

Genen als moleculaire

Biochemische netwerken, bestaande uit interagerende biomoleculen zoals eiwitten en DNA, zijn de analoge computers van het leven: ze stellen de cel in staat verschillende soorten rekenkundige berekeningen uit te voeren. Voor ons onderzoek hebben wij een combinatie van bio-informatica, analytische theorieën en computersimulaties gebruikt om inzicht te krijgen in de ontwerpprincipes van genetische netwerken – circuits van genen die elkaar aan- en uitzetten. Onze bio-informatiecastudie laat zien dat in de bacterie *Escherichia coli* genen die elkaar aansturen veel dichter bij elkaar op het DNA liggen dan op basis van toeval verwacht mag worden. Uit de theoretische en numerieke analyses blijkt dat dit de cel resistenter maakt tegen biochemische ruis en haar in staat stelt sterker te reageren op veranderingen in haar omgeving. De resultaten suggereren dat genregulatie, ofwel het proces van het op gezette tijden aan- en uitzetten van genen, bepaalt hoe het genoom ruimtelijk georganiseerd is. Rosalind Allen, Patrick Warren en Pieter Rein ten Wolde

voerd door biomoleculen, zoals eiwitten en DNA, die chemisch en fysisch met elkaar interageren. Deze zogenaamde biochemische netwerken worden aangestuurd door toevalsprocessen. De componenten van een netwerk – de eiwitten en het DNA – bewegen zich namelijk vaak op een willekeurige manier door de cel om elkaar te vinden. Bovendien zijn de interacties stochastisch van aard: chemische en fysische interacties zijn kansprocessen. Deze factoren zijn vooral belangrijk wanneer de concentraties van de componenten laag zijn. In de levende cellen is dit vaak het geval, met als gevolg dat de biochemische reacties die aan een moleculair netwerk ten grondslag liggen vaak veel ruis vertonen. Hoe de netwerken desondanks informatie op een betrouwbare manier kunnen verwerken, blijft daarom een belangrijke vraag.

Deze vraag is bijzonder relevant voor zogenaamde genetische netwerken. Genen zorgen voor de productie van eiwitten. Als zo'n eiwit gevormd wordt, spreken we van het tot expressie brengen van een gen (zie het kader). De meeste eiwitten vormen enzymen die biochemische reacties katalyseren, zoals de omzetting van suikermoleculen. Sommige eiwitten echter, zogenaamde transcriptiefactoren, kunnen hun eigen synthese of de expressie van andere genen – en dus de synthese van andere eiwitten – reguleren (zie het kader). Dergelijke netwerken van interagerende genen vormen het rekenkundige hart van elke levende cel. Genetische netwerken bepalen onder andere hoe een cel zich moet differentiëren tijdens de ontwikkeling van het embryo en stellen haar in staat om signalen uit de omgeving te integreren (zie kader).

De concentraties van de transcriptiefactoren zijn vaak uitzonderlijk laag – in een cel is vaak niet meer dan een tiental kopieën van één type aanwezig.

BIOCHEMISCHE ANALOGE COMPUTERS

De meeste organismen kunnen reageren op, en zich aanpassen aan veranderingen in hun omgeving. Hiertoe kunnen zij signalen uit hun omgeving detecteren, doorgeven en versterken, en ook verschillende signalen integreren om bijvoorbeeld patronen in het voedselaanbod te herkennen. Organismen kunnen daarom als informatieverwerkende machines beschouwd worden. Zelfs simpele organismen, zoals de bacterie *Escherichia coli*, kunnen relatief complexe berekeningen uitvoeren, zoals: IF de suikersoort lactose aanwezig is AND NOT glucose, THEN gebruik lactose (zie ook het kader 'Genregulatie en lactoseconsumptie').

Biologen gebruiken deze terminologie uit de Booleaanse logica vaker. De terminologie suggereert dat de berekeningen deterministisch van aard zijn en geen fouten toelaten. In de levende cel worden berekeningen echter uitge-

tenwolde@amolf.nl



Rosalind Allen (1976) studeerde scheikunde aan Cambridge University. Na haar masters-opleiding aan de University of Pennsylvania, verrichtte zij haar promotieonderzoek aan Cambridge University onder begeleiding van prof. Hansen. Zij is nu postdoc in de groep van Pieter Rein ten Wolde op AMOLF in Amsterdam.



Patrick Warren (1965) studeerde natuurkunde aan Cambridge University. Hij promoveerde als theoretisch fysicus aan het Cavendish Laboratory in Cambridge onder begeleiding van prof. Ball. Na een postdoc in de groep van prof. Pusey in Edinburgh is hij sinds twaalf jaar werkzaam bij Unilever in Port Sunlight (VK).



Pieter Rein ten Wolde (1970) studeerde fysische chemie aan de Rijksuniversiteit Groningen. Zijn promotieonderzoek verrichtte hij op AMOLF onder begeleiding van Daan Frenkel. Na een postdoc in de groep van prof. Chandler aan de University of California at Berkeley startte hij in 2001 een onderzoeksgroep Biochemische Netwerken op AMOLF.

(sub)microprocessors

Bovendien is ook het DNA in zeer lage concentraties aanwezig. Recente experimenten hebben dan ook aangetoond dat genexpressie zeer veel ruis vertoont [1–4]. Toch voeren de genetische netwerken hun rekenkundige taken vaak op een heel betrouwbare manier uit. De ontwikkeling van het embryo bijvoorbeeld is gewoonlijk uitermate precies. Hoe genetische netwerken deze hoge nauwkeurigheid kunnen bereiken, wordt nog erg slecht begrepen.

RUIMTELIJKE NETWERKMOTIEVEN

Bio-informatica is een geschikt instrument om meer inzicht in deze vraag te krijgen, omdat ze ons in staat stelt patronen – ‘motieven’ – in netwerken te identificeren [5]. Wij hebben onlangs gekeken naar de ruimtelijke verdeling van genen op het DNA van de bacterie *Escherichia coli* [6]. Figuur 1 toont een kaart van het netwerk van de bacterie en, ter vergelijking, een kaart van een random netwerk. Deze kaarten suggereren dat genen die elkaar aansturen dichter bij elkaar op het DNA liggen dan je op basis van toeval zou mogen verwachten.

Om de ruimtelijke verdeling van genen te kwantificeren hebben we gebruik gemaakt van een grootheid die ook veel gebruikt wordt in theorieën van vloeistoffen. Dit is de paar-correlatiefunctie $g(s)$, die gegeven wordt door het aantal paren in het *E. coli*-netwerk waarbij de twee genen op een afstand s van elkaar liggen (gemeten langs het genoom), gedeeld door het ‘ideale gas’-resultaat, dat overeenkomt met een willekeurige verdeling van de genen over het DNA. Als de genen in het *E. coli*-netwerk willekeurig verdeeld zouden zijn over het genoom, dan is $g(s) = 1$ voor alle afstanden s . De paar-correlatiefunctie meet dus in hoeverre het *E. coli*-netwerk afwijkt van een random netwerk.

Figuur 2a laat zien dat in *E. coli* de genen die elkaar aansturen, en paren

van genen die gecoreguleerd worden door een derde gen, ruimtelijk veel dichter bij elkaar liggen dan in een random netwerk. De figuur toont bovendien dat de dracht van de correlatie

zeer kort is, namelijk enkele honderden baseparen. Uit figuur 2 blijkt ook dat paren van genen die elkaar niet aansturen en ook niet aangestuurd worden door dezelfde transcriptiefac-

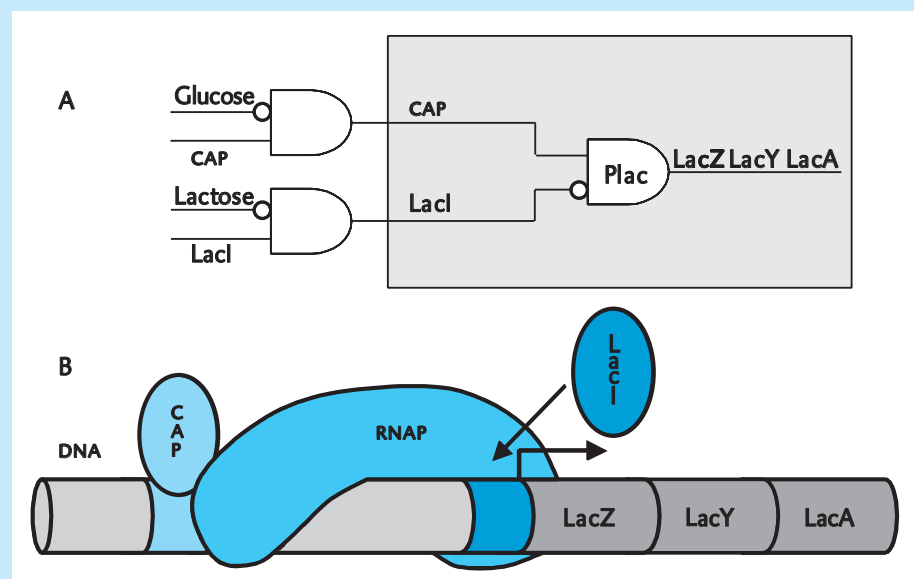
Genregulatie en lactoseconsumptie

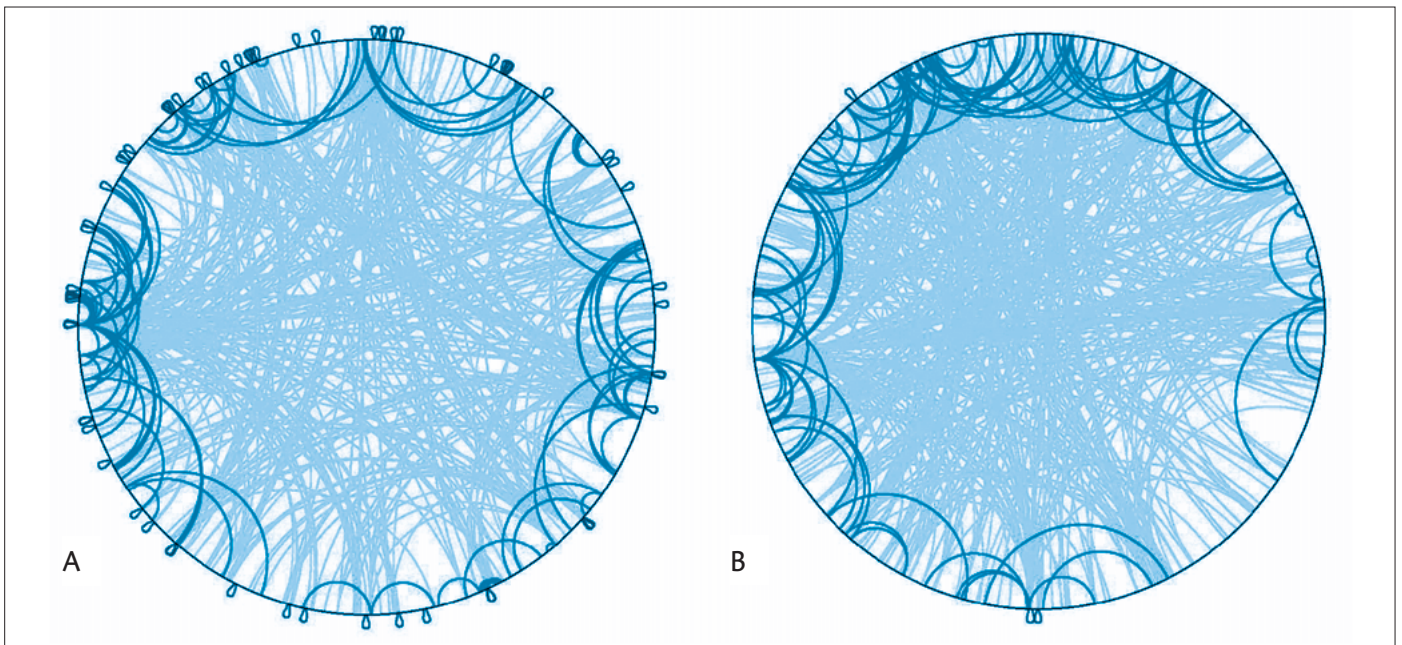
Glucose is de meest voordelige brandstof voor de bacterie *Escherichia coli*. Alleen in afwezigheid van glucose mag de cel lactose gebruiken. Om lactose op te kunnen nemen, moeten drie eiwitten worden gesynthetiseerd. Hiertoe moeten de voor deze eiwitten coderende genen LacZ, LacY, LacA, die op het DNA liggen, tot expressie worden gebracht. Dit is de taak van het enzym RNA-polymerase (RNAP), dat aan het DNA kan binden en dan langs het DNA kan lopen om het DNA te vertalen in eiwitten. De binding van RNAP aan het DNA wordt gereguleerd door twee eiwitten, LacI en CAP, die zelf ook aan het DNA kunnen binden. Hun bindingsplaats wordt een operator genoemd. LacI bindt het DNA in de afwezigheid van lactose, terwijl CAP het DNA bindt in de aanwezigheid van glucose. RNAP kan bovendien alleen maar aan het DNA binden wanneer CAP al aan het DNA gebonden is, maar LacI niet – LacI en RNAP sluiten el-

kaar uit, omdat hun DNA-bindingsplaatsen overlappen.

(A) geeft het elektronisch netwerkdiagram weer: het circuit bestaat uit drie zogenaamde NOT-IF-poorten, met als resultaat dat de drie genen alleen significant tot expressie worden gebracht wanneer lactose aanwezig is en glucose niet.

(B) geeft de moleculaire architectuur van de genetische poort (het grijze veld in (A)) weer. Deze set van drie genen wordt altijd ‘van links naar rechts’ tot expressie gebracht. Andere genen kunnen echter ook van ‘rechts naar links’ tot expressie worden gebracht; het RNAP bindt dan ‘rechts’ van de genen en de genetische code wordt dan vanaf de andere streng van het DNA vertaald. Eén eiwit kan, door bijvoorbeeld tussen twee RNAP’s aan het DNA te binden, twee sets van genen tegelijkertijd aansturen (zie ook figuur 3).





Figuur 1
 Het genetische netwerk van *E. coli* (a) vergeleken met een random netwerk (b). De cirkel geeft het circulaire DNA van *E. coli* weer en de lijnen tonen de paren van genen die elkaar aansturen. Kleuren coderen voor afstand s : donkerblauw lusjes aan de buitenkant: $s < 10$ kbp; blauw: $10 \text{ kbp} < s < 500$ kbp; lichtblauw: $s > 500$ kbp.

tor, wél min of meer willekeurig over het genoom zijn verdeeld.

Om de ruimtelijke organisatie verder te onderzoeken hebben we gekeken naar de oriëntatie waarin genen tot expressie worden gebracht. Het genoom bestaat uit dubbelstrengs DNA; bovendien worden sommige genen vanaf de ene streng, en dus in de ene richting, tot expressie gebracht, terwijl andere genen vanaf de andere streng, en dus in de tegenovergestelde richting, tot expressie worden gebracht (zie het kader en figuur 3). Een paar van genen kan daarom in een gemeenschappelijke richting, in convergente of in divergente richtingen vertaald worden. Figuur 3b laat zien dat paren van genen die gecoreguleerd worden of elkaar aansturen, op lange afstanden in willekeurige richting tot expressie worden gebracht. Op korte afstanden, daarentegen, worden deze genen voorname-

lijk in divergente richtingen tot expressie gebracht.

De database-analyses laten dus zien dat genenparen die gekoppeld zijn in het genetische netwerk van *E. coli* a) veel dichter bij elkaar liggen dan je op basis van toeval zou mogen verwachten, en b) op korte afstand een divergente oriëntatie aannemen. De combinatie van beide observaties suggereert dat genregulatie een belangrijke factor is in de ruimtelijke organisatie van het genetische netwerk van *E. coli*.

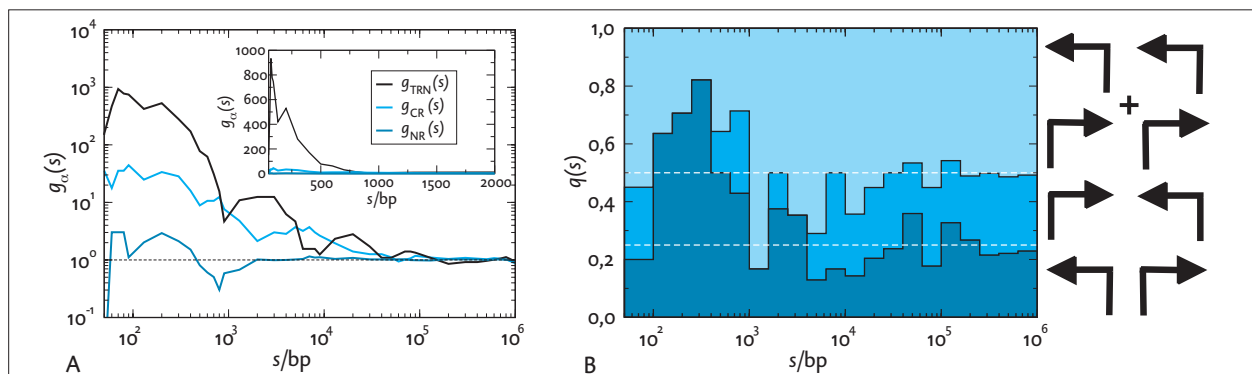
BETROUWBAAR REKENEN

Wij hebben analytische theorie en computersimulaties gebruikt om het scenario waarin genregulatie de ruimtelijke organisatie van het genoom bepaalt, nader te bestuderen. Een inspectie van de divergente genenparen liet zien dat hun zogenaamde *operatoren*

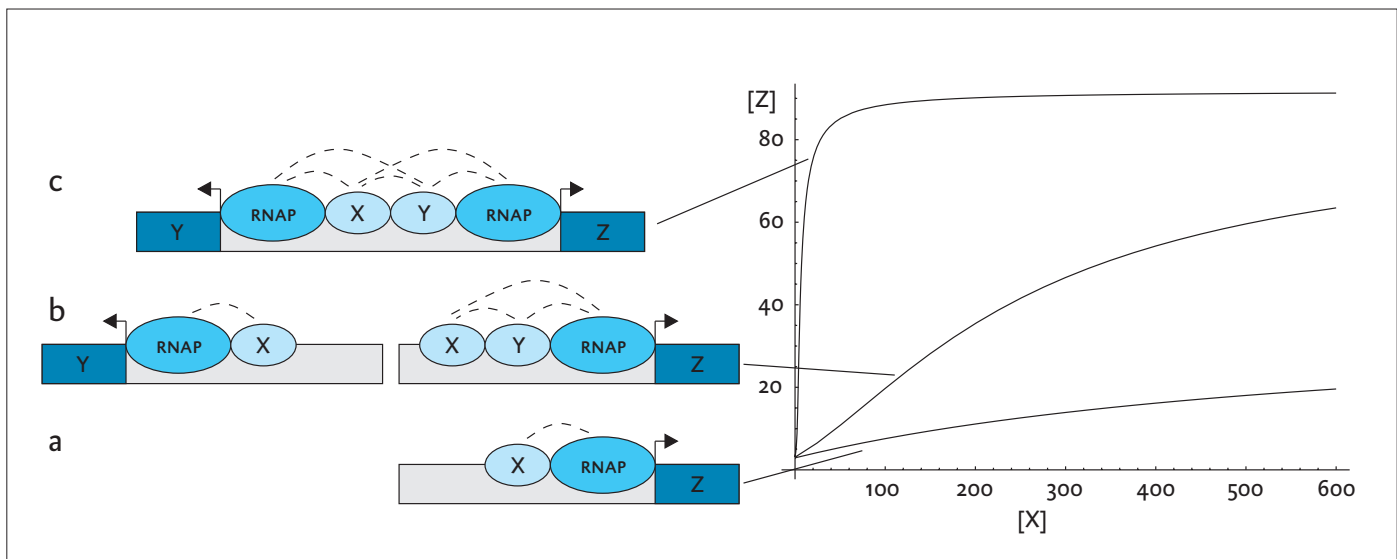
vaak overlappen. Een operator is een stuk DNA waar de transcriptiefactoren kunnen binden om het gen 'aan' of 'uit' te zetten (zie het kader). De operatoren van twee genen kunnen elkaar dus overlappen indien de genen in divergente richtingen vertaald worden (zie figuur 3). Dit biedt de cel een aantal mogelijkheden. Zo kan één gebonden transcriptiefactor bijvoorbeeld twee genen tegelijkertijd aan- of uitzetten, of één gen aanzetten en het andere uit. Overlappende operatoren stellen de cel dus in staat om twee genen op een gecorreleerde, dan wel anti-gecorrleerde manier aan te sturen.

Om naar de consequenties van (anti) correlatie in genexpressie te kijken, hebben we twee netwerkmotieven geanalyseerd. Uri Alon en zijn medewerkers hebben laten zien dat veel biochemische netwerken opgebouwd zijn uit relatief kleine deelnetwerken – de net-

8



Figuur 2
 a) Paar-correlatiefunctie voor paren van genen die elkaar aansturen ($g_{\text{TRN}}(s)$), voor gecoreguleerde paren ($g_{\text{CR}}(s)$), en voor de overige genenparen ($g_{\text{NR}}(s)$); voor een random netwerk is $g(s) = 1$ voor alle afstanden s .
 b) Oriëntatiestatistiek voor de netwerkparen als functie van de afstand. Donkerblauw correspondeert met de fractie divergente genen, blauw met convergente genen, en lichtblauw met tandemgenen (= gemeenschappelijke richting). De gestippelde witte lijnen corresponderen met een random oriëntatieverdeling. s/bp staat voor de afstand gemeten in basenparen.



Figuur 3
 Respons van een *feed forward-module* (FFM) als functie van de ruimtelijke architectuur. Eiwit x zet gen y aan, en de eiwitten x en y zetten samen gen z aan. b) Conventionele FFM; c) FFM met overlappende operatoren: één molecuul x (y) kan tegelijkertijd zowel gen y als gen z aanzetten. a) Ter vergelijking een structuur waarbij de expressie van z alleen wordt aangestuurd

door eiwit x. Rechts de concentratie van z (de output) als functie van de concentratie van x (de input). De input-output-relatie laat zien dat een FFM kan fungeren als een versterker en dat overlappende operatoren de performance van de versterker aanzienlijk kunnen verhogen. De gestippelde lijnen geven interacties tussen moleculen weer.

werkmotieven – die onderling met elkaar verbonden zijn [5]. Deze netwerk-motieven zijn de rekenkundige eenheden van een biochemisch netwerk; elk netwerk-motief kan één of meerdere rekenkundige taken uitvoeren. Het eerste motief dat we hebben bekeken is de zogenaamde *feed-forward-module*. Een *feed-forward-module* is een netwerk-motief waarbij een transcriptiefactor x een gen y aanstuurt, en de transcriptiefactoren x en y tezamen het gen z aansturen (zie figuur 3). *Feed-forward-modules* kunnen verschillende soorten berekeningen uitvoeren. Ze kunnen bijvoorbeeld tijdelijke signalen filteren, maar ook inkomende signalen versterken. Onze database-analyses lieten zien dat in de *feed-forward-modules* van *E. coli* de *downstream* genen y en z vaak naast elkaar op het genoom liggen en in divergente richtingen worden getranscribeerd (figuur 3). Deze architectuur stelt de eiwitten x en y in staat de genen y en z tegelijkertijd aan te zetten. Onze theoretische berekeningen laten zien dat dit tot een scherpere respons van het netwerk leidt (figuur 3). Overlappende operatoren stellen de cel dus in staat om sterker op veranderingen in haar omgeving te reageren. Het tweede motief dat we hebben bestudeerd bestaat uit twee genen, die elkaar wederzijds onderdrukken (x onderdrukt de expressie van y, en omgekeerd). Dit netwerk kan twee stabiele toestanden aannemen – corresponderend met veel gesynthetiseerd x en weinig gevormd y, en *vice versa*. Het netwerk kan dus als een genetische schakelaar fungeren.

Normaliter zetten inkomende signalen (bijvoorbeeld voedselmoleculen) dit soort schakelaars in een andere stand door een transcriptiefactor zodanig te modificeren dat deze zijn gen niet meer kan aansturen. Een schakelaar kan echter door biochemische ruis ook spontaan omklappen: als de schakelaar bijvoorbeeld in stand x staat, en er dus veel eiwit x aanwezig is, dan is er een kleine kans dat y toch gesynthetiseerd wordt. De reden is dat een molecuul niet continu aan het DNA gebonden kan blijven (als dat wel zou kunnen, dan zou het ook oneindig veel energie kosten de schakelaar om te laten klappen). Een belangrijke vraag is dus hoe de stabiliteit van genetische schakelaars beïnvloed wordt door biochemische ruis.

Wij hebben met behulp van stochastische simulaties de stabiliteit van een conventionele schakelaar vergeleken met die van een schakelaar met overlappende operatoren, waarbij de transcriptiefactoren elkaar wederzijds uitsluiten op het DNA [7,8]. De berekeningen laten zien dat in beide gevallen de stabiliteit zeer sterk toeneemt met het aantal tot expressie gebrachte moleculen [7]. Dit verklaart mogelijk waarom in levende cellen de concentraties van de transcriptiefactoren vaak zeer laag zijn. De berekeningen laten ook zien dat de ‘exclusieve’ schakelaars ordes van grootte stabiel zijn dan conventionele schakelaars [7,8]. Overlappende operatoren kunnen dus de stabiliteit van genetische netwerken aanzienlijk verhogen.

Het moleculaire karakter van de componenten leidt er toe dat genetische netwerken worden aangedreven door processen die inherent stochastisch van aard zijn. Tegelijkertijd maakt moleculaire lego het ook mogelijk van deze netwerken geavanceerde rekenmachines te maken. Zo onthulden de database-analyses dat de operatoren van twee genen die elkaar aansturen, of die gecoreguleerd worden, elkaar vaak overlappen, waardoor deze genen ófwel op een gecorreleerde, ófwel op een anti-gecorreleerde manier tot expressie kunnen worden gebracht. De theoretische berekeningen en de computersimulaties hebben laten zien dat dit soort ontwerpen de respons van genetische netwerken sterk kan verhogen en de netwerken meer resistent kan maken tegen biochemische ruis. Samen suggereren deze resultaten dat genregulatie de drijvende kracht is voor de ruimtelijke organisatie van het genoom.

REFERENTIES

- 1 H.H. McAdams en A. Arkin, *Proc. Natl. Acad. Sci. USA* **94** (1997), 814.
- 2 Elowitz en S. Leibler, *Nature* **403** (2000), 335.
- 3 E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman en A. van Oudenaarden, *Nat. Gen.* **31** (2002), 69.
- 4 M.B. Elowitz, A.J. Levine, E.D. Siggia en P.S. Swain, *Science* **297** (2002), 1183.
- 5 S.S. Shen-Orr, R. Milo, S. Mangan en U. Alon, *Nat. Gen.* **31** (2002), 64.
- 6 P.B. Warren en P.R. ten Wolde, *J. Mol. Biol.* **342** (2004), 1379.
- 7 P.B. Warren en P.R. ten Wolde, *Phys. Rev. Lett.* **92** (2004), 128101.
- 8 R.J. Allen, P.B. Warren en P.R. ten Wolde, *Phys. Rev. Lett.* **94** (2005), 018104.