI}}}{1+\frac{c_{G6P}}{K_{G6P}\left(1+\frac{c_{G6P}}{K_{F6PG6P}}\right)}+\frac{c_{F6P}}{K_{F6P}}} \\
r_9 &= r_{9maxf}\frac{c_{ADP}}{c_{ADP}+K_{ADP}}
\end{aligned}
\tag{6.1}
$$

$$
\begin{aligned}
dc_{GL} &= \left((-r_1) + \frac{F}{V}\left(c_{GL_{feed}} - c_{GL}\right)\right)dt + \sigma_{11}d\omega \\
dc_{F16B} &= \left((r_7 - r_2) - \frac{F}{V}c_{F16B}\right)dt + \sigma_{22}d\omega \\
dc_{DHAP} &= \left((r_2 - r_6) - \frac{F}{V}c_{DHAP}\right)dt + \sigma_{33}d\omega \\
dc_{G3P} &= \left((r_2 - r_3) - \frac{F}{V}c_{G3P}\right)dt + \sigma_{44}d\omega \\
dc_{PYR} &= \left((r_3 - r_4) - \frac{F}{V}c_{PYR}\right)dt + \sigma_{55}d\omega \\
dc_{LAC} &= \left((r_4) - \frac{F}{V}c_{LAC}\right)dt + \sigma_{66}d\omega \\
dc_{ATP} &= \left((-r_1 - r_7 + 2r_3 - r_5) - \frac{F}{V}c_{ATP}\right)dt + \sigma_{77}d\omega \\
dc_{NAD} &= \left((-r_3 + r_4 + r_6) - \frac{F}{V}c_{NAD}\right)dt + \sigma_{88}d\omega \\
dc_{PO4} &= \left((-r_3 + r_5 + r_9) - \frac{F}{V}c_{PO4}\right)dt + \sigma_{99}d\omega \\
dc_{G6P} &= \left((r_1 - r_8) - \frac{F}{V}c_{G6P}\right)dt + \sigma_{1010}d\omega \\
dc_{ADP} &= \left((r_1 + r_7 - 2r_3 + r_5 - r_9) - \frac{F}{V}c_{ADP}\right)dt + \sigma_{1111}d\omega \\
dc_{NADH} &= \left((r_3 - r_4 - r_6) - \frac{F}{V}c_{NADH}\right)dt + \sigma_{1212}d\omega \\
dc_{F6P} &= \left((r_8 - r_7) - \frac{F}{V}c_{F6P}\right)dt + \sigma_{1313}d\omega \\
dc_{AMP} &= \left((r_9) - \frac{F}{V}c_{AMP}\right)dt + \sigma_{1414}d\omega \\
dc_{GP} &= \left((r_6) - \frac{F}{V}c_{GP}\right)dt + \sigma_{1515}d\omega \qquad (6.2) \\
dV &= Fdt + \sigma_{1616}d\omega \qquad (6.3)
\end{aligned}
$$

6.4 Decoupled qualitative experimental design

Using the ideas and the results from the qualitative experimental design analysis from chapter 4 here a different approach for qualitative experimental design is pursued. The main motivation for this different approach is that for more complex models such as Model III, is much more difficult to apply the method-

ology described in chapter 4.

A decoupled qualitative experimental design is introduced and illustrated on Model III. In this version of the methodology the target is to determine which states have to be measured in order to render all the network reaction rates identifiable. Moreover, process inputs are considered from the beginning of the analysis. The second phase, where the focus is on assessing the structural parameter identifiability the problem is subdivided and each measured state is paired with a reaction rate. In the Tables 6.3–6.11 a collection of potential sets of identifiable parameters is obtained and the final decision belongs to the model developer in selecting a parameter set to be quantitatively identified.

The core of the first approach presented in chapter 4 is represented by the analytical solution of a system of algebraic equations. When formulating this system of algebraic equations, Lie derivatives for all the measured states are considered simultaneously, thus nonlinear interactions between the model states are embedded in the solution. Therefore the obtained analytical solution represents a global test for identifiability. The approach used in this section, when formulating the system of algebraic equations considers Lie derivatives obtained from one measured state at a time. The obtained analytical solution is thus influenced by just one state. After obtaining the analytical solutions for all the measured states a set of identifiable parameters is obtained. This set of parameters is not obtained as a solution of a unique system of algebraic equations, thus the solution does not represent a global test for identifiability.

The first phase of the analysis used to identify the sets of measured states that will render all reaction rates identifiable is identical to the one presented in Section 4.3. Regarding the second phase of the algorithm which consists of determining the identifiable parameters set for each of the identifiable reaction rates a different approach is considered. The idea is to divide the original problem into smaller subproblems. For each identifiable reaction rate, a measured state will be assigned. The Lie derivatives will be considered for each of the measured states at a time. When solving for the parameters, the parameter sets to be considered will be formed just with parameters related to the considered kinetic equation. The procedure is applied for all measured states and considered inputs. As a results, a map of the identifiable parameters will be obtained showing the identifiable parameter sets. In order to illustrate this approach a more detailed enzymatic reaction model will be used.

6.4.1 Identifiable rates

The reaction identifiability criterion Brendel et al. (2006) was applied to identify all the possible sets of measured states rendering all the reaction rates identifiable. A MATLAB script was used to automate the search of identifiable sets of identifiable reaction rates by screening all the possible combinations. A total of 1028 combinations were found to render the 9 reaction rates identifiable. Table 6.1 contains, the combinations containing the originally four measured species only, are listed. Measurements of NAD, NADH and GP have

not been considered for practical reasons.

No.	Measurement sets								
1	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{DHAP}	c_{G3P}	c_{PYR}	c_{AMP}	c_{F6P}
2	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{F16B}	c_{DHAP}	c_{PYR}	c_{LAC}	c_{F6P}
3	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{DHAP}	c_{G3P}	c_{PO4}	c_{PYR}	c_{LAC}
4	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{F16B}	c_{DHAP}	c_{PO4}	c_{PYR}	c_{LAC}
5	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{F16B}	c_{DHAP}	c_{G3P}	c_{PYR}	c_{LAC}
6	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{DHAP}	c_{G3P}	c_{PO4}	c_{PYR}	c_{F6P}
7	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{F16B}	c_{DHAP}	c_{PO4}	c_{PYR}	c_{F6P}
8	c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{F16B}	c_{DHAP}	c_{G3P}	c_{PYR}	c_{F6P}

Table 6.1: Combinations of measured species rendering all the nine reaction rates identifiable

6.4.2 Identifiable parameters

The analysis is illustrated with the first set of measured states from Table 6.1. A reaction rate is associated to each of the measured states. The matching is based on the reaction network at hand, in the sense that the rate influencing the measured rate was selected for pairing.

For each corresponding state equation the first three order Lie derivatives are computed. Thus, a system of three algebraic equations can be obtained for each corresponding state equation. A maximum of 3 parameters, can be selected simultaneously to be solved for each of the measured states. Table 6.2 contains the selected pairing between the reaction rates and the measured species for the first measurement set of the Table 6.1.

Measured sets								
c_{GL}	c_{ATP}	c_{G6P}	c_{ADP}	c_{DHAP}	c_{G3P}	c_{PYR}	c_{AMP}	c_{F6P}
r_1	r_5	r_8	r_3	r_6	r_2	r_4	r_9	r_7

Table 6.2: Correspondence between the reaction rates and the measured species

After generating and solving the systems of algebraic equations certain subsets of parameters could be rendered identifiable. For each pair of identifiable reaction rate and measured compound, the identifiable parameters are given in Tables 6.3 – 6.11.

r_1	r_{1maxf}	r_{1maxb}	K_{ADP}	K_{G6P}	K_{eqHK}	K_{ATP}	K_{GL}
GL	no solution:						
	4,6,9	3,7,8	1,2,3,4	1,5,6,7	2,5,8,9		
	unique solution:						
	1	1					
	multiple solutions:						

Table 6.3: Analysis for subsets of two parameters taken simultaneously

Various combinations of (un)/identifiable sets of parameters have been obtained. In Tables 6.3–6.11 each parameter set has been labeled using a number.

Table 6.3 contains the results of the analysis for the pairing GL as measured state and r_1 as identifiable reaction rate. After solving the systems of algebraic equations, three situations can occur: no solutions, a unique solution, and multiple solutions. For instance in the row labeled "unique solution" there is a label denoted 1, below r_{1maxf} and r_{1maxb}, which means that the set of two parameters rendered a unique solution while the labels 4,6,9 below the label "no solution" means that for the corresponding subsets no solution was found.

r_2	r_{2maxf}	r_{2maxb}	K_{F16B}	K_{eqALDO}	K_{DHAP}	K_{G3P}	K_{iG3P}
$G3P$	no solution:						
	unique solution:						
	1,2,4,7	1,3,5,8	7,8,9			4,5,6	2,3,6,9
	11	12	10,14		11,12,13,14	10	13
	multiple solutions:						
					1	1	

Table 6.4: Analysis for subsets of two parameters

r_3	r_{3maxf}
ADP	no solution:
	unique solution:
	1
	multiple solutions:

Table 6.5: Analysis for subsets of one parameter

r_4	r_{4maxf}	r_{4maxb}	K_{eqLDH}	K_{PYR}	K_{NADH}	K_{NAD}	K_{LAC}
PYR	no solution:						
	1,3	1,2	2,3				
	unique solution:						
	4,8	3,7	2,6,9	1	5,9	5,6,7,8	1,2,3,4
	11,14	10,13		12,13,14	10,11,12		
	multiple solutions:						
				3,4	2,4	1,3	1,2

Table 6.6: Analysis for subsets of two parameters taken simulateously

r_5	r_{5maxf} K_{ATP}
ATP	no solution:
	unique solution:
	1
	multiple solutions:
	1

Table 6.7: Analysis for subsets of one parameter

From the collection of potentially identifiable parameter vectors from the decoupled qualitative experimental design analysis one set of potentially identifiable parameters for estimation is selected and given in Table 6.12.

r_6	$r_6 maxf$
$DHAP$	no solution: unique solution: 1 multiple solutions: 1

Table 6.8: Analysis for subsets of one parameter

r_7	r_{7maxf}	K_{ADP}	K_{ATP}	K_{F6P}
$F6P$	no solution: unique solution: 1,2	1,3	2,3	
	multiple solutions:			

Table 6.9: Analysis for subsets of two parameters

r_8	r_{8maxf}	r_{8maxb}	K_{F6PG6P}	K_{F6P}	K_{G6P}	K_{eqPGI}
$G6P$	no solution: 2,3	1,3		1,2		
	unique solution: 3	2		1,2,3		1
	multiple solutions: 3	2	1,2,3,4	1		4

Table 6.10: Analysis for subsets of two parameters taken simultaneously

r_9	r_{9maxf}	K_{ADP}
AMP	no solution:	1
	unique solution: 1	
	multiple solutions:	

Table 6.11: Analysis for subsets of one parameter

Parameter	Parameter	Parameter
r_{1maxf}	r_{3maxf}	K_{F6P}
r_{2maxf}	r_{4maxf}	r_{8maxf}
K_{F16B}	K_{PYR}	r_{9maxf}
K_{G3P}	r_{5maxf}	

Table 6.12: One parameter vector set considered qualitatively identifiable from the measurement set GL, G6P, DHAP, ATP, ADP, AMP, PYR

Table 6.12 contains in fact only a very limited subset of the full parameter set. In order to render more globally identifiable parameters, as the qualitative experimental design pointed out it is necessary to measure more compounds and to perturb the Glucose input. These two points were used to qualitatively design a new experiment. Therefore in Experiment 30A a Glucose input perturbation was introduced and three more compounds i.e. ADP, AMP, PYR

have been measured.

6.5 Parameter estimation using data from Exp 30 A

Data collected from Experiment 30A, representing two similar experiments (two identical reactors I and II) are used for estimation and validation. Since the experimental conditions have been identical, the data represents two different stochastic realizations of the process. Each realization has been measured twice, thus four sets are available from Experiment 30A reactor I, set 1 and 2 and reactor II, set 1 and 2. These data includes a pulse perturbation of the Glucose feed, at time 3 hours as a consequence of the results of the qualitative experimental design.

The raw data are reproduced in Section 6.6.1 in Figure 6.6 and in the Appendix B. The data from the first reactor (reactor I) are used for estimating the parameter vector given in Table 6.12. The value of the Glucose feed pulses is F = 1E-4 $[L/min]$ with a duration of 1 minute and with a constant feed substrate concentration $c_{GL\,feed} = 400\ [mM/L]$. As mentioned in Table 6.13, the initial volume of the reactor is 1E-2 $[L]$.

The fixed model parameters together with the initial concentration for each of the compounds and their values are given in Table 6.13. For the saturation constants, an average of the numerical values found in the literature have been used. For the difussion terms of the mass balances corresponding to unmeasured compounds, small values have been considered. The assumption was that the equations for these compounds are correct.

Name	Parameter	Name	Parameter	Name	Parameter
c_{GL0}	1.0E-5	V_0	1.0E-2	K_{LAC}	1.291
c_{F16B0}	1.0E-5	K_{GL}	0.08	r_{6maxf}	16.64
c_{DHAP0}	1.0E-5	K_{ATP}	0726	r_{7maxf}	3.9651
c_{G3P0}	1.0E-5	r_{1maxb}	0.04	r_{8maxb}	1
c_{PYR0}	1.0E-5	K_{eqHK}	7177	Keq_{PGI}	0.312
c_{LAC0}	1.0E-5	K_{G6P}	0.64	K_{F6PG6P}	0.2
c_{ATP0}	11.5	K_{ADP}	0.479	σ_{22}	1.0E-2
c_{NAD0}	5.75	r_{2maxb}	0.1368	σ_{44}	1.0E-2
c_{PO40}	11.1	K_{eqALD}	0.1439	σ_{66}	1.0E-2
c_{G6P0}	1.0E-5	K_{DHAP}	0.96	σ_{88}	1.0E-2
c_{ADP0}	1.0E-5	K_{iG3P}	0.325	σ_{99}	1.0E-2
c_{NADH0}	1.0E-5	K_{NADH}	0.2535	σ_{1212}	1.0E-2
c_{F6P0}	1.0E-5	r_{4maxb}	0.012	σ_{1313}	1.0E-2
c_{AMP0}	1.0E-5	K_{eqLDH}	10.0	σ_{1515}	1.0E-2
c_{GP0}	1.0E-5	K_{NAD}	2.507	σ_{1616}	1.0E-2

Table 6.13: Fixed parameters value for Model III

Name	Estimate	Name	Estimate	Name	Estimate
r_{1maxf}	8.24098E-02	r_{8maxf}	1.33909E+00	S_{11}	8.56135E-02
r_{2maxf}	1.54521E+00	r_{9maxf}	1.91085E-02	S_{22}	2.42433E-04
K_{F16B}	1.89134E-01	σ_{11}	2.08442E-05	S_{33}	3.93489E-05
K_{G3P}	6.51939E-01	σ_{33}	6.27604E-02	S_{44}	2.91626E-02
r_{3maxf}	1.13599E-04	σ_{55}	7.17167E-02	S_{55}	4.81779E-05
r_{4maxf}	1.13402E-01	σ_{77}	4.82363E-01	S_{66}	9.18731E-05
K_{PYR}	1.63708E-02	σ_{1010}	1.31088E-05	S_{77}	1.67310E-03
r_{5maxf}	4.03031E-02	σ_{1111}	1.66903E-01		
K_{F6P}	1.23720E-01	σ_{1414}	8.85992E-02		

Table 6.14: Estimation results for Model III, using data from Experiment 30A, reactor I set 1 and 2

The estimation has been performed using the differential evolution algorithm described in chapter 5, with the following parameters $NP = 5$, $CR = 0.5$ and $F = 0.8$. The computer program CTSM has been modified accordingly to use this optimizer. The performance of the model has been assessed by using a validation dataset. The validation data set is represented by the data collected during Experiment 30A, reactor II, set 1. The estimation results are given in Table 6.14.

The plots for pure simulation (blue line) and one step ahead prediction (red line) are given below in Figure 6.4.

Since the optimization algorithm does not have a convergence criterion, the evolution of the cost function along the number of generations is plotted in figure 6.5.

Analyzing the plots in Figure 6.4, the quality of one step ahead prediction is fairly good, while the pure simulation does not seem to perform well for these experiments. The most significant difference between the pure simulation and one step ahead prediction in the case of AMP and PYR. The experimental data series used for validation for Pyruvate, and plotted in Figure 6.4.g contains few missing samples before the pulse, which were set to zero. However the one step-ahead prediction follows the trend observed for the Pyruvate data for reactor II, data used for estimation, see Figure 6.6.g.

Significant difference between pure simulation and one step ahead prediction occurs also for, ATP and ADP. The diffusion terms for all the four corresponding drift equations are fairly large indicating potential model deficiencies. The poor fit can be related to the values of the fixed parameters as well. However to improve the information content of the data it is desirable to perform a quantitative experimental design.

A very detailed model (**Model IV**) has been developed containing all the known enzymatic reactions. The model equations as well as the reaction network are given in Appendix A.

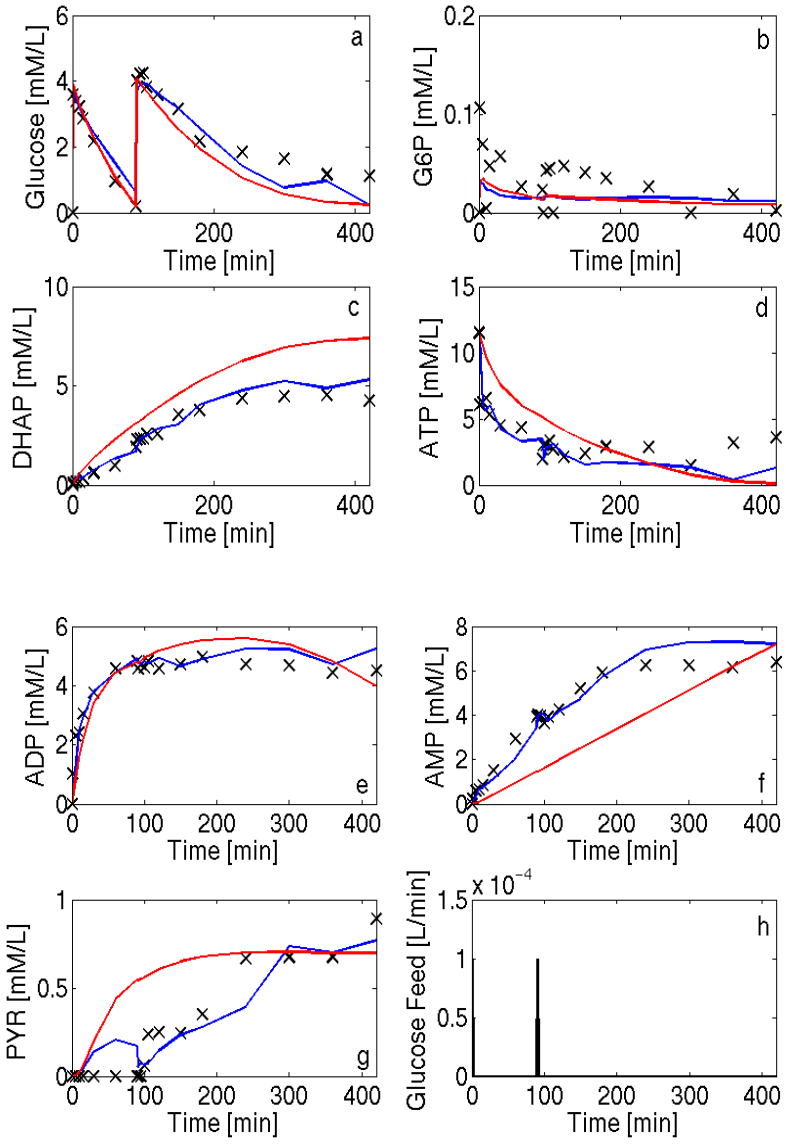


Figure 6.4: Simulation of Model III against data from Experiment 30A, reactor II, set 1; pure simulation-red, one step prediction-blue. Data from Experiment 30A, reactor I, set 1 and 2 used for estimation

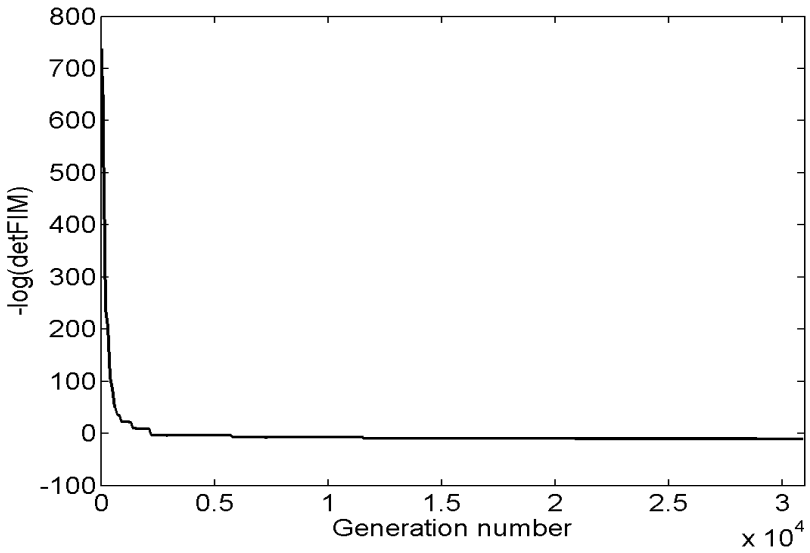


Figure 6.5: Progression of the convergence using $NP = 5$, $N = 100$ and $F = 0.79$, $CR = 0.8$

6.6 Quantitative experimental design for Model III of SBT

This section presents the application of the quantitative experimental design method presented in chapter 5 and tested for the simple bio-reactor model, for Model III of the simplified SBT as developed and presented in Section 6.3. The starting point for this investigation is the available experimental data collected prior to experimental design. Firstly, the amount of information available for parameter estimation present in the given data sets will be evaluated by computing the experimental design criterion i.e. the negative of the determinant of the FIM. Secondly, a quantitative experimental design will be performed under similar experimental conditions, with respect to the variance of the measurements.

6.6.1 Evaluating the initial experimental data

The raw experimental data sets already available from the laboratory are plotted in Figure 6.6. These are the data from Experiment 30A, reactor I and II, sets 1 ans 2 for each of the reactor. Two of the four data series in Figure 6.6.g for the Pyruvate contains few missing data samples before the pulse, which were replace with zero values.

Table 6.15 contains the experimental conditions used in Experiment 30. Initial conditions of the measured compounds, sampling times, switiching time as well as the input value for the Glucose feed are given.

Variable	Values
Initial conc $[mM/L]$	1.0E-05, 1.0E-05, 1.0E-05, 11.5, 1.0E-05, 1.0E-05, 1.0E-05
Sampling times	1.0, 5.0, 10.0, 15.0, 30.0, 60.0
$[min]$	90.0, 91.0, 95.0, 100.0, 105.0, 120.0
	150.0, 180.0, 240.0, 300.0, 360.0, 420.0
Switching time $[min]$	0.0, 1.0, 90.0, 91.0
Glucose feed-rate $[L/min]$	1.0E-04, 0.0, 1.0E-04, 0.0

Table 6.15: Experimental conditions used in Experiment 30A

The standard deviation has been computed and the values will be used to evaluate the experimental design criterion. Table 6.16 gives the experimentally determined variances of the measurements for this data sets.

The input profile consists of pulses of Glucose modelled as four constant intervals. Two of them are proportional with the value of the pulses, while the other two are zero. The value of the Glucose feed pulses is F = 1E-4 $[L/min]$ with a duration of 1 minute and a constant feed substrate concentration $c_{GL\,feed}$ = 400 $[mM/L]$. Using this information and the experimental conditions, i.e. initial values of the states, sampling times, now the information content will be evaluated. The value of the objective function evaluated under the conditions listed above, was $-ln(det(FIM))$ = -49.2898 or $det(FIM)$ = 2.54856E+21.

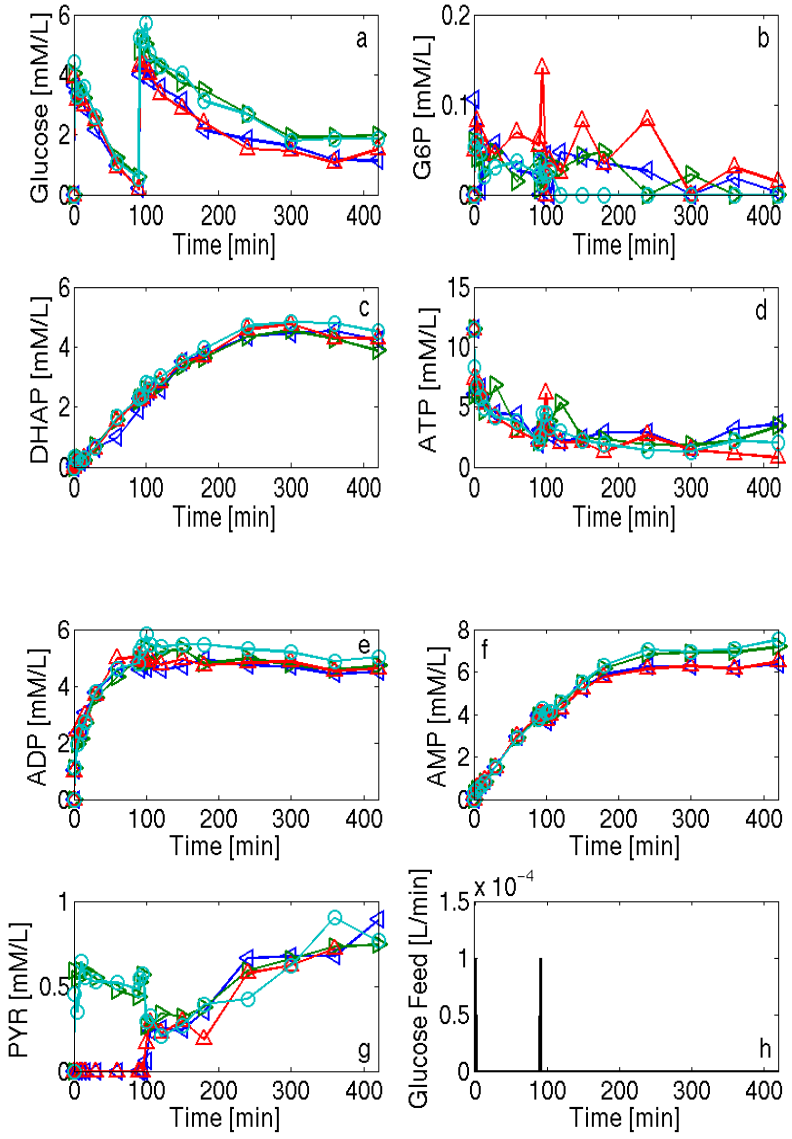


Figure 6.6: Data collected in Experiment 30A, set 1 and 2, reactor I and II

6.6.2 Optimal experimental design

The differential evolution algorithm presented in 5.2.1 was run with a large number of generations NG with $NP = 20$, $F = 0.9$, and $CR = 0.9$. However, due to time limitation the algorithm has been stopped after 1900 generations

Measured compound $[mM/L]$	Mean variance
c_{GL}	1.6890E-01
c_{G6P}	6.2325E-04
c_{DHAP}	2.3457E-02
c_{ATP}	6.6210E-01
c_{ADP}	7.2002E-02
c_{AMP}	5.9781E-02
c_{PYR}	4.8315E-02

Table 6.16: Computed mean variance for data collected in Experiment 30A, set 1 and 2, reactor I and II

and the results obtained are given as plots. The minimum distance between sampling time was set to $\Delta t_{spl}^{min} = 1$ minute while the maximum distance was set to $\Delta t_{spl}^{max} = 402$ minutes. The minimum distance between switching time was set to $\Delta t_{sw}^{min} = 5$ minutes and the maximum distance $\Delta t_{sw}^{max} = 402$ minutes. The experiment duration was assumed to be similar to the one of the Experiment 30, i.e. 420 minutes. For the Glucose feed flow-rate a maximum value of 1.1E-03 $[L/min]$, higher than the value of the pulses perfomed in Experiment 30. The feed Glucose concentration remained constant as before at $c_{GL\,feed} = 400\ [mM/L]$. The value of the objective function value reached to $-ln(det(FIM))$=-9.66E+01 or $det(FIM)$=8.62E+41. The simulated output profiles and the sampling points for the measured compounds using the optimal experimental settings are plotted in Figure 6.7.

Table 6.17 presents the experimental conditions obtained after 1900 generations.

Variable	Values
Initial conc $[mM/L]$	1.27E+01, 1.43E+01, 6.2E+00, 5.11E-02
	1.38E+01, 2.01E+00, 5.24E+00
Sampling times	9.36E-02, 6.63E+00, 3.26E+01, 6.13E+01, 9.78E+01, 1.37E+02
$[min]$	2.60E+02, 2.72E+02, 2.9E+02, 3.06E+02, 3.21E+02, 3.8E+02
	3.99E+02, 4.03E+02, 4.08E+02, 4.17E+02, 4.18E+02, 4.2E+02
Switching time $[min]$	2.21E+02, 3.64E+02, 3.99E+02, 4.2E+02
Glucose feed-rate $[L/min]$	1.096E-03, 5.51E-04, 6.51E-04, 7.82E-04

Table 6.17: Optimized experimental conditions obtained after 1900 generations

Since the optimization algorithm does not have a measure of convergence, the evolution of the cost function along the number of generations is presented in Figure 6.8.

The obtained results shown above, indicates a grouping of the sampling points in three series. The first group is placed at the beginning of the experiment, then the second group in the middle of the experiment and the last group at the end of the experiment. The minimum difference between two measurements times constraints is active between the third and second last sampling times .

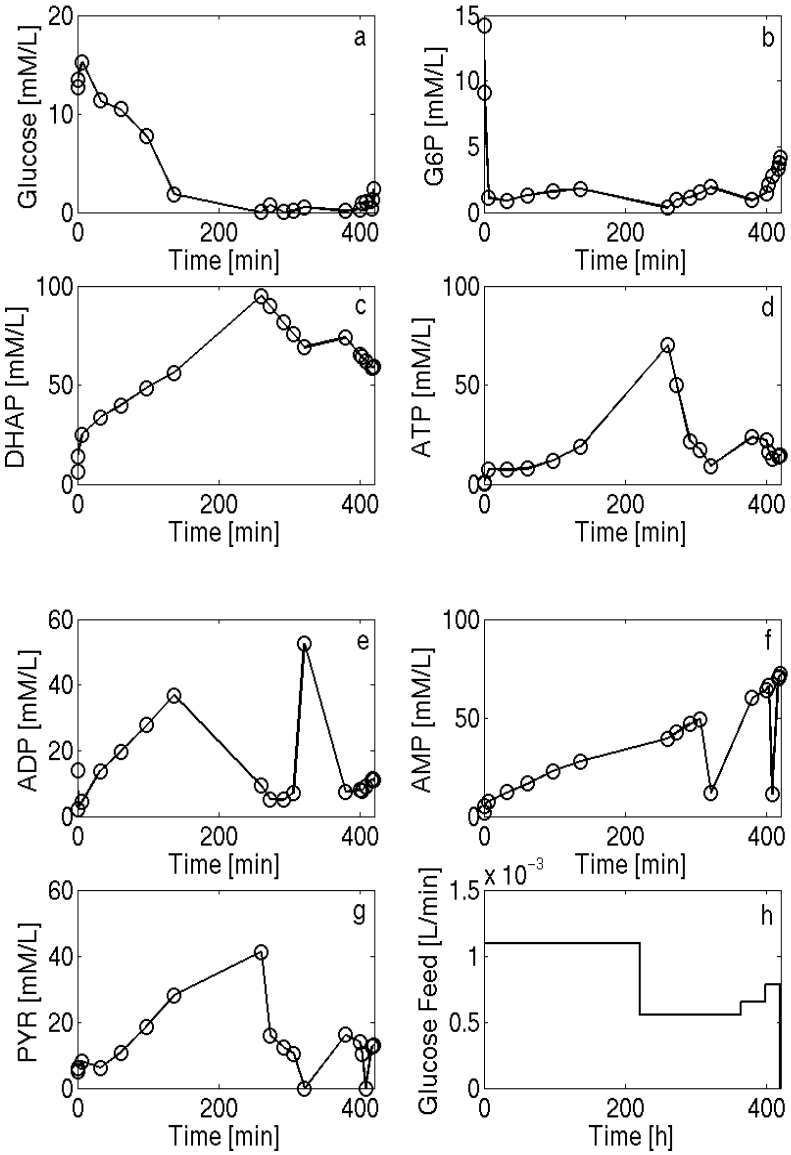


Figure 6.7: Simulated profiles for optimal experimental design for Model III, obj func =-9.66E+01 after 1900 generations

The feed flow-rate of Glucose, Figure 6.7.h is very high in the first half of the experiment, reaching the upper limit, then is reduced significantly for approximately 150 minutes and then is increased again first to approximately half

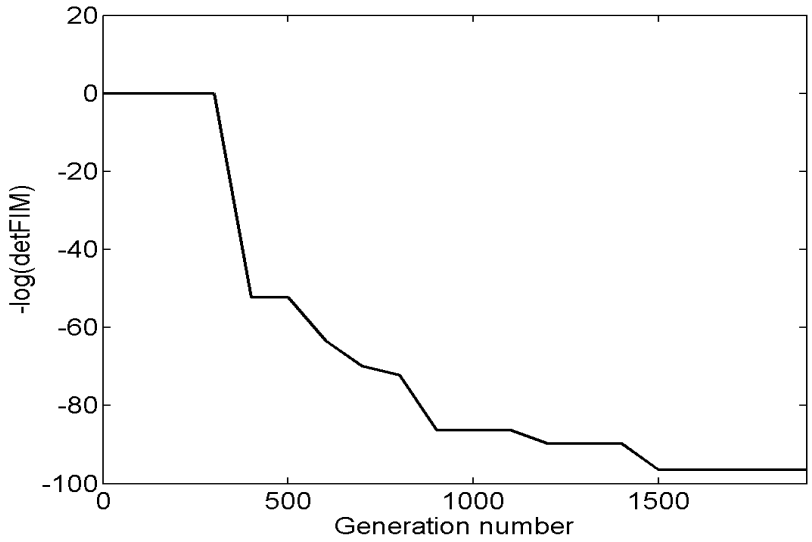


Figure 6.8: Progression of the convergence using $NP = 20$, $N = 100$ and $F = 0.9$, $CR = 0.9$

of the initial value and then at the end to almost the same value as in the beginning of the experiment. All the design variables, e.g the initial values of the concentrations, the input profile, and the sampling times have been varied significantly during the optimization. It needs to be mentioned that the measurement times were the same for all the measured compounds, thus the optimizer went for a trade off between the sensitivities of the measurements and the input profile, this might explain the large value of the Glucose concentration. The obtained metabolite concentrations are also significantly higher than the pulsed fed batch Experiment 30. Therefore it is desirable to investigate the availability of enzymes in the SBT such that the optimization can be suitably constrained.

It would be interesting however to allow in the optimization different measurement times for each measured compound, however by doing this the number of design variables increases significantly.

6.7 Conclusion

An improved grey-box stochastic modelling framework has been proposed. A grey-box stochastic model for an enzymatic reaction network has been validated using the proposed improved grey-box stochastic modelling framework. The model has been developed by combining first principle engineering knowledge with experimental data in an iterative way.

The qualitative experimental design performed in section 6.4 has been used in defining the input perturbation as well as selection of extra measured compounds applied in Experiment 30A. The qualitative experimental design results have been applied in the selection of the parameters to be estimated as well. Here the qualitative experimental design has been used mere as qualitative identifiability analysis.

The performance of Model III, in terms of pure simulation and one step ahead prediction could be further improved, thus it shows that there is a need for quantitative experimental design for parameter estimation. In order to address this issue, a quantitative experimental design for parameter estimation has been performed as well. Clearly, a very different design was obtained thus enriching the informational content of the data. It would be desirable to perform this optimal designed experiment in the laboratory and then use the collected data to (re-)estimate the model parameters.

7

Software issues

This section discusses the software tools used or developed for the work presented in the thesis.

7.1 CTSM software

The CTSM program (Kristensen et al. 2004a) is a program developed previously by a joint collaboration between CAPEC, Department of Chemical Engineering and Department of Informatics and Mathematical Modelling at Technical University of Denmark. The program aims mainly at estimating parameters of models described by stochastic differential equations. However, it can be used for performing state estimation, filtering or simulation of stochastic differential equations.

In this work, the parameter estimation results reported in chapter 2 and 6 were obtained using CTSM.

Moreover, in order to handle experimental data obtained under different conditions, e.g. different initial conditions simultaneously, the program has been extended in order to allow this possibility. A parameter or a subset of the model parameters can be estimated either globally across all the experimental data sets or individually as having a different values for each experiment. Thus rendering the program much more flexible. The program has been also extended by including a global optimizer based on the differential evolution algorithm presented in chapter 5.

7.2 EXPDSGN

The algorithm presented in chapter 5 has been implemented in a software tool called EXPDSGN using the FORTRAN programming language. The program has been utilized to determine the optimal experimental design for the case studies presented in chapters 5–6. The routines computing the log-likelihood function for a system of stochastic differential equations based on the Kalman Filter approach has been reused from the CTSM code. The optimization routine has been developed starting with a freely available on-line MATLAB^{TM} implementation which was converted into FORTRAN code. Furthermore the

extension for handling constraints discussed in chapter 5 has been implemented also.

7.3 Maple code

Parts of the identifiability analysis algorithm developed in chapters 3 and 4 have been implemented as MAPLE^{TM} scripts. Particularly the steps calculating the Lie derivatives and solving the nonlinear algebraic system of equations symbolically have been implemented. However since the user has to intervene if the machine gets stuck while solving symbolically the system of algebraic equations, implementation of a fully automated algorithm was prohibited.

8

Conclusions and Future work

This chapter reviews the general hypotheses defined in the beginning of the thesis in chapter 1, then briefly resumes the new contributions. Finally the chapter is closed by a section collecting all the suggestions for future work.

8.1 General conclusions on hypotheses

This section will briefly review the hypotheses and each of them will be commented here.

The first hypothesis was: *In order to facilitate efficient optimization of a production process it is most beneficial to develop a reasonably accurate process model where information about model uncertainty is available.*

At the end of chapter 6 a grey box stochastic model has been obtained. However in order to really validate this hypothesis a more accurate model would be required. It was considered that by investigating the model simulations would possible to help the SBT development by indicating which genes could be knocked out.

The second hypothesis: *A structural identifiability analysis step, preferably before the estimation step will guide the model parameters estimation and will reduce both the time and the number of estimations by focusing only on the theoretically identifiable process parameters.*

The results obtained in chapter 3 regarding the structurally identifiable parameters improved the parameter estimation steps performed in chapters 2 and 6 by considerably reducing the parameter vector to be estimated. The results obtained supports the validity of the hypothesis.

The third hypothesis: *A qualitative experimental design step will guide the experimentalist to focus on measuring only the relevant states and on perturbing only the relevant inputs in order to identify structurally identifiable process parameters.*

The first series of experimental data contained only two measured compounds, way below the minimal necessary number of measured compounds for identifying the parameters. After informing the experimentalist, more compounds were measured and it was possible to estimate more parameters of the model. Moreover, after performing the qualitative experimental design study in chapter 4, in order to render more parameters identifiable, the need for an input (Glucose) perturbation became clear. Once the input perturbation was introduced in the experiments, the new experimental data allowed estimation of extra parameters. It can be concluded that the hypothesis was at least partly confirmed.

The last hypothesis was: *A quantitative experimental design step for grey box stochastic models will improve the parameter estimation step by providing the experimental data containing the maximum amount of information.*

The experimental design step performed in chapter 6 provided inputs for the next series of experiments. The limited time framework made the iterations to stop here. Thus the last hypothesis could not be proved in practice. In principle it can at least be proved by estimating the parameters using the simulated data.

Some of the suggestions for future work will address inherently the validation and prove of the last two hypotheses.

8.2 New contributions

This section briefly reviews the new contributions of this thesis.

A new and more systematic and practical methodology for parameter identifiability analysis and qualitative experimental design is developed and presented in chapters 3–4. The scope of the methodology is limited to dynamic models originating from reaction networks since, the stoichiometric data is necessary in the first stage of this methodology. However, the theory behind is not new.

Another novelty introduced in this thesis is quantitative experimental design for stochastic differential equation. The new methodology presented in chapter 5 has been implemented in software tool called EXPDSGN. However the theory behind most of the elements used in this methodology are not new. For example the algorithm for computing the Fisher Information Matrix has been already introduced by Spall (2003). The optimization algorithm used has been introduced few years ago but is one of the latest research developments in the area of global stochastic optimization. The criteria and the approach for the design of experiments is a classic one.

With respect to applications, chapter 6 contains grey-box stochastic models for the system of bio-transformations.

8.3 Suggestions for future work

This section gathers all the possible suggestions for future work.

8.3.1 Further modelling

One direction for future work regarding the modelling part of the work is to investigate model IV mentioned in chapter 6 and presented in Appendix A which represents a maximum extension of model given in section 6.3. Of course a minimum condition is to obtained more extensive experimental data. These new data sets should be obtained after applying the methodology for qualitative experimental design. By performing this task before collecting more experimental data actually would help proving the third hypothesis further. The experimental data should be reacher in terms of data points for each of the collected time series in order to really be able to use the grey-box modelling framework, i.e. in order to be able to estimate the variance of the model states.

Another natural continuation is to perform an actual experiment using the optimal experimental design conditions obtained in chapter 6 and then re-estimate the model parameter in order to really prove the last hypothesis and implicitly the first one, by having a better and more detailed validated model.

Some extensions for quantitative experimental design can be suggested.

8.3.2 Parameter independent experimental design

The algorithm presented in chapter 5, has one limitation regarding the uncertainty of the model parameter values. In essence, the assumption is, that the model parameters values are the true ones. This assumption is in practice unrealistic, and usually the interval where the true value of the parameters is located is approximately known, or perhaps something about the distribution can be said.

The *Expected value* criterion mentioned above in Section5.1 is one way to deal with this situation.

One other approach to deal with this situation is the so-called robust framework for design of experiments as defined by Asprey and Macchietto (2000), Asprey and Macchietto (2002) which converts the Expected value design into a *maxmin* problem. Thus, a further extension of the quantitative experimental design is to incorporate the SD criterion, presented in 5, into an outer loop where together with an extra optimization problem whose variables are the model parameters will render a criterion providing the maximum amount of information for any value of the model parameters. This is in fact the similar to the robust design criterion as described in Asprey and Macchietto (2000), Asprey and Macchietto (2002) for systems described by DAE or ODE's. As for one of the future developments it will be interesting to investigate this criterion for stochastic differential equations. The time frame limit prevented this extension in the present work.

8.3.3 Experimental design for model discrimination

In the work presented in this thesis different models have been developed and presented starting with a simple case and then building in more and more complexity. When discriminating and modifying the model, various reasonings have been applied. Either extra physical knowledge, or experimental conditions and data or statistical test and residual analysis have been applied during the process of model development. The situation is often where there are more model candidates available with similar performance with respect to the available experimental data and therefore in this situation it is useful to design experiments to differentiate among the available models. The idea of experimental design for model discrimination is to maximize the gap between the outputs of the two models by manipulating the inputs, the switching times, the initial value of the states and the sampling times. This class of experimental design for dynamic models was introduced by Asprey and Macchietto (2000). This problem can be reformulated as follows: the log-likelihood function is computed for each of the to model candidates and then a criterion related to the ratio (or difference) of the two log-likelihood functions is maximized.

The development of the algorithm and the software tool for the class of models represented by stochastic differential equations can constitute an interesting future work. However the main difficulty is the evaluation of the log-likelihood function for the stochastic differential equations. In the work presented in this thesis, an Iterated Extended Kalman Filter based implementation (Kristensen et al. 2004a) is used to solve the propagation equations and to evaluate the log-likelihood function.

A

A very detailed kinetic model

This appendix presents a very detailed model which represents the highest level of detail for this enzymatic reaction network. Much more data are required to estimate and validate this model. Figure A.1 presents the complete reaction network. For development of the kinetic models, numerical values and the kinetic expressions have been cited from Chassagnole et al. (2002) and Teusink et al. (2000) .

The model equations with the extensions and additional kinetic data are reproduced in Eqs. A.1–A.2. The reason to include this model in the study is that it is important to evaluate the kinetic parameters as the SBT performs at quite different conditions than the ones found in a living cell. This model validation represent a future work since the model can only be validated once more compounds in the SBT experiments will be measured.

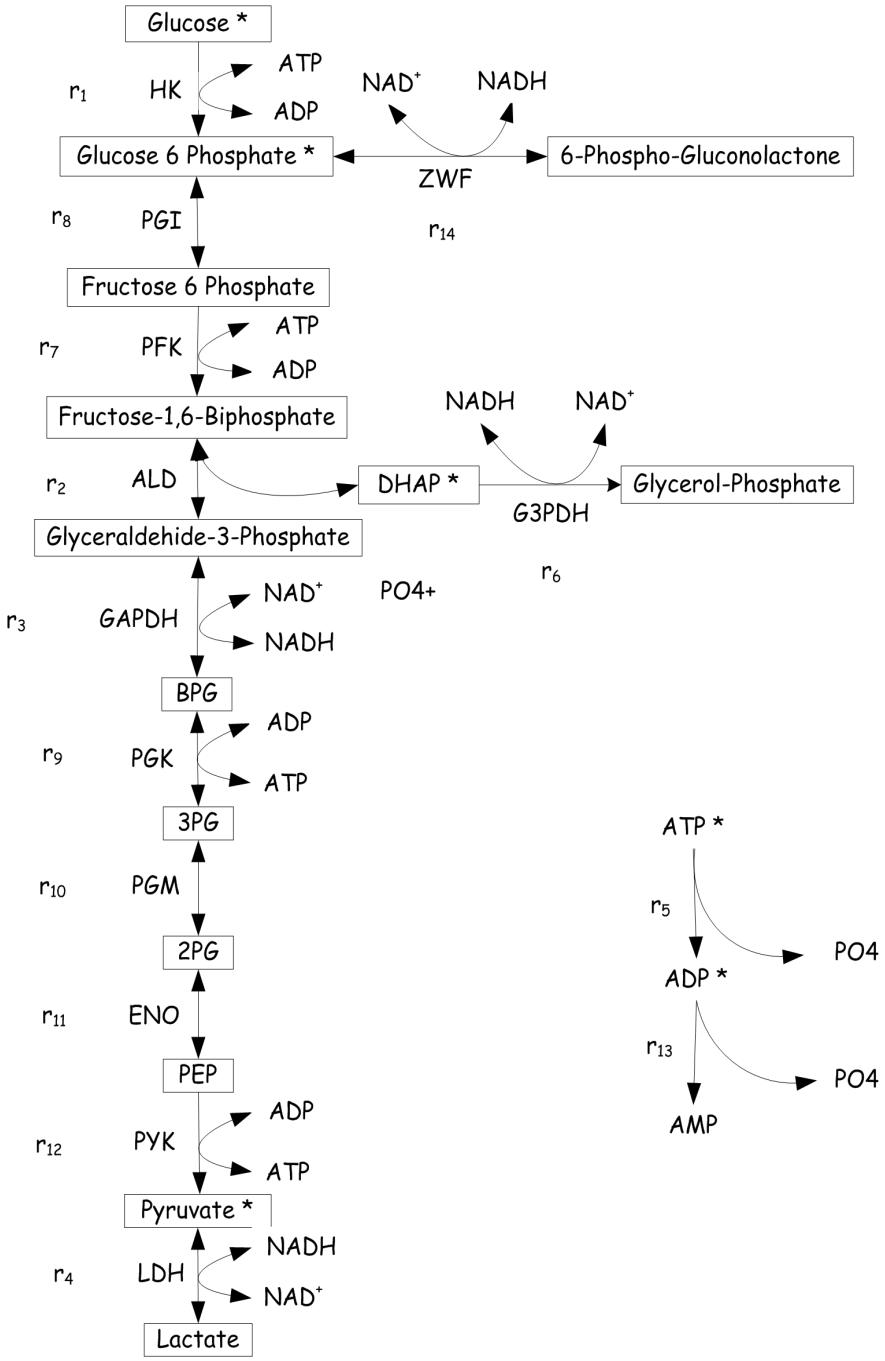


Figure A.1: The simplified reaction network used in this analysis

$$
\begin{aligned}
r_1 &= r_{1maxhk}\frac{\frac{c_{GL}c_{ATP}}{K_{GLhk}K_{ATPhk}}\left(1-\frac{c_{G6P}c_{ADP}}{c_{GL}c_{ATP}}\frac{1}{K_{eqhk}}\right)}{\left(1+\frac{c_{GL}}{K_{GLhk}}+\frac{c_{G6P}}{K_{G6Phk}}\right)\left(1+\frac{c_{ATP}}{K_{ATPhk}}+\frac{c_{ADP}}{K_{ADPhk}}\right)}\\
r_2 &= r_{2maxaldo}\frac{\frac{c_{F16B}}{K_{F16Baldo}}\left(1-\frac{c_{DHAP}c_{G3P}}{c_{F16B}}\frac{1}{K_{eqaldo}}\right)}{1+\frac{c_{DHAP}}{K_{iDHAPaldo}}+\frac{c_{F16B}}{K_{F16Baldo}}+\frac{c_{DHAP}}{K_{DHAPaldo}}+AA}\\
AA &= \frac{c_{G3P}}{K_{G3Paldo}}+\frac{c_{F16B}c_{G3P}}{K_{F16Baldo}K_{iG3Paldo}}+\frac{c_{DHAP}c_{G3P}}{K_{DHAPaldo}K_{G3P}}\\
r_3 &= r_{3maxgapdh}\frac{\frac{c_{G3P}c_{NAD}}{K_{G3Pgapdh}K_{NADgapdh}}\left(1-\frac{c_{PGP}c_{NADH}}{c_{G3P}c_{NAD}}\frac{1}{K_{eqgapdh}}\right)}{\left(1+\frac{c_{G3P}}{K_{G3Pgapdh}}+\frac{c_{NAD}}{K_{NADgapdh}}\right)\left(1+\frac{c_{PGP}}{K_{PGPgapdh}}+\frac{c_{NADH}}{K_{NADHgapdh}}\right)}\\
r_4 &= r_{4maxldh}\frac{\frac{c_{PYR}c_{NADH}}{K_{PYRldh}K_{NADHldh}}\left(1-\frac{c_{LAC}c_{NAD}}{c_{PYR}c_{NADH}}\frac{1}{K_{eqldh}}\right)}{\left(1+\frac{c_{NAD}}{K_{NADldh}}+\frac{c_{NADH}}{K_{NADHldh}}\right)\left(1+\frac{c_{LAC}}{K_{LACldh}}+\frac{c_{PYR}}{K_{PYRldh}}\right)}\\
r_5 &= r_{5max}\frac{c_{ATP}}{K_{ATP5}+c_{ATP}}\\
r_6 &= r_{6maxg3pdh}\frac{c_{DHAP}}{K_{DHAPg3pdh}+c_{DHAP}}\\
A &= 1+\frac{c_{PEP}}{K_{PEPpfk}}+\frac{c_{ADP}}{K_{ADPpfkB}}+\frac{c_{AMP}}{K_{AMPpfkB}}\\
B &= 1+\frac{c_{ADP}}{K_{ADPpfkA}}+\frac{c_{AMP}}{K_{AMPpfkA}}\\
C &= 1+\frac{L_{pfk}}{1+c_{F6P}\left(\frac{B}{K_{F6PpfkS}A}\right)^{n_{pfk}}}\\
r_7 &= r_{7maxpfk}\frac{c_{F6P}c_{ATP}}{\left(c_{ATP}+K_{ATPpfkS}\left(1+\frac{c_{ADP}}{K_{ADPpfkC}}\right)\right)\left(c_{F6P}+K_{F6PpfkS}\frac{A}{B}\right)\ (C)}\\
r_8 &= r_{8maxpgi}\frac{\frac{c_{G6P}}{K_{G6Ppgi}}\left(1-\frac{c_{F6P}}{c_{G6P}}\frac{1}{K_{eqpgi}}\right)}{1+\frac{c_{G6P}}{K_{G6Ppgi}}+\frac{c_{F6P}}{K_{F6Ppgi}\left(1+\frac{c_{6PG}}{K_{F6Ppgi,i6PG}}\right)}+\frac{c_{6PG}}{K_{G6Ppgi,i6PG}}}\\
r_9 &= r_{9maxpgk}\frac{\frac{c_{ADP}c_{BPG}}{K_{ADPpgk}K_{BPGpgk}}\left(1-\frac{c_{ATP}c_{3PG}}{c_{ADP}c_{PGP}}\frac{1}{K_{eqpgk}}\right)}{\left(1+\frac{c_{ADP}}{K_{ADPpgk}}+\frac{c_{BPG}}{K_{BPGpgk}}\right)\left(1+\frac{c_{ATP}}{K_{ATPpgk}}+\frac{c_{3PG}}{K_{3PGpgk}}\right)}\\
r_{10} &= r_{10maxpgm}\frac{\frac{c_{3PG}}{K_{3PGpgm}}\left(1-\frac{c_{2PG}}{c_{3PG}}\frac{1}{K_{eqpgm}}\right)}{1+\frac{c_{3PG}}{K_{3PGpgm}}+\frac{c_{2PG}}{K_{2PGpgm}}}\\
r_{11} &= r_{11maxeno}\frac{\frac{c_{2PG}}{K_{2PGeno}}\left(1-\frac{c_{PEP}}{c_{2PG}}\frac{1}{K_{eqeno}}\right)}{1+\frac{c_{2PG}}{K_{2PGeno}}+\frac{c_{PEP}}{K_{PEPeno}}}\\
r_{12} &= r_{12maxpyk}\frac{\frac{c_{PEP}}{K_{PEPpyk}}\left(\frac{c_{PEP}}{K_{PEPpyk}}+1\right)^{n_{pyk}-1}\left(\frac{c_{ADP}}{c_{ADP}+K_{ADPpyk}}\right)}{L_{pyk}\left(\frac{\frac{c_{ATP}}{K_{ATPpyk}}+1}{\frac{c_{F16B}}{K_{F16Bpyk}}+\frac{c_{AMP}}{K_{AMPpyk}}+1}\right)^{n_{pyk}}+\left(1+\frac{c_{PEP}}{K_{PEPpyk}}\right)^{n_{pyk}}}\\
r_{13} &= r_{13max}\frac{c_{ADP}}{K_{ADP}+c_{ADP}}\\
r_{14} &= r_{14maxg6pdh}\frac{c_{G6P}c_{NAD}}{(c_{G6P}+K_{G6Pg6pdh})\left(1+\frac{c_{NADH}}{K_{iG6Pg6pdh}}\right)\left(K_{NADg6pdh}\left(1+\frac{c_{NADH}}{K_{iNADg6pdh}}\right)+c_{NAD}\right)}
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
dc_{GL} &= -r_1 + \sigma_{11} d\omega \\
dc_{F16B} &= r_7 - r_2 + \sigma_{22} d\omega \\
dc_{DHAP} &= r_2 - r_6 + \sigma_{33} d\omega \\
dc_{G3P} &= r_2 - r_3 + \sigma_{44} d\omega \\
dc_{PYR} &= r_{12} - r_4 + \sigma_{55} d\omega \\
dc_{LAC} &= r_4 + \sigma_{66} d\omega \\
dc_{ATP} &= -r_1 - r_7 - r_5 + r_9 + r_{12} + \sigma_{77} d\omega \\
dc_{NAD} &= -r_3 - r_{14} + r_4 + r_6 + \sigma_{88} d\omega \\
dc_{PO4} &= -r_3 + r_5 + r_{13} + \sigma_{99} d\omega \\
dc_{G6P} &= r_1 - r_8 + \sigma_{1010} d\omega \\
dc_{ADP} &= -r_9 - r_{12} + r_1 + r_7 + r_5 + \sigma_{1111} d\omega \\
dc_{NADH} &= r_3 - r_4 + r_1 4 + r_6 + \sigma_{1212} d\omega \\
dc_{F6P} &= r_8 - r_7 + \sigma_{1313} d\omega \\
dc_{AMP} &= r_{13} + \sigma_{1414} d\omega \\
dc_{BPG} &= -r_9 + r_3 + \sigma_{1515} d\omega \\
dc_{3PG} &= r_9 - r_{10} + \sigma_{1616} d\omega \\
dc_{2PG} &= r_{10} - r_{11} + \sigma_{1717} d\omega \\
dc_{PEP} &= r_{11} - r_{12} + \sigma_{1818} d\omega \\
dc_{6PG} &= r_{14} + \sigma_{1919} d\omega \\
dc_{PG} &= r_6 + \sigma_{2020} d\omega
\end{aligned}
\tag{A.2}
$$

B

Experimental data

The experiments have been conducted at ETH Zurich (Schümperli et al. 2007). In phase I, fed-batch (semi-batch) fermentations of *E. coli* LJ110 *tpi* are conducted until the optical density (OD) in the bioreactor reaches a preset value of 600. The broth is centrifuged and the cells are resuspended in SBT-buffer (100 mM HEPES, 0.84 mM KCl, 1 mM ZnSO4 and at pH = 7). The cells are disrupted by high-pressure homogenization. The remaining solids are eliminated by centrifugation/filtration and the liquid extract is recovered. The total protein concentration is determined by Bradford analysis and adjusted to the desired concentration by dilution with SBT buffer. The liquid extract contains the enzymes and compounds present in the cell at the time when the fermentation was stopped. In phase II, a volume of 5 ml of SBT extract is used for each experiment. Defined amounts of Hexokinase and Lactate-DH as well as ATP and NAD^+, are added. The reactions are initiated by adding the substrate glucose. Samples are collected according to a previously defined time plan. The experiments are terminated after 300 or 360 minutes. First, the proteins are removed by precipitation with HCl (1M) followed by centrifugation. The samples are analyzed by enzymatic assays. First Glucose-6-Phosphate is determined, and then Glucose and Glucose-6-Phosphate are determined together by addition of Glucose-6-Phosphate-dehydrogenase and Hexokinase to form $NADPH$ which is determined spectro-photometrically. Di-Hydroxy-Acetone-Phosphate is determined by addition of Glycerol-3-Phosphate-dehydrogenase and measuring the $NADH$ consumption spectro-photometrically. ATP is measured as well. Each experiment is defined by the SBT extract design, by the initial concentrations of Glucose, cofactors, Phosphate and enzymes, by the sampling time and by the end time. The samples were collected frequently in the beginning (between 1-15 minutes) and less frequent towards the end of the batch and the total time was 360 minutes. The initial concentration of the enzymes, cofactors, phosphate and glucose is also known, with some uncertainty.

B.1 Experimental procedure

SBT production or phase 1

- The E-coli tpi KO mutant fermentation is performed using a 3 Liters reactor and operated in fed-batch mode.
- When the optical density (OD) reaches the value of 20 units the cultivation is stopped.
- The reactor broth is pumped out and cooled to $4°C$.

Kinetic experiments or phase 2

- The broth is centrifuged and the solid (the cells) is recovered
- The cells are suspended in the SBT buffer. The SBT buffer composition is: 100 mM HEPES, 0.84 mM KCl, 5mM MgCl, 1 mM ZnSO4 at pH = 7
- Cell walls are rupted by high-pressure homogenization
- The solid particles are eliminated using filtration and the liquid is recovered
- Total protein concentration is measured and the desired protein concentration is corrected by adding SBT buffer solution
- Take a volume of 5 ml and feed it in the beaker
- Add the components (enzymes + ATP + NAD+ + PO4) that correspond to the desired concentration
- Start reaction by adding the glucose amount corresponding to its desired initial concentration
- Take samples after 15, 30, 60, 90, 120, 180, 240, 300 etc min.
- Prepare the sample for enzymatic analysis
  - Add HCl to change the pH to 1 and precipitate all the proteins
  - Bring the pH back to neutral
  - Filter by centrifugation to remove the precipitated proteins
- Perform enzymatic assay of the samples

B.2 Experiment 1

4 parallel measurements have been taken for each point in time from 2 reactors. The reactors were 2 beakers in a shaker.

Comp Exp. No.	**Exp. 1 A**	**Exp. 1 B**
Hexokinase U/ml	1	1
Lactate-DH U/ml	1	1
Glc mM	11.1	11.1
PO4 mM	11.1	11.1
ATP mM	11.5	11.5
NAD+ mM	11.5	11.5
Prot.tot. mg/ml	20	20
T	30	30
Strain	LJ110-tpi	wild type

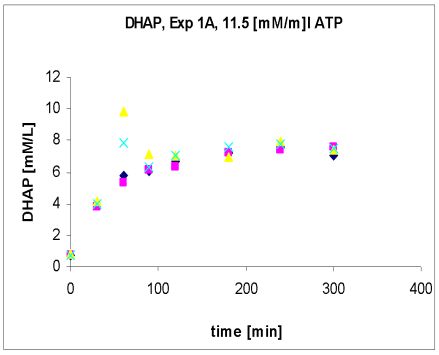
Table B.1: Experiment 1 description

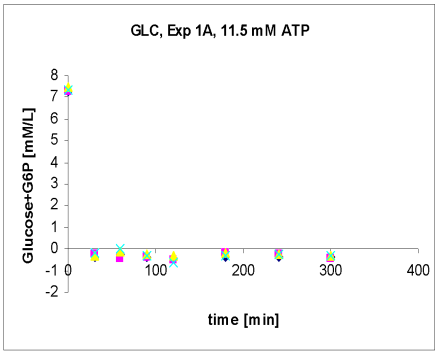
Time	DHAP				GL			
0	0.7	0.69	0.82	0.72	7.28	7.26	7.5	7.34
30	3.92	3.81	4.11	3.98	-0.39	-0.27	-0.31	-0.16
60	5.75	5.33	9.82	7.86	-0.27	-0.44	-0.13	0.04
90	6.05	6.18	7.1	6.31	-0.38	-0.37	-0.28	-0.31
120	6.67	6.34	7.01	7.05	-0.44	-0.53	-0.32	-0.66
180	7.18	7.26	6.96	7.58	-0.37	-0.16	-0.21	-0.33
240	7.6	7.42	7.9	7.77	-0.41	-0.25	-0.21	-0.28
300	7.02	7.59	7.39	7.45	-0.35	-0.41	-0.28	-0.33

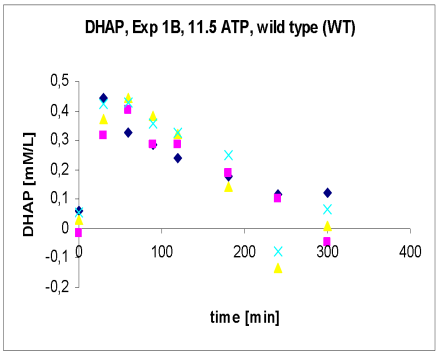
Table B.2: Experiment 1A

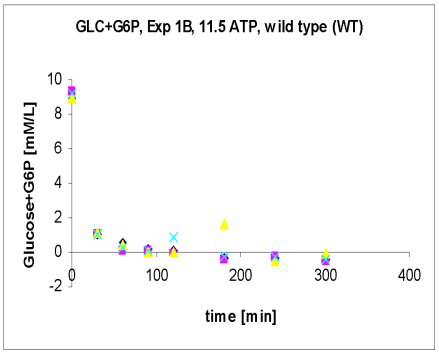
Time	DHAP				GL			
0	0.06	-0.02	0.03	0.06	8.93	9.26	8.94	9.22
30	0.44	0.32	0.37	0.42	1.06	1.03	1.12	1.04
60	0.32	0.4	0.44	0.43	0.5	0.07	0.44	0.25
90	0.28	0.28	0.38	0.36	0.19	0.04	0	0.09
120	0.24	0.29	0.32	0.32	0.09	-0.08	-0.03	0.9
180	0.18	0.19	0.14	0.25	-0.35	-0.46	1.62	-0.28
240	0.12	0.1	-0.13	-0.08	-0.39	-0.26	-0.48	-0.26
300	0.12	-0.05	0.01	0.07	-0.45	-0.49	-0.13	-0.47

Table B.3: Experiment 1B


DHAP, Exp 1A, 11.5 [mM/m]l ATP
DHAP [mM/L]
time [min]


GLC, Exp 1A, 11.5 mM ATP
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 1B, 11.5 ATP, wild type (WT)
DHAP [mM/L]
time [min]


GLC+G6P, Exp 1B, 11.5 ATP, wild type (WT)
Glucose+G6P [mM/L]
time [min]

B.3 Experiment 2

In experiment 2 the protein concentration was varied from 2 [mg/ml] to 20 [mg/ml] in three batch experiments.

Comp Exp. No.	**Exp. 2 A**	**Exp.2 B**	**Exp. 2 C**
Hexokinase U/ml	1	1	1
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11.1	11.1
ATP mM	11.5	11.5	11.5
NAD+ mM	11.5	11.5	11.5
Prot.tot. mg/ml	20	10	2
T	30	30	30
Strain	LJ110-tpi	LJ110-tpi	LJ110-tpi

Table B.4: Experiment 2 description

Time	DHAP				GL			
0	0.34	0.24	0.24	0.34	8.7	8.98	10.44	10.37
30	2.72	2.63	2.98	2.82	-0.13	-0.23	-0.34	-0.51
60	4.19	1.56	4.22	4.19	-0.4	-0.66	-0.41	-0.55
90	5.29	5.24	5.06	5.05	-0.64	-0.92	-0.43	-0.53
120	5.96	6.04	5.55	5.1	-1.77	-0.67	-0.58	-0.64
180	6.78	6.77	6.06	6.36	-0.59	-0.9	-0.7	-0.59
240	7.25	7.22	6.35	6.3	-0.46	-0.56	-0.08	-0.36
300	7.39	7.37	6.77	6.75	-0.21	-0.63	-0.71	-0.7

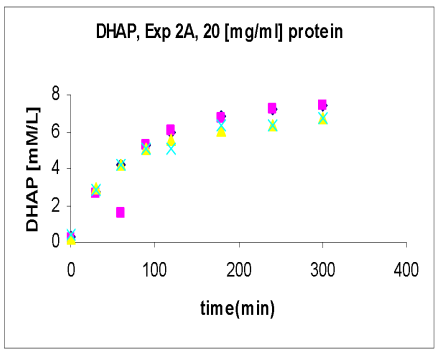
Table B.5: Experiment 2A, 20 [mg/ml]

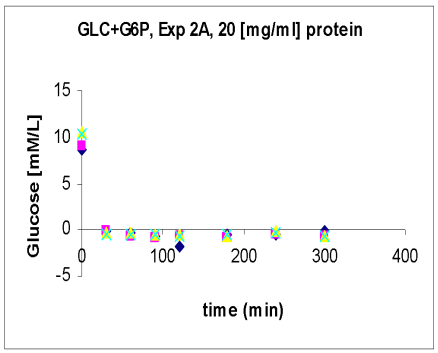
Time	DHAP				GL			
0	0.13	0.1	0.13	0.07	9.26	9.66	11.61	11.54
30	1.51	1.54	1.59	1.43	1.7	1.75	3.08	3.06
60	2.62	2.44	2.4	2.39	0.08	-0.04	0.88	0.81
90	3.04	3	3.12	2.98	-0.31	-0.31	-0.06	-0.07
120	3.61	3.69	3.63	3.61	-0.21	-0.42	-0.36	-0.09
180	4.26	4.19	4.36	4.24	-0.42	-0.33	-0.15	-0.11
240	4.66	4.59	4.78	4.65	-0.29	-0.07	-0.14	0.18
300	4.74	4.64	4.95	4.78	-0.28	-0.4	-0.31	0.06

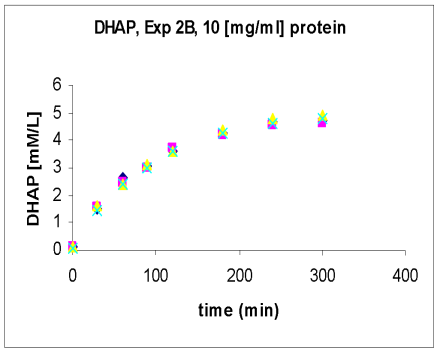
Table B.6: Experiment 2B, 10 [mg/ml]

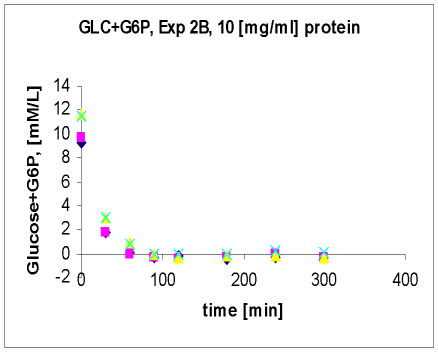
Time	DHAP				GL			
0	0	-0.46	-0.05	-0.04	9.85	10.41	10.18	10.64
30	0.29	-1.59	0.16	0.23	6.02	5.86	5.76	6.26
60	0.56	0.31	0.53	0.55	4.96	4.79	4.43	4.38
90	0.76	0.65	0.83	0.83	3.63	3.62	3.51	3.64
120	0.96	1.03	0.8	1.05	2.86	2.79	2.88	3.11
180	1.37	1.34	1.39	1.45	1.98	1.51	2.07	1.94
240	1.68	1.58	1.29	1.3	1.22	1.34	0.82	0.88
300	1.95	2.05	2.03	1.9	0.83	1.04	0.85	0.77

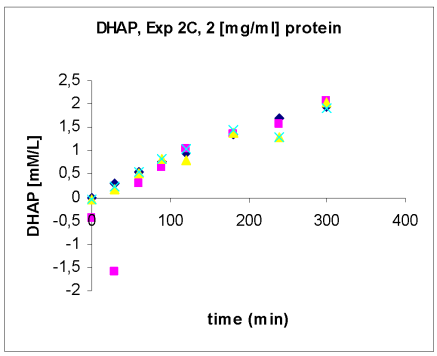
Table B.7: Experiment 2C, [2 mg/ml]

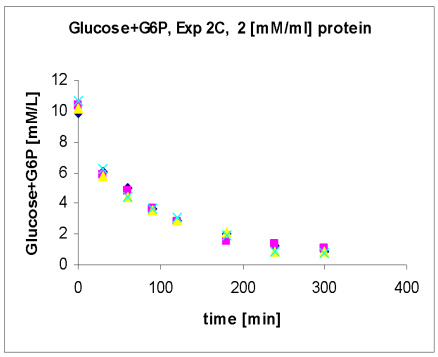

DHAP, Exp 2A, 20 [mg/ml] protein
DHAP [mM/L]
8
6
4
2
0
0
100
200
300
400
time(min)


GLC+G6P, Exp 2A, 20 [mg/ml] protein
Glucose [mM/L]
15
10
5
0
-5
0
100
200
300
400
time (min)


DHAP, Exp 2B, 10 [mg/ml] protein
DHAP [mM/L]
6
5
4
3
2
1
0
0
100
200
300
400
time (min)


GLC+G6P, Exp 2B, 10 [mg/ml] protein
Glucose+G6P, [mM/L]
14
12
10
8
6
4
2
0
-2
0
100
200
300
400
time [min]


DHAP, Exp 2C, 2 [mg/ml] protein
DHAP [mM/L]
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
0
100
200
300
400
time (min)


Glucose+G6P, Exp 2C, 2 [mM/ml] protein
Glucose+G6P [mM/L]
12
10
8
6
4
2
0
0
100
200
300
400
time [min]

B.4 Experiment 3

In experiment 3 the temperature was varied and while keeping the same amount of hexokinase.

Comp Exp. No.	**Exp. 3 A**	**Exp. 3 B**	**Exp. 3 C**
Hexokinase U/ml	1	1	1
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11.1	11.1
ATP mM	11.5	11.5	11.5
NAD+ mM	11.5	11.5	11.5
Prot.tot. mg/ml	10	10	10
T	37	30	25
Strain	LJ110-tpi	LJ110-tpi	LJ110-tpi

Table B.8: Experiment 3 description

Time	DHAP				GL			
0	0.22	0.14	0.19	0.09	8.97	8.94	9.27	9.26
15	1.66	1.68	1.62	1.6	1.91	1.72	1.77	1.65
30	2.69	2.62	2.65	2.64	0.21	0.14	0.13	0.11
60	4.1	4.06	3.99	3.81	0.39	-0.09	0.21	-0.22
90	5.06	5.03	4.85	4.74	-0.12	-0.25	-0.27	-0.46
120	5.72	5.76	5.25	5.3	-0.09	0.02	-0.34	-0.24
180	6.97	6.75	5.85	5.78	-0.27	0.04	-0.34	-0.33
240	6.45	6.31	5.93	5.89	0.21	-0.38	-0.24	-0.32
300	7.35	7.19	6.33	6.22	1.78	-0.21	-0.33	-0.3

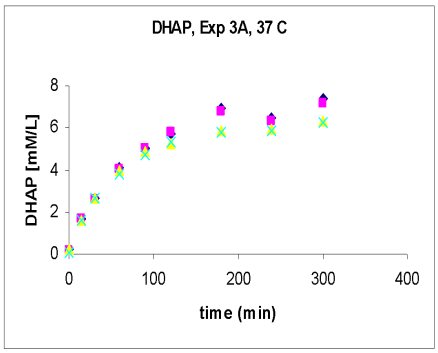
Table B.9: Experiment 3A, 37 C

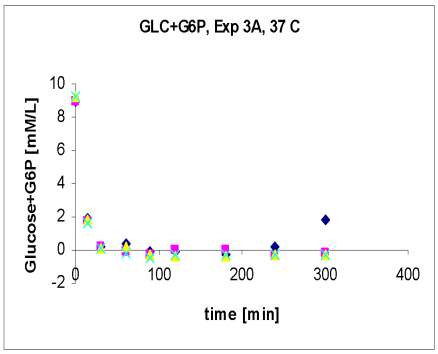
Time	DHAP				GL			
0	0.19	0.09	0.18	0.15	8.7	8.8	8.39	8.24
15	1.18	1.1	1.17	1.08	3.33	3.43	3.18	3.4
30	1.87	1.93	1.82	1.83	1.02	1.3	0.82	0.97
60	2.96	2.83	2.79	2.78	-0.15	-0.18	-0.49	-0.22
90	3.82	3.55	3.62	3.59	-0.23	-0.27	-0.14	-0.28
120	4.26	4.12	4.09	4.14	-0.31	-0.27	-0.45	-0.29
180	4.95	5.02	4.79	4.85	-0.22	0.02	-0.4	-0.27
240	5.39	5.21	5.11	4.99	-0.1	-0.27	-0.54	-0.27
300	5.71	5.64	5.4	5.15	-0.27	-0.27	-0.25	-0.04

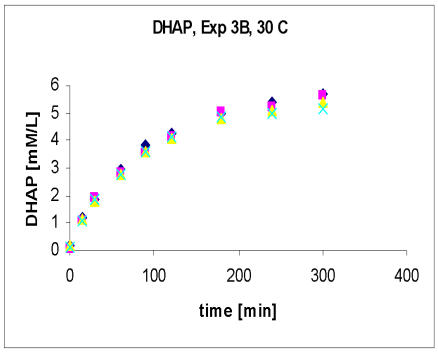
Table B.10: Experiment 3B, 30 C

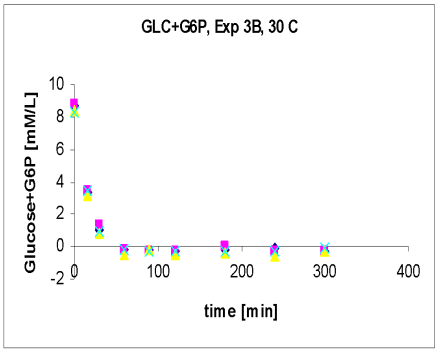
Time	DHAP				GL			
0	0.05	0.05	-0.01	-0.09	9.1	9.69	8.98	8.83
15	0.73	0.73	0.7	0.73	4.08	4.4	4.22	4.16
30	1.46	1.46	1.23	1.22	2.88	2.41	2.19	2.34
60	2.14	2.14	2.1	2.02	0.5	0.36	0.2	0.35
90	2.72	2.72	2.63	2.63	-0.07	0.03	-0.1	-0.06
120	3.18	3.18	3.19	3.17	-0.11	-0.35	-0.2	-0.14
180	3.87	3.87	3.92	3.89	-0.34	-0.24	-0.47	-0.24
240	4.23	4.23	4.38	4.38	-0.39	-0.37	-0.35	-0.21
300	4.53	4.53	4.53	4.54	-0.17	-0.22	-0.31	-0.12

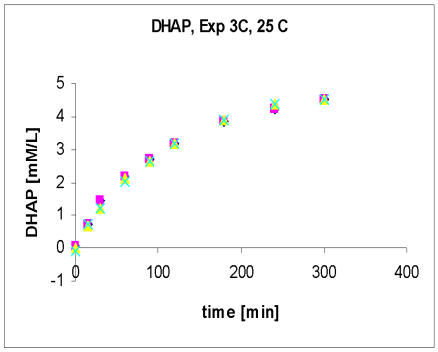
Table B.11: Experiment 3C, 25 C

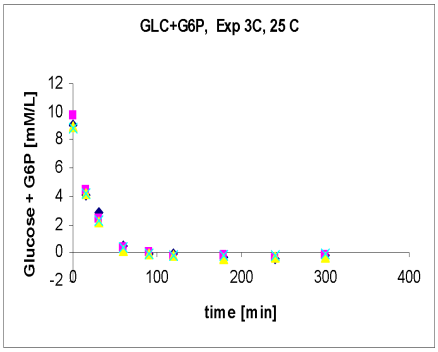

DHAP, Exp 3A, 37 C
DHAP [mM/L]
time (min)


GLC+G6P, Exp 3A, 37 C
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 3B, 30 C
DHAP [mM/L]
time [min]


GLC+G6P, Exp 3B, 30 C
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 3C, 25 C
DHAP [mM/L]
time [min]


GLC+G6P, Exp 3C, 25 C
Glucose + G6P [mM/L]
time [min]

B.5 Experiment 4

For experiment 4 the amount of hexokinase has been reduced from 1 [U/ml] to 0.06 [U/ml]. In this experiment the amount of ATP has been varied from 11.5 [mM/ml] to 1.15 [mM/ml] stepwise.

Comp Exp. No.	**Exp. 4 A**	**Exp. 4 B**	**Exp. 4 C**
Hexokinase U/ml	0.06	0.06	0.06
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11.1	11.1
ATP mM	11.5	5.75	1.15
NAD+ mM	11.5	11.5	11.5
Prot.tot. mg/ml	10	10	10
T	37	37	37
Strain	LJ110-tpi	LJ110-tpi	LJ110-tpi

Table B.12: Experiment 4 description

Time	DHAP				GL			
0	0.13	-0.07	-0.02	-0.08	11.25	11.74	11.28	11.28
15	1.24	1.27	1.29	1.33	4.48	4.65	4.84	4.9
30	2.49	2.33	2.37	2.55	2.1	1.83	2.18	2.2
60	4.05	3.97	4.01	4.35	0.03	0	0.08	-0.01
90	5.08	4.84	5.12	4.92	-0.13	-0.11	-0.17	0.14
120	5.5	5.45	5.88	5.77	-0.41	0.01	-0.01	0.12
180	6.06	5.96	6.66	6.52	-0.14	-0.21	-0.1	-0.13
240	6.84	6.78	7.61	7.05	0.67	-0.17	0.02	0.06
300	7.39	7.44	8.04	7.96	-0.27	-0.37	0.09	-0.12

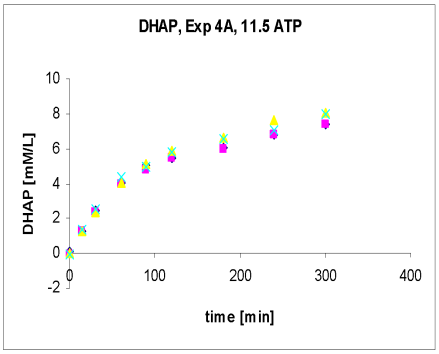
Table B.13: Experiment 4A, 11.5 mM ATP

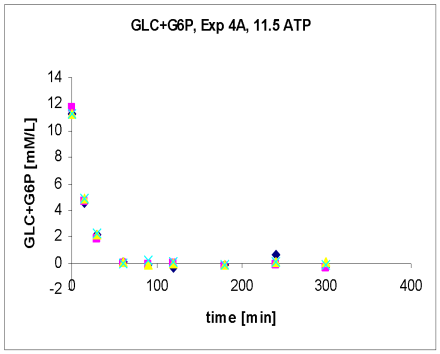
Time	DHAP				GL			
0	0.13	0.14	0.1	0.1	28.49	28.66	20.19	19.74
15	1.22	1.14	0.94	0.93	5.12	5.23	5.79	5.73
30	1.89	1.81	1.57	1.57	4.04	4.43	4.97	4.69
60	2.97	2.77	2.42	2.38	3.1	3.15	4.07	3.86
90	3.58	3.48	2.92	2.9	2.38	1.88	3.11	3.06
120	3.97	3.89	3.26	3.24	1.83	1.82	2.56	2.17
180	4.36	4.2	3.69	3.62	0.98	0.88	1.77	1.78
240	4.56	4.33	3.91	3.89	0.24	0.28	1.33	1.04
300	4.63	4.4	3.87	3.92	0.06	0	0.5	0.67

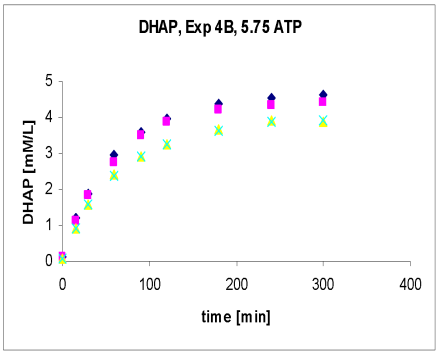
Table B.14: Experiment 4B, 5.75 mM ATP

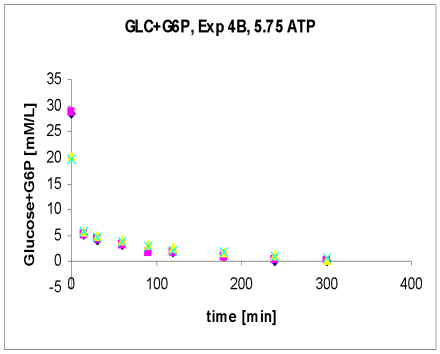
Time	DHAP				GL			
0	0.1	0.1	0.08	0.14	18.68	18.34	12.06	11.82
15	0.5	0.37	0.39	0.35	7.67	8.35	8.74	8.9
30	0.67	0.68	0.62	0.62	7.99	7.8	8.56	8.55
60	1.16	1.17	1.07	1.12	7.56	7.58	7.83	7.97
90	1.61	1.39	1.49	1.49	6.69	6.6	6.98	7.02
120	1.92	1.78	1.9	1.87	6.04	6.01	6.2	6.29
180	2.4	2.25	2.37	2.23	4.76	4.75	4.97	4.91
240	2.85	2.77	2.75	2.69	3.85	3.91	4.77	3.72
300	2.97	3.03	2.88	2.87	2.82	3.11	2.86	2.65

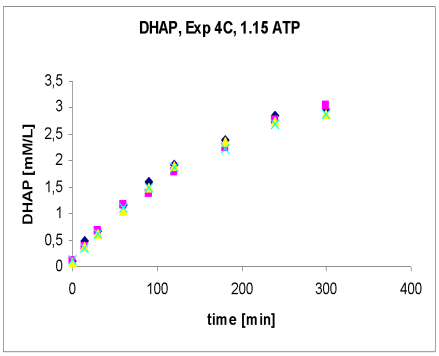
Table B.15: Experiment 4C, 1.15 ATP

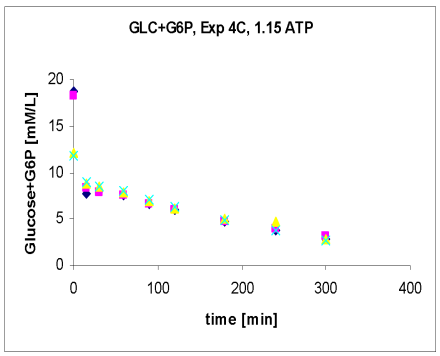

DHAP, Exp 4A, 11.5 ATP
DHAP [mM/L]
time [min]


GLC+G6P, Exp 4A, 11.5 ATP
GLC+G6P [mM/L]
time [min]


DHAP, Exp 4B, 5.75 ATP
DHAP [mM/L]
time [min]


GLC+G6P, Exp 4B, 5.75 ATP
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 4C, 1.15 ATP
DHAP [mM/L]
time [min]


GLC+G6P, Exp 4C, 1.15 ATP
Glucose+G6P [mM/L]
time [min]

B.6 Experiment 5

Comp Exp. No.	**Exp. 5 A**	**Exp. 5 B**	**Exp. 5 C**
Hexokinase U/ml	0.06	0.06	0.06
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11.1	11.1
ATP mM	5.57	5.75	5.75
NAD+ mM	11.5	5.75	1.15
Prot.tot. mg/ml	10	10	10
T	37	37	37
Strain	LJ110-tpi	LJ110-tpi	LJ110-tpi

Table B.16: Experiment 5 description

Time	DHAP				GL			
0	0.06	0.1	0.05	0.03	9.08	9.49	9.66	9.6
15	0.85	0.87	0.89	0.94	6.1	5.82	5.81	5.81
30	1.49	1.43	1.51	1.53	5.12	5.27	5.25	4.97
60	2.28	2.3	2.39	2.45	3.96	3.96	3.97	3.96
90	2.79	2.78	2.97	2.93	3.47	3.66	3.09	3.13
120	3.18	3.12	3.22	3.08	2.88	3.08	2.78	2.32
180	3.64	3.62	3.64	3.56	2.58	2.7	2.01	1.85
240	4.15	4.13	3.7	3.7	2.28	2.34	1.19	1.14
300	4.5	4.42	3.64	3.6	1.82	1.94	0.66	0.72

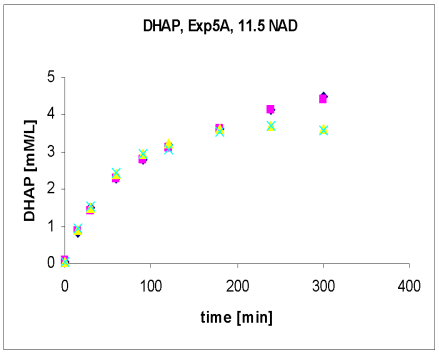
Table B.17: Experiment 5A, 11.5 mM NAD

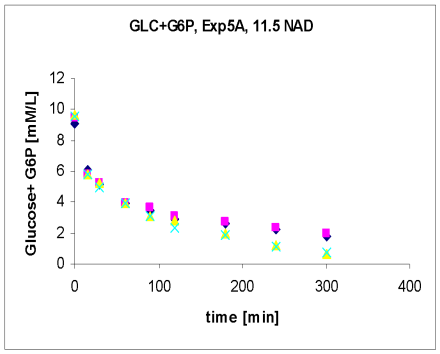
Time	DHAP				GL			
0	0.03	0.07	0	0	10.12	10.38	10.48	10.56
15	0.92	0.91	0.88	0.9	5.6	5.95	6.06	6.02
30	1.46	1.37	1.49	1.53	4.86	4.94	4.93	4.76
60	2.2	2			3.89	3.86		
90	2.75	2.56			3.31	3.32		
120	2.86	2.86	2.9	2.82	2.53	2.67	2.67	2.72
180	3.22	3.1	3.4	3.19	2.42	1.93	2.1	2.07
240	3.15	3.3	3.36	3.92	1.28	1.33	1.28	1.42
300	3.16	3.09	3.32	3.97	0.69	1.04	0.9	0.81

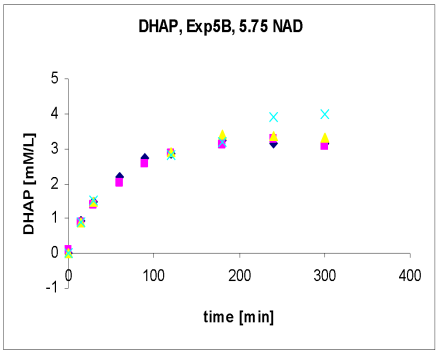
Table B.18: Experiment 5B, 5.75 mM NAD

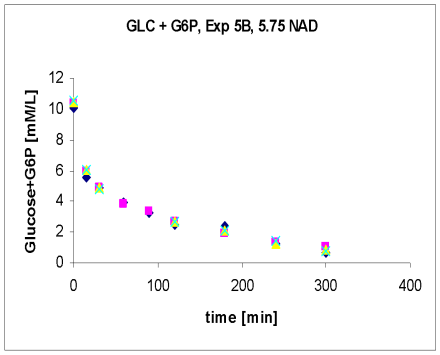
Time	DHAP				GL			
0	-0.17	0.07	0.09	0.07	10.19	10.36	10.57	10.77
15	0.87	0.86	0.81	0.89	5.77	5.91	5.8	5.93
30	1.42	1.24	1.41	1.41	5.03	4.67	5.25	4.91
60	2.17	2.2	2.04	2.11	3.56	3.78	3.99	4.19
90	2.6	2.67	2.69	2.6	3.38		3.05	3.5
120	2.98	2.97	2.81	2.88	2.71	2.87	2.85	3.05
180	3.27	3.18	3.2	3.21	1.95	1.72	2.14	2.09
240	3.49	3.24	3.36	3.26	1.21	1.64	1.5	1.69
300	3.55	3.61	3.64	3.52	1.04	1.04	1.51	1.05

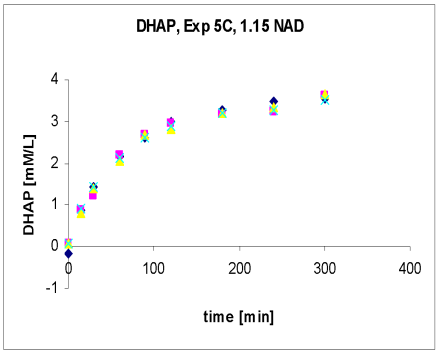
Table B.19: Experiment 5C, 1.15 mM NAD

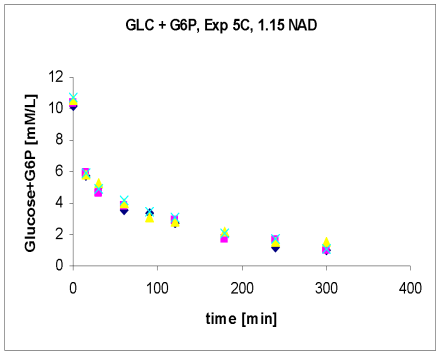

DHAP, Exp5A, 11.5 NAD
DHAP [mM/L]
time [min]


GLC+G6P, Exp5A, 11.5 NAD
Glucose+ G6P [mM/L]
time [min]


DHAP, Exp5B, 5.75 NAD
DHAP [mM/L]
time [min]


GLC + G6P, Exp 5B, 5.75 NAD
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 5C, 1.15 NAD
DHAP [mM/L]
time [min]


GLC + G6P, Exp 5C, 1.15 NAD
Glucose+G6P [mM/L]
time [min]

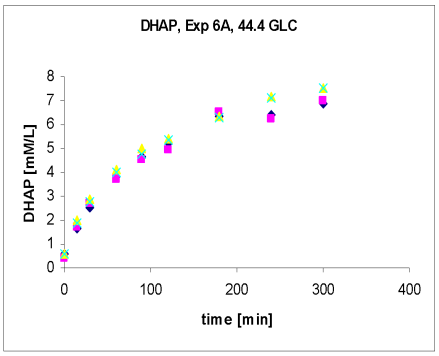
B.7 Experiment 6

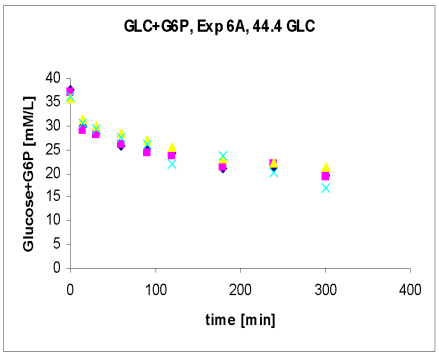
Comp Exp. No.	Exp. 6 A
Hexokinase U/ml	0.06
Lactate-DH U/ml	1
Glc mM	44.4
PO4 mM	44.4
ATP mM	11.5
NAD+ mM	11.5
Prot.tot. mg/ml	10
T	37
Strain	LJ110-tpi

Table B.20: Experiment 6 description

Time	DHAP				GL			
0	0.61	0.41	0.59	0.56	37.74	37.02	35.94	36.07
15	1.64	1.73	1.97	1.87	29.93	28.95	31.29	30.61
30	2.54	2.69	2.85	2.77	28.47	28.09	29.93	29.4
60	3.83	3.73	4.11	4.03	25.81	26.02	28.39	27.43
90	4.66	4.53	4.99	4.77	25.22	24.2	27.03	26.19
120	5.21	4.94	5.43	5.38	24.29	23.67	25.37	21.84
180	6.33	6.51	6.37	6.32	21.09	21.46	22.99	23.78
240	6.39	6.26	7.2	7.13	21.42	21.93	22.3	20.14
300	6.89	6.98	7.54	7.55	19.57	19.27	21.22	16.86

Table B.21: Experiment 6A, 44 mM GL





B.8 Experiment 7

Comp Exp. No.	**Exp. 7 A**	**Exp. 7 B**	**Exp. 7 C**
Hexokinase U/ml	0.06	0.06	0.06
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11.1	11.1
ATP mM	11.5	11.5	11.5
NAD+ mM	11.5	11.5	11.5
Prot.tot. mg/ml	10	10	10
T	37	37	37
Strain	LJ110-tpi	LJ110-tpi	LJ110-tpi

Table B.22: Experiment 7 description

Time	DHAP				GL			
0	0.55	0.41	0.53	0.58	9.6	9.7	9.88	9.85
15	1.69	1.77	1.68	1.75	5.12	4.61	5.98	5.69
30	2.82	2.75	3.1	2.83	2.61	2.44	3.71	3.48
60	4.29	4.32	4.42	4.08	0.51	0.78	1.4	1.3
90	5.05	5.04	5.06	5.18	0.09	0.08	0.5	0.38
120	5.49	5.24	5.29	5.59	-0.44	-0.24	0.05	-0.04
180	5.84	5.68	5.95	5.81	-0.49	-0.15	0.15	-0.15
240	5.8	5.74	6.1	5.88	-0.5	-0.43	-0.81	-0.44
300	5.94	5.91	6.03	5.98	-0.47	-0.37	-0.35	-0.53

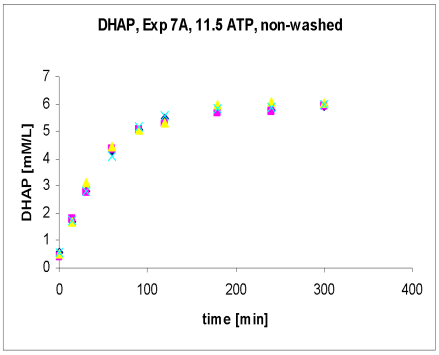
Table B.23: Experiment 7A, nonwashed extract

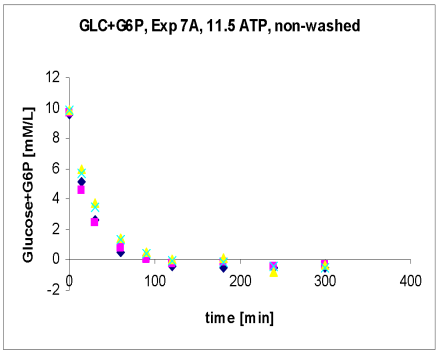
Time	DHAP				GL			
0	0.49	0.38	0.45	0.4	8.73	8.94	9.25	8.76
15	1.15	1.19	1.26	1.13	5.51	5.45	5.67	5.71
30	2.07	2.02	2.13	2.05	2.96	3.05	3.66	3.34
60	3.22	3.22	3.3	3.21	1.3	1.02	1.09	1.12
90	4.14	4	4.24	4.12	0.31	0.07	0.25	0.17
120	4.63	4.68	4.57	4.44	0.02	-0.16	-0.32	-0.15
180	5.04	4.82	5.33	4.78	-0.16	-0.16	0.2	-0.19
240	5.25	5.12	5.49	5.05	-0.19	-0.33	-0.03	-0.3
300	5.24	5.42	5.56	5.17	-0.14	-0.22	-0.03	-0.21

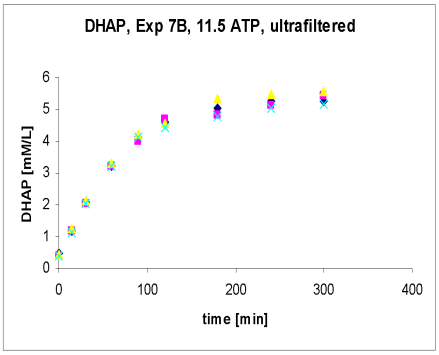
Table B.24: Experiment 7B, ultrafiltered extract

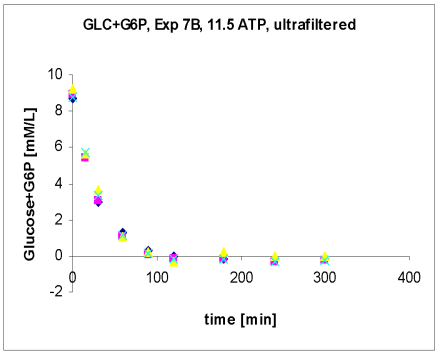
Time	DHAP		GL	
0	-0.08	0.01	8.62	7.98
15	0.52	0.48	5.79	7.93
30	1.17	1.16	4.23	4.07
60	2.08	1.93	2.24	2.27
90	2.68	2.47	1.09	1.31
120	3	2.93	0.45	0.7
180	3.52	3.54	0.08	0.31
240	3.95	4.23	-0.41	-0.18
300	4.4	4.22	-0.12	-0.18

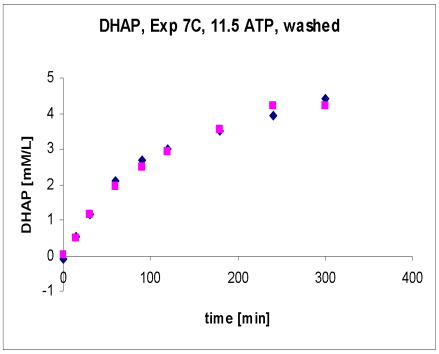
Table B.25: Experiment 7C, washed extract

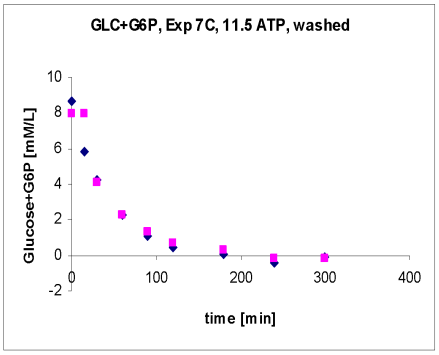

DHAP, Exp 7A, 11.5 ATP, non-washed
DHAP [mM/L]
time [min]


GLC+G6P, Exp 7A, 11.5 ATP, non-washed
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 7B, 11.5 ATP, ultrafiltered
DHAP [mM/L]
time [min]


GLC+G6P, Exp 7B, 11.5 ATP, ultrafiltered
Glucose+G6P [mM/L]
time [min]


DHAP, Exp 7C, 11.5 ATP, washed
DHAP [mM/L]
time [min]


GLC+G6P, Exp 7C, 11.5 ATP, washed
Glucose+G6P [mM/L]
time [min]

B.9 Experiment 8

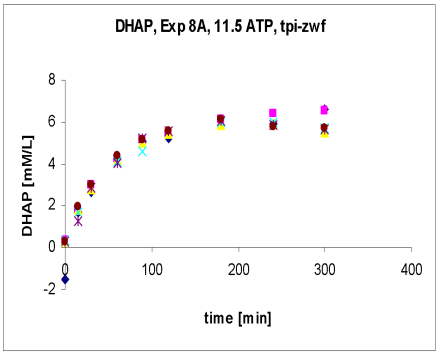
Before experiment 8 a new batch was executed with a double mutant LJ110-tpi-zwf and then used for the next SBT experiments.

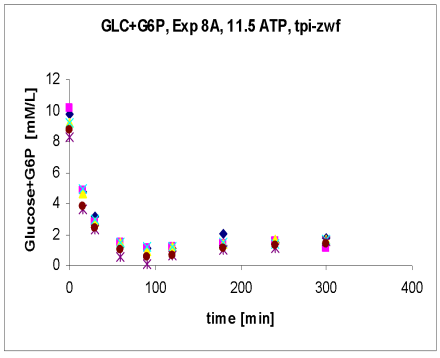
Comp Exp. No.	Exp. 8A
Hexokinase U/ml	0.06
Lactate-DH U/ml	1
Glc mM	11.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
Prot.tot. mg/ml	10
T	37
Strain	LJ110 tpi-zwf

Table B.26: Experiment 8 description

Time	DHAP						GL					
0	-1.52	0.38	0.32	0.37	0.32	0.31	9.74	10.15	9.2	9.18	8.28	8.75
15	1.67	1.8	1.84	1.71	1.27	1.94	4.72	4.82	4.7	4.94	3.59	3.81
30	2.65	3.03	2.81	2.83	2.85	2.98	3.2	2.76	2.75	2.94	2.28	2.4
60	4.22	4.28	4.15	4.1	4.07	4.36	1.24	1.47	1.46	1.48	0.53	1
90	5.09	5.17	5.02	4.62	5.2	5.18	1.11	1.1	0.94	1.17	0.11	0.55
120	5.23	5.42	5.42	5.6	5.57	5.6	1.11	1.18	1.22	1.21	0.65	0.66
180	5.81	6.14	5.83	5.98	6.08	6.11	2.02	1.44	1.39	1.44	1	1.15
240	5.93	6.4	5.91	5.96	5.87	5.76	1.5	1.57	1.68	1.44	1.11	1.31
300	6.64	6.52	5.49	5.68	5.64	5.69	1.74	1.16	1.69	1.64	1.57	1.36

Table B.27: Experiment 8A, double mutant





B.10 Experiment 9

In experiment 9, production of 5-deoxy-5-ethyl-xylulose-1-phosphate with the LJ110 tpi zwf was investigated.

Comp Exp. No.	**Exp 9A**	**Exp 9B**
Hexokinase U/ml	0.06	0.06
Lactate-DH U/ml	1	1
Glc mM	11.1	11.1
PO4 mM	11.1	11.1
ATP mM	11.5	11.5
NAD+ mM	5.75	5.75
Prot.tot. mg/ml	10	10
T	37	37
Strain	LJ110 tpi-zwf	LJ110 tpi-zwf

Table B.28: Experiment 9 description

LJ110 tpi zwf. Reactor system changed. Stirred double-walled beakers. Added Aldolase reaction. Production phase starting at 182 min, butanal added

Time	DHAP		GL	
0	0.49	0.39	7.81	0.66
5	0.76	0.81	7.46	7.68
10	1.23	1.29	5.28	4.61
15	1.75	1.83	4.34	4.36
30	2.4	3.09	2.64	2.76
60	4.33	4.57	1.86	2.07
90	4.96	5.31	1.59	1.68
120	5.71	5.84	1.58	1.02
180	6.51	6.88	1.35	1.36
182	6.25	6.44	1.86	1.89
185	6.06	6.24	1.79	1.69
190	6.04	6.41	1.43	1.61
195	6.11	6.12	1.6	1.71
210	5.83	6.18	1.62	1.93
240	5.38	6.03	1.62	1.62
270	5.2	5.73	1.6	1.77
305	5.07	5.34	1.7	1.48

Table B.29: Experiment 9A, double mutant

The production of 5-deoxy-5-ethyl-xylulose-1-phosphate was initiated by adding an extra reactant, butanal in large amount [10mM/ml], which was supposed to consume the DHAP present in the reactor. The performance of the reaction was not so good. The possible causes are:

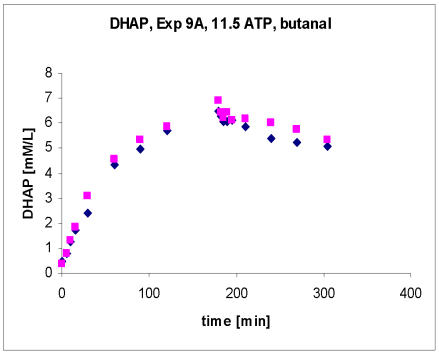
- somehow the pool of the aldolase presents in the extract, got saturated or blocked by one of the reactants, the reaction producing DHAP run at an acceptable rate;
- the pool of enzyme was just to small to cope with the large amount of butanal
- butanal, being a non natural reactant made the reaction to be less efficient than the one with G3P (cleavage of F16B to DHAP and G3P)

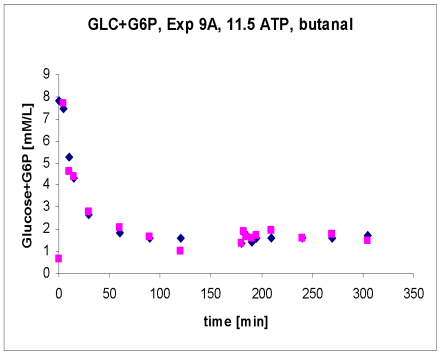
LJ110 tpi zwf. Reactor system changed. Stirred double-walled beakers. Added Aldolase reaction. Production phase starting at 182 min, butanal and RAMA added.

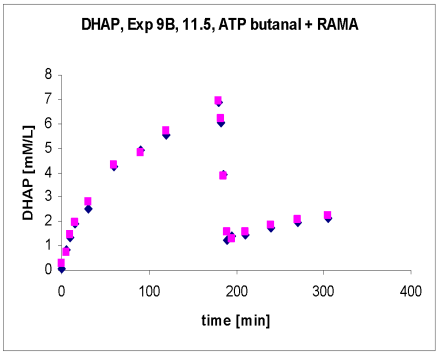
Time	DHAP		GL	
0	0.06	0.29	8.99	8.56
5	0.81	0.7	7.15	6.91
10	1.32	1.46	5.4	5.52
15	1.91	1.95	4.6	4.63
30	2.52	2.77	3.25	3.34
60	4.27	4.3	1.99	2.05
90	4.95	4.8	1.87	2.06
120	5.56	5.73	1.64	1.85
180	6.87	6.96	1.61	1.67
182	6.04	6.2	1.61	1.6
185	3.93	3.84	1.74	1.81
190	1.21	1.57	1.4	1.45
195	1.41	1.31	1.19	0.96
210	1.47	1.54	0.72	0.52
240	1.72	1.82	0.23	0.16
270	1.97	2.08	0.3	0.2
305	2.1	2.25	0.21	0.31

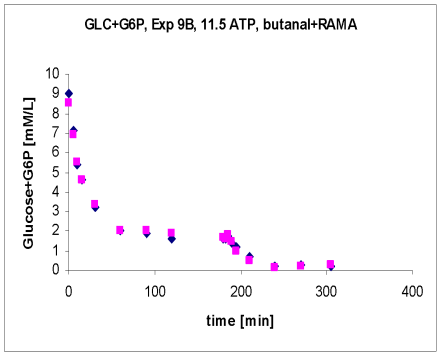
Table B.30: Experiment 9B, double mutant

A second experiment was performed, when besides butanal, extra amount of aldolase enzyme was added (RAMA- rabbit muscle aldolase, type I of aldolase). The production phase was started at 182 min.


DHAP, Exp 9A, 11.5 ATP, butanal
DHAP [mM/L]
time [min]
0
1
2
3
4
5
6
7
8
0
100
200
300
400


GLC+G6P, Exp 9A, 11.5 ATP, butanal
Glucose+G6P [mM/L]
time [min]
0
1
2
3
4
5
6
7
8
9
0
50
100
150
200
250
300
350


DHAP, Exp 9B, 11.5, ATP butanal + RAMA
DHAP [mM/L]
time [min]
0
1
2
3
4
5
6
7
8
0
100
200
300
400


GLC+G6P, Exp 9B, 11.5 ATP, butanal+RAMA
Glucose+G6P [mM/L]
time [min]
0
1
2
3
4
5
6
7
8
9
10
0
100
200
300
400

B.11 Experiment 10

Before experiment 10 a new LJ110 tpi batch was produced. In experiment 10 the new reactor system with stirred double-walled beakers was used. G6P was measured for the first time independently. The amount of hexokinase and the temperature were maintained constant at the previous values. The ATP concentration was varied.

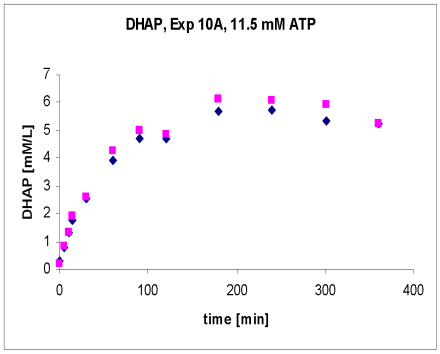
Experiments where G6P was measured also.

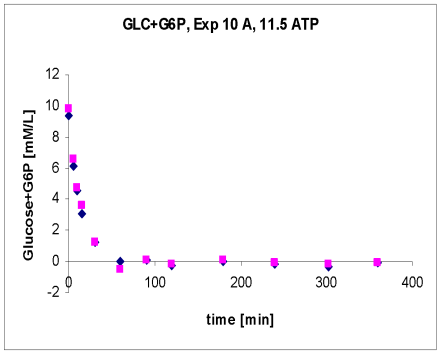
Comp Exp. No.	**Exp 10A**	**Exp 10B**
Hexokinase U/ml	0.06	0.06
Lactate-DH U/ml	1	1
Glc mM	11.1	11.1
PO4 mM	11.1	11.1
ATP mM	11.5	5.75
NAD+ mM	5.75	5.75
Prot.tot. mg/ml	10	10
T	37	37
Strain	LJ110 tpi	LJ110 tpi

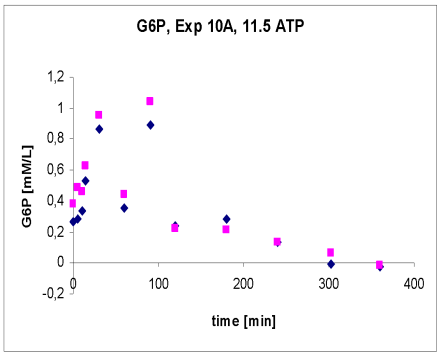
Table B.31: Experiment 10 description

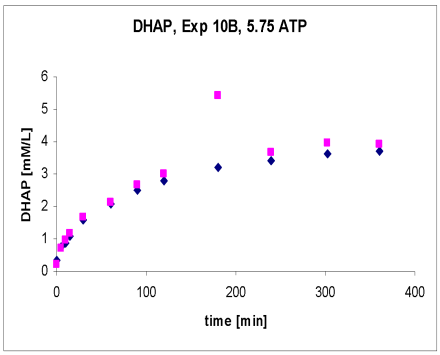
Time	DHAP		GL+G6P		G6P	
0	0.27	0.2	9.39	9.83	0.27	0.38
5	0.8	0.81	6.13	6.58	0.28	0.49
10	1.3	1.33	4.52	4.78	0.33	0.46
15	1.76	1.9	3.07	3.58	0.53	0.63
30	2.57	2.61	1.2	1.21	0.86	0.95
60	3.91	4.24	0.05	-0.51	0.36	0.45
90	4.7	4.99	0.07	0.08	0.89	1.04
120	4.72	4.86	-0.23	-0.15	0.24	0.22
180	5.69	6.12	0.05	0.13	0.28	0.21
240	5.72	6.07	-0.19	-0.1	0.13	0.13
302	5.32	5.9	-0.31	-0.13	-0.01	0.06
360	5.25	5.21	-0.08	-0.09	-0.02	-0.02

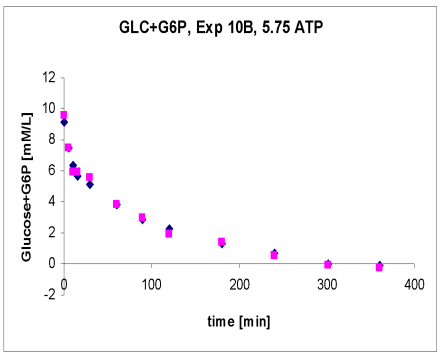
Table B.32: Experiment 10A, 11.5 mM ATP, Reactor I

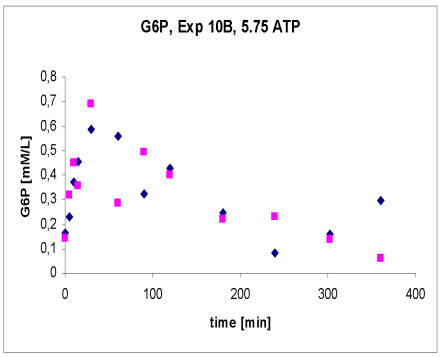

DHAP, Exp 10A, 11.5 mM ATP
DHAP [mM/L]
time [min]


GLC+G6P, Exp 10 A, 11.5 ATP
Glucose+G6P [mM/L]
time [min]


G6P, Exp 10A, 11.5 ATP
G6P [mM/L]
time [min]


DHAP, Exp 10B, 5.75 ATP
DHAP [mM/L]
time [min]


GLC+G6P, Exp 10B, 5.75 ATP
Glucose+G6P [mM/L]
time [min]


G6P, Exp 10B, 5.75 ATP
G6P [mM/L]
time [min]

Time	DHAP		GL+G6P		G6P	
0	0.32	0.22	9.09	9.6	0.17	0.14
5	0.73	0.71	7.45	7.49	0.23	0.32
10	0.88	0.97	6.32	5.92	0.37	0.45
15	1.1	1.15	5.64	5.95	0.45	0.36
30	1.57	1.67	5.16	5.57	0.59	0.69
60	2.09	2.12	3.81	3.81	0.56	0.29
90	2.51	2.66	2.9	2.99	0.32	0.5
120	2.79	2.99	2.27	1.88	0.43	0.4
180	3.22	5.42	1.28	1.37	0.25	0.22
240	3.42	3.66	0.66	0.56	0.08	0.23
302	3.6	3.95	-0.01	-0.08	0.16	0.14
360	3.7	3.91	-0.08	-0.28	0.29	0.06

Table B.33: Experiment 10B, 5.75 mM ATP, Reactor II

B.12 Experiment 11

In experiment 11 the amount of hexokinase was further reduced from 0.06 [U/ml] to 0.01 [U/ml] and the ATP concentration was varied again.

Comp Exp. No.	**Exp 11A**	**Exp 11B**
Hexokinase U/ml	0.01	0.01
Lactate-DH U/ml	1	1
Glc mM	11.1	11.1
PO4 mM	11.1	11.1
ATP mM	11.5	5.75
NAD+ mM	5.75	5.75
Prot.tot. mg/ml	10	10
T	37	37
Strain	LJ110 tpi	LJ110 tpi

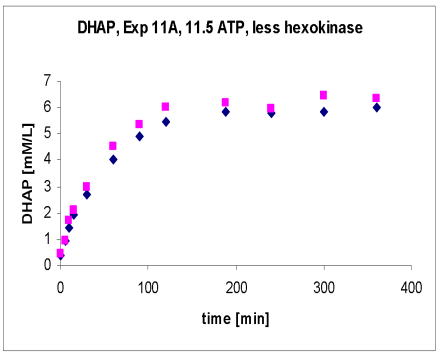
Table B.34: Experiment 11 description

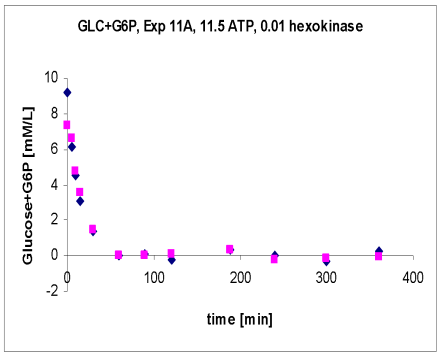
Time	DHAP		GL+G6P		G6P		ATP	
0	0.41	0.46	9.22	7.37	0.29	0.2	6.95	7.25
5	0.92	0.95	6.17	6.65	0.33	0.4	2.88	3.36
10	1.44	1.72	4.55	4.77	0.5	0.57	1.36	1.32
15	1.93	2.09	3.11	3.59	0.5	0.59	0.73	0.85
30	2.71	2.97	1.37	1.46	0.58	0.76	0.28	0.28
60	4.04	4.54	-0.02	-0.02	0.45	0.8	0.23	0.29
90	4.92	5.34	0.06	-0.01	0.09	0.32	0.22	0.38
120	5.46	6	-0.19	0.05	0.33	0.25	0.16	0.1
188	5.84	6.16	0.38	0.34	0.32	0.14	0.1	0.09
240	5.77	5.93	0	-0.25	0.18	0.05	0.06	0.06
300	5.84	6.46	-0.31	-0.15	0.14	0.03	0.05	0.05
361	5.99	6.36	0.28	-0.03	-0.01	0.04	-0.01	0.03

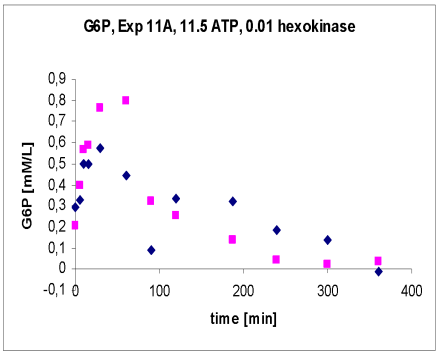
Table B.35: Experiment 11A, 11.5 mM ATP, reactor I

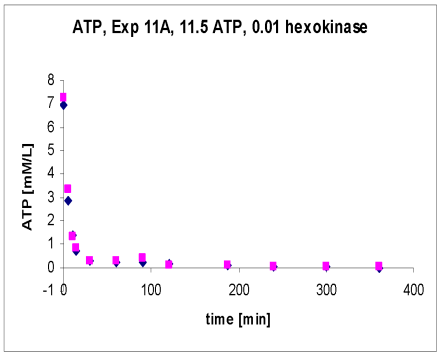
Time	DHAP		GL+G6P		G6P		ATP	
0	0.38	0.38	9.07	10.57	0.27	0.22	3.01	3.36
5	0.7	0.79	7.3	8.02	0.39	0.35	0.53	0.59
10	0.95	1.02	6.62	6.95	0.44	0.46	0.2	0.19
15	1.04	1.03	5.6	6.24	0.31	-0.25	0.11	0.18
30	1.65	1.83	4.93	5.47	0.38	0.37	0.04	0.13
60	2.3	2.59	3.68	4.19	0.32	0.47	0.04	0.07
90	2.84	3.05	3.01	3.06	0.54	0.44	0.05	-0.02
120	3.03	3.34	2.32	2.9	0.75	0.96	0.09	0.01
188	3.2	3.34	1.01	1.1	0.29	0.2	0.06	0.02
240	3.56	3.92	0.39	0.4	0.08	0.22	0.03	0
300	3.89	4.24	-0.06	0.03	0.23	0.2	0.02	0.02
361	3.88	4.16	1.02	0.17	0.13	0.02	0.03	0.06

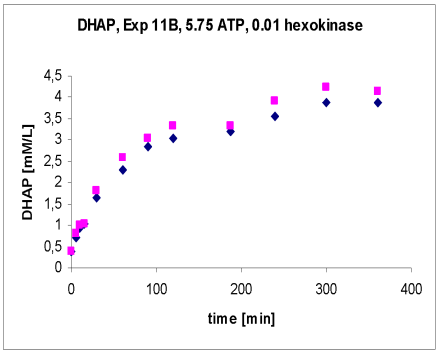
Table B.36: Experiment 11B, 5.75 mM ATP, reactor II

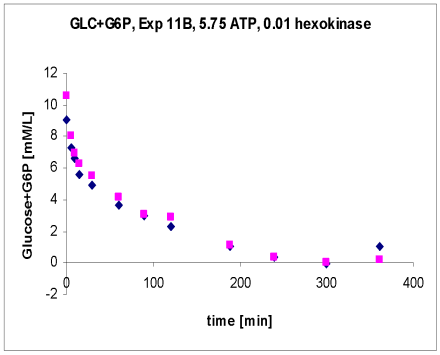


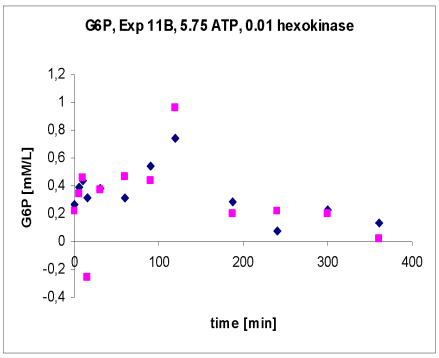


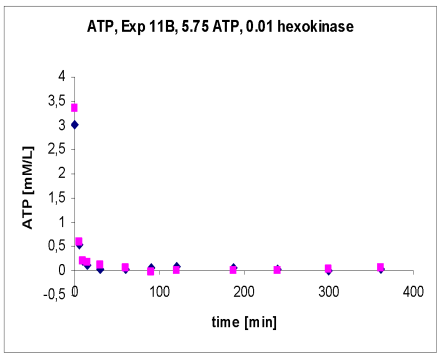





DHAP, Exp 11B, 5.75 ATP, 0.01 hexokinase
DHAP [mM/L]
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
0
100
200
300
400
time [min]


GLC+G6P, Exp 11B, 5.75 ATP, 0.01 hexokinase
Glucose+G6P [mM/L]
12
10
8
6
4
2
0
-2
0
100
200
300
400
time [min]


G6P, Exp 11B, 5.75 ATP, 0.01 hexokinase
G6P [mM/L]
1,2
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
0
100
200
300
400
time [min]


ATP, Exp 11B, 5.75 ATP, 0.01 hexokinase
ATP [mM/L]
4
3,5
3
2,5
2
1,5
1
0,5
0
-0,5
0
100
200
300
400
time [min]

B.13 Experiment 12

For experiment 12 the hexokinase concentration was kept constant at 0.01 [U/ml]. In this experiment the Glucose initial concentration was doubled (from 11.1 to 22.2 [mM/ml]). Three types of experiments have been performed.

- an experiment with unwashed extract, like in the previous experiments
- an experiment with the extracted prepared as follows. Crude extract was filtered with an ultra-filter. The concentrate was diluted to original volume with buffer and the filtrate was discarded. The procedure was repeated three times.
- an experiment in which the crude extract was filtered and then the concentrate was mixed again with filtrate

LJ110 tpi. Stirred double-walled beakers.Reactor IV, filtered I

Comp Exp. No.	**Exp 12A**	**Exp 12B**	**Exp 12C**
Hexokinase U/ml	0.01	0.01	0.01
Lactate-DH U/ml	1	1	1
Glc mM	22.2	22.2	22.2
PO4 mM	11.1	11.1	11.1
ATP mM	11.5	11.5	11.5
NAD+ mM	5.75	5.75	5.75
Prot.tot. mg/ml	10	10	10
T	37	37	37
Strain	LJ110 tpi	LJ110 tpi	LJ110 tpi

Table B.37: Experiment 12 description

Time	DHAP		GL+G6P		G6P		ATP	
0	0.44	0.33	17.13	18	0.58	0.53	6.78	7.01
5	0.77	0.88	16.02	16.85	0.25	0.16	2.79	2.92
10	1.17	1.03	14.1	14.71	0.59	0.38	1.85	2.16
15	1.47	1.56	14.49	15.64	0.57	0.49	1.3	1.48
30	2.07	2.32	12.02	13.1	0.92	0.83	0.48	0.53
60	3.48	3.64	11.28	11.89	1.16	1.13	0.08	0.09
90	4.11	4.31	10.81	11.57	1.25	1.18	0.1	0.07
120	4.67	4.7	9.13	9.76	1.07	0.96	0.1	0.13
180	5.16	5.36	9.96	10.72	1.28	1.4	0.05	1.25
259	5.91	6.06	9.07	10.23	1	1.01	0.12	0.87
300	6.25	6.87	8.28	8.7	1.17	1.24	0.19	1.1
360	5.63	5.91	6.55	6.85	-0.01	-0.02	0.01	-0.13

Table B.38: Experiment 12A, unwashed extract, Reactor I

Time	DHAP		GL+G6P		G6P		ATP	
0	-0.05	-0.05	16.22	17.66	0.38	0.47	7.11	7.88
5	0.21	0.24	17.69	18.99	0.42	0.69	5.59	6.39
10	0.6	0.57	15.78	16.22	0.61	0.63	3.91	4.03
15	0.89	0.96	15.26	15.68	0.67	0.71	3.24	3.49
30	1.54	1.62	12.74	13.38	0.94	0.99	1.55	1.65
60	2.73	2.83	11.71	12.59	1.54	1.66	0.48	0.57
90	3.41	3.6	10.04	10.37	1.63	1.76	0.23	0.29
120	4.16	4.34	9.06	9.41	1	1.04	0.27	0.28
180	4.66	5.08	7.18	7.82	0.73	0.76	0.28	0.33
259	5.75	6.08	6.51	7.07	0.78	0.79	0.25	0.3
300	5.41	5.61	5.77	6.69	0.73	0.77	0.28	0.26
360	6.17	6.26	5.24	5.01	0.63	0.67	0.2	0.21

Table B.39: Experiment 12B, washed extract I, Reactor II

Time	DHAP		GL+G6P		G6P		ATP	
0	-0.01	-0.09	17.71	18.32	0.26	0.3	6.7	7.01
5	0.15	0.16	16.91	18.66	0.4	0.41	5.14	5.38
10	0.58	0.71	16.98	17.59	0.54	0.58	4.61	4.94
15	0.95	0.91	15.71	15.71	0.53	0.57	2.92	3.13
30	1.7	1.73	13.65	14.41	0.75	0.8	1.63	1.75
60	2.76	2.71	11.42	12.28	1.17	1.28	0.7	0.47
90	3.3	3.45	10.09	10.81	1.35	1.44	0.24	0.22
120	3.9	4.14	8.8	9.15	0.69	0.72	0.23	0.29
180	5.08	5.46	8.95	9	0.37	0.82	0.34	0.36
259	5.72	6.18	7.04	7.94	0.3	0.64	0.31	0.28
300	5.76	6.38	6.64	7.26	0.27	0.63	0.25	0.28
360	5.97	6.3	6.57	6.54	0.55	0.61	0.21	0.19

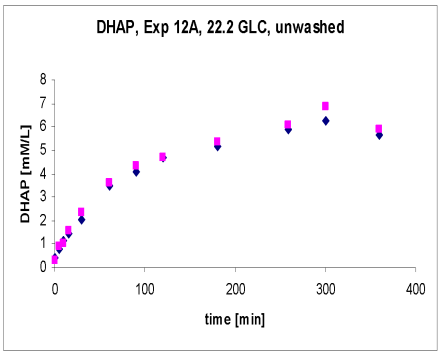
Table B.40: Experiment 12A, washed extract II ,Reactor III

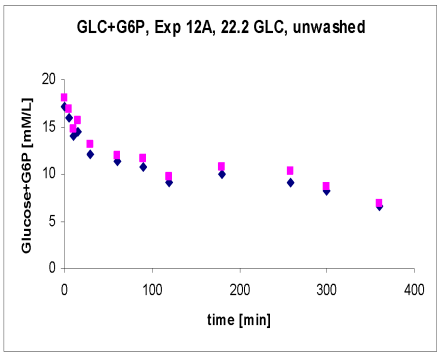
Time	DHAP		GL+G6P		G6P		ATP	
0	0.34	0.28	17.01	17.28	0.11	0.13	6.66	7.07
5	0.51	0.58	16.67	16.79	0.14	0.14	4.74	5.15
10	0.95	0.98	15.08	15.73	0.23	0.24	3.06	3.23
15	1.27	1.43	15.15	15.35	0.3	0.3	2.43	2.63
30	2.11	2.22	13.7	14.75	0.55	0.58	0.99	1.21
60	3.44	3.39	11.08	12.46	0.96	1.05	0.2	0.27
90	4.08	4.26	10.22	11.61	1.07	1.15	0.19	0.14
120	12.21	12.35	26.53	25.49	1.5	1.43	1.46	1.54
180	4.89	5.39	8.33	8.8	0.8	0.89	0.14	0.15
259	5.92	6.06	7.71	8.62	0.82	0.9	0.13	0.16
300	6.01	6.62	7.01	7.41	0.77	0.84	0.11	0.14
360	6.11	6.6	6.72	7.3	0.82	0.85	0.1	0.12

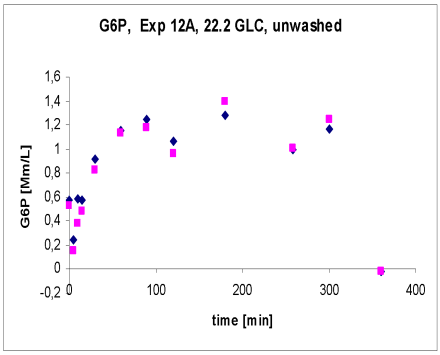
Table B.41: Experiment 12C, filtered extract I, Reactor IV

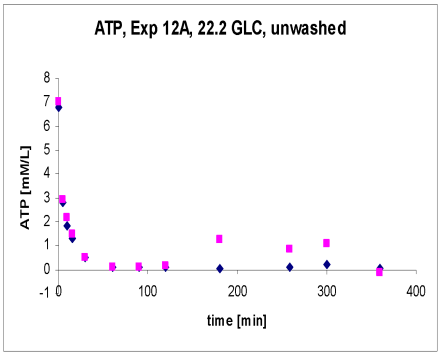
Time	DHAP		GL+G6P		G6P		ATP	
0	0.35	0.33	20.51	21.4	0.09	0.09	6.42	6.68
5	0.64	0.61	19.24	19.62	0.06	0.07	4.34	4.63
10	1	1.09	18.68	19.66	0.26	0.37	2.72	2.89
15	1.36	1.38	17.7	19.02	0.25	0.29	2.08	2.27
30	2.21	2.32	16.54	17.01	0.59	0.64	0.78	0.83
60	3.4	3.63	14.98	15.65	0.91	0.99	0.26	0.27
90	4.22	4.32	13.95	15.19	0.72	1.12	0.14	0.16
120	12.8	11.57	32.86	33.97	1.35	1.31	1.24	1.26
180	5.31	5.53	11.64	11.25	0.7	0.81	0.21	0.2
259	6.33	6.8	11.04	12.33	0.76	0.85	0.16	0.14
300	6.64	7.41	11.9	11.69	0.78	0.88	0.19	0.21
360	6.74	9.02	12.59	12.68	0.87	0.96	0.23	0.24

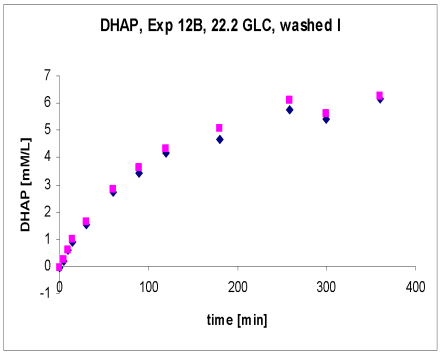
Table B.42: Experiment 12C, filtered extract II, Reactor V

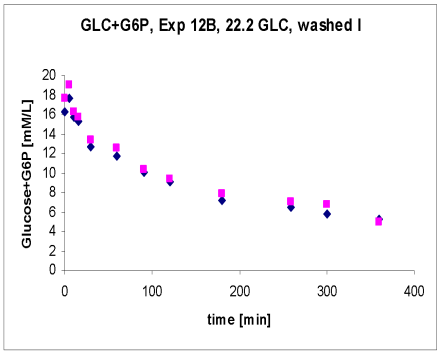


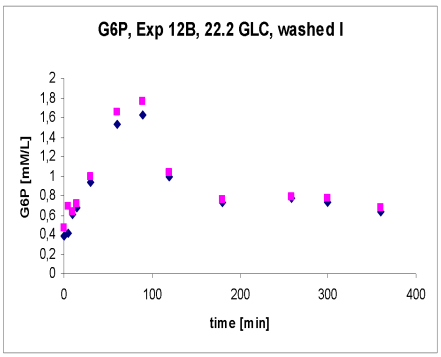


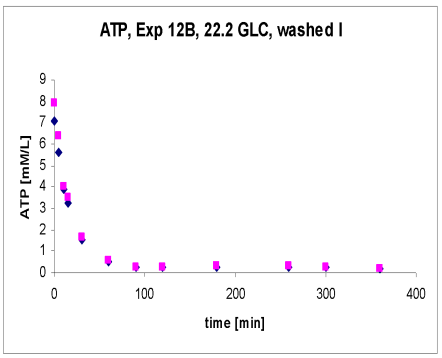


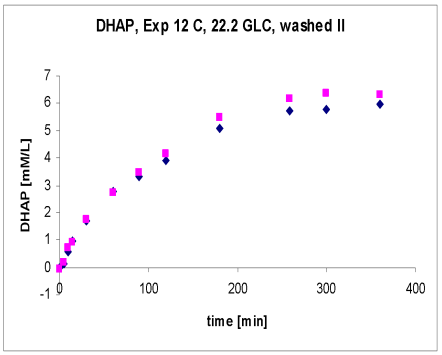


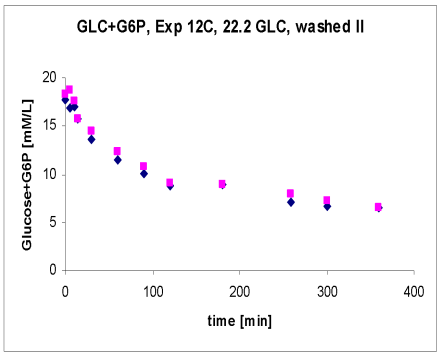

DHAP, Exp 12B, 22.2 GLC, washed I
DHAP [mM/L]
time [min]

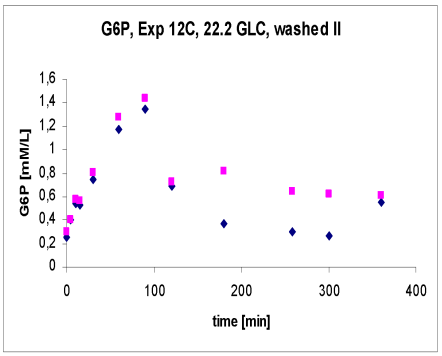

GLC+G6P, Exp 12B, 22.2 GLC, washed I
Glucose+G6P [mM/L]
time [min]

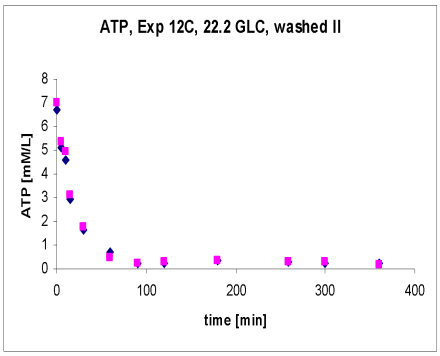

G6P, Exp 12B, 22.2 GLC, washed I
G6P [mM/L]
time [min]

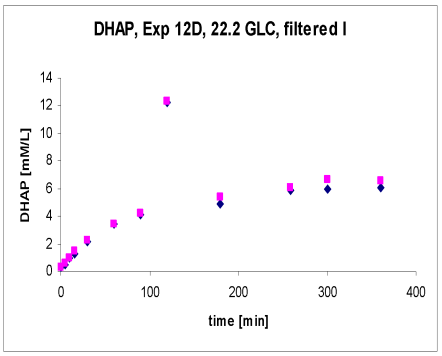

ATP, Exp 12B, 22.2 GLC, washed I
ATP [mM/L]
time [min]

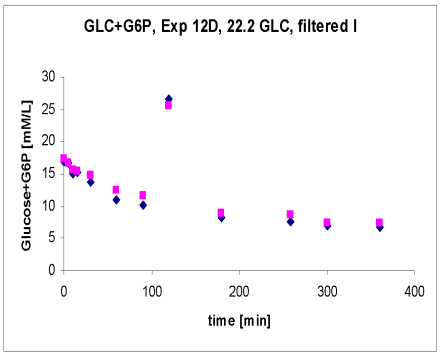

DHAP, Exp 12 C, 22.2 GLC, washed II
DHAP [mM/L]
time [min]

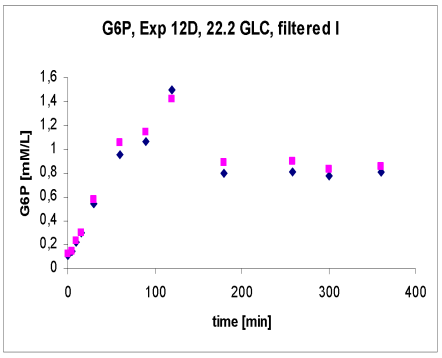

GLC+G6P, Exp 12C, 22.2 GLC, washed II
Glucose+G6P [mM/L]
time [min]

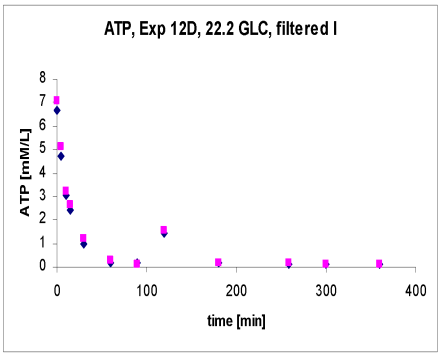

G6P, Exp 12C, 22.2 GLC, washed II
G6P [mM/L]
time [min]

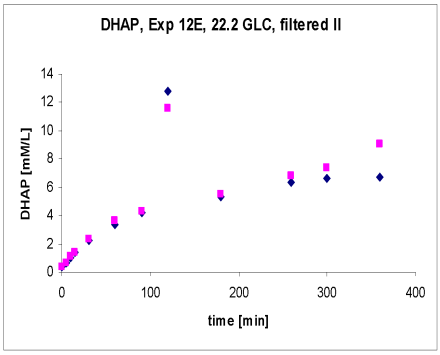

ATP, Exp 12C, 22.2 GLC, washed II
ATP [mM/L]
time [min]

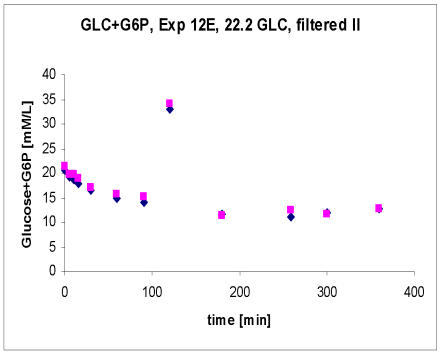

DHAP, Exp 12D, 22.2 GLC, filtered I
DHAP [mM/L]
time [min]

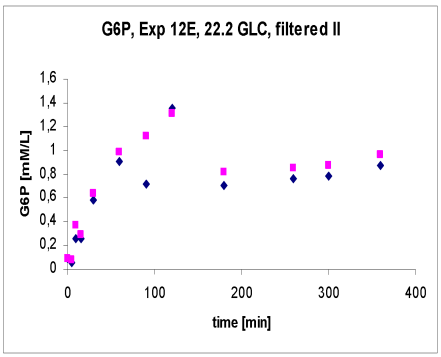

GLC+G6P, Exp 12D, 22.2 GLC, filtered I
Glucose+G6P [mM/L]
time [min]

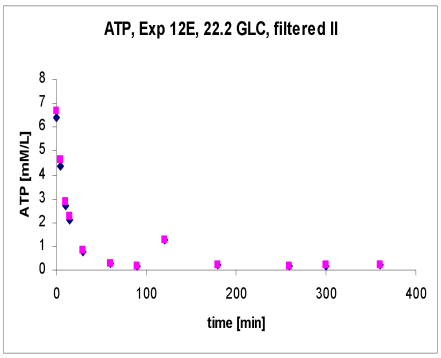

G6P, Exp 12D, 22.2 GLC, filtered I
G6P [mM/L]
time [min]


ATP, Exp 12D, 22.2 GLC, filtered I
ATP [mM/L]
time [min]


DHAP, Exp 12E, 22.2 GLC, filtered II
DHAP [mM/L]
time [min]


GLC+G6P, Exp 12E, 22.2 GLC, filtered II
Glucose+G6P [mM/L]
time [min]


G6P, Exp 12E, 22.2 GLC, filtered II
G6P [mM/L]
time [min]


ATP, Exp 12E, 22.2 GLC, filtered II
ATP [mM/L]
time [min]

B.14 Experiment 17

A new experiment, exp 17, was performed after ATP sink was eliminated in the way described and the initial concentration of ATP was varied. Five reactors (stirred, double-walled beakers) were used. One reactor for the experiment with an initial concentration of ATP of 11.5 [mM/ml], two reactors (II and III) with an initial concentration of 5.75 [mM/ml] and two reactors (IV and V) with 1.15 [mM/ml]. First time when measuring ADP and AMP. AMP measurements are not reliable. Time delays between ATP and glucose additions and first sampling: ATP 15 - 20 s; glucose 10 s.

Comp Exp. No.	**Exp 17A**	**Exp 17BC**	**Exp 17DE**
Hexokinase U/ml	0.01	0.01	0.01
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11.1	11.1
ATP mM	11.5	5.75	1.15
NAD+ mM	5.75	5.75	5.75
Prot.tot. mg/ml	37	37	37
Strain	LJ110 tpi	LJ110 tpi	LJ110 tpi

Table B.43: Experiment 17 description

Time	DHAP		G6P+Gl		G6P		ATP		ADP	
0	0.56	0.53	7.79	7.91	0.18	0.24	6.51	6.27	4.51	0.11
5	1.02	1.07	6.8	7.21	0.57	0.6	3.27	3.6	5.18	-0.03
10	1.45	1.63	4.88	4.94	0.61	0.62	1.76	2.14	4.51	4.49
15	2	2.12	3.96	3.99	0.82	0.81	1.25	1.37	4.18	4.22
30	2.99	3.31	2.15	2.07	1.26	1.32	0.67	0.52	2.87	2.93
60	4.62	4.95	0.44	0.03	1.08	1.2	0.32	0.57	2.04	1.88
90	5.55	6.06	-0.15	-0.24	0.74	0.78	0.49	0.41	1.86	2.16
120	6.42	6.88	-0.22	-0.19	0.57	0.74	0.05	0.31	1.88	2
180	7.02	6.99	0.11	0.19	0.49	0.52	0.45	0.19	2.11	2.16
259	7.38	7.64	0.16	0.25	0.37	0.44	0.3	0.26	1.87	1.92
300	7.82	8.06	-0.3	0.11	0.18	0.19	0.16	-0.02	1.63	1.59
360	8.17	8.35	-0.24	-0.4	0.19	0.17	0.43	0.03	1.58	1.57

Table B.44: Experiment 17A, Stirred double-walled beakers. Reactor I, 11.5 mM ATP

Time	DHAP		G6P+Gl		G6P		ATP		ADP	
0	0.5	0.52	7.68	7.69	0.15	0.18	2.16	2.49	3.5	3.43
5	0.78	0.85	7.09	7.5	0.43	0.49	0.91	1.14	2.76	2.96
10	1.35	1.22	7.22	6.48	0.39	0.45	0.84	0.84	2.44	2.85
15	1.5	1.76	5.72	5.87	0.74	0.81	0.44	0.22	1.64	1.8
30	2.19	2.39	4.62	4.84	0.88	0.94	0.33	0.32	0.94	1.31
60	3.18	3.3	3.86	4.06	0.93	0.93	0.22	0.36	0.73	0.78
90	4.09	4.42	3.45	3.86	0.92	0.96	0.31	0.1	1.06	1.22
120	5.01	5.48	2.12	2.16	0.63	0.59	0.05	-0.24	0.94	1.14
180	5.45	5.6	1.35	1.32	0.64	0.62	-0.02	-0.07	0.96	1.03
259	6.74	7.01	0.33	0.33	0.53	0.52	0.49	-0.11	0.96	0.98
300	7.2	7.39	-0.13	-0.23	0.25	0.3	0.04	0	1.21	1.18
360	7.26	7.45	-0.16	-0.04	0.17	0.16	0.08	-0.08	1.37	1.34

Table B.45: Experiment 17B. Stirred double-walled beakers. Reactor II, 5.75 mM ATP

Time	DHAP		G6P+Gl		G6P		ATP		ADP	
0	0.51	0.53	7.02	6.77	0.13	0.11	1.19	1.52	3.5	5.77
5	0.74	0.85	6.31	6.99	0.35	0.26	-0.44	0.05	3.21	3.17
10	1.13	1.16	5.71	5.97	0.53	0.47	-0.6	-0.96	2.42	2.6
15	1.24	1.5	4.94	5.31	0.58	0.6	-0.75	-0.68	2.05	2.12
30	2.03	2.27	4.43	6.72	0.81	0.83	-0.94	-0.83	1.43	1.38
60	2.89	3.26	3.51	3.45	0.83	0.89	-0.98	-1.02	1.16	1.06
90	3.98	4.25	2.66	2.6	0.71	0.76	-0.91	-0.97	1.06	0.88
120	4.42	4.75	1.44	1.58	0.58	0.61	-1.05	-0.9	1.12	1.09
180	4.99	5.02	1.08	0.94	0.58	0.7	-0.35	-0.13	1.03	1.15
259	5.8	6.27	-0.08	-0.37	0.29	0.34	-0.05	-0.09	1.02	1.08
300	6.18	6.77	-0.35	-0.6	0.13	0.13	-0.22	-0.32	1.19	1.24
360	6.64	7.26	-0.79	-0.63	0.09	0.07	-0.22	-0.19	1.14	1.2

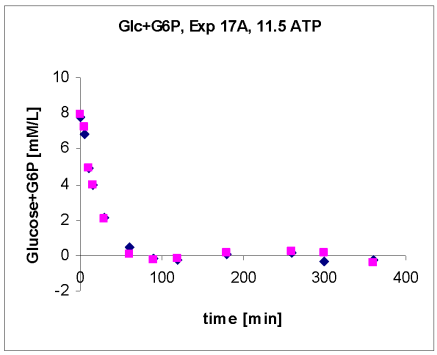
Table B.46: Experiment 17B. Stirred double-walled beakers. Reactor III, 5.75 mM ATP

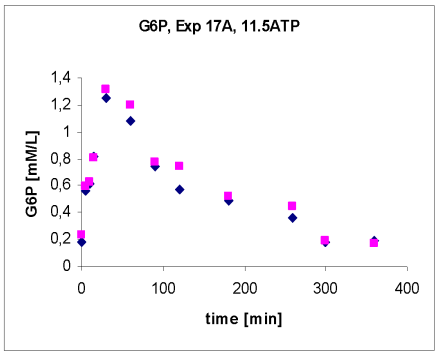
Time	DHAP		G6P+Gl		G6P		ATP		ADP	
0	0.32	0.54	9.56	9.76	0.06	0.07	-0.57	-0.33	0.24	1.56
5	0.62	0.72	8.28	8.41	0.21	0.2	-0.76	-0.53	0.86	0.92
10	0.78	0.9	7.76	7.92	0.25	0.27	-1.18	-0.77	0.76	0.77
15	0.91	1.1	7.63	7.73	0.34	0.35	-1.03	-0.7	0.66	0.67
30	1.36	1.58	6.92	7.27	0.36	0.39	-1.08	-0.82	0.58	0.63
60	1.91	1.97	6.35	6.9	0.43	0.48	-1.05	-0.81	0.62	0.72
90	2.35	2.44	6.29	6.75	0.48	0.43	-1.2	-0.72	0.52	0.71
120	2.94	3.22	5.4	6.22	0.36	0.33	-1.03	-0.71	0.65	0.82
180	3.25	3.34	4.74	5.01	0.37	0.38	-0.21	-0.35	0.71	0.74
259	4.33	4.31	3.12	3.18	0.29	0.34	-0.4	-0.38	0.73	0.74
300	5.07	5.55	1.77	1.84	0.27	0.3	-0.35	-0.48	0.71	0.72
360	6.15	5.84	0.61	0.75	0.27	0.29	-0.43	-0.43	0.78	0.81

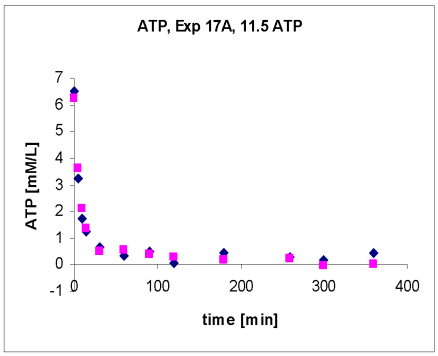
Table B.47: Experiment 17C. Stirred double-walled beakers.Reactor IV, 1.15 mM ATP

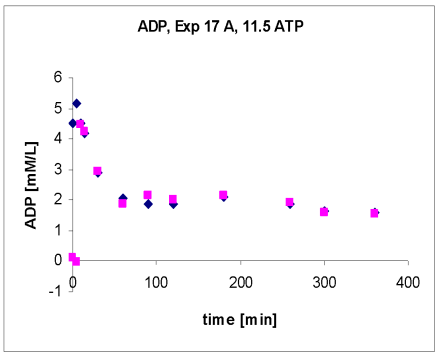
Time	DHAP		G6P+Gl		G6P		ATP		ADP	
0	0.51	0.55	8.48	9.24	0.06	0.07	0.26	0.38	1.59	1.69
5	0.62	0.99	7.8	8.06	0.23	0.25	0.29	-0.11	1.11	1.03
10	0.8	0.96	7.87	8.11	0.25	0.28	0.34	0	0.89	0.79
15	1.03	1.02	7.85	8.16	0.32	0.33	0.36	0.08	0.87	0.78
30	1.31	1.43	6.86	7.26	0.32	0.38	0.24	0.1	0.85	0.63
60	1.84	1.97	6.54	6.69	0.43	0.45	0.13	0.01	0.75	0.66
90	2.32	2.42	6.42	6.5	0.42	0.45	0.01	-0.03	0.88	0.75
120	2.95	3.25	5.71	5.75	0.4	0.4	-0.07	-0.05	0.94	0.81
180	3.52	3.52	4.04	3.9	0.28	0.3	0.06	0.3	0.8	0.79
259	4.17	4.64	2.39	2.75	0.25	0.29	0.08	-0.42	0.71	0.76
300	5.39	5.41	1.69	1.84	0.3	0.34	0.07	0.08	0.82	0.79
360	5.99	0.13	0.54	0.95	0.3	0.32	0.05	0.05	0.86	0.88

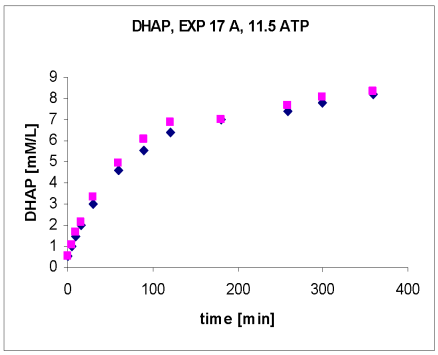
Table B.48: Experiment 17C. Stirred double-walled beakers. Reactor V, 1.15 mM ATP

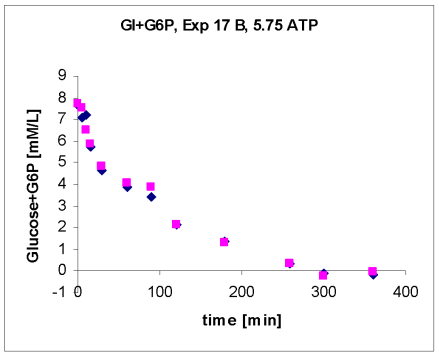

Glc+G6P, Exp 17A, 11.5 ATP
Glucose+G6P [mM/L]
time [min]

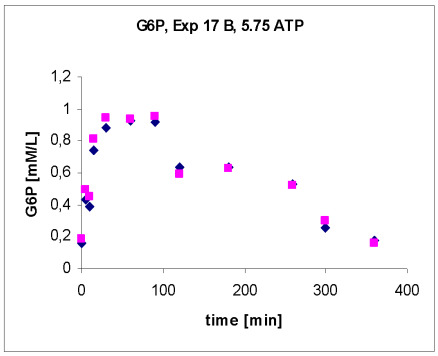

G6P, Exp 17A, 11.5ATP
G6P [mM/L]
time [min]

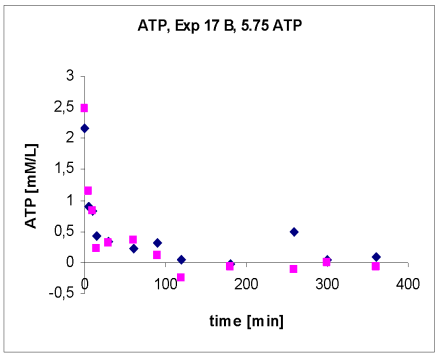

ATP, Exp 17A, 11.5 ATP
ATP [mM/L]
time [min]

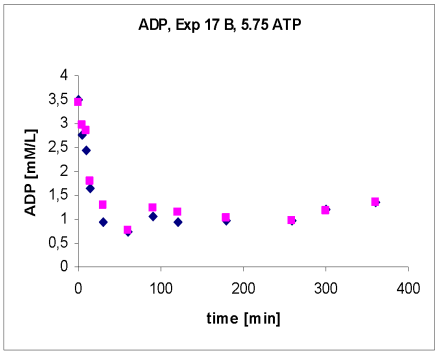

ADP, Exp 17 A, 11.5 ATP
ADP [mM/L]
time [min]

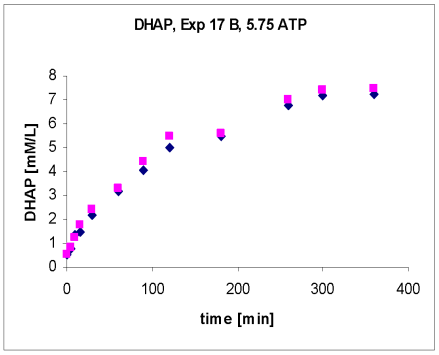

DHAP, EXP 17 A, 11.5 ATP
DHAP [mM/L]
time [min]

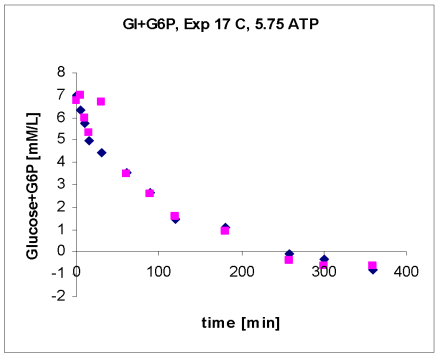

Gl+G6P, Exp 17 B, 5.75 ATP
Glucose+G6P [mM/L]
time [min]

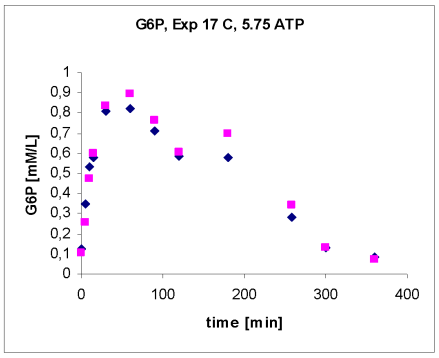

G6P, Exp 17 B, 5.75 ATP
G6P [mM/L]
time [min]

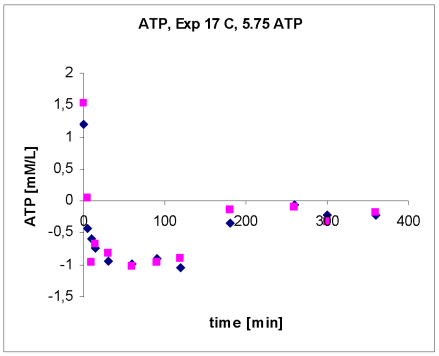

ATP, Exp 17 B, 5.75 ATP
ATP [mM/L]
time [min]

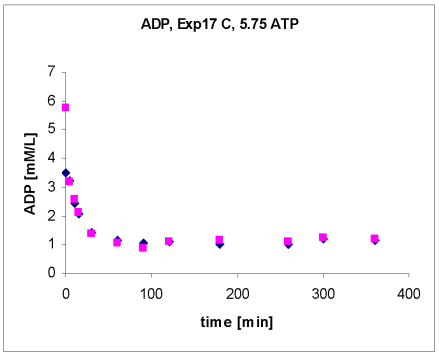

ADP, Exp 17 B, 5.75 ATP
ADP [mM/L]
time [min]

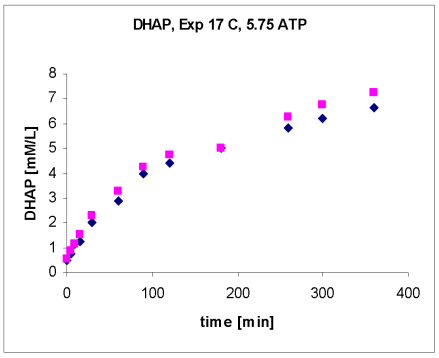

DHAP, Exp 17 B, 5.75 ATP
DHAP [mM/L]
time [min]

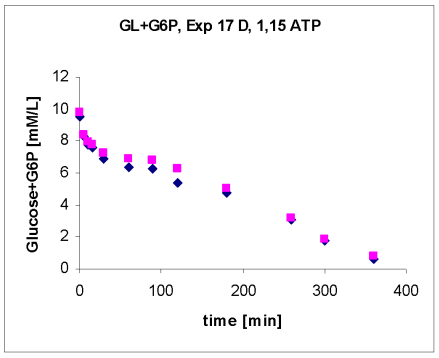

Gl+G6P, Exp 17 C, 5.75 ATP
Glucose+G6P [mM/L]
time [min]

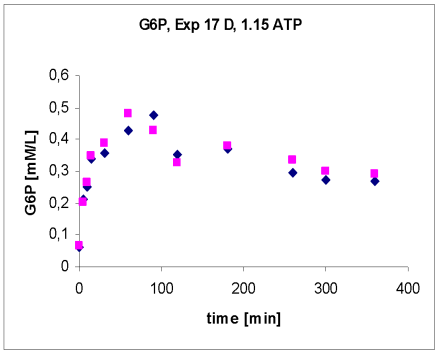

G6P, Exp 17 C, 5.75 ATP
G6P [mM/L]
time [min]

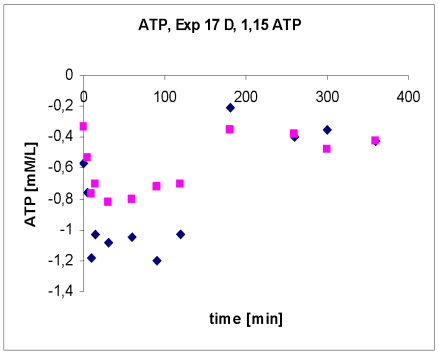

ATP, Exp 17 C, 5.75 ATP
ATP [mM/L]
time [min]

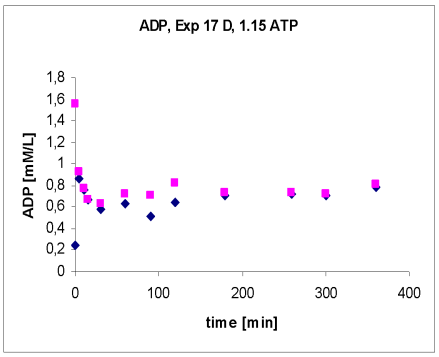

ADP, Exp17 C, 5.75 ATP
ADP [mM/L]
time [min]

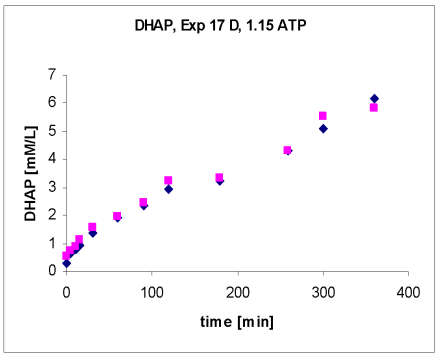

DHAP, Exp 17 C, 5.75 ATP
DHAP [mM/L]
time [min]

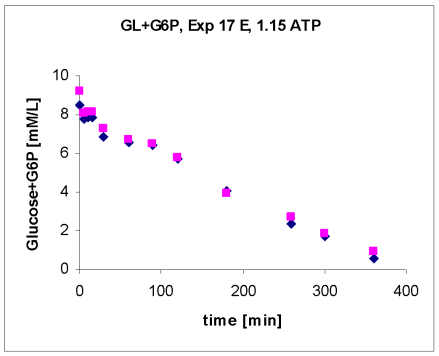

GL+G6P, Exp 17 D, 1,15 ATP
Glucose+G6P [mM/L]
time [min]

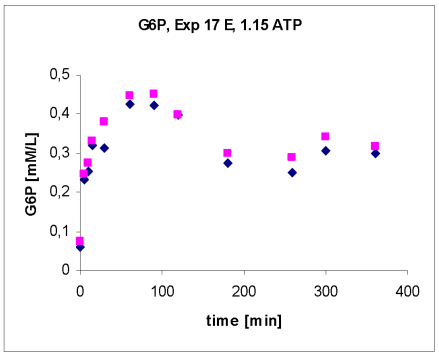

G6P, Exp 17 D, 1.15 ATP
G6P [mM/L]
time [min]

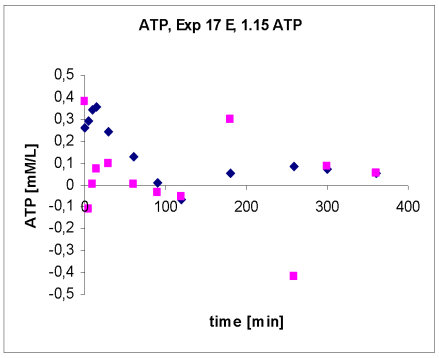

ATP, Exp 17 D, 1,15 ATP
ATP [mM/L]
time [min]

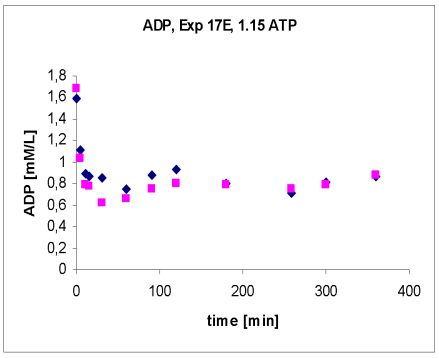

ADP, Exp 17 D, 1.15 ATP
ADP [mM/L]
time [min]

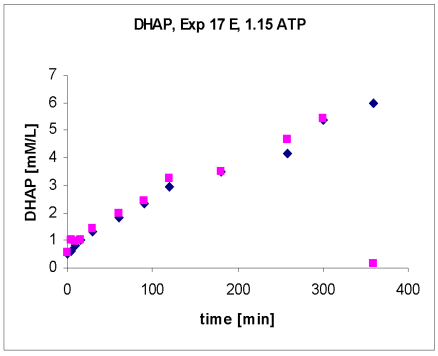

DHAP, Exp 17 D, 1.15 ATP
DHAP [mM/L]
time [min]


GL+G6P, Exp 17 E, 1.15 ATP
Glucose+G6P [mM/L]
time [min]


G6P, Exp 17 E, 1.15 ATP
G6P [mM/L]
time [min]


ATP, Exp 17 E, 1.15 ATP
ATP [mM/L]
time [min]


ADP, Exp 17E, 1.15 ATP
ADP [mM/L]
time [min]


DHAP, Exp 17 E, 1.15 ATP
DHAP [mM/L]
0
1
2
3
4
5
6
7
0
100
200
300
400
time [min]

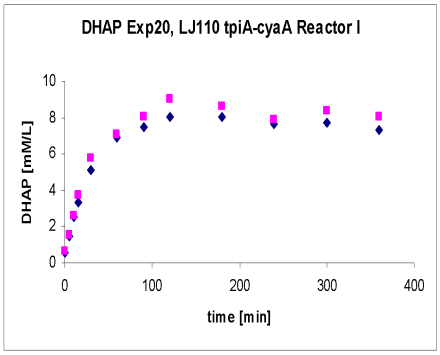
B.15 Experiment 20

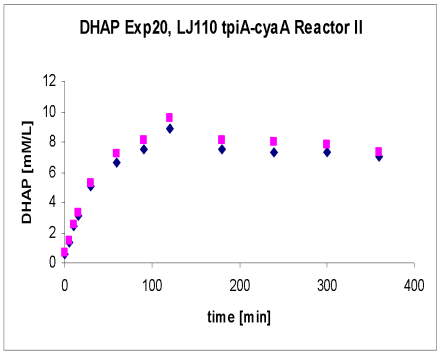
Comp Exp. No.	**Exp 20A**
Hexokinase U/ml	0.01
Lactate-DH U/ml	1
Glc mM	11.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
Prot.tot. mg/ml	10
T	37
Strain	LJ110 tpiA-cyaA

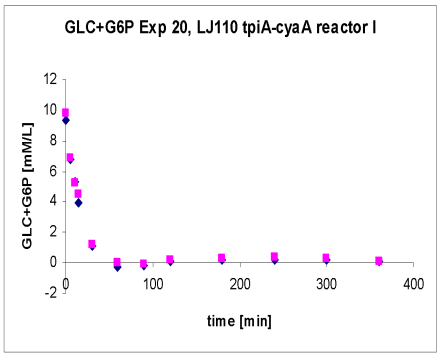
Table B.49: Experiment 20 description

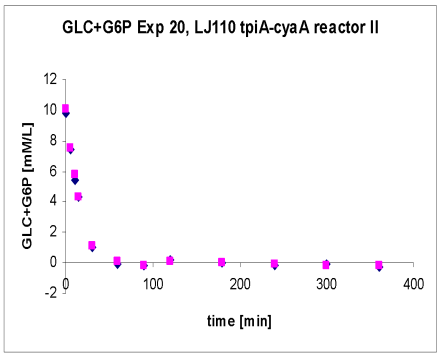
Time	DHAP		GL+G6P		G6P		ATP	
0	0.57	0.61	9.36	9.79	0.59	0.64	8.57	8.92
5	1.44	1.52	6.79	6.86	0.51	0.42	4.63	4.95
10	2.56	2.6	5.28	5.2	0.47	0.52	2.89	3.02
15	3.36	3.74	3.94	4.52	0.52	0.52	2.02	2.36
30	5.1	5.8	1.14	1.16	0.71	0.81	0.76	0.84
60	6.89	7.07	-0.27	-0.03	0.43	0.41	0.49	0.65
90	7.49	8.03	-0.21	-0.06	0.38	0.4	0.23	0.37
120	8.01	9	0.11	0.16	0.41	0.4	0.17	0.32
180	8.04	8.59	0.2	0.26	0.35	0.34	-0.09	-0.36
240	7.68	7.91	0.18	0.37	0.25	0.28	-0.17	-0.07
300	7.75	8.35	0.19	0.25	0.2	0.21	-0.19	0.02
360	7.35	8.06	0.14	0.07	0.14	0.13	-0.15	0.12

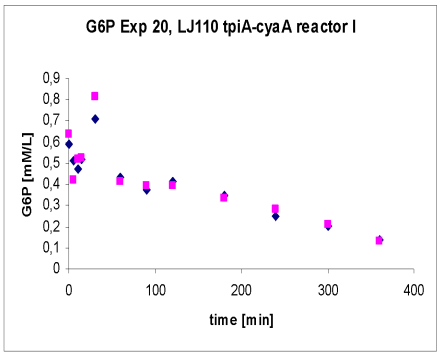
Table B.50: Experiment 20A, reactor I

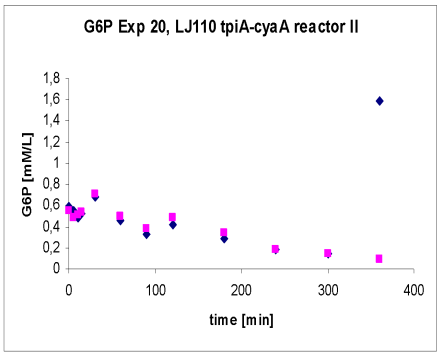

DHAP Exp20, LJ110 tpiA-cyaA Reactor I
DHAP [mM/L]
time [min]

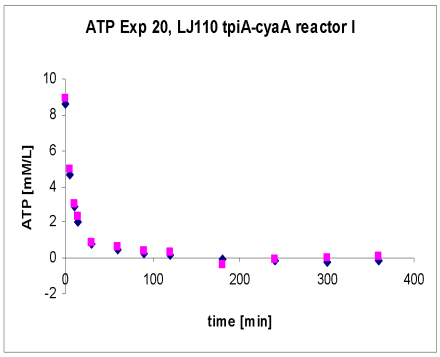

DHAP Exp20, LJ110 tpiA-cyaA Reactor II
DHAP [mM/L]
time [min]

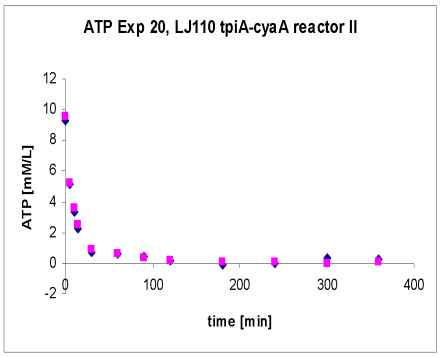

GLC+G6P Exp 20, LJ110 tpiA-cyaA reactor I
GLC+G6P [mM/L]
time [min]


GLC+G6P Exp 20, LJ110 tpiA-cyaA reactor II
GLC+G6P [mM/L]
time [min]


G6P Exp 20, LJ110 tpiA-cyaA reactor I
G6P [mM/L]
time [min]


G6P Exp 20, LJ110 tpiA-cyaA reactor II
G6P [mM/L]
time [min]


ATP Exp 20, LJ110 tpiA-cyaA reactor I
ATP [mM/L]
time [min]


ATP Exp 20, LJ110 tpiA-cyaA reactor II
ATP [mM/L]
time [min]

Time	DHAP		GL+G6P		G6P		ATP	
0	0.62	0.64	9.8	10.04	0.59	0.56	9.26	9.53
5	1.35	1.43	7.39	7.49	0.55	0.48	5.15	5.19
10	2.47	2.56	5.41	5.75	0.48	0.51	3.35	3.61
15	3.09	3.35	4.33	4.34	0.52	0.54	2.24	2.52
30	5.06	5.26	1.04	1.1	0.68	0.72	0.71	0.89
60	6.65	7.19	-0.11	0.1	0.46	0.5	0.63	0.58
90	7.54	8.1	-0.16	-0.14	0.33	0.39	0.41	0.35
120	8.83	9.58	0.17	0.15	0.43	0.49	0.16	0.14
180	7.51	8.11	0	-0.03	0.29	0.35	-0.07	0.12
240	7.3	7.97	-0.13	-0.1	0.18	0.18	0.01	0.08
300	7.29	7.78	-0.1	-0.14	0.15	0.14	0.31	0
360	7.04	7.3	-0.29	-0.21	1.59	0.09	0.25	0.07

Table B.51: Experiment 20A, reactor II

B.16 Experiment 23

Comp Exp. No.	**Exp 23A**
Hexokinase U/ml	0.01
Lactate-DH U/ml	1
Glc mM	11.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
PPi mM	40.00
Prot.tot. mg/ml	10
T	37
Strain	LJ110 tpiA-cyaA

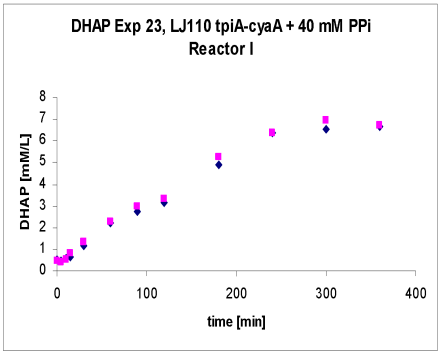
Table B.52: Experiment 23 description

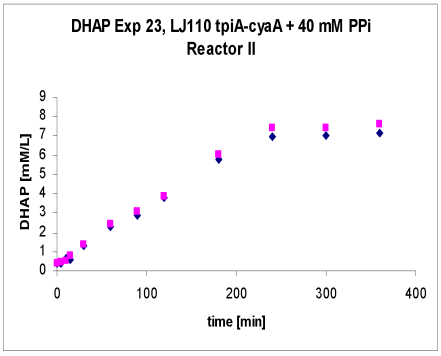
Time	DHAP		GL+G6P	
0	0.5	0.5	9.06	9.48
5	0.44	0.43	7.96	8.22
10	0.57	0.53	7.78	8.18
15	0.67	0.81	6.91	6.87
30	1.18	1.35	5.14	5.89
60	2.25	2.3	3.49	3.32
90	2.75	2.96	2.12	2.12
120	3.15	3.35	1.41	1.39
180	4.88	5.23	-0.82	-0.77
240	6.36	6.35	-0.95	-0.92
300	6.54	6.97	-0.53	-1.01
360	6.67	6.7	-1.26	-0.98

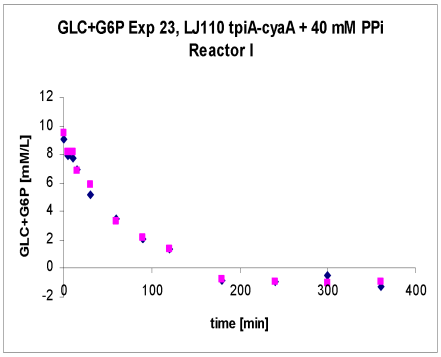
Table B.53: Experiment 23A, reactor I

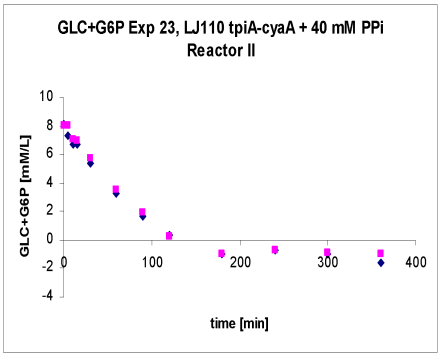
Time	DHAP		GL+G6P	
0	0.41	0.38	8.13	8.04
5	0.4	0.48	7.35	8.06
10	0.64	0.54	6.72	7.1
15	0.61	0.78	6.76	6.98
30	1.31	1.39	5.37	5.74
60	2.29	2.44	3.27	3.54
90	2.86	3.1	1.66	1.89
120	3.82	3.87	0.33	0.26
180	5.79	6.07	-1.01	-0.95
240	6.93	7.45	-0.72	-0.76
300	7.06	7.42	-0.99	-0.94
360	7.18	7.63	-1.64	-0.99

Table B.54: Experiment 23A, reactor II









B.17 Experiment 24

Comp Exp. No.	**Exp 24A**	**Exp 24B**	**Exp 24C**
Hexokinase U/ml	0.01	0.01	0.01
Lactate-DH U/ml	1	1	1
Glc mM	11.1	11.1	11.1
PO4 mM	11.1	11,1	11.1
ATP mM	11.5	5.75	1.15
NAD+ mM	5.75	5.75	5.75
cAMP mM	10.00	10.00	10.00
Prot.tot. mg/ml	10	10	10
T	37	37	37
Strain	LJ110 tpiA-cyaA	LJ110 tpiA-cyaA	LJ110 tpiA-cyaA

Table B.55: Experiment 24 description

Time	DHAP		GL+G6P		G6P		ATP		LAC	
0	0.35	0.26	9.54	10.67	0.41	0.82	7.04	7.87	1.32	1.27
5	1	1.22	8.04	8.25	0.35	0.73	4.29	4.42	1.03	1.04
10	2.03	1.95	6.34	6.77	0.6	0.64	3.24	3.03	1.17	1.06
15	2.65	2.82	5.36	5.42	0.62	0.65	2.45	2.44	1.24	1.21
30	4.28	4.69	2.59	2.61	0.65	0.72	1.14	1.11	1.32	1.29
60	6.53	6.77	0.46	0.41	0.72	0.75	0.37	0.35	1.19	1.12
90	7.87	7.93	0.26	0.45	0.68	0.7	0.33	0.53	1.17	1.11
120	7.98	8.58	0	0.02	0.45	0.46	0.19	0.15	1.16	1.06
180	8.32	8.8	0.46	0.55	0.47	0.51	-0.13	-0.14	0.84	0.8
259	7.42	7.67	-0.08	0.1	0.17	0.2	-0.24	-0.19	0.75	0.66
300	6.78	7.53	-0.14	-0.16	0.11	0.12	-0.26	-0.29	0.52	0.58
360	6.6	7.05	-0.28	-0.21	0.08	0.09	-0.15	-0.27	0.48	0.53

Table B.56: Experiment 24A, 11.5 mM ATP, reactor I

Time	DHAP		GL+G6P		G6P		ATP		LAC	
0	0.4	0.45	9.2	9.7	0.62	0.63	6.23	6.36	1.48	1.39
5	1.01	1.08	7.6	8.22	0.68	0.72	4.31	4.5	1.26	1.21
10	1.72	1.81	6.13	6.32	0.53	0.56	2.86	2.98	1.37	1.35
15	2.46	2.7	5.16	5.24	0.58	0.6	2.36	2.33	1.68	1.03
30	4.46	4.71	2.77	2.85	0.76	0.77	0.97	0.91	2.5	1.44
60	6.29	6.57	0.7	0.63	0.87	0.91	0.47	0.27	3.22	1.78
90	7.39	7.37	0.78	0.76	0.66	0.7	0.49	0.19	2.83	1.75
120	7.77	8.1	-0.08	-0.01	0.47	0.52	0.45	0.05	1.8	1.22
180	7.6	8.6	0.09	0.04	0.3	0.31	-0.15	-0.22	0.88	0.86
259	7.45	8.04	0.51	0.55	0.24	0.26	-0.13	-0.14	0.88	0.96
300	7.12	9.63	0.04	-0.23	0.11	0.12	-0.24	-0.27	0.66	0.78
360	6.64	7.11	-0.33	-0.29	0.08	0.08	-0.06	-0.12	0.73	0.75

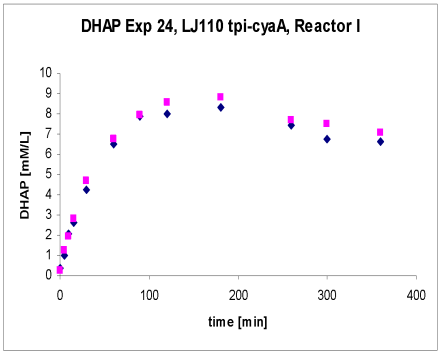
Table B.57: Experiment 24A, 11.5 mM ATP, reactor II

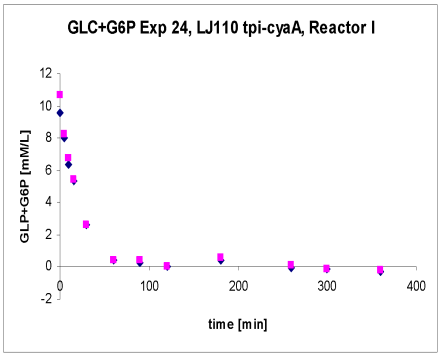
Time	DHAP		GL+G6P		G6P		ATP		LAC	
0	0.51	0.53	7.02	6.77	0.13	0.11	1.19	1.52	3.5	5.77
5	0.74	0.85	6.31	6.99	0.35	0.26	-0.44	0.05	3.21	3.17
10	1.13	1.16	5.71	5.97	0.53	0.47	-0.6	-0.96	2.42	2.6
15	1.24	1.5	4.94	5.31	0.58	0.6	-0.75	-0.68	2.05	2.12
30	2.03	2.27	4.43	6.72	0.81	0.83	-0.94	-0.83	1.43	1.38
60	2.89	3.26	3.51	3.45	0.83	0.89	-0.98	-1.02	1.16	1.06
90	3.98	4.25	2.66	2.6	0.71	0.76	-0.91	-0.97	1.06	0.88
120	4.42	4.75	1.44	1.58	0.58	0.61	-1.05	-0.9	1.12	1.09
180	4.99	5.02	1.08	0.94	0.58	0.7	-0.35	-0.13	1.03	1.15
259	5.8	6.27	-0.08	-0.37	0.29	0.34	-0.05	-0.09	1.02	1.08
300	6.18	6.77	-0.35	-0.6	0.13	0.13	-0.22	-0.32	1.19	1.24
360	6.64	7.26	-0.79	-0.63	0.09	0.07	-0.22	-0.19	1.14	1.2

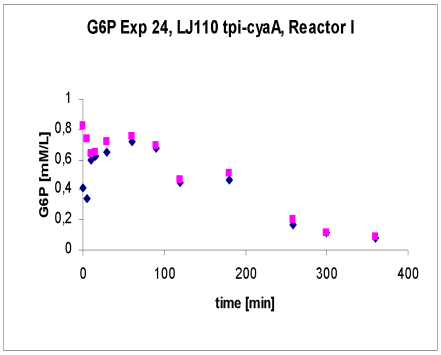
Table B.58: Experiment 24B, 5.75 mM, reactor III

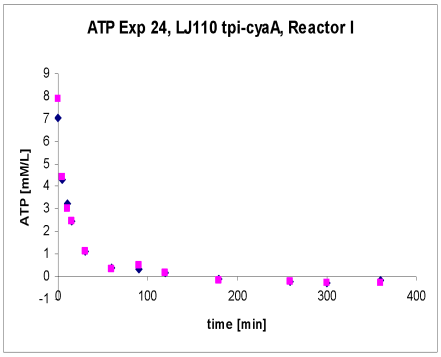
Time	DHAP		GL+G6P		G6P		ATP		LAC	
0	0.32	0.54	9.56	9.76	0.06	0.07	-0.57	-0.33	0.24	1.56
5	0.62	0.72	8.28	8.41	0.21	0.2	-0.76	-0.53	0.86	0.92
10	0.78	0.9	7.76	7.92	0.25	0.27	-1.18	-0.77	0.76	0.77
15	0.91	1.1	7.63	7.73	0.34	0.35	-1.03	-0.7	0.66	0.67
30	1.36	1.58	6.92	7.27	0.36	0.39	-1.08	-0.82	0.58	0.63
60	1.91	1.97	6.35	6.9	0.43	0.48	-1.05	-0.81	0.62	0.72
90	2.35	2.44	6.29	6.75	0.48	0.43	-1.2	-0.72	0.52	0.71
120	2.94	3.22	5.4	6.22	0.36	0.33	-1.03	-0.71	0.65	0.82
180	3.25	3.34	4.74	5.01	0.37	0.38	-0.21	-0.35	0.71	0.74
259	4.33	4.31	3.12	3.18	0.29	0.34	-0.4	-0.38	0.73	0.74
300	5.07	5.55	1.77	1.84	0.27	0.3	-0.35	-0.48	0.71	0.72
360	6.15	5.84	0.61	0.75	0.27	0.29	-0.43	-0.43	0.78	0.81

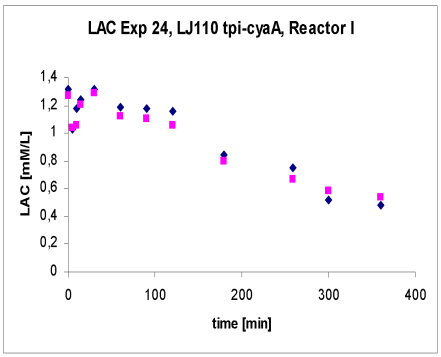
Table B.59: Experiment 24B, 5.75 mM, reactor IV

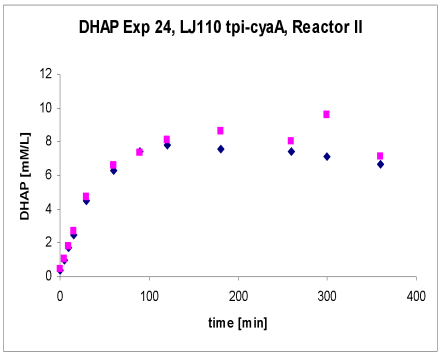

DHAP Exp 24, LJ110 tpi-cyaA, Reactor I
DHAP [mM/L]
time [min]

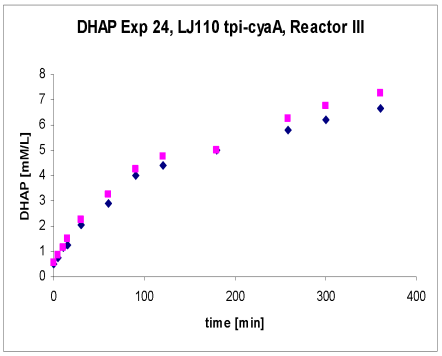

GLC+G6P Exp 24, LJ110 tpi-cyaA, Reactor I
GLP+G6P [mM/L]
time [min]

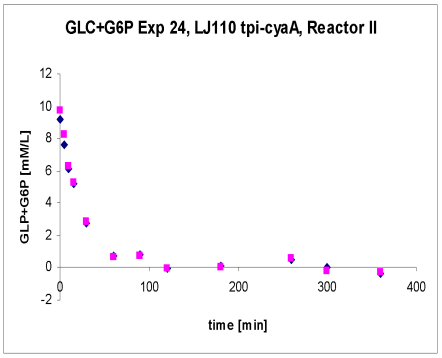

G6P Exp 24, LJ110 tpi-cyaA, Reactor I
G6P [mM/L]
time [min]

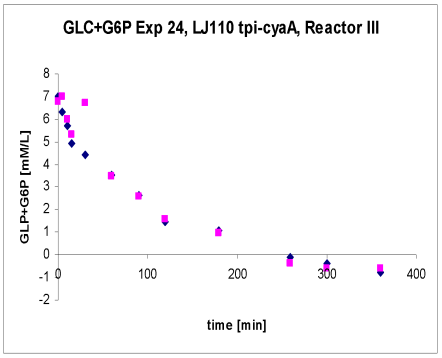

ATP Exp 24, LJ110 tpi-cyaA, Reactor I
ATP [mM/L]
time [min]

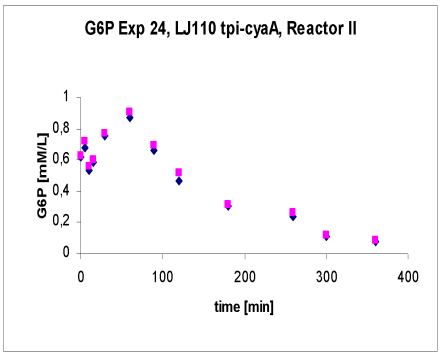

LAC Exp 24, LJ110 tpi-cyaA, Reactor I
LAC [mM/L]
time [min]

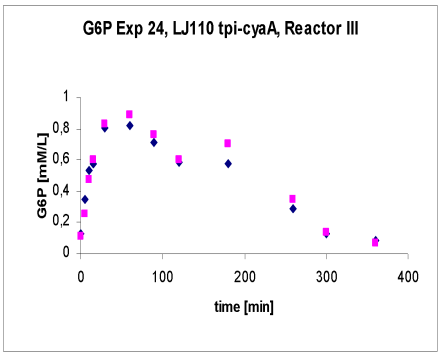

DHAP Exp 24, LJ110 tpi-cyaA, Reactor II
DHAP [mM/L]
time [min]

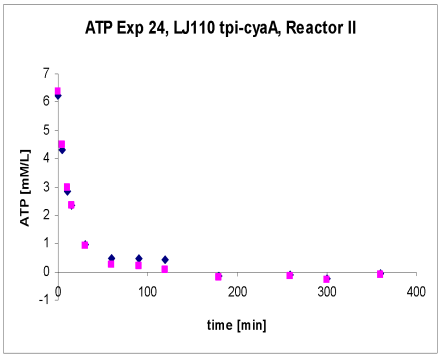

DHAP Exp 24, LJ110 tpi-cyaA, Reactor III
DHAP [mM/L]
time [min]

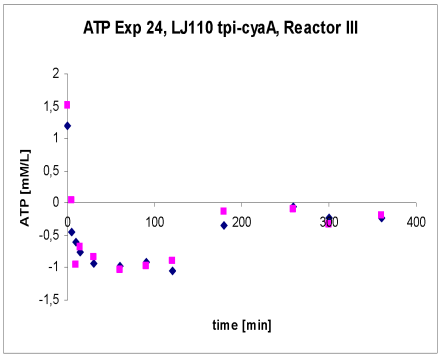

GLC+G6P Exp 24, LJ110 tpi-cyaA, Reactor II
GLP+G6P [mM/L]
time [min]

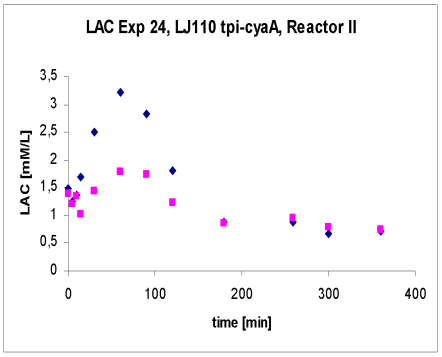

GLC+G6P Exp 24, LJ110 tpi-cyaA, Reactor III
GLP+G6P [mM/L]
time [min]

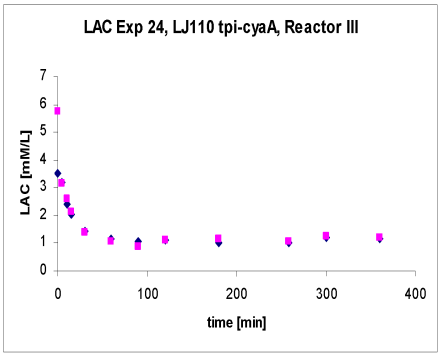

G6P Exp 24, LJ110 tpi-cyaA, Reactor II
G6P [mM/L]
time [min]

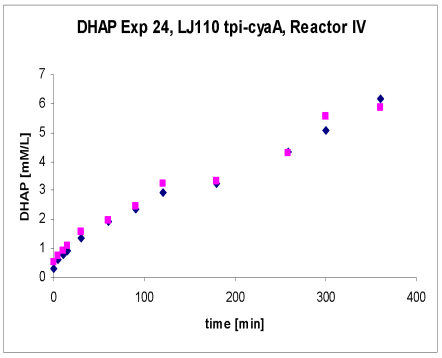

G6P Exp 24, LJ110 tpi-cyaA, Reactor III
G6P [mM/L]
time [min]

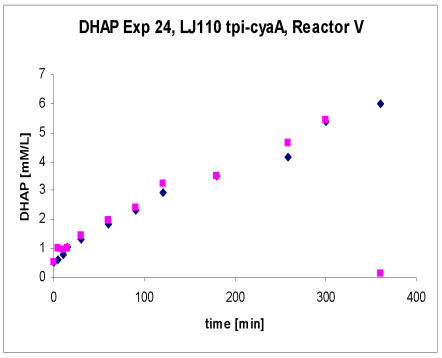

ATP Exp 24, LJ110 tpi-cyaA, Reactor II
ATP [mM/L]
time [min]

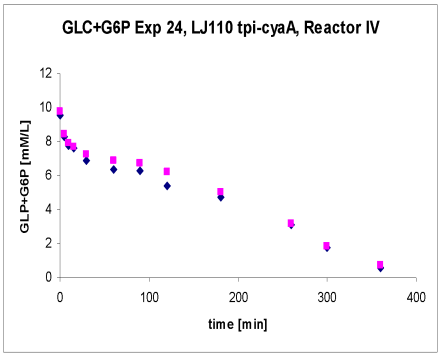

ATP Exp 24, LJ110 tpi-cyaA, Reactor III
ATP [mM/L]
time [min]

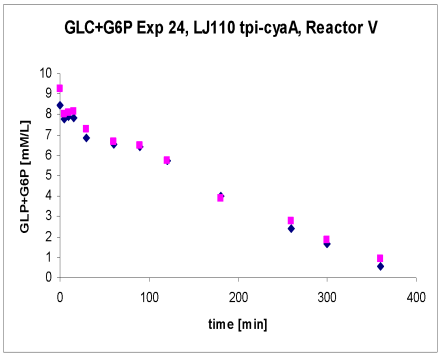

LAC Exp 24, LJ110 tpi-cyaA, Reactor II
LAC [mM/L]
time [min]

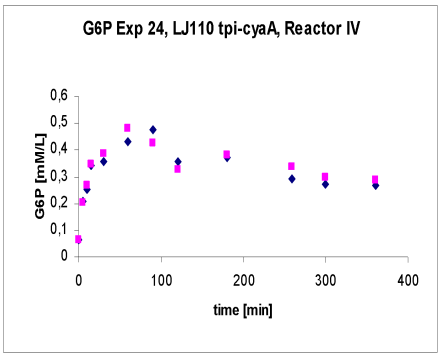

LAC Exp 24, LJ110 tpi-cyaA, Reactor III
LAC [mM/L]
time [min]

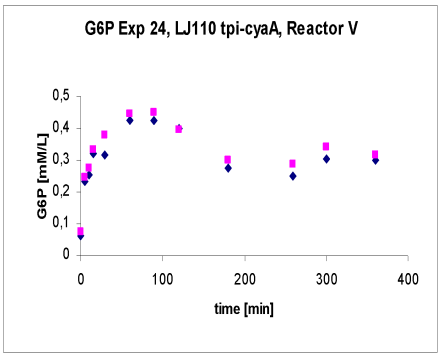

DHAP Exp 24, LJ110 tpi-cyaA, Reactor IV
DHAP [mM/L]
time [min]


DHAP Exp 24, LJ110 tpi-cyaA, Reactor V
DHAP [mM/L]
time [min]


GLC+G6P Exp 24, LJ110 tpi-cyaA, Reactor IV
GLP+G6P [mM/L]
time [min]

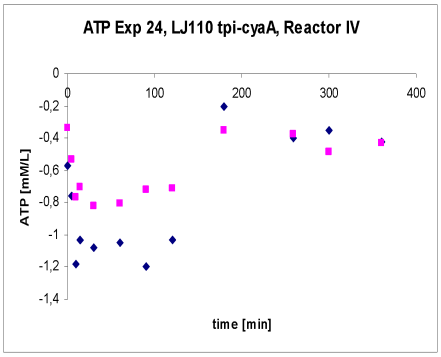

GLC+G6P Exp 24, LJ110 tpi-cyaA, Reactor V
GLP+G6P [mM/L]
time [min]

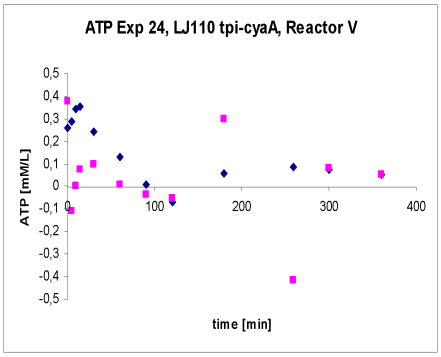

G6P Exp 24, LJ110 tpi-cyaA, Reactor IV
G6P [mM/L]
time [min]

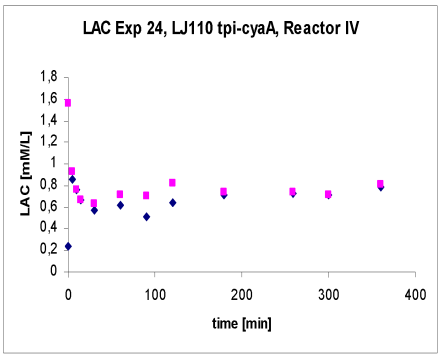

G6P Exp 24, LJ110 tpi-cyaA, Reactor V
G6P [mM/L]
time [min]

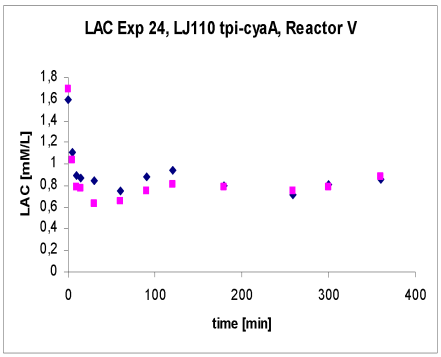
Time	DHAP		GL+G6P		G6P		ATP		LAC	
0	0.51	0.55	8.48	9.24	0.06	0.07	0.26	0.38	1.59	1.69
5	0.62	0.99	7.8	8.06	0.23	0.25	0.29	-0.11	1.11	1.03
10	0.8	0.96	7.87	8.11	0.25	0.28	0.34	0	0.89	0.79
15	1.03	1.02	7.85	8.16	0.32	0.33	0.36	0.08	0.87	0.78
30	1.31	1.43	6.86	7.26	0.32	0.38	0.24	0.1	0.85	0.63
60	1.84	1.97	6.54	6.69	0.43	0.45	0.13	0.01	0.75	0.66
90	2.32	2.42	6.42	6.5	0.42	0.45	0.01	-0.03	0.88	0.75
120	2.95	3.25	5.71	5.75	0.4	0.4	-0.07	-0.05	0.94	0.81
180	3.52	3.52	4.04	3.9	0.28	0.3	0.06	0.3	0.8	0.79
259	4.17	4.64	2.39	2.75	0.25	0.29	0.08	-0.42	0.71	0.76
300	5.39	5.41	1.69	1.84	0.3	0.34	0.07	0.08	0.82	0.79
360	5.99	0.13	0.54	0.95	0.3	0.32	0.05	0.05	0.86	0.88

Table B.60: Experiment 24C, 1.15 mM, reactor V









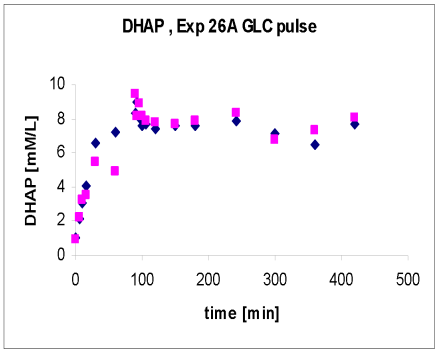
B.18 Experiment 26

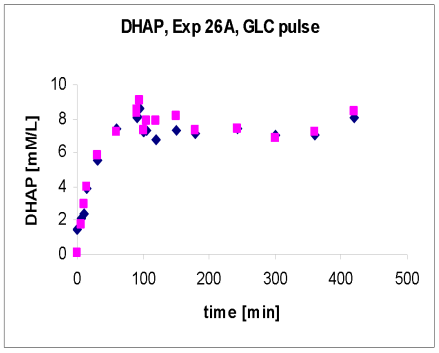
Comp Exp. No.	**Exp 26A**
Hexokinase U/ml	0.01
Lactate-DH U/ml	1
Glc mM	11.1/11.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
Prot.tot. mg/ml	10
T	37
Strain	LJ110 tpiA-cyaA

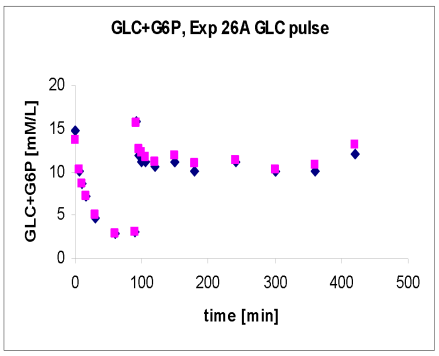
Table B.61: Experiment 26 description

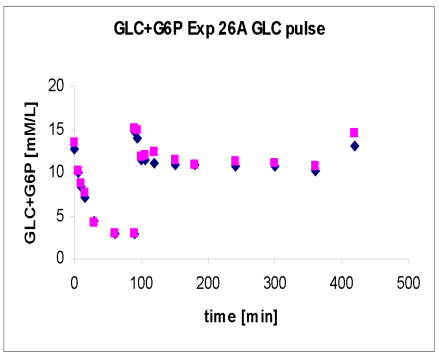
Time	DHAP		GL+G6P		G6P		ATP		PYR	
0	0.98	0.9	14.69	13.72	0.74	0.8	9	9.34	1.16	0.98
6	2.13	2.22	10.1	10.35	0.67	0.72	4.6	4.79	1.44	1.35
10	3.08	3.28	8.72	8.72	0.74	0.72	3.52	3.62	1.92	1.96
15	4.06	3.55	7.24	7.23	0.7	0.74	2.46	2.47	2.27	2.42
30	6.55	5.47	4.75	5.05	0.85	0.86	2.21	2.31	2.53	2.77
60	7.27	4.86	2.82	2.85	0.44	0.46	0.63	0.64	2.08	2.11
90	8.3	9.44	2.99	2.99	0.46	0.44	0.61	0.4	1.54	1.81
91	8.97	8.14	15.79	15.63	0.63	0.61	0.66	0.56	1.64	1.77
95	8.06	8.87	11.84	12.7	0.79	0.8	0.63	0.46	1.46	1.63
100	7.6	8.11	11.19	12.22	0.68	0.76	0.64	0.31	1.48	1.14
105	7.66	7.83	11.18	11.66	0.73	0.8	0.65	0.21	1.37	1.3
120	7.38	7.8	10.71	11.14	0.76	0.78	0.15	0.22	0.4	0.44
150	7.6	7.66	11.1	11.93	1.06	1.15	0.4	0.44	0.13	0.17
180	7.57	7.86	10.15	10.95	0.74	0.73	0.15	0.15	0.24	-0.02
242	7.9	8.36	11.21	11.4	0.64	0.65	0.16	0.19	0.57	0.55
300	7.14	6.75	10.08	10.26	0.14	0.36	0.1	0.17	0.34	0.18
360	6.44	7.3	10.01	10.85	0.29	0.3	0.16	-0.04	0.48	0.27
420	7.69	8.03	12.01	13.23	0.43	0.44	0.33	0.19	0.58	0.74

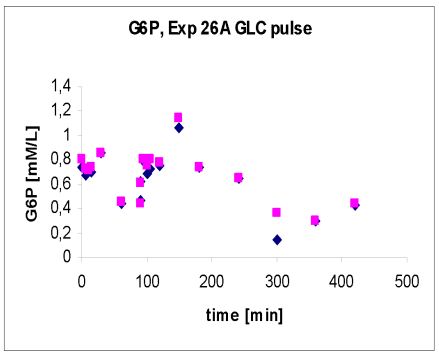
Table B.62: Experiment 26A, 11,5 mM, reactor I

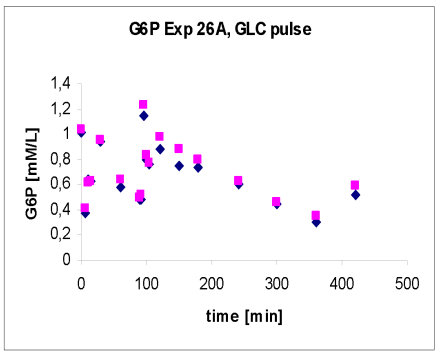

DHAP , Exp 26A GLC pulse
DHAP [mM/L]
time [min]

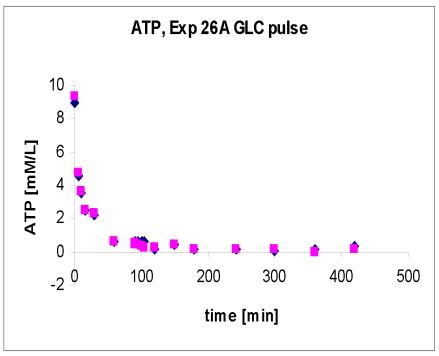

DHAP, Exp 26A, GLC pulse
DHAP [mM/L]
time [min]

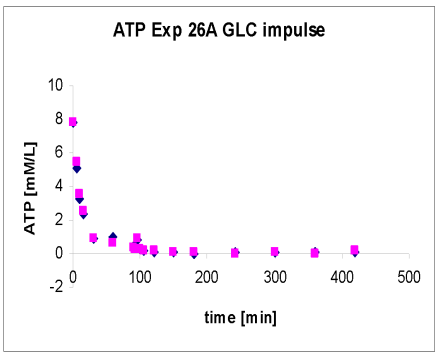

GLC+G6P, Exp 26A GLC pulse
GLC+G6P [mM/L]
time [min]


GLC+G6P Exp 26A GLC pulse
GLC+G6P [mM/L]
time [min]


G6P, Exp 26A GLC pulse
G6P [mM/L]
time [min]

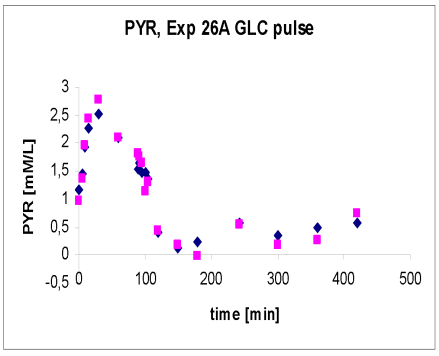

G6P Exp 26A, GLC pulse
G6P [mM/L]
time [min]

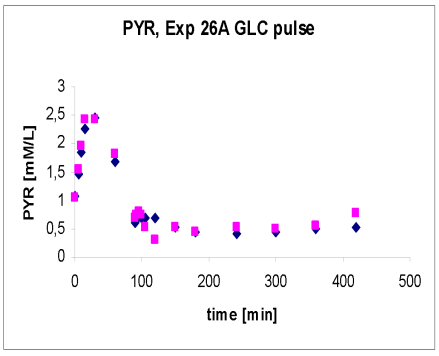

ATP, Exp 26A GLC pulse
ATP [mM/L]
time [min]


ATP Exp 26A GLC impulse
ATP [mM/L]
time [min]

Time	DHAP		GL+G6P		G6P		ATP		PYR	
0	1.48	0.06	12.68	13.46	1.01	1.03	7.8	7.84	1.08	1.05
6	2.16	1.74	10.03	10.17	0.38	0.41	5.11	5.47	1.45	1.53
10	2.39	2.96	8.31	8.76	0.64	0.61	3.31	3.5	1.84	1.95
15	3.88	3.98	7.18	7.58	0.63	0.63	2.39	2.58	2.25	2.42
30	5.57	5.82	4.28	4.2	0.94	0.96	0.89	0.89	2.44	2.43
60	7.42	7.24	2.98	2.97	0.58	0.64	0.96	0.65	1.68	1.81
90	8.14	8.48	2.98	3	0.49	0.5	0.44	0.36	0.6	0.7
91	8.03	8.29	14.73	15.1	0.49	0.52	0.32	0.3	0.74	0.74
95	8.64	9.1	14.09	14.95	1.14	1.23	0.84	0.94	0.71	0.8
100	7.2	7.36	11.43	11.86	0.79	0.83	0.23	0.28	0.68	0.75
105	7.34	7.87	11.38	12.09	0.77	0.77	0.18	0.16	0.7	0.51
120	6.72	7.84	11.15	12.35	0.88	0.98	0.1	0.17	0.69	0.31
150	7.28	8.15	10.83	11.44	0.75	0.88	0.11	0.1	0.53	0.54
180	7.15	7.32	10.92	10.99	0.74	0.8	0.02	0.08	0.45	0.45
242	7.39	7.43	10.7	11.24	0.6	0.63	0.07	-0.03	0.42	0.53
300	7.01	6.81	10.68	11.16	0.45	0.46	0.08	0.08	0.44	0.5
360	7.06	7.21	10.19	10.73	0.3	0.35	0.07	0.03	0.5	0.55
420	8.06	8.38	13.08	14.61	0.52	0.59	0.13	0.17	0.53	0.77

Table B.63: Experiment 26A, 11,5 mM, reactor II





B.19 Experiment 27

Comp Exp. No.	Exp 27A
Hexokinase U/ml	0.01
Lactate-DH U/ml	1
Glc mM	5/6.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
Prot.tot. mg/ml	10
T	37
Strain	LJ110 tpiA-cyaA

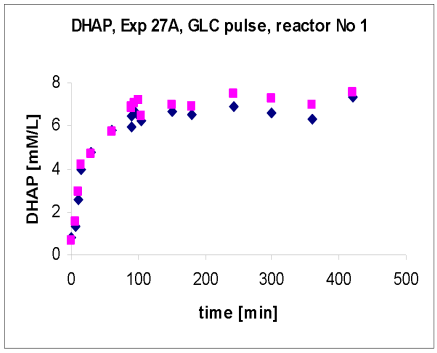
Table B.64: Experiment 27 description

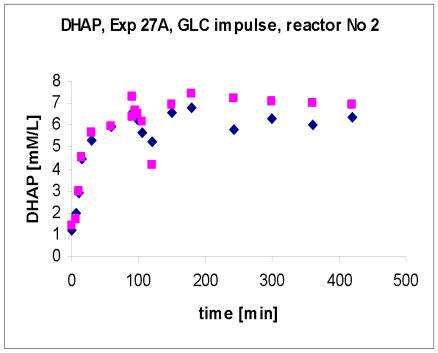
Time	DHAP		GL+G6P		G6P		ATP		ADP		AMP		PYR	
0	0.8	0.66	4.51	4.11	0.3	0.23	9.03	9.66	3.36	3.71	1.28	1.22	1.25	0.84
6	1.33	1.54	2.01	1.89	0.48	0.56	4.52	5.24	3.32	3.52	1.79	3.08	1.19	1.6
10	2.58	2.93	0.34	0.21	0.3	0.33	3.43	3.99	3.53	3.85	2.54	4.78	1.75	2.56
15	3.97	4.15	0.13	0.11	0.15	0.16	4.65	5.26	3.35	3.96	2.26	3.73	2.95	2.75
30	4.78	4.72	0.14	0.14	0.12	0.12	2.2	2.58	2.56	2.77	2.39	4	2.99	2.84
60	5.81	5.74	0.08	-0.05	0.1	0.13	0.85	1.02	1.8	2.03	2.35	4.78	0.44	0.84
90	6.43	6.89	0.12	0.17	0.26	0.27	1.53	1.94	2.23	2.33	1.85	5.54	0.75	0.99
91	5.97	6.86	6.62	7.12	0.27	0.29	0.52	0.76	1.55	2.49	1.64	5.38	0.61	0.88
95	6.7	7.01	8.08	8.52	0.53	0.57	1.11	1.07	1.63	1.91	4.53	7.45	1.06	0.7
100	6.48	7.21	6.2	6.58	0.4	0.43	0.6	0.55	2.19	2.2	3.21	6.09	0.91	0.94
105	6.23	6.44	6.3	6.29	0.48	0.53	0.88	0.84	2.07	2.17	1.85	5.5	0.61	0.34
120									1.08	1.08	-0.53	-0.89	0.41	0.41
150	6.69	6.97	4.95	5.87	0.47	0.5	0.4	0.15	1.61	1.65	3.68	6.27	0.73	0.76
180	6.56	6.92	4.58	4.72	0.44	0.47	0.43	0.2	1.49	1.51	4.06	7.7	-0.15	-0.53
242	6.89	7.5	4.7	5.06	0.51	0.58	0.61	0.4	1.6	1.6	1.72	6.07	0.66	-0.35
300	6.62	7.26	4	4.07	0.36	0.37	0.41	0.2	1.46	1.48	3.21	7.42	-0.23	-0.96
360	6.32	6.96	3.8	3.36	0.3	0.31	0.36	-0.01	1.4	1.44	3.22	7.42	-0.9	0.22
420	7.37	7.57	4.43	3.07	0.44	0.5	0.6	0.27	1.63	1.63	2.98	1.78	0.59	0.31

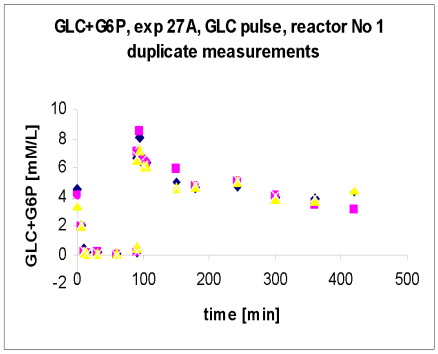
Table B.65: Experiment 27A, 11,5 mM ATP, reactor I

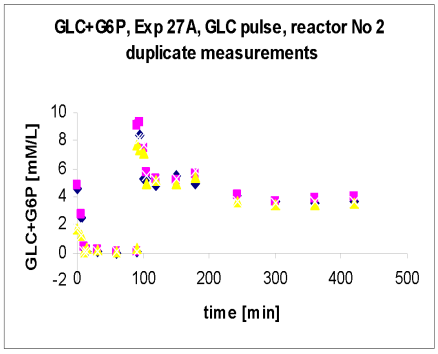
Time	DHAP		GL+G6P		G6P		ATP		ADP		AMP		PYR	
0	1.19	1.41	4.59	4.83	0.57	0.59	9.24	9.82	6.29	7.24	2.33	2.37	1.38	1.07
6	1.95	1.71	2.53	2.75	0.3		7.44	-0.29	5.86	6.05	2.13	1.95	1.89	1.89
10	2.88	3	0.28	0.46	0.19	0.21	6.28	6.99	4.96	5.48	0.75	1.63	2.13	2.12
15	4.49	4.55	0.1	0.21	0.21	0.25	7.99	8.93	6.28	6.51	2.26	2.23	2.79	2.77
30	5.3	5.64	0.08	0.14	0.17	0.18	5.25	5.87	3.72	4.57	2.21	1.99	3.31	3.1
60	5.92	5.95	-0.07	0.05	0.12	0.12	1.01	1.09	1.82	2.11	2.83	3.15	0.98	0.87
90	6.6	6.36	0.08	0.13	0.14	0.15	0.51	0.67	2.59	1.84	3.33	3.57	0.56	0.83
91	6.52	7.29	9.27	9.11	0.35	0.37	1.53	1.72	0.09	2.31	0.06	3.3	0.81	0.73
95	6.63	6.69	8.43	9.34	0.45	0.48	1.06	1.25	1.13	1.94	3.01	3.69	0.88	0.34
100	6.25	6.54	5.29	7.51	0.61	0.68	0.81	1.03	2.36	2.53	3.2	2.38	0.63	0.37
105	5.68	6.18	5.31	5.72	0.36	0.4	0.17	0.32	1.95	2.05	3.7	4.03	0.27	0.38
120	5.26	4.21	4.85	5.27	0.43	0.43	0.3	0.24	1.7	1.75	3.64	4.11	0.7	0.22
150	6.56	6.93	5.51	5.2	0.52	0.59	0.43	0.36	1.8	1.73	4.1	2.54	0.71	0.22
180	6.82	7.41	4.9	5.62	0.58	0.67	0.45	0.41	0.36	1.71	2.35	2.98	0.65	-0.62
242	5.83	7.25	3.99	4.17	0.33	0.36	0.15	0.16	3.03	1.49	4.54	4.31	0.71	0.82
300	6.27	7.05	3.6	3.68	0.3	0.33	0.22	0.17	0.24	1.44	6.48	2.76	0.31	-0.32
360	5.99	6.99	3.56	3.84	0.31	0.3	0.24	0.18	1.06	1.47	5.19	4.01	-0.24	-0.57
420	6.34	6.97	3.61	3.99	0.27	0.31	0.31	0.39	1.64	1.47	3.99	2.87	0.63	0.63

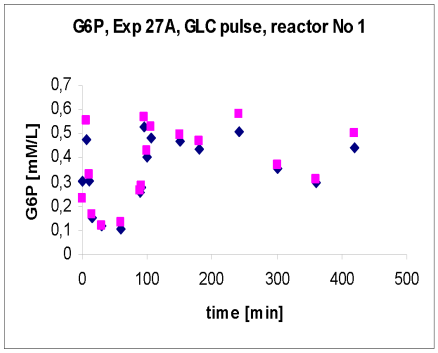
Table B.66: Experiment 27A, 11,5 mM ATP, reactor II

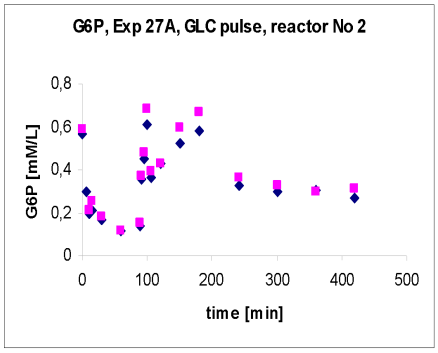

DHAP, Exp 27A, GLC pulse, reactor No 1
DHAP [mM/L]
time [min]

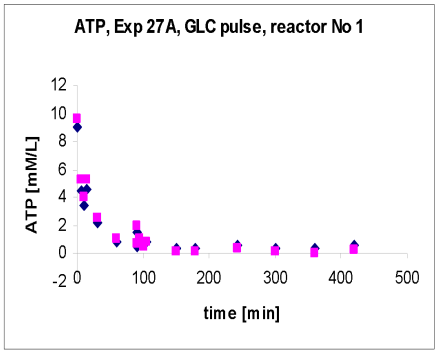

DHAP, Exp 27A, GLC impulse, reactor No 2
DHAP [mM/L]
time [min]

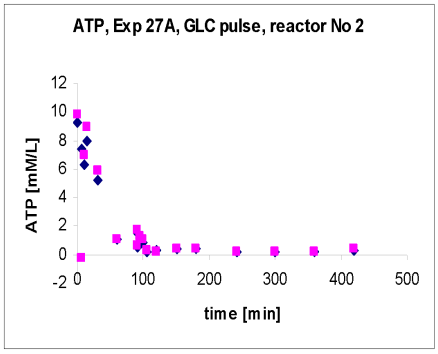

GLC+G6P, exp 27A, GLC pulse, reactor No 1
duplicate measurements
GLC+G6P [mM/L]
time [min]

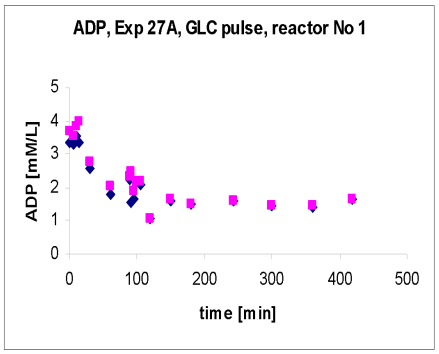

GLC+G6P, Exp 27A, GLC pulse, reactor No 2
duplicate measurements
GLC+G6P [mM/L]
time [min]

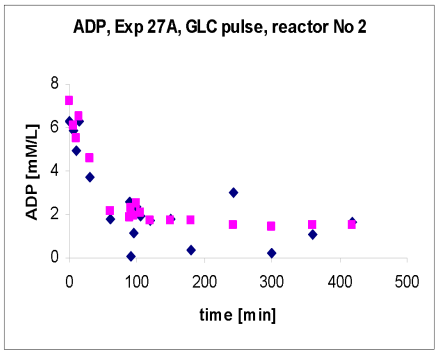

G6P, Exp 27A, GLC pulse, reactor No 1
G6P [mM/L]
time [min]

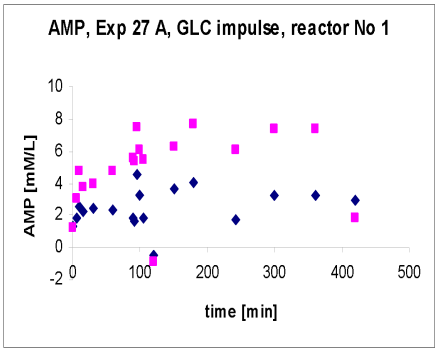

G6P, Exp 27A, GLC pulse, reactor No 2
G6P [mM/L]
time [min]

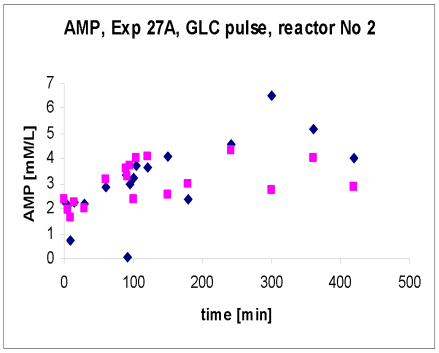

ATP, Exp 27A, GLC pulse, reactor No 1
ATP [mM/L]
time [min]

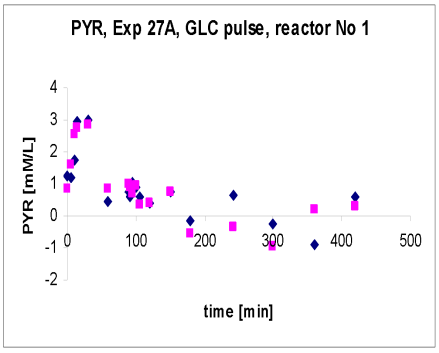

ATP, Exp 27A, GLC pulse, reactor No 2
ATP [mM/L]
time [min]

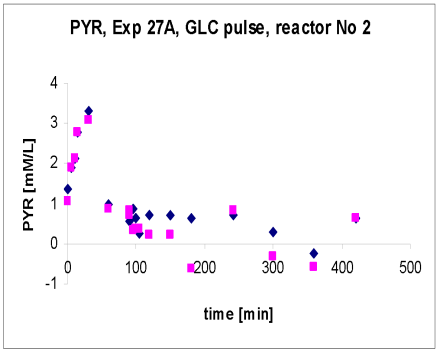

ADP, Exp 27A, GLC pulse, reactor No 1
ADP [mM/L]
time [min]


ADP, Exp 27A, GLC pulse, reactor No 2
ADP [mM/L]
time [min]


AMP, Exp 27 A, GLC impulse, reactor No 1
AMP [mM/L]
time [min]


AMP, Exp 27A, GLC pulse, reactor No 2
AMP [mM/L]
time [min]


PYR, Exp 27A, GLC pulse, reactor No 1
PYR [mM/L]
time [min]


PYR, Exp 27A, GLC pulse, reactor No 2
PYR [mM/L]
time [min]

B.20 Experiment 28

Comp Exp. No.	Exp 28A
Hexokinase U/ml	0.01
Lactate-DH U/ml	1
Glc mM	5/6.1
PO4 mM	11.1
ATP mM	11.5
NAD+ mM	5.75
Prot.tot. mg/ml	10
T	37
Strain	LJ110 tpiA-cyaA

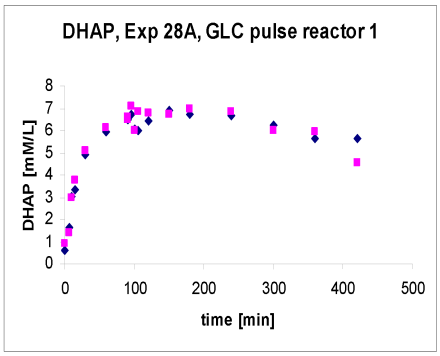
Table B.67: Experiment 28 description

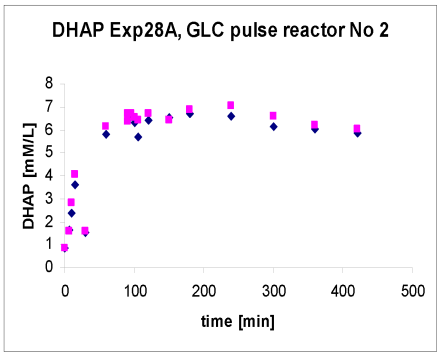
Time	DHAP		GL+G6P		G6P		ATP		ADP		AMP		PYR	
0	0.6090	0.9243	3.7652	3.8339	0.1783	0.1803	8.0000	8.4286	3.3176	3.6377	0.9245	1.0643	1.1464	1.2446
6	1.6325	1.4094	1.5733	1.8348	0.2618	0.2415	4.5000	4.7837	3.8437	3.9603	1.7620	1.8869	1.3265	1.4719
10	3.0328	2.9632	-0.1012	-0.2616	0.1824	0.1885	3.3155	3.5020	3.6725	3.9801	2.6545	2.9163	2.3255	2.2293
15	3.3173	3.7758	-0.3494	-0.4850	0.1375	0.1355	4.5873	4.9167	3.8362	4.1935	2.1621	2.3808	2.5548	2.7677
30	4.9017	5.0696	-0.3112	-0.4391	0.0560	0.0846	2.5714	2.9266	2.9231	3.1836	2.3570	2.5861	2.5056	2.5752
60	5.9314	6.1320	-0.5556	-0.2272	0.0662	0.0866	1.1250	1.2659	2.0645	2.2184	2.7720	2.8910	1.3388	1.3572
90	6.6254	6.4821	-0.4945	-0.6129	0.0988	0.1049	0.3968	0.4187	1.5459	1.7221	3.0919	3.4072	1.2262	1.1853
91	6.4923	6.5824	6.3313	7.3279		0.2659		0.3909	1.5658	1.6600	3.1722	3.3998	1.1525	1.2528
95	6.7052	7.1146	6.1384	6.4783	0.2598	0.2843	0.2113	0.0238	1.3722	1.5136	3.1692	3.7003	1.2344	1.1975
100	6.0481	6.0113	5.9847	5.9504	0.3647	0.3851	0.4812	0.6843	0.9417	0.9888	2.7765	3.2912	0.9959	1.0082
105	5.9990	6.8751	5.6411	5.5838	0.3607	0.4157	0.2014	0.3006	0.8201	0.9367	2.9550	3.3492	1.0471	1.1556
120	6.4248	6.7932	5.2516	5.6869	0.4401	0.4422	0.2391	0.3383	0.7134	0.7258	3.8936	4.1897	1.1105	0.8567
150	6.8915	6.7175	5.0492	4.9995	0.4463	0.4789	0.1300	0.3482	0.6117	0.5893	3.5842	4.1822	0.5906	0.6192
180	6.7318	6.9959	4.6902	4.6807	0.4381	0.4463	0.3284	0.2530	0.4851	0.5695	2.2559	2.5995	0.4841	0.5087
240	6.6643	6.8301	3.9628	4.0544	0.4116	0.3831	0.3442	0.2550	0.3586	0.5521	2.0907	2.4954	0.8690	0.7380
300	6.2160	6.0235	3.5733	4.0506	0.3525	0.3546	0.3423	0.2927	0.3337	0.4330	1.5850	2.3198	0.8895	0.9202
360	5.6264	5.9498	3.5294	3.7203	0.2710	0.2873	0.2431	0.2867	0.3660	0.3412	1.9851	2.2707	0.5763	0.4411
420	5.6530	4.5681	3.5045	3.2621	0.2221	0.2343	0.2946	0.1677	0.2866	0.3238	1.3351	1.3708	0.8117	0.8915

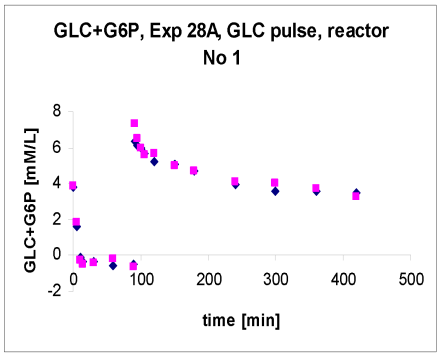
Table B.68: Experiment 28A, 11,5 mM ATP, reactor I

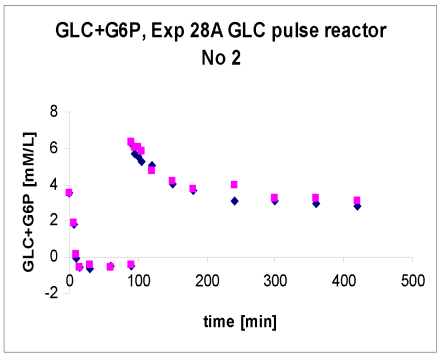
Time	DHAP		GL+G6P		G6P		ATP		ADP		AMP		PYR	
0	0.8628	0.8444	3.5628	3.5666	0.1803	0.1763	7.2163	8.0099	2.9231	3.3486	0.7847	0.8323	1.2774	1.3634
6	1.6305	1.6018	1.8081	1.8578	0.1111	0.1212	5.2579	5.7996	4.1017	4.3821	1.7977	1.8423	1.3265	1.0072
10	2.3675	2.8199	-0.0668	0.1737	0.2455	0.2822	3.5575	3.9643	3.7667	4.0223	2.5712	2.6843	2.1720	2.1474
15	3.6162	4.0686	-0.5327	-0.5289	0.0785	0.0805	3.4246	3.8194	3.7692	4.1538	2.9416	3.0428	2.4565	2.6428
30	1.5322	1.5773	-0.5995	-0.4296	0.0397	0.0540	4.0179	4.6885	3.4938	4.0124	1.9197	2.1607	0.7390	0.6919
60	5.7820	6.1525	-0.4907	-0.5461	0.0662	0.0703	1.4524	1.7857	2.6873	2.6650	3.1157	3.0041	1.4084	1.4166
90	6.4309	6.7114	-0.5136	-0.4067	0.0907	0.0886	0.4464	0.7381	1.7444	1.7345	3.1052	3.2942	1.1750	1.1853
91	6.6581	6.3920	6.2492	6.3484	0.1987	0.1946	0.5833	0.6230	1.7742	1.8412	3.0338	3.3492	1.2139	1.2590
95	6.5926	6.7195	5.7126	6.0678	0.2944	0.3169	0.2014	0.5645	1.5732	1.5707	3.2466	3.6571	1.2078	1.0604
100	6.3122	6.5189	5.5704	6.0477	0.4157	0.4300	0.2411	0.3423	1.2494	0.9640	3.2004	3.8089	1.1904	1.1699
105	5.7144	6.4289	5.2401	5.8683	0.4014	0.4238	0.0962	0.0823	0.9938	0.7333	3.2064	3.7508	0.9058	1.1720
120	6.4002	6.7155	5.0243	4.7399	0.4361	0.4626	0.0169	0.0665	0.8722	0.5546	3.6229	4.1391	1.1331	1.1372
150	6.5148	6.4125	4.0773	4.2205	0.4585	0.4789	0.0228	0.1002	0.5918	0.2618	3.3373	3.9383	0.9836	0.9775
180	6.7093	6.8895	3.7165	3.7604	0.4503	0.4646	-0.0228	0.0109	0.4801	0.2742	2.5861	3.1409	0.9836	0.9488
240	6.5906	7.0430	3.1303	3.9609	0.3790	0.3974	0.0327	-0.0288	0.0856	0.0335	3.3254	4.0231	0.7523	0.8772
300	6.1382	6.5926	3.0845	3.2697	0.3362	0.3647	0.1776	-0.0446	-0.0112	0.0385	2.0833	3.0026	0.6029	0.9345
360	6.0092	6.1730	2.9928	3.2773	0.2731	0.2710	0.0625	-0.0069	-0.1452	-0.1154	2.0000	2.3793	0.7892	0.7748
420	5.8454	6.0297	2.8535	3.1418	0.2140	0.2466	0.1280	-0.1200	-0.1154	-0.0558	1.7501	2.4314	0.7707	0.7175

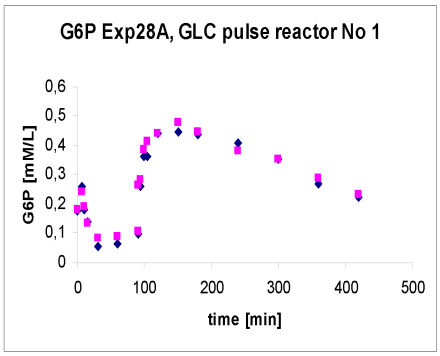
Table B.69: Experiment 28A, 11,5 mM ATP, reactor II

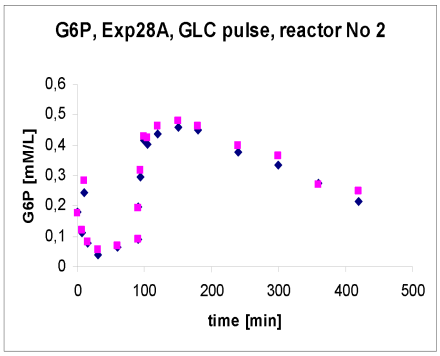

DHAP, Exp 28A, GLC pulse reactor 1
DHAP [mM/L]
time [min]

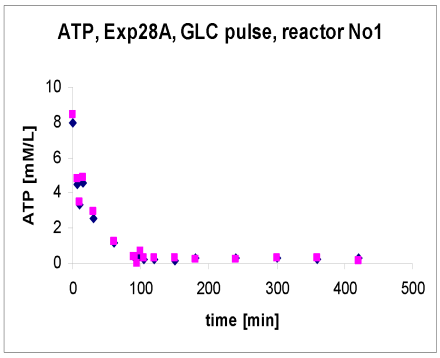

DHAP Exp28A, GLC pulse reactor No 2
DHAP [mM/L]
time [min]

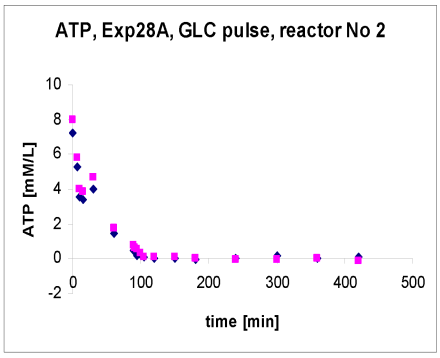

GLC+G6P, Exp 28A, GLC pulse, reactor No 1
GLC+G6P [mM/L]
time [min]

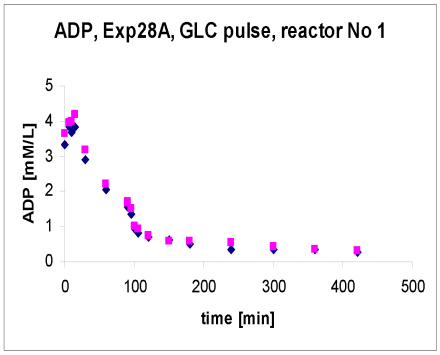

GLC+G6P, Exp 28A GLC pulse reactor No 2
GLC+G6P [mM/L]
time [min]

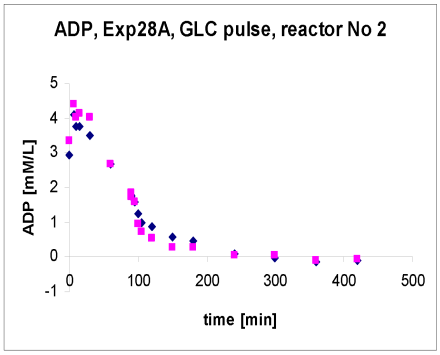

G6P Exp28A, GLC pulse reactor No 1
G6P [mM/L]
time [min]

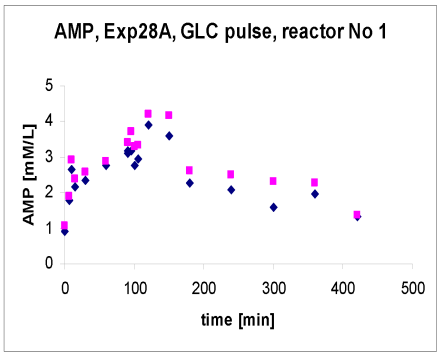

G6P, Exp28A, GLC pulse, reactor No 2
G6P [mM/L]
time [min]

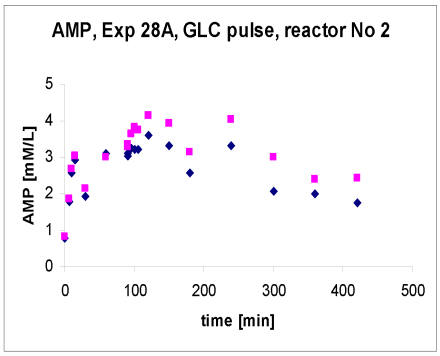

ATP, Exp28A, GLC pulse, reactor No1
ATP [mM/L]
time [min]

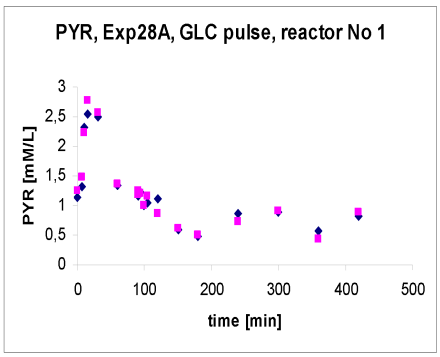

ATP, Exp28A, GLC pulse, reactor No 2
ATP [mM/L]
time [min]

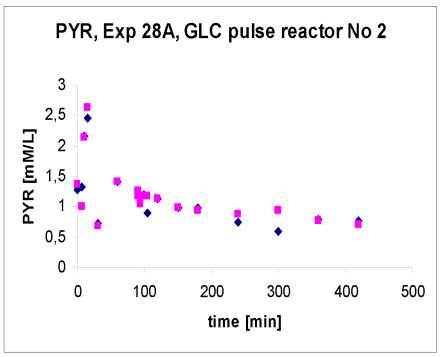

ADP, Exp28A, GLC pulse, reactor No 1
ADP [mM/L]
time [min]


ADP, Exp28A, GLC pulse, reactor No 2
ADP [mM/L]
time [min]


AMP, Exp28A, GLC pulse, reactor No 1
AMP [mM/L]
time [min]


AMP, Exp 28A, GLC pulse, reactor No 2
AMP [mM/L]
time [min]


PYR, Exp28A, GLC pulse, reactor No 1
PYR [mM/L]
time [min]


PYR, Exp 28A, GLC pulse reactor No 2
PYR [mM/L]
time [min]

B.21 Experiment 30

Comp Exp. No.	Exp 30A	Exp 30B
Hexokinase U/ml	0.001	0.001
Lactate-DH U/ml	0.1	0.1
Glc mM	4/4	11.5
PO4 mM	11.1	11.1
ATP mM	11.5	1.15
NAD+ mM	5.75	5.75
Prot.tot. mg/ml	1	1
T	37	37
Strain	LJ110 tpiA-cyaA	LJ110 tpiA-cyaA

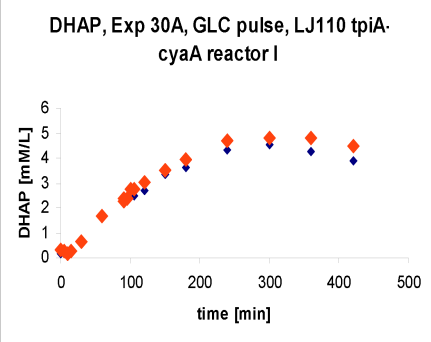
Table B.70: Experiment 30 description

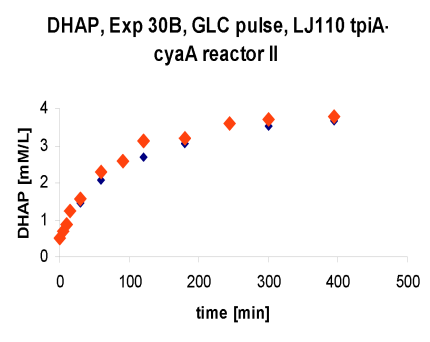
Time	DHAP		GL+G6P		G6P		ATP		ADP		AMP		PYR	
0	0.1412	0.3112	4.0897	4.4506	0.0591	0.0530	5.9514	8.2550	1.1017	1.1290	0.2923	0.3042	0.5865	0.4514
5	0.2436	0.2436	3.5952	3.3317	0.0408	0.0591	6.3224	6.7173	1.9404	1.9355	0.4619	0.4589	0.5374	0.3490
10	0.2682	0.1740	3.4119	3.5532	0.0408	0.0469	6.0724	5.7411	2.1414	2.3896	0.5586	0.6032	0.5906	0.6479
15	0.1945	0.2825	3.2420	3.6105	0.0245	0.0204	4.6875	4.9097	2.7022	2.7295	0.7981	0.8085	0.5865	0.5517
30	0.6919	0.6551	2.5356	2.6406	0.0469	0.0306	6.7966	4.1419	3.5558	3.8114	1.5270	1.5240	0.5455	0.5251
60	1.5599	1.6540	1.2258	1.1169	0.0143	0.0367	3.1121	3.7986	4.3275	4.6129	2.8955	2.8747	0.4739	0.5210
90	2.2006	2.2907	0.6224	0.5728	0.0408	0.0224	2.1062	2.0387	4.9504	4.8065	3.9487	3.7836	0.4350	0.4800
91	2.2702	2.3603	4.8000	5.2525	0.0061	0.0387	2.2212	2.5526	4.8288	5.2233	3.8118	3.9100	0.5210	0.5476
95	2.3173	2.3705	5.1990	5.1800	0.0143	0.0122	3.0446	4.4573	4.8462	5.4144	3.8542	4.2871	0.5619	0.5701
100	2.5691	2.7779	5.0578	5.7279	0.0367	0.0224	2.8304	4.4931	5.3102	5.8263	3.9301	4.0625	0.2651	0.2958
105	2.5056	2.7677	4.5442	4.7160	0.0224	0.0265	3.7669	3.2867	4.8362	5.4640	3.9048	4.1056	0.2692	0.3224
120	2.7001	3.0031	4.2730	4.3036	0.0285	-0.1121	5.3244	2.9673	5.2060	5.3821	4.4968	4.5712	0.3367	0.2057
150	3.3429	3.5189	3.7613	4.0382	0.0428	-0.1651	2.4256	2.2371	5.3002	5.4764	5.4652	5.4920	0.3204	0.2692
180	3.6213	3.9427	3.5246	3.1217	0.0469	-0.2466	2.3304	1.8919	4.7916	5.4690	6.1554	6.3146	0.3756	0.3941
240	4.3419	4.6981	2.6845	2.6749	-0.5298	-0.2343	1.8284	1.3780	4.9801	5.3102	6.8159	7.0569	0.5906	0.4248
300	4.5445	4.8373	1.9418	1.7928	0.0224	-0.1977	1.8105	1.3125	4.7444	5.1935	6.8903	6.9468	0.6581	0.6213
360	4.2600	4.7984	1.9303	1.8692	-0.1100	-0.1365	2.1954	2.1954	4.5558	4.8834	6.9379	7.1075	0.7339	0.9038
420	3.8710	4.5015	1.9742	1.8597	-0.1060	0.0571	3.3879	2.0506	4.7171	5.0323	7.1685	7.5195	0.7421	0.7666

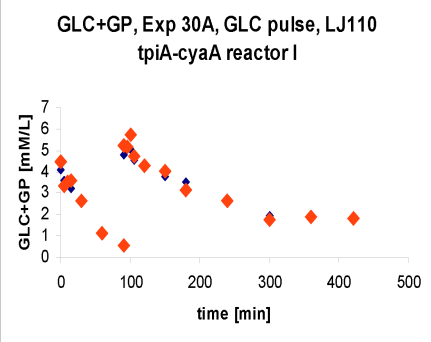
Table B.71: Experiment 30A, GLC pulse, reactor I

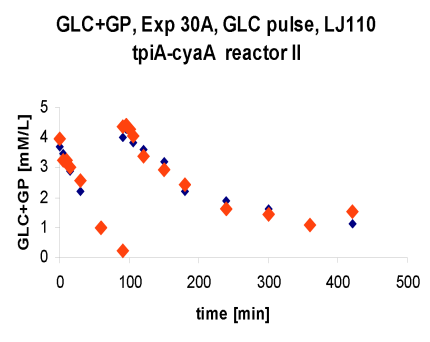
Time	DHAP		GL+G6P		G6P		ATP		ADP		AMP		PYR	
0	0.1024	0.0655	3.7060	3.9714	0.1060	0.0489	6.0823	7.3919	1.0248	0.9876	0.2477	0.2878	-0.3531	-0.4432
5	0.1515	0.1351	3.4711	3.2611	0.0693	0.0815	6.2450	7.3085	2.3002	2.4119	0.5883	0.5779	-0.3142	-0.4248
10	0.1597	0.1801	3.2229	3.2592	0.0041	0.0652	6.5784	6.3978	2.3970	2.6650	0.6493	0.6999	-0.3531	-0.3101
15	0.2968	0.2866	2.9041	3.0301	0.0469	0.0530	5.3165	5.2728	3.0521	2.9777	0.8650	0.9141	-0.3715	-0.3367
30	0.6203	0.6121	2.2205	2.5451	0.0571	0.0448	4.4911	4.1379	3.7717	3.6675	1.5136	1.4927	-0.3449	-0.3224
60	0.9887	1.5455	0.9661	0.9852	0.0265	0.0693	4.3581	2.9792	4.5633	4.9752	2.9416	3.0026	-0.3961	-0.3552
90	1.9038	2.2477	0.2291	0.2406	0.0224	0.0550	1.9613	2.1637	4.8238	5.0124	3.9219	3.9368	-0.3736	-0.3879
91	2.2805	2.2784	4.0115	4.3570	0.0000	0.0632	3.0466	2.2054	4.5583	5.0521	4.0022	4.0736	-0.4350	-0.4432
95	2.3214	2.3030	4.2635	4.4258	0.0428	0.1406	3.0526	3.2768	4.7320	4.9231	3.9598	4.0446	-0.3245	-0.7994
100	2.3603	2.5302	4.2807	4.2883	0.0448	-0.0183	3.3165	6.1478	4.6055	4.8883	3.6489	3.7486	0.0583	0.1648
105	2.5466	2.5077	3.8243	4.0706	-0.0224	0.0387	2.6558	3.0744	4.8313	4.9752	3.9078	3.9583	0.2344	0.2897
120	2.5568	2.8127	3.6200	3.3776	0.0469	0.0245	2.1121	2.0605	4.5583	4.7494	4.2321	4.2678	0.2487	0.2242
150	3.5230	3.4698	3.1866	2.9365	0.0408	0.0815	2.3879	2.1081	4.6998	4.9454	5.1856	5.1990	0.2426	0.2938
180	3.7503	3.6909	2.1900	2.4325	0.0346	0.0346	2.8938	1.2292	4.9529	4.7395	5.8936	5.7717	0.3511	0.1853
240	4.3398	4.5752	1.8807	1.6191	0.0265	0.0835	2.8839	2.6677	4.7270	4.8313	6.2373	6.1376	0.6643	0.5742
300	4.4463	4.7492	1.6305	1.4625	-0.0082	-0.0265	1.4673	1.4474	4.6849	4.8660	6.2536	6.2878	0.6745	0.6233
360	4.5261	4.3112	1.1819	1.0883	0.0183	0.0306	3.1696	1.0784	4.4169	4.5360	6.1540	6.1287	0.6745	0.7216
420	4.2293	4.2866	1.1322	1.5236	0.0020	0.0143	3.5665	0.7649	4.5161	4.6179	6.3697	6.4753	0.8895	-0.0706

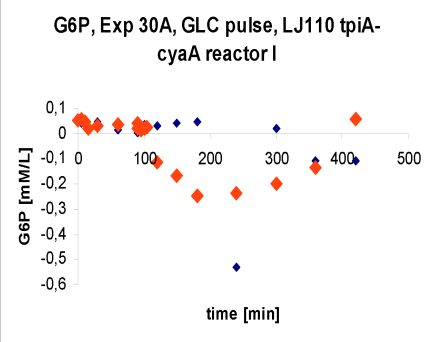
Table B.72: Experiment 30A, GLC pulse, reactor II

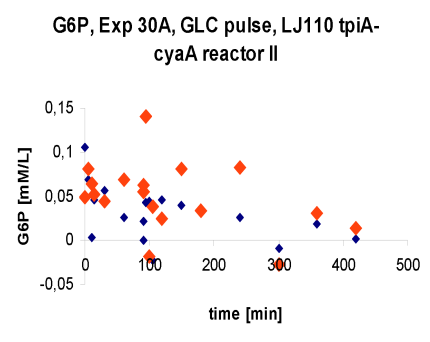

DHAP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
DHAP [mM/L]
time [min]

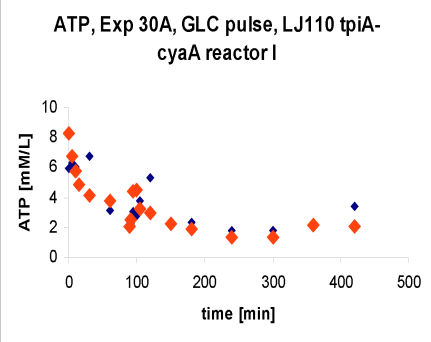

DHAP, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor II
DHAP [mM/L]
time [min]

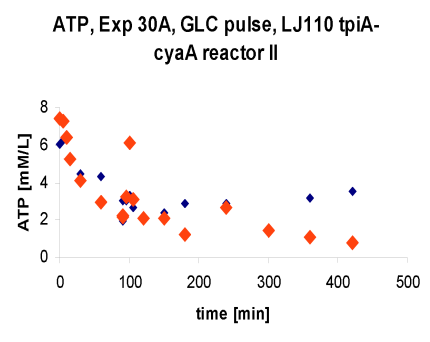

GLC+GP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
GLC+GP [mM/L]
time [min]

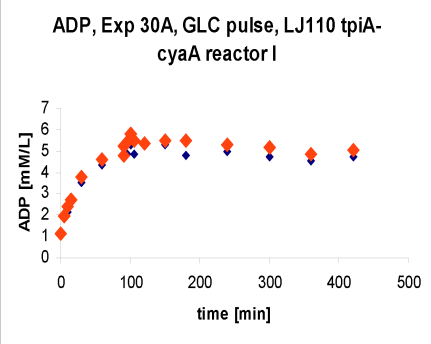

GLC+GP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor II
GLC+GP [mM/L]
time [min]

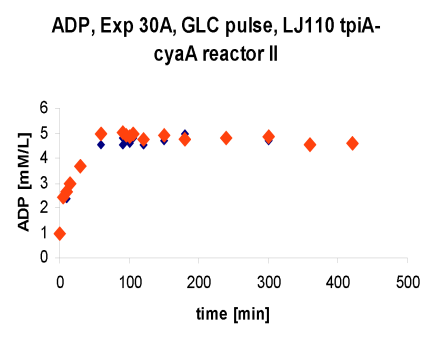

G6P, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
G6P [mM/L]
time [min]

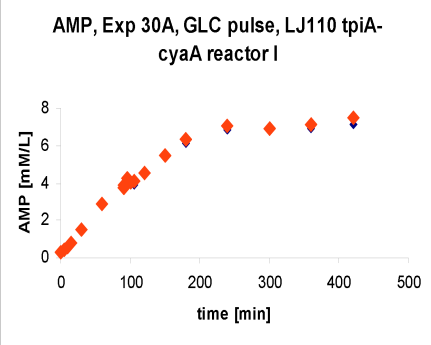

G6P, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor II
G6P [mM/L]
time [min]

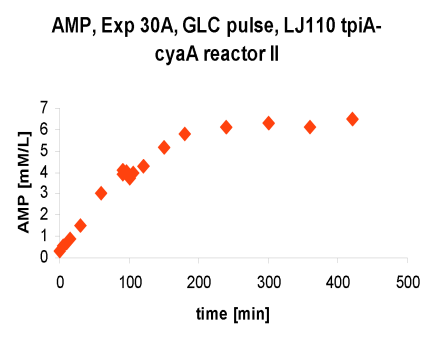

ATP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
ATP [mM/L]
time [min]

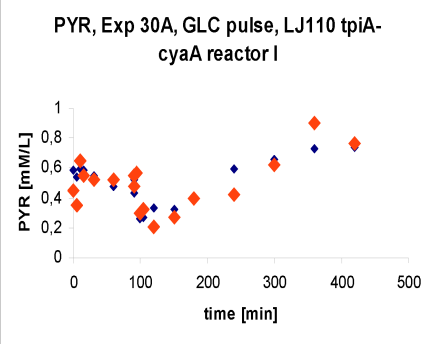

ATP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor II
ATP [mM/L]
time [min]

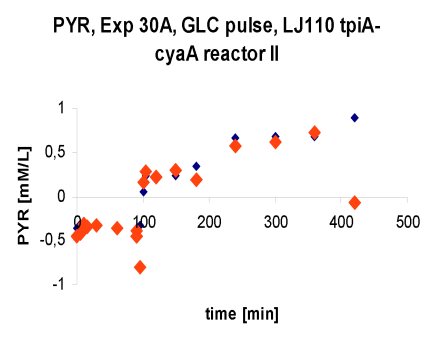

ADP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
ADP [mM/L]
time [min]


ADP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor II
ADP [mM/L]
time [min]


AMP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
AMP [mM/L]
time [min]


AMP, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor II
AMP [mM/L]
time [min]


PYR, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor I
PYR [mM/L]
time [min]

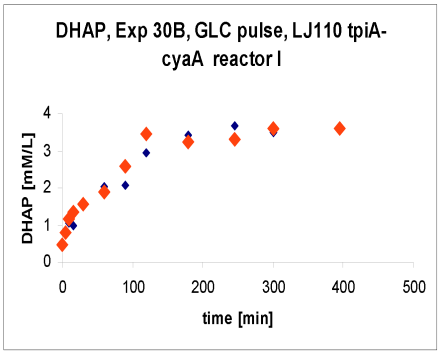

PYR, Exp 30A, GLC pulse, LJ110 tpiA-cyaA reactor II
PYR [mM/L]
time [min]

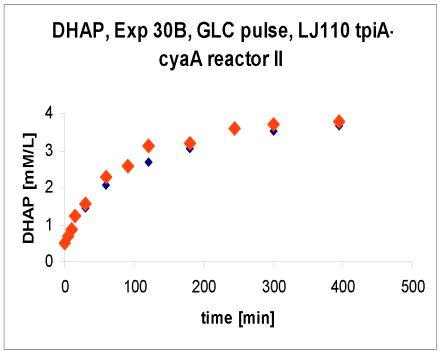
Time	DHAP		GL+G6P		G6P		ATP	
0	0.5589	0.4872	10.3045	10.1995	0.1805	0.1958	0.4534	0.5823
5	0.7431	0.7963	9.7470	9.8807	0.1396	0.1430	0.2827	0.3462
10	1.0542	1.1668	9.8501	10.0640	0.1005	0.1243	0.2113	0.2867
15	0.9969	1.3306	8.9814	9.6649	0.1056	0.1226	0.3383	0.0982
30	1.5844	1.5783	8.1852	8.8516	0.1107	0.1396	0.2470	0.2232
60	2.0409	1.9017	7.9313	8.0802	0.1107	0.1158	0.3165	0.1875
90	2.0757	2.5834	6.9422	7.9370	0.1073	0.1175	0.1379	0.1597
120	2.9478	3.4371	6.7074	6.8430	0.0971	0.1005	0.1935	0.0863
180	3.4207	3.2508	5.0005	5.3461	0.0885	0.0971	0.1091	0.1488
245	3.6704	3.3060	4.0439	3.9370	0.0988	0.0851	0.1667	0.1607
300	3.5005	3.6049	2.6940	2.7532	0.0698	0.0681	0.1548	0.0476
395	3.5292	3.5865	1.4453	1.6115	0.1192	0.1260	0.2143	0.1706

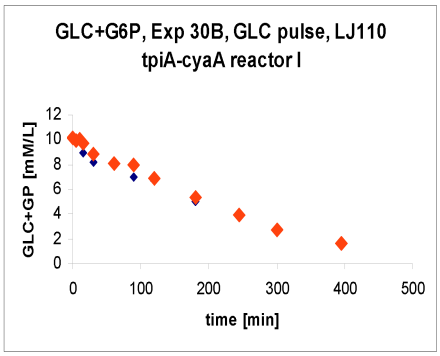
Table B.73: Experiment 30B, 1.15 mM ATP, reactor I

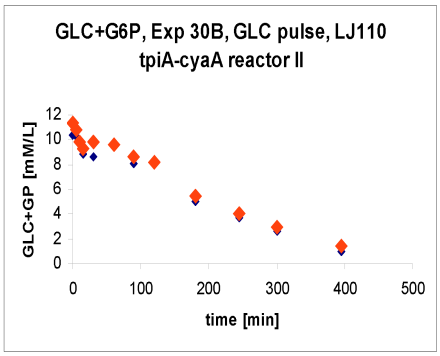
Time	DHAP		GL+G6P		G6P		ATP	
0	0.5404	0.5179	10.3618	11.3699	0.0954	0.0971	0.3244	0.3790
5	0.7083	0.6940	10.3064	10.8143	0.1056	0.1141	0.1617	0.1885
10		0.8823	9.7795	9.8024	0.1277	0.1328	0.1359	0.1409
15	1.1750	1.2242	8.8573	9.2850	0.1158	0.1277	0.1181	0.1607
30	1.4473	1.5763	8.6243	9.8043	0.1022	0.1107	0.0982	0.1468
60	2.0655	2.2968	9.8692	9.6172	0.1260	0.1328	0.1597	0.1825
90	2.6080	2.5977	8.0516	8.6644	0.1294	0.1379	0.1260	0.1250
120	2.7062	3.1341	7.9981	8.2310	0.1362	0.1379	0.1379	0.1746
180	3.0542	3.1955	5.0158	5.4874	0.1192	0.1209	0.0813	-0.0119
245	3.5517	3.5967	3.7079	4.0248	0.1141	0.1124	0.0913	0.0734
300	3.5169	3.7216	2.5967	2.9575	0.0885	0.0903	0.0675	-0.0139
395	3.6643	3.7748	0.9871	1.4014	0.0698	0.0766	0.0179	0.0496

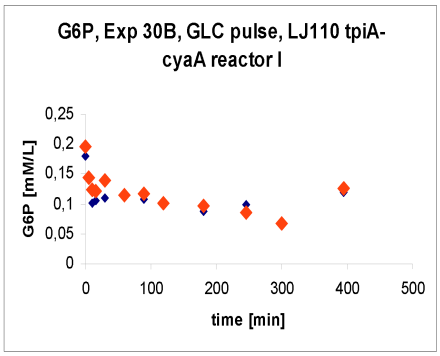
Table B.74: Experiment 30B, 1.15 mM ATP, reactor II

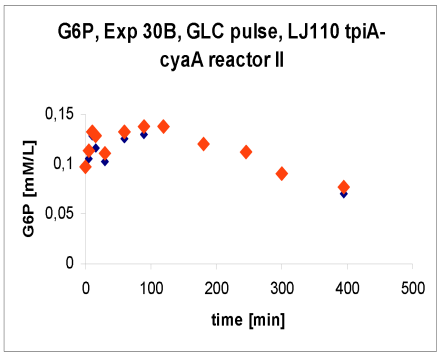

DHAP, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor I
DHAP [mM/L]
time [min]

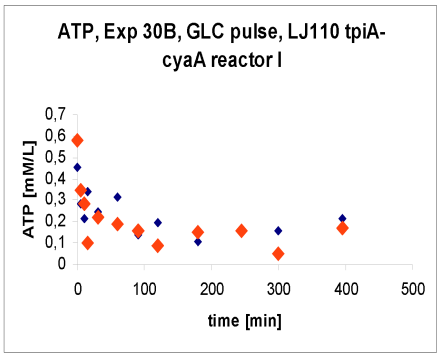

DHAP, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor II
DHAP [mM/L]
time [min]

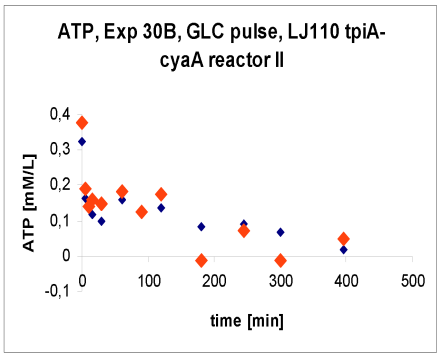

GLC+G6P, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor I
GLC+GP [mM/L]
time [min]


GLC+G6P, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor II
GLC+GP [mM/L]
time [min]


G6P, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor I
G6P [mM/L]
time [min]


G6P, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor II
G6P [mM/L]
time [min]


ATP, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor I
ATP [mM/L]
time [min]


ATP, Exp 30B, GLC pulse, LJ110 tpiA-cyaA reactor II
ATP [mM/L]
time [min]

B.22 Data analysis

This section contains some general observations that could be inferred from the data plots.

B.22.1 Duplicate experiments

Experiments 2B and 3B are in fact duplicates. DHAP maximum concentration in experiment 2B is about 5 [mM/ml] value fairly similar to the one in experiment 3B. The measurements of the concentration of the sum of Glucose and G6P reaches zero in both experiments in the interval between 60 and 80 minutes, displaying a reproducibility of the experiment. Two other experiments which are duplicates are 5A and 4B. In this case the maximum concentration of DHAP is almost the same in the two experiments, the profiles seem similar as well. For the second measurement the profiles look similar, although there were some errors in the first measurement point. Exp 5B and 10B are also duplicates. The DHAP measurements looks in agreements, a sensible maximum higher values can be noticed for Exp 10B while a smaller one for the second measurement (G6P+Glucose). It can be concluded that generally the measurements are reliable.

B.22.2 Hexokinase influence

A part of the experiments investigated the *hexokinase* influence. *Hexokinase* is the enzyme responsible for the reaction converting glucose into glucose-6-phosphate. Experiments 3B, 4A makes the step change in the *hexokinase* concentration for 1 [U/ml] to 0.06 [U/ml]. Comparing the maximum value of the DHAP concentration in the two experiments a sensible increase could be noticed in the experiment with less *hexokinase*. A second series of experiments in which the *hexokinase* concetration was reduced even further from 0.06 [U/ml] to 0.01 [U/ml] is related to experiments 10A-11A and 10B-11B. Comparing again the maximum value of the DHAP concentration in the first two experiments the values are fairly similar considering the variance of the measurements. The same small difference could be noticed in the second series, for the second case the concentration is slightly larger. The *hexokinase* concentration does not seem to affect too much the DHAP final concentration (yield).

B.22.3 Temperature influence

Experiments 3A-B-C are investigating the influence of the temperature upon the system. Analyzing the plots obtained with data from these experiments, clearly the temperature has a positive influence on the reaction rates and looks like for $37°C$ the best yield is obtained. No experiments above this temperature were performed. Starting with this experiment the operating temperature was

chosen to be 37 for the following experiments.

B.22.4 Protein concetration

The influence of the protein concentration can be seen be analyzing series 2 of experiments (exp2A-2C). Clearly, the more proteins are present in the reactor the faster the reactions are. The value of 10 [mg/ml] has been chosen and used further.

B.22.5 ATP influence

ATP cofactor plays an important role in the first reactions producing DHAP, and is being produced in the reactions after DHAP. The series of experiments 4A-4C investigated the yield for various level of ATP. A similar series Exp 11A-Exp 11B, Exp 10A-Exp 10B, and finally the series Exp 17A-Exp 17C. All these experiments revealed that the reactions are highly influenced by the fresh ATP added to the reactor and that there is an unexpected sink of ATP. Exp 13 investigates the ATP behavior in buffer only. Clearly there is no influence. Exp 14 shows the ATP degradation in a solution of buffer and crude extract and crude extract but washed. AtpD was suspected to be responsible for ATP fast consumption. In experiment 15 a comparison using extract from two different mutants, first using **LJ110-tpi** and second using **W3110-atpD**. In experiment 16, ATP, ADP and AMP were measured separately. By studying the evolution of the sum of the three adenosine compounds it can be observed a decreasing curve (not shown). The conclusion was that there is a reaction degrading ATP. They have identified the *adenylate cyclase* to be the responsible reaction and the regeneration cycle could not overcome it. In experiment 17 the reaction responsible for ATP consumption was inhibited with PPi and cAMP. The DHAP yield was almost maximum for the experiment with 11.1 [mM/ml] ATP initial concentration, while for the experiments with less initial ATP the yield was much higher compared with their counterparts experiments.

B.22.6 NAD+ influence

Series Exp 5A-5C presents the results obtained after varying the NAD^+ concentration. NAD^+ is involved in the reaction occurring after DHAP production and plays an important role in the ATP regenerating system. Analyzing the plots corresponding to these experiments it can be concluded that the system is able to regenerate NAD^+ quite well, and in the following experiments a reduced amount of this cofactor was used.

B.22.7 DHAP-aldolase reaction

In some experiments, Ex6A for example it can be noticed that there is some sort of saturation of the reaction producing DHAP. Two extra experiments have

been performed, with a double mutant in order to investigate the influence of an extra reactant which reacts with DHAP and in this way the DHAP concentration will be reduced. In experiment 9A butanal was added in large execess. It can noticed that the DHAP is indeed consumed, while no significant response can be seen in the Glucose+G6P measurement. IN exp 9B after adding extra amounts of aldolase the reaction consuming DHAP is much faster as can be seen in plots from Exp 9B. Once the concentration of free DHAP was reduced (consumed in the reaction), the reactions consuming Glucose and G6P have been reactivated as can be seen from the plot showing the evolution of Glucose and G6P. It can be concluded that the reaction producing DHAP is inhibited by DHAP.

B.22.8 Glucose influence

Glucose is the main reactant in DHAP production and partial investigation was performed about the influence that glucose has on the reaction network. By increasing the amount of Glucose relative to the cofactors a slightly higher DHAP yield was obtained but the marginal yield was insignificant. The conclusion is drawn based on data from Experiments 6A and 12A-B. It seems that the ATP regenerating system is not able to produce enough ATP to keep up with the consumption in the first reactions. The effect is not yet clarified, the saturation on DHAP or possibly on Glucose could also have an influence.

Nomenclature

Symbols

$\hat{F}_m^r$	matrix containing the reaction rates , see equation (3.1), page 54
CR	real values crossover factor, tunning parameter , see equation (5.4), page 100
e_k	white noise process , see equation (2.2), page 22
F	likelihood function , page 108
F	constant tunning parameter , see equation (5.4), page 100
f	RHS unaffected by inputs , see equation (3.3), page 55
G	gradient of the likelihood function , page 108
g	RHS affected by inputs , see equation (3.3), page 55
h	measurement equations vector , see equation (3.3), page 55
I	identity matrix , see equation (3.2), page 55
i	index of state , see equation (3.3), page 55
j	index of inputs , see equation (3.3), page 55
k	arbitrary sampling time
L	Lie derivative
l_o	maximum Lie derivation order , page 59
m	number of inputs , see equation (3.3), page 55
$M(\theta)$	input output map if the model , see equation (3.3), page 55
m_{lo}	maximum Lie derivation order , page 59
N	stoichiometric matrix , page 54
N	number of samples , see equation (5.10), page 102
N_m	stoichiometric matrix of the measured species , page 54
N_m^+	the Moore-Penrose inverse of N_m , see equation (3.1), page 54
N_P	population size , page 99

n_R number of reactions , page 54

n_S number of species , page 54

N_{Su} number of unmeasured species , page 54

n_{Sm} number of measured species , page 54

N_u stoichiometric matrix of the unmeasured species , page 54

N_φ number of optimization variables , page 99

P_G population matrix containing NP vectors , see equation (5.1), page 99

R matrix containing the reaction rates , see equation (3.1), page 54

r_i random indexes used to select the proper value , see equation (5.4), page 100

r_j reaction rate j , page 55

R^i_{kk-1} covariance matrix of the measurements , see equation (2.13), page 24

s vector containing the generating series coefficients , see equation (3.5), page 56

t $[min]$ time

u process inputs vector

$\varphi^{'}_{i,G}$ population matrix of trial candidates , see equation (5.3), page 99

$\varphi^{'}_{j,i,G+1}$ elements of the trials vectors , see equation (5.4), page 100

V parameters neighborhood space , page 55

V posterior covariance of the parameters , see equation (5.7), page 102

$V_{i,G+1}$ mutated vectors , see equation (5.4), page 100

$v_{j,i,G+1}$ elements of the mutated vectors , see equation (5.4), page 100

Vo matrix containing the volume measurements , see equation (3.1), page 54

x process state vector

y process outputs vector

Greek letters

$\hat{\theta}$ process parameters vector , see equation (3.3), page 55

Δ^r identifiability criterion matrix , see equation (3.2), page 55

Δt_{spl}^{max} maximum allowed difference between two successive sampling points , page 98

Δt_{swl}^{max} maximum allowed difference between two successive switching points , page 98

Δt_{spl}^{min} minimum allowed difference between two successive sampling points , page 98

Δt_{swl}^{min} minimum allowed difference between two successive switching points , page 98

ω_t standard Wiener process , see equation (2.2), page 22

Φ experimental design variables vector space , see equation (5.7), page 102

Σ_{pre} prior covariance matrix , see equation (5.8), page 102

φ experimental design variables vector , see equation (5.7), page 102

φ_{ED} expected value design criterion , see equation (5.7), page 102

φ_{WC} worst case design criterion , see equation (5.11), page 102

Σ matrix containing the diffusion terms , see equation (2.2), page 22

σ state independent diffusion term , see equation (2.2), page 22

$\Sigma_{\hat{\theta}}$ covariance matrix of the measurements , see equation (2.13), page 24

Θ model parameter vector space , page 102

θ model parameters vector , see equation (3.3), page 55

θ^* plant parameters vector , see equation (3.3), page 55

References

Asprey, S. and Macchietto, S.: 2000, Statistical tools for optimal dynamic model building, *Computers and Chemical Engineering* **24**, 1261–1267.

Asprey, S. and Macchietto, S.: 2002, Designing robust optimal dynamic experiments, *Journal of Process Control* **12**, 545–556.

Atkinson, A. C. and Donev, A.: 1996, *Optimum Experimental Design*, Clrendon Press Oxford Science Publications.

Audoly, S., Bellu, G., D Angio, L., Saccomani, M. P. and Cobelli, C.: 2001, Global identifiability of nonlinear models of biological systems, *IEEE Transactions on Biomedical Engineering* **48**(1), 55–65.

Bali, M. and Thomas, S. R.: 2001, A modelling study of feedforward activation in human erythrocyte glycolysis, *C.R. Acad. Sci. Paris, Sciences de la vie / Life Sciences* pp. 185–199.

Bauer, I., Bock, H. G., Körkel, S. and Schlöder, J. P.: 2000, Numerical methods for optimum experimental design in dae systems, *Journal of Computational and Applied Mathematics* **120**, 1–25.

Benabbas, L., Asprey, S. P. and Macchietto, S.: 2005, Curvature-based methods for designing optimally informative experiments in multiresponse nonlinear dynamic situations, *Ind. Eng. Chem. Res.* **44**, 7120–7131.

Bonvin, D. and Rippin, D. W. T.: 1990, Target factor aanalysis for the identification of stoichiometric models, *Chemical Engineering science* **45**(12), 3417–3426.

Brendel, M., Bonvin, D. and Marquardt, W.: 2006, Incremental identification of kinetic models for homogeneous reactions systems, *Chemical Engineering Science* **61**, 5404–5420.

Brun, R., Kühnia, M., Siegrist, H., Gujer, W. and Reichert, P.: 2002, Practical identifiability of asm2d parametersãsystematic selection and tuning of parameter subsets, *Water Research* **36**, 4113–4127.

Chassagnole, C., Noisommit-Rizzi, N., Schmidt, J. W., Mauch, K. and Reuss, M.: 2002, Dynamic modeling of the central carbon metabolism of escherichia coli, *Biotechnology and Bioengineering* **79**(1), 53–73.

Chen, B. H. and Asprey, S. P.: 2003, On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations, *Industrial Engineering and Chemical Research* **42**, 1379–1390.

Chen, B. H., Bermingham, S., Neumann, A. H., Kramer, H. J. M. and Asprey, S. P.: 2004, On the design of optimally informative experiments for dynamic crystallization process modeling, *Ind. Eng. Chem. Res.* **43**, 4889–4902.

Dochain, D., Vanrolleghem, P. A. and Van Daele, M.: 1995, Structural identifiability of biokinetic models of activated sludge respiration, *Water Research* **29**(11), 2571–2578.

Fliess, M.: 1980, Generating series for discrete-time nonlinear systems, *IEEE Transcations on automatic control* **25**(5), 984–985.

Fliess, M. and Lagarrigue, L.: 1980, Functional expansions in nonlinear-system theory, *Scietific papers of the institute of technical cybernetics of the Technical University of Wroclaw.*

Fliess, M., Lamnabhi, M. and Lamnabhi-Lagarrigue, F.: 1983, An algebraic approach to nonlinear functional expansions, *IEEE Transactions on circuits and systems* **CAS-30**(8), 554–570.

Gadkar, K. G., Gunawan, R. and Doyle III, F. J.: 2005, Iterative approach to model identification of biological networks, *BMC Bioinformatics* **6**, 154–174.

Hanson, T. L. and Fromm, H. J.: 1967, Rat skeletal muscle hexo.kinase ii. kinetic evidence for a second hexokinase in muscle tissues, *The Journal of Biological Chemistry* **242**(3), 501–508.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. and Kummer, U.: 2006, Copasi - a complex pathway simulator, *Bioinformatics* **22**, 3067–3074.

Jazwinski, A. H.: 1970, *Stochastic Processess and Filtering Theory*, Academic Press, New York, USA.

Kedem, G.: 1980, Automatic differentiation of computer programs, *ACM Transactions on Mathematical Software* **6**(2).

Kloeden, P. E. and Platen, E.: 1999, *Numerical Solution of Stochastic Differential Equations*, Springer.

Kontoravdi, C., Asprey, S. P., Pistikopoulos, E. N. and Mantalaris, A.: 2005, Application of global sensitivity analysis to determine goals for design of experiments: An example study on antibody-producing cell cultures, *Biotechnol. Prog.* **21**, 1128–1135.

Körkel, S., Bauer, I., Bock, H. G. and Schlöder, J. P.: 1999, A sequential approach for nonlinear optimum experimental design in dae systems, *Proc. of Int. Workshop on Scientific Computing in Chemical Engineering* .

Kremling, A., Fischer, S., Sauter, T., Bettenbrock, K. and Gilles, E.: 2004, Time hierarchies in the escherichia coli carbohydrate uptake and metabolism, *BioSystems* **73**, 57–71.

Kristensen, N. R., Madsen, H. and Jørgensen, S. B.: 2004a, A method for systematic improvement of stochastic grey-box models, *Computers and Chemical Engineering* **28**, 1431–1449.

Kristensen, N. R., Madsen, H. and Jørgensen, S. B.: 2004b, Parameter estimation in stochastic grey-box models, *Automatica* **40**, 225–237.

Kukkonen, S. and Lampinen, J.: 2005, Gde3: The third evolution step of the generalized differential evolution, *The 2005 IEEE Congress on Evolutionary Computation*, IEEE, pp. 443–450.

Kukkonen, S. and Lampinen, J.: 2006, Constrained real-parameter optimization with generalized differential evolution, *The 2006 IEEE Congress on Evolutionary Computation*, IEEE, pp. 207–214.

Lampinen, J.: 2002, A constraint handling approach for the differential evolution algorithm, *The 2002 IEEE Congress on Evolutionary Computation*, IEEE, pp. 1468–1473.

Ljung, L. and Glad, T.: 1994, On global identifiability for arbitrary model parametrizations, *Automatica* **30**(2), 265–276.

Madsen, H. and Holst, J.: 1995, Estimation of continuous-time models for the heat dynamics of a building, *Energy and Buildings* **22**, 67–79.

Madsen, H. and Melgaard, H.: 1991, The mathematical and numerical methods used in ctlsm, *Technical Report 7*, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Melgaard, H. and Madsen, H.: 1993, Ctlsm - a program fro parameter estimation in stochastic differential equations, *Technical Report 1*, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, B. and Madsen, H.: 1998, Identification of a linear continuous time stochastic model of the heat dynamics of a greenhouse, *J. agric. Engng Res.* **71**, 249–256.

Ning, J., Purich, D. L. and Fromm, H. J.: 1969, Studies on the kinetic mechanism and allosteric nature of bovin brain hexokinase, *The Journal of Biological Chemistry* **244**(14), 3840–3846.

Peeters, R. L. and Hanzon, B.: 2005, Identifiability of homogeneous systems using the state isomorphism approach, *Automatica* **41**, 513–529.

Petersen, B.: 2000, *Calibration, Identifiability and Optimal Experimental design of Activated Sludge Models*, PhD thesis, Faculteit Landbouwkunige en Toegepaste Biologische Wetenschappen, Gent University, Belgium.

Pohjanpalo, H.: 1978, System identifiability based on the power series expansion of the solution, *Mathematical Biosciences* **41**, 21–33.

Price, K. V.: 1999, *New Ideas in Optimization*, McGraw-Hill, chapter An Introduction to Diferential Evolution, pp. 79–103.

Price, K. V., Storn, R. and Lampinen, J. A.: 2005, *Differential Evolution: A Practical Approach to Global Optimization*, Springer.

Rao, C.: 1973, *Linear statistical inference and its applications*, Whiley, New York.

Rizzi, M., Baltes, M., Theobald, U. and Reuss, M.: 1996, In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: Ii. mathematical model, *Biotechnology and Bioengineering* **55**(4), 592–608.

Rodriguez-Fernandez, M., Kucherenko, S., Pantelides, C. and Shah, N.: 2007, Optimal experimental design based on global sensitivity analysis, *in* V. Plesu and P. Agachi (eds), *17th European Symposium on Computer Aided Process Engineering - ESCAPE 17*, Elsevier B.V.

Saccomani, M. P.: 2004, Some results on parameter identification of nonlinear systems, *Cardiovascular Engineering: An International Journal* **4**(1), 95–102.

Saccomani, M. P., Audoly, S. and d' Angio, L.: 2003, Parameter identifiability of nonlinear systems: the role of initial conditions, *Automatica* **39**, 619–632.

Sadegh, P., Holst, J., Madsen, H. and Melgaard, H.: 1995, Experiment design for grey box identification, *International Journal of Adaptive Control and Signal Processing* **9**(6), 491–507.

Sadegh, P., Meelgaard, H., Madsen, H. and Holst, J.: 1994, Optimal experiment design for identification of grey-box models, *American Control Conference* .

Sarmiento Ferrero, C., Chai, Q., Diez, M. D., Amrani, S. H. and Lie, B.: 2006, Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data, *Modeling, Identification and Control* **27**.

Schümperli, M., Pellaux, R. and Panke, S.: 2007, Chemical and enzymatic routes to dihydroxyacetone phosphate, *Appl Microbiol Biotechnol* **75**, 33–45.

Sidoli, F. R., Mantalaris, A. and Asprey, S. P.: 2005, Toward global parametric estimability of a large-scale kinetic single-cell model for mammalian cell cultures, *Ind. Eng. Chem. Res.* **44**, 868–878.

Spall, J. C.: 2003, Monte carlo-based computation of the fisher information matrix in nonstandard settings, *Proceedings of the American Control Conference*, IEEE.

Spall, J. C.: 2004, Cramer-rao bounds and monte carlo calculation of the fisher information matrix in difficult problems, *Proceeedings of the 2004 American Control Conference*, IEEE.

Spall, J. C.: 2005, Monte carlo of the fisher information matrix in nonstandard settings, *Journal of Computational and Graphical Statistics* **14**(4), 889–909.

Storn, R.: 1996, On the usage of differential evolution for function optimization, *IEEE* pp. 519–523.

Storn, R.: 1997, Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization* **11**, 341–359.

Storn, R. and Price, K.: 1995, Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces, *Technical Report 95-012*, International Computer Science Institute, Berkeley. available on-len at *http://www.icsi.berkeley.edu/techreports/1995.abstracts/tr-95-012.html*.

Storn, R. and Price, K.: 1996, Minimizing the real functions of the icec'96 contest by differential evolution, *IEEE* pp. 842–844.

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., van der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., van Dam, K., Westerhoff, H. V. and Snoep, J. L.: 2000, Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry, *European Journal of Biochemistry* **267**, 5313–5329.

Tornøe, C. W., Jacobsen, J. L., Pedersen, O., Hansen, T. and Madsen, H.: 2004, Grey-box modelling of pharmacokinetic /pharmacodynamic systems, *Journal of Pharmacokinetics and Pharmacodynamics* **31**(5), 401–417.

Vajda, S., H., R., Walter, E. and Lecourtier, Y.: 1989, Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models, *Chemical Engineering Community* **83**, 191–219.

Viola, R. E., Raushe, F. M., Rendina, A. R. and Cleland, W. W.: 1982, Substrate synergism and the kinetic mechanism of yeast hexokinase, *Biochemistry* **21**, 1295–1302.

Walter, E., Braems, I., Jaulin, L. and Kieffer, M.: 2004, *Numerical Software with Result Verification*, Springer Berlin / Heidelberg, chapter Guaranteed Numerical Computation as an Alternative to Computer Algebra for Testing Models for Identifiability, pp. 124–131.

Walter, E. and Pronzato, L.: 1987, *Identifiability of Parametric Models*, Pergamon, Oxford, chapter Robust experiment design: between qualitative and quantitative identifiabilities, pp. 104–113.

Walter, E. and Pronzato, L.: 1990, Qualitative and quantitative experiment design for phenomenological models-a survey, *Automatica* **26**(2), 195–213.

Walter, E. and Pronzato, L.: 1996, On the identifiability and distinguishability of nonlinear parametric models, *Mathematics and Computers in Simulation* **42**, 125–134.

Walter, E. and Pronzato, L.: 1997, *Identification of Parametric Models from Experimental Data*, Springer. Translated by authors with help from Jhon Norton.

Zewe, V. and Fromm, H. J.: 1965, Kinetic studies of rabbit muscle lactate dehydrogenase. 11. mechanism of the reaction, *Biochemistry* **4**(4), 782–792.

Zgiby, S. M., Thomson, G. J., Qamar, S. and Berry, A.: 2000, Exploring substrate binding and discrimination in fructose 1,6-bisphosphate and tagatose 1,6-bisphosphate aldolases, *European Journal of Biochemistry* **267**, 1858–1868.