

Regelbaserad modellering som verktyg för biologisk forskning

Maria Ström-Bestor

Kandidatavhandling i datavetenskap

Handledare: Ion Petre, Sepinoud Azimi

Naturvetenskaper och teknik

Åbo Akademi

2016

Referat

Modellering har blivit ett allt viktigare verktyg för biologisk forskning under de senaste decennierna. Komplexa, dynamiska biologiska processer modelleras och simuleras för att samla ihop och visualisera information samt förutspå händelser under olika förhållanden. Matematiska modeller som grundar sig på differentialekvationer har föreslagits och använts i flera årtionden, men på senare år har regelbaserade modeller väckt intresse. Istället för att konstruera ekvationer som beskriver molekylers beteende, handlar regelbaserad modellering om att specificera strukturerade objekt med relevanta attribut och formulera regler för deras interaktioner. Objekten kan representera t.ex. biomolekyler eller hela celler. Denna avhandling illustrerar regelbaserad modellering som ett verktyg för biologisk forskning med en enkel regelbaserad modell för cellcykeln hos jästsvampen *Schizosaccharomyces pombe*. En cells cellcykel är en strikt reglerad serie av processer som en cell genomgår för att dela sig i två kopior. *S. pombe* är en välstuderad encellig organism och används som modellorganism för eukaryoter, d.v.s. organismer vars celler förvarar sitt genetiska material i en cellkärna.

Innehållsförteckning

1	Inledning	1
2	Modellering av biologiska system	3
2.1	Varför modellera biologiska system?	3
2.2	Faktorer att beakta vid modellering inom biologi	3
2.3	Regelbaserad modellering	5
3	Cellcykeln	6
3.1	Cellcykeln består av en serie distinkta faser	6
3.2	Cellcykelns reglering	7
3.3	Celldelning och parningstyper hos <i>Schizosaccharomyces pombe</i>	8
4	Cellcykeln hos <i>S. pombe</i> modellerad med BioNetGen	10
4.1	Önskade egenskaper hos modellen	10
4.2	En matematisk modell som utgångspunkt	10
4.3	BioNetGen	12
4.4	En regelbaserad modell med BioNetGen	12
4.4.1	Konstruktionen av den regelbaserade modellen	12
4.4.2	Parametrar (parameters)	13
4.4.3	Objekt och attribut (molecule types)	13
4.4.4	Starttillstånd (seed species)	14
4.4.5	Relevant utdata (observables)	14
4.4.6	Regler (reaction rules)	15
4.4.7	Metoder (actions)	16
4.5	Simuleringar och resultat	16
5	Diskussion	17
6	Referenser	18

Bilagor: Bilaga 1: Cellcykelmodellen med BioNetGen

1 Inledning

En organism, vare sig den består av en cell eller flera, är beroende av komplicerade nätverk av signaler. Ett nätverk kan vara begränsad till cellens inre, där en signal sprids som en kaskad av reaktioner mellan molekyler. Det kan vara fråga om samverkan mellan celler i en flercellig organism eller mellan en organism och dess miljö. Oberoende av typen av samspel, är det fråga om väldigt komplicerade system.

När man betraktar ett signaleringssystem inom en cell är det ofta fråga om proteiner som reagerar med varandra [1]. Ett dylikt protein kan tänkas binda till andra proteiner och bilda komplex bestående av två eller flera molekyler. Dessa kan i sin tur reagera med andra molekyler i form av enskilda proteiner eller proteinkomplex. För att simulera en sådan process, bör man även beakta hastigheterna av dessa reaktioner samt bildnings- och nedbrytningshastigheter hos varje komponent i processen.

Denna komplexitet betyder att redan korta signalkaskader med ett fåtal molekyler kräver avancerade matematiska beräkningar för simulering [2]. Biologiska system kan beskrivas med t.ex. ordinära differentialekvationer, d.v.s. ekvationer som innefattar både en funktion och en eller flera av dess derivator, men dessa är mödosamma eller rentav omöjliga att återanvända mellan olika system, samt svåra att förstå utan en djupgående matematisk bakgrund.

Ett nyare alternativ till ren matematik är s.k. regelbaserad modellering [3] [1]. Istället för ekvationer beskrivs samspelet mellan molekyler med enkla regler, ofta utgående från den syntax som används för att beskriva kemiska reaktioner. En reaktion kan t.ex. beskriva hur två molekyler binder till varandra under specifika förhållanden, eller hastigheterna för aktivering och deaktivering en molekyls reaktivitet.

Flera ramverk och språk har utvecklats för att fungera som mer lättförståeliga och anpassningsbara gränssnitt till den bakomliggande matematiken. Ett av dessa är verktyget BioNetGen som använder språket BioNetGen Language (BNGL) [4]. Detta arbete beskriver regelbaserad modellering med BioNetGen och jämför denna typ av modell med direkt användning av differentialekvationer genom att använda cellcykeln som exempel. Cellcykeln är en livsavgörande process för alla organismer

som har studerats i detalj under flera decennier. Detta arbete exemplifierar processen med jästsvampen *Schizosaccharomyces pombe*, som ofta används som modellorganism inom biologin [5].

2 Modellering av biologiska system

2.1 Varför modellera biologiska system?

Vare sig en biolog studerar liv på molekylär eller ekologisk nivå, är det fråga om biologiska system med hög komplexitet och empirisk forskning som kostar både tid och pengar. I takt med framsteg inom molekylärbiologi och datorteknik har matematiska modeller därmed blivit allt viktigare verktyg, vilket har lett till flera tvärvetenskapliga discipliner som kombinerar dessa tre områden [6]. Ett viktigt exempel är systembiologi, som studerar samverkan inom biologiska system utgående från tanken att helheten är mer än summan av dess enskilda delar [7]. Denna del av biologin försöker utreda hur beteendet hos ett system kan uppstå från det fysiska samspelet och vilka interaktioner är nödvändiga för processen.

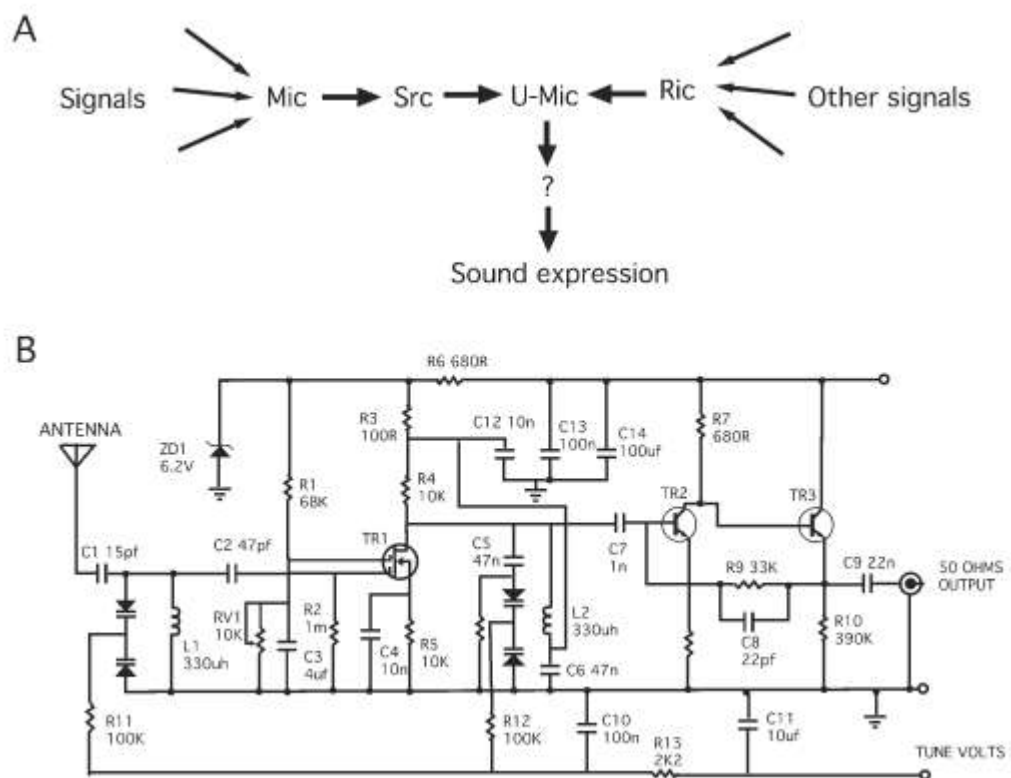
Det är viktigt att notera att biologiska modeller inte kan ersätta empirisk forskning. De är verktyg som används för att förutspå händelser utgående från den information som finns tillgänglig och för att undersöka konsekvenser av antaganden, såsom hur en process påverkas av olika förhållanden [3]. Dessa förutsägelser kan användas som grund för att utforma hypoteser som kan testas med empiriska experiment. Fördelen med preliminär modellering är att klart oanvändbara hypoteser kan förkastas i ett tidigt skede. En modell behöver alltså inte vara perfekt, utan den kan och bör anpassas till frågeställningen. Modeller kan dock också användas för att samla ihop och visualisera känd information [3].

2.2 Faktorer att beakta vid modellering inom biologi

En av de största utmaningarna vid modellering är den kombinatoriska komplexiteten hos biomolekyler med ett flertal aktiva ställen och som deltar i flera olika reaktioner. Cellsignaler är ett flitigt studerat exempel [1] [8], som illustrerar detta problem. En cells processer styrs av signaler som förmedlas av komplexa nätverk av reaktioner mellan biomolekyler. En biomolekyl – exempelvis ett protein – kan ha flera s.k. aktiva ställen som kan reagera med andra biomolekyler, antingen självständigt eller påverkade av signaler från andra aktiva ställen. Ofta är det fråga om tillfälliga bindningar till andra biomolekyler, vilket påverkar deras aktivitet, men

proteiner kan även genomgå långvariga förändringar under sin livstid. Systemen är alltså väldigt dynamiska.

Språket som används för att beskriva händelser, både visuellt och i form av text, har också stor betydelse. I insändaren "Can a biologist fix a radio?" från år 2002 [9] gjorde molekylärbiologen Yuri Lazebnik en humoristisk jämförelse av tillvägagångssätten hos biologer och ingenjörer för att beskriva processer. Han använde ett hypotetiskt exempel på hur en biolog skulle namnge och beskriva delar av en radio jämfört med ett kretsdiagram. Avsikten med texten var att poängtera behovet av ett enhetligt, formellt språk för att beskriva biologiska processer i stil med kretsdiagram, vilket Figur 1 illustrerar.



Figur 1: Biologers och ingenjörers sätt att beskriva processer enligt Lazebnik. A) En biologs beskrivning av en radio. B) En ingenjörers beskrivning av en radio.
Bild: Figure 3 i Lazebnik 2002 [9].

Figur 1 belyser språkets roll vid förmedling av information. Bild B innehåller en massa användbar information, medan det skulle vara omöjligt att skapa en vettig modell utgående från bild A. När man betraktar ren matematisk modellering ur denna synvinkel, märks liknande brister. Hur mycket kan man säga om ett systems aktivitet

genom att bara se på en samling differentialekvationer? Ett mer beskrivande språk för att uttrycka processer skulle underlätta förståelsen av processen i fråga och därmed underlätta både samarbete och spridning av modellering som verktyg bland forskare. Det är här man finner styrkan av regelbaserade språk.

2.3 Regelbaserad modellering

Under de senaste decennierna har regelbaserad modellering av biologiska system blivit ett mer och mer populärt alternativ. Som namnet anger, handlar det om att definiera fysiska interaktioner som enkla regler, uttryckta m.h.a. specialiserade verktyg som har utvecklats för ändamålet [3].

Det centrala hos regelbaserad modellering är objekt och deras samverkan. Inom biologin är detta ett användbart förhållningssätt som kan tillämpas för att simulera processer vid olika nivåer, från molekylärt samspel som signalering inuti och mellan celler [8] [10] [2] [1] till beteende på organismnivå och ekologiska samband mellan olika arter [11].

Ett objekt i en biologisk regelbaserad modell kan representera en biomolekyl, en cell eller en organism. Samverkan mellan olika objekt anges som regler för att förändra attribut hos dessa objekt [4]. Komplexa reaktionssystem reduceras till enkla, lättförstådda regler. I många verktyg följer reglerna ofta den notation som används för kemiska reaktioner [2]. En reaktion som beskriver två molekyler som binder till varandra och bildar ett komplex kan exempelvis beskrivas så här, där k_1 och k_2 beskriver hastigheten (eng. *kinetic rate*) för att bilda respektive bryta ner dimeren:



När simuleringen körs, översätter verktyget reglerna till de ekvationer som behövs för att generera nätverket av reaktioner.

En klar fördel med en dylik notation är att den är bekant för biologer, vilket sänker tröskeln för att använda modellering som verktyg. Detta sätt att representera händelserna ger också en tydlig bild av vad som sker: två molekyler reagerar med varandra och producerar en tredje typ av molekyl. Därmed fungerar en serie händelser beskrivna på detta sätt även som en dokumentation av processen.

3 Cellcykeln

3.1 Cellcykeln består av en serie distinkta faser

Cellcykeln består av en serie distinkta faser som en cell genomgår för att dela sig själv i två kopior [12]. Hos encelliga celler är det alltså fråga om reproduktion. Hos flercelliga organismer handlar det dessutom dels om att bilda en komplett organism vid utvecklingsstadiet, dels om att ersätta döda celler i den fullvuxna organismen. Det senare fallet är dock beroende av celltyp; t.ex. nervceller delar sig inte, medan hudceller ständigt förnyas. Fel i processen kan ha allvarliga följder för organismen, inklusive celldöd och hejdlös celldelning, vilket är en vanlig egenskap hos cancerceller.

Cykels detaljer varierar mellan olika organismer och mellan olika skeden av en cells livstid, men vissa grundläggande mekanismer är gemensamma för alla organismer [12]. För att kunna dela sig i två kopior, måste en cell kopiera sitt genetiska material (DNA, deoxiribonukleinsyra) och distribuera det i två identiska delar. De flesta celler måste även duplicera andra nödvändiga organeller och makromolekyler, d.v.s. öka sin massa för varje omgång av cellcykeln.

En eukaryot cell kännetecknas av en cellkärna som innehåller cellens DNA [5]. Hos eukaryota celler består cellcykeln i allmänhet av fyra faser (Figur 2) [12]. Under S-fasen dupliceras det genetiska materialet – därav bokstaven S, som står för *syntes* av DNA. Under M-fasen delas själva cellen itu. Bokstaven M hänvisar till *mitos*, delning av cellkärnan, vilket sker i denna fas. Mellan dessa två viktiga faser finns två G-faser, G1 och G2, vilka ger cellen mer tid för tillväxt, vilket kan behövas för fördubbling av annat cellulärt innehåll än DNA.



Figur 2: Cellcykelns fyra faser hos en eukaryot cell. De två huvudsakliga faserna, S (DNA syntes) och M (delning), separeras av två G-faser, som ger tid för tillväxt vid behov.
Bild: Alun Bestor (2016)

3.2 Cellcykelns reglering

För en lyckad genomgång av cellcykeln, måste de olika faserna ske i rätt tid och ordning. Cellcykeln är strikt reglerad och påverkas av både yttre och inre faktorer [12]. Fel i regleringen kan leda till att DNA förloras, modifieras eller fördelas ojämnt mellan dottercellerna, vilket ofta är fallet hos cancerceller.

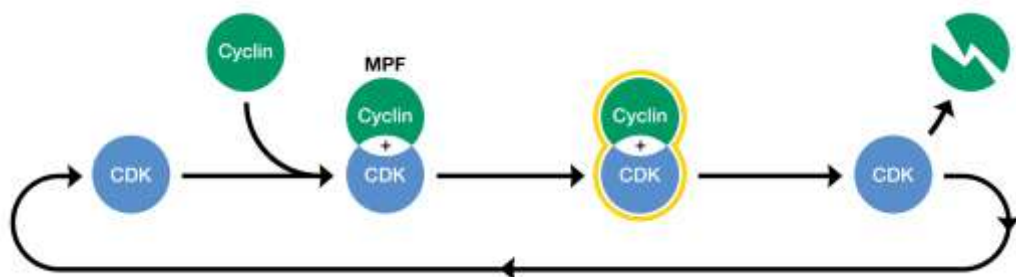
Vid vissa skeden i cykeln finns kontrollpunkter där progressionen kan stoppas tillfälligt [12]. Detta kan ske ifall cellen inte är redo för nästa steg, ifall den yttre miljön är ogynnsam eller om cellen är skadad. Om pausen beror på cellskador, har cellen i allmänhet en chans att reparera skadorna eller gå i apoptos, om reparation inte är möjlig. Apoptos kallas även programmerad celldöd, ty celler som går i apoptos genomgår en strikt reglerad process som leder till cellens död.

De flesta eukaryotiska celler har tre kontrollpunkter [12]. Startpunkten (även kallad *restriction point*) finns i slutet av G1-fasen. När cellen passerar denna punkt, går den in i S-fasen och startar en ny genomgång av cykeln. De två andra återfinns vid övergången till M-fasen och i ett senare skede av samma fas, strax innan den fysiska separationen startar.

Regleringen av cellcykeln är beroende av en familj av likartade proteiner som kallas cyklin-beroende kinaser (Cdk, eng. *cyclin-dependent kinase*) [12]. Såsom namnet

anger, regleras dessa proteiner av en annan proteinfamilj som kallas cykliner. Varje kontrollpunkt regleras av en specifik kombination av ett Cdk-protein och ett cyclin.

För att kunna aktiveras, måste Cdk-proteinerna måste binda till cyclin och bilda ett komplex som kallas MPF (eng. *maturation promoting factor*) (Figur 3). MPF-komplexet kan existera i aktiv och inaktiv form. För att komma förbi en kontrollpunkt i cellcykeln krävs en tillräcklig koncentration av aktivt MPF i cellen. Andra proteiner är också involverade i processen, men cyclin och Cdk i form av MPF är de huvudsakliga aktörerna [12] [13].



Figur 3: Proteinerna Cdk och cyclin bildar komplexet MPF genom att binda till varandra. När MPF aktiveras i tillräckliga mängder, startas en signalkaskad som för cellcykeln i en ny fas. När MPF inte längre behövs, bryts bindingen. Cyclin bryts ner, medan Cdk återanvänds. Bild: Alun Bestor (2016)

3.3 Celldelning och parningstyper hos *Schizosaccharomyces pombe*

Jästsvampen *S. pombe* är en encellig organism som används som modellorganism inom biologin [5]. Dess cellcykel har en enklare regleringsmekanism än t.ex. däggdjur, ty den har endast ett Cdk-protein vid namn cdc2 [12]. Proteinet cdc2 kan binda till fyra olika cykliner [5]. Varje cyclin har sina egna målproteiner som de i sin tur aktiverar och därmed leder signalkaskader startade av olika cykliner till olika resultat. På detta vis kan MPF-komplexet skilja på och reglera olika faser i cellcykeln.

En cells DNA är ordnat i kromosomer. En cell med en enkel uppsättning kromosomer kallas haploid, medan en diploid cell har en dubbel kromosomuppsättning. *S. pombe* kan existera i både haploid och diploid form. Detta

betyder att jästsvampen kan föröka sig både sexuellt och asexuellt. Normalt fortplantar sig *S. pombe* via celldelning, men den kan även föröka sig sexuellt via parning mellan två motsatta parningstyper, M och P [14]. I sådana fall bildas en diploid avkomma, som genast genomgår meios (reduktionsdelning) i M-fasen istället för mitos för att producera fyra haploida dotterceller. Två av dem är av typen P, medan de två andra är av typen M. En jästcell kan även byta parningstyp i S-fasen vid behov.

4 Cellcykeln hos *S. pombe* modellerad med BioNetGen

4.1 Önskade egenskaper hos modellen

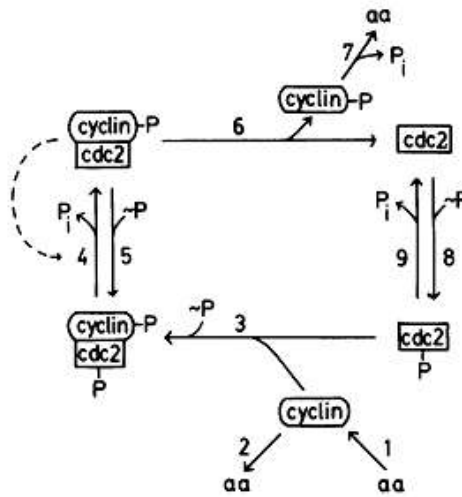
För att utvärdera användningen av verktyget BioNetGen för regelbaserad modellering, utformade jag en modell som innefattar regleringsmekanismen av *S. pombe*s cellcykel, byte av parningstyp och parning av motsatta typer. Som utgångspunkt användes en modell som beskrevs av Tyson år 1991 [13]. Byte av parningstyp beskrivs enligt en modell föreslagen av Klar år 1992 [14].

Modellens utformning beror på målsättningen med arbetet, d.v.s. hurdana frågor man vill besvara. I detta fall var det huvudsakliga målet att översätta de existerande modellerna i regelbaserad form, så jag valde enkla kvantitativa mål: att förutsäga totala antalet celler eller antalet celler av en specifik parningstyp vid ett specifikt stadium av simulationen. Modellen ska alltså simulera celltillväxt och -delning, samt skilja på objekt av olika parningstyp.

4.2 En matematisk modell som utgångspunkt

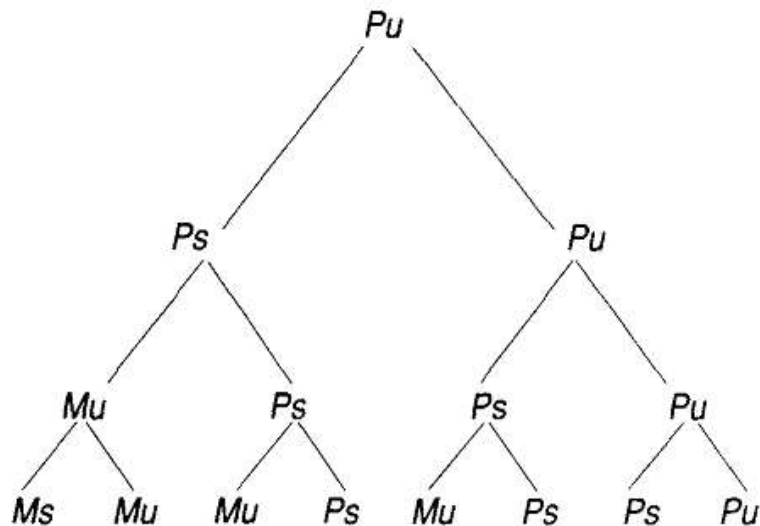
Tyson (1991) föreslog en matematisk modell för att beskriva celldelning hos jästsvampen *S. pombe* [13]. Modellen är starkt förenklad och beaktar endast cyklin, cdc2 och komplexet MPF som de bildar. Dessutom används endast en cyklinmolekyl för att representera hela familjen av cykliner. Detta är möjligt eftersom cyklinerna reagerar med cdc2 på samma sätt och skiljer sig från varandra genom att aktivera olika målproteiner [12], som inte beaktas i denna modell. Ytterligare förenklingar finns i form av ihopslagna steg i processen, t.ex. sker aktivering av cyklin i samma steg som bindning till cdc2.

Figur 4 ger en överblick av regleringsmekanismen. I denna modell hålls mängden cdc2 konstant, medan cyclin bildas före bindning till cdc2 för att bilda MPF och nedbryts efter separering från komplexet. Varje numrerad pil får en motsvarande regel i BioNetGen-modellen. Aktivering av en molekyl beror på fosforylering eller defosforylering (bindning till eller separering från en fosfatgrupp), vilket anges med P respektive P_i .



Figur 4: Samverkan mellan cyclin och cdc2 i cellcykeln enligt Tyson (1991). P och P_i anger fosforylering respektive defosforylering, vilket har betydelse för molekylernas aktivitet. Bild: Fig. 1 i Tyson 1991 [13].

Modellen ska även simulera byte av parningstyp hos *S. pombe*. I detta fall valde jag att modellera byte av parningstyp endast vid haploid celledelning enligt Klar (1992) [14]. Parningsbytet följer ett enkelt mönster som illustreras i Figur 5. En cell som ej kan byta typ ger upphov till två dotterceller av samma typ, varav en har förmågan att byta typ. En cell med en bytbar typ ger upphov till en identisk dottercell och en som är av motsatt typ utan förmåga att byta.



Figur 5: Byte av parningstyp vid celledelning hos *S.pombe*. De två parningstyperna är P och M. En cell kan ha förmågan att byta typ (s) eller saknar förmågan (u). Bild: Fig. 2 i Klar 1992 [14].

4.3 BioNetGen

BioNetGen är ett verktyg för att bygga och simulera regelbaserade modeller av biokemiska system [4]. Den första versionen utgavs år 2004.

En modell skapad i BioNetGen kan delas in i sex huvudsakliga delar [4]. Först definieras de *parametrar* (eng. *parameters*) som beskriver systemets dynamik, såsom konstanter för reaktionshastigheter och värden för objektens utgångskoncentrationer. De olika *typerna av objekt* (eng. *molecule types*) namnges och beskrivs med attribut (eng. *components*). Attributen kan t.ex. vara bindnings- eller aktiveringsställen hos den molekyl objektet representerar. *Starttillståndet* (eng. *seed species*) för systemet definieras med en lista av objekt och deras utgångskoncentrationer. Data som modellen producerar anges som koncentrationer av reaktionernas produkter. Man kan iaktta bl.a. koncentrationen av namngivna objekt eller objekt med specifika attribut genom att specificera *relevant utdata* (eng. *observables*). *Regler* (eng. *reaction rules*) för reaktionerna bör naturligtvis definieras, samt *metoderna* (eng. *actions*) för att generera och simulera modellen.

För att beskriva modeller med detta verktyg används ett BioNetGen-specifikt språk, BNGL. BioNetGen utvecklades för att modellera intracellulära biokemiska processer, vilket innebär att de regler som kan uttryckas med BNGL är begränsade till typiska reaktioner mellan biomolekyler [8] [4]. De fem grundläggande omvandlingarna av ett objekts attribut är bildning och brytning av en bindning, förändring av ett attributs tillstånd samt bildning och nedbrytning av en molekyl. För varje regel anges en reaktionshastighet – eller två, ifall regeln beskriver en reaktion som kan gå i båda riktningarna.

4.4 En regelbaserad modell med BioNetGen

4.4.1 Konstruktionen av den regelbaserade modellen

För att skapa en regelbaserad modell, bör man först identifiera de relevanta objekten och deras attribut. Sedan definieras de regler som beskriver systemet utgående från utveckling hos objekten och samverkan mellan attribut under simulationens gång. Dessutom behövs reaktionshastigheter för varje reaktion.

Modellen som beskrivs i detta avsnitt är uppställd enligt den indelning som används med BioNetGen. För varje del anges den engelska BNGL-termen inom parentes. Alla kodexempel är skrivna i BNGL och den kompletta versionen av modellen finns i Bilaga 1.

4.4.2 Parametrar (parameters)

Modellens parametrar består av de reaktionshastigheter som anges av Tyson i [13], angivna som enheter per minut. I denna del definierade jag även utgångskoncentrationer för objekten.

4.4.3 Objekt och attribut (molecule types)

För modellering av cellcykeln hos *S. pombe* behövs naturligtvis ett objekt som representerar själva jästcellen. Objektet behöver attribut för volym, fas av cellcykeln, celltyp och förmågan att byta typ. För denna modell är det onödigt att beskriva faserna S och G2 som skilda tillstånd, så de kombineras till en tillväxtfas. Objekt för de proteiner som styr regleringen behövs också, d.v.s. cyklin, cdc2 och deras komplex MPF. Enligt Tysons modell, representeras cyklinfamiljen av ett cyklinobjekt. Cyklin- och MPF-objekten behöver två attribut: aktivitet och bindningsställe.

Jag definierade objekten så här:

```
cell (vol~g~d, phase~g1~sg2~m, cType~m~p, switch~u~s)
cyclin (s1, aa1~u~p)
cdc2 (s2, aa2~u~p)
```

Attributen anges inom parentes och separeras med kommatecken. För varje attribut anges möjliga tillstånd. Hos `cell`-objektet kan cellvolymen `vol` anta tillstånden `g` (växande) och `d` (dubbel storlek). Cellcykelns faser anges med tillstånden `g1`, `sg2` (kombinerad s- och g2-fas) och `m` hos attributet `phase`. Celltypen och förmågan att byta typ anges med `cType` respektive `switch`, enligt beteckningen använd av Klar i [14]. Proteinobjekten `cyclin` och `cdc2` har vardera ett bindningsställe (`s1` respektive `s2`) och ett aktiveringsställe (`aa1` respektive `aa2`), varav det senare kan anta aktivt (`p`) eller inaktivt (`u`) tillstånd.

4.4.4 Starttillstånd (seed species)

I starttillståndet definieras de objekt som bör finnas i början av simuleringen, samt deras attribut och utgångskoncentrationer på följande vis:

```
cell(vol~g, phase~g1, cType~m, switch~u) mType
```

Exemplet ovan visar utgångspunkten för celler av typen M utan förmåga att byta typ, i fas G1 med växande volym. Reaktionshastigheten anges av `mType`, som är definierad i parameterdelen.

4.4.5 Relevant utdata (observables)

Simuleringar med BioNetGen kan följa utvecklingen av objekt med specifika egenskaper. I det enklaste fallet kan man ange ett objekt utan att bestämma attributen, vilket ger den totala koncentrationen av det objektet.

```
Molecules Cyclin          cyclin()
```

Exemplet ovan ger utdata för alla former av cyklin-objektet: aktivt, inaktivt och bundet till `cdc2`. Ordet `Cyclin` används som etikett för identifiering bland resultaten. `Molecules` är ett av två nyckelord för typen av objekt; det andra är `Species`. `Molecules` anger antalet kopior som passar det angivna mönstret, medan `Species` räknar de objekt och kombinationer av objekt som innehåller det mönstret.

Ifall man är intresserad av specifika reaktionsprodukter, kan ett eller flera attribut bestämmas. Här definieras celler av typ M som ej kan byta typ under etiketten `CellType_Mu`:

```
Species CellType_Mu      cell(cType~m, switch~u)
```

Man kan även specificera molekyler som är bundna till varandra. I denna modell definierade jag MPF-komplexet på följande sätt:

```
Species ActiveMPF      cyclin(aa1~p) .cdc2(aa2~u)  
Species InactiveMPF    cyclin(aa1~p) .cdc2(aa2~p)
```

MPF-komplexet definieras som cyklin bundet till `cdc2`. Hos aktivt MPF är `cdc2` icke-fosforylerat, vilket anges med tillståndet `u` för attributet `aa2` på första raden, medan inaktivt MPF har fosforylerat `cdc2` (tillstånd `p` för motsvarande attribut på andra raden). I båda fallen är cyklinet fosforylerat.

4.4.6 Regler (reaction rules)

För cellcykelmodellen kan de relevanta händelserna delas in i två grupper. Själva cellen går igenom tillväxt, övergång mellan faser och celledelning. Regleringsmekanismen bör simulera syntetisering och nedbrytning av cyclin, bildning av inaktivt MPF, aktivering av MPF och separation av MPF-komplexets delar.

Övergång från fas G1 till den kombinerade tillväxtfasen S och G2 och vidare till M är densamma för alla typer av celler och beskrivs m.h.a. förändring i tillståndet av cellens volym och fas.

```
cell(vol~g,phase~g1) -> cell(vol~g,phase~sg2) k6
cell(vol~g,phase~sg2) -> cell(vol~d,phase~m) k6
```

Utgångssituationen finns till vänster om pilen och resultatet till höger. Reaktionshastigheten k_6 är definierad i parameterdelen.

Övergången från M till G1 är mer komplicerad, för den bör även beskriva både celledelning och de resulterande typerna av dotterceller. Här är det fråga om både förändring av attributtillstånd och bildning av en molekyl.

```
cell(vol~d,phase~m,cType~p,switch~u) -> \
  cell(vol~g,phase~g1,cType~p,switch~u) + \
  cell(vol~g,phase~g1,cType~p,switch~s) k6
```

För båda dottercellerna ändras volymens tillstånd från dubbel till växande (vol från d till g) och fasen från M till G1. Denna regel visar celledelningen hos en cell av typen P som inte kan byta typ, så båda dottercellerna är av typ P, men en av dem får förmågan att byta typ (switch~s). Reglerna för de tre resterande fallen (celler av typerna Ps, Mu och Ms) följer samma mönster.

Reglerna för regleringsmekanismen är mer varierande. Ett exempel är bildandet av inaktivt MPF från aktivt cdc2 och inaktivt cyclin, vilket illustrerar hur man beskriver bindning mellan två molekyler.

```
cyclin(s1,aa1~u) + cdc2(s2,aa2~p) -> \
  cyclin(s1!1,aa1~p).cdc2(s2!1,aa2~p) k3
```

Cyclin och cdc2 binder till varann genom bindningställena s_1 och s_2 , vilket anges med en punkt mellan de två objekten och etiketten !1. Siffran 1 är ett namn för

bindningen i fråga som anger vilka molekyler är bundna till varandra, vilket har betydelse när ett objekt har två eller fler bindingar.

4.4.7 Metoder (actions)

Innan simuleringen kan startas, måste simuleringsparametrarna definieras. Med BioNetGen handlar det dels om att generera ett nätverk av de specificerade reaktionerna och dels om att simulera detta nätverk.

4.5 Simuleringar och resultat

[Sätts till om/när jag är nöjd med modellen innan den slutliga versionen lämnas in.]

5 Diskussion

BNGL följer bekant notation för kemiska reaktioner och i många fall kan man intuitivt förstå vad en regel beskriver. Dess simplicitet gör BioNetGen till ett attraktivt alternativ för att bekanta sig med regelbaserad modellering. Trots enkelheten är verktyget välanpassat för sitt ändamål, d.v.s. att skapa och simulera reaktionsnätverk mellan biomolekyler.

BNGL är ett begränsat språk. Man kan beskriva endast fem olika typer av reaktioner som är typiska för reaktioner mellan molekyler. Detta är speciellt tydligt i de engelska namnen som jag i denna avhandling översatte till objekt och attribut för att beskriva det jag försökte åstadkomma med min modell. De engelska termerna *molecule type* och *component* betyder egentligen *typ av molekyl* och *komponent*. När man anger de möjliga tillstånden för *komponenter* hos en *typ av molekyl*, måste alla möjliga tillstånd specificeras. Detta är vettigt när det är fråga om molekyler i end signalkaskad, men gör det svårt att beskriva en egenskap vars tillstånd varierar kontinuerligt, såsom cellvolymen i cellcykelexemplet.

Cellcykelmodellen innefattar två olika nivåer. Reglerna för regleringsmekanismen beskriver händelser på proteinnivån, medan byte av parningstyp sker på cellnivån. BioNetGen utvecklades med tanke på reaktionsnätverk mellan molekyler på samma nivå, som t.ex. intracellulära signaleringskaskader. Därmed är det inte ett idealiskt verktyg för modellering av system som länkar ihop reaktioner på flera nivåer. I den nuvarande versionen beskrivs i princip två skilda modeller, vars enda samband består av en gemensam reaktionshastighet. Jag saknar ett sätt att beskriva sambandet på ett hierarkiskt vis, som kan hantera och utnyttja nivåskillnaderna.

Trots begränsningarna hos BioNetGen och BNGL när det gäller exempelmodellen, finns det många viktiga fördelar med regelbaserad modellering. Regelbaserade modeller är mer flexibla än rent matematiska modeller, för man kan enkelt lägga till, ta bort och justera både regler och molekyler. En existerande modell kan vidareutvecklas för nya ändamål eller anpassas för modellering av närbesläktade system.

6 Referenser

- [1] L. A. Chylek, L. A. Harris, C.-S. Tung, J. R. Faeder, C. F. Lopez och W. S. Hlavacek, "Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 6, nr 1, pp. 13-36, 2014.
- [2] C. Maus, S. Rybacki och A. M. Uhrmacher, "Rule-based multi-level modeling of cell biological systems," *BMC Systems Biology*, vol. 5, nr 166, 2011.
- [3] L. A. Chylek, E. C. Stites, R. G. Posner och W. S. Hlavacek, "Innovations of the Rule-Based Approach," i *Systems Biology*, A. Prokop och B. Csukás, Red., Springer Netherlands, 2013, pp. 273-300.
- [4] J. R. Faeder, M. L. Blinov och W. S. Hlavacek, "Rule-Based Modeling of Biochemical Systems with BioNetGen," i *Systems Biology*, I. V. Maly, Red., Totowa, New Jersey: Humana Press, 2009, pp. 113-167.
- [5] B. A. Moser och P. Russell, "Cell cycle regulation in *Schizosaccharomyces pombe*," *Current Opinion in Microbiology*, vol. 3, nr 6, pp. 631-636, 2000.
- [6] K. Powell, "All systems go," *The Journal of Cell Biology*, vol. 165, nr 3, pp. 299-303, 2004.
- [7] H. Kitano, "Systems Biology: A Brief Overview," *Science*, vol. 295, nr 5560, pp. 1662-1664, 2002.
- [8] J. R. Faeder, M. L. Blinov, B. Goldstein och W. S. Hlavacek, "Rule-Based Modeling of Biochemical Networks," *Complexity*, vol. 10, nr 4, pp. 22-41, 2005.
- [9] Y. Lazebnik, "Can a biologist fix a radio?--Or, what I learned while studying apoptosis," *Cancer Cell*, vol. 2, nr 3, pp. 179-182, 2002.
- [10] M. Hwang, M. Garbey, S. A. Berceli och R. Tran-Son-Tay, "Rule-Based Simulation of Multi-Cellular Biological Systems - A Review of Modeling Techniques," *Cellular and Molecular Bioengineering*, vol. 3, nr 2, pp. 285-294, 2009.
- [11] D. L. DeAngelis och W. M. Mooij, "Individual-Based Modeling of Ecological and Evolutionary Processes," *Annual Review of Ecology, Evolution, and Systematics*, vol. 36, pp. 147-168, 2005.
- [12] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts och P. Walter, *Molecular Biology of the Cell*, 6:e red., New York: Garland Science, 2014, pp. 963-1034.
- [13] J. J. Tyson, "Modeling the cell division cycle: cdc2 and cyclin interactions," *Proc. Natl. Acad. Sci. USA*, vol. 88, nr 16, pp. 7328-7332, 1991.
- [14] A. J. S. Klar, "Developmental choices in mating-type interconversion in fission yeast," *Trends Genet.*, vol. 8, nr 6, pp. 208-213, 1992.

Bilaga 1: Cellcykelmodellen med BioNetGen

```

begin parameters
  #initial concentrations
  cyclin0  0
  cdc20    200
  mType    20
  pType    0

  #steps (Tyson model, steps 1-9), per min
  k1      0.015  #synthesis of cyclin
  k2      0.0    #breakdown of cyclin
  k3      200    #formation of preMPF
  k4      50     #activation of MPF (adjustable, 10-1000)
  k5      0     #deactivation of MPF
  k6      4     #breakdown of MPF (adjustable, 0.1-10)
  k7      0.6   #dephosph of cyclin
  k8      1000  #phosph of cdc2 (>>k9)
  k9      100   #dephosph of cdc2 (>>k6)

#   Td      116      #time to double cell volume in min
end parameters

begin molecule types
  #volume: growing, double; switch: unswitchable, switchable
  cell(vol~g~d,phase~g1~sg2~m,cType~m~p,switch~u~s)

  #cell cycle control mechanism
  cyclin(s1,aa1~u~p)
  cdc2(s2,aa2~u~p)

  #creation and destruction of molecules
  trash()
  I()
end molecule types

begin seed species
  cell(vol~g,phase~g1,cType~m,switch~u)      mType
#  cell(vol~g,phase~g1,cType~p,switch~u)      pType
  cyclin(s1,aa1~u)                          cyclin0
  cdc2(s2,aa2~u)                             cdc20
  I()                                         1    #for synthesis
end seed species

begin observables
  Species      ActiveMPF      cyclin(aa1~p).cdc2(aa2~u)
  Species      InactiveMPF    cyclin(aa1~p).cdc2(aa2~p)
  Molecules    Cyclin         cyclin()
  Molecules    CellType_Mu    cell(cType~m,switch~u)
  Molecules    CellType_Pu    cell(cType~p,switch~u)
  Species      CellType_Ms    cell(cType~m,switch~s)
  Species      CellType_Ps    cell(cType~p,switch~s)
  Species      Phase_G1      cell(phase~g1)
  Species      Phase_S_G1    cell(phase~sg2)
  Species      Phase_M       cell(phase~m)
  Species      Cells_tot     cell()
end observables

begin reaction rules
  # CELL GROWTH
  #phase g1 to s/g2

```

```

cell(vol~g,phase~g1) -> cell(vol~g,phase~sg2)          k6
#phase sg2 to m
cell(vol~g,phase~sg2) -> cell(vol~d,phase~m)          k6

# M PHASE, TYPE SWITCHING
#phase m to g1, Pu type
cell(vol~d,phase~m,cType~p,switch~u) -> \
    cell(vol~g,phase~g1,cType~p,switch~u) + \
    cell(vol~g,phase~g1,cType~p,switch~s) k6
#phase m to g1, Ps type
cell(vol~d,phase~m,cType~p,switch~s) -> \
    cell(vol~g,phase~g1,cType~p,switch~s) + \
    cell(vol~g,phase~g1,cType~m,switch~u) k6
#phase m to g1, Mu type
cell(vol~d,phase~m,cType~m,switch~u) -> \
    cell(vol~g,phase~g1,cType~m,switch~u) + \
    cell(vol~g,phase~g1,cType~m,switch~s) k6
#phase m to g1, Ms type
cell(vol~d,phase~m,cType~m,switch~s) -> \
    cell(vol~g,phase~g1,cType~p,switch~u) + \
    cell(vol~g,phase~g1,cType~m,switch~s) k6

# CYCLIN, CDC2 AND MPF
#synthesis and breakdown of cyclin
I() -> I() + cyclin(s1,aa1~u)                          k1
cyclin(aa1~u) -> trash()                               k2
#form inactive MPF ("preMPF")
cyclin(s1,aa1~u) + cdc2(s2,aa2~p) -> \
    cyclin(s1!1,aa1~p).cdc2(s2!1,aa2~p) k3
#activate MPF (dephosphorylation of cdc2)
cyclin(aa1~p).cdc2(aa2~p) <-> cyclin(aa1~p).cdc2(aa2~u) \
    k4, k5
#destroy active MPF
cyclin(s1!1,aa1~p).cdc2(s2!1,aa2~u) -> cyclin(s1,aa1~p) + \
    cdc2(s2,aa2~u) k6
#dephosphorylation of cyclin
cyclin(aa1~p) -> trash()                               k7
#phosphorylation of cdc2
cdc2(aa2~u) <-> cdc2(aa2~p)                          k8, k9
end reaction rules

begin actions
    generate_network({overwrite=>1, TextReaction=>1});
    simulate_ode({suffix=>"ode", t_start=>0, t_end=>200, \
        n_steps=>10});
    saveConcentrations();
end actions

```