



Universiteit
Leiden
The Netherlands

Freedom of additional signals on genes: on the combination of DNA mechanics, genetics and translation speed

Zuiddam, M.

Citation

Zuiddam, M. (2022, April 6). *Freedom of additional signals on genes: on the combination of DNA mechanics, genetics and translation speed. Casimir PhD Series*. Retrieved from <https://hdl.handle.net/1887/3281818>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3281818>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

DNA is de drager van verschillende vormen van informatie. De meest bekende van deze vormen is klassieke genetische informatie: genen op het DNA coderen voor eiwitten en RNAs. Deze genetische code bestaat uit reeksen van basen, afgekort A, T, C en G. Een basenreeks kan informatie bevatten die nodig is om een specifiek eiwit te bouwen. In deze dissertatie wordt een tweede informatielaag bestudeerd: de mechanica van DNA. Een reeks bestaande uit A's en T's zal anders buigen dan een reeks G's en T's. Dit komt doordat de intrinsieke vorm en flexibiliteit van DNA afhangen van de gekozen basen. Een belangrijk gevolg van deze mechanische informatielaag betreft de positionering van nucleosomen.

Nucleosomen bestaan uit 147 basenparen (bp) DNA gewikkeld om een eiwitkern, als een draad om een spoel. Nucleosomen zijn de fundamentele bouwblokken waaruit chromosomen bestaan. Ze zorgen ervoor dat het DNA compact is, en ze zijn verantwoordelijk voor de hogere-ordestructuren van DNA. Voor nucleosomen is ook een belangrijke taak weggelegd voor de regulatie van DNA. Ze kunnen fysiek voorkomen dat transcriptiefactoren (eiwitten die transcriptie reguleren) de transcriptiefactorbindingsplaatsen op het DNA bereiken, simpelweg voor zich op zo'n plaats te bevinden. Een cel kan de nucleosomen (of, om precies te zijn, de 'histonstaarten' van nucleosomen) scheikundig modificeren, waardoor een bindingsplaats weer beschikbaar wordt. Deze modificaties kunnen zelfs geërfd worden door de nakomelingen van een organisme. Dit is een voorbeeld van zogeheten epigenetica. Doordat een nucleosoom een bindingsplaats kan blokkeren of vrij kan maken, kan het dienen als een aan/uit-knop. De locatie van zulke knoppen is zeer belangrijk.

De locaties van nucleosomen op het DNA worden beïnvloed door de mechanische informatie op DNA. Het DNA op een nucleosoom moet op een specifieke manier gebogen zijn, een buiging die bij sommige DNA-reeksen makkelijk bereikt wordt, maar bij andere juist niet. Dit feit leidt tot een nucleosoompositiecode, vaak gerepresenteerd door de kans om een dinucleotidestap (een base gevolgd door een andere base) aan te treffen op verschillende locaties in het nucleosoom. Zo zijn bijvoorbeeld de kansen om dinucleotidestappen TT, AA of TA aan te treffen het hoogst daar waar de kleine groef van het DNA gericht is op de eiwitkern, terwijl de kans op een GC-stap het hoogst is op plekken waar de grote groef op de eiwitkern gericht is.

Er bestaan modellen die deze regels kunnen reproduceren maar ze niet kunnen verklaren. Vooral de regels over de GC-stap blijken tegenintuïtief te zijn, omdat de kansen om GC aan te treffen het hoogst zijn daar waar de buigingsenergie van GC maximaal is. Het eerste belangrijke resultaat in deze dissertatie is een fysische verklaring voor deze regels. In Hoofdstuk 2 wordt dit probleem aangepakt door gebruik te maken van een model dat de voornaamste nucleosoompositieregels reproduceert en ook analytisch op te lossen is. Gebruikmakende van de transfermatrixmethode

en een nieuwe benaderingsmethode slagen we er in om de mechanica achter de GC-regels te ontsluiten, alsook om methodisch de andere regels te verklaren.

Een andere aanpak is om niet deze dinucleotideregels te bestuderen, maar om te kijken naar hoeveel mogelijkheid er is om lange stukken DNA mechanische informatie te laten bevatten. In Hoofdstuk 3 pakken wij dit aan door de DNA-reeksen te bestuderen die óf het allerbest óf het allerslechtst zijn in het aantrekken van nucleosomen. We beschrijven alle mogelijke DNA-reeksen in een gewogen graaf. Voor de gewichten maken we gebruik van het probabilistische model van Tompita et al. [10]. Door gebruik te maken van een kortstepad algoritme vinden we de DNA-sequenties die de hoogste en laagste energiekosten hebben om onderdeel te zijn van een nucleosoom. Gebruikmakende van een k -kortstepad algoritme vinden we k van de hoogste en laagste mogelijke energieën. De twee grote voordelen van deze methoden (ten opzichte van bijvoorbeeld Monte-Carlosimulaties) zijn dat kortstepad algoritmes exact en snel zijn.

DNA in echte, levende organismen heeft niet de luxe om alleen mechanische signalen te bevatten. Echt DNA moet informatie bevatten voor RNA's en eiwitten. Theoretisch gezien kunnen deze twee informatielagen, genetisch en mechanisch, naast elkaar bestaan op hetzelfde stuk DNA. Een eiwit bestaat namelijk uit één of meer aminozuurketens. De eigenschappen van een eiwit hangen af van het type aminozuur en de volgorde van de aminozuren in de ketens. De aminozuren zijn geëncodeerd op het DNA in zogeheten codons (een drietal van basen). In veel gevallen zijn er synonieme codons, verschillende codons die coderen voor hetzelfde aminozuur. Dit wordt de ontaarding van de genetische code genoemd. Door deze ontaarding kunnen verschillende DNA-sequenties voor hetzelfde eiwit coderen én verschillende mechanische eigenschappen hebben. Daarom kan een eiwitcoderend stuk DNA ook mechanische informatie bevatten.

Gebruikmakende van de ontaarding van de genetische code hebben we onderzocht hoeveel vrijheid DNA heeft om beide vormen van informatie te bevatten. We hebben dit gedaan door grafen te gebruiken die alle synonieme manieren bevatten om te coderen voor hetzelfde eiwit. Een kortste pad in zo'n graaf geeft nu de DNA-sequenties met de hoogste en laagste energiekosten, terwijl het DNA nog steeds codeert voor het oorspronkelijke eiwit. Bovendien presenteren wij een heuristische methode die gunstige nucleosoomposities kan creëren bovenop een gen. Deze methode demonstreren we op het genoom van *Saccharomyces cerevisiae*, bakkersgist. Ook deze methode gebruikt graafrepresentaties van genen en kortstepad algoritmes. We onderzoeken *alle* posities op *alle* eiwitcoderende genen van gist, en slagen erin om nucleosoompositioneringssignalen te creëren met een precisie van één enkel basepaar voor 99,897% van alle mogelijke nucleosoomposities.

Tot dusver hebben we twee verschillende informatielagen voorbij zien komen. Hierbij hebben we een *derde* informatielaag over het hoofd gezien: translatiesnelheid. Translatiesnelheid is de snelheid waarmee een eiwit gemaakt wordt. Deze snelheid heeft belangrijke gevolgen voor de resulterende eiwitten, aangezien het niet alleen beïnvloed hoe snel eiwitten geproduceerd kunnen worden, maar ook hoe de aminozuurketen zich vouwt tijdens translatie. De vouwing beïnvloed de functie en kwaliteit van een eiwit. De snelheid waarmee een aminozuur wordt toegevoegd aan de keten hangt af van de codon op het DNA (en ook van het celtype en het soort organisme). Ook al coderen synonieme codons voor hetzelfde aminozuur, toch maakt

de keuze van het codon uit voor het resulterende translatielandschap. Hierdoor kan het aanpassen van een gen, ook al blijft de genetische code intact, leiden tot niet-werkende eiwitten. Om deze reden voegen we in Hoofdstuk 4 de translatiesnelheid toe aan onze analyse. We presenteren verschillende aanpakken die translatiesnelheid incorporeren gebruikmakende van de graafrepresentaties die we in Hoofdstuk 3 geïntroduceerd hebben. Deze aanpakken gebruiken hetzij *pruning*, waarbij knooppunten uit een graaf worden weggesnoeid die zouden lijden tot een te verschillend translatiesnelheidslandschap, hetzij veranderen de gewichten in de graaf zodat die zowel translatiesnelheid als nucleosoomenergie bevatten.

We kunnen de laatste aanpak zelfs gebruiken in de context van genetisch gemodificeerde organismes. Wanneer een eiwitcoderend stuk DNA van het ene organisme in een ander organisme geplaatst wordt, kan het translatielandschap erg verschillend zijn. Het aanpassen van dit stuk DNA kan ervoor zorgen dat het translatiesnelheidslandschap enigszins hersteld wordt, maar hierdoor kan dit aangepaste DNA nu juist zijn mechanische signaal verloren zijn. Daarom beschrijven we in deze dissertatie een heuristische methode die laat zien hoe het DNA aangepast kan worden zodanig dat, in een ander organisme, de genetische informatie behouden blijft terwijl het mechanische landschap en het translatiesnelheidslandschap sterk lijken op de versies in het oorspronkelijke organisme. Ook deze methode is gestoeld op het gebruik van grafen en kortstepadalgoritmes.

In het laatste deel van de dissertatie bespreken we hoe genetica en mechanica in genomen van echte organismen zijn gemultiplexed (bij multiplexing is er sprake van meerdere signalen op hetzelfde medium, in dit geval: DNA). Eerder hebben we laten zien dat nucleosoomsignalen kunnen bestaan in eiwitcoderende regio's van genen, zelfs als translatiesnelheid meegenomen wordt. Het is echter belangrijk om aan te merken dat genen ook niet-coderende stukken DNA bevatten, zoals introns. Het is dus mogelijk dat de mechanische signalen op DNA simpelweg geëncodeerd worden door deze niet-coderende stukken. Om dit uit te zoeken introduceren we een klassificatiemethode voor verschillende typen multiplexing. Deze methode stelt ons in staat erachter te komen welke organismen mechanische informatie bevatten op eiwitcoderend DNA, en welke organismen andere strategieën gebruiken. De methode maakt onderscheid tussen de contributies van verschillende regio's aan het totale nucleosoompositioneringssignaal. Deze regio's zijn exonen (coderende regio's), intronen en UTRs (beide niet-coderende regio's). Ook introduceren we het concept van intraregionale signalen, die zich bevinden op één enkele regio, en interregionale signalen, de signalen die voortkomen uit de inherente verschillen tussen regio's.

We bestuderen de Transcription Start Sites (TSSs), plaatsen op het DNA waarvan bekend is dat ze sterke mechanische signalen bevatten, van een groot aantal organismen. Met onze klassificatiemethode laten we zien dat, voor veel organismen zoals vissen en veel planten, interregionale signalen dominant zijn in de buurt van een TSS. Voor mensen en veel andere dieren blijkt het introngedeelte (niet-coderend) van het intraregionale signaal te domineren. Bij rijst en andere granen overheerst het exongedeelte (coderend) van het intraregionale signaal. Het mechanische signaal van rijst, wat groter is dan dat van mens, laat zien dat zelfs in echte genomen coderend DNA sterke mechanische signalen kan bevatten.

Voor rijst laten we zien dat een groot gedeelte van dit nucleosoompositioneringssignaal niet veroorzaakt wordt door synonieme codons maar door de aminozu-

urkeuze. We zijn erachter gekomen dat de aminozuursequentie een significant effect heeft op het gemiddelde mechanische landschap van rijst. Hierdoor vermoeden wij dat mechanische informatie in sommige gevallen belangrijker is dan eiwitcoderende informatie, en dat er zelfs evolutionaire competitie kan zijn tussen deze twee informatielagen.

We waren ook in de gelegenheid om translatiesnelheid aan deze analyse te verbinden. Resultaten van rijst en mens doen ons vermoeden dat er nog een competitie bestaat: tussen mechanische informatie en translatiesnelheid. Alles bij elkaar leidt dit tot de hypothese dat genetica, mechanica en translatiesnelheid een competitie met elkaar aangaan.