

University of Groningen

## Numerical simulations of proteins

Ramirez Palacios, Carlos

DOI:  
[10.33612/diss.196789826](https://doi.org/10.33612/diss.196789826)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2022

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Ramirez Palacios, C. (2022). *Numerical simulations of proteins: molecular dynamics, docking, and deep learning*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.  
<https://doi.org/10.33612/diss.196789826>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### 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.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Chapter 11

## Samenvatting

Computersimulaties maken de studie van bio-moleculaire systemen in tijd- en lengteschalen mogelijk die anders onbereikbaar zouden zijn door experimentele technieken. Een belangrijk aspect bij het maken van rekenmodellen is dat ze toetsbare voorspellingen moeten doen. Validatie zorgt ervoor dat de numerieke modellen de werkelijkheid weergeven en dat we een goed begrip hebben van de mechanismen waaruit waarneembare eigenschappen ontstaan. In dit proefschrift gebruik ik drie numerieke methoden om bio-moleculaire systemen (eiwitten) te modelleren: moleculaire dynamica, moleculaire docking, en diep leren. De eerste twee methoden behandelen het systeem als een verzameling deeltjes die op elkaar inwerken volgens vooraf ingestelde regels (krachtveld of energiefunctie). Diep leren is een data-gebaseerde methode en het systeem wordt behandeld afhankelijk van de architectuur en het type weergave dat wordt gebruikt. Hoewel moleculaire dynamica en docking goed beproefde methoden zijn met een overvloed aan bestaande literatuur en expertise, is diep leren een opkomend veld met het potentieel om een revolutie teweeg te brengen in de manier waarop we bio-moleculaire systemen simuleren. Kort gezegd gaat dit werk over drie computermethodes om eiwitten te modelleren die worden toegepast op katalyse. Details over experimenteel werk voor elk hoofdstuk zijn te vinden in de respectieve publicatie.

In **Hoofdstuk 2** wordt een computationeel protocol voor het voorspellen van de enantioselectiviteit van Vf TA ontwikkeld en gevalideerd op een benchmark dataset van verbindingen. We laten zien dat het modelleren van één tussenproduct van de transamineringsreactie voldoende is om nauwkeurige voorspellingen te doen over de globale transamineringsreactie. Een combinatie van docking en moleculaire dynamica maakt een betere bemonstering van de conformationele ruimte mogelijk en levert een nauwkeurig maar rekenkundig goedkoop algoritme op voor het modelleren van de Vf-TA-selectiviteit. Bovendien is aangetoond dat er een correlatie is tussen de Rosetta Interface Energie en de katalytische activiteit van Vf-TA, die zou kunnen worden benut bij inspanningen op het gebied van enzymengineering. Dit idee wordt verder onderzocht in de **Hoofdstukken 3** en **4**.

In **Hoofdstuk 3** werd de Rosetta-interface-energie gebruikt als de objectieve functie om het herontwerp van een thermostabiele Pj-TA-variant te begeleiden. De Pj-TA-mutanten werden computationeel gemodelleerd en experimentele medewerkers

voerden de laboratoriumverificatie uit van de ontworpen mutanten. De resultaten bevestigden dat de voorgestelde methode voor het modelleren van  $\omega$ -TA's eiwitengineering kan ondersteunen; niet alleen was het slagingspercentage hoog (68 van de 70 varianten vertoonden katalytische activiteit), maar ook de gemeten opbrengsten waren in overeenstemming met de verwachte waarden. Verdere rationalisatie van de structuur-activiteitsrelatie werd gepresenteerd met behulp van de gekoppelde complexen. Ten slotte werden ns-schaal moleculaire dynamica-simulaties uitgevoerd op de aangemeerde complexen om de waterverplaatsing na substraatbinding te meten, en een zwakke correlatie tussen mutanten werd waargenomen. Het doel van het werk dat in dit hoofdstuk wordt gepresenteerd is tweeledig: het valideren van een rekenkundig goedkope methode voor het herontwerpen van het substraatbereik van *Pj*-TA, en het aantonen van het potentieel van de thermostabiele *Pj*-TA-varianten in de synthese van chirale aminen.

**Hoofdstuk 4** is een klein hoofdstuk, waarin we meer datasets laten zien waarmee de enzymatische activiteit van *Vf*-TA en *Pj*-TA werd voorspeld met behulp van de Rosetta Interface Energie. Er werd een rationele analyse van de gedockte structuren uitgevoerd en we laten zien dat hoewel *Vf*-TA en *Pj*-TA vergelijkbaar zijn in de identiteit van de residuen die op de bindingsplaats worden gevonden, de conformaties die die residuen doorlopen om een nieuw substraat te accommoderen, verschillend zijn en dat is wat hun reactieprofielen anders maakt. Het werk is een verder bewijs dat *Pj*-TA het potentieel heeft van een breder substraatspectrum.

In **Hoofdstuk 5** wordt aangetoond dat het destijds nieuw ontwikkelde Martini 3 krachtveld kan worden gebruikt bij onbevooroordeelde binding van kleine organische moleculen aan de bindingsplaats van eiwitten. In het hoofdstuk wordt de binding van liganden aan twee enzymen beschreven (*Vf*-TA en Src-kinase), en het volledige manuscript bevat zeven showcase-systemen. De toepassing is opwindend omdat het de mogelijkheid opent om bindingspaden te bestuderen, die rekenkundig te duur zijn voor atomistische simulaties. Zoals vermeld in **Hoofdstuk 3**, begint er belangstelling te ontstaan voor het modelleren van niet alleen het Michaelis-complex, maar ook het proces dat leidt tot zijn vorming en de rol van het oplosmiddel in dat proces. Bovendien geeft het aan dat het Martini 3-krachtveld nauwkeurig genoeg is om interacties tussen kleine organische moleculen en eiwitten te modelleren. Het waarborgen van de nauwkeurigheid van Martini is belangrijk omdat het vertrouwen geeft dat simulaties van veel grotere systemen (bijvoorbeeld een hele cel) de werkelijkheid vertegenwoordigen.

**Hoofdstuk 6** is een klein hoofdstuk, waarin de moleculaire modellering van weer een ander PLP-afhankelijk enzyme wordt getoond, een diaminopimelaatdecarboxylase (DAPDCs) van *Thermotoga maritima*, om de activiteitsprofielen te rationaliseren. Sommige van de opnieuw ontworpen DAPDC's katalyseerden de omzetting van 2-

aminopimelzuur naar 6-aminocapronzuur, een bouwsteen in de synthese van nylon-6. De uitdagingen bij het modelleren van decarboxylasen waren dat er nauwelijks publicaties over de structuur-activiteitsrelatie beschikbaar waren en dat het reactiemechanisme niet volledig is opgehelderd. We veronderstelden dat de  $C_{\alpha}$ - $CO_2^-$  splitsing de beperkende stap in de reactie zou zijn, en modelleerden dus het externe aldimine-tussenproduct van de decarboxylering. De verschuiving in katalytische activiteiten van de E315X-mutanten werd gerationaliseerd op basis van het optreden van bindende houdingen die worden gevonden in moleculaire dynamische trajecten. Er wordt enige correlatie aangetoond tussen de frequentie van reactieve bindingshoudingen en enzymactiviteit, maar er is ruimte voor verbetering. In **Hoofdstuk 9** wordt aangetoond dat de analyse van moleculaire dynamica-trajecten kan worden gedaan met zeer weinig handmatige tussenkomst en kan helpen bij het definiëren van interessante bindingsposities.

**Hoofdstuk 7** gaat over de moleculaire modellering van een *N*-glycosyltransferase van *Actinobacillus pleuropneumoniae* (ApNGT). Dit hoofdstuk is sterk gebaseerd op resultaten van experimentele technieken. De experimentele medewerkers hadden vastgesteld dat ApNGT de glycosylering van adhesinefragmenten op een semiprocessieve manier katalyseert, en moleculaire modellering werd gebruikt om te proberen het mechanisme te beschrijven waaruit processiviteit voortvloeit. De uitdagingen bij het modelleren van ApNGT waren dat er geen kristalstructuur in complex met glucose of het peptide (adhesinefragment) bekend is en dat het reactiemechanisme niet goed wordt begrepen. Rosetta werd gebruikt om de eiwit-peptide- of ligand-peptidecomplexen te modelleren. We veronderstelden dat de processiviteit voortkomt uit de ondiepe substraatbindingsgroef in ApNGT, en een schuifmechanisme wordt voorgesteld.

In **Hoofdstuk 8** wordt aangetoond dat het mogelijk is om een convolutioneel neuraal netwerk van een spectrale grafiek te trainen om de bindingsenergieën van combinatorische bibliotheken van *Vf*-TA-mutanten te voorspellen. Het neurale netwerk werd getraind op een kleine dataset van mutanten met bindingsenergieën verkregen via traditionele moleculaire modelleringsmethoden. Binnen een paar uur leerde het neurale netwerk de ingewikkelde synergetische interacties tussen mutatieplaatsen en werd het nauwkeurig in het evalueren van onzichtbare varianten. Het nut van deze benadering is dat het getrainde neurale netwerk verschillende ordes van grootte sneller is dan de traditionele moleculaire modelleringsmethode bij het screenen van nieuwe varianten, waardoor de tijd die nodig is om één variant te scannen van ~20 minuten tot <1 milliseconde wordt gebracht. De korte evaluatietijd maakt het mogelijk om de volledige zoekruimte te screenen. Verder is aangetoond dat de nauwkeurigheid van de voorspellingen verder kan worden vergroot door injectie van functie-inbeddingen verkregen uit een vooraf getraind taalmodel. Voor zover ik weet,

is dit de eerste keer dat een dergelijke benadering wordt gerapporteerd in combinatorische bibliotheken.

In **Hoofdstuk 9** worden de moleculaire dynamische trajecten gegenereerd in **Hoofdstuk 2** opnieuw bekeken om het potentieel aan te tonen van het gebruik van neurale netwerken om trajecten te analyseren. Vergelijkbaar met wat werd gepresenteerd in **Hoofdstuk 2**, hebben we een reeks simulatietrajecten en willen we deze classificeren in "reactief" of "niet-reactief", afhankelijk van het enantiomeer dat de bindingsplaats van het enzym bezet. De getrainde modellen bereikten een hoge nauwkeurigheid bij het onderscheiden van de twee klassen. Dit hoofdstuk is demonstratief in plaats van toepassingsgericht. Ik laat zien dat het gebruik van een neuraal netwerk om geometrische criteria van belang te identificeren, de handmatige interventie die nodig is om strategieën zoals die uit **Hoofdstuk 2** of **6** uit te voeren, zou kunnen verminderen.

Zoals de lezer kan zien, was het gemeenschappelijke thema in de hoofdstukken het modelleren van eiwitten door middel van numerieke methoden. Het promotiewerk was niet beperkt tot een bepaald onderwerp of onderzoeksvraag, maar eerder om de mogelijkheden van computationele modellering te verkennen. Het merendeel van de publicaties is het resultaat van samenwerkingen met experimentele medewerkers, wat in hoge mate dient als validatie dat de computermethodes de werkelijkheid vertegenwoordigen. Over het algemeen was het doel van experimentele medewerkers om het enzym aan te passen om een gewenst substraat te accepteren, en mijn doel was om algoritmen te ontwerpen om de uitkomst van de enzymatische reactie te voorspellen en te modelleren. Verderop in het proefschrift maakt de lezer kennis met machinale leermethoden om eiwitten te modelleren of te helpen bij het modelleren ervan.

Numerieke simulaties van moleculaire systemen worden steeds beter in het reproduceren van de werkelijkheid. Dit is mogelijk gemaakt door zowel de toename van rekenkracht in de afgelopen decennia, de ontwikkeling van efficiëntere algoritmen en een beter begrip van de mechanismen die moleculaire interacties sturen. Onderzoeksgroepen die moleculaire systemen onderzoeken, zijn steeds meer geïnteresseerd in het reproduceren van hun werk via computersimulaties. Desalniettemin is het belangrijk om te blijven innoveren, en ik zou zeggen dat de recente ontwikkelingen op het gebied van machinaal leren dit enorm zullen vergemakkelijken, en zelfs het zetje kunnen zijn dat nodig is om de volgende stap in de simulatie van de werkelijkheid te bereiken.