



# Dynamisk modellering av hur feedback via cytosoliskt Mig1 reglerar *SUC2* i SNF1-signalvägen

Dynamic modelling of feedback through cytosolic Mig1 and its regulation of *SUC2* in the SNF1 signaling pathway

*Kandidatarbete inom civilingenjörsutbildningen vid Chalmers*

Sophia Axillus

Leo Benson

Viktor Hasselgren

Frida Moberg



# Dynamisk modellering av hur feedback via cytosoliskt Mig1 reglerar *SUC2* i SNF1-signalvägen

*Kandidatarbete i matematik inom civilingenjörsprogrammet Bioteknik vid Chalmers*

Sophia Axillus      Leo Benson  
Viktor Hasselgren      Frida Moberg

Handledare: Johannes Borgqvist  
Marija Cvijovic  
Sebastian Persson

Institutionen för Matematiska vetenskaper  
CHALMERS TEKNISKA HÖGSKOLA  
GÖTEBORGS UNIVERSITET  
Göteborg, Sverige 2020



## Förord

Det här är ett kandidatarbete utfört av Sophia Axillus, Leo Benson, Viktor Hasselgren och mig, Frida Moberg, samtliga läser på civilingenjörsprogrammet inom Bioteknik vid Chalmers tekniska högskola. Projektet är utformat av Cvijovic Lab och utfört på Institutionen för Matematiska vetenskaper.

Redan under hösten 2018 stötte vi på Johannes Borgqvist som övningsledare i kursen Tillämpad matematik, han nämnde då att hans forskargrupp erbjöd möjligheten att skriva kandidatarbete hos dem. Sophia och Leo blev intresserade och bestämde med Johannes att de skulle ha ett möte när valet av kandidatarbete närmade sig. Ett år gick och när det var dags för möte hade även jag hört om det spännande projektet som vankades och följde med på ett möte som kom att övertyga oss alla om att det var det projektet vi ville göra. Några månader senare fick vi våra kandidatgrupper och då stod även Viktor på listan och vår grupp var komplett!

Sedan tog arbetet fart och det märktes snabbt att vi utgjorde en grupp där alla hade lika höga ambitioner och många gånger samma målbild. Framför allt ville vi få chansen att utveckla våra programmeringskunskaper. För att vara effektiva delades uppgifterna in i fyra områden där varje person ansvarade för ett område enligt:

|                                    |        |
|------------------------------------|--------|
| Biologin kring signalvägen SNF1    | Leo    |
| ODE-modeller och Stabilitetsanalys | Frida  |
| Identitets- och Känslighetsanalys  | Sophia |
| Optimering                         | Viktor |

Även under skrivandet av rapporten har vi utgått från ovanstående fördelning. Vid sidan av rapporten har vi fört en loggbok över våra individuella prestationer. Trots uppdelningen har vi genom hela arbetsgången haft ett nära samarbete för att alla ska känna sig delaktiga i hela projektet. Under perioden då vi läste in oss på respektive område hade vi till exempel bokklubb varje vecka där vi kort presenterade det viktigaste av det vi läst. Till och med när pandemin Covid-19 slog mot Sverige och stoppade all platsundervisning på Chalmers bibehöll vi en tät kontakt genom korta dagliga mötet via nätet. Dessutom har vi vid flera tillfällen jobbat tillsammans då områdena gått in i varandra, som att Leos kunskaper om signalvägen krävdes när jag skulle sätta ihop ODE-modellerna och Sophia och Viktor som har skrivit ihop sina koder i ett program.

Utöver samarbetet inom gruppen har vi haft ett mycket nära samarbete med våra handledare Johannes Borgqvist, Sebastian Persson och Marija Cvijovic på Cvijovic Lab, som hjälpt och väglett oss genom arbetet. Deras dörr har alltid stått öppen, de har alltid svarat snabbt på mail och dessutom har vi varit med på deras veckoliga gruppmöten där vi fått tips från hela forskargruppen. Det har inte bara varit väldigt hjälpsamt utan också väldigt roligt och fått oss att känna oss som en del av forskargruppen och för det vill vi framföra ett stort tack.

Det som vi presenterar i den här rapporten är en viktig men än dock liten del av det vi lärt oss under den här våren. Vi har lärt oss så mycket mer, som att hantera nya verktyg inom modellering, koda i Python samt fått en bild av hur det kan se ut att arbeta med forskning, något som vi definitivt kommer ta med oss i framtiden.

## Populärvetenskaplig presentation

Vad kan vi göra för att leva ett långt liv? Inte bara sett till antal år, utan hur kan vi få många friska och glada år? Idag lider många av åldersrelaterade sjukdomar på grund av vår långa livslängd och det kan försämra livskvaliteten något enormt. Det är därför av stort intresse bland forskare och inom sjukvården att bättre förstå mänskligt åldrande.

För att få mer förståelse för denna fråga behöver vi veta hur åldrande fungerar på cellnivå. Anledningen till det är att om vi vill förstå åldrande hos en population behöver vi veta hur individerna lever och för att förstå individen behöver vi kunskap om tillståndet hos cellerna. Det behövs alltså kännedom om de cellulära komponenterna för att förstå åldrande i grund och botten. Det är känt att åldrande hos celler handlar om cellens välstånd; vilket kan beskrivas som ansamlingen av skador hos cellen. Ordet "skada" syftar här på ändringar i DNA som i sin tur betyder förändring av proteiner som kan ge en nedsättning av cellulära funktioner och på sikt leda till att cellen dör [1]. Det är en tydlig process men det är däremot svårt att definiera vilka funktioner och mekanismer som är involverade i åldrandet.

I samhället idag har vi alla hört debatterna kring hur människan på bästa sätt kan öka sin livskvalité och livslängd. Problemet i många fall är att handlingar som bidrar positivt också kan ge negativa konsekvenser, till exempel är solljus avgörande för vårt välmående men det ökar även vår risk för hudcancer. Anledningen till varför det finns många motstridande argument är för att det inte alltid är självklart hur mekanismerna fungerar. Antalet unika proteiner som har upptäckts hos människor är nästan 6000, varav flera kan interagera med hundratals andra proteiner [2]. Det blir snabbt svårt att hitta mekanismerna som ansvarar för specifika reaktioner som forskare vill undersöka. För att studera denna komplexitet krävs förenklingar. En vanlig förenkling är att använda sig av något som kallas modellorganism, en enklare organism som kan verka som modell för ett tillstånd eller sjukdom hos människor.

En ofta använd modellorganism är bagerijäst, *Saccharomyces cerevisiae* (*S. cerevisiae*) [3]. Genom studier av *S. cerevisiae* har den encelliga organismen visat sig vara ett ovärderligt verktyg för att utveckla forskningsfältet genetik och bredda vår förståelse av eukaryoter (celler som har cellkärna) [3]. Jäst används ofta som modell för mänskligt åldrande och är fördelaktig på många sätt eftersom den, precis som människor, åldras och dör. Dessutom är en livscykel hos jäst relativt kort jämfört med om åldrandeprocessen hos ett djur skulle undersökas.

Åldrande sker hos så gott som alla organismer och många av de processer som är livsavgörande på cellnivå är välbevarade genom evolutionen, inte minst de som kallas näringsigenkännande processer [4]. Det vill säga cellens förmåga att hantera bristen eller överskottet av näring. Cellens förmåga att reagera på miljöförändringar involverar det som kallas för signalvägar, givet en yttre signal ges en respons i genuttryck och fler eller färre antal av ett visst protein bildas. Näringsigenkännande är en livsviktig process och kan ses som basen till evolutionära ändringar hos celler vid knappa förhållanden [4]. Det har även visats att det är involverat i flera sjukdomar såsom cancer och ämnesomsättningsjukdomar [4]. Genen som ansvarar för näringsigenkännande kallas för *AMPK* hos människor och *SNF1* i jäst. Förutom dess koppling till sjukdomar så har ökad AMPK-aktivitet kopplats till en ökad livslängd eftersom genen även styr resistens mot miljörelaterad stress, vilket orsakar cellulära skador och därmed också att vi åldras [5]. I detta arbete undersöks möjliga reaktionsvägar som SNF1-signalvägen har i jäst. Arbetet kommer innefatta reaktionsvägar och proteiner involverade längre ned i signalvägen men trots det måste vi veta hur andra bitar samspelar och hur allting börjar.

Huvudpersonen i SNF1-signalvägen är proteinet SNF1 och det ansvarar för en omställning i cellens ämnesomsättning så att den kan leva på alternativa kolkällor [6]. Jäst föredrar de enkla sockerarterna glukos och fruktos som huvudkälla för energi men kan även växa på andra varianter. Två exempel är maltos, som består av två sammansatta glukosmolekyler, och sackaros, som består av en glukosmolekyl bunden till en fruktosmolekyl. Att använda dem kräver dock lite mer jobb då de först behöver spjälkas. Vid spjälkningen använder sig cellen av enzymer, proteiner som katalyserar kemiska reaktioner, till exempel att klippa isär två molekyler. För att cellen ska spara energi finns inte alltid de enzymer som är nödvändiga för spjälkningen, utan cellen tillverkar dem endast vid behov. Det här behovet känner cellen då glukoshalterna sjunker drastiskt och då aktiveras SNF1-signalvägen. Allting börjar med proteinet SNF1, som vid aktivering kan liknas vid en marionettspelare där själva cellen är dess docka. När proteinet aktiveras styr det genuttrycket genom transkriptionsfaktorer. Transkriptionsfaktorer ansvarar i sin tur för specifika gensekvenser och avgör om dessa uttrycks eller inte. I fallet med sjunkande glukoshalter leder det till att enzymer som kan hantera

de alternativa kolkällorna börjar produceras. En utav transkriptionsfaktorerna i signalvägen kallas för Mig1 och den styr uttrycket av, bland annat, genen *SUC2* [7]. När genen är aktiverad uttrycker *SUC2* ett enzym som behövs för spjälkning av sackaros. Då glukoshalten sänks drastiskt kan det antas vara självklart att genuttrycket av *SUC2* borde öka. Det intressanta är att ökningen bara sker temporärt. Experiment har visat att uttrycket sjunker igen efter en viss tid och det är inte känt varför. Baserat på det som påvisats av tidigare forskning borde enzymet fortsätta bildas eller stanna vid ett högt uttryck på grund av Mig1-aktiviteten. Hur kan vi få information om vad som egentligen händer? Komplexiteten hos signalvägar är att vi endast har indata (yttre stimulans, exempelvis glukoskoncentration) och en utsignal (ändrat genuttryck i cellen, exempelvis proteinaktivitet), men vad som händer där emellan är obestämt. Eftersom signalvägar inte kan ge mycket data om vad som händer inom cellen behövs alternativa metoder för vidare undersökning.

Det kan bli en stor utmaning att se mellanstegen genom experimentella försök då vi inte har data, men här är matematiken ett starkt verktyg [1]. Den ger oss möjligheten att ta hänsyn till flera interaktioner samtidigt och det låter oss förstå processer som annars är omöjliga att åskåda. Med hjälp av matematisk modellering kan vi visualisera en process genom att göra teoretiska antaganden och sedan utvärdera resultaten med hjälp av statistik [1, 8]. Rent praktiskt sätts matematiska samband upp baserat på biologiska antaganden, vilket ger oss ett system av ekvationer. Ekvationerna innehåller konstanter som inte är kända och då används det som kallas parameteruppskattning. Genom parameteruppskattning passas en teoretisk modell mot experimentella data genom att finna den bästa passformen och minimera avvikelser mot datapunkterna. Det görs med hjälp av datorer och olika typer av algoritmer. Passningen behöver sedan utvärderas med hjälp av statistik för att se om modellen kan användas för att förutspå nya resultat eller om den behöver justeras. Här är det viktigt att ta hänsyn till att det handlar om modeller, inte verkligheten. Matematiken kan vara ett bra verktyg för att hjälpa oss förstå och ge en bild över hur det en process kan se ut men den ger inte alltid en rättvis bild över hur verkligheten faktiskt ser ut.

Det här kandidatarbetet undersöker om Mig1 kan vara involverad i det minskade *SUC2*-uttrycket genom att sätta upp matematiska modeller. Datan som används är från experimentella försök på jäst där cellerna utsätts för en plötslig sänkning av glukoshalten. Arbetet görs för att kunna bidra med insikter om AMPK/SNF1-signalvägen som kan leda till nya experimentella försök och i vidare forskning bidra med kunskap om människors åldrande. Men det finns betydligt fler gener som kan och kommer att undersökas med hjälp av matematisk modellering i takt med att tekniken blir bättre i rasande fart. Så vad kan vi göra för att uppnå ett hållbart åldrande? Det återstår att se, men med hjälp av modellering och modellorganismer är vi på god väg att ta reda på det.

### Sammanfattning

Den näringsigenkännande SNF1-signalvägen är viktig för att bagerijäst, *Saccharomyces cerevisiae*, ska kunna anpassa sig till olika näringsförhållanden. Signalvägen involverar transkriptionsfaktorn Mig1 som repressar genen *SUC2* vid höga glukosnivåer. Därför förväntas en ökning av enzymet Suc2 som genen producerar när glukosnivån sjunker, experimentella data visar att ökningen sker, men också att den endast är temporär. I det här projektet undersöks hypotesen att minskningen av genuttrycket är genom feedback-inhibering av cytosoliskt Mig1. Undersökningen görs genom att skapa olika matematiska modeller innehållande ordinära differentialekvationer för att beskriva experimentell data där jästcellerna utsätts för en plötslig sänkning av extracellulärt glukos.

Tre modeller togs fram; en fullständig med alla komponenter som ansågs vara essentiella och två reducerade varianter där olika hastighetskonstanter antogs vara lika. Modellerna ställdes upp under biologiska antaganden och enligt massverkans lag. Strukturell identifierbarhetsanalys utfördes som ett första steg och den fullständiga modellen var oidentifierbar. Det vill säga att skattningarna från en parameteruppskattning inte är trovärdiga, och därför gjordes vidare analyser på de två reducerade modellerna, vilka var strukturellt identifierbara. En stabilitetsanalys utfördes även på modellerna och det kunde konstateras att den fullständiga modellen samt en av de reducerade var stabila med potential för att oscillera. Efter parameteruppskattning via Gauss-Newtons metod kunde de passas mot datan och här visade samtliga modeller en bra passning för Suc2 men inte för Mig1.

Validering av modellerna skedde sedan med hjälp av residual-, känslighet-, och numerisk identifierbarhetsanalys. Resultatet visade att ingen av de reducerade modellerna kan användas för prediktioner då parametrarna hade stora varianser samt höga korrelationer och därför är deras skattningar inte trovärdiga. Slutsatsen som drogs var därför att dessa modeller inte kan användas för att bekräfta eller dementera om cytosoliskt Mig1 skulle inhibera *SUC2*.

## Abstract

The nutrient sensing SNF1 signaling pathway in baker's yeast, *Saccharomyces cerevisiae*, is important in order for the cells to adjust to changes in nutrient availability. The pathway involves Mig1, a transcription factor which represses the gene *SUC2* when glucose levels are high. Therefore it is suspected that the *SUC2* expression will increase when glucose levels drop. However, our experiments show that this increase only occurs temporarily. This project investigates the hypothesis that the sudden decrease is caused by cytosolic Mig1. It is examined by creating different mathematical models that consists of a system of ordinary differential equations. These equations are then used in order to describe experimental data where the yeast cells have been exposed to a sudden drop in extracellular glucose level.

The biological analysis resulted in three different models. The first one containing all components that were deemed essential and two reduced variants where two different rate constants were assumed to be equal. The models were constructed according to biological reasonable assumptions and in line with the law of mass action. As a first step, structural identifiability analysis was conducted and the result was that the first model was unidentifiable (parameters could not be fitted correctly) and therefore further analyses were only applied to the reduced models which were structurally identifiable. A stability analysis was also carried out and the result was that the complete model as well as one of the reduced variants were stable after a period of time and possess the ability to oscillate under certain conditions. After parameter estimations with the Gauss-Newton method, the models could be fitted against the data and it resulted in a good fit for *Suc2* but a bad fit for *Mig1*.

A validation of models followed and was carried out by using residual, sensitivity and numerical identifiability analysis. The result showed that none of models could be used to make predictions of new results. This is because the estimated parameters had large variances and high correlations. The conclusion from this was that these models could neither be used to determine or contradict that cytosolic Mig1 inhibits *SUC2*.

## Ord och begrepp

**ATP:** en högennergiförening cellen använder som energi för intracellulära processer

**Cytoplasma:** det som återfinns inuti cellen, vilket innefattar cellkomponenter och *cytosolen* (cellvätska)

**Enzym:** proteiner som katalyserar kemiska reaktioner

**Fosforylering:** aktivering/deaktivering av ett protein genom bindning till en fosfatgrupp, motsatsen *defosforylering* är när fosfatgruppen avlägsnas

**Genuttryck:** en process där informationen av en gen används för att bilda en produkt

**Inhibering:** hämning av en reaktionsväg eller repression av en gen som leder till att mindre produkt bildas, vid *Feedback-inhibering* orsakas inhiberingen av produkten som bildas i den hämmade reaktionsvägen

**Jämvikt:** när ett system, till exempel en kemisk reaktion, uppnått ett läge som är oförändrat över tid, motsvarar engelskans *steady state*

**Nukleus:** en avgränsad del av cellen där arvsmassan finns, även kallad *cellkärna*

**Ortolog:** proteiner med samma funktion i olika arter

**Repressor:** ett protein som hindrar avläsning av en specifik DNA-sekvens

***Saccharomyces cerevisiae*:** det latinska namnet för bagerijäst, en encellig eukaryot organism

***SUC2*:** *SUC2*-genen i *Saccharomyces cerevisiae*

**Suc2:** uttrycket av *SUC2*, proteinet som bildas är ett enzym som katalyserar nedbrytningen av alternativa kolkällor

**Systembiologi:** ett interdisciplinärt område mellan biologi, matematik och datavetenskap

**Tillstånd:** en godtycklig aktör i en matematisk modell, till exempel ett protein eller en gen

# Innehåll

|          |                                                                                                                         |           |
|----------|-------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Inledning</b>                                                                                                        | <b>1</b>  |
| 1.1      | Syfte . . . . .                                                                                                         | 2         |
| <b>2</b> | <b>Teori</b>                                                                                                            | <b>2</b>  |
| 2.1      | SNF1-signalvägen . . . . .                                                                                              | 3         |
| 2.2      | Generella principer för modellering . . . . .                                                                           | 3         |
| 2.3      | Strukturell identifierbarhet . . . . .                                                                                  | 4         |
| 2.4      | Optimering . . . . .                                                                                                    | 5         |
| 2.5      | Stabilitetsanalys . . . . .                                                                                             | 5         |
| 2.6      | Numerisk identifierbarhet . . . . .                                                                                     | 7         |
| 2.7      | Känslighetsanalys . . . . .                                                                                             | 7         |
| <b>3</b> | <b>Data och metod</b>                                                                                                   | <b>8</b>  |
| 3.1      | Data . . . . .                                                                                                          | 8         |
| 3.2      | Strukturell identifierbarhet . . . . .                                                                                  | 8         |
| 3.3      | Parameteruppskattning med Gauss-Newtons metod . . . . .                                                                 | 8         |
| 3.3.1    | Avvikelse från Gauss-Newtons metod . . . . .                                                                            | 9         |
| 3.3.2    | Struktur på optimeringsalgoritmen . . . . .                                                                             | 10        |
| 3.3.3    | Beskrivning av optimering . . . . .                                                                                     | 10        |
| 3.4      | Stabilitetsanalys, numerisk identifierbarhetsanalys och känslighetsanalys . . . . .                                     | 10        |
| <b>4</b> | <b>Resultat</b>                                                                                                         | <b>11</b> |
| 4.1      | Den underliggande mekanismen för Modell 1 består av fosforylering, defosforylering och en feedback-loop . . . . .       | 11        |
| 4.1.1    | Dynamiken för Modell 1 är stabil och kan antingen oscillera eller gå mot styv jämvikt . . . . .                         | 13        |
| 4.1.2    | Givet data för Suc2 och Mig1 är inte nedbrytningen av Mig1 identifierbar . . . . .                                      | 13        |
| 4.1.3    | Störst inverkan har nedbrytnings- och inhiberingstermer . . . . .                                                       | 14        |
| 4.2      | Den konstruerade modellen reducerades till att ha en likvärdig nedbrytning av Mig1 och X . . . . .                      | 14        |
| 4.2.1    | $\hat{\theta}_1 = 0$ och därför visar stabilitetsanalysen att dynamiken för Modell $k_6 = k_7$ inte är stabil . . . . . | 15        |
| 4.2.2    | Stor varians hos parametrarna för Mig1 . . . . .                                                                        | 15        |
| 4.2.3    | Den observerade datan är okänslig för ändring av parametrar . . . . .                                                   | 16        |
| 4.3      | En alternativ reducering testades där SUC2 och X har en likvärdig inhibering . . . . .                                  | 16        |
| 4.3.1    | Dynamiken hos Modell $k_3 = k_5$ är stabil och kommer oscillera kring jämviktsläget . . . . .                           | 17        |
| 4.3.2    | Den reducerade modellen är strukturellt identifierbar men inte numeriskt identifierbar . . . . .                        | 18        |
| 4.3.3    | Känslighetsanalys för den reducerade modellen visar en känslighet för ändring av $\hat{\theta}_2$ . . . . .             | 18        |
| <b>5</b> | <b>Diskussion</b>                                                                                                       | <b>19</b> |
|          | <b>Referenser</b>                                                                                                       | <b>21</b> |
| <b>6</b> | <b>Appendix</b>                                                                                                         | <b>23</b> |
| A        | Härledning av kostnadsfunktion . . . . .                                                                                | 23        |
| B        | Härledning av Newton-Raphsons metod . . . . .                                                                           | 25        |
| C        | Härledning av Gauss-Newtons metod . . . . .                                                                             | 26        |
| D        | Data . . . . .                                                                                                          | 28        |
| E        | Test av optimeringsalgoritm på testmodell . . . . .                                                                     | 29        |
| F        | Utförd stabilitetsanalys på samtliga modeller . . . . .                                                                 | 31        |
| F.1      | Modell 1 . . . . .                                                                                                      | 31        |
| F.2      | Modell $k_6 = k_7$ . . . . .                                                                                            | 33        |
| F.3      | Modell $k_3 = k_5$ . . . . .                                                                                            | 34        |
| G        | Utförd strukturell identifierbarhetsanalys på samtliga modeller . . . . .                                               | 35        |
| G.1      | Modell 1 . . . . .                                                                                                      | 35        |
| G.2      | Modell $k_6 = k_7$ . . . . .                                                                                            | 35        |
| G.3      | Modell $k_3 = k_5$ . . . . .                                                                                            | 36        |

# 1 Inledning

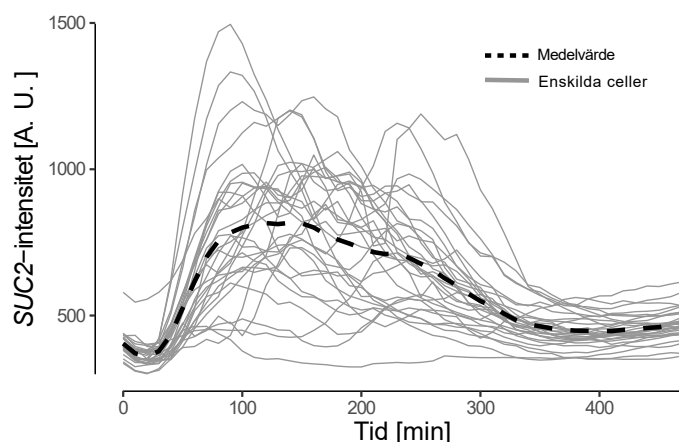
Mekanismerna för åldrande visar sig hos nästan alla levande organismer och ofta på olika sätt, men åldrandets definition är inte alltid helt självklar. Den generella och svepande definitionen av åldrande som "ansamling av skada med tid" kommer från det åldrande som observeras vardagligen, och är starkt förknippat med förlust av funktion. Vare sig det gäller skador på DNA-nivå, proteinnivå, cellnivå eller organismnivå leder skadorna med tid oundvikligen till avvikelse från normala funktioner, och därmed åldrande [1]. Det kan från denna definition resoneras att åldrande, ansamling av skada, är resultatet av de underliggande och flyktiga rötterna till uppkomsten och bibehållelsen av skadan.

Sökandet efter åldrandets rötter är en av många forskningsområden som för systembiologerna och matematikerna har lett in på studiet av signalvägar. En signalväg är ett komplext nätverk innehållande stafettliknande moment som leder till en intracellulär adaptiv reaktion på den yttre miljön [9]. I de ständigt föränderliga miljöer som organismer befinner sig i är reglering av cellmetabolismen essentiell för dess överlevnad. Regleringen görs möjlig av flera viktiga komponenter hos cellen. I många fall är det först genom cellens förmåga att omvandla de yttre miljöfaktorerna, till exempel olika nivåer av pH, salt, vatten, eller näring, till hanterliga signaler som efterföljande intracellulära processer kan äga rum [9]. Homeostas, bibehållandet av likartad miljö inuti cellen relativt den externa miljön, är ett fenomen vars definition är tydligt kopplad till signalvägar [1]. Konstanta förhållanden i dynamiska system som celler innebär ett stort behov av reglering, vilket är ett välbekant område för signalvägar. Fenomenet är också relaterat till biologiskt åldrande [5, 10].

Fastän det finns tydliga skillnader mellan olika organismers mekanismer för åldrande finns det också stråk av likhet för vissa signalvägar som har hittats [11]. Det utnyttjas vid studier av mekanismerna i området där likheterna infinner sig utan att behöva experimentera på människor, utan istället på organismen där likheten ligger. Från slutsatser kring åldrande hos denna modellorganism kan sedan paralleller dras till mänskligt åldrande, och åldrandets rötter blir förhoppningsvis lite tydligare [3].

Den åldringsrelaterade proteinklassen AMPK/SNF1 (AMP-aktiverat Kinas/Sucrose Non-Fermenting1) innehåller olika protein som är evolutionärt konserverade mellan eukaryota organismer [11]. I däggdjur är AMP-aktiverat kinas (AMPK) ett proteinkomplex som spelar en roll i åldrandeprocessen genom att bibehålla energihomeostas i cellen [5]. Komplexets otrolog i den knoppande jästen *Saccharomyces cerevisiae*, Sucrose Non-Fermenting 1 (SNF1), spelar också rollen som bevarare av homeostas men misstänks göra det genom alternativa mekanismer till energiavkänning [6]. Det är känt att ett slutsteg i signalvägen förknippad med SNF1 är repressionen av *SUC2*-genen [7], som är viktig för cellens möjlighet att bryta ner kolkällor som inte är glukos [12]. Nyckeln till denna repression tros vara ett annat protein, Mig1, men det är inte helt känt genom vilka mekanismer som Mig1 reglerar uttrycket av *SUC2* [13].

Något som däremot är känt sedan tidigare är att *SUC2*-uttrycket för jästceller ökar under förhållanden av glukosvält [7]. Nyligen framtagen experimentell data över *SUC2*-uttrycket, se Figur 1, visar dock att denna ökning endast är temporär, ett beteende som kan vara kopplat till en hittills okänd mekanism för Mig1.



**Figur 1:** Uttryck av *SUC2*-genen efter plötslig sänkning av glukoshalt (vid  $t=0$ ) från 4 % till 0,2 %. De grå linjerna visar uttrycken för 30 individuella celler och den streckade linjen visar medelvärdet.

Inom många fält av biologisk forskning, inte minst åldrande, är det av stort intresse att förstå mekanismerna bakom intressanta signalvägar. Om det på proteinnivå är känt hur en cell reagerar på extern stimulans och hur den kopplade signalvägen ser ut kan det bidra till grundforskning och potentiellt leda till farmakologiska tillämpningar som kan justera signalvägen på önskvärt sätt. På grund av signalvägars komplexitet är det oftast ett mångfacetterat problem att dechiffrera och kartlägga systemet i fråga på en meningsfull nivå [10]. Följaktligen har en mängd tillvägagångssätt utvecklats med basis i flera olika fält med sina egna unika metoder. Bland dem finns ett kraftfullt verktyg som kombinerar biologisk kunskap och teorier kring signalvägar med experimentell data, nämligen matematisk modellering [1]. Med hjälp av det kan nyfunnen kunskap från de primärt matematiska respektive biologiska fälten användas tillsammans mot utveckling av nuvarande teorier. Vidare finns det en effektivitetsfråga som modellering besvarar. Att utföra biologiskt laborationsarbete är kostsamt, både tidsmässigt och monetärt; matematisk modellering och simulering tillåter verifiering av teorier där de enda begränsande faktorerna är hårdvara samt tiden det tar att konstruera modeller. På så sätt utnyttjas de värdefulla biologiska resurserna noggrant medan tyngden av arbetet kan läggas på att översätta kumulativ biologisk litteratur till matematiska modeller. Från ett systembiologiskt perspektiv ger kraften av matematisk modellering på så vis en lösning på svårigheterna att kombinera riktade och specifika biologiska upptäckter till en helhet.

De matematiska modellerna är typiskt formulerade som ordinära differentialekvationer (ODE) [1, 8, 14] som beskriver hur derivatan för respektive del i systemet beror av de olika ingående delarna i systemet. Många gånger är ODE:er lätta att sätta upp jämfört med ekvationer som beskriver hela systemet som ofta är omöjliga att sätta upp [1]. ODE:er har dock nackdelen att de kan vara beräkningsmässigt kostsamma att räkna ut jämfört med ekvationer som beskriver hela systemet, men då beräkningarna kan utföras av datorer är det inte ett stort problem. De ODE:er som satts upp innehåller ett flertal obekanta koefficienter som måste skattas för att modellen ska kunna passas till datan. Modellerna kan därefter analyseras matematiskt genom fem huvudsakliga delar [8]. Den första är strukturell identifierbarhet, som innebär att bestämma om modellens parametrar är teoretiskt möjliga att skatta utan hänsyn till tillgänglig data. Den andra är stabilitetsanalys, vilken innefattar att hitta en modells egenvärden och bestämma dess karaktär för att få en bild av modellens dynamiska beteende. Den tredje är skattning av parametrar och syftar till att optimera modellens parametrar så att den bäst passar datan. Sedan följer numerisk identifierbarhet som ger information om med vilken precision modellens parametrar kunde skattas efter datan och ger därmed information om hur trovärdiga skattningarna är. Den femte och sista analysen är känslighetsanalys, och den är till för att utröna vilka av modellens parametrar som verkar ha störst inverkan på modellen.

## 1.1 Syfte

Syftet med kandidatarbetet är att undersöka hypotesen att minskningen av genuttrycket hos *SUC2* sker genom feedback-inhibering av cytosoliskt Mig1. Utifrån en föreslagen mekanism konstrueras tidsberoende ODE-modeller som ska utgöra en matematisk spegling av den biologiska mekanismen. Modellen kommer därefter valideras med hjälp av stabilitetsanalys, strukturell och numerisk identifierbarhetsanalys samt känslighetsanalys. Tillsammans är det gemensamma målet att i första hand ge ett mått på hur bra en modell fångar beteendet av datan, och att i andra hand ge ledtrådar om hur modellen kan utvecklas vidare.

## 2 Teori

Det är kombinationen av biologi och matematik som är kärnan i systembiologin där arbetsgången sker i en cyklisk process [15]. Inledningsvis görs experiment för att samla in data, sedan konstrueras en matematisk modell som kan efterlikna datan. Därefter testas modellen och om den är bra kan den användas för att prediktera nya resultat, om den inte är bra justeras den, testas igen och så vidare. För att göra en bra modell som väl beskriver experimentell data krävs god kunskap om den biologiska process som studeras, samt kännedom om teorin för de matematiska verktyg som används för att bygga modellen, optimera den, testa den och slutligen avgöra om den på ett tillförlitligt sätt beskriver datan. Följande stycken kommer därför gå igenom SNF1-signalvägen, generella principer för modellering, optimering, strukturell och numerisk identifierbarhet samt känslighets- och stabilitetsanalys.

## 2.1 SNF1-signalvägen

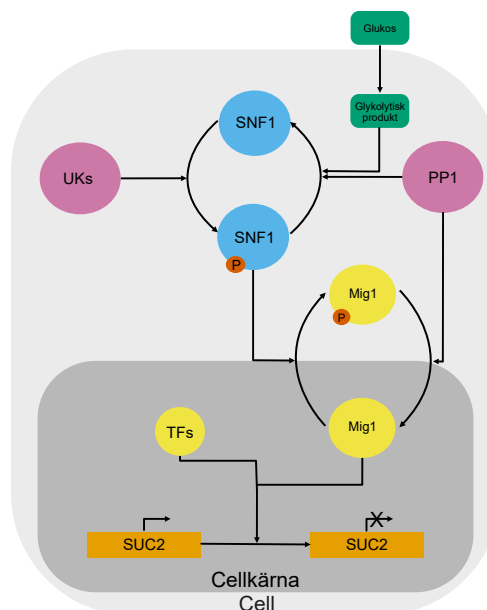
Som tidigare diskuterat är proteinkomplexen AMPK och SNF1 varandras ortologer i däggdjur respektive bagerijäst, och deras signalvägar har likheter. Till skillnad från AMPK är det inte lika välförstått hur SNF1 agerar på omgivningen för att utföra sina homeostasbevarande funktioner. En möjlighet är att SNF1 agerar via avkänning av näring i form av glukos, troligtvis relativt tidigt i glukosmetabolismens förlopp [13].

SNF1 är centralt i SNF1-signalvägen som reglerar genuttrycket av gener involverade i nedbrytning av icke-jäsbara kolkällor. För att den av cellen önskvärda förändringen av genuttryck ska ske behövs flera mellansteg som ofta är kopplade till aktivering och deaktivering av olika protein, vilka i sin tur för vidare signalen längre ned i systemet till dess att den nått sin destination, genen. I SNF1-signalvägen börjar ledet med SNF1. Aktivering och deaktivering av SNF1 sker genom fosforylering respektive defosforylering av komplexet. Främst katalyseras aktiveringen via några proteiner tillsammans kallade uppströmskinasen (UKs), och deaktiveringen via proteinkomplexet Glc7/Reg1 (PP1) [16]. Vidare verkar det som att den främsta inkörningsporten för glukospåverkan på SNF1 sker genom någon av dessa två aktiverings/deaktiveringsfaktorer, möjligen genom repression av UKs eller aktivering av PP1 [13]. Det finns experimentella resultat som vidareutvecklar denna misstanke. Rubenstein et. al. talar emot att mekanismen för glukosavkänning sker genom direkta aktiverings/deaktiveringsmekanismer för UKs eller PP1. Istället verkar deras resultat tyda på att glukos gör SNF1 mer tillgängligt för defosforylering via PP1, och att aktiveringen av SNF1 genom UKs är oberoende av glukos [17].

Längre ned i signalvägen agerar aktivt SNF1 vid förhållanden av låg glukosmängd genom fosforylering av proteinet Mig1 [18], som tros vara viktig i regleringen av genuttryck. Vid förhållanden med hög glukosmängd defosforyleras Mig1 glukosberoende genom PP1 och även genom en annan okänd glukosberoende mekanism [19]. I sitt defosforylerade tillstånd gynnas transporten av Mig1 från cytoplasman till cellkärnan [20], där proteinet binder specifikt till regioner som kodar för protein nödvändiga för nedbrytning av alternativa glukoskällor, däribland *SUC2*-genen [21]. I cellkärnan inhiberar Mig1 transkriptionen av *SUC2* tillsammans med några andra transkriptionsfaktorer (TFs).

Fosforyleringen av Mig1 via SNF1 korrelerar med upplösningen av det *SUC2*-bindande komplexet [22], samt exporten av Mig1 från cellkärnan till cytoplasman [13]. Fastän det är tydligt att SNF1 fosforylerar Mig1 finns det osäkerheter kring hur SNF1 kommer åt att fosforylera Mig1. En hypotes som besvarar den osäkerheten är att import och export av Mig1 sker kontinuerligt och med olika hastigheter beroende på glukosmängd [20]. På grund av det kontinuerliga in- och utflödet av Mig1 kan aktiverat SNF1 då fosforylera Mig1 som genom denna mekanism tillkommer cytoplasman. En alternativ teori till hur SNF1 fosforylerar Mig1 är att komplexet själv vid låga glukoshalter förflyttas till cellkärnan och fosforylerar nukleärt Mig1 där [13]. Hädanefter kommer den första hypotesen angående fosforyleringen av Mig1 genom SNF1 att utgå ifrån.

En sammanfattning över signalvägen illustreras i Figur 2. Alla reaktionspilar har ritats med grund i litteratur, men vissa är mer trovärdiga än andra. Alla antaganden görs med en grad av osäkerhet som bör hållas i åtanke.



Figur 2: Schematisk bild över SNF1-signalvägen.

## 2.2 Generella principer för modellering

När de biologiska mekanismerna som bygger upp ett system är kända kan en matematisk modell över systemet konstrueras. Som ett första steg är det viktigt att formulera en tydlig frågeställning, eftersom många processer som nämnt kan vara mycket komplexa finns det annars en risk för att modellen inte ger svar på det som efterfrågades [1]. I nästa steg definieras systemet som modellen ska beskriva, vilket innebär att det undersöks vilka komponenter och interaktioner som är viktiga och ska inkluderas i modellen. Vidare görs

antaganden för initialvärden och bedömningen om några av modellens parametrar kan anses konstanta [1]. Först efter att dessa steg har avklarats kan den matematiska modellen ta form. Den första modellen för en process är typiskt antingen enkel för att sedan byggas på och utökas, eller mer komplex för att sedan reduceras ner och förenklas. Båda alternativen sker som en del i det iterativa arbetet. Många modeller inom systembiologi består av kopplade system av differentialekvationer. Det beror på att differentialekvationer kan beskriva tids- och rumsberoende hos en process. Om modellen endast beror av en variabel, till exempel tid, är ekvationen en ordinär differentialekvation, en ODE [1]. Generellt skrivs en ODE på formen

$$\frac{dA}{dt} = [\text{produktion av } A] - [\text{nedbrytning av } A] \quad (1)$$

där  $A$  är ett protein och  $t$  är tid.

Om modellen ska beskriva en metabolisk väg är det vanligt att använda approximationer baserade på Michaelis-Menten-kinetik. Det är dock inte lämpligt för signalvägar, då Michaelis-Menten-kinetik gör antagandet att substratet finns i stort överskott i förhållande till mängden enzym. Så är inte fallet för signalvägar, där mängderna av katalys- och substratmolekyler ofta befinner sig i samma storleksordningar [1]. Istället baseras modeller av signalvägar oftast på *massverkans lag*. Den säger att sannolikheten att reaktanterna kolliderar, och således reagerar i en kemisk reaktion, är proportionell mot reaktionshastigheten och mot koncentrationen av reaktanterna upphöjt till deras molära förhållande i reaktionen [10]. Ett enkelt exempel på hur massverkans lag kan appliceras är följande kemiska reaktion



där  $S_1$  och  $S_2$  är substrat som bildar produkten  $P$  med hastighetskonstanten  $k_1$ . Åt andra hållet sönderfaller  $P$  till substraten med hastighetskonstanten  $k_2$ . Eftersom reaktion 2 är av andra ordningen har hastighetskonstanterna enheten ”per koncentration per tidsenhet”. Vidare definieras hastighetskonstanterna alltid som positiva [10]. Reaktionen kan nu beskrivas med två hastighetsekvationer, enligt ekvation 3

$$r_1 = k_1 \cdot S_1 \cdot S_2 \quad (3a)$$

$$r_2 = k_2 \cdot P^2 \quad (3b)$$

där reaktionshastigheten  $r$  mäts i ”mängd per tidsenhet”, mängden kan vara koncentration, aktivitet eller liknande. Genom att implementera en signalvägs olika reaktionshastigheter i ODE:er, enligt definitionen i ekvation 1, genereras ett ODE-system som agerar som en modell över signalvägen [1]. Ett ODE-system för exemplet presenteras i ekvation 4

$$\frac{dS_1}{dt} = \frac{dS_2}{dt} = r_2 - r_1 \quad (4a)$$

$$\frac{dP}{dt} = r_1 - r_2 \quad (4b)$$

där  $S_1$ ,  $S_2$  och  $P$  är modellens tillstånd och  $k_i$ , som är inkluderade i  $r_i$  enligt ekvation 3, är modellens parametrar. Sådana system löses och verifieras sedan numeriskt då de generellt är icke-linjära [1].

## 2.3 Strukturell identifierbarhet

En viktig del i utformningen av en modell är att vara säker på att en modells parametrar är strukturellt identifierbara, det vill säga att de är möjliga att bestämma, givet utdata som är helt fri från brus [8]. Om någon parameter inte är strukturellt identifierbar ger det kunskap om ifall parametern bör skrivas om eller tas bort från modellen, eftersom den inte kan bestämmas utifrån modellens struktur. Det kallas för strukturell identifierbarhet då endast modellens struktur samt utdata är avgörande. Att testa strukturell identifierbarhet är ett bra hjälpmedel då det ger kunskap om vilka parametrar som är oidentifierbara redan i början av modellutformningen och undviker problem i optimeringsprocessen [8].

## 2.4 Optimering

En färdig modell passas till den experimentella datan genom att anpassa dess parametrar. Passningen bygger på en maximum likelihood-skattning (ML-skattning) som går ut på att anpassa parametrarna  $\theta \in \mathbb{R}^k$  sådana att sannolikheten att den modellerade datan  $\hat{y}(\theta) \in \mathbb{R}^{n \cdot m}$  kommer från samma fördelning som den experimentella datan  $y \in \mathbb{R}^{n \cdot m}$  maximeras [23]. Härledningen av ML-skattningen kan ses i appendix A och resulterar i kostnadsfunktionen  $S(\theta) \in \mathbb{R}$  i ekvation 5

$$S(\theta) = \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}^2 \quad (5)$$

där  $r_{\psi,i}$  är residualen definierad som  $y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta)$ ,  $\psi$  och  $i$  är tillståndet respektive tidpunkten. Syftet med optimeringen är således att minimera  $S(\theta)$  för att passa modellen till datan. Minimeringen går inte att utföra analytiskt och därför måste numeriska metoder användas [24]. Det finns flera olika tillvägagångssätt, bland annat med *steepest descent*-metoden och *Newton-Raphsons metod*. Skillnaden mellan de två metoderna är att *steepest descent*-metoden approximerar funktionen som ska minimeras som ett polynom av första graden medan *Newton-Raphsons metod* approximerar den som ett polynom av andra graden. *Newton-Raphsons metod* ger därför generellt sett en bättre approximation och därmed en effektivare minimering [24]. En härledning av *Newton-Raphsons metod* kan ses i appendix B och ger den nedåtgående riktningen  $p \in \mathbb{R}^k$  i ekvation 6.

$$p = -(H(\theta))^{-1} \cdot \nabla S(\theta) \quad (6)$$

där  $H(\theta) \in \mathbb{R}^{k \times k}$  är Hessianen för kostnadsfunktionen och  $\nabla S(\theta) \in \mathbb{R}^k$  är gradienten av kostnadsfunktionen. Hessianen kan vara beräkningskrävande och det är därför av intresse att approximera den och genom detta förenkla *Newton-Raphsons metod*. En av metoderna som bygger på *Newton-Raphsons metod* är Gauss-Newtonns metod, vilken är fördelaktig i att den är anpassad för att minimera kostnadsfunktioner bestående av summor [25]. De antaganden och omskrivningar som Gauss-Newtonns metod bygger på kan ses i appendix C och resulterar i ekvation 7.

$$p = (J^T J)^{-1} \cdot J^T r \quad (7)$$

där  $J \in \mathbb{R}^{n \cdot m \times k}$  är Jacobimatrisen och  $r \in \mathbb{R}^{n \cdot m}$  är residualvektorn, vilka beskrivs i ekvation 8.

$$J = \begin{bmatrix} \frac{\partial \hat{y}_1(t_1, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_1(t_1, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_1(t_1, \theta)}{\partial \theta_k} \\ \frac{\partial \hat{y}_1(t_2, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_1(t_2, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_1(t_2, \theta)}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_1(t_m, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_1(t_m, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_1(t_m, \theta)}{\partial \theta_k} \\ \frac{\partial \hat{y}_2(t_1, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_2(t_1, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_2(t_1, \theta)}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_n(t_m, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_n(t_m, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_n(t_m, \theta)}{\partial \theta_k} \end{bmatrix} \quad r = \begin{bmatrix} r_{1,1} \\ r_{1,2} \\ \vdots \\ r_{1,m} \\ r_{2,1} \\ r_{2,2} \\ \vdots \\ r_{n,m} \end{bmatrix} \quad (8)$$

## 2.5 Stabilitetsanalys

Som tidigare nämnt är de verktyg som används för att lösa eller verifiera modeller numeriska metoder. Därför kan det vara svårt att få en känsla för vad lösningen ska vara och om resultaten är rimliga. Då är det bra att komplettera med att genomföra en stabilitetsanalys. Det är en analytisk metod som utnyttjar Taylorutveckling för att linjärisera systemet och lösa ut dess egenvärden [26]. Egenvärdenas tecken, samt om de är reella eller komplexa, ger sedan svar på om systemet efter lång tid kommer gå mot jämvikt, oscillera (börja svänga) eller gå mot oändligheten [26]. Vidare kan en stabilitetsanalys visa om ett stabilt jämviktsläge endast uppnås inom vissa intervall på parametrarna och bidrar på så sätt med begränsningar till vad parametrarna kan anta för värden, något som underlättar i parameterskattningen.

Det första steget i stabilitetsanalysen är att hitta systemets jämviktslägen. Det vill säga, lösa ut vilka koncentrationer,  $A_i$ , som uppfyller att  $\frac{dA_i}{dt} = 0$ . Sedan görs Taylorutveckling av varje funktion kring jämvikt, vilket

resulterar i en Jacobimatrix som liknar den i ekvation 8, men i stabilitetsanalysen beskriver den istället  $A_i$  utvärderad vid jämvikt och har dimensionerna  $n \times n$ , där  $n$  är antalet tillstånd i modellen [1]. Ur Jacobimatriken härleds den karakteristiska ekvationen, ekvation 9, från vilken systemets egenvärden kan beräknas

$$\det(J - \lambda I) = 0 \quad (9)$$

där  $\lambda$  är egenvärdet och  $I$  är identitetsmatrisen. Ekvationen löses för  $\lambda$ , som nämnt ovan avgör om systemet är stabilt eller instabilt, och oscillerande eller styvt.

Eftersom Jacobimatriken här är en  $n \times n$ -matrix kommer den genom ekvation 9 att ge upphov till ett polynom av grad  $n$  enligt

$$a_0 \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n = 0 \quad (10)$$

där  $a_i$  är reellvärda koefficienter. Det följer att ekvationen har  $n$  antal rötter och därmed har systemet  $n$  egenvärden. Rötterna till polynomet kan vara reellt positiva, reellt negativa eller komplexa, där reella rötter ger ett styvt system och komplexa rötter ger ett oscillerande system. Ett nödvändigt och tillräckligt villkor för att systemet är stabilt är att samtliga egenvärden har en strikt negativ realdel [1].

Om systemen är stora kommer det ge polynom av hög grad som många gånger är svåra att lösa. Då finns det några matematiska satser som kan användas för att visa om ett system är stabilt eller inte utan att behöva lösa polynomet. En sats som kan användas för att snabbt se om ett system har reella och positiva egenvärden är Descartes teckenregel. Den säger att ett polynom på samma form som ekvation 10 maximalt har samma antal reella positiva rötter som antalet gånger koefficienterna byter tecken [27]. Det vill säga att för ett stabilt system måste samtliga koefficienter ha samma tecken. Descartes teckenregel ger dock ingen information om huruvida realdelen på eventuella imaginära rötter är negativ eller inte. För att ta reda på det kan Routh-Hurwitzs stabilitetskriterie användas [26]. Först skapas Hurwitz matrix  $H$  av koefficienterna  $a_i$  från det karakteristiska polynomet, ekvation 10. Matrizen byggs genom att sätta koefficienter med udda index i första kolumnen och sedan minska index med ett åt höger, koefficienter där  $i < 0$  eller  $i > n$  ersätts med nollor, enligt

$$H = \begin{bmatrix} a_1 & a_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & a_0 & 0 & 0 & \dots & 0 \\ a_5 & a_4 & a_3 & a_2 & a_1 & a_0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & a_n \end{bmatrix}. \quad (11)$$

$H$ 's diagonala principalunderdeterminanter ges då av

$$\Delta_1 = a_1, \Delta_2 = \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix}, \Delta_3 = \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix}, \Delta_n = \begin{vmatrix} a_1 & a_0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & \dots & 0 \\ a_5 & a_4 & a_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_n \end{vmatrix}.$$

Routh-Hurwitzs stabilitetskriterie säger sedan att, givet  $a_i > 0$ , är samtliga rötters realdelar strikt negativa om och endast om  $\Delta_1 > 0$ ,  $\Delta_2 > 0$ ,  $\Delta_3 > 0$ , ...,  $\Delta_n > 0$  uppfylls [26]. Kriteriet innebär att många determinanter måste beräknas för högre graders polynom. Liénard-Chiparts kriterie är en modifiering av Routh-Hurwitzs stabilitetskriterie som minskar antalet determinanter som behöver beräknas [28]. Kriteriet säger att om ett av följande villkor är uppfyllt är det ett nödvändigt och tillräckligt villkor för att systemet är stabilt:

1.  $a_n > 0, a_{n-2} > 0, \dots; \Delta_1 > 0, \Delta_3 > 0, \dots$
2.  $a_n > 0, a_{n-2} > 0, \dots; \Delta_2 > 0, \Delta_4 > 0, \dots$
3.  $a_n > 0, a_{n-1} > 0, a_{n-3} > 0, \dots; \Delta_1 > 0, \Delta_3 > 0, \dots$
4.  $a_n > 0, a_{n-1} > 0, a_{n-3} > 0, \dots; \Delta_2 > 0, \Delta_4 > 0, \dots$

där  $a_i$  är definierade på samma sätt som för Routh-Hurwitzs stabilitetskriterie [28].

## 2.6 Numerisk identifierbarhet

Efter parameteruppskattning är det användbart att veta med vilken precision en parameter har skattats och då används numerisk identifierbarhetsanalys [8]. Det här är en viktig egenskap att undersöka då det krävs att modellen är numeriskt identifierbar för att modellens prediktioner ska vara tillförlitliga. Numerisk identifierbarhet kan undersökas med hjälp av kovariansmatrisen för parametrarna,  $\text{COV}(\hat{\theta})$ . Eftersom, för skattade parametrar  $\hat{\theta}$  är en parameter  $\hat{\theta}_i$  numeriskt identifierbar då den har en låg parametervarians samt ett värde nära noll i korrelationsmatrisen. Parametervariansen ges av diagonalelementen i kovariansmatrisen [8]. Kovariansmatrisen kan i sin tur, approximeras med Fishers informationsmatris,  $F(\hat{\theta})$ . Enligt Cramér-Raos sats fås då en nedre gräns

$$\text{COV}(\hat{\theta}) \geq F^{-1}(\hat{\theta}) \quad (12)$$

där  $F(\hat{\theta})$  är en summering av Jacobianen i kvadrat över tid viktad med den reciproka variansen [8]. Vidare kan approximationen göras att  $\hat{H} \approx F(\hat{\theta})$  [8]. Det går därmed att beräkna kovariansmatrisen genom att ta fram Hessianen för de optimala parametrarna. Hessianen kan sedan tas fram via Jacobianen med antagandena som görs i Gauss-Newtons metod då residualerna antas vara små, se Appendix C.

För att lättare visualisera variationen hos stora system kan det vara till fördel att använda variationskoefficienter,  $\sigma_i = \frac{\sqrt{\text{COV}(\hat{\theta}_i, \hat{\theta}_i)}}{\hat{\theta}_i}$ , där  $\text{COV}(\hat{\theta}_i, \hat{\theta}_i)$  är diagonalelementet i kovariansmatrisen för den  $i$ :te parametern. Variationskoefficienten visar på förhållandet mellan den skattade standardavvikelsen och parametervärdet. Fortsättningsvis är det även av intresse att se hur parametrarna korrelerar genom att ta fram korrelationsmatrisen. Det görs genom att skala om kovariansmatrisen enligt

$$\text{COR}(\hat{\theta}_i, \hat{\theta}_j) = \frac{\text{COV}(\hat{\theta}_i, \hat{\theta}_j)}{s_i \cdot s_j} \quad (13)$$

där  $s_i, s_j$  är roten ur värdet av kovariansmatrisens diagonalelement för parameter  $\hat{\theta}_i$  och  $\hat{\theta}_j$ . Värdena i korrelationsmatrisen är mellan -1 och 1 och här innebär ett värde nära 1 att två parametrar samvarierar och medför att det inte går att urskilja dem från varandra, de blir alltså numeriskt oidentifierbara [8].

## 2.7 Känslighetsanalys

Vid vidare utveckling av en modell är det intressant att veta vilka parametrar som har störst inverkan på modellens tillstånd. För att undersöka det kan en känslighetsanalys utföras, det vill säga att variationer i modellens tillstånd i respons till variationer i parametrarna kvantifieras [8]. För att illustrera detta används ofta känslighetsmatrisen  $S_{\hat{\theta}}$  enligt

$$S_{\hat{\theta}} = \begin{bmatrix} \frac{\partial \hat{y}_1(t_1)}{\partial \hat{\theta}_1} & \dots & \frac{\partial \hat{y}_1(t_m)}{\partial \hat{\theta}_1} & \frac{\partial \hat{y}_2(t_1)}{\partial \hat{\theta}_1} & \dots & \frac{\partial \hat{y}_2(t_m)}{\partial \hat{\theta}_1} & \dots & \frac{\partial \hat{y}_n(t_1)}{\partial \hat{\theta}_1} & \dots & \frac{\partial \hat{y}_n(t_m)}{\partial \hat{\theta}_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_1(t_1)}{\partial \hat{\theta}_k} & \dots & \frac{\partial \hat{y}_1(t_m)}{\partial \hat{\theta}_k} & \frac{\partial \hat{y}_2(t_1)}{\partial \hat{\theta}_k} & \dots & \frac{\partial \hat{y}_2(t_m)}{\partial \hat{\theta}_k} & \dots & \frac{\partial \hat{y}_n(t_1)}{\partial \hat{\theta}_k} & \dots & \frac{\partial \hat{y}_n(t_m)}{\partial \hat{\theta}_k} \end{bmatrix} \quad (14)$$

som innehåller alla partiella derivator av modellens olika tillstånd med avseende på parametrarna. Genom den här matrisen kan det noteras om, exempelvis, en liten ändring i en parameter ger en stor ändring i tillståndet. Det betyder då att modellen är känslig för en variation i den parametern, alltså är parametern mer identifierbar [8]. Eftersom  $S_p$  är tidsvarierande och svår att analysera är det användbart att använda ett viktat medelvärde av den så kallade relativa känslighetsmatrisen över tid. Det brukar betecknas root mean square (RMS) som beräknas enligt

$$\text{RMS}[s_{ij}(t_1, t_2, \dots, t_N)] = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N \left( \frac{\partial \hat{y}_i(t_k)}{\partial \hat{\theta}_j} \right)^2 \cdot \left( \frac{\hat{\theta}_j}{\hat{y}_i(t_k)} \right)^2} \quad (15)$$

där  $s_{ij}$  är den relativa känsligheten av tillståndet  $\hat{y}_i$  och parameter  $\hat{\theta}_j$ ,  $N$  är antalet tidssteg över ett specifikt tidsintervall.

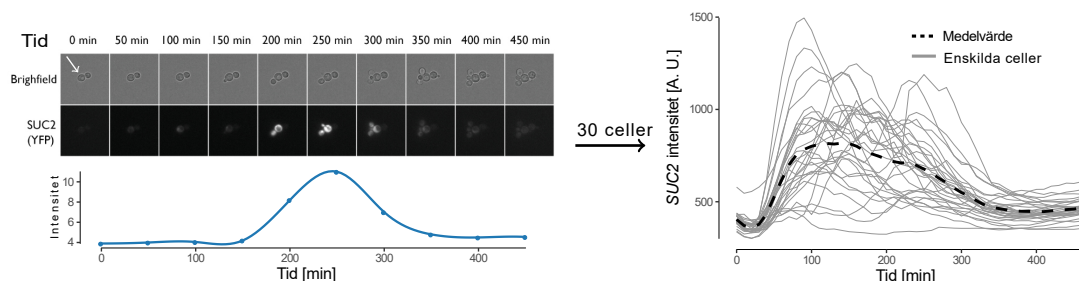
### 3 Data och metod

Vid den matematiska analysen av modellerna användes flera färdiga koncept som direkt implementerades, undantaget parameteruppskattningen där vissa anpassningar och utvecklingar gjordes av Gauss-Newtons metod. För optimeringen formulerades en egen iterativ optimeringsalgoritm med hjälp av en ODE-lösare. Hur de färdiga koncepten implementerades och detaljer kring parameteruppskattningen och optimeringsalgoritmen presenteras nedan och all kod som användes hittas på GitHub (<https://github.com/axillus/Kandidat2020>). Närmast följer ett avsnitt om hur datan till arbetet togs fram.

#### 3.1 Data

Den experimentella datan som användes för att optimera och testa modeller togs fram och behandlades av Niek Welkenhuysen från Cvijovic Lab, och var därför inte en del av detta arbete. 30 stycken *Saccharomyces Cerevisiae*-celler odlades på en kontrollerad och justerbar mängd glukos med konstant tillförsel. Cellerna hade modifierats så att mätningar av Mig1 och Suc2 kunde göras. Mig1-protein i cellen hade ett fluorescerande protein bundet till sig vilket möjliggjorde kvantifiering, medan *SUC2* hade ett annat fluorescerande protein bundet till gensekvensen som istället visade på genuttrycket. Glukoskoncentrationen justerades ned, från glukoshalten 4% till 0,2%, varefter tidsförloppet för datan började. Mikroskopbilder av cellerna togs med regelbundna intervall, vilka blev grund till datan. Efterbehandlingen av bilderna omfattade bland annat att mäta intensiteten av fluorescens från bilderna för Mig1 och Suc2, och att översätta fluorescensen till ett numeriskt värde, samt att minska förutspåeligt brus i datan. För varje tidsserie togs det också fram ett medelvärde över alla cellers fluorescensvärden. Det var dessa medelvärden som användes som data i alla matematiska analyser i detta arbete. Ett exempel av flödet från fluorescensbilder till färdig data illustreras i Figur 3.

Den slutgiltiga datan innefattade tidsserier över Mig1 och Suc2. För Mig1 är den närmaste tolkningen av datan att värdena motsvarar koncentrationen av Mig1 i cellkärnan vid en viss tidpunkt. Värdena av Suc2 kan tolkas som storleken av uttrycket för *SUC2*-genen vid en viss tidpunkt. Figurer över datan för Mig1 och Suc2 kan ses i större format i Appendix D.



**Figur 3:** Från fluorescensbilder (t.v.) till en tidsserie över uttryck för tillståndet Suc2 (t.h.). T.v. innebär *Brighfield* och (*YFP*) mikroskopbilder tagna med vanligt ljus respektive UV-ljus. Från varje cells uttryck togs sedan ett medelvärde fram, vilket är datan som användes i detta arbete. Figuren är, med tillåtelse från författarna, adapterad från [29].

#### 3.2 Strukturell identifierbarhet

För strukturell identifierbarhetsanalys av alla modeller användes paketet IdentifiabilityAnalysis (version 0.9.0) [30, 31] i Wolfram Mathematica (version 12.0.0.0).

#### 3.3 Parameteruppskattning med Gauss-Newtons metod

Gauss-Newtons metod, beskriven i avsnitt 2.4, användes för att formulera en iterativ optimeringsalgoritm som sedan implementerades i Python3. Ett test av den använda optimeringsalgoritmen gjordes, vilken finns i E. Testet visade att den använda optimeringsalgoritmen fungerade och hittade optimumet snabbare än *steepest descent*-metoden.

### 3.3.1 Avvikelser från Gauss-Newtons metod

För att anpassa optimeringen till de modeller som undersöks samt göra optimeringen snabbare och mer robust gjordes följande anpassningar av Gauss-Newtons metod.

#### Begränsningar av sökområdet

De parametrar som togs fram är hastighetskonstanter i kemiska reaktioner och därför finns det en viktig begränsning i vilka värden de kan anta, de kan inte vara mindre än noll. Kännedom om hastighetskonstanternas storlek var inte tillräckligt stor för att kunna sätta en övre begränsning.

#### Avsaknad av derivator

Då modellen är uppbyggd av system av ODE:er är de derivator som krävs för att ta fram den nedåtgående riktningen ofta otillgängliga. Därför togs dessa fram med hjälp av finita differenser. För att få en bra approximation av derivatan måste en lagom stor steglängd,  $h$ , väljas. Här användes  $h = \sqrt{\epsilon_{ps}}$ , där  $\epsilon_{ps}$  är maskintoleransen, då detta värde har visat ge en bra approximation [32].

#### Hantering av singulära- och indefinita matriser

För att  $\mathbf{p}$  ska vara en nedåtgående riktning behöver approximationen av Hessianen vara positivt semidefinit [24]. Dessutom behöver matrisen vara inverterbar vilket ställer kravet att den inte är singulär och då inte har nollvärda egenvärden [33]. Sammantaget ger det att ekvation 16 behöver uppfyllas.

$$\lambda_j > 0, \quad \forall j = 1, 2, \dots, k \quad (16)$$

där  $\lambda_j$  är de  $k$  stycken egenvärdena för Hessianen. Om ekvationen inte är uppfylld kan den nedåtgående riktningen,  $\mathbf{p}$ , inte tas fram med Gauss-Newtons metod. En lösning på det är att i stället använda *steepest descent*-metoden men då den metoden ofta är långsam är det inte önskvärt. En bättre lösning är Levenberg Marquardts metod [34]. I denna rapport användes en version av denna metod där matrisens egenvärden skiftas sådant att matrisen återigen blir positivt definit, det görs med ekvation 17

$$\begin{aligned} \mathbf{H}_{skiftad} &= \mathbf{H} + \gamma \mathbf{I} \\ \gamma &= -\lambda_{min} + h \end{aligned} \quad (17)$$

där  $\lambda_{min}$  är det minsta egenvärdet för Hessianen  $\mathbf{H}$  [24, 35]. Det kan ses att då  $\gamma \rightarrow \infty$  går  $\mathbf{H} \rightarrow \mathbf{I}$  och  $\mathbf{p}$  går mot *steepest descent*-lösningen. Det ska dock nämnas att en singulär Hessian kan vara ett tecken på att modellen är felformulerad, framför allt vid linjär regression. Vid icke-linjär regression är denna koppling inte lika säker och det kan därför fortfarande vara av intresse att fortsätta parameteruppskattningen [36].

#### Bestämning av steglängd

Det är inte säkert att ett helt steg i den nedåtgående riktningen  $\mathbf{p}$  resulterar i en minskning i kostnadsfunktionen i enlighet med ekvation 18.

$$S(\boldsymbol{\theta} + \mathbf{p}) < S(\boldsymbol{\theta}) \quad (18)$$

Dock kan  $\mathbf{p}$  fortfarande vara en nedåtgående riktning men det behövs ett kortare steg för att få en minskning av kostnadsfunktionen. Därmed behövs en steglängd,  $\alpha$ , hittas som uppfyller ekvation 19.

$$S(\boldsymbol{\theta} + \alpha \mathbf{p}) < S(\boldsymbol{\theta}) \quad (19)$$

För att bestämma steglängden på ett tillräckligt bra sätt användes en *line search*-metod där  $\alpha_{start}$  är 1 och  $\alpha$  sedan halveras i varje iteration tills  $S(\boldsymbol{\theta} + \alpha \mathbf{p}) < S(\boldsymbol{\theta})$  uppfylls [24]. Om  $\alpha < h$  avbröts iterationen.

#### Viktning

Eftersom delar av datan kan vara mer intressant att anpassa modellen efter än andra delar kan optimeringen viktas för göra den anpassningen mer lönsam. Viktningen görs genom att residualerna för de olika datapunkterna multipliceras med en vikt av vald storlek, varpå kostnadsfunktionen beräknas som i ekvation 20.

$$S_{viktad}(\boldsymbol{\theta}) = \sum_{\psi=1}^n w_{\psi} \sum_{i=1}^m r_{\psi,i}^2 \quad (20)$$

där  $w_{\psi}$  är vikten för respektive del av datan. Gradienten ges då av

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \theta_j} = 2 \sum_{\psi=1}^n w_{\psi} \sum_{i=1}^m r_{\psi,i} \cdot \frac{\partial \hat{y}_{\psi}(t_i, \boldsymbol{\theta})}{\partial \theta_j} \quad (21)$$

och Hessianen blir

$$\frac{\partial^2 S(\theta)}{\partial \theta_j \partial \theta_l} = 2 \sum_{\psi=1}^n w_{\psi} \sum_{i=1}^m \frac{\partial \hat{y}_{\psi}(t_i, \theta)}{\partial \theta_j} \cdot \frac{\partial \hat{y}_{\psi}(t_i, \theta)}{\partial \theta_l}. \quad (22)$$

I den använda optimeringen valdes vikten för Mig1 till 1 och vikten för Suc2 till 10, för att bättre anpassa modellen till de experimentella värdena för Suc2.

### Anpassning för att hitta globala optimumet

Kostnadsfunktionen kan ha många lokala minimum och Gauss-Newtons metod kan bara hitta lokala extrempunkter. För att öka chansen att det globala optimumet blir funnet kördes iterationen flera gånger med olika startgissningar.

### 3.3.2 Struktur på optimeringsalgoritmen

Formuleringen av kostnadsfunktionen och den nedåtgående riktningen i avsnitt 2.4 med de anpassningar angivna i avsnitt 3.3.1 användes för att formulera följande struktur på optimeringsalgoritmen.

1. Välj en startgissning av de olika parametrarna  $\theta_1, \theta_2, \dots, \theta_k$ , där  $k$  är antalet parametrar.
2. Använd en ODE-lösare för att ta fram modellerade värden för värdena på parametrarna.
3. Använd en ODE-lösare för att ta fram modellerade värden för en förskjutning,  $h$ , i värdet på respektive parameter.
4. Ta fram Jacobimatrisen med derivator beräknade med finita differenser baserat på värden från steg 2 och 3.
5. Ta fram residualmatrisen från den experimentella datan och värden från steg 2.
6. Beräkna derivatan och Hessiananapproximationen.
7. Beräkna den nedåtgående riktningen  $p$ .
8. Sätt de nya parametrarna som  $\theta_{ny} = \theta + \alpha p$  med steglängd  $\alpha > 0$  sådan att  $S(\theta + \alpha p) < S(\theta)$ .
9. Upprepa steg 2 - 8 tills dess att något av följande kriterier är uppfyllt.
  - (a) Kostnadsfunktionen är tillräckligt liten, det vill säga  $S(\theta) \approx 0$ .
  - (b) Det inte går att hitta en nedåtgående riktning sådan att  $S(\theta + \alpha p) < S(\theta)$ .
10. Upprepa steg 1 - 9 flera gånger med olika startgissningar för att öka chansen att det globala maximumet blir funnet.

### 3.3.3 Beskrivning av optimering

Strukturen ovan implementerades i Python 3.7.7 med PyCharm som IDE. De system av ODE:er som utgjorde modellerna löstes som begynnelsevärdesproblem med Scipy (version 1.4.1) [37]. För matrisberäkningar och diverse andra beräkningar användes NumPy (version 1.18.1) [38]. Den experimentella datan hämtades in med hjälp av Pandas (version 1.0.3) [39] och slutligen ritades datan upp med Matplotlib (version 3.1.3) [40]. Vid implementeringen användes i vissa fall 3D matriser, där det i avsnitt 2.4 används matriser på formen  $\mathbb{R}^{n \times m \times k}$  eller  $\mathbb{R}^{n \cdot m}$  användes då  $\mathbb{R}^{n \times m \times k}$  respektive  $\mathbb{R}^{n \times m}$  istället.

## 3.4 Stabilitetsanalys, numerisk identifierbarhetsanalys och känslighetsanalys

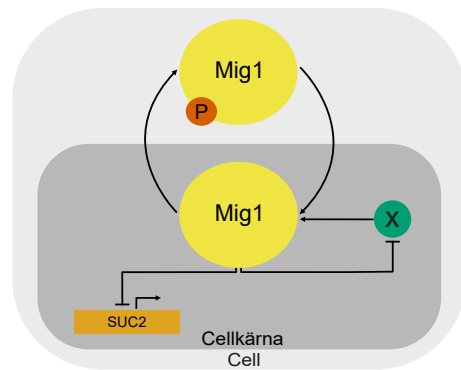
Stabilitetsanalysen, den numeriska identifierbarhetsanalysen och känslighetsanalysen implementerades i Python 3.7.7 med PyCharm som IDE. För beräkningar med matriser och summor användes NumPy (version 1.18.1) [38]. Framtagning av parameterintervallen i stabilitetsanalysen samt heat maps för korrelationsmatriserna gjordes med hjälp av Seaborn (version 0.10.1) [41] och matplotlib (version 3.1.3) [40].

## 4 Resultat

Resultatet grundas i hypotesen för SNF1-signalvägen från syftesformuleringen i avsnitt 1.1. Hypotesen gestaltas som en mekanism där Mig1 reglerar sin egen produktion genom repression av genen *MIG1*, en mekanism skulle kunna förklara varför *SUC2*-uttrycket verkar vara semi-transient och möjligen cykliskt som ses i Figur 1. Utifrån litteratur och SNF1-signalvägen, beskriven i Figur 2, undersöktes vilka av signalvägens delar som kunde vara relevanta för hypotesen. De delarna valdes ut och inkluderades som tillstånd i Modell 1. Efter optimering och verifiering av den första modellen gjordes mindre justeringar vilket resulterade i två nya modeller, Modell  $k_6 = k_7$  och Modell  $k_3 = k_5$ . Modellerna presenteras nedan tillsammans med respektive matematiska analys.

### 4.1 Den underliggande mekanismen för Modell 1 består av fosforylering, defosforylering och en feedback-loop

De tillgängliga mätvärdena gav information om Suc2, Mig1, och Hxk1. I referens till SNF1-signalvägen i Figur 2 koncentrerades modellen därför till den nedre delen av signalvägen för att fånga Mig1p och Mig1 som i sin tur kunde agera på *SUC2*. Detta gjordes genom att ta bort flera tillstånd ur Figur 2 och ersätta dem med en hastighetskonstant. Uppifrån och ned gjordes resonemang för varför proteinet i fråga togs bort, samt hur dess påverkan modellerades istället. För glukos gjordes resonemanget att det i princip agerar snabbt och direkt på SNF1 i samband med PP1. Glukos och PP1 modellerades därför som en del av defosforyleringen från SNF1p till SNF1. Fosforyleringen av SNF1 genom UKs verkar, enligt tidigare försök, inte bero av glukoskoncentrationen och modellerades därför som en del av fosforyleringen av SNF1 till SNF1p. Fosforyleringen av SNF1 till SNF1p modellerades som att vara en del av fosforyleringen av Mig1 till Mig1p. Vidare modellerades TFs som en del av Mig1-repressionen av *SUC2*. En tillsats gjordes också till modellen. X introducerades som ett tillstånd som kan liknas vid genen *MIG1*, och gav ett sätt att kunna modellera feedbacken av Mig1. X fungerade också för att modellera nybildandet av Mig1 med en viss tidsförskjutning. Resonemangen ovan landade i en biologisk modell innehållande fyra olika tillstånd: Mig1, Mig1p, Suc2 och X. Modellen presenteras i Figur 4.



**Figur 4:** Schematisk figur över den biologiska modellen för den förenklade SNF1-signalvägen

För att sedan beskriva den biologiska modellen med en matematisk modell applicerades massverkans lag, ekvation 3. På så vis erhålls hastighetsekvationer som var och en beskriver varje enskilt steg. Tabell 1 innehåller samtliga hastighetsekvationer med respektive parameter samt det biologiska resonemanget bakom.

**Tabell 1:** Hastighetsekvationerna med respektive parameter som beskriver varje steg i den biologiska modellen samt det biologiska resonemanget bakom.

| Hastighetsekvation<br>[mängd per<br>tidsenhet] | Parameter<br>[per<br>tidsenhet] | Biologisk betydelse                                                                                                                                                                     |
|------------------------------------------------|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $r_1 = k_1 \cdot Mig1p$                        | $k_1$                           | Defosforylering av Mig1p till Mig1.                                                                                                                                                     |
| $r_2 = k_2 \cdot Mig1$                         | $k_2$                           | Fosforylering av Mig1 till Mig1p.                                                                                                                                                       |
| $r_3 = \frac{k_3}{Mig1+0,1}$                   | $k_3$                           | Produktion av Suc2 som inhiberas av Mig1. Genom den inversa proportionaliteten kommer Suc2-produktionen minska när Mig1 ökar. 0,1 adderas i nämnaren för att undvika division med noll. |
| $r_4 = k_4 \cdot Suc2$                         | $k_4$                           | Nedbrytning av Suc2.                                                                                                                                                                    |
| $r_5 = \frac{k_5}{Mig1+0,1}$                   | $k_5$                           | Produktion av X som inhiberas av Mig1. Genom den inversa proportionaliteten kommer X-produktionen minska när Mig1 ökar. 0,1 adderas i nämnaren för att undvika division med noll.       |
| $r_6 = k_6 \cdot X$                            | $k_6$                           | Produktion av Mig1 från X.                                                                                                                                                              |
| $r_7 = k_7 \cdot Mig1$                         | $k_7$                           | Nedbrytning av Mig1.                                                                                                                                                                    |

Hastighetsekvationerna kombineras sedan, enligt ekvation 1, vilket genererar ODE:er som visar hur uttrycket av Mig1, Mig1p, Suc2 och X varierar över tid. Tillsammans bildar ODE:erna den matematiska modellen, Modell 1, se ekvation 23. I samma ekvation presenteras startvärdet för varje tillstånd, där det för Mig1 och Suc2 är baserat på den givna datan. Startvärdena för Mig1p och X antogs vara noll då det precis innan  $t_0$  varit hög glukosnivå och därmed bör uttrycket av de båda tillstånden varit lågt.

$$\frac{dMig1}{dt} = r_1 - r_2 + r_6 - r_7, \quad Mig1(t_0) = 2,649 \quad (23a)$$

$$\frac{dMig1p}{dt} = -r_1 + r_2, \quad Mig1p(t_0) = 0 \quad (23b)$$

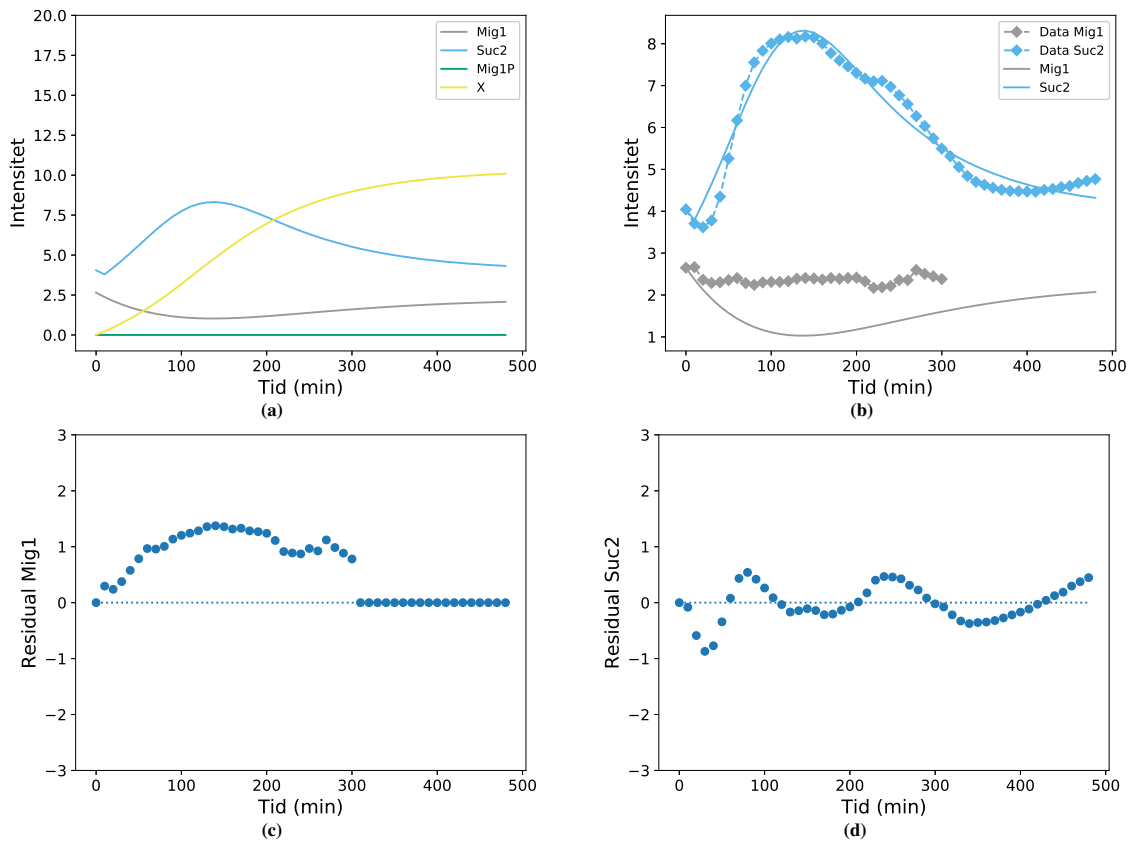
$$\frac{dSuc2}{dt} = r_3 - r_4, \quad Suc2(t_0) = 4,09(\pm 0,01) \quad (23c)$$

$$\frac{dX}{dt} = r_5 - r_6, \quad X(t_0) = 0 \quad (23d)$$

Optimering av Modell 1 resulterade, efter 12 startgissningar, i en bästa parameteruppskattning presenterad i Tabell 2. Med de värdena insatta i modellen erhöles aktiviteterna som ses i Figur 5a och mer specifika jämförelser mellan datan och modellerade värden för Mig1 och Suc2 kan ses i Figur 5b.

**Tabell 2:** Kostnadsfunktionen, den viktade kostnadsfunktionen samt de skattade parametrarna för Modell 1.

| $S(\hat{\theta})$ | $S_{viktad}(\hat{\theta})$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ | $\hat{\theta}_5$     | $\hat{\theta}_6$     | $\hat{\theta}_7$     |
|-------------------|----------------------------|------------------|------------------|------------------|------------------|----------------------|----------------------|----------------------|
| 37,9              | 534                        | 12,0             | 0,00             | 46,6             | 4,97             | $5,91 \cdot 10^{-2}$ | $2,46 \cdot 10^{-3}$ | $1,13 \cdot 10^{-2}$ |

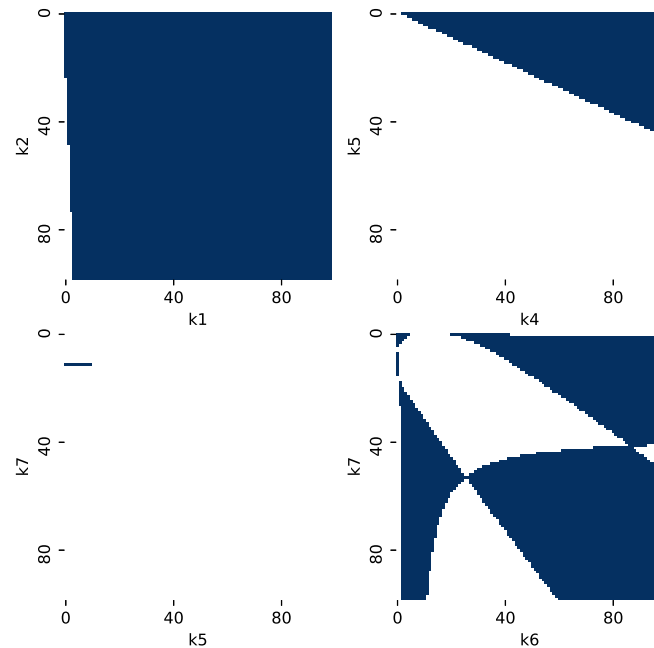


**Figur 5:** Resultat för Modell 1. (a) De olika aktiviteterna givna av Modell 1 med de optimerade parametrarna. Det kan ses att aktiviteten för Mig1p är noll genom hela tidsintervallet vilket beror på att hastighetskonstanten för reaktionen i vilken Mig1 fosforyleras är noll. (b) Jämförelser mellan de modellerade och de experimentellt framtagna värdena för Mig1 respektive Suc2. För Mig1 ger modellen en betydligt större minskning än den experimentella datan visar. Vid  $t=140$  min ger modellen att intensiteten av Mig1 börjar öka igen. Den experimentella datan visar en viss minskning i intensitet i början för att sedan ligga någorlunda konstant tills  $t=200$  min där intensiteten först minskar för att sedan öka igen. Efter  $t=300$  finns det ingen experimentell data att jämföra med. För Suc2 följer modellen någorlunda den experimentella datan men med vissa avvikelser. Modellen fångar till viss del den initiala minskningen i intensiteten men missar "puckeln" vid  $t=250$  min samt ökningen mot slutet. (c) Residualanalys för Mig1. Alla residualer är positiva och det finns en tydlig trend i residualernas storlek vilket tyder på ett systematiskt fel. Efter  $t=300$  min finns inga uppmätta värden för Mig1. (d) Residualanalys för Suc2. Residualerna är någorlunda jämt fördelade runt noll men det finns en oscillerande trend vilket kan tyda på ett systematiskt fel.

Det kan ses att modellen fångar datan för Suc2 bättre än den för Mig1. Detta är extra tydligt i Figur 5c och 5d där residualen mellan respektive modell och dess data är uppritad.

#### 4.1.1 Dynamiken för Modell 1 är stabil och kan antingen oscillera eller gå mot styv jämvikt

Efter genomförd stabilitetsanalys, se Appendix F.1, konstateras att för  $k_1$ ,  $k_4$ ,  $k_6$  och  $k_7 > 0$  är Modell 1 stabil. Analysen visar att egenvärdena antingen blir komplexa med negativ realdel eller reellt negativa. De båda möjligheterna bidrar till att modellen antingen går mot en styv jämvikt eller kommer oscillera beroende på värdet hos parametrarna. Figur 6 visar att  $k_5$  har stor betydelse för när modellen börjar svänga.



**Figur 6:** Det blå området motsvarar en styv modell och det vita en oscillerande modell. Graferna visar vilka intervall för respektive parameter som ger oscillerande eller styv modell. Resterande parametrar hålls konstanta vid det optimerade värdet från Tabell 2

Med de skattade parametervärdena i Tabell 2 går Modell 1 mot en styv jämvikt. Det innebär att respektive tillstånd kommer gå mot jämviktsläget som presenteras i Tabell 3. Vid jämförelse med resultatet som presenteras i Figur 5a ses att intensiteten går mot just dessa värden.

**Tabell 3:** Hur mycket som finns, enligt stabilitetsanalysen, av respektive tillstånd när systemet i Modell 1 uppnått jämvikt.

| Tillstånd | Mängd vid jämvikt [Intensitet] |
|-----------|--------------------------------|
| Mig1      | 2,24                           |
| Mig1p     | 0,0                            |
| Suc2      | 4,57                           |
| X         | 11,7                           |

#### 4.1.2 Givet data för Suc2 och Mig1 är inte nedbrytningen av Mig1 identifierbar

Strukturell identifierbarhetsanalys utfördes, se Appendix G.1, och där framkommer det att parametern  $k_7$  inte är identifierbar. Det medför att Modell 1 inte heller är numeriskt identifierbar och därför har inte en sådan analys utförts på denna modell. Eftersom just  $k_7$  inte är identifierbar kan det tolkas som att datan inte kan beskriva någon nedbrytning av Mig1. Resterande parametrar som tillsammans beskriver fosforylering, defosforylering och en feedback-loop kan däremot identifieras givet den data som finns och därför kan resterande struktur bevaras.

#### 4.1.3 Störst inverkan har nedbrytnings- och inhiberingstermer

Vid känslighetsanalys på de optimerade parametrarna, se Tabell 4, kan det noteras att  $\hat{\theta}_4$  har en större påverkan på Suc2 och likvis har  $\hat{\theta}_5$ ,  $\hat{\theta}_6$  och  $\hat{\theta}_7$  det på Mig1. Alltså är modellen mer känslig för ändringar av de här parametrarna. Med modellens utseende och biologiska mening i åtanke är det här logiskt med tanke på att  $k_4$  och  $k_7$  är nedbrytningstermer direkt kopplade till Suc2 och Mig1. Mer intressant är hur parametern  $k_5$  är kopplad till Mig1. Mig1 är positivt ökande med ökande intensiteter av X medan det reciproka gäller för X. Det här kan orsaka en svängning och känslighetsanalysen visar därför precis som stabilitetsanalysen, att  $k_5$  spelar en stor roll i när den här svängningen sker. Notera även att  $\hat{\theta}_1$  inte påverkar modellens värden alls, vilket gör dess skattning mindre trovärdig.

**Tabell 4:** RMS-värden för Suc2 och Mig1 i Modell 1

| RMS              | Suc2  | Mig1                 |
|------------------|-------|----------------------|
| $\hat{\theta}_1$ | 0,00  | 0,00                 |
| $\hat{\theta}_2$ | -     | -                    |
| $\hat{\theta}_3$ | 0,988 | $2,20 \cdot 10^{-6}$ |
| $\hat{\theta}_4$ | 2,08  | $2,12 \cdot 10^{-7}$ |
| $\hat{\theta}_5$ | 0,473 | 0,505                |
| $\hat{\theta}_6$ | 0,328 | 0,351                |
| $\hat{\theta}_7$ | 0,461 | 0,496                |

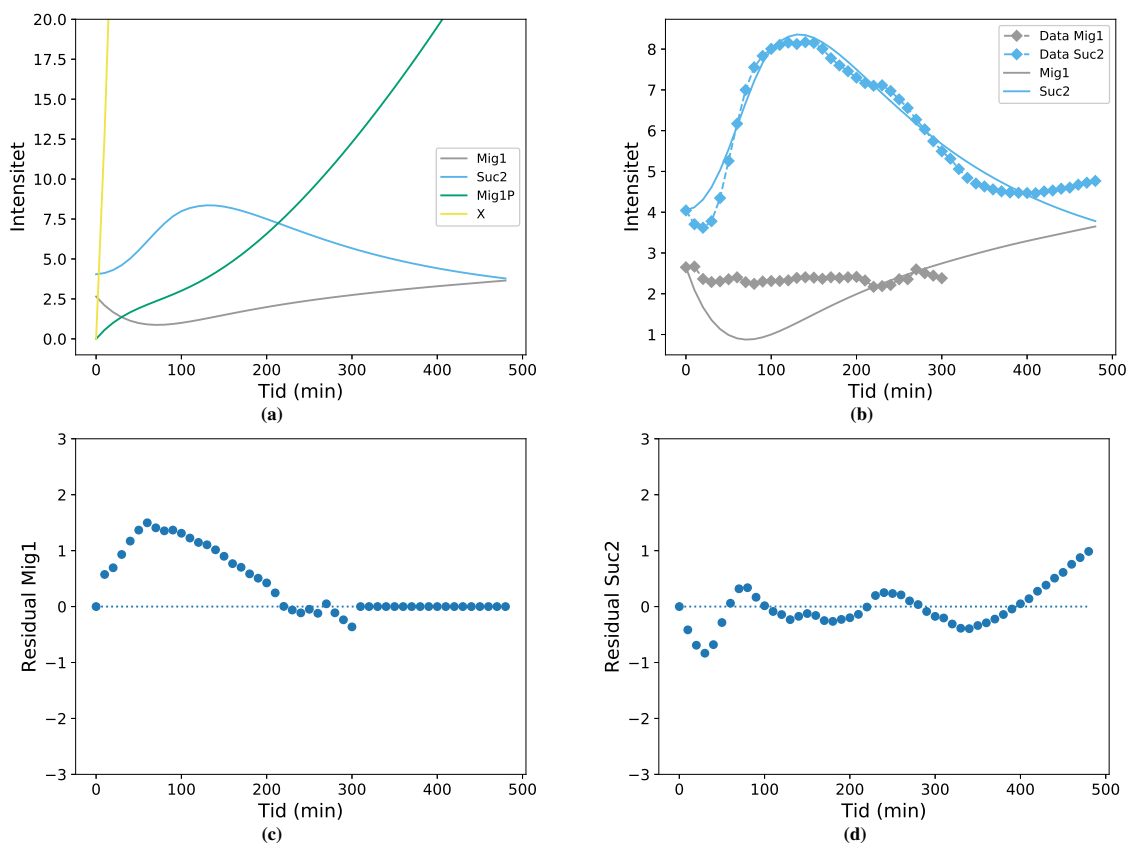
## 4.2 Den konstruerade modellen reducerades till att ha en likvärdig nedbrytning av Mig1 och X

Eftersom Modell 1 inte var strukturellt identifierbar gjordes två olika försök att reducera den. I det första försöket antogs nedbrytningen av Mig1 och X ske med samma hastighet, eftersom X kan ses som en omskrivning av Mig1 uttryckt av genen *MIG1*. Matematiskt resulterar det i att  $k_6 = k_7$ , vilket leder till att  $r_7 = k_6 \cdot Mig1$ , resterande hastighetsekvationer förblir på samma form som i Tabell 1. Med detta insatt i ekvation 23 och med samma antaganden som presenteras där erhålls Modell  $k_6 = k_7$ .

Optimering av Modell  $k_6 = k_7$  resulterade, efter 59 startgissningar, i en bästa parameteruppskattning som kan ses i Tabell 5. Med de värdena insatta i modellen erhöles aktiviteterna som ses i Figur 7a och mer specifika jämförelser mellan datan och modellerade värden för Mig1 och Suc2 kan ses i Figur 7b. Det kan ses att modellen fångar datan för Suc2 bättre än den för Mig1. Detta är extra tydligt i Figur 7c och 7d där beloppet av residualen för Suc2 är mycket mindre än det för Mig1 vid de flesta tidpunkter.

**Tabell 5:** Kostnadsfunktionen, den viktade kostnadsfunktionen samt de skattade parametrarna för Modell  $k_6 = k_7$ .

| $S(\hat{\theta})$ | $S_{viktad}(\hat{\theta})$ | $\hat{\theta}_1$ | $\hat{\theta}_2$     | $\hat{\theta}_3$ | $\hat{\theta}_4$     | $\hat{\theta}_5$ | $\hat{\theta}_6$     | $\hat{\theta}_7$ |
|-------------------|----------------------------|------------------|----------------------|------------------|----------------------|------------------|----------------------|------------------|
| 29,3              | 691                        | 0,00             | $2,39 \cdot 10^{-2}$ | 0,122            | $1,03 \cdot 10^{-2}$ | 3,16             | $1,26 \cdot 10^{-4}$ | -                |



**Figur 7:** Resultat för Modell  $k_6 = k_7$ . (a) De olika aktiviteterna givna av Modell  $k_6 = k_7$ . Det kan ses att intensiteten för X drar iväg snabbt i början och ökar under tidsintervallet till ett värde runt 700 (visas inte i figuren). Även Mig1p drar iväg runt  $t=150$  min och når en intensitet på 25 (visas inte i figuren). (b) Jämförelse mellan de modellerade och de experimentellt framtagna värdena för Mig1 och Suc2. För Mig1 ger modellen en betydligt större minskning än den experimentella datan visar. Vid  $t=80$  min ger modellen att intensiteten av Mig1 börjar öka igen och vid  $t=280$  min är den modellerade intensiteten högre än den uppmätta. Den experimentella datan visar en viss minskning i intensitet i början för att sedan ligga någorlunda konstant tills  $t=200$  min där intensiteten först minskar för att sedan öka igen. För Suc2 följer modellen den experimentella datan någorlunda men med vissa avvikelser. Modellen fångar inte den initiala minskningen i intensiteten och missar "puckeln" vid  $t=250$  min. Modellen ger även en fortsatt minskning av intensiteten av Suc2 mot slutet medan datan visar på en ökning efter  $t=400$  min. (c) Residualanalys för Mig1. Alla residualer fram till  $t=220$  är positiva och det finns en tydlig trend i residualernas storlek vilket tyder på ett systematiskt fel. Efter  $t=300$  min finns inga uppmätta värden för Mig1. (d) Residualanalys för Suc2. Residualerna är någorlunda jämt fördelade runt noll men det finns en oscillerande trend vilket kan tyda på ett systematiskt fel. Mot slutet kan det även ses en tydlig uppgående trend.

#### 4.2.1 $\hat{\theta}_1 = 0$ och därför visar stabilitetsanalysen att dynamiken för Modell $k_6 = k_7$ inte är stabil

Stabilitetsanalysen visar att för de skattade parametrarna är den reducerade modellen inte stabil. Eftersom skattningen  $\hat{\theta}_1 = 0$  innebär att  $k_1 = 0$ , kommer jämviktskoncentrationen av Mig1p gå mot oändligheten, se Appendix F.2.

#### 4.2.2 Stor varians hos parametrarna för Mig1

Modellen visar sig vara strukturellt identifierbar, se Appendix G.2, och numerisk identifierbarhetsanalys utfördes. Genomförd analys visar att variationskoefficienterna för de optimerade parametrarna gav noterbart olika resultat för Suc2 och Mig1, se Tabell 6. Parametern  $k_1$  sattes till  $\hat{\theta}_1 = 0$  i optimeringen och därför är variationskoefficienten inte definierad. Det kan noteras att det främst är  $\hat{\theta}_3$  och  $\hat{\theta}_4$  som har en väldigt stor standardavvikelse, ungefär hundra gånger större än parametervärdet. Gällande Suc2 är det främst  $\hat{\theta}_5$  och  $\hat{\theta}_6$  som har hög varians, dessa är även höga för Mig1. Det finns alltså en stor osäkerhet för nästan alla parametrar i denna modell. Även vid vidare analys med hjälp av korrelationsmatriserna, som presenteras i Figur 8, så existerar det en hög korrelation mellan flera parametrar. Resultatet är att modellen inte är numeriskt identifierbar och kan därmed inte användas för att göra prediktioner.

### 4.2.3 Den observerade datan är okänslig för ändring av parametrar

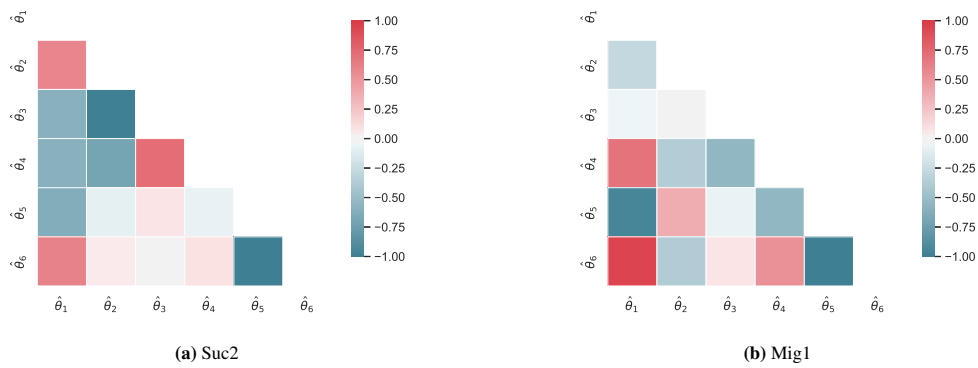
Från känslighetsanalysen återfanns låga RMS-värden för alla parametrar, se Tabell 7. För Suc2 är det ingen noterbar skillnad bland parametrarna till skillnad från Mig1 där storleksordningen för parametrarna skiljer sig med en faktor  $10^3$ . För Mig1 är RMS för  $\hat{\theta}_3$  och  $\hat{\theta}_4$  väldigt låga, vilket kan ha gjort parametrarna svåra att identifiera.

**Tabell 6:** Variationskoefficienter för Suc2 och Mig1 då  $k_6 = k_7$

| $\sigma_i$       | Suc2  | Mig1  |
|------------------|-------|-------|
| $\hat{\theta}_1$ | -     | -     |
| $\hat{\theta}_2$ | 0,606 | 0,387 |
| $\hat{\theta}_3$ | 0,527 | 283   |
| $\hat{\theta}_4$ | 0,255 | 111   |
| $\hat{\theta}_5$ | 7,26  | 23,5  |
| $\hat{\theta}_6$ | 7,12  | 23,2  |

**Tabell 7:** RMS-värden för Suc2 och Mig1 då  $k_6 = k_7$

| RMS              | Suc2  | Mig1                 |
|------------------|-------|----------------------|
| $\hat{\theta}_1$ | -     | -                    |
| $\hat{\theta}_2$ | 0,393 | 0,481                |
| $\hat{\theta}_3$ | 0,881 | $3,13 \cdot 10^{-4}$ |
| $\hat{\theta}_4$ | 0,945 | $2,30 \cdot 10^{-3}$ |
| $\hat{\theta}_5$ | 0,435 | 0,537                |
| $\hat{\theta}_6$ | 0,425 | 0,523                |



**Figur 8:** Heat maps som representerar korrelationsmatriserna för Suc2 (a) och Mig1 (b) för Modell  $k_6 = k_7$ . Här är ett värde nära 1 representerat med en mörkare färg av röd och ett värde nära -1 representerat med en mörkare färg av blå.

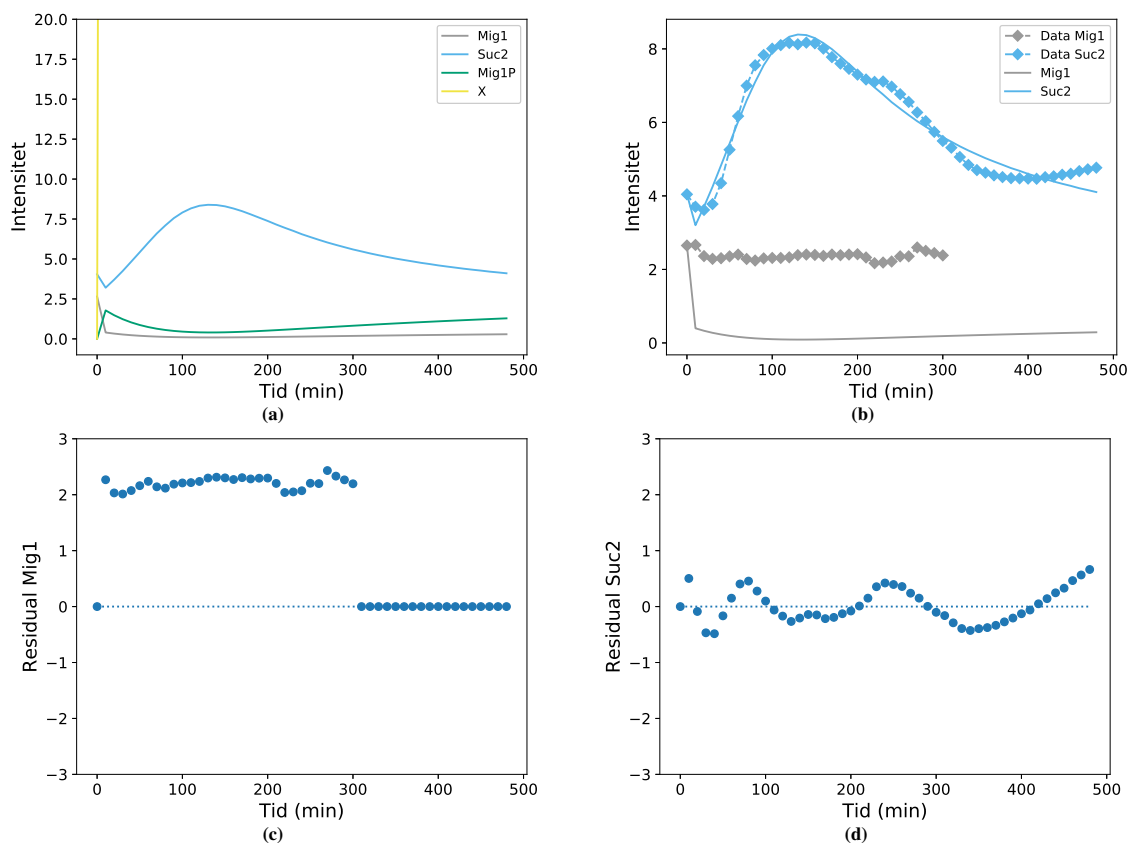
### 4.3 En alternativ reducering testades där SUC2 och X har en likvärdig inhibering

I det andra försöket att reducera Modell 1 antogs att uttrycket av Mig1 inhiberar SUC2 och X lika mycket. Antagandet baseras på att det är samma protein som utför inhiberingen, och därför kan inhiberingen ske på ett likvärdigt sätt. I den matematiska modellen betyder det att  $k_3 = k_5$ , vilket leder till att  $r_5 = \frac{k_3}{Mig1+0,1}$ , resterande hastighetsekvationer förblir på samma form som i Tabell 1. Med detta insatt i ekvation 23 och med samma antaganden som presenteras där erhålls Modell  $k_3 = k_5$ .

Optimering av Modell  $k_3 = k_5$ , resulterade efter 30 startgissningar, i en bästa parameteruppskattning som kan ses i Tabell 8. Med de värdena insatta i modellen erhöles aktiviteterna som ses i Figur 9a och mer specifika jämförelser mellan datan och modellerade värden för Mig1 och Suc2 kan ses i Figur 9b. Det kan ses att modellen fångar datan för Suc2 bättre än den för Mig1. Detta är extra tydligt i Figur 9c och 9d där residualen mellan respektive modell och dess data är uppritad.

**Tabell 8:** Kostnadsfunktionen, den viktade kostnadsfunktionen samt de skattade parametrarna för Modell  $k_3 = k_5$ .

| $S(\hat{\theta})$ | $S_{viktad}(\hat{\theta})$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ | $\hat{\theta}_5$ | $\hat{\theta}_6$     | $\hat{\theta}_7$ |
|-------------------|----------------------------|------------------|------------------|------------------|------------------|------------------|----------------------|------------------|
| 151               | 584                        | 2,26             | 10,0             | 11,6             | 7,28             | -                | $1,58 \cdot 10^{-6}$ | 0,103            |



**Figur 9:** Resultat för Modell  $k_3 = k_5$ . (a) De olika aktiviteterna givna av Modell  $k_3 = k_5$ . Det kan ses att intensiteten för X drar iväg snabbt i början och ökar under tidsintervallet till ett värde runt 20 000 (visas inte i figuren). Mig1p ökar snabbt initialt för att sedan minska fram till  $t=150$  min varefter intensiteten ökar igen. (b) Jämförelse mellan de modellerade och de experimentellt framtagna värdena för Mig1 och Suc2. För Mig1 ger modellen att intensiteten minskar snabbt i början för att sedan sakta in och börja öka långsamt vid  $t=150$  min. Den experimentella datan visar en viss minskning i intensitet i början för att sedan ligga någorlunda konstant tills  $t=200$  min där intensiteten först minskar för att sedan öka igen. För Suc2 följer modellen någorlunda den experimentella datan men med vissa avvikelser. Modellen fångar någorlunda den initiala minskningen i intensitet men missar "puckeln" vid  $t=250$  min. Modellen ger även en fortsatt minskning av intensiteten av Suc2 mot slutet medan datan visar på en ökning efter  $t=400$  min. (c) Residualanalys för Mig1. Alla residualer är positiva och det finns en tydlig trend då residualerna ligger på ett någorlunda konstant värde, detta tyder på ett systematiskt fel. Efter  $t=300$  min finns inga uppmätta värden för Mig1. (d) Residualanalys för Suc2. Residualerna är någorlunda jämt fördelade runt noll men det finns en oscillerande trend vilket kan tyda på ett systematiskt fel. Mot slutet kan det även ses en viss uppgående trend.

#### 4.3.1 Dynamiken hos Modell $k_3 = k_5$ är stabil och kommer oscillera kring jämviktsläget

Stabilitetsanalysen visar att modellen är stabil för de skattade parametrarna. Till skillnad från Modell 1 så kommer den att oscillera oavsett parametervärde, vilket innebär att den gör det för den optimerade parameterskattningen. Innebörden av det är att intensiteten inte helt stabiliseras vid ett värde utan istället svänger kring värdet med en avtagande oscillation. De värden som respektive tillstånd svänger kring beräknas i Appendix F.3 och presenteras i Tabell 9. Där ses att mängden för X är mycket större än de övriga, något som även Figur 9a visar. Ur Figur 9a ser det ut som att X går mot oändligheten men stabilitetsanalysen visar att så inte är fallet, utan X kommer att stabiliseras men vid ett mycket högre värde än övriga tillstånd.

**Tabell 9:** Hur mycket som finns, enligt stabilitetsanalysen, av respektive tillstånd när systemet i Modell  $k_3 = k_5$  uppnått oscillerande jämvikt.

| Tillstånd | Mängd vid jämvikt [Intensitet] |
|-----------|--------------------------------|
| Mig1      | 10,6                           |
| Mig1p     | 46,9                           |
| Suc2      | 0,23                           |
| X         | $1,06 \cdot 10^6$              |

#### 4.3.2 Den reducerade modellen är strukturellt identifierbar men inte numeriskt identifierbar

Resultatet från identifierbarhetsanalys vid reduceringen  $k_3 = k_5$  visar att modellen är strukturellt identifierbar och detta presenteras i Appendix G.3. Därför utförs även en numerisk identifierbarhetsanalys på de optimerade parametrarna  $\hat{\theta}$  där resultatet visas i Tabell 10 och Figur 10. Från variationskoefficienterna ses det att värdet överskrider 1 på en del parametrar för Mig1, detta betyder att standardavvikelsen är större än det skattade parametervärdet. Framför allt kan det noteras att variationskoefficienten för  $\hat{\theta}_4$  har en standardavvikelse som är 200 gånger större än parametervärdet. Med andra ord finns det en stor osäkerhet i parametrarna för Mig1. Däremot är variationskoefficienterna för Suc2 inom rimliga gränser. Vidare undersökning av korrelationsmatrisen ger ett liknande resultat, se Figur 10. Här finns en stark korrelation mellan flera olika parametrar vilket gör modellen oidentifierbar. Därmed kan inte dessa parameterpar urskiljas från varandra och de verkliga värdena går inte att bestämma numeriskt. Det gäller både för Suc2, som har stark korrelation mellan många parametrar, och Mig1 som har ett par.

#### 4.3.3 Känslighetsanalys för den reducerade modellen visar en känslighet för ändring av $\hat{\theta}_2$

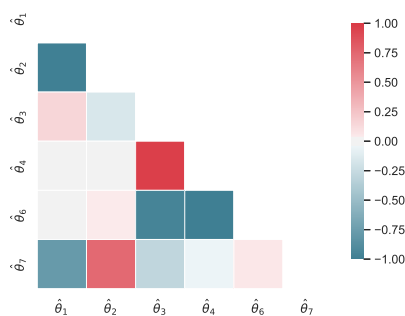
Från känslighetsanalysen kan det åskådas att både Mig1 och Suc2 har störst värden i RMS för  $\hat{\theta}_2$ , se Tabell 11, som representerar fosforylering av Mig1 till Mig1p. Resterande värden är i samma storleksordning, med undantaget  $\hat{\theta}_4$  för Mig1, således verkar de ha en likvärdig effekt på modellen. Parametern  $\hat{\theta}_4$  har ett värde som är en tusendel av de mer influentiella parametrarna vilket kan göra den icke-identifierbar. Tolkningen av att  $\hat{\theta}_2$  har ett relativt högt värde ges av modellstrukturen. Där syns det att eftersom  $k_3 = k_5$  kommer Mig1-intensiteten ha en större påverkan på X vilket Mig1 ökar positivt med. Likaså kommer Suc2 börja sjunka vid en viss tidpunkt beroende på  $k_3$ . I sin tur är Mig1-intensiteten beroende av  $k_2$ , som skattades till ett relativt stort tal, se Tabell 8.

Tabell 10: Variationskoefficienter för Suc2 och Mig1

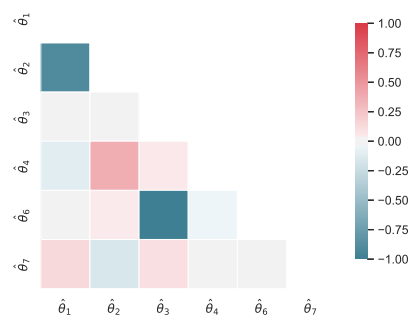
| $\sigma_i$       | Suc2   | Mig1  |
|------------------|--------|-------|
| $\hat{\theta}_1$ | 0,196  | 1,66  |
| $\hat{\theta}_2$ | 0,0425 | 0,421 |
| $\hat{\theta}_3$ | 0,715  | 13,7  |
| $\hat{\theta}_4$ | 0,680  | 289   |
| $\hat{\theta}_6$ | 0,681  | 13,7  |
| $\hat{\theta}_7$ | 0,290  | 1,45  |

Tabell 11: RMS-värden för Suc2 och Mig1 då  $k_3 = k_5$

| RMS              | Suc2  | Mig1                 |
|------------------|-------|----------------------|
| $\hat{\theta}_1$ | 0,705 | 0,753                |
| $\hat{\theta}_2$ | 3,09  | 3,19                 |
| $\hat{\theta}_3$ | 0,662 | 0,620                |
| $\hat{\theta}_4$ | 0,989 | $2,36 \cdot 10^{-3}$ |
| $\hat{\theta}_6$ | 0,394 | 0,619                |
| $\hat{\theta}_7$ | 0,398 | 0,680                |



(a) Suc2



(b) Mig1

Figur 10: Heat maps som representerar korrelationsmatriserna för Suc2 (a) och Mig1 (b) i Modell  $k_3 = k_5$ . Här är ett värde nära 1 representerat med en mörkare färg av röd och ett värde nära -1 representerat med en mörkare färg av blå.

## 5 Diskussion

Resultatet visar att Modell 1 inte klarar av att beskriva den biologiska mekanism som presenterades i syftet, där hypotesen var att cytosoliskt Mig1 reglerar Suc2-produktion vid svält av jästceller. En anledning till att modellen inte fungerar för att spegla beteendet är att optimering av Modell 1 skattar värdet av parametern  $k_2$  till  $\hat{\theta}_2 = 0$ , se Tabell 2. Det innebär att inget Mig1p skapas och intensiteten av Mig1p är lika med noll under hela förloppet. Biologiskt kan det tolkas som att inget Mig1 fosforyleras eller exporteras till cytoplasman, vilket inte överstämmer med hur jästceller reagerar på svält [20]. I sin tur hade detta inneburit att den Mig1-beroende inhiberingen av *SUC2* inte beror av fosforylerings- eller spatialtillståndet av Mig1, något som går emot tidigare teori. Resultatet kan också tolkas som att fosforyleringen av Mig1 spelar en så liten roll i regleringen av *SUC2* att de enda signifikanta mekanismerna för *SUC2*-repression genom Mig1 är dess nybildning genom X ( $k_6$ ) och dess egna nedbrytning ( $k_7$ ). Om samma resonemang dras längre skulle detta betyda att flera av mekanismerna uppströms i signalvägen, bland annat SNF1:s reglering av Mig1, inte spelar någon roll i hur Mig1 påverkar uttrycket av *SUC2*, vilket inte stämmer överens med tidigare teori. Ur en optimeringssynpunkt beror skattningen  $\hat{\theta}_2 = 0$  förmodligen på att det inte finns data för Mig1p att jämföra med vilket gör att jämvikten mellan Mig1 och Mig1p är svår att sätta upp. Det blir även svårt att balansera denna mot nedbrytningen respektive nybildningen av Mig1. Modell 1 är inte heller strukturellt identifierbar på grund av att  $k_7$  inte uppvisar identifierbarhet, se Appendix G.1, vilket gör det svårt att dra slutsatser, speciellt kring oidentifierbara parametrar.

Vidare tyder residualerna för de modellerade tillstånden Mig1 och Suc2 i Figur 5c respektive Figur 5d på att Modell 1 inte följer beteendet hos datan helt. Det är förväntat till viss grad på grund av eventuellt brus i datan som inte kunde filtreras bort, se avsnitt 3.1, men residualerna uppvisar också vissa mönster som tyder på att modellen är otillräcklig. Residualerna för Suc2 verkar oscillera med tiden, vilket kan betyda att det finns ett snabbt oscillerande beteende av Suc2 som inte fångas av modellen. En snabb analys av datan genom Figur 3 visar dock att mitten av tidsserien är ett område med stor variation i Suc2 mellan enskilda celler. Därför är det möjligt att små mönster under det intervallet, till exempel det minskande avtagandet av medelvärde runt tiden 200 min, är uppkommet av slumpmässiga variationer. Oavsett divergerar Suc2 för modellen gentemot datan starkt även i slutet av tidsserien, vilket tyder på att det finns mekanismer som inte fångas av modellen.

Eftersom  $k_7$  inte var strukturellt identifierbar utformades den första reducerade modellen med antagandet att nedbrytningen av Mig1 och X sker med samma hastighet, alltså att  $k_6 = k_7$ . Problemet som uppstod var att skattningen av hastighetskonstanten  $k_1$  gav  $\hat{\theta}_1 = 0$ . Då  $k_1$  reglerar hastigheten för defosforylering av Mig1p resulterar skattningen i att Mig1, till en början, sjunker väldigt snabbt, vilket i sin tur resulterar i att X ökar fort i jämförelse med resterande värden. Det här visas i Figur 7a som illustrerar en snabb uppgång för både Mig1p och X. Dessutom kommer Mig1p inte stabilisera sig enligt stabilitetsanalysen då det inte längre uppstår en jämvikt mellan Mig1 och Mig1p. Det resulterar även i att nedbrytningen av Mig1 och fosforyleringen av Mig1, från optimeringens synpunkt, är likvärdiga för att minska intensiteten av Mig1. Därför blir parametrarna  $k_2$  och  $k_6$  svåra att skatta.

För övrigt är passningen även här bra för Suc2 men modellen fångar inte uppgången mot slutet. Gällande Mig1 är passningen dålig och ger en alltför parabolisk form med stor initial minskning av Mig1. Vid visualisering av residualerna kan det synas att det existerar trender bland felen, speciellt för Mig1, se Figur 7c och 7d. Det här felet visar sig även hos den numeriska identifierbarhetsanalysen. Här är variationskoefficienterna högre än önskvärt, framför allt för Mig1 men även för några parametrar i Suc2. Slutresultatet blev en stor variation bland nästan alla parametrar och deras skattning är därför inte trovärdig. Det här styrks även av korrelationsmatrisen där höga korrelationer kunde åskådas i båda tillstånden vilket gör Modell  $k_6 = k_7$  oidentifierbar. Vid känslighetsanalysen visas liknande resultat då en del parametrar har betydligt lägre RMS-värden än andra, vilket gör skattningar svåra att lita på.

I den sista modellen, Modell  $k_3 = k_5$ , ger stabilitetsanalysen en mycket hög jämviktskoncentration för X men trots det var modellen stabil, alltså kommer koncentrationen till slut stabilisera sig om än vid ett högt värde. Vid jämförelse med experimentell data passar Suc2 datan men modellen lyckas inte fånga kurvan i slutet av datan, se Figur 9b. För Mig1 passar datan dåligt och även formen av kurvan lyckades inte fånga beteendet, det här kan ses i Figur 9b. Den trenden som fanns i felet som uppvisas i tidigare modell syns även här vid undersökning av residualerna, se Figur 9c och 9d. Vidare gav optimeringen av Modell  $k_3 = k_5$  en väldigt snabb ökning av intensiteten av X. Det beror på att antagandet  $k_3 = k_5$  ger att nybildandet av X är lika stort som nybildandet av Suc2 och då det finns data att jämföra med för Suc2 men inte för X kommer

parametern  $k_3$  att anpassas för att passa Suc2 datan. Nedbrytningen av X till Mig1 regleras av  $k_6$  men den parametern kommer endast att anpassas för att passa Mig1 datan, även om den passning inte blir bra.

Precis som för ovanstående modeller är inte Modell  $k_3 = k_5$  numeriskt identifierbar. Även om variationskoefficienterna är inom rätt intervall för Suc2 visar korrelationsmatrisen att en del parametrar är starkt korrelerade. Liknande beteende syns hos Mig1 som har stora parametervarianser och starkt korrelerade parametrar.

Ett problem som uppstår i samtliga modeller är att de inte tycks fånga hur uttrycket av Suc2 vänder uppåt igen mot slutet av provtagningarna, se Figur 5b, 7b och 9b. Stabilitetsanalysen visar dock på att Modell 1 kan oscillera för vissa parametervärden och Modell  $k_3 = k_5$  gör det för alla parametervärden. Det innebär att de har potential att fånga beteendet hos Suc2, vilket kan vara intressant vid vidare utveckling av modellen. Att modellen har den oscillerande egenskapen är mycket rimligt ur biologisk synpunkt då det är ett vanligt förekommande fenomen, till exempel vid feedback-inhibering [1]. Om modellen fortsätter utvecklas kan även den mycket höga jämviktskoncentrationen för X vara av intresse att modellera om. Eftersom koncentrationen är beräknad efter de skattade parametrarna kan villkor vid parametereskattningen leda till att den inte ökar lika snabbt. Ett villkor som skulle kunna införas är att  $k_6$  måste vara högre än vad det är nu, se Tabell 8, då det är den parametern som styr nedbrytningen av X. Ett perspektiv att ha i åtanke är att de olika tillstånden (Mig1, Mig1p, Suc2, X) har modellerats utifrån vissa biologiska antaganden, och tillstånden kan därför få viss subjektiv innebörd som är viktigt att vara medveten om. För att illustrera detta, se ekvationerna 23a och 23b: mellan Mig1 och Mig1p är skillnaden att det sistnämnda är det fosforylerade tillståndet av det första, och därmed den inaktiva formen av Mig1. I modellernas uppbyggnad finns det dock utrymme för alternativa tolkningar. En sådan tolkning är att Mig1p som tillstånd i modellen inte nödvändigtvis är fosforylerat Mig1, utan kanske endast Mig1 som inte kommer åt att inhibera *SUC2*, till exempel cytosoliskt Mig1. Hastighetskonstanterna  $k_1$  och  $k_2$  skulle i detta fall kunna modellera hur snabbt Mig1 åker in respektive ut ur cellkärnan, till skillnad från att modellera hur snabbt Mig1 defosforyleras respektive fosforyleras. Det kan också noteras att dessa tolkningar inte behöver utesluta varandra.

Det kan även riktas en viss skepticism mot datan för Mig1 eftersom den inte följer beteendet som tidigare publikationer, såsom Bendrioua et. al., har funnit [20]. Dock är datan insamlad under en längre tid än i tidigare publikationer samt att bara mängden Mig1 i cellkärnan är uppmätt, inte förhållandet mellan cellkärnan och cytoplasman. Skillnaderna försvårar jämförelsen mellan den använda datan för Mig1 och data använt i tidigare publikationer. Vid en närmare undersökning av den använda datan, se Appendix D Figur 1, kan det ses att för de första 60-70 minuterna skiljer sig datan inte mycket från den använda i Bendrioua et. al. [20], förutom att den är mer utdragen. Om det finns fel i den använda datan kan det ha orsakat problem i optimeringen då det riktiga beteendet av Mig1 inte finns bland datan som ska passas mot våra antaganden. Det är också möjligt att de antaganden som görs och mekanismer som sätts upp i rapporten inte är giltiga under hela det undersökta tidsintervallet.

Sammantaget passar alla modeller Suc2-datan bra, de lyckas inte fånga det oscillerande beteendet till fullo men de har potential att göra det och gällande Mig1 passade ingen av modellerna datan. Alla modeller är oidentifierbara vilket innebär att de inte kan användas för prediktioner. Däremot var två av modellerna stabila och de kan oscillera. Eftersom de även passar datan för Suc2 kan de bli bra om de justeras till att bli numeriskt identifierbara. Det kan konstateras att resultatet inte visar på något som stödjer eller motsäger en inhibering av Suc2 via cytosoliskt Mig1 i SNF1-signalvägen. Däremot finns utvecklingsmöjligheter, kanske i form av liknande modeller med andra parametrar eller i form av nya experiment som fångar tillståndens beteende bättre.

## Referenser

1. Klipp, E., Herwig, R., Kowald, A., Wierling, C. & Lehrach, H. *Systems Biology in Practice, Concepts, Implementation and Application* ISBN: 9783527310784. doi:10.1002/3527603603 (WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim).
2. Wishart, D. S. *m.fl.* HMDB 4.0: the human metabolome database for 2018. eng. *Nucleic acids research* **46**, D608–D617. ISSN: 1362-4962. doi:10.1093/nar/gkx1089 (jan. 2018).
3. Botstein, D. & Fink, G. R. Yeast: an experimental organism for 21st Century biology. eng. *Genetics* **189**, 695–704. ISSN: 1943-2631. doi:10.1534/genetics.111.130765 (nov. 2011).
4. Efeyan, A., Comb, W. C. & Sabatini, D. M. Nutrient-sensing mechanisms and pathways. *Nature* **517**, 302–310. ISSN: 1476-4687. doi:10.1038/nature14190 (2015).
5. Salminen, A. & Kaarniranta, K. AMP-activated protein kinase (AMPK) controls the aging process via an integrated signaling network. *Ageing Research Reviews* **11**, 230–241. ISSN: 1568-1637. doi:10.1016/J.ARR.2011.12.005 (april 2012).
6. Hedbacker, K. & Carlson, M. SNF1/AMPK pathways in yeast. eng. *Frontiers in bioscience : a journal and virtual library* **13**, 2408–2420. ISSN: 1093-9946. doi:10.2741/2854 (jan. 2008).
7. Welkenhuysen, N., Schnitzer, B., Österberg, L. & Cvijovic, M. Robustness of nutrient signaling is maintained by interconnectivity between signal transduction pathways. *Frontiers in Physiology* **10**, 1–13. ISSN: 1664042X. doi:10.3389/fphys.2018.01964 (2019).
8. DiStefano III, J. *Dynamic Systems Biology and Simulation* ISBN: 9788578110796 (Academic Press, 2015).
9. Alberts, B. *m.fl.* *Molecular Biology of the Cell* 6th. ISBN: 9780815344643 (Garland Publishing Inc, 2014).
10. Klipp, E. & Liebermeister, W. Mathematical modeling of intracellular signaling pathways. *BMC Neuroscience* **7**, 1–16. ISSN: 14712202. doi:10.1186/1471-2202-7-S1-S10 (2006).
11. Hardie, D. G. AMP-activated/SNF1 protein kinases: Conserved guardians of cellular energy. *Nature Reviews Molecular Cell Biology* **8**, 774–785. ISSN: 14710072. doi:10.1038/nrm2249 (2007).
12. Lutfiyya, L. L. *m.fl.* Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*. *Genetics* **150**, 1377–1391. ISSN: 00166731 (1998).
13. Broach, J. R. Nutritional control of growth and development in yeast. *Genetics* **192**, 73–105. ISSN: 00166731. doi:10.1534/genetics.111.135731 (2012).
14. Raue, A. *m.fl.* Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLoS ONE* **8** (utg. Hernandez-Lemus, E.) e74335. ISSN: 1932-6203. doi:10.1371/journal.pone.0074335 (sept. 2013).
15. Kitano, H. Computational systems biology. *Nature* **420**, 206–210. ISSN: 1476-4687. doi:https://doi.org/10.1038/nature01254 (2002).
16. McCartney, R. R., Garnar-Wortzel, L., Chandrashekarappa, D. G. & Schmidt, M. C. Activation and inhibition of Snf1 kinase activity by phosphorylation within the activation loop. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1864**, 1518–1528. ISSN: 18781454. doi:10.1016/j.bbapap.2016.08.007 (2016).
17. Rubenstein, E. M. *m.fl.* Access denied: Snf1 activation loop phosphorylation is controlled by availability of the phosphorylated threonine 210 to the PP1 phosphatase. *Journal of Biological Chemistry* **283**, 222–230. ISSN: 00219258. doi:10.1074/jbc.M707957200 (2008).
18. Treitel, M. A., Kuchin, S. & Carlson, M. Snf1 Protein Kinase Regulates Phosphorylation of the Mig1 Repressor in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **18**, 6273 LP–6280. doi:10.1128/MCB.18.11.6273 (nov. 1998).
19. Shashkova, S., Wollman, A. J. M., Leake, M. C. & Hohmann, S. The yeast Mig1 transcriptional repressor is dephosphorylated by glucose-dependent and -independent mechanisms. *FEMS Microbiology Letters* **364**. ISSN: 0378-1097. doi:10.1093/femsle/fnx133 (juni 2017).
20. Bendrioua, L. *m.fl.* Yeast AMP-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *Journal of Biological Chemistry* **289**, 12863–12875. ISSN: 1083351X. doi:10.1074/jbc.M114.547976 (2014).
21. Ostling, J., Carlberg, M. & Ronne, H. Functional domains in the Mig1 repressor. eng. *Molecular and cellular biology* **16**, 753–761. ISSN: 0270-7306. doi:10.1128/mcb.16.3.753 (mars 1996).

22. Papamichos-Chronakis, M., Gligoris, T. & Tzamarias, D. The Snf1 kinase controls glucose repression in yeast by modulating interactions between the Mig1 repressor and the Cyc8-Tup1 co-repressor. *EMBO Reports* **5**, 368–372. ISSN: 1469221X. doi:10.1038/sj.embor.7400120 (2004).
23. Devore, J. L. *Probability and Statistics for Engineering and the Sciences* 9. utg. ISBN: 9781305251809 (Cengage Learning, Boston, 2015).
24. Andreasson N Evgrafov A, P. M. *An Introduction to Continous Optimization* 1st. ISBN: 9789144044552 (Studentlitteratur).
25. Hartley, H. O. The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares. *Technometrics* **3**, 269–280. ISSN: 15372723. doi:10.1080/00401706.1961.10489945 (1961).
26. Teschl, G. *Ordinary differential equations and dynamical systems* ISBN: 0821883283. doi:10.1090/gsm/140 (American Mathematical Soc., 2012).
27. Kostov, V. P. *Polynomials, sign patterns and Descartes' rule of signs* 2017. doi:10.21136/MB.2018.0091–17.
28. Jeltsch, R. & Mansour, M. *Stability Theory: Hurwitz Centenary Conference Centro Stefano Franscini, Ascona, 1995* ISBN: 978-3-0348-9208-7. doi:https://doi.org/10.1007/978-3-0348-9208-7 (Birkhäuser, 1995).
29. Persson, S., Welkenhuysen, N., Shashkova, S. & Cvijovic, M. Fine-tuning of energy levels regulates SUC2 via a SNF1-dependent feedback loop. *Manuskript under granskning* (2020).
30. Karlsson, J., Anguelova, M. & Jirstrand, M. An Efficient Method for Structural Identifiability Analysis of Large Dynamic Systems. *IFAC Proceedings Volumes* **45**, 941–946. ISSN: 1474-6670. doi:10.3182/20120711-3-BE-2027.00381 (juli 2012).
31. Raue, A., Karlsson, J., Saccomani, M. P., Jirstrand, M. & Timmer, J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **30**, 1440–1448. ISSN: 1367-4803. doi:10.1093/bioinformatics/btu006 (jan. 2014).
32. Hairer, E., Wanner, G. & Nørsett, S. *Solving Ordinary Differential Equations I* 3. utg. ISBN: 978-3-540-56670-0. doi:10.1007/978-3-540-78862-1 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1993).
33. Lay, D. C., Lay, S. R. & McDonald, J. J. *Linear Algebra and its Applications* 5th. ISBN: 9781292092232 (Pearson, Harlow, 2016).
34. Moré, J. J. *The Levenberg-Marquardt algorithm: Implementation and theory* i *Numerical Analysis* (utg. Watson, G. A.) (Springer Berlin Heidelberg, Berlin, Heidelberg, 1978), 105–116. ISBN: 978-3-540-35972-2. doi:https://doi.org/10.1007/BFb0067700.
35. Chiang, C.-Y. & Lin, M. M. The eigenvalue shift technique and its eigenstructure analysis of a matrix. *Journal of Computational and Applied Mathematics* **253**, 235–248. ISSN: 0377-0427. doi:10.1016/J.CAM.2013.04.024 (dec. 2013).
36. Gill, J. & King, G. *What to Do When Your Hessian Is Not Invertible* 2004. doi:10.1177/0049124103262681.
37. Virtanen, P. *m.fl.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272. ISSN: 15487105. doi:10.1038/s41592-019-0686-2 (2020).
38. Travis, O. *A guide to NumPy* 1. utg. (Trelgol Publishing USA, 2006).
39. McKinney, W. *Data Structures for Statistical Computing in Python* i *Proceedings of the 9th Python in Science Conference* **1697900** (2010), 51–56. doi:10.25080/Majora-92bf1922-00a.
40. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95. doi:10.1109/MCSE.2007.55 (2007).
41. Waskom, M. *m.fl.* mwaskom/seaborn: v0.10.1 (April 2020). doi:10.5281/ZENODO.3767070 (april 2020).

## 6 Appendix

### A Härledning av kostnadsfunktion

Parametrarna i modellen, och därmed modellen själv, passas till datan med hjälp av en maximum likelihood (ML) skattning. ML skattningar går ut på att maximera sannolikheten att den modellerade datan kommer från samma fördelning som den experimentella datan. Beräkningen av sannolikheten bygger på att felet,  $\varepsilon = y - \hat{y}(\theta)$ , mellan observationen,  $y$ , och det förväntade värdet,  $\hat{y}(\theta)$ , är normalfördelat, givet tillräckligt många observationer, enligt  $\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$ . Sannolikheten att få ett experimentellt värde givet en parameteruppsättning kan då skrivas som

$$\mathcal{L}(\theta) = p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\left(-\frac{\varepsilon^2}{2\sigma^2}\right)\right\} \quad (\text{A.1})$$

Om flera observationer görs kan ofta de respektive mätfele vara oberoende från varandra. Därmed kan ekvation A.1 utvecklas till ekvation A.2 [23].

$$\begin{aligned} \mathcal{L}(\theta) &= p(\mathbf{y}|\theta) \\ &= p(y_1|\theta) \cdot p(y_2|\theta) \cdot \dots \cdot p(y_m|\theta) \\ &= \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{\left(-\frac{\varepsilon_i^2}{2\sigma_i^2}\right)\right\} \end{aligned} \quad (\text{A.2})$$

För den använda modellen finns flera serier av utvärden,  $\hat{y}_\psi(t_i, \theta)$  och flera serier av experimentella värden,  $y_\psi(t_i)$ . Detta ger  $\varepsilon_{\psi,i} = y_\psi(t_i) - \hat{y}_\psi(t_i, \theta)$  och  $\mathcal{L}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n|\theta)$ . Även här antas det att fele är oberoende och detta kan då skrivas enligt ekvation A.3 [23].

$$\mathcal{L}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n|\theta) = \mathcal{L}(\varepsilon_1|\theta) \cdot \mathcal{L}(\varepsilon_2|\theta) \cdot \dots \cdot \mathcal{L}(\varepsilon_n|\theta) = \prod_{\psi=1}^n \mathcal{L}(\varepsilon_\psi|\theta). \quad (\text{A.3})$$

Vilket ger

$$\mathcal{L}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n|\theta) = \prod_{\psi=1}^n \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_{\psi,i}^2}} \exp\left\{\left(-\frac{(y_\psi(t_i) - \hat{y}_\psi(t_i, \theta))^2}{2\sigma_{\psi,i}^2}\right)\right\}. \quad (\text{A.4})$$

Maximering av  $\mathcal{L}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n|\theta)$  med avseende på  $\theta$  ger då de  $\theta$  som bäst anpassar modellen efter den experimentella datan. Det kan även vara användbart att ta logaritmen av ML-skattningen då omskrivningar från produkter till summor då kan göras och även den exponentiella funktionen förenklas [8]. I denna rapport kommer den negativa logaritmen av ML-skattningen att användas, vilket tillåter följande omskrivningar.

$$S(\theta) = -\log(\mathcal{L}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n | \theta)) = \quad (\text{A.5})$$

$$-\log \left( \prod_{\psi=1}^n \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_{\psi,i}^2}} \exp \left\{ \left( -\frac{(y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2}{2\sigma_{\psi,i}^2} \right) \right\} \right) = \quad (\text{A.6})$$

$$-\log \left( \prod_{\psi=1}^n \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_{\psi,i}^2}} \cdot \prod_{\psi=1}^n \prod_{i=1}^m \exp \left\{ \left( -\frac{(y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2}{2\sigma_{\psi,i}^2} \right) \right\} \right) = \quad (\text{A.7})$$

$$-\log \left( \prod_{\psi=1}^n \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_{\psi,i}^2}} \right) - \log \left( \prod_{\psi=1}^n \prod_{i=1}^m \exp \left\{ \left( -\frac{(y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2}{2\sigma_{\psi,i}^2} \right) \right\} \right) = \quad (\text{A.8})$$

Det kan då ses att

$$-\log \left( \prod_{\psi=1}^n \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_{\psi,i}^2}} \right) \quad (\text{A.9})$$

inte beror på  $\theta$  och därför inte kommer att vara intressant för minimeringen. Därför kan minimeringen av  $S(\theta)$  skrivas om som

$$\min_{\theta} S(\theta) = \min_{\theta} \left( -\log \left( \prod_{\psi=1}^n \prod_{i=1}^m \exp \left\{ \left( -\frac{(y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2}{2\sigma_{\psi,i}^2} \right) \right\} \right) \right) = \quad (\text{A.10})$$

$$\min_{\theta} \left( \sum_{\psi=1}^n \sum_{i=1}^m -\log \left( \exp \left\{ \left( -\frac{(y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2}{2\sigma_{\psi,i}^2} \right) \right\} \right) \right) = \quad (\text{A.11})$$

$$\min_{\theta} \left( \sum_{\psi=1}^n \sum_{i=1}^m \frac{(y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2}{2\sigma_{\psi,i}^2} \right). \quad (\text{A.12})$$

I den här rapporten kommer  $\sigma_{\psi,i} = 1 \quad \forall \psi, i$  att antas, vilket ger den slutliga förenklingen

$$S(\theta) = \sum_{\psi=1}^n \sum_{i=1}^m (y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta))^2 = \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}^2 \quad (\text{A.13})$$

Där  $r_{\psi,i}$  är residualen.  $S(\theta)$  kommer i denna rapport att refereras till som kostnadsfunktionen.

## B Härledning av Newton-Raphsons metod

Newton-Raphsons metod bygger på en andra ordningens Taylorutveckling.

$$S(\theta + p) = S(\theta) + \nabla S(\theta)^T \cdot p + \frac{1}{2} \cdot p^T \cdot \nabla^2 S(\theta) \cdot p \quad (\text{B.1})$$

Största minskningen från ett steg  $p$  fås av

$$\min_p S(\theta + p) \Rightarrow \frac{d}{dp} (S(\theta + p)) = 0 \quad (\text{B.2})$$

detta ger

$$\frac{d}{dp} \left( S(\theta) + \nabla S(\theta)^T \cdot p + \frac{1}{2} \cdot p^T \cdot \nabla^2 S(\theta) \cdot p \right) = 0 \quad (\text{B.3})$$

$$\nabla S(\theta) + \nabla^2 S(\theta) \cdot p = 0 \quad (\text{B.4})$$

$$\nabla^2 S(\theta) \cdot p = -\nabla S(\theta) \quad (\text{B.5})$$

Och slutligen

$$p = -(\nabla^2 S(\theta))^{-1} \cdot \nabla S(\theta) = -(\mathbf{H}(\theta))^{-1} \cdot \nabla S(\theta) \quad (\text{B.6})$$

där  $\mathbf{H}$  är Hessianen och  $p$  är en nedåtgående riktning.

## C Härledning av Gauss-Newtons metod

Gauss-Newtons metod bygger på Newton-Raphsons metod med följande omskrivningar och antaganden [8].

Kostnadsfunktionen beräknas som

$$S(\theta) = \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}(\theta)^2 \quad (\text{C.1})$$

där  $r_{\psi,i}(\theta) = y_{\psi}(t_i) - \hat{y}_{\psi}(t_i, \theta)$ . Gradienten av kostnadsfunktionen fås då, enligt kedjeregeln, av

$$\begin{aligned} \frac{\partial S(\theta)}{\partial \theta_j} &= 2 \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}(\theta) \cdot \frac{\partial r_{\psi,i}(\theta)}{\partial \theta_j} \\ &= -2 \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}(\theta) \cdot \frac{\partial \hat{y}_{\psi}(t_i, \theta)}{\partial \theta_j} \end{aligned} \quad (\text{C.2})$$

vilket kan skrivas som

$$\nabla S(\theta) = -2 \begin{bmatrix} \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}(\theta) \cdot \frac{\partial \hat{y}_{\psi}(t_i, \theta)}{\partial \theta_1} \\ \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}(\theta) \cdot \frac{\partial \hat{y}_{\psi}(t_i, \theta)}{\partial \theta_2} \\ \vdots \\ \sum_{\psi=1}^n \sum_{i=1}^m r_{\psi,i}(\theta) \cdot \frac{\partial \hat{y}_{\psi}(t_i, \theta)}{\partial \theta_k} \end{bmatrix} \quad (\text{C.3})$$

där  $k$  är antalet parametrar i  $\theta$ . Genom att sätta

$$\mathbf{r} = \begin{bmatrix} r_{1,1}(\theta) \\ r_{1,2}(\theta) \\ \vdots \\ r_{1,m}(\theta) \\ r_{2,1}(\theta) \\ r_{2,2}(\theta) \\ \vdots \\ r_{n,m}(\theta) \end{bmatrix} \quad (\text{C.4})$$

och

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \hat{y}_1(t_1, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_1(t_1, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_1(t_1, \theta)}{\partial \theta_k} \\ \frac{\partial \hat{y}_1(t_2, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_1(t_2, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_1(t_2, \theta)}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_1(t_m, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_1(t_m, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_1(t_m, \theta)}{\partial \theta_k} \\ \frac{\partial \hat{y}_2(t_1, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_2(t_1, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_2(t_1, \theta)}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_n(t_m, \theta)}{\partial \theta_1} & \frac{\partial \hat{y}_n(t_m, \theta)}{\partial \theta_2} & \cdots & \frac{\partial \hat{y}_n(t_m, \theta)}{\partial \theta_k} \end{bmatrix} \quad (\text{C.5})$$

kan gradienten skrivas som

$$\nabla S(\boldsymbol{\theta}) = -2\mathbf{J}^T \mathbf{r}. \quad (\text{C.6})$$

Approximationen av Hessianen tas fram på följande sätt

$$\frac{\partial^2 S(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l} = 2 \sum_{\psi=1}^n \sum_{i=1}^m \frac{\partial r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_j} \cdot \frac{\partial r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_k} + r_{\psi,i}(\boldsymbol{\theta}) \cdot \frac{\partial^2 r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l}. \quad (\text{C.7})$$

Gauss-Newtons metod fås genom att anta att

$$\left| r_{\psi,i}(\boldsymbol{\theta}) \cdot \frac{\partial^2 r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right| \ll \left| \frac{\partial r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_j} \cdot \frac{\partial r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_k} \right| \quad (\text{C.8})$$

vilket ger

$$\begin{aligned} \frac{\partial^2 S(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} &= 2 \sum_{\psi=1}^n \sum_{i=1}^m \frac{\partial r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_j} \cdot \frac{\partial r_{\psi,i}(\boldsymbol{\theta})}{\partial \theta_k} \\ &= 2 \sum_{\psi=1}^n \sum_{i=1}^m \frac{\partial \hat{y}_{\psi}(t_i, \boldsymbol{\theta})}{\partial \theta_j} \cdot \frac{\partial \hat{y}_{\psi}(t_i, \boldsymbol{\theta})}{\partial \theta_k}. \end{aligned} \quad (\text{C.9})$$

Med  $\mathbf{J}$  ansatt i ekvation C.5 kan då Hessianen  $\mathbf{H}$  för kostnadsfunktionen skrivas som

$$\mathbf{H} = 2\mathbf{J}^T \mathbf{J}. \quad (\text{C.10})$$

Utgående från Newton-Raphsons metod fås då

$$\mathbf{p} = -\mathbf{H}(\boldsymbol{\theta})^{-1} \cdot \nabla f(\boldsymbol{\theta}) = -(2\mathbf{J}^T \mathbf{J})^{-1} \cdot -2\mathbf{J}^T \mathbf{r} \quad (\text{C.11})$$

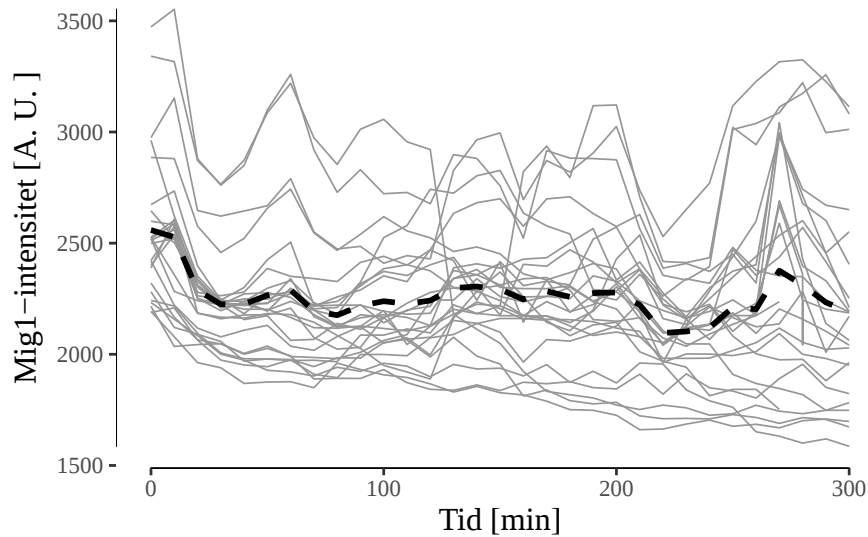
vilket slutligen ger

$$\mathbf{p} = (\mathbf{J}^T \mathbf{J})^{-1} \cdot \mathbf{J}^T \mathbf{r} \quad (\text{C.12})$$

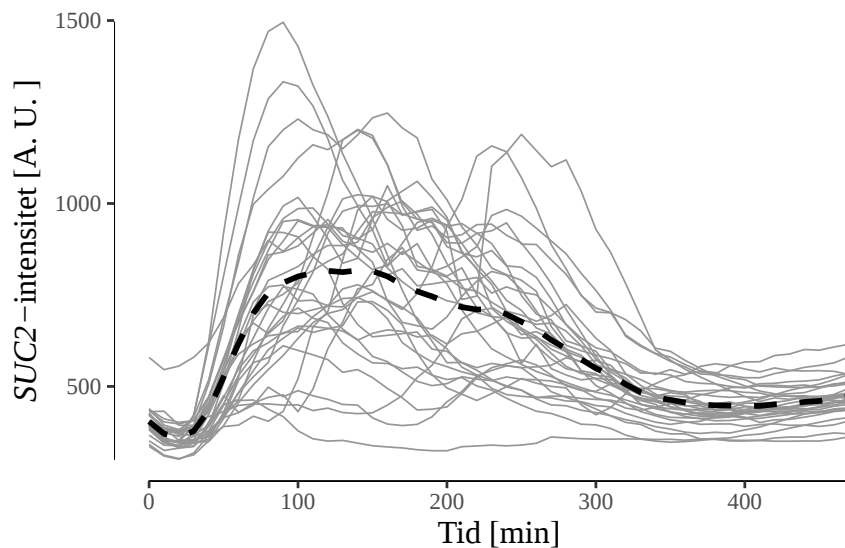
där  $\mathbf{p}$  är en nedåtgående riktning.

## D Data

Datan använd i denna rapport är tagen från 30 *S. cerevisiae* celler som utsattes för svält i form av en minskning av glukoshalten från 4% till 0,2%. Datan som insamlades var i form av flourescensbilder för mängden Mig1 i cellkärnan samt uttrycken av Suc2 samt Hxk1. Datan för de olika cellerna sammanställdes till medelvärden vilket var det som modellerades i denna rapport. Då enbart data för Mig1 och Suc2 användes i rapporten är endast den inkluderad här i Figur 1 respektive 2.



**Figur 1:** Använd data för Mig1 taget som medelvärde från 30 *S. cerevisiae* celler.



**Figur 2:** Använd data för Suc2 taget som medelvärde från 30 *S. cerevisiae* celler.

## E Test av optimeringsalgorithm på testmodell

För att testa om den framtagna optimeringsalgoritmen faktiskt kan hitta de korrekta parametrarna, under ideala förhållanden, och minimera kostnadsfunktionen användes den på en testmodell där de sanna parametrarna är kända. Med testmodellen jämfördes även Gauss-Newtons metod till *steepest descent*-metoden. All kod som användes hittas på GitHub (<https://github.com/axillus/Kandidat2020>). Modellen hade formen

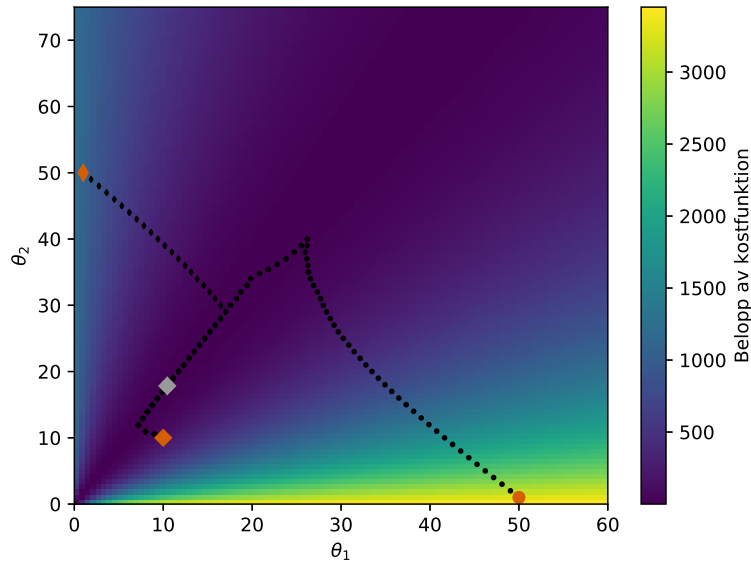
$$\begin{aligned} X &\rightleftharpoons Y \\ X_0 &= 4.043530 \\ Y_0 &= 2.649860 \end{aligned}$$

Vilket kan uttryckas som

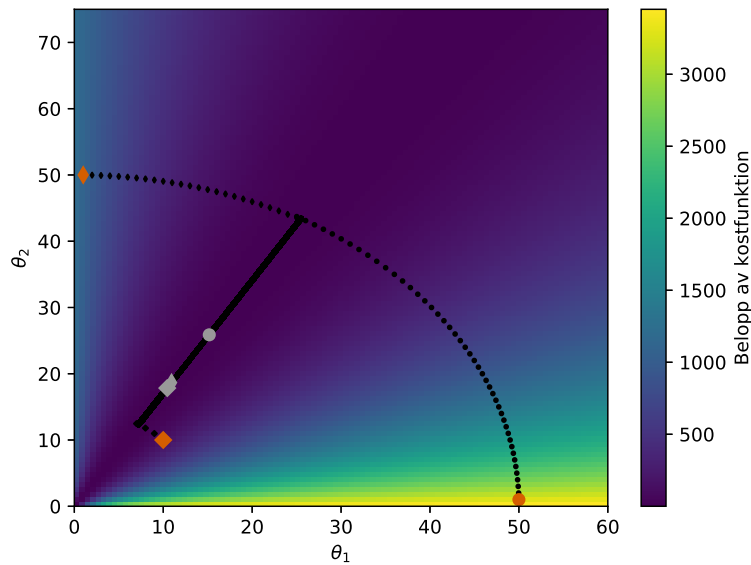
$$\begin{aligned} \frac{dX}{dt} &= Y \cdot \theta_2 - X \cdot \theta_1, & X_0 &= 4.043530 \\ \frac{dY}{dt} &= X \cdot \theta_1 - Y \cdot \theta_2, & Y_0 &= 2.649860 \end{aligned}$$

Tre olika startgissningar testades

$$\begin{cases} \theta_1 = 50 \\ \theta_2 = 1 \end{cases}, \begin{cases} \theta_1 = 1 \\ \theta_2 = 50 \end{cases}, \begin{cases} \theta_1 = 10 \\ \theta_2 = 10 \end{cases} \quad (\text{E.1})$$



**Figur 1:** Optimering med Gauss-Newtons metod utifrån tre olika startgissningar. Orange markörer (◆ ● ◆) markerar respektive startgissning och de gråa markörerna (◆ ● ◆) markerar de optimerade parametrarna, i figuren överlappar de varandra och därmed syns bara en grå markör. Alla tre startgissningar har slutat i de korrekta parametrarna,  $\theta_1 = 10.454$  och  $\theta_2 = 17.837$ . Det tog olika många iterationer att nå det korrekta värdet beroende på startgissningen. För startgissning  $\theta_1 = 50$  och  $\theta_2 = 1$  (◆) tog det 76 iterationer, för  $\theta_1 = 1$  och  $\theta_2 = 50$  (●) tog det 47 iterationer och för  $\theta_1 = 10$  och  $\theta_2 = 10$  (◆) tog det 24 iterationer. Det kan ses att metoden tar steg av liknande steglängd genom hela optimeringen.



**Figur 2:** Optimering med *steepest descent*-metoden utifrån tre olika startgissningar. Orange markörer (♦ ● ♦) markerar respektive startgissning och de gråa markörerna (♦ ● ♦) markerar de optimerade parametrarna. Bara en av startgissningarna,  $\theta_1 = 10$  och  $\theta_2 = 10$ , har slutat i de korrekta parametrarna,  $\theta_1 = 10.454$  och  $\theta_2 = 17.837$ , detta tog 58566 iterationer. Ytterligare en startgissning,  $\theta_1 = 1$  och  $\theta_2 = 50$ , kom nära med de optimerade parametrarna  $\theta_1 = 10.936$  och  $\theta_2 = 18.648$ , detta tog 78422 iterationer. Den sista startgissningen,  $\theta_1 = 50$  och  $\theta_2 = 1$ , slutade i de optimerade parametrarna  $\theta_1 = 15.178$  och  $\theta_2 = 25.875$ , detta tog 27476 iterationer. Som kan ses tog denna metod väldigt korta och många steg nere i "dalen" som går diagonalt genom figuren.

Som kan ses når Gauss-Newtons metod det sanna värdet snabbare och säkrare. Dock är detta en väldigt enkel modell med bara två parametrar, för att kunna illustrera stegningen enklare. Datan i modellen har inte heller något brus, vilket inte är fallet med riktig experimentell data. Modellen verkar även bara ha en extrempunkt vilket gör att en lokal optimering funkar bra. Inget av ovanstående gäller för datan och modellen som används i denna rapport vilket gör att optimeringen blir svårare och osäkrare. På grund av brus i datan samt många lokala minimum är det även svårt att veta om det globala minimumet är nått, detta görs endast genom att ta den lägsta viktade kostnadsfunktionen från de olika optimeringarna.

## F Utförd stabilitetsanalys på samtliga modeller

Stabilitetsanalys utfördes på samtliga modeller. Nedan följer en fullständig härledning av stabilitetsanalys för Modell 1 samt beräkningar av de viktigaste resultaten för de reducerade modellerna.

### F.1 Modell 1

Stegvis stabilitetsanalys genomförd på Modell 1, se ekvation 23. Inledningsvis löses modellen för att hitta jämviktsläge, det vill säga

$$0 = \frac{dMig1}{dt} = k_1 \cdot Mig1p - k_2 \cdot Mig1 + k_6 \cdot X - k_7 \cdot Mig1 \quad (F.1a)$$

$$0 = \frac{dMig1p}{dt} = -k_1 \cdot Mig1p + k_2 \cdot Mig1 \quad (F.1b)$$

$$0 = \frac{dSuc2}{dt} = \frac{k_3}{Mig1 + 0,1} - k_4 \cdot Suc2 \quad (F.1c)$$

$$0 = \frac{dX}{dt} = \frac{k_5}{Mig1 + 0,1} - k_6 \cdot X. \quad (F.1d)$$

Ekvationerna ger följande uttryck för varje koncentration vid jämvikt.

$$F.1b \Rightarrow Mig1p = \frac{k_2}{k_1} \cdot Mig1 \quad (F.2a)$$

$$F.1c \Rightarrow Suc2 = \frac{k_3}{k_4} \cdot \frac{1}{Mig1 + 0,1} \quad (F.2b)$$

$$F.1d \Rightarrow X = \frac{k_5}{k_6} \cdot \frac{1}{Mig1 + 0,1} \quad (F.2c)$$

Genom att ersätta Mig1p, Suc2 och X i F.1a med deras beroende på Mig1 kan koncentrationen vid jämvikt för Mig1 beräknas.

$$\begin{aligned} 0 &= k_2 \cdot Mig1 - k_2 \cdot Mig1 + \frac{k_5}{Mig1 + 0,1} - k_6 \cdot Mig1 \\ &\Rightarrow k_5 - k_6 \cdot Mig1(Mig1 + 0,1) = 0 \\ &\Rightarrow Mig1 = -\frac{0,1}{2} \pm \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_7}} \end{aligned} \quad (F.3)$$

Eftersom koncentrationer ej kan vara negativa kommer lösningen där rottermen subtraheras inte vara möjlig. Resultatet för Mig1 kan sedan sättas in i ekvation F.2 och ger följande uttryck för respektive koncentration.

$$Mig1 = \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_7}} - \frac{0,1}{2} \quad (F.4a)$$

$$Mig1p = \frac{k_2}{k_1} \cdot \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_7}} - \frac{0,1}{2} \quad (F.4b)$$

$$Suc2 = \frac{k_3}{k_4} \cdot \frac{1}{\sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_7}} + \frac{0,1}{2}} \quad (F.4c)$$

$$X = \frac{k_5}{k_6} \cdot \frac{1}{\sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_7}} + \frac{0,1}{2}} \quad (F.4d)$$

Med de skattade parametrarna, Tabell 2, fås jämviktskoncentrationerna som presenteras i Tabell 3.

Nästa steg är att ta fram Jacobianen där varje element kommer från ekvation F.1 som deriveras med avseende på varje koncentration enligt

$$J = \begin{bmatrix} -k_2 - k_7 & k_1 & 0 & k_6 \\ k_2 & -k_1 & 0 & 0 \\ \frac{k_3}{a} & 0 & -k_4 & 0 \\ \frac{-k_5}{a} & 0 & 0 & -k_6 \end{bmatrix} \quad (F.5)$$

där  $a = (Mig1 + 0, 1)^2$  vilket alltid är  $> 0$ . Egenvärdena,  $\lambda$ , fås nu genom att lösa den karakteristiska ekvationen

$$0 = \det(J - \lambda I) = \begin{vmatrix} -k_2 - k_7 - \lambda & k_1 & 0 & k_6 \\ k_2 & -k_1 - \lambda & 0 & 0 \\ \frac{k_3}{a} & 0 & -k_4 - \lambda & 0 \\ \frac{-k_5}{a} & 0 & 0 & -k_6 - \lambda \end{vmatrix}. \quad (F.6)$$

Eftersom den tredje kolonnen innehåller flest nollor beräknas determinanten utifrån den, vilket ger

$$\det(J - \lambda I) = (-k_4 - \lambda) \underbrace{\begin{vmatrix} -k_2 - k_7 - \lambda & k_1 & k_6 \\ k_2 & -k_1 - \lambda & 0 \\ \frac{-k_5}{a} & 0 & -k_6 - \lambda \end{vmatrix}}_B \quad (F.7)$$

$$B = (-k_2 - k_7 - \lambda) \underbrace{\begin{vmatrix} -k_1 - \lambda & 0 \\ 0 & -k_6 - \lambda \end{vmatrix}}_C - k_1 \underbrace{\begin{vmatrix} k_2 & 0 \\ \frac{k_5}{a} & -k_6 - \lambda \end{vmatrix}}_D + k_6 \underbrace{\begin{vmatrix} k_2 & -k_1 - \lambda \\ \frac{k_5}{a} & 0 \end{vmatrix}}_E \quad (F.8)$$

$$C = k_1 k_6 + k_1 \lambda + k_6 \lambda + \lambda^2, D = -k_2 k_6 - k_2 \lambda, E = -\frac{k_1 k_5}{a} - \frac{k_5 \lambda}{a} \quad (F.9)$$

Sammantaget ger ekvationerna F.7, F.8 och F.9

$$\begin{aligned} \det(J - \lambda I) &= \lambda^4 + \\ &+ \underbrace{(k_1 + k_2 + k_4 + k_6 + k_7)}_{a_1} \lambda^3 + \\ &+ \underbrace{(k_2 k_4 + k_4 k_7 + k_1 k_4 + k_4 k_6 + k_2 k_6 + k_1 k_7 + k_6 k_7 + k_1 k_6 + \frac{k_5 k_6}{a})}_{a_2} \lambda^2 + \\ &+ \underbrace{(k_2 k_4 k_6 + k_1 k_4 k_7 + k_4 k_6 k_7 + k_1 k_4 k_6 + k_1 k_6 k_7 + \frac{k_4 k_5 k_6}{a} + \frac{k_1 k_5 k_6}{a})}_{a_3} \lambda + \\ &+ \underbrace{k_1 k_4 k_6 k_7 + \frac{k_1 k_4 k_5 k_6}{a}}_{a_4} \end{aligned} \quad (F.10)$$

Eftersom alla parametrar per definition är positiva kommer det aldrig vara någon teckenväxling mellan termerna i polynomet. Enligt Decartes teckenregel innebär det att lösningen innehåller 0-2 reellt negativa rötter och 2-4 komplexa rötter. För att titta närmre på tecknet hos realdelen till de komplexa rötterna används Routh-Hurwitz stabilitetskriterie. För Modell 1 blir det:

1.  $a_i > 0$
2.  $\Delta_1 = a_0 > 0$
3.  $\Delta_2 = a_1 a_2 - a_3 > 0$
4.  $\Delta_3 = a_1 a_2 a_3 - a_4 a_1^2 - a_3^2 > 0$
5.  $\Delta_4 = a_4 \Delta_3 > 0 \Rightarrow a_4 > 0$

Där  $\Delta_3$  innehåller många termer och är därför inte är trevlig att beräkna. Därför används Liénard-Chiparts kriteriets andra villkor,  $a_n > 0, a_{n-2} > 0, \dots; \Delta_2 > 0, \Delta_4 > 0, \dots$ , som inte kräver en beräkning av  $\Delta_3$ . Det är ekvivalent med att uppfylla villkor 1, 3 och 5 ovan. Eftersom alla parametrar  $k_i$  är definierade som positiva uppfylls kriterie 1 och 5 om  $k_1, k_4, k_6$  och  $k_7 \neq 0$ . För kriterie 3 kommer  $a_1 a_2$  innehålla alla termer som  $a_3$  består av, plus ytterligare 38 termer och kommer därför vara större än 0. Sammantaget betyder det att Modell 1 är stabil om  $k_1, k_4, k_6$  och  $k_7 \neq 0$ . Om modellen går mot ett styvt jämviktsläge eller svänger avgörs av värdet på parametrarna. I figur 6 som presenteras i resultatet, visas exempel på några intervall som ger styva och oscillerande modeller.

Med de skattade parametrarna, Tabell 2, fås två komplexa rötter och två reella rötter, se Tabell 12. Vilket innebär att dynamiken för Modell 1 går mot en styv jämvikt vid de optimalt skattade parametrarna. Metoden som användes för att lösa ut egenvärdena från ekvation F.10 var analytiskt men skrevs in i Python3 för att underlätta beräkningarna. Koden finns på GitHub (<https://github.com/axillus/Kandidat2020>).

**Tabell 12:** Beräknade egenvärden med parameterskattningen för Modell 1

|             |              |
|-------------|--------------|
| $\lambda_1$ | -12.74-7.75j |
| $\lambda_2$ | -12.74+7.75j |
| $\lambda_3$ | -21.25       |
| $\lambda_4$ | -4.24        |

## F.2 Modell $k_6 = k_7$

I stabilitetsanalysen för Modell  $k_6 = k_7$  ersätts  $k_7$  med  $k_6$  vilket ger jämviktsuttrycken i ekvation F.11.

$$Mig1 = \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_6}} - \frac{0,1}{2} \quad (F.11a)$$

$$Mig1p = \frac{k_2}{k_1} \cdot \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_6}} - \frac{0,1}{2} \quad (F.11b)$$

$$Suc2 = \frac{k_3}{k_4} \cdot \frac{1}{\sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_6}} + \frac{0,1}{2}} \quad (F.11c)$$

$$X = \frac{k_5}{k_6} \cdot \frac{1}{\sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_5}{k_6}} + \frac{0,1}{2}} \quad (F.11d)$$

Då parameterskattningen presenterad i Tabell 5, resulterade i att  $\hat{\theta}_1 = 0$ , är systemet inte stabilt eftersom det leder till division med 0 i ekvation F.11b.

### F.3 Modell $k_3 = k_5$

I stabilitetsanalysen för Modell  $k_3 = k_5$  ersätts  $k_5$  med  $k_3$  vilket ger jämviktsuttrycken i ekvation F.12.

$$Mig1 = \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_3}{k_7} - \frac{0,1}{2}} \quad (F.12a)$$

$$Mig1p = \frac{k_2}{k_1} \cdot \sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_3}{k_7} - \frac{0,1}{2}} \quad (F.12b)$$

$$Suc2 = \frac{k_3}{k_4} \cdot \frac{1}{\sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_3}{k_7} + \frac{0,1}{2}}} \quad (F.12c)$$

$$X = \frac{k_3}{k_6} \cdot \frac{1}{\sqrt{\left(\frac{0,1}{2}\right)^2 + \frac{k_3}{k_7} + \frac{0,1}{2}}} \quad (F.12d)$$

Med de skattade parametrarna, Tabell 8, fås jämviktskoncentrationerna som presenteras i Tabell 9.

Vidare studeras stabiliteten genom att ersätta  $k_5$  med  $k_3$  i ekvation F.10. Det resulterar inte i några större förändringar utan samtliga kriterier för stabilitet uppnås med kriteriet, precis som för ovanstående modeller, att  $k_1, k_4, k_6$  och  $k_7 \neq 0$ . Vid beräkning av egenvärden visar sig resultatet alltid bli fyra komplexa tal och därmed kommer dynamiken i Modell  $k_3 = k_5$  alltid vara oscillerande. De beräknade egenvärdena för den optimala paramterskattningen presenteras i Tabell 13.

**Tabell 13:** Beräknade egenvärden med parameterskattningen för Modell  $k_3 = k_5$

|             |                  |
|-------------|------------------|
| $\lambda_1$ | -14.75+9.51j     |
| $\lambda_2$ | -14.75-9.51j     |
| $\lambda_3$ | -24.60-2.07e-19j |
| $\lambda_4$ | -4.90+1.09e-18j  |

## G Utförd strukturell identifierbarhetsanalys på samtliga modeller

I detta projekt används Wolfram Mathematica-paketet IdentifiabilityAnalysis [30, 31]. I vidare ekvationer sätts  $x_1 = \text{Mig1}$ ,  $x_2 = \text{Mig1p}$ ,  $x_3 = \text{Suc2}$ ,  $x_4 = X$ . Satta som utdata är Mig1 och Suc2.

### G.1 Modell 1

Resultatet från identifierbarhetsanalysen av den fullständiga modellen. För den här modellen blev  $k_7$  inte identifierbar vilket gör modellen oidentifierbar.

```
In[1]:= Needs["IdentifiabilityAnalysis`"]

deq = {x1'[t] == k1 x2[t] - k2 x1[t] + k6 x4[t] - k7 x1[t], x1[0] == 1,
       x2'[t] == -k1 x2[t] + k2 x1[t], x2[0] == 0,
       x3'[t] == k3 / (x1[t] + 0.01) - k4 x3[t], x3[0] == 4.09,
       x4'[t] == k5 / (x1[t] + 0.01) - k6 x4[t], x4[0] == 0
      };

In[7]:= iad = IdentifiabilityAnalysis[{deq, {x1[t], x3[t]}},
                                     {x1, x2, x3, x4}, {k1, k2, k3, k4, k5, k6, k7}, t]

Out[7]:= IdentifiabilityAnalysisData[False, <>]

In[6]:= iad["NonIdentifiableParameters"]

Out[6]:= {k7}
```

### G.2 Modell $k_6 = k_7$

Resultatet från strukturell identifierbarhetsanalys av den reducerade modellen  $k_6 = k_7$ . Slutsatsen drogs att modellen var strukturellt identifierbar.

```
In[1]:= Needs["IdentifiabilityAnalysis`"]

In[2]:= deq = {x1'[t] == k1 x2[t] - k2 x1[t] + k6 x4[t] - k6 x1[t], x1[0] == 1,
              x2'[t] == -k1 x2[t] + k2 x1[t], x2[0] == 0,
              x3'[t] == k3 / (x1[t] + 0.01) - k4 x3[t], x3[0] == 4.09,
              x4'[t] == k5 / (x1[t] + 0.01) - k6 x4[t], x4[0] == 0
             };

In[3]:= iad = IdentifiabilityAnalysis[{deq, {x1[t], x3[t]}},
                                     {x1, x2, x3, x4}, {k1, k2, k3, k4, k5, k6}, t]

Out[3]:= IdentifiabilityAnalysisData[True, <>]
```

### G.3 Modell $k_3 = k_5$

Vid strukturell identifierbarhetsanalys av Modell  $k_3 = k_5$  blev resultatet att modellen var strukturellt identifierbar

```
In[1]:= Needs["IdentifiabilityAnalysis`"]

In[2]:= deq = {x1'[t] == k1 x2[t] - k2 x1[t] + k6 x4[t] - k6 x1[t], x1[0] == 1,
              x2'[t] == -k1 x2[t] + k2 x1[t], x2[0] == 0,
              x3'[t] == k3 / (x1[t] + 0.01) - k4 x3[t], x3[0] == 4.09,
              x4'[t] == k3 / (x1[t] + 0.01) - k5 x4[t], x4[0] == 0
            };

In[3]:= iad = IdentifiabilityAnalysis[{deq, {x1[t], x3[t]}},
                                     {x1, x2, x3, x4}, {k1, k2, k3, k4, k5, k6}, t]

Out[3]:= IdentifiabilityAnalysisData[True, <>]
```