



Universiteit
Leiden
The Netherlands

Topological decoding of biomolecular fold complexity

Scalvini, B.

Citation

Scalvini, B. (2023, July 5). *Topological decoding of biomolecular fold complexity*. Retrieved from <https://hdl.handle.net/1887/3629563>

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/1887/3629563>

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

SAMENVATTING

In de jaren zeventig voorspelde de Nobelprijswinnende scheikundige Linus Pauling het belang van de vorm van moleculen bij het reguleren van hun fysiologisch gedrag. Tegenwoordig, vijftig jaar later, wordt biopolymeertopologie erkend als een sleutelfactor die biofysische eigenschappen beïnvloedt, zoals de snelheid en routes van eiwitvouwing, evenals de bindingsdynamiek. Hoewel de wederzijdse invloed tussen structuur en functie algemeen wordt erkend, blijven de mechanismen van deze interacties grotendeels een open vraag, ver verwijderd van het begrip dat de wetenschappelijke gemeenschap heeft verworven over de impact van chemische en biologische variabelen. Dit kennishiaat is gedeeltelijk te wijten aan het ontbreken van een eenduidig kader voor de topologische karakterisering van biopolymeren. Hoofdstuk 1 onderzoekt een verscheidenheid aan populaire toepassingen van topologie op het gebied van gevouwen moleculaire ketens, waaronder eiwitten, RNA, DNA en synthetische polymeren. In het bijzonder onderzoeken we de voor- en nadelen van de topologie van ruggraat-kruisingen (knooptechnologie) en contactgebaseerde beschrijvingen - specifiek het relatief nieuwe kader van Circuit Topologie (CT). Circuit Topologie berust op de definitie van drie fundamentele topologische relaties tussen contactparen, namelijk serie (S), parallel (P) en kruis (X), gedefinieerd op basis van de onderlinge posities van hun contactpunten. Hoofdstuk 2 heeft als doel de rol van circuit-topologie als voorspeller van vouwsnelheid bloot te leggen en de voorspellende kracht te vergelijken met traditionele geometrische beschrijvingen zoals eiwitgrootte en contactvolgorde. We vinden dat CT-relaties, ontleend aan de inheemse 3D-structuur van een eiwit, een aanvulling kunnen zijn op andere voorspellers door nauwkeurige schattingen van de vouwsnelheid te geven en een theoretisch inzicht te geven in de topologische principes van eiwitvouwing.

Echter, niet alle eiwitten hebben stabiele inheemse structuren waaruit topologische parameters kunnen worden afgeleid. Intrinsiek ongeordende eiwitten (IDPs) zijn betrokken in de ontwikkeling van verschillende ziektes maar het is lastig om hun structuur te bepalen. In Hoofdstuk 3 breiden we de theoretische en computationele pijplijn van CT uit voor het afleiden van conformationele kenmerken uit dynamische en ongestructureerde systemen. We hebben een methode ontwikkeld genaamd Dynamische Circuit Topologie die verder gaat dan het beschrijven van eiwitvouwingen. Met deze methode kunnen ook ongestructureerde polymeren vergeleken worden en tussen verschillende IDPs kan een topologische gelijkheidsscore bepaald worden. Deze techniek maakt onder andere een nieuwe representatie van het vouwtraject op de topologische ruimte mogelijk, die in staat is om tijdelijke toestanden te benadrukken en mogelijk te relateren aan

belangrijke conformities voor farmaceutische toepassingen.

Theoretische kaders hebben als doel om enerzijds ons begrip van fenomenen te verbeteren en anderzijds om experimentele verkenning te ondersteunen en te begeleiden. In de afgelopen jaren hebben technieken voor forcespectroscopie met één molecuul een ongekend inzicht gegeven in topologische overgangen in moleculen, waardoor tijdelijke toestanden in moleculaire processen zoals eiwitvouwing direct kunnen worden waargenomen. De interpretatie van dergelijke experimenten kan echter soms een uitdaging zijn, vooral vanwege de complexiteit en heterogeniteit van interacties die kunnen optreden tijdens het vouwen onder fysiologische omstandigheden. Eiwitvouwing, in de cellulaire omgeving, is inherent een samenwerkingsproces. Moleculaire chaperonnes zijn een klasse eiwitten met een verscheidenheid aan functies om het vouwproces te ondersteunen. Terwijl Optical Tweezers (OT) en andere enkelmolecuulstechnieken de eiwit-chaperonne-interactie in isolatie met specifieke chaperonne-systemen diepgaand hebben bestudeerd, ontbreekt nog steeds een volledig beeld van wat er gebeurt in het cellulaire cytoplasma. Hoofdstuk 4 beoogt de eerste stap in deze richting te zetten, door OT trekexperimenten uit te voeren op eiwitten en DNA in een complexe omgeving, namelijk het verdunde *E coli*-cytosol. In onze cytosoloplossing bevinden zich chaperonnes en andere moleculen die de vouwdynamiek in fysiologische concentraties beïnvloeden. Daarom onderzoeken we het effect van de gecombineerde chaperonne-machinerie van het bacteriële cytosolische interactoom op DNA en Maltose Binding Protein (MBP), hier gebruikt als een topologische sensor. We trekken ook een parallel tussen onze computationele en experimentele onderzoekslijn, door een voorbeeld te geven van hoe het CT-formalisme kan worden gebruikt om Force Spectroscopie-gegevens te analyseren en te visualiseren. Zowel chaperonnes als DNA zijn belangrijke doelwitten voor geneesmiddelen, en het karakteriseren van hun dynamische eigenschappen en interacties in de cellulaire omgeving zou inzichtelijke informatie opleveren over de kinetiek van geneesmiddelbinding.

Eiwitten zijn niet de enige biopolymeren met complexe architecturen die fundamenteel zijn voor het uitvoeren van hun functie. Het genoom vouwt ook om te functioneren, zij het via verschillende mechanismen. De 3D-schikking van het menselijk genoom in de kern is de sleutel tot het corrigeren van genexpressie en regulatie, en een belangrijk onderdeel van epigenetische veranderingen in gezondheid en ziekte-toestanden. In Hoofdstuk 5 ontwikkelen we een multi-schaal heterogeniteitsanalyse op basis van Circuit Topology om het 3D-vouwen van eencellige genoomstructuren te karakteriseren die experimenteel zijn afgeleid door Hi-C. We laten het potentieel van CT zien voor de identificatie van statistisch verschillende topologische toestanden in de chromosomen, terwijl we ook sterk

geconserveerde motieven in chromatine-looping benadrukken. Deze onderzoekslijn zou licht kunnen werpen op de ingewikkelde relatie tussen genomische activiteit en chromosoomorganisatie, en ook een praktisch hulpmiddel kunnen bieden voor het identificeren van ziektegerelateerde structurele biomarkers.

De flexibiliteit van het CT-kader berust op de vrijheid die het biedt voor contactdefinitie, die specifiek kan worden aangepast aan het systeem zonder verlies van generaliteit. Als zodanig kan Circuit Topology worden toegepast op elk ketenachtig object waar het mogelijk is om een richting en intra- (of inter-, in het geval van multi-keten systemen) ketencontacten te definiëren. Deze polymeerachtige objecten kunnen abstract zijn en hoeven niet per se tot het biologische veld te behoren. In hoofdstuk 6 onderzoeken we een van deze toepassingen, waarbij we het CT-formalisme gebruiken om de communicatiestijl in geschreven teksten te karakteriseren in een dataset bestaande uit waarheidsgetrouwe en valse nieuwsartikelen. Hier behandelen we tekst als een keten van zinnen en definiëren contacten tussen zinnen op basis van semantische gelijkens: die zinnen die dicht bij elkaar liggen in betekenis creëren een contact in de tekstuele keten. Semantische gelijkens wordt gekwantificeerd door woord-embeddings, dit zijn numerieke attributen van tekstuele elementen verkregen uit state-of-the-art vooraf getrainde machine learning-modellen. We demonstreren dat er in onze dataset statistische variatie is in de topologie van semantische gelijkens tussen waarheidsgetrouwe en valse nieuwsberichten, overeenkomend met twee verschillende communicatiestijlen.

Tot slot trekken we in Hoofdstuk 7 onze conclusies en bespreken we mogelijke toekomstige toepassingen en kansen op basis van de huidige theoretische en computationele staat van het CT-framework. Drie belangrijke richtingen worden afgebakend, namelijk de inclusie van ruggengraat-topologie, de lokalisatie van CT-parameters en de mogelijke integratie van CT in machine learning-pijpleidingen om zijn voorspellende potentieel te benutten.