

# VU Research Portal

## RNA in Formation

Smit, S.

2008

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Smit, S. (2008). *RNA in Formation: Computational Studies on RNA Structure and Evolution*.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## RNA in Formatie: Computationale Studies van de Structuur en Evolutie van RNA

Dit proefschrift gaat over de structuur en evolutie van ribonucleïnezuren, oftewel RNA (ribonucleic acid). RNA moleculen zijn te vinden in levende cellen en vervullen daar belangrijke taken, zoals het opslaan en overbrengen van genetische informatie of het op gang brengen van chemische reacties. RNAs zijn lange ketens van nucleotiden, die elk weer bestaan uit een suiker (ribose), een fosfaat en één van vier organische basen (aangeduid met de letters U, C, A, of G). Een nucleotide wordt geïdentificeerd op grond van de base. De keten van basen, ook wel sequentie genoemd, beschrijft het molecuul. RNA moleculen vouwen zich op in drie-dimensionale structuren net zoals eiwitten dat doen. Deze structuur bepaalt welke functie het molecuul heeft en hoe deze uitgevoerd wordt. Het is dus belangrijk om te begrijpen hoe deze structuur gevormd wordt.

RNA moleculen zijn cruciaal voor het functioneren van een cel en zijn dus aanwezig in alle organismen op aarde. RNAs met dezelfde functie in meerdere organismen zijn echter niet allemaal identiek. De structuur die tevens de functie bepaalt, is globaal gezien hetzelfde, maar er kunnen verschillen zijn in de sequentie. Deze verschillen zijn ontstaan door evolutie, het principe van overerving met variatie. Het genetisch materiaal in de nakomelingen van een organisme is net weer anders dan dat in de ouder. Veranderingen die de fitheid van het organisme aantasten, zullen tegengehouden worden door natuurlijke selectie (bijvoorbeeld: het individu sterft of kan zich niet voortplanten). Daarentegen leveren veranderingen waardoor het organisme beter functioneert een evolutionair voordeel op. Dit evolutieproces vindt in feite plaats op moleculair niveau in de cel door veranderingen in het genetisch materiaal. Met de huidige technieken in de biologie kunnen we evolutie in macromoleculen, zoals RNA, bestuderen om fundamenteel inzicht in het proces van evolutie te krijgen.

Dit proefschrift bevat meerdere studies van de evolutie van RNA moleculen. We bestuderen wat de invloed van de structuur is op de evolutie van een molecuul. Ook presenteren we twee praktische toepassingen die met RNA structuren te maken hebben. Dit werk valt onder de noemer bioinformatica, een vorm van toegepaste informatica. Het doel van studies in dit vakgebied is om biologische vragen te beantwoorden door middel van (grootschalige) data analyse. Specifiek gaat het om het uitvoeren van computationele analyses (i.e. met behulp van een computer) en het onwikkelen

van nieuwe methoden of algoritmen, rekenkundige procedures die een bepaalde taak uitvoeren. Hieronder geven we een korte uitleg over het werk dat in het proefschrift gepresenteerd wordt.

#### Evolutie van ribosomale RNA moleculen

Darwin ontwikkelde zijn evolutietheorie door het vergelijken van het uiterlijk en gedrag van individuele organismen en soorten. Tegenwoordig kunnen we evolutie op moleculair niveau bestuderen. De genetische informatie van een cel staat opgeslagen in het DNA en elke keer dat een cel zich deelt of een organisme zich voortplant, kunnen er kleine veranderingen in dit DNA optreden. Deze veranderingen zullen doorwerken in andere moleculen die door het DNA gecodeerd worden, zoals RNA en eiwitten. Het vergelijken van sequenties van moleculen uit verschillende organismen vertelt iets over de evolutionaire relatie tussen hen. Als de sequenties heel verschillend zijn, zit er dus een lange tijd van evolutie tussen (mogelijk miljoenen jaren), maar als de sequenties veel op elkaar lijken, moeten de twee organismen korter geleden een gemeenschappelijke voorouder gehad hebben. Uit dit soort vergelijkingen weten we bijvoorbeeld hoe genetisch verschillend mensen en chimpanzees zijn en hoe lang geleden hun gemeenschappelijke voorouder leefde. Met name voor dit soort tijdschattingen zijn modellen van de evolutie nodig die beschrijven welke veranderingen in het DNA hoe vaak voorkomen.

Een molecuul dat uitermate geschikt is voor het bepalen van dit soort evolutionaire relaties tussen soorten is het ribosomale RNA (rRNA). Het ribosoom is namelijk een essentieel onderdeel in elke cel: het is de moleculaire machine die eiwitten produceert. Dit apparaat is zó belangrijk dat het aanwezig is in alle organismen op aarde en waarschijnlijk al aanwezig was in de eerste vormen van leven. Het is ook nog eens aanwezig in grote aantallen in elke cel, dus de sequentie ervan kan relatief makkelijk bepaald worden. Veel van wat we weten over moleculaire evolutie is gebaseerd op vergelijkingen van deze rRNA moleculen. De evolutiemodellen kunnen mogelijk verfijnd worden door in meer detail naar deze moleculen te kijken.

In hoofdstuk 2 en 3 hebben we de richting en snelheid van veranderingen in RNA vergeleken en bestudeerd hoe dit verschilt tussen bepaalde onderdelen van structuur en tussen groepen organismen (bijvoorbeeld de zoogdieren of gammaproteo-bacteriën). Een belangrijk resultaat is dat de samenstelling en de snelheid van evolutie verschilt per onderdeel van de structuur. Ook tussen verschillende groepen organismen worden verschillen gevonden. Als evolutiemodellen rekening zouden houden met deze verschillen, zouden de verwantschappen tussen soorten nauwkeuriger bepaald kunnen worden.

#### Evolutie van genetische boodschappers

Het ribosomale RNA dat hierboven beschreven staat is slechts één van de vele soorten RNA. Een ander, veelvuldig voorkomend RNA is “messenger RNA” (mRNA), oftewel boodschappers. De functie van deze moleculen is om de genetische instructies (genen) in het DNA van een organisme aan de ribosomen door te geven, waar ze omgezet

worden in eiwitten. Een mRNA molecuul is dus in feite een afdruk van een gen. De (vrijwel) universele genetische code beschrijft precies hoe de nucleotiden in het mRNA molecuul vertalen naar een eiwit. Deze code zit zodanig in elkaar dat fouten in het mRNA molecuul niet altijd leiden tot fouten in het eiwit. Dit komt doordat meerdere stukjes sequentie (codons) voor hetzelfde aminozuur (bouwsteen van eiwitten) coderen. Toch gebruiken organismen deze redundante codons niet met dezelfde frequentie, maar laten ze specifieke voorkeuren zien. De reden achter dit selectieve gebruik van de genetische code is veel minder goed bestudeerd dan de genetische code zelf.

In hoofdstuk 4 hebben we de evolutie van deze genetische boodschappers onder de loep genomen. Specifiek hebben we getest of het codongebruik in biologische mRNA moleculen samenhangt met het voorkomen van fouten in eiwitten, net als in de genetische code zelf het geval is. Als de boodschappers geëvolueerd zouden zijn om deze eigenschap te optimaliseren, zouden we verwachten dat biologische mRNA moleculen inderdaad beter bestand zijn tegen fouten dan willekeurig gemaakte RNA moleculen. Dit blijkt echter niet het geval te zijn. Sterker nog, willekeurig gemaakte RNA moleculen waren beter bestand tegen deze fouten. Dit leidde tot de interessante speculatie dat RNA moleculen misschien wel geselecteerd zijn om een bepaald niveau van variatie toe te staan. Evolutie bestaat per slot van rekening ook alleen bij gratie van variatie: zonder verandering kan iets ook nooit beter worden.

### Structuur voorspelling

Een grote uitdaging wat betreft RNA moleculen en eiwitten is het voorspellen van de structuur en de functie van het molecuul op grond van de keten van bouwstenen. Deze ketens kunnen we met experimentele technieken “aflezen” en als ons begrip van het vouwingsproces compleet zou zijn, zouden we kunnen bepalen welke vorm en functie het molecuul heeft zonder experimenten uit te voeren. Deze voorspellingstechnieken zijn uitermate belangrijk in de medische wereld voor bijvoorbeeld het begrijpen en mogelijk repareren van defecten in eiwitten. Een andere belangrijke toepassing is het doelgericht ontwerpen van medicijnen of andere moleculen die een specifieke functie uitvoeren.

Voorspellingsmethoden voor RNA structuren bestaan in alle soorten en maten en behalen wisselend succes op verschillende typen RNA. In hoofdstuk 5 presenteren we een nieuwe methode om RNA structuren te voorspellen voor een groep van biologisch gerelateerde sequenties. We maken gebruik van evolutionaire patronen in de compositie van het molecuul (beschreven in hoofdstuk 2), die nog nooit voor dit doeleinde zijn ingezet. Deze patronen blijken een rijke bron van informatie te zijn en we beschrijven hoe deze precies gebruikt kunnen worden voor het voorspellen van een RNA structuur.

### Knopen in het RNA

Modellen van RNA structuren worden in allerlei toepassingen gebruikt, bijvoorbeeld voor het zoeken naar nieuwe RNA moleculen in het genetisch materiaal van een orga-

nisme. In sommige structuren komen zogenaamde (pseudo)knopen voor die berekeningen erg moeilijk maken. Deze worden dus vaak verwijderd voordat de analyse met de structuur uitgevoerd wordt. Het probleem met het verwijderen van deze knopen is dat veel onderzoekers dit vaak ad hoc en op een verschillende manier doen, wat tot onduidelijkheid leidt en het moeilijk maakt om studies te reproduceren. In hoofdstuk 6 beschrijven we verschillende criteria die gebruikt kunnen worden voor het definiëren en verwijderen van pseudoknopen. Alle methoden kunnen resulteren in een verschillende knoop-vrije structuur, dus onderzoekers moeten goed nadenken over welke methode het meest geschikt is voor hun toepassing. Bovendien moeten ze goed documenteren welke methode ze gebruikt hebben.

#### Bioinformatica software

De software die is ontwikkeld is om de analyses die in dit proefschrift staan beschreven uit te voeren, heeft bijgedragen aan de opbouw van een software pakket (PyCogent) voor “comparative genomics” (vergelijkende genoomanalyse). Dit type software is hard nodig in de bioinformatica omdat de hoeveelheden biologische data razend snel groeien en de analyses dus steeds groter en ingewikkelder worden. Het PyCogent pakket zorgt ervoor dat analyses snel opgezet kunnen worden en niet iedereen steeds dezelfde taken hoeft uit te voeren. Veel standaard concepten en methodes uit de bioinformatica zijn namelijk geïmplementeerd en direct bruikbaar. We hopen deze software in de toekomst verder te ontwikkelen. Het doel is om dit pakket niet alleen binnen het onderzoek in te zetten, maar ook voor educatieve doeleinden te gebruiken zodat de volgende generatie bioinformatica studenten een vliegende start kan krijgen.