

Inhoudsopgave

1. Inleiding	2
2. Typogenetics	5
2.1. Genexpressie	5
2.1.1. Het centrale dogma van de moleculaire biologie: DNA → RNA → eiwit	5
2.1.2. DNA	5
2.1.3. DNA → RNA: Transcriptie	7
2.1.4. RNA → eiwit: Translatie.....	8
2.1.5. Eiwitten	9
2.1.6. Eiwitstructuur	10
2.2. Werking van Typogenetics	10
2.3. Zelfreproductie	15
2.4. Onderzoek met Typogenetics.....	17
2.4.1. Het genetisch algoritme	18
2.4.2. Determinisme	20
3. Verschillen tussen genexpressie en Typogenetics	21
3.1. DNA: De consensussequentie.....	23
3.2. RNA: Intronen	25
3.3. Eiwit: Eiwitstructuur	26
3.4. Concluderend	28
4. Abstractie	30
4.1. Holisme en reductionisme.....	30
4.2. Het samen zwermen van vogels.....	32
5. Conclusie	33

1. Inleiding

“An abstraction is one thing that represents several real things equally well.” ~ Edsger W. Dijkstra

Abstractie is het weglaten van alle niet essentiële informatie of aspecten om meer fundamentele structuren zichtbaar te maken. Abstractie kan gebruikt worden om complexe systemen beter te begrijpen. De onderdelen die niet nodig zijn worden weggelaten waardoor het systeem transparanter en overzichtelijker wordt. Het citaat van Dijkstra laat zien dat abstraheren in de praktijk een stuk moeilijker is dan op het eerste gezicht lijkt. Ten eerste is het vaak niet duidelijk welke dingen je precies probeert te abstraheren; een systeem bestaat uit verschillende onderdelen en het onderscheid is niet altijd even duidelijk. Daarnaast is de term *“equally well”* niet erg wetenschappelijk meetbaar. In de praktijk betekent dit dat abstraheren meer een kunst is dan een wetenschap.

In deze scriptie zullen wij Typogenetics¹ bekijken als case study voor een abstractie. Typogenetics is een formeel systeem bedacht door Douglas Hofstadter. Typogenetics is een abstractie van genexpressie. Met genexpressie doelen we op het proces van overdracht van informatie tussen DNA, waarin de informatie ligt opgeslagen, en een eiwit. In ons paper zullen we aan de hand van het systeem Typogenetics bekijken welke moeilijkheden men treft bij het maken van een goede abstractie, in het geval van Typogenetics dus een goede representatie van genexpressie. Met een goede representatie bedoelen we het volgende. Genexpressie is een systeem met een bepaalde mate van complexiteit en in de abstractie willen we deze complexiteit behouden. Met een goede representatie bedoelen we niet dat het zich op de dezelfde functionele manier gedraagt. We zijn dus niet geïnteresseerd naar de input en output maar naar de manier waarop informatie wordt gemanipuleerd. Het antwoord op deze vraag is niet te beantwoorden met behulp van één wetenschapsgebied. De kennis van de complexiteit van het systeem genexpressie komt van de kant van de moleculaire biologie. Cognitieve Kunstmatige Intelligentie (CKI) kan de mate complexiteit beoordelen. Binnen de CKI worden systemen gecategoriseerd naar uitdrukingskracht. Met uitdrukingskracht wordt bedoeld met het vermogen van een systeem om bepaalde functies uit te voeren.

Ons doel is niet om kritiek te leveren op het werk van Hofstadter. Hofstadter heeft nooit het doel gehad om een allesomvattende abstractie te maken van genexpressie. Het concept Typogenetics

¹ Douglas R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid*, 1978, Chapter XVI, Self-Ref and Self-Rep

komt uit Hofstadter's boek *Gödel, Escher, Bach, an Eternal Golden Braid* en met dit boek wil hij onderbouwen wanneer betekenisloze onderdelen samen een betekenisvol geheel vormen. Typogenetics is bedacht om zijn standpunten te onderbouwen en niet zozeer om een perfecte abstractie van genexpressie te maken. Dit wordt ook door Hofstadter zelf bevestigd: "*Of course, Typogenetics involves many simplifications, and therefore is useful primarily for didactic purposes*".²

Desondanks is het toch interessant om een kritische blik te werpen op Typogenetics. Ten eerste is Typogenetics omarmd binnen het onderzoek naar Artificial Life^{3 4 5 6 7}, een onderzoeksgebied van CKI. Artificial Life houdt zich bezig met de vraag hoe het zo kan zijn dat uit een aantal simpele chemische elementen complexe structuren ontstaan die samen een systeem vormen dat zichzelf in stand kan houden. Onderzoek met dit soort chemische processen is problematisch vanwege de extreem kleine schaal waarop ze plaatsvinden en dat resultaten moeilijk te reproduceren zijn. Het is daarom moeilijk een goede beschrijving te geven van de processen die plaatsvinden. Artificial Life heeft Typogenetics daarom gebruikt als geabstraheerde versie van genexpressie. Typogenetics is geschikt voor dit soort onderzoek omdat het makkelijk te simuleren is op een computer. Hoe dit soort onderzoek met Typogenetics werkt zal later beschreven worden nadat we het systeem zelf hebben uitgelegd.

Daarnaast is het ook interessant omdat Typogenetics het reductionisme van biologie combineert met het holisme van CKI. Reductionisme is een manier om bepaalde wetenschappelijke fenomenen te bestuderen. Het reduceren van complexe systemen tot de interacties van de onderdelen van dat systeem kan leiden tot een beter begrip van het geheel. Binnen de biologie is reductionisme noodzakelijk om moleculair-biologische systemen uit te leggen. De cel bestaat uit allerlei complexe chemische structuren. Door de werking van de afzonderlijke structuren te onderzoeken, kan een totaalbeeld van de werking van de cel als geheel worden verkregen. De structuren zijn moleculen en voor zover bekend worden processen in de cel voornamelijk bepaald door chemische en natuurkundige interacties van diezelfde moleculen. Als al deze interacties in kaart kunnen worden gebracht, kunnen we precies begrijpen welke processen in de cel plaatsvinden en wat deze in gang zetten. De reductionistische inslag van de biologie wordt doorgevoerd tot op allerkleinste niveau.

² Douglas R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid*, 1978, p. 504

³ V. Kvasnička J. Pospíchal T. Kaláb, A Study of Replicators and Hypercycles by Typogenetics, In *Advances in Artificial Life*, pages 37-54, Springer Berlin / Heidelberg, 2001

⁴ Kyubum Wee, Woosuk Lee, A Extensions and Variations on Construction of Autoreplicators in Typogenetics, In *Advances in Artificial Life*, pages 685-694, Springer Berlin / Heidelberg, 2005

⁵ Chohwa Gwak Kyubum Wee, Construction of Hypercycles in Typogenetics with Evolutionary Algorithms, In *Advances in Artificial Life*, pages 1060-1068, Springer Berlin / Heidelberg, 2007

⁶ Harold C. Morris. Typogenetics: A logic for artificial life. In Christopher G. Langton, editor, *Artificial Life*, pages 369–395, Redwood City, California, 1989. Addison-Wesley.

⁷ Louis Varetto. Typogenetics: An artificial genetic system. *Journal of Theoretical Biology*, 160:185–205, 1993

Francis Crick, een invloedrijke wetenschapper die onder andere de structuur van DNA heeft ontrafeld: *'The ultimate aim of the modern movement in biology is to explain all biology in terms of physics and chemistry'*⁸. Heden ten dagen wordt met behulp van bekende chemische variabelen (interacties) voorspeld hoe processen in de cel verlopen.

CKI kijkt meer van uit een holistisch standpunt naar de werking van een levende cel. Holisme is het idee dat de eigenschappen van een systeem niet kunnen worden verklaard door de som van zijn delen. Een CKI'er is helemaal niet zo geïnteresseerd in de werking van de verschillende onderdelen van de cel. Het is veel interessanter dat de cel een levend systeem is terwijl het uit alleen maar levenloze componenten bestaat. Het 'levend' zijn van de cel is een eigenschap die niet kan worden toegekend aan de onderdelen van de cel. Het ontstaan van dit soort eigenschappen wordt emergentie genoemd. Emergentie is een concept verwant aan holisme. Emergentie is het ontstaan van eigenschappen die niet kunnen worden toegeschreven aan de losse componenten van het systeem.

⁸ Crick, F ,Molecular Biology of the Gene 1966

2. Typogenetics

Om de abstractie van Typogenetics te onderzoeken moeten we eerst een overzicht geven van de werking van genexpressie. Daarna volgt een uitleg van het systeem Typogenetics en de manier waarop daar onderzoek mee wordt gedaan. Het is van belang een goed beeld te hebben van de werking van genexpressie zodat het proces van abstractie zichtbaar wordt.

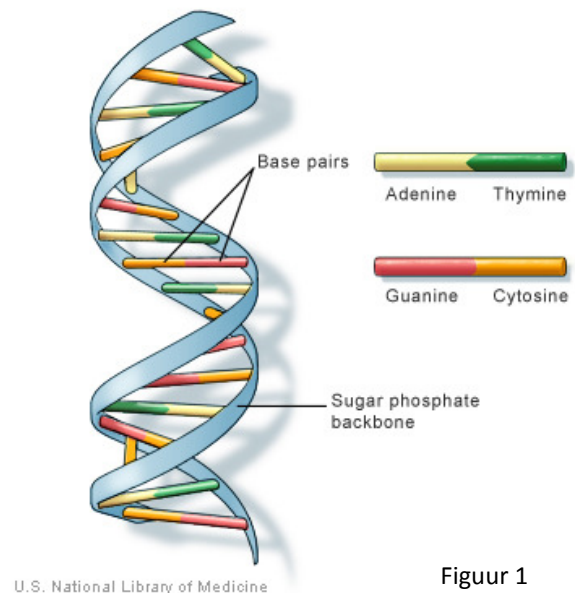
2.1. Genexpressie

2.1.1. Het centrale dogma van de moleculaire biologie: DNA → RNA → eiwit

Genexpressie is het aflezen van genen en het omzetten van de genen in functionele eiwitten. Dit gaat volgens het principe DNA → RNA → eiwit, ook wel het centrale dogma van de moleculaire biologie genoemd. In 1958 werd het dogma geformuleerd door James Watson en Francis Crick.⁹ Aan de hand van dit dogma zal genexpressie worden uitgelegd.

2.1.2. DNA

In de chemische structuur van het DNA ligt de informatie opgeslagen die de cel nodig heeft om te groeien, een metabolisme in stand houden, te delen en andere processen die de cel levend houden.¹⁰ DNA is een lineair polymeer (een molecuul bestaande uit zich repeterende eenheden) waarvan de eenheden bestaan uit vier verschillende nucleotiden, die op iedere mogelijke volgorde aan elkaar gekoppeld kunnen worden. Deze ketens kunnen miljoenen nucleotiden lang worden en zijn voor ieder individueel organisme uniek.



Figuur 1

⁹ Crick, F., *Of Molecules and Men*, 1966

¹⁰De beschrijving van het centrale dogma is gedaan met behulp van de boeken '*Biology*', door Campbell en Reece, 8th edition (2007), en Brown, T.A., *Genomes 3*, Garland Sciences Publishing (2007), beide opgenomen in de literatuurlijst.

De nucleotiden van DNA bestaan uit een suikergroep, een fosfaatgroep en een base. De suikergroep van nucleotide één, bindt aan de fosfaatgroep van nucleotide twee, enzovoorts. De base van elke nucleotide speelt geen rol in deze reactie, maar zorgt voor de verschillen tussen de vier nucleotiden: de vier verschillende bases zijn Adenosine, Cytidine, Guanosine en Thyminine (afgekort A, C, G en T). Cytidine en Thyminine hebben een andere chemische structuur dan Adenosine en Guanine, daarom worden C en T *pyrimidines* genoemd, en A en G purines.

Wanneer basen aan elkaar gekoppeld zijn tot een keten, is een base in staat waterstofbruggen (een zwakke chemische verbinding) te vormen met een andere base, die zich in een andere keten bevindt. Deze andere keten wordt de complementaire keten genoemd, omdat de baseparing altijd verloopt via het principe: A combineert met T, en C met G, dus een pyrimidine met een purine. Na voltooiing van alle chemische reacties zijn er dus twee ketens ontstaan, elk bestaande uit nucleotiden waarvan de basen complementair binden aan de andere keten (zie figuur 1). De structuur die ontstaat wordt de *double helix* genoemd, omdat vanwege de chemische eigenschappen van de betrokken atomen er een bepaald hoek in de keten ontstaat. Deze dubbele helix, in iedere menselijke cel aanwezig, omvat de genetische code en daarmee alle erfelijke informatie van de cel.

Een van de belangrijke kenmerken als gevolg van de specifieke baseparing, is dat er perfecte kopieën gemaakt kunnen worden (want C combineert altijd met G, en A altijd met T). De kopie die dan ontstaat is echter wel complementair. Bijvoorbeeld: de keten CCAGGA wordt, wanneer deze gekopieerd wordt GGTCCT. Er zijn twee belangrijke gebeurtenissen waarin DNA gekopieerd wordt. De eerste betreft DNA-replicatie: wanneer een cel deelt, zal ook de erfelijke informatie van die cel gekopieerd moeten worden. Hiertoe wordt DNA-replicatie in gang gezet tijdens een bepaald 'setpoint' in de cel-cyclus. Het tweede betreft de transcriptie van genen. Genen zijn sequenties in het DNA die coderen voor een bepaald eiwit, of ander transcript, die op hun beurt diverse functies uitvoeren in de cel.

De historie van het begrijpen van dit principe gaat meer dan 70 jaar terug: in 1941 kwamen Edward Tatum en George Beadle met hun 'One gene, one enzyme' hypothese.¹¹ Hiermee doelen ze op een enkel gen dat codeert voor een enkel enzym dat een bepaalde biochemische reactie katalyseert in de cel. Deze hypothese werd ondermijnd door bevindingen in het lab. Uit diverse experimenten bleek dat de processen ingewikkelder waren, en dat de transcriptie van genen niet kon plaatsvinden zonder hulp van andere processen. Dit laat zien dat genexpressie een complex systeem is waarbij

¹¹ Beadle GW, Tatum EL., "Genetic Control of Biochemical Reactions in Neurospora". 1941, p. 499–506.

veel verschillende moleculen een rol spelen. Het gegeven dat een gen via RNA wordt vertaald naar een eiwit, vormt nog steeds de basis voor hedendaagse moleculaire genetica, met als twee belangrijkste processen transcriptie en translatie.

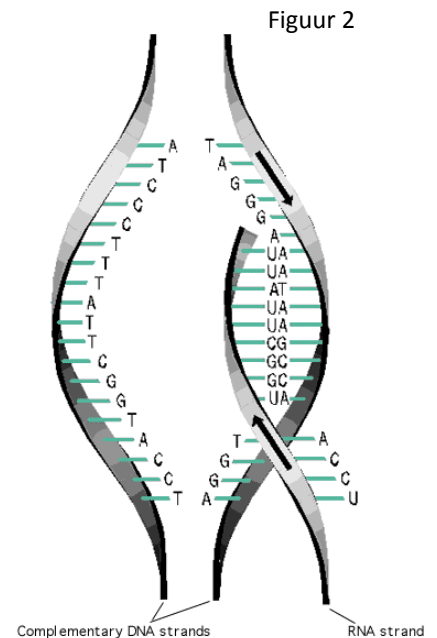
2.1.3. DNA → RNA: Transcriptie

Transcriptie is het proces waarbij de informatie die in de sequentie van het DNA ligt wordt omgezet in een RNA keten, waarna de informatie kan worden verspreid in de cel. Op elk willekeurig moment in de celcyclus bevinden zich honderdduizenden transcripten in het cytoplasma. Al deze afgeschreven sequenties, in de vorm van mRNA (messenger RNA), zijn ontstaan tijdens de transcriptie van genen. Dit proces gebeurt in twee fasen: het eerste is transcriptie initiatie, het tweede synthese en verwerking van het RNA.

Transcriptie initiatie gebeurt met behulp van DNA-bindende eiwitten. Deze eiwitten binden aan een bepaalde signaalsequentie, de 'promoter', die vlakbij het gen gelegen is. Daarna werken ze als signaal voor het enzym RNA-polymerase, dat nu kan binden aan het complex van DNA en eiwitten. Was het eerst nog een *gesloten promoter complex*, nu is het, door de binding van RNA polymerase een *open promoter complex*. De twee complementaire ketenen van het DNA zijn los van elkaar gekomen en het transcriptie initiatiecomplex is gevormd.

De volgende stap is de synthese van het mRNA. Doordat de keten van het DNA dat als *template* dient is vrijgekomen van zijn complementaire keten, kunnen nieuwe complementaire nucleotides worden toegevoegd. RNA-polymerase, het enzym dat de nieuwe keten maakt, laat alleen nucleotides toe die net iets anders zijn in structuur dan DNA: de base Thymine is vervangen door Uracil en in plaats van een 'desoxyribose' is de ribose nu met een extra zuurstof. Vandaar dat het transcript de naam RNA draagt in plaats van DNA.

RNA-polymerase katalyseert dus de volgende processen: het loskomen van de twee DNA-ketens, faciliteren van binding van een RNA-nucleotide met een DNA-nucleotide, en de chemische reactie tussen twee RNA-nucleotiden. Hierdoor ontstaat weer een keten. De volgorde van die keten wordt bepaald door de volgorde van de nucleotiden van het DNA, omdat C alleen met G paart en A alleen met, in dit geval, U. Zo wordt een tegenovergestelde sequentie verkregen.



Het RNA-polymerase schuift nu verder over de keten, telkens naar een nieuw deel van de DNA sequentie. Dit proces wordt elongatie genoemd en zorgt ervoor dat de RNA keten zijn lengte krijgt. Aan de achterkant van het polymerase-enzym laat het RNA weer los van het DNA, waarna het DNA zich weer sluit tot zijn normale dubbele helix structuur. Zo worden, afhankelijk van de lengte van het gen, honderden RNA-nucleotides aan elkaar gekoppeld, totdat er terminatie optreedt en de keten volledig loskomt van het DNA. Het DNA krijgt weer zijn normale conformatie (ook alle DNA-bindende eiwitten laten weer los).

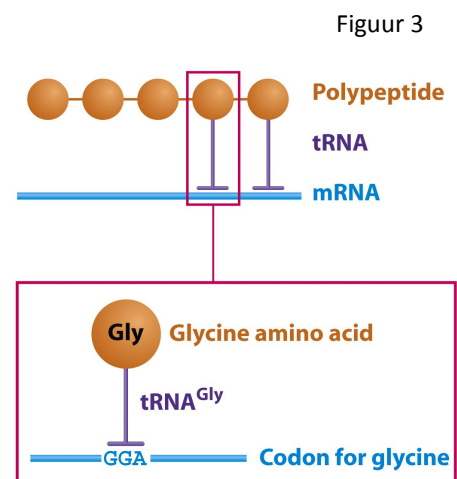
Deze keten van RNA is vrij instabiel, mede omdat er geen complementaire keten aan gekoppeld zit. De keten wordt in deze staat pre-RNA genoemd, omdat er nog allerlei *post-translational modifications* (PTM) moeten worden uitgevoerd, waardoor het RNA stabiel wordt. Het pre-RNA heet nadat alle PTM's zijn uitgevoerd 'messenger RNA' (mRNA) en via een kernporie verlaat de keten de kern en kan in het cytoplasma translatie optreden.

2.1.4. RNA → eiwit: Translatie

Translatie is het proces waarin de genetische code, die nu is opgeslagen in een mRNA keten, wordt omgezet in een eiwit. Het proces translatie gaat volgens een aantal stappen. Eerst moet de genetische code, nu opgeslagen in een mRNA molecuul, worden ontcijferd. Daarna wordt met behulp van het ribosoom het eiwit gevormd door middel van een groeiende keten van aminozuren, die de bouwstenen zijn van eiwitten.

Naast mRNA's, spelen nog verschillende andere RNA's een rol in de cel. Bij het translatieproces is vooral tRNA betrokken, dat staat voor transfer RNA. Deze RNA vormt zowel een fysieke link (de tRNA's binden aan het mRNA en aan de groeiende keten van aminozuren) als een informatieve link (de tRNA's zorgen ervoor dat de aminozuurvolgorde overeenkomstig is met de mRNA sequentie).

De structuur van een tRNA molecuul is als volgt: aan de staartzijde van het molecuul bevindt zich een *anticodon*, een vrije RNA sequentie van drie nucleotiden lang. Deze drie nucleotiden zijn specifiek voor het aminozuur wat aan de kopzijde van het tRNA molecuul zit. Een tRNA met anticodon GAG, bijvoorbeeld, bezit altijd het aminozuur Leucine (Leu). Het mRNA wordt afgelezen per drie nucleotiden, ook wel tripletten, of *codons*

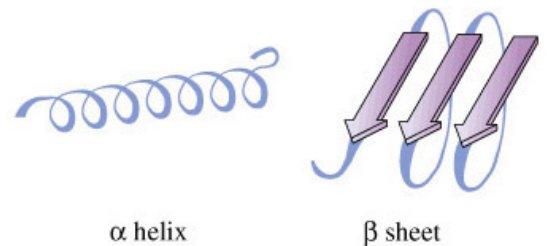


genoemd. Iedere sequentie van drie codeert voor een aminozuur. Met behulp van een speciaal enzym, het ribosoom, worden alle aminozuren aan elkaar vastgemaakt zodat er een keten ontstaat (zie figuur 3). Dus als een mRNA 330 nucleotiden lang is, zal er na translatie een eiwit van 110 aminozuren ontstaan. De volgorde van de aminozuren wordt dus indirect bepaald door de genetische code:

DNA	GAGGAGGAG
mRNA	CUCCUCCUC
tRNA's	GAG-GAG-GAG
Aminozuren	Leu-Leu-Leu

2.1.5. Eiwitten

Welke aminozuren er in het eiwit zitten, en de manier waarop deze gerangschikt zijn, bepalen de eigenschappen van het eiwit. Er zijn 20 verschillende aminozuren, elk met hun eigen specifieke kenmerken. Een aminozuurketen bestaat, evenals een keten van DNA of RNA uit een *backbone*, in dit geval bestaande uit peptidebindingen. Wat elk aminozuur onderscheidt is een functionele groep. Omdat eiwitten verschillende functies hebben in de cel, bestaan ze ook uit verschillende aminozuurconstructies. Er zijn twee basiskenmerken van eiwitten: of ze zijn *hydrofoob*, of ze zijn *hydrofiel*. Dit betekent respectievelijk waterafstotend en wateroplosbaar. Een eiwit dat een functie heeft in het cytoplasma (per definitie een waterige oplossing) moet dus vooral uit hydrofiel aminozuren bestaan, terwijl een membraaneiwit alleen vast kan blijven in het membraan als hydrofobe aminozuren contact maken met de fosfolipiden (vetten) waaruit het membraan is opgebouwd.



Een eiwit bestaat echter altijd uit een combinatie van eigenschappen. Dus ook een hydrofiel eiwit, met een functie in het cytoplasma, heeft hydrofobe aminozuren. Deze zitten dan meestal aan de binnenkant van het eiwit. Een eiwit is een driedimensionale structuur, en lijkt na volledige totstandkoming in het geheel niet meer op de keten zoals die direct na de translatie was.

Net als bij DNA bestaat Typogenetics ook uit vier nucleotiden, die samen twee baseparen vormen. Deze worden gerepresenteerd doormiddel van de letters A,T en C,G. Het systeem van Typogenetics bestaat uit ketens van deze letters. Een aantal voorbeelden zijn:

GGC

ATGCCA

TATATATATAT

tabel 1

De ketens kunnen worden afgelezen, een vergelijkbaar proces als translatie. Elke twee letters representeren een 'Typogenetisch' aminozuur. (zie tabel 1). In de werkelijkheid worden aminozuren door drie basen gerepresenteerd. Maar voor het formele systeem maakt dit niks uit, drie bases zou

	A	C	G	T
A	STOP	cut _s	del _s	swi _r
C	mvr _s	mvl _s	cop _r	off _l
G	ina _s	inc _r	ing _r	int _l
T	rpy _r	rpu _l	lpy _l	lpu _l

alleen maar meer mogelijkheden geven. Door voor twee bases te kiezen blijft het systeem overzichtelijker. Al deze aminozuren worden met elkaar verbonden en vormen een Typogenetisch enzym. Binnen de biologie maakt men onderscheid tussen transcriptie en translatie. Transcriptie is het aflezen van het DNA naar mRNA, translatie is het vormen van enzymen aan de hand van de afgelezen informatie. In de werkelijkheid is het vormen van enzym heel ingewikkeld en bestaat uit meerdere processen. Bij Typogenetics zijn deze processen geabstraheerd in bovenstaande tabel. Bij Typogenetics is dus nog maar een proces en is dus het onderscheid tussen translatie en transcriptie betekenisloos geworden. Om die reden zullen we nu als we het over Typogenetics hebben praten over transcriptie en zullen de term translatie achterwege laten.

Een voorbeeld van transcriptie:

De keten

CGTCCTGAGG

Aminozuren

CG-TC-CT-GA-GG

Vormen het enzym

cop-rpu-swi-ina-ing

AA representeert de stop functie. Hierbij wordt het huidige enzym afgesloten en een nieuwe enzym gestart.

De keten

CGTCAACTGAGG

Vormt het enzym

cop-rpu

en

swi-ina-ing

tabel 2

Elk aminozuur representeert een bepaalde operatie (zie tabel 2). Een enzym is dus als het ware een takenpakket . Dus het takenpakket van het enzym *cut-rpu-swi-ina-ing* is:

Aminozuren	
cut	Cut strand(s).
del	Delete a base from the strand.
swi	Switch enzyme to other strand.
mvr	Move one unit to right.
mvl	Move one unit to left.
cop	Turn on copy mode.
off	Turn of copy mode.
ina	Insert A to the right of this unit.
inc	Insert C to the right of this unit.
ing	Insert G to the right of this unit.
int	Insert T to the right of this unit.
rpy	Search for nearest pyrimidine to the right.
rpu	Search for nearest purine to the right.
lpy	Search for nearest pyrimidine to the left.
lpu	Search for nearest purine to the left.

Ga in copy mode

Zoek de eerste purine aan de rechter kant

Switch na de andere kant

Zet een A aan de rechterkant.

Zet een G aan de rechterkant.

De meeste functies spreken voor zich maar enkelen moeten nog worden uitgelegd. Ten eerste de copy mode: als copy mode aan is en het enzym beweegt zich over de keten, dan zal zich een complementaire keten vormen naast de huidige keten. Complementair betekent voor A een T en visa versa en voor G een C en visa versa. Stel een enzym beweegt zich over onze keten met copy mode aan, dan gebeurt er dit:

CGTCCTGAGG

Copy mode...

*GCAGGACTCC**

CGTCCTGAGG

Als laatste nog de switch. Bij deze operatie springt het enzym over naar de andere keten gemaakt door copy mode.

GCAGGACTCC*
CGTCCTGAGG

Switch

GCAGGACTCC*
CGTCCTGAGG

Bij de complementaire ketens staat een asterisk. Dit om aan te geven dat de gekopieerde keten in spiegelbeeld is. Dus de twee uiteindelijke ketens zijn CGTCCTGAGG en CCTAGGACG. Dit is analoog aan het feit dat een enkele DNA keten een bovenkant en onderkant heeft.

Het laatste wat we nog nodig hebben het startpunt van het enzym. Hierbij spelen de subscripts uit tabel 1 een rol. De letters *s,l,r* staan respectievelijke voor straight, left en right. Dit bepaalt in welke richting het volgende aminozuur gebonden zal worden: ongewijzigd (recht door), een hoek naar links of een hoek naar rechts. De dispositie van het eerste aminozuur maakt niks uit omdat links, rechts en recht door relatief zijn. Dus bij ons enzym CGTCCTGAGG hoort de structuur:

swi – ina
| |
cop – rpu ing

De bindingsvoorkeur van het enzym wordt bepaald door de stand van het eerste en laatste segment. (zie tabel 3). Bij ons enzym is het eerste segment (*cop – rpu*) naar rechts en het laatste segment (*ina – ing*) naar onder gericht. Dus ons enzym heeft een voorkeur voor de letter G. Het enzym zal dus beginnen bij de letter G van een bepaalde keten. Als een keten

tabel 3		
Eerste segment	Laatste segment	Bindingsletter
⇒	⇒	A
⇒	↑	C
⇒	↓	G
⇒	⇐	T

meerdere letters G heeft is niet gedefinieerd welke letter er door het enzym gekozen wordt. Dus het startpunt is dan op een willekeurige G in de keten.

Deze structuur is analoog aan de tertiaire structuur die we in DNA vinden, namelijk de aminozuren in de ruimte. Deze tertiaire structuur van een eiwit bepaalt ook de mogelijkheid tot aanhechtingspunten en is afhankelijk van de aminozuren in de keten. Alleen is de structuur zo sterk versimpeld bij Typogenetics dat de overeenkomsten hierbij ophouden.

Voorbeeld:

Nu zal er een volledige ronde van het systeem Typogenetics worden voorgedaan:

Onze keten is:

CGTCCTGAGG

De transcriptie naar het enzym:

cop-rpu-swi-ina-ing

Met als structuur:

swi – ina

| |

cop – rpu ing

Dus de bindingsletter van ons enzym is:

G

Stel dat ons enzym zich bindt aan de eerste G van onze keten:

CGTCCTGAGG

Stap 1: Ga in Copymode

C*

CGTCCTGAGG

Stap 2: Zoek de eerste purine aan de rechterkant:

CAGGAC*
CGTCCTGAGG

Stap 3: Switch naar de andere kant:

CAGGACA*
CGTCCTGAGG

Stap 4: Zet een A een de rechterkant

CAGGACA*
CGTCCTGAGG

Stap 5: Zet een G een de rechterkant

CAGGACGA*
CGTCCTGAGG

Klaar!

Dus nu dat het enzym klaar hebben we twee ketens. De originele CGTCCTGAGG en de nieuwe AGCAGGAC (let op spiegelbeeld). Het hoeft niet altijd zo te zijn dat de originele keten nog intact is.

2.3. Zelfreproductie

Een interessante vraag is of Typogenetics de mogelijkheid heeft om zich zelf in stand te houden, net zoals het werkelijke DNA. Deze vraag werd ook gesteld door Hofstadter. Een tactiek om zichzelf in stand te houden is zelfreproductie.

Zelfreproductie is iets anders dan zelfperpetuatie. Bij zelfperpetuatie keert de keten cyclisch weer terug naar zijn originele toestand. De keten wordt dus wel in stand gehouden maar niet gedupliceerd. Zelfperpetuatie is niet een moeilijk te verkrijgen eigenschap. Er zijn een aantal hele

triviale voorbeelden mogelijk. Een voorbeeld is de keten CACCCACC. Deze keten produceert een enzym dan geen invloed op de configuratie van de keten. Er vernadert dus niks en de keten wordt dus in stand gehouden. Ook het de keten uit ons voorbeeld is een triviale zelfperpetuator. Een minder triviaal voorbeeld: TCAGCGATGTCTAGAGCT¹².

Zelfreproductie is nooit door Hofstadter zelf aangetoond. Andere hebben later wel ketens bedacht die wel voor zelfreproductie zorgen. Tevens zijn er doormiddel van computersimulaties ketens gevonden die in staat zijn tot zelfreproductie^{13 14 15 16}. Andrew Snare¹⁷ beschrijft een methode die gebaseerd is op een methode van Morris:

De keten moet een enzym produceren die de gehele keten kopieert, op die manier wordt de keten gedupliceerd en is er sprake van zelfreproductie. Dit kan door middel van het enzym *cop-swi-rpy*. Dit enzym gaat in copy-mode, switched naar de andere kant en gaat dan zoeken naar de eerste volgende pyrimidine. Er is geen pyrimidine (er zijn zelfs geen bases). Dus het enzym zal de gehele keten traverseren met als gevolg dat de gehele keten gekopieerd wordt. De keten die bij dit enzym hoort is CGATTA. Maar de kopie die van de keten is niet de kopie die we willen hebben. Ten eerste is de keten gespiegeld en daarnaast zijn de individuele bases gewisseld met hun complementaire base van het basepaar (A is T geworden, C is G geworden). De kopie die we nu hebben van CGATTA is TAATCG. Een truc die je nu kan toepassen is door deze twee ketens aan elkaar vast te plakken. Hierdoor krijg je een soort Typogenetisch palindroom. Een palindroom is een woord dat het zelfde woord leest als je het achterwaarts leest. Bijvoorbeeld: lepel en koortsmeetsysteemstrook. Een Typogenetisch palindroom werkt hetzelfde als een gewoon palindroom alleen zijn de baseparen van de tweede helft omgedraaid. Een Typogenetisch palindroom bestaat dus ook altijd een even aantal element in tegenstelling tot een gewoon palindroom. Als het enzym de keten CGATTATAATCG traverseert, wordt de complementaire keten dus CGATTATAATCG.

¹² Harold C. Morris. Typogenetics: A logic for artificial life. In Christopher G. Langton, editor, Artificial Life, p 380, Redwood City, California, 1989. Addison-Wesley.

¹³ V. Kvasnička J. Pospíchal T. Kaláb, A Study of Replicators and Hypercycles by Typogenetics, In Advances in Artificial Life, pages 37-54, Springer Berlin / Heidelberg, 2001

¹⁴ Kyubum Wee, Woosuk Lee, A Extensions and Variations on Construction of Autoreplicators in Typogenetics, In Advances in Artificial Life, pages 685-694, Springer Berlin / Heidelberg, 2005

¹⁵ Harold C. Morris. Typogenetics: A logic for artificial life. In Christopher G. Langton, editor, Artificial Life, pages 369–395, Redwood City, California, 1989. Addison-Wesley

¹⁶ Louis Varetto. Typogenetics: An artificial genetic system. Journal of Theoretical Biology, 160:185–205, 1993

¹⁷ Andrew Snare, Typogenetics, Thesis for Bachelor of Science (Computer Science) Honours Clayton Campus, 1999

Het probleem is nu dat niet meer het originele enzym wordt geproduceerd bij de translatie van de keten. De sequentie die achter de originele sequentie geplakt is, heeft invloed op het Typogenetisch eiwit. Dit kan worden opgelost door de stop-functie in het enzym te stoppen. Het is natuurlijk van belang dat onze keten wel nog een Typogenetisch palindroom blijft. Dus: CGATTA-AA-TT-TAATCG. Bij de translatie blijft ontstaat nu de enzymen. *cop-swi-rpy* en *lpu-rpy-swi-cop*. Dat er een tweede enzym ontstaat is niet erg zolang het kopieer enzym maar eerst wordt toegepast.

Het enige waar nu nog voor gezorgd moet worden is dat het enzym zich kan binden aan de eerste letter van de keten, anders wordt niet de gehele keten gekopieerd. Op het moment kan dit niet omdat de bindingsvoorkeur van het enzym A is, en de eerste letter is een C. Dit kan worden opgelost door een aminozuur toe te voegen aan het enzym die wel de bindingsvoorkeur aanpast maar niet de functionaliteit van het aminozuur. Dit kan bijvoorbeeld door de *off* functie toe te voegen aan het enzym. Hierdoor wordt de bindingsvoorkeur C maar blijft het enzym toch nog de zelfde functionaliteit hebben. De nieuwe keten is dus: CGATTA-CT-AATT-AG-TAATCG. De letters CT en AG zijn nodig om het Typogenetisch palindroom in stand te houden.

Het mooie van deze methode is dat elke willekeurige keten van aminozuren in het midden van de keten kan worden geplaatst zolang het maar een

CGATTACTAA ~ R~ TTAGTAATCG

Typogenetisch palindroom is. We hebben nu dus als

het waren een wrapper gecreëerd voor een zelfreproducerende keten. Typogenetics kan zich dus net zoals echt DNA reproducen. Merk op dat de zelfreproductie een emergente eigenschap is. In het programma van het enzym staat nergens de directe opdracht zich zelf te kopiëren.

2.4. Onderzoek met Typogenetics

Zoals al gezegd probeert men in de studie naar Artificial Life beter te begrijpen hoe complexe structuren ontstaan uit simpele structuren. Men onderzoekt dan hoe bijvoorbeeld zelfreproductie kan plaatsvinden door interactie tussen simpele structuren. Een andere fenomeen waar men ook naar zoekt is de hypercyclus. Simpel uitgelegd is een hypercyclus het effect dat simpele structuren elkaars reproductie katalyseren. Waarom deze eigenschappen interessant zijn voor Artificial Life gaat te ver om hier te beschrijven maar het heeft te maken met het feit dat ze het toenemen van complexiteit binnen systeem kunnen verklaren.

Fenomenen zoals zelfreproductie en hypercycli zijn voorbeelden van emergentie. Het probleem met onderzoek naar deze fenomenen is dat emergente eigenschappen moeilijk constructief te benaderen zijn. Dit komt omdat bij constructief benaderen sprake is van een *bottom-up approach* waarbij je bij met niks begint en de eigenschap stap voor stap opbouwt. De fenomenen die we willen onderzoeken, zijn echter voorbeelden van emergente eigenschappen die eigenlijk alleen op holistische manier benaderd kunnen worden. De auteurs van een studie in Typogenetics zeggen:

“It is demonstrated that a construction of autoreplicators and hypercycles belongs to very complicated combinatorial problems and therefore an effort to generate them by a systematic constructive enumeration is hopeless.”¹⁸

2.4.1. Het genetisch algoritme

Een methode om toch emergente eigenschappen te vinden is door gebruik te maken van een genetisch algoritme^{19 20 21}. Het genetisch algoritme is geïnspireerd door de evolutietheorie²². Het genetisch algoritme zoekt een oplossing voor een bepaald probleem door middel een populatie van individuen. Elk van deze individuen representeert een mogelijke oplossing. De oplossing is de genetische code van het individu. Deze individuen kunnen worden geëvolueerd in een simulatie. In de simulatie hebben individuen met een betere oplossing een grotere kans tot overleven dan individuen met een minder goede oplossing. Dit is analoog aan de *survival of the fittest*, waarbij organismes met een gunstigere genetische opmaak een grotere kans hebben tot overleven. Daarnaast kunnen deze individuen ook muteren en kan er cross-over plaats vinden tussen verschillende individuen waardoor er nieuwe individuen ontstaan. Bij mutatie vindt een willekeurige verandering aan de genetische code aan het individu. Bij cross-over wordt de genetische code van twee individuen gecombineerd in een nakomeling. Deze nieuwe individuen kunnen betere maar ook slechtere oplossingen hebben dan de oude individuen. De individuen worden gewaardeerd door middel van de fitnessfunctie die een discrete waarde geeft aan individu.

¹⁸ V. Kvasnička J. Pospíchal T. Kaláb, A Study of Replicators and Hypercycles by Typogenetics, In Advances in Artificial Life, pages 37-54, Springer Berlin / Heidelberg, 2001

¹⁹ V. Kvasnička J. Pospíchal T. Kaláb, A Study of Replicators and Hypercycles by Typogenetics, In Advances in Artificial Life, pages 37-54, Springer Berlin / Heidelberg, 2001

²⁰ Kyubum Wee, Woosuk Lee, A Extensions and Variations on Construction of Autoreplicators in Typogenetics, In Advances in Artificial Life, pages 685-694, Springer Berlin / Heidelberg, 2005

²¹ Chohwa Gwak Kyubum Wee , Construction of Hypercycles in Typogenetics with Evolutionary Algorithms , In Advances in Artificial Life, pages 1060-1068, Springer Berlin / Heidelberg, 2007

²² Gary William Flake, The Computational Beauty of Nature, 1998 p339 t/m p360

Hier een pseudo-code van het genetisch algoritme:

```
Initialiseer een populatie van willekeurige individuen ter grote van N  
Herhaal tot de beste individu voldoet aan de gestelde criteria: {  
    Evalueer alle individuen van de van de populatie door middel van de fitness functie  
    Herhaal N keer:{  
        Doe een van onderstaande:  
            -Kies een individu naar zijn fitnesswaarde, plaats deze in de nieuwe  
            populatie.  
            - Kies een individu naar zijn fitnesswaarde, laat mutatie plaats vinden  
            plaats deze in de nieuwe populatie.  
            -Kies twee individuen naar zijn fitnesswaarde,cross-over tussen de  
            individuen, plaats nakomeling in nieuwe populatie  
        }  
    Vervang oude populatie door nieuwe populatie  
    }
```

Het voordeel van het genetisch algoritme is dat er parallel naar verschillende oplossingen gezocht kan worden. Elke individu representeert een mogelijke oplossing en doormiddel simulatie blijven vanzelf de meest geschikte oplossingen over. Er is wel een aantal beperkingen. Ten eerste is het niet altijd mogelijk een oplossing te representeren in een genetische code waarop operaties als mutatie en cross-over op kunnen worden uitgevoerd. Daarnaast kan het ook moeilijk om een goede fitnessfunctie te construeren omdat het niet altijd mogelijk is een discrete waarde toe te kennen. Een ander nadeel van het genetisch algoritme is dat de oplossing die ermee gevonden worden niet echt begrijpbaar zijn. In de oplossing is vaak geen duidelijke structuur te herkennen dit komt natuurlijk omdat ze door een willekeurig proces zijn verkregen. De oplossingen kunnen daarom ook maar beperkt helpen bij het beter begrijpen van het probleem.

Typogenetics is een ideale kandidaat voor het genetisch algoritme. Als het genetisch algoritme wordt toegepast op Typogenetics wordt een individu gerepresenteerd door een willekeurige keten. Op deze ketens kunnen moeiteloos de mutatie en cross-over operatie op worden uitgevoerd. De vraag of een fitnessfunctie mogelijk is kan niet definitief beantwoorden omdat deze fitnessfunctie afhangt van te onderzoeken fenomeen. Een voorbeeld van een werkbare fitnessfunctie voor het fenomeen zelfreproductie toegepast op Typogenetics:

$$Fitness(S) = \frac{1}{2}(2 - d(S, R') - d(S', R))^{23}$$

Waarbij $R = \underline{S}$, de gecomplementeerde keten van S. En replicatie(S)=R' en replicatie(R)=S'. $d(x,y)$ is de mate van afwijking tussen x en y gerepresenteerd met een getal tussen 0 en 1. De fitness waarde ligt dus tussen 0 en 1. Bij een perfecte replicator is $d(S, R')=1$ en $d(S', R)=1$. Hieruit volgt dat de fitnesswaarde van een perfecte replicator één is.

2.4.2. Determinisme

Voor met Typogenetics moet een aanpassing worden gemaakt aan de versie van Hofstadter. Zoals Hofstadter namelijk Typogenetics heeft gemodelleerd is het niet geheel deterministisch. Determinisme binnen een formeel systeem betekent dat alle actie volgens vaste regels verlopen. Dit betekent dat het systeem dat bij dezelfde input altijd dezelfde output volgt. Het feit dat Typogenetics deterministisch is komt doordat hij sommige onderdelen van het systeem bewust arbitrair heeft gelaten. Tevens is het zo dat Hofstadter het systeem vrij informeel beschrijft waardoor niet alle grensgevallen worden behandeld. Een voorbeeld van een arbitraire eigenschap is het aangrijpingspunt van het enzym. Een enzym heeft een voorkeur voor een bepaalde base. Als er meerdere exemplaren van deze base zich in de keten bevinden, wat praktisch altijd het geval is, dan kan het enzym zich op verschillende plaatsen binden. Wat als gevolg heeft dat er meerdere uitkomsten mogelijk zijn. Een voorbeeld van adeterminisme, wat veroorzaakt is door het niet beschrijven van een grensgeval door Hofstadter, is de volgorde van enzymen. Bij het aflezen van een keten kan het zo zijn dat er meerdere enzymen (zie: AA stopfunctie) worden gevormd. Hofstadter heeft niet gedefinieerd in welke volgorde deze enzymen dan worden toegepast op de keten. Het gevolg hiervan is dus dat het systeem niet deterministisch is. Bij de beschreven onderzoeken kiest men er altijd voor om Typogenetics volledig deterministisch te maken²⁴. De reden hiervoor is simpelweg dat de gevonden resultaten consequent moeten zijn. Met andere woorden je wilt dat een bepaalde keten altijd het zelfde resultaat geeft. Vanuit een computationaliteitsstandpunt is het ook zo dat adeterministische systemen kunnen worden gesimuleerd door middel van deterministische systemen. Dus hier uit volgt dat eigenschappen gevonden in het deterministische systeem ook gelden voor het adeterministische systeem²⁵

²³ V. Kvasnička J. Pospíchal T. Kaláb, A Study of Replicators and Hypercycles by Typogenetics, In Advances in Artificial Life, pages 37-54, Springer Berlin / Heidelberg, 2001

²⁴ Het volledig deterministisch maken is voor het eerst gedaan door Louis Varetto. Typogenetics: An artificial genetic system. Journal of Theoretical Biology, 160:185–205, 1993 en daarna door alle ander onderzoekers overgenomen.

²⁵ Christos Papadimitriou, Computational Complexity, 1994 p45 t/m p51

3. Verschillen tussen genexpressie en Typogenetics

Zoals al gezegd heeft Hofstadter veel onderdelen van genexpressie weggelaten in zijn abstractie. Abstraheren hoeft niet slecht te zijn als de eigenschappen die je interessant vindt maar behouden blijven. Abstraheren gebeurt door middel van onderdelen weg te laten die niet essentieel zijn voor het systeem. Het is de kunst om precies die onderdelen te behouden die zorgen voor de eigenschappen die je wilt abstraheren. Bij Typogenetics heeft Hofstadter een aantal onderdelen weggelaten waardoor er eigenschappen verloren zijn gegaan. Het wegvallen van deze eigenschappen heeft als gevolg dat Typogenetics als geheel minder uitdrukingskracht heeft dan genexpressie.

Er zijn drie belangrijke onderdelen die Hofstadter weggelaten heeft bij het abstractieproces. Ten eerste heeft Hofstadter de het proces van DNA → RNA → eiwit helemaal weg gemodelleerd tot een enkele stap. Bij Typogenetics gaat het van DNA meteen naar eiwit. Het proces van DNA → RNA → eiwit is een zeer complex proces. Hier gebeurt veel regulatie. Dit is niet te modelleren binnen Typogenetics.

Ten tweede is Genexpressie in de werkelijkheid veel dynamischer dan Typogenetics. Bij Typogenetics is er sprake van een serieel proces. Hiermee bedoelen we dat de operaties een voor een verlopen. In tegenstelling tot genexpressie waarbij er sprake van parallelle processen. In de werkelijkheid is het niet zo dat de processen een voor een verlopen. De cel kan vergeleken worden met een groot chemische vat waar vele duizenden operaties per seconde plaats vinden die allemaal met elkaar samenhangen en elkaar beïnvloeden.

Ten derde is buiten de aminozuursequentie ook de ruimtelijke structuur van belang bij genexpressie. De manier waarop Hofstadter de structuur van een eiwit heeft gemodelleerd in zijn model is erg versimpeld en geeft niet de essentie van eiwit weer. Ook de structuur van het DNA zelf is buiten beschouwing gelaten, wat invloed heeft op de genexpressie. Een gevolg van het weglaten van de ruimtelijke structuur van eiwitten bij Typogenetics, is dat er geen rekening wordt gehouden met andere structurele eigenschappen dan de bindingsvoorkeur. In Typogenetics bepaalt de structuur namelijk de bindingsvoorkeur van een eiwit. In de werkelijkheid bepaalt de structuur ook de werking van het eiwit. Verandering van ruimtelijke structuur van het eiwit verandert ook de werking van het eiwit.

Deze drie verschillen zorgen ervoor dat in Typogenetics een aantal eigenschappen verloren zijn gegaan die ook vanuit CKI standpunt interessant zijn die er dus voor zorgen dat Typogenetics minder uitdrukingskracht heeft.

Ten eerste is Typogenetics een contextvrij model. Wat betekent dat een bepaalde keten altijd, in welke situatie dan ook dezelfde uitkomst zal hebben. In werkelijkheid heeft DNA een aantal contextafhankelijke eigenschappen. Dit betekent dus dat een bepaalde keten niet altijd het dezelfde uitkomst heeft, afhankelijk van de context waarin het zich bevindt. Dit betekent dus bijvoorbeeld dat een bepaalde keten een andere eiwit kan produceren afhankelijk van de context. Contextafhankelijk is een interessante eigenschap omdat een primitieve vorm is van geheugen. Een keten 'weet' in welke context het zich bevindt want afhankelijk van deze context heeft het verschillend gedrag.

Daarnaast kent genexpressie een bepaalde vorm van adeterminisme. Bij bepaalde processen is er sprake van een kansfactor. In tegenstelling tot Typogenetics wat geheel deterministisch is. Althans, bij onderzoek door middel van het systeem. Bij Typogenetics zullen processen altijd op te zelfde manier plaatsvinden. Doordat er in een cel vele duizenden processen per seconde plaats vinden is het verloop van genexpressie wel voorspelbaar. Dit is vergelijkbaar met het gooien van een munt. Bij één keer gooien is de uitkomst niet te voorspellen, maar bij duizenden keren gooien is de uitkomst vrij zeker de helft kop en de helft munt. Het is mogelijk om een kansfactor in te bouwen in Typogenetics maar daardoor zullen de resultaten van Typogenetics alleen maar inconsequent worden. Bij een kansfactor is het van belang dat een bepaald proces vele malen herhaald wordt. Dit is niet praktisch uit te voeren in Typogenetics vanwege zijn seriële karakter. Je kan de vraag stellen of het erg is dat Typogenetics we deterministisch is want zoals al eerder gezegd zijn deterministische systemen equivalent aan adeterministische systemen. (adeterminisme is te modelleren in deterministische systemen) Het antwoord is dat adeterministische systemen in een aantal gevallen veel praktischer zijn. Waarom dit zo is, is moeilijk om hier uit te leggen maar Papadimitriou zegt:

“Nondeterminism is a central concept in complexity theory because of its affinity not so much with computation itself, but with the application of computation, most notably logic, combinatorial optimization and artificial intelligence”²⁶

²⁶ Christos Papadimitriou, Computational Complexity, 1994p49

Als laatste is het zo dat het effect van de configuratie van een keten op de werking ervan veel meer holistisch is dan Typogenetics. Dit soort holistische effecten geeft een grotere kans op emergente eigenschappen en dit zijn de verschijnselen die wetenschappers willen onderzoeken.

We zullen nu een aantal voorbeelden geven die het verschil tussen de uitdrukingskracht van Typogenetics en het werkelijke systeem van genexpressie. Er zijn vele voorbeelden mogelijk, we hebben hier gekozen om uit ieder onderdeel van het centrale dogma een proces te nemen die zorgt voor de discrepantie.

3.1. DNA: Consensussequentie

Bij de initiatie van transcriptie spelen meerdere factoren een rol. Een sleutelrol is weggelegd voor de *promoter*. De promoter is een sequentie die upstream (vlak vóór) het gen gelegen is en dient als een regulerende factor. Zonder promoter kan RNA polymerase niet binden en kan het DNA niet worden afgeschreven.²⁷

Figuur 4

RNA polymerase kan alleen binden aan de promoter als de sequentie voldoende herkenbaar is voor het DNA-bindende deel van RNA polymerase. De promotersequentie is echter niet altijd hetzelfde. Als RNA polymerase bijvoorbeeld de meeste

Promoter	-35 box
Consensus	5'-TTGACA-3'
Lactose operon	5'-TTTACA-3'
Tryptophan operon	5'-TTGACA-3'

bindingsaffiniteit heeft met de sequentie TTTACG, zou het goed kunnen zijn dat de promoter van een bepaald gen de sequentie TTAACG heeft. De gemiddelde van alle verschillende promotersequenties wordt een *consensussequentie* genoemd. De variëteit van de promotersequentie heeft grote invloed op de effectiviteit van de gentranscriptie. Effectiviteit wordt in het geval van transcriptie gedefinieerd als het aantal afschrijvingen per seconde. De promotersequentie van een gen ligt vast, en is dus niet reguleerbaar. Dat betekent dat in principe het aantal initiaties per seconde vastligt per gen. Het DNA-bindende deel van RNA polymerase, echter is wel veranderlijk en reageert op invloeden van buitenaf.

In bacteriën bindt het σ -subunit aan de promoter, waarna gentranscriptie begint. In normale omstandigheden is dit σ^{70} , die ook wel de 'housekeeping subunit' wordt genoemd. De RNA

²⁷ Geiduschek, E.P., Kassavetis, G.A. (2001), *The RNA polymerase III transcription apparatus*, J. Mol. Biology vol. 310: p. 1-26

polymerase schrijft genen, die hoge affiniteit hebben met σ^{70} , in grote hoeveelheden af. Genen met promoters die minder affiniteit met σ^{70} hebben, worden dus ook minder effectief afgeschreven. Maar wanneer de bacterie in een omgeving komt waar het organisme hogere temperaturen moet weerstaan, wordt een RNA polymerase gemaakt met een andere σ -subunit. σ^{32} , in dit geval, kan weer beter binden aan promotersequenties van andere genen, die de bacterie helpen de stress te doorstaan.²⁸

Bij humane genen geldt ongeveer hetzelfde: het transcriptie-initiatiecomplex bestaat uit tientallen eiwitten die RNA polymerase uiteindelijk het gen laten afschrijven. Sommige van deze eiwitten, transcriptiefactoren geheten, zijn gevoeliger voor bepaalde promotersequenties dan andere. Zo kan de cel de afschrijving van genen reguleren.²⁹

Weggevalen eigenschappen:

De consensussequentie is een voorbeeld van het adeterminisme van DNA. Doordat de ruimtelijke structuur die niet gemodelleerd is in Typogenetics, ervoor zorgt dat een keten niet altijd wordt afgelezen. Het principe van een consensussequentie zorgt er dus voor dat er sprake is van een kansfactor bij het aflezen van een DNA keten. In tegenstelling tot Typogenetics waarin een keten altijd zal worden afgelezen. Het zou mogelijk zijn om een kansfactor te modelleren in Typogenetics maar dan zoals al eerder gezegd wordt Typogenetics minder bruikbaar. Een interessant gegeven is hier dat er informatie is opgeslagen in de ruimtelijke structuur van het DNA. Namelijk in welke mate een bepaalde stuk DNA moet worden afgelezen. Informatie wordt dus niet alleen op het niveau van basevolgorde opgeslagen zoals bij Typogenetics.

Daarnaast is het $\sigma^{70} / \sigma^{32}$ complex een voorbeeld van contextafhankelijkheid. Afhankelijk van de context, de cel bevindt zich in een omgeving met een hogere temperatuur, wordt het DNA anders afgeschreven. Bij Typogenetics is dit om verschillende redenen niet gemodelleerd. Ten eerste vindt er bij Typogenetics geen interactie plaats met de omgeving. Daarnaast is het binnen Typogenetics niet mogelijk om dezelfde keten op verschillende manieren af te lezen.

²⁸ Gruber TM, Gross CA. (2003). *Multiple sigma subunits and the partitioning of bacterial transcription space*. Annu Rev Microbiol. Vol 57: 441-66

²⁹ Brown, T.A., *Genomes 3*, Garland Sciences Publishing (2007): p. 354-362

3.2. RNA: Intronen

Wanneer een gen is afgeschreven, dat wil zeggen wanneer een bepaald stuk dubbelstrengs DNA is omgezet in enkelstrengs RNA (pre-mRNA), worden zoals eerder gezegd enkele *post-translational modifications* uitgevoerd in de kern van de cel. Één van de belangrijkste aanpassingen die het RNA ondergaat is het verwijderen van intronen. Een intron is een stuk RNA dat wel wordt afgeschreven en vertaald naar RNA, maar niet de celkern verlaat om uiteindelijk te worden omgezet in een eiwit. Wanneer het RNA, nadat alle PTM's zijn uitgevoerd zijn structuur heeft gekregen (het heet dan mRNA), is de sequentie dus korter dan de DNA-template was. De stukken RNA die worden omgezet in een eiwit heten exonen.³⁰

Om deze aanpassing uit te voeren beschikt de cel over een geavanceerd *splicing apparatus*, (het spliceosoom) die de RNA sequentie splitst, een intron eruit knipt en later weer de exonen aan elkaar vastmaakt. De eiwitten die verantwoordelijk zijn voor de chemische reacties die noodzakelijk zijn heten snRNP's en herkennen een bepaalde sequentie in de intronen. Eenmaal gebonden brengen deze snRNP's de uiteindes van de intronen (nu nog een deel van de RNA sequentie) dicht bij elkaar. Als de afstand klein genoeg is hydrolyseren ('knippen') de snRNP's in het RNA, zowel aan het uiteinde van exon 1 als aan het begin van exon 2. Hierdoor komt het intron vrij en worden de uiteindes van de exonen weer aan elkaar vast gemaakt.³¹

Net als bij de promotorsequenties eerder genoemd, is de basenvolgorde van het RNA die herkend wordt door de snRNP's een consensussequentie. SnRNP's binden niet altijd aan de sequentie. Als gevolg hiervan worden sommige exonen overgeslagen in het 'knipproces'. Op deze manier kan een bepaalde DNA sequentie voor verschillende eiwitten coderen. Dit fenomeen wordt *alternative splicing* genoemd. In figuur 5 is dit schematisch weergegeven: de genummerde donkerblauwe delen zijn intronen die wel of niet uit de sequentie worden geknipt.³²

Figuur 5

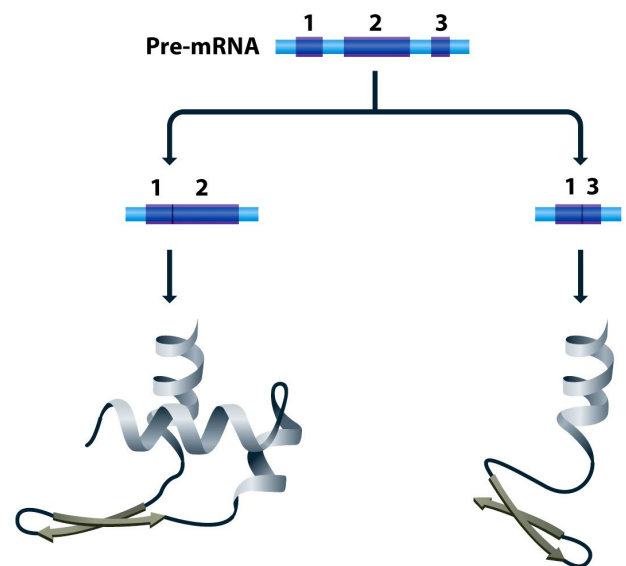


Figure 6-5 Genomes 3 (© Garland Science 2007)

³⁰ Gilbert, Walter (1978). *Why genes in pieces*. Nature vol. **271**: p. 501

³¹ Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72: 291-336

³² Brown, T.A., *Genomes 3*, Garland Sciences Publishing (2007): p. 354-362

Of een intron er wel of niet wordt uitgeknipt heeft dus voor een belangrijk deel met de het concept consensussequentie te maken. Er is echter een groot aantal eiwitten die splicing kunnen bevorderen of onderdrukken. Deze SR eiwitten (*stimulation/repression*) hebben invloed op de constructie van het spliceosoom, en bepalen zo welke variant van het gen in de cel op een willekeurige tijd overheersen. Deze samenstelling van deze SR eiwitten kan verschillen.³³

Weggevalen eigenschappen:

Hier is er weer sprake consensussequentie. Er is dus een kansfactor dus er is sprake van adeterminisme. Alleen is het hier niet de kans of de keten wordt afgelezen maar welke vorm het eiwit krijgt. Het splicing apparaat zorgt ervoor dat een dezelfde DNA keten uiteindelijk verschillende eiwitten kan vormen. Een DNA keten kan ambigu zijn, een keten kan meerdere eiwitten representeren. Dit is niet het geval bij Typogenetics.

Daarnaast zijn de SR-eiwitten een voorbeeld van context afhankelijkheid, de samenstelling van SR-eiwitten in de cel is afhankelijk van de omgeving waarin de cel zich bevindt. De eiwitten die zich in de omgeving van de mRNA keten bevinden hebben invloed op de uiteindelijk vorm die het eiwit krijgt. Bij Typogenetics hebben eerder gevormde eiwitten geen invloed op de manier waarop een eiwit wordt gevormd, het is contextvrij.

3.3. Eiwit: Eiwitvouwing

Nadat de mRNA keten met behulp van bijbehorende tRNA's is omgezet tot een aminozuursequentie, wordt de lange keten van aminozuren direct gevouwen tot een bepaalde structuur. Het proces van eiwitvouwing wordt voornamelijk door twee elementen beïnvloed³⁴: interactie met aminozuren verderop in de keten, en interactie met water. Zoals eerder beschreven bestaan er twee hoofdklassen aminozuren: hydrofobe en hydrofiele. De hydrofiele aminozuren zullen waterstofbruggen kunnen maken met het water in de omgeving. Deze aminozuren ondervinden daardoor geen drijvende kracht om een andere conformatie aan te nemen. De hydrofobe echter, hebben de neiging om samen te pakken om zo min mogelijk contact te maken met de watermoleculen in de omgeving. Bovendien zijn hydrofobe aminozuren in staat om met andere hydrofobe aminozuren zwakke chemische bindingen aan te gaan waardoor het

³³ Ritter, P. (1995) *Biochemistry: A Foundation*, p. 710

³⁴ Dinner, A.R., *et al*, *Understanding protein folding via free-energy surfaces from theory and experiment*, Trends Biochem. Sci. 2000, vol. 25, p. 331-339

‘samengeklonterde’ geheel nog stabiel wordt. Door de energetisch meest gunstige conformatie aan te nemen wordt het eiwit ‘gevouwen’ tot een driedimensionale structuur.

Deze structuur is essentieel voor het functioneren van het eiwit. De functie van een eiwit wordt namelijk bepaald door de domeinen die het bezit. Verschillende domeinen (bijvoorbeeld een DNA-bindend domein, zoals we zagen bij RNA polymerase) en hoe deze georganiseerd zijn in de driedimensionale ruimte bepalen de functie. De meeste eiwitten hebben een katalyserende werking. Dat wil zeggen, ze helpen chemische reacties plaats te vinden in de cel. Er zijn vele structurele eigenschappen die deze reactie in gang brengen, bijvoorbeeld het dicht bij elkaar houden van substraten die gekatalyseerd moeten worden. Het is dan erg belangrijk dat in de driedimensionale structuur van het eiwit twee *substrate binding-domains* te vinden zijn op een afzienbare afstand tot elkaar.

Wanneer een aminozuurketen in een reageerbuis wordt gebracht, ontstaat bij de meeste eiwitten meteen de *native structure*, de structuur zoals deze zich in de levende cel bevindt. Dit komt door de zojuist beschreven chemische interacties tussen eigen aminozuren en de waterige omgeving en de energetisch meest gunstige staat waarin het eiwit wordt gebracht. Toch is het in de biologie nog altijd bijna onmogelijk om een driedimensionale structuur te voorspellen aan de hand van een aminozuursequentie. Bepaalde domeinen zijn meestal wel te voorspellen (het motief Cys – X – X – Cys – 12 maal X – His – X – X – His bijvoorbeeld, heeft altijd een DNA-bindende werking,³⁵ maar een totale predictie is nog niet mogelijk. In de moleculaire biologie is *Structure Prediction* een belangrijk concept. Heden ten dagen worden allerlei computerprogramma’s ingezet die met behulp van gegeven variabelen zoals het hydrofobe effect, interacties met eigen aminozuren en zuurgraad van het water proberen een driedimensionale structuur te geven van een aminozuurketen. Deze voorspellingen zijn zoals gezegd echter lang niet accuraat. Wanneer men in staat zou zijn alle chemische interacties en drijvende krachten volledig te kunnen beschrijven bestaat in theorie de mogelijkheid dat we in staat zijn een honderd procent accurate voorspelling van de tertiaire structuur kunnen doen.

Er zijn echter ook eiwitten in de cel die per definitie niet te voorspellen zijn. Dit zijn eiwitten met één of meerdere ‘chameleon sequences’. Een kameleon sequentie is een korte keten (5-7 aminozuren lang) in het eiwit die veranderlijk is, bijvoorbeeld wanneer het eiwit in een andere chemische omgeving komt, of wanneer het bindt aan een ander eiwit. In de loop der jaren is het aantal

³⁵ Aitken, A., *Protein Consensus Sequence Motifs*, Mol. Biotechnol. 1999, vol. 12, p. 241-253

ontdekte kameleon sequenties gestaag gestegen.³⁶ Een typische kameleon sequentie is een keten van 5 aminozuren lang die kan *switchen* tussen een α -helix en een β -sheet. Onder invloed van de chemische omgeving, maar ook van de tertiaire structuur van het gehele eiwit kan de conformatie veranderen. In figuur ___ is dit schematisch weergegeven: de gele α -helix links kan onder bepaalde omstandigheden een β -sheet worden. De functie van het eiwit verandert hierdoor meestal ook: een voorbeeld is het DNA-bindende eiwit in gist, MAT α 2, dat bepaalt welk *mating type* de gist is (mannelijk of vrouwelijk). Dit is namelijk afhankelijk van het soort genen dat wordt afgeschreven. Door een andere conformatie aan te nemen, α -helix of β -sheet, kan het DNA bindende eiwit MAT α 2 sommige genen blokkeren. Het gevolg is een ander eiwitpatroon in de cel (genen coderen immers voor eiwitten) en zo ook een ander *mating type*.

Weggevalen eigenschappen:

Het eiwitvouwen is een voorbeeld van holisme in DNA dat niet aanwezig is in Typogenetics. In Typogenetics is de structuur reductionair te verklaren. Elk aminozuur geeft een bepaalde richting aan de structuur. Al deze richtingen op geteld geeft de uiteindelijke structuur. In de werkelijkheid is er bij de vorming van het eiwit sprake van een keten die om zichzelf heen vouwt. Het gevolg is dat de structuur niet meer reductionair te verklaren is zoals bij Typogenetics, de uiteindelijke vorm kan verklaard door de afzonderlijke aminozuren. Dit betekent dus dat als sequentie A zorgt voor structuur A' en sequentie B voor structuur B' het niet zo is dat sequentie AB zorgt voor structuur A'B'. De sequentie van A en van B beïnvloeden elkaar waardoor er een nieuwe structuur ontstaat. Het gevolg is dat een kleine aanpassing in de sequentie grote gevolgen kan hebben voor de uiteindelijke structuur.

De chameleon sequentie is weer een goed voorbeeld van context afhankelijkheid. Onder invloed van de chemische omgeving of eiwitten in de omgeving zal de RNA keten verschillende eiwitten vormen.

3.4 Ter conclusie

Hofstadter heeft met behulp van een aantal simpele regels een formeel systeem ontworpen dat in de basis een redelijk adequate weergave is van het fenomeen genexpressie. Namelijk, de genetische code staat aan de basis van de synthese van een eiwit. In Typogenetics gaat de weg van DNA naar eiwit in één stap. In de cel gaat het in veel meer stappen, één van de voorbeelden is dat er tussen en DNA en eiwit nog de mRNA keten zit. Het verdwijnen van deze kenmerken is in beginsel geen

³⁶ Mezei, M., *Chameleon Sequences in the PDB*, Protein Eng. 1998, vol. 11, p. 411-414

probleem. Als de genetische code TTAATGAAA een eiwit met de sequentie Leu-Met-Lys geeft, waarom is het dan belangrijk om te weten dat er in het proces van genexpressie nog een stap tussen zit (van TTAATGAAA naar AAUUACUUU en dan pas naar Leu-Met-Lys)? Maar voor onderzoek in Artificial Life is dit niet voldoende, omdat men hier geïnteresseerd is in de manier waarop het gebeurt en niet zozeer naar de input en output. Het versimpelen van het centrale dogma (het overslaan van mRNA) heeft als gevolg dat de manier waarop een eiwit wordt gevormd voor een groot deel is weggefallen.

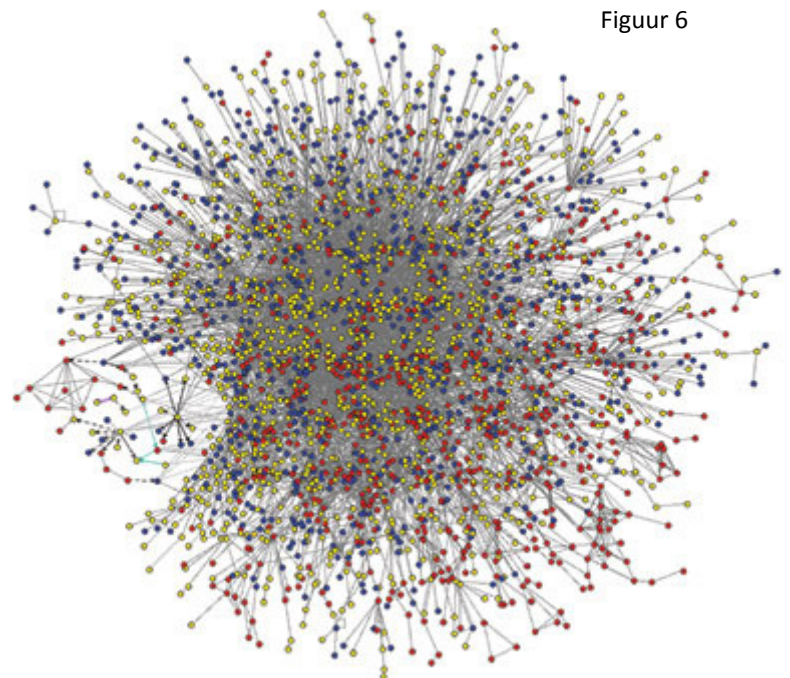
Concluderend is genexpressie veel meer een complex systeem die te omschrijven is als een evenwichtssituatie waarbij parallelle processen plaatsvinden en niet enkel in seriële processen is te definiëren. In de weg van DNA naar eiwit zijn dus enkele voor Artificial Life interessante eigenschappen weggelaten.

4. Abstractie

Bij het construeren van een formeel systeem, maak je zoals gezegd gebruik van abstractie. Dit om het systeem begrijpelijker en overzichtelijker te maken en zodat het toepasbaar wordt. Reductionisme en holisme zijn van belang bij dit abstractieproces. Wanneer in de wetenschap fenomenen bestudeerd worden, komen vaak de termen reductionisme en holisme naar voren. Strikt genomen verwijzen deze termen naar de vraag of de eigenschappen van een systeem respectievelijk wel of niet voorspelbaar zijn op grond van een beschrijving van de afzonderlijke componenten. Wij zullen de termen reductionisme gebruiken wanneer we spreken over het systeem opdelen in afzonderlijke onderdelen. Soms schiet reductionisme te kort, omdat bepaalde eigenschappen niet aan de afzonderlijke onderdelen toe te schrijven zijn. Dan vereist het een holistische aanpak.

4.1. Holisme en reductionisme

Bij het verklaren van fenomenen wordt in de biologie veel gebruik gemaakt van reductionisme. Het reduceren van verklaringen voor deze fenomenen tot de kleinst mogelijke eenheden brengt is een algemeen geaccepteerde methode voor het ontwikkelen van een verklarende theorie in de biologie.³⁷ Zo wordt met zo min mogelijk elementen zoveel mogelijk verklaard of beschreven. Biologische verschijnselen als voorbeeld, zijn vaak enorm complex. Wanneer deze verschijnselen verklaard moeten worden door een theorie, wordt de voorkeur gegeven aan een holistische aanpak,³⁸ omdat de verklarende werking van die theorie vaak afneemt bij het weglaten van belangrijke elementen. Een voorbeeld uit de biologie is het onderzoeken van eiwit-eiwit interacties *in vitro*. In een dergelijk experiment worden de eiwitten uit de cel gehaald en in een reageerbuis worden de interacties onderzocht. Een groot voordeel van dit experiment is dat het goed



Figuur 6

³⁷ Reductionism in Biology First published Tue May 27, 2008

³⁸ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1299179/>

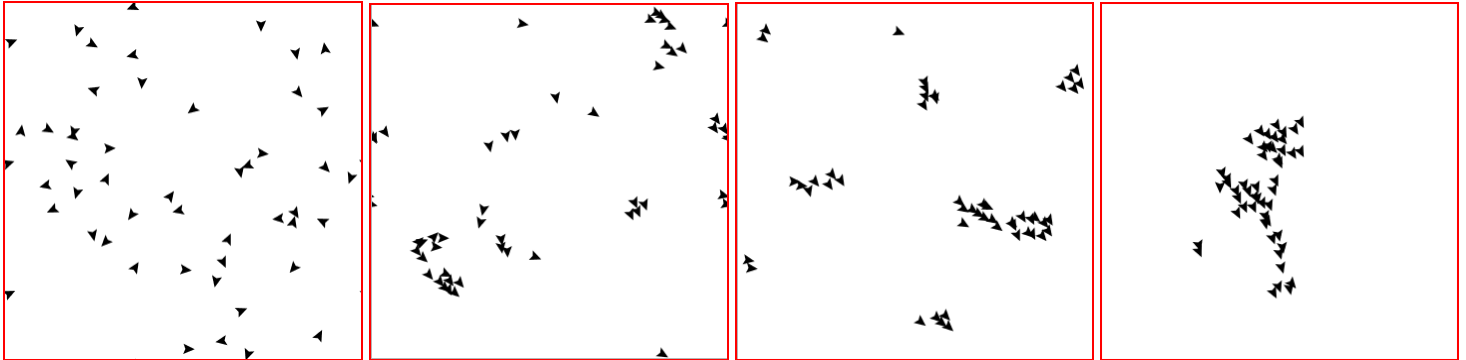
mogelijk is om de variabelen (concentraties van eiwit, zuurgraad, temperatuur etc.) te controleren. Op die manier kan verfijnd worden onderzocht hoe de eiwitten elkaar beïnvloeden of interacties aangaan. In de hedendaagse biologie worden alle bekende interacties bij elkaar gevoegd om zo een beeld te krijgen van hoe de eiwitten in de cel werken en welke processen er gaande zijn. Figuur 6 geeft een voorbeeld van een dergelijk *interaction web*. Vanuit een reductionistisch standpunt zijn de afzonderlijke interacties van de eiwitten genoeg om het geheel te verklaren. Het nadeel van zulke experimenten is echter dat de te onderzoeken interacties los staan van de omgeving waar die interacties onder normale omstandigheden plaatsvinden: de cel. Als er een interactie is aangetoond is het nog niet zeker te zeggen of dit ook in de cel gebeurt. Als een bepaald proces in de cel bijvoorbeeld van drie eiwitten afhankelijk is, en alle interacties van die drie eiwitten zijn bekend, kan je met de bekende variabelen dus voorspellen en simuleren wat er gaat gebeuren. Omdat de cel echter zo complex is, en er ook nog factoren zijn die de eiwitten zelf beïnvloeden (pH bijvoorbeeld), is het vaak toch moeilijk een dergelijke simulatie te maken.³⁹ Reductionisme werkt in de moleculaire biologie vooral erg goed op zeer kleine schaal (het bepalen van de structuur van een DNA-bindend domein in een eiwit bijvoorbeeld), omdat het aantal variabelen dan is te overzien.

Het totale beeld van het centrale dogma van de moleculaire biologie is er nog niet, er worden steeds nieuwe ontdekkingen gedaan. Het is onmogelijk om alle variabelen van de cel (alle miljarden eiwitten en andere moleculen) mee te nemen wanneer je op zoek bent naar een beschrijving van genexpressie. Het doel is wel het vinden van een alomvattende beschrijving van genexpressie en alle processen die daar omheen plaatsvinden. Reductionisme is daarbij zoals gezegd onvermijdelijk. Hieruit kunnen we het volgende concluderen: bij een abstractie probeer je zoveel mogelijk onderdelen mee te nemen zodat bijvoorbeeld emergente eigenschappen niet verloren gaan. Dit conflicteert met een ander criterium voor een goede abstractie: het systeem moet zo simpel mogelijk blijven (weinig regels in het formele systeem), zodat transparantie en overzichtelijkheid gerealiseerd kan worden. Dit is dus een dilemma: aan de ene kant wil je het systeem zo uitgebreid mogelijk maken, aan de andere kant zo simpel mogelijk. Hoe kan je, wanneer je gaat abstraheren, aan beide criteria voldoen? Het volgende formele systeem is een voorbeeld van een goede abstractie.

³⁹ Kellenberger, E., The evolution of molecular biology, 2004, EMBO reports, vol. 5 no. 6

4.2 Het samen zwermen van vogels.

Figuur 7



Het samen zwermen van vogels is een interessant fenomeen van uit Artificial Life omdat er sprake is van zelf organisatie. Een zwerm volgens bestaat uit een groep individuen die samen één entiteit vormen. Dit gebeurt zonder dat de groep aangestuurd wordt door een individu en zonder onderlinge communicatie. Aan het eind van de jaren tachtig bedacht Craig Reynolds een systeem waarin hij het zwermgedrag van vogels heeft gemodelleerd^{40 41}. In figuur 7 is een simulatie van Reynolds' model gegeven⁴². In de eerste stap heeft elk individu een eigen richting. Naarmate de simulatie voortduurt clusteren de individuen samen. Het gedrag van de individuen wordt beschreven door slechts drie regels:

1. Ontwijken: beweeg weg van individuen die te dicht bij zijn. (dit om botsingen te voorkomen)
2. Kopieer: pas eigen snelheid en richting aan, aan de snelheid en richting van de andere individuen die in de buurt zijn.
3. Centreer: beweeg naar het centrum van de groep (het centrum is een gunstiger plaats omdat men door soort genoten beschermt wordt van predatoren etc.).

Het interessante van deze regels is dat het samen zwermen nergens direct staat beschreven. Het samen zwermen is dus een emergente eigenschap. Daarnaast dingen als de richting van de zwerm niet gedefinieerd in de regels. Het is dus een goede abstractie: het systeem bevat slechts drie regels en toch wordt het complexe gedrag verrassend goed gemodelleerd.

⁴⁰ Gary William Flake, *The Computational Beauty of Nature*, 1998. p270 t/m p275

⁴¹ Wilensky, U. (1999). NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL

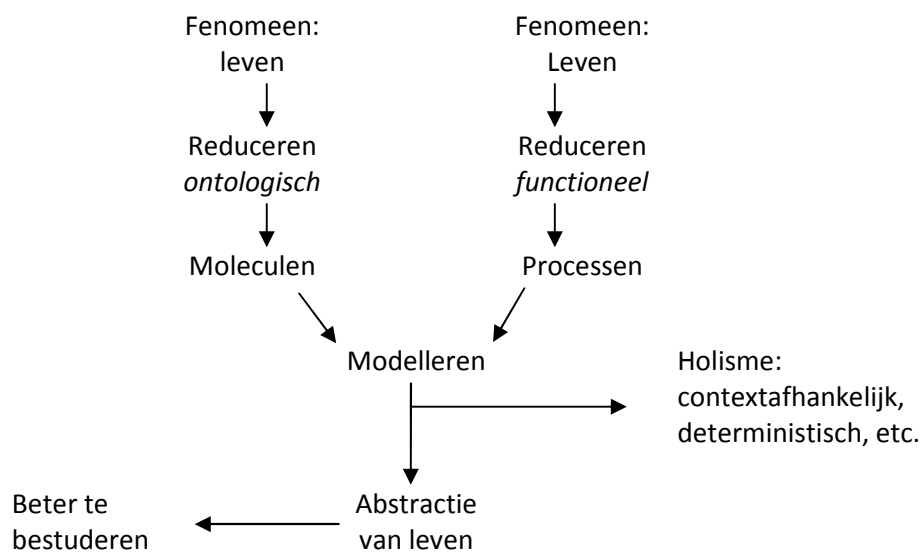
⁴² Wilensky, U. (1998). NetLogo Flocking model. <http://ccl.northwestern.edu/netlogo/models/Flocking>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

5. Conclusie

“Everything should be made as simple as possible, but no simpler” ~Albert Einstein

Bij het abstractieproces van een complex systeem als genexpressie, zijn er hoofdzakelijk twee concepten die je wil abstraheren. Aan de ene kant is er het ontologische concept. Het natuurlijke systeem reduceer je tot de afzonderlijke onderdelen. In het geval van genexpressie reduceer je dus de cel tot de afzonderlijke moleculen (DNA, RNA en eiwit), waardoor je precies kan beschrijven uit welke onderdelen de cel bestaat. Aan de andere kant wil je, wanneer je een dergelijk systeem abstraheert ook de functionaliteit uiteindelijk modelleren. Bij deze vorm van reductie is het belangrijk dat de functie van de onderdelen centraal staat. Als voorbeeld kan een mRNA keten genomen worden: bij reductie op een ontologische manier beschrijf je mRNA als een keten bestaande uit aan elkaar gekoppelde nucleïnezuren, terwijl wanneer je reduceert op functioneel niveau je mRNA beschrijft als de *messenger* die tussen DNA en eiwit in zit.

Wanneer deze twee concepten samenbrengt om tot het modelleren te komen, moeten er keuzes worden gemaakt. Want bij het bekijken van alle onderdelen blijkt het bijvoorbeeld contextafhankelijk te zijn of adeterministisch. Hoe meer je op een holistische manier kijkt, hoe meer onderdelen je mee moet nemen. Maar de abstractie voer je uit om een beter begrip te krijgen van het natuurlijke systeem. Om deze reden wil je zoveel mogelijk onderdelen weglaten die niet essentieel zijn om het systeem te begrijpen. In onderstaand schema is een schematisch overzicht gegeven van het abstractie proces zoals wij dat in voorgaande hoofdstukken beschreven.



De onderste term, 'Abstractie van leven' is het uiteindelijke doel van Artificial Life. Het onderzoek naar Typogenetics is hier zoals gezegd onderdeel van. Hofstadter heeft echter tijdens het abstractieproces elementen weggelaten die voor het onderzoek naar Artificial Life van groot belang zijn gebleken. We hebben in dit paper drie voorbeelden gegeven: ten eerste ging Hofstadter ervan uit dat alle informatie lag opgeslagen in de nucleotidenvolgorde. Bij het bekijken van de consensussequentie echter, bleek dat er ook informatie ligt opgeslagen in de ruimtelijke structuur van het DNA: namelijk in welke mate een bepaalde keten DNA moet worden afgelezen. In de biologie speelt deze zogenaamde effectiviteit van gentranscriptie een belangrijke rol, maar wanneer je een adequate representatie wil geven van gentranscriptie is het ook van belang. Ten tweede hield Hofstadter er geen rekening mee dat er tijdens de RNA fase van genexpressie nog modulaties plaatsvinden die de structuur en de functie van het uiteindelijke eiwit fundamenteel kunnen veranderen. We hebben laten zien aan de hand van het concept intronen/exonen dat de context van waarin de cel zich bevindt van grote invloed is. Als laatste hebben we het verschil onderzocht tussen de totstandkoming van de structuren van eiwitten bij Typogenetics en genexpressie. In de werkelijkheid blijkt dat de synthese van een eiwit een exact vertaald mRNA is, maar dat bij de vouwing erg veel factoren een rol spelen. Het gevolg is dat kleine aanpassingen in de omgeving van de aminozuren grote gevolgen kunnen hebben voor de uiteindelijke structuur.

We hebben het systeem Typogenetics gebruikt als een case study om te laten zien hoe moeilijk het is om van een natuurlijk systeem als genexpressie een formeel systeem te maken dat alle eigenschappen van het natuurlijke systeem behoudt. Het reduceren van het systeem in afzonderlijke onderdelen gaat altijd gepaard met een dilemma: wat neem je mee, en wat laat je weg? Hofstadter heeft de keuze gemaakt om bovenstaande eigenschappen weg te laten. Deels omdat Typogenetics voor zijn doeleinden niet zo veelomvattend hoefde te zijn, maar ook deels omdat zowel de consensussequentie als de intronen pas ná het verschijnen van zijn boek Gödel, Escher, Bach zijn ontdekt.

De vraag is nu of het mogelijk is om een natuurlijk systeem zoals genexpressie goed genoeg te kunnen begrijpen om er een adequate abstractie van te kunnen van maken. Het probleem is dat natuurlijk systemen op willekeurige manier zijn ontstaan net zoals de oplossingen gevonden door een genetisch algoritme. De systemen zijn niet geconstrueerd dus een duidelijk structuur ontbreekt vrijwel altijd. De functionaliteit van deze systemen is emergent. Net zoals de oplossingen die genetische algoritmes bieden een emergente aard hebben. Dit betekent dat het praktisch onmogelijk is om zeker te weten dat je een goed begrip hebt van alle 'dingen' uit het citaat van Dijkstra om een goede abstractie te maken.