



ACADEMIA ROMÂNĂ
Școala de Studii Avansate a Academiei Române
Institutul de Biochimie

REZUMATUL TEZEI DE DOCTORAT

**TEHNICI DE BIOINFORMATICĂ ȘI BIOCOMPUTING
PENTRU ANALIZA ȘI DEZVOLTAREA INTERACȚIILOR
LIGAND-COMPUS BIOACTIV**

CONDUCĂTOR DE DOCTORAT:
Dr. Andrei-Jose Petrescu

DOCTORAND:
Șulea Teodor Asvadur

Cuprins

PARTEA I-a – INTRODUCERE GENERALĂ	4
CAPITOLUL 1 – scopul Lucrării	4
PARTEA II-a – PREZICEREA ȘI MODELAREA STRUCTURII PROTEINELOR	7
CAPITOLUL 2 – EVALUAREA LIMITELOR NOILOR METODE DE PREDICȚIE STRUCTURALĂ BAZATE PE INTELIGENȚĂ ARTIFICIALĂ	7
1. Introducere	7
2. Familia extinsă a A. Thaliana	9
3. ZAR1 – Comparația dintre modele și structura rezolvată	9
4. Analiza modelelor setului extins de CNL-uri	11
5. Corectarea rezultatelor predicției prin preprocesarea intrărilor în sisteme cu mai multe stări	12
6. Concluzii	13
PARTEA III-a – DINCOLO DE PREDICȚIA STRUCTURALĂ – EȘANTIONAREA CONFORMAȚIONALĂ PRIN SIMULĂRI MOLECULARE	14
CAPITOLUL 3 – EȘANTIONARE ÎMBUNĂTĂȚITĂ PRIN UTILIZAREA EȘANTIONĂRII GIBBS PE GRADELE DE LIBERTATE INTERNE	14
1. Materiale și metode	14
2. Rezultate și discuții – Dipeptida de alanină	15
3. Rezultate și discuții – Glycoproteina E2N1	17
4. Concluzii	18
PARTEA IV-a – APLICAREA MODELĂRII ȘI SIMULĂRII PENTRU STUDIEREA INTERACȚIUNILOR MOLECULARE COMPLEXE	19
CAPITOLUL 4 – MODELAREA COMPLEXĂRII RECEPTORULUI TROMBOPOIETINEI DE CĂTRE CALRETICULINA MUTANTĂ	19
1. Introducere	19
2. Modelarea prin omologie a TpoR	19
3. Evaluarea interacțiunii CRM1 a TpoR cu calreticulina mutantă	21
4. Concluzii	22
CAPITOLUL 5 – OXIDAREA EPITOPULUI TIROZINAZEI ȘI EFECTUL ACESTEIA ASUPRA REACTIVITĂȚII CELULELOR T ÎN MELANOM	23
1. Introducere	23
2. Metode – Modelarea complexului ternar, generarea de parametrii pentru YM2M6D, HREX	23
3. Modelarea complexului ternar HLA/YMD/TCR	24
4. Calculul energiei libere de legare	25
5. Concluzii	27
CAPITOLUL 6 – INTERACȚIUNEA CANALULUI VOLTAJ-DEPENDENT NAV1.5 CU CENOBAMATUL	28
1. Introducere	28
2. Generarea modelului	28
3. Rezultate – Modelare și simulare	29

4. Rezultate – Calculul de energie liberă de legare	30
5. Concluzii	30
Concluzii generale	30
Contribuții Personale	32
Referințe	33

PARTEA I-A – INTRODUCERE GENERALĂ

CAPITOLUL 1 – SCOPUL LUCRĂRII

Dezvoltările tehnologice în generarea datelor biologice, precum și soluțiile mai mari de stocare a datelor și gestionarea mai rapidă a acestora au fost un beneficiu al secolului XXI (Carugo and Djinović-Carugo, 2023). Cercetătorii au acum sarcina copleșitoare de a analiza această cantitate masivă de date, pentru care trebuie dezvoltate metode avansate (și făcute cât mai ușor de utilizat posibil) pentru a transforma datele în cunoștințe utile. Începând cu mijlocul deceniului 2010, noile tehnologii de cryo-EM au revoluționat domeniul și au permis obținerea unor informații structurale despre sisteme biomoleculare din ce în ce mai complexe; și, deoarece interpretarea datelor cryo-EM depinde în mod critic de modelele moleculare precise, a doua revoluție a venit la începutul anilor 2020 sub forma platformelor de modelare automate bazate pe AI – o schimbare fără precedent în domeniu. Având în vedere acest context, scopul lucrării prezentate a fost:

- a) de a examina limitele și de a îmbunătăți capabilitățile noii generații de tehnologii AI în predicția structurală a proteinelor;
- b) de a contribui, dincolo de predicția structurală, la dezvoltarea de noi instrumente pentru explorarea spațiului conformațional al sistemelor moleculare – relevante pentru descrierea interacțiunilor și proceselor biomoleculare în condiții biologice normale
- c) de a dezvolta fluxuri de lucru computaționale pentru rezolvarea unor probleme structurale complexe legate de mai multe complexe proteine-proteine și proteine-ligand de relevanță medicală.

Lucrarea care acoperă aceste obiective este structurată în cinci capitole grupate în trei părți după cum urmează: Prima parte, care cuprinde două capitole, se concentrează pe modelarea moleculară in-silico și rezultatele obținute pe unele familii relevante de proteine folosind noile metode de AI, atât în versiunea lor automată, cât și personalizată. Această primă parte începe cu un scurt capitol teoretic, care prezintă tehnicile de modelare prin omologie, atât cele „personalizate” cât și cele mai recente bazate pe „rețele neuronale”. Al doilea capitol al acestei părți este dedicat unei cercetări aprofundate a capabilităților și limitelor noilor metode AI, care au devenit uzuale în ultimii ani ai studiilor mele de doctorat și care au primit recent Premiul Nobel în 2024. Deși aceste metode sunt cunoscute pentru performanțele lor bune în cazuri simple, le-am testat în cazul mai provocator al proteinelor multidomeniu și multistare, luând drept exemplu familia de receptori cu coiled-coil a

domeniului de oligomerizare legat de nucleotid (NOD-like), despre care grupul nostru (Departamentul de Bioinformatică și Biochimie Structurală – DBSB) a acumulat o vastă expertiză în ultimii 15 ani. Rezultatele noastre indică faptul că prin ajustarea fină a ceea ce a fost introdus ca intrare în rețeaua neuronală, platforma AlphaFold a reușit să modeleze distinct formele „active” și „inactive” ale proteinei. Acest rezultat indică faptul că metoda noastră restrictivă de intrare ar putea fi utilă în proteinele multidomeniu și multistare pentru a ghida fluxul automat de modelare către conformații specifice pentru proteinele pentru care există multiple conformații rezolvate.

În timp ce modelarea predictivă structurală este indispensabilă ca o descriere preliminară virtuală a sistemelor biomoleculare, aceasta nu este suficientă pentru a explica interacțiunile și procesele care au loc în aceste sisteme, care în condiții biologice naturale sunt ghidate de legile termodinamicii. Având în vedere că moleculele sunt într-o schimbare termică perpetuă, dezvoltarea de tehnici noi și mai rapide pentru a descrie posibilele lor stări este de mare importanță. În acest context, a doua parte a tezei mele descrie contribuțiile noastre la dezvoltarea Robosample, o nouă platformă software de eșantionare moleculară, bazată pe o metodă publicată anterior care se încadrează în categoria „Enhanced Sampling” și care, având în vedere dimensiunea tot mai mare a sistemelor structurale rezolvate (Berman, Vallat and Lawson, 2020), a fost utilizată tot mai mult în ultima decadă (Shen, Zhou and Shi, 2023). Ideea de eșantionare moleculară bazată pe algoritmi robotici a apărut din colaborarea productivă dintre Dr. Laurențiu Spiridon din departamentul nostru (DBSB-IBAR) și Prof. David Minh de la Illinois Institute of Technology și director asociat al Center for Interdisciplinary Scientific Computation, colaborare care a dus la dezvoltarea metodei originale.

Cel mai semnificativ beneficiu al eșantionării spațiului conformațional al sistemelor moleculare la temperatura camerei este posibilitatea de a estima precis energia liberă de legare a complexelor moleculare. De aceea, ultima parte a tezei este dedicată rezolvării a trei probleme complexe relevante în medicina moleculară prin combinarea modelării și simulării pentru calcularea energiilor libere de legare ale complexelor proteine-peptid și proteine-ligand în sisteme de relevanță medicală. Alături de pașii de modelare și simulare, această lucrare se bazează pe un set divers de tehnici de calcul al energiei libere, care variază de la metode de punct final (MM-GBSA) la metode bazate pe transformări alchimice (HRE/MBAR).

Primul capitol al acestei părți este dedicat modelării receptorului de trombopoietină (TpoR – necunoscută la acel moment) și interacțiunii acestuia cu calreticulina mutată, o

interacțiune care s-a dovedit responsabilă pentru ~40% din neoplasmele mieloproliferative. Această lucrare este rezultatul unei colaborări îndelungate între DBSB-IBAR și grupul de semnalizare celulară de la Université Catholique de Louvain, condus de Prof. Ștefan N. Constantinescu, care a furnizat constrângerile experimentale HDX și a validat predicțiile noastre structurale, predicții care au fost confirmate recent prin rezultate cryo-EM pe structura TpoR. (Tsutsumi *et al.*, 2023a; Sarson-Lawrence *et al.*, 2024).

Aceasta este urmată de un capitol care descrie utilizarea acestor calcule ale energiei libere în imunobiologie într-o lucrare care a rezultat din colaborarea departamentului nostru DBSB cu Departamentul de Biologie Moleculară și Celulară al IBAR, legată de recunoașterea de către receptorul T-celular (TCR) a unui sistem HLA încărcat cu epitopul tirozinazei YMD (369-YMDGTMSQV-377), care a fost dovedit a fi relevant în melanom. În acest studiu, am evaluat mai întâi formarea complexului HLA-YMD și apoi am observat complexul ternar care se formează atunci când sistemul HLA-YMD interacționează cu regiunea hiper-variabilă a TCR, fiind utilizate metode alchimice pentru a eșantiona spațiul său conformațional și utilizând Multistate Bennett Acceptance Ratio pentru a calcula energia liberă de legare.

Ultimul capitol din această parte se concentrează pe utilizarea estimărilor energiei libere în farmacologie și este dedicat unei lucrări realizate în cadrul colaborării dintre departamentul nostru DBSB și grupul Prof. Bogdan Amuzescu de la Facultatea de Biologie a Universității din București, pe interacțiunea canalului de ioni NaV 1.5 cu cenobamat, o moleculă mică utilizată ca medicament antiepileptic. În această lucrare, modelarea canalului de ioni a fost realizată de echipa Prof. Amuzescu, precum și docking-ul inițial al liganului cenobamat pe canalul de ioni, în timp ce noi am realizat generarea structurilor mutante, construcția sistemului (adică construirea bilei lipidice, încorporarea diferitelor canale în aceasta, parametrizarea sistemului) și simulările de dinamică moleculară. Folosind traiectoria generată, energiile libere de legare au fost calculate între formele native/mutante ale receptorului și molecula de cenobamat utilizând cadre la intervale egale din simularea MD.

PARTEA II-A – PREZICEREA ȘI MODELAREA STRUCTURII PROTEINELOR

CAPITOLUL 2 – EVALUAREA LIMITELOR NOILOR METODE DE PREDICȚIE STRUCTURALĂ BAZATE PE INTELIGENȚĂ ARTIFICIALĂ

1. Introducere

Biologia structurală a cunoscut un impuls transformator la începutul anilor 2020, odată cu apariția tehnicilor automate de modelare pe bază de Deep Learning, cum ar fi AlphaFold2 și 3 (AF2/3) (Jumper *et al.*, 2021; Abramson *et al.*, 2024), RoseTTAFold All-Atom (Krishna *et al.*, 2024) (RFAA) sau OmegaFold (Wu *et al.*, 2022) (OF) care au depășit metodologiile tradiționale de modelare prin omologie. Metodele AF și RFAA se bazează în esență pe aceiași pași ca modelarea tradițională prin omologie, care includ alinierea secvenței multiple (MSA), căutarea de șabloane și corelarea acestora în construirea de modele ale unei secvențe țintă pe baza unui antrenament masiv al rețelelor neuronale, având ca scop optimizarea structurii locale și a contactelor aminoacidice. Pe de altă parte, OmegaFold adoptă o abordare diferită, prezicând structura unei proteine direct din secvența sa primară unică, utilizând un model de limbaj proteic bazat pe un transformator inspirat de geometrie. Unul dintre cele mai importante avantaje ale acestor metode este viteza cu care sunt propuse modele structurale pentru o secvență dată. În câteva ore pentru AF și RFAA sau chiar în câteva minute pentru OF, oricine poate obține modele prezise pentru o secvență dată. Multe recenzii subliniază, de asemenea, precizia mare a acestor modele automate, în special în cazul țintelor cu un singur domeniu, cu similaritate foarte mare față structuri experimentale deja rezolvate. Pe de altă parte, aceste proceduri automate generează modele tip „cutie neagră”, care nu permit nicio intervenție umană bazată pe constrângeri experimentale suplimentare sau informații, sau pe intuiția și experiența dobândită a unui cercetător care, în special în cazurile de omologie îndepărtată, ar putea fi relevante.

Prin contrast, modelarea manuală tradițională prin omologie este mult mai consumatoare de timp și poate duce la structuri mai puțin optimizate, dar are avantajul de a fi mult mai flexibilă, permițând cercetătorilor, pe de o parte, să ia în considerare constrângeri experimentale sau orice alte observații experimentale și, pe de altă parte, să îmbunătățească modelele prin testarea ipotezelor experimentale într-o manieră de încercare și eroare, fiind

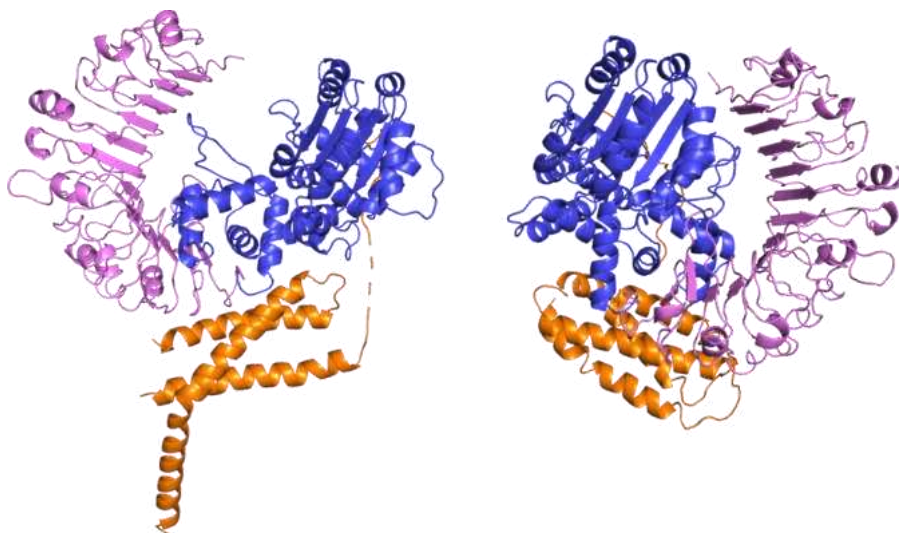
astfel ghidată de o abordare euristică. Astfel, modelarea euristică este mai probabil să abordeze mai bine țintele de omologie îndepărtată sau sisteme mai complexe, cum ar fi structurile proteice dependente de stare, proteinele cu mai multe domenii sau sistemele heteromoleculare.

În acest context, având în vedere că aproximativ 60% din proteinele cunoscute cuprind secvențe proteice complexe care prezintă mai mult de un domeniu structural, este foarte relevant să determinăm cât de bune sunt modelele generate de noile platforme de predicție structurală bazate pe AI pentru astfel de sisteme și să înțelegem mai bine limitările acestora în generarea unor modele structurale plauzibile și să găsim modalități de a le spori capacitățile de predicție, mai ales atunci când astfel de proteine au demonstrat că adoptă multiple conformații pe parcursul ciclului lor funcțional. Cu alte cuvinte, este foarte instructiv să înțelegem cum funcționează predictorii automați în cazurile în care în bazele de date structurale aceeași proteină a fost rezolvată în mai multe stări structurale și modelarea se confruntă cu o problemă de mapare a unei singure secvențe la mai multe structuri.

Un studiu de caz bun pentru a evalua performanța noilor platforme AI pe sisteme proteice mai complexe, în cazul cărora modelarea euristică s-a dovedit a fi eficientă, este familia de proteine multistare și multidomeniu de receptori CNL NOD-Like (Coiled-Coil Nucleotide-Binding Oligomerization Domain-Like). În esență, aceste proteine constau din trei domenii canonice cu extensii N-terminale sau C-terminale:

- (1) un „conector” CC (Coiled Coil)
- (2) un „comutator” NBD (Nucleotide Binding Domain)
- (3) un „senzor” LRR (Leucine-Rich Repeat)

La rândul lor, la activare (indusă de un efector molecular al unui patogen), comutatorul NBD suferă o tranziție conformațională internă în care subdomeniul Arc2 se rotește la 180° în jurul regiunii NBS-Arc1, prin eliberarea cofactorului ADP (specific stării inactivate) și schimbarea acestuia cu un ATP care stabilizează conformația activă.



Figură 2-1. Conformații active (6J5T) și inactive (6J5W) ale proteinei ZAR1

2. Familia extinsă a *A. Thaliana*

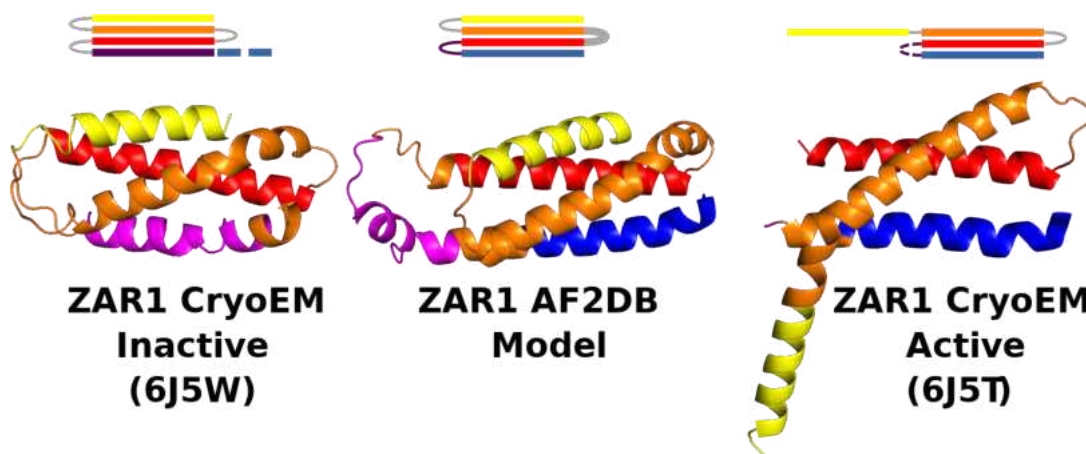
Pentru a testa performanța modelării asistate de inteligență artificială, am selectat un set reprezentativ de proteine CNL din *Arabidopsis thaliana*. Un total de 1257 de secvențe au fost extrase din baza de date NLRscape (Martin *et al.*, 2023), pe care le-am filtrat în funcție de organizarea domeniilor, eliminând toate secvențele care nu conțineau toate cele 9 motive CNL. Setul rezultat a fost grupat folosind MMSeqs2 la un prag de acoperire de 70% și identitate de 70%, rezultând 36 de grupuri, dintre care 4 au fost eliminate.

Cele 32 de secvențe reprezentative au fost apoi modelate folosind versiunile instalate local ale AlphaFold2 (AF2) și OmegaFold (OF); precum și online, utilizând AlphaFold3 (AF3) și RoseTTAFold All-Atom (RFAA) prin intermediul serverelor NeuroSnap. În plus, modelele acestor secvențe au fost recuperate din AlphaFold Database (AFDB) pentru comparație (Varadi *et al.*, 2024). Având în vedere că noile implementări web ale AF3 și RFAA permit modelarea proteinelor împreună cu liganzii lor, le-am utilizat pentru a modela cele 32 de proteine CNL atât în formă liberă, cât și în interacțiune cu ADP și ATP. Toate structurile au fost minimizate energetic folosind OpenMM (Eastman *et al.*, 2024). Calitatea modelelor a fost evaluată folosind MolProbity (Williams *et al.*, 2018). Modelele generate cu RFAA au avut o calitate foarte slabă, iar încercările de îmbunătățire prin recoacere simulată nu au avut succes. Totuși, aceste modele au fost stabile și s-au menținut într-un bazin conformațional de maximum 2Å.

3. ZAR1 – Comparația dintre modele și structura rezolvată

Deoarece structura proteinei ZAR1 a fost deja rezolvată experimental, aceasta oferă un punct de referință valoros pentru a evalua direct acuratețea programelor de modelare

menționate anterior. Domeniul CC a fost caracterizat atât în conformația sa „activă”, cât și în cea „inactivă”. Așa cum este prezentat în Figură 2-2 , cele două forme prezintă diferențe structurale semnificative. Totuși, predicțiile implicite generate de AF2, AF3, RoseTTAFold All-Atom și OmegaFold tind să producă modele hibride, care combină elemente din ambele conformații, rezultând astfel în structuri implauzibile.



Figură 2-2. Comparație între forma activă și cea inactivă rezolvate prin cryo-EM și modelul obținut de pe baza de date AlphaFold (AFDB)

Model Name	CC	NBS-Arc1	ARC2	LRR
AF2 – Database	14.46	1.86	1.63	1.45
AF2 – Active Control	1.39	0.49	0.42	0.53
AF2 – Inactive Control	18.51	1.86	1.51	0.84
AF2 – Active MSA	1.55	0.59	0.61	0.96
AF2 – Inactive MSA	14.40	1.79	1.53	0.97
AF3 – ADP	14.41	1.80	1.64	1.42
AF3 – ATP	14.39	1.89	1.58	1.33
AF3 – No Ligand	14.45	1.92	1.59	1.51
RFAA – ADP	14.42	2.04	1.64	1.86
RFAA – ATP	14.50	2.05	1.78	2.02
RFAA – No Ligand	14.45	1.92	1.65	1.96
OmegaFold	14.48	1.87	1.33	3.11

Tabel 2-1. RMSD-ul (Å) pentru fiecare domeniu, calculat față de structura rezolvată a stării “active”

Model Name	CC	NBS-Arc1	Arc2	LRR
AF2 – Database	12.42	1.33	0.88	1.34
AF2 – Active Control	19.19	1.89	1.45	0.78
AF2 – Inactive Control	0.80	0.69	0.35	0.47
AF2 – Active MSA	18.48	1.88	1.51	0.90
AF2 – Inactive MSA	12.54	0.86	0.44	0.75
AF3 – ADP	11.99	1.31	0.98	1.22
AF3 – ATP	12.11	1.35	0.87	1.16
AF3 – No Ligand	12.08	1.26	0.93	1.37
RFAA – ADP	12.74	1.41	1.02	1.77
RFAA – ATP	12.65	1.53	1.04	1.92
RFAA – No Ligand	12.36	1.42	1.11	1.89
OmegaFold	12.99	1.24	0.87	3.04

Tabel 2-2 RMSD-ul ($\text{\AA}$) pentru fiecare domeniu, calculat față de structura rezolvată a stării “inactive”

Model Name	RMSD vs active	RMSD vs inactive
AF2 – Database	22.158	6.004
AF2 – Active Control	0.832	23.036
AF2 – Inactive Control	22.612	0.675
AF2 – Active MSA	1.271	22.873
AF2 – Inactive MSA	21.997	5.837
AF3 – ADP	22.119	5.785
AF3 – ATP	22.076	5.879
AF3 – No Ligand	22.191	5.889
RFAA – ADP	22.406	6.756
RFAA – ATP	22.646	6.792
RFAA – No Ligand	22.424	6.027
OmegaFold	22.374	6.442

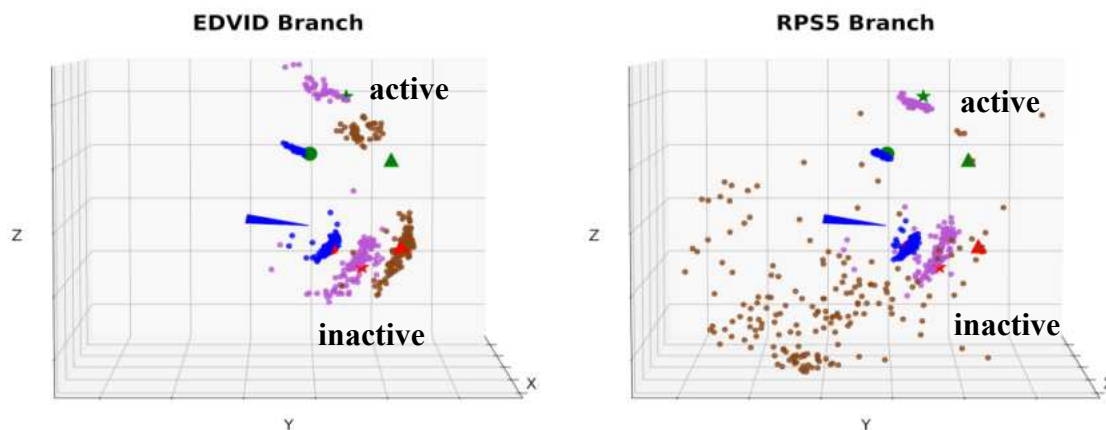
Tabel 2-3. RMSD global ($\text{\AA}$) pentru modelele cu cel mai mare scor din fiecare metodă de generare, în raport cu structura cristalografică activă/inactivă

La compararea valorilor RMSD globale, reiese clar tendința modelelor AI de a reprezenta proteinele CNL în starea „inactivă”, care este mai compactă. Această preferință ar putea orienta rețeaua neuronală către forma ce minimizează suprafața expusă și este energetic mai favorabilă. În plus, modelele prezintă o tendință de supracompactare a structurii, observabilă prin legarea mai strânsă dintre domenii, conform măsurătorilor realizate cu ajutorul software-ului Prodigy. (Vangone and A. M. Bonvin, 2015; Xue *et al.*, 2016)

4. Analiza modelelor setului extins de CNL-uri

Celelalte secvențe CNL au fost modelate utilizând aceeași procedură descrisă pentru ZAR1 mai sus. Având în vedere că sistemul are multiple stări posibile, a fost important să studiem conformația generată de fiecare software pentru fiecare secvență. Am mapat poziția centrului de masă al domeniilor CC, Arc2 și LRR, precum și a planului format de motiful

„VG” (începutul domeniului NBS), centrele de masă ale domeniilor NBS și Arc1 (planul de origine). Folosind această reprezentare 3D, putem distinge cu succes conformațiile „active” de cele „inactive”. Această distribuție este prezentată în Figură 2-3, pentru un subset de modele.



Figură 2-3. Distribuția domeniilor CC (puncte maronii), Arc2 (puncte albastre) și LRR (puncte maronii) în raport cu aceleași domenii ale structurilor experimentale (triunghiul, cercul și steaua verde reprezintă CC activ, ARC2 și LRR, respectiv; triunghiul, cercul și steaua roșie reprezintă CC inactiv, ARC2 și LRR, respectiv).

Așa cum se prezintă, toate modelele de pe ramura „EDVID” adoptă fie o conformație „inactivă”, fie una „activă”, cu o tendință puternică către modelul compact „inactiv”. De asemenea, modelarea în prezența ATP nu generează conformația „activă”, indicând faptul că specificitatea ligandului nu este luată în considerare de niciunul dintre AF3 sau RFAA. În contrast, modelele de pe ramura „RPS5” prezintă o distribuție foarte diferită și mult mai dispersată decât ramura „EDVID”. O analiză detaliată a ramurii „RPS5” arată că domeniul CC nu interacționează cu domeniile NBD sau LRR, indicând un mecanism de semnalizare diferit. Pentru proteinele CNL din alte familii (*Solanum tuberosum*, *Triticum aestivum* și *Hordeum vulgare*), tendința pentru forma „inactivă” rămâne valabilă.

5. Corectarea rezultatelor predicției prin preprocesarea intrărilor în sisteme cu mai multe stări

Pentru a ghida rețeaua către o conformație plauzibilă și pentru a o îndepărta de structura „chimerică” a domeniului CC prezentată mai sus, am îmbogățit/restricționat datele de intrare din care AlphaFold2 extrage caracteristici. În primul rând, adăugând structuri recente de proteine Nod-like receptor din PDB și clasificându-le ca „active” sau „inactive”, am reușit să ghidăm rețeaua către una dintre aceste conformații. În al doilea rând, având în vedere că atât domeniile CC, cât și cele LRR au arhitecturi bine definite – desemnări CATH (Waman *et al.*, 2024) 1.20 Up-Down Bundle și 3.80.10 Leucine-Rich Repeat, respectiv – baza de date PDB locală a AF2 a fost filtrată pentru a conține doar proteine care conțin cel

puțin unul dintre aceste domenii, doar acest set redus fiind utilizat. Mai mult, având în vedere că MSA generat ar trebui să reflecte caracteristici specifice proteinelor NLR, au fost utilizate doar secvențe NLR ca date de intrare.

În ceea ce privește secvența ZAR1, modelele generate cu aceste metode de filtrare a datelor de intrare sunt mult mai apropiate de structurile cristalizate, măsurate prin RMSD (Tabel 2-1 și Tabel 2-2): domeniul CC este modelat conform conformației active în ambele seturi „active”, în timp ce forma „inactivă” este corect modelată atunci când nu sunt alimentate date MSA în rețea (Control inactiv) – așa cum este vizibil din valorile globale RMSD (Tabel 2-3). Totuși, toate celelalte domenii au un RMSD mai mic față de structurile lor rezolvate experimental decât are orice alt model generat față de structurile rezolvate.

Setul extins de CNL a beneficiat, de asemenea, de această filtrare a datelor de intrare, deoarece pentru toate secvențele au fost construite modele distincte „active” și „inactivă”.

6. Concluzii

La nivelul domeniilor, structurile pliate ale modulelor NBD-Arc1, Arc2 și LRR, care aparțin topologiilor CATH 3.40.50.300, 1.10.533.10 și 3.80.10, respectiv, au fost prezise cu o precizie de mai puțin de 2 Å de toate platformele evaluate. Cu toate acestea, în regiunea CC, RMSD-ul față de structura experimentală este mai mare de 12 Å. Analiza soluțiilor furnizate de platformele AI pentru această regiune sugerează că acestea combină informații structurale din ambele structuri ADP-Inactive și ATP-Active ale ZAR1. La nivel global, s-a demonstrat că metodele bazate pe AI favorizează starea mai compactă și inactivă. Doar prin filtrarea selectivă a datelor de intrare cu date structurale specifice, au putut fi generate modele pentru starea activă a tuturor proteinelor CNL. Este de remarcat faptul că includerea (sau excluderea) ligandului nu a favorizat rețeaua pentru nicio conformație, ceea ce indică faptul că procesele AF3 și RFAA iau în considerare poziția ligandului doar după modelarea receptorului înconjurător, ignorând astfel orice efect „indus fit”. Incorporarea datelor despre ligand în aceste procese reprezintă un pas important în modelarea proteinelor cu multiple stări metastabile. Lucrarea descrisă aici conturează un protocol pentru ghidarea modelării rapide bazate pe AI către conformațiile dorite, fără costul computațional al reantrenării unei rețele neuronale. Vorbind despre reantrenare, este important de menționat că AlphaFold (sau mai exact DeepMind) nu furnizează scripturi de antrenament sau date de antrenament, ceea ce face imposibilă generarea de rețele antrenate specific. În acest sens, au existat eforturi pentru deschiderea acestor tipuri de software, cum ar fi OpenFold (Ahdriz *et al.*, 2024) care permite reantrenarea și produce rezultate comparabile cu AlphaFold2.

În plus, acesta oferă o metodă eficientă din punct de vedere computațional de generare a proteinelor în conformații multimerice (precum conformația „activă” CNL), fără costul computațional al rulării AlphaFold2-Multimer, care pentru modelarea unei structuri pentamerice ar fi prohibitiv de ridicat.

PARTEA III-A – DINCOLO DE PREDICȚIA STRUCTURALĂ – EȘANTIONAREA CONFORMAȚIONALĂ PRIN SIMULĂRI MOLECULARE

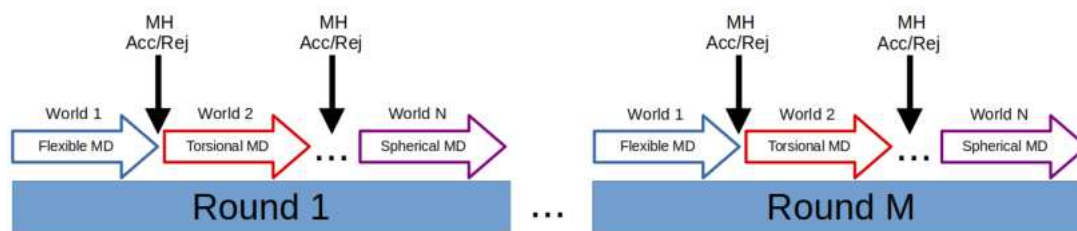
CAPITOLUL 3 – EȘANTIONARE ÎMBUNĂȚITĂ PRIN UTILIZAREA EȘANTIONĂRII GIBBS PE GRADELE DE LIBERTATE INTERNE

1. Materiale și metode

Software-ul descris în această parte este Robosample, dezvoltat de Spiridon et al. (Spiridon *et al.*, 2020). Acesta aplică algoritmi de dinamică constrânsă de înaltă performanță pentru simularea sistemelor biologice relevante. Folosește un amestec de Gibbs sampling blocat cu HMC – Constrained Dynamics Hamiltonian Monte Carlo (CDHMC) – pentru a explora eficient spațiul conformațional.

Utilizează coordonate interne de legătură/unghi/torsiune (Bond/Angle/Torsion - BAT) pentru a explora mult mai eficient spațiile conformaționale ale sistemelor biologice complexe decât simulările tradiționale MD. Articolul original despre Constrained Dynamics Hamiltonian Monte Carlo (CDHMC) (Spiridon and Minh, 2017) detaliază modul în care utilizarea Gibbs sampling împreună cu coordonatele BAT permite recuperarea corectă a distribuției Boltzmann.

În Robosample, atomii sau grupurile de atomi sunt grupate pentru a forma „corpuri rigide”, ceea ce înseamnă că niciuneia dintre coordonatele lor interne nu îi este permis să varieze. Corpuri rigide diferite sunt legate între ele prin diferite articulații (precum Pin, Spherical sau Cartesian). În acest fel, un grafic robotic (corpuri rigide și diferite tipuri de articulații) este suprapus peste graficul chimic (atomi și legături). Prin utilizarea diferitelor combinații de definiții ale corpurilor rigide și tipuri de articulații, se pot explora diferite subseturi ale spațiului conformațional. Prin explorarea împreună a gradelor de libertate corelate, se poate obține o explorare mai eficientă a spațiului conformațional.



Figură 3-1. Reprezentare schematică a unei simulări Robosample. Formată din M runde, fiecare cu N lumi.

Pentru a evalua eficiența Robosample, am realizat simulări pe două sisteme moleculare: dipeptidul de alanină și un model mai complex care include primul polizaharid al proteinei E2 din virusul hepatitei C (E2N1). În acest model, polizaharidul este atașat unui peptid derivat din E2 [-12 + 6] la locul ASN17. Se crede că acest polizaharid joacă un rol în protejarea virusului de sistemul imunitar al gazdei prin întârzierea sau reducerea recunoașterii de către anticorpii anti-E2 (Prentoe *et al.*, 2019).

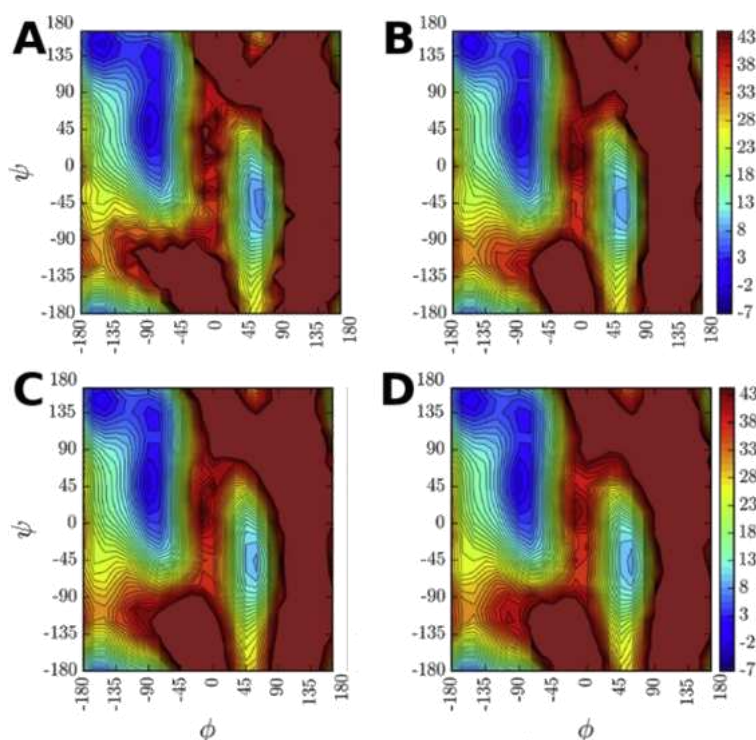
Sistemul dipeptidului de alanină a fost simulat în mai multe combinații de blocuri Gibbs:

- Toate legăturile rotabile au fost făcute flexibile, cu articulații Pin (TD), Cylinder (CYL) sau Ball – „Toate Flexibile”
- Doar legăturile N-C α și C α -C au fost făcute flexibile, cu articulații Pin (TD), Cylinder (CYL) sau Ball – „Dinamică Ramachandran”

Pentru a asigura ergodicitatea, cele șase lumi de mai sus au fost simulate împreună cu o lume cartesiană.

2. Rezultate și discuții – Dipeptida de alanină

În ciuda dimensiunii sale mici comparativ cu alte biomolecule, dipeptidul de alanină (N-acetilalanină-N-metilamidă) are o suprafață a energiei potențiale (PES) extrem de complexă, făcându-l un sistem dificil pentru tehnicile de eșantionare moleculară. Maparea spațiului său de configurație pe unghiurile dihedrale ϕ și ψ ajută la păstrarea maximelor și minimelor cheie ale PES-ului, reducând în același timp complexitatea. Drept urmare, potențialul forței medii în domeniul ϕ/ψ este utilizat frecvent pentru a evalua acuratețea și eficiența metodelor de eșantionare (Montgomery Pettitt and Karplus, 1985). Toate simulările au reușit să recupereze suprafața energiei libere a sistemului ALA2.



Figură 3-2 Suprafețe de energie liberă ale ALA2, recuperate folosind patru tipuri de simulări:: A – Complet Flexibilă, B - Mixed-RamaTD, C - Mixed-RamaBall, D- Mixed-RamaCYL

Faptul că suprafețele energiei libere sunt consistente între ele arată că simulările CDHMC pot recupera suprafețele energiei libere folosind articulații Pin, cilindrice sau sferice, indiferent de dimensiunea corpurilor.

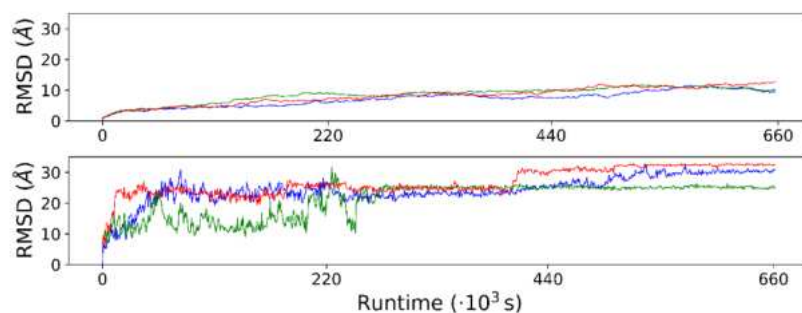
Timpurile medii de trecere a primelor treceri între bazinele de energie liberă oferă o imagine asupra eficienței relative a simulărilor realizate. După cum se poate observa în Tabel 3-1, simularea complet flexibilă are cele mai lente tranziții generale. Simulările „All-Bonds” eșantionează mai bine decât cea complet flexibilă, regimul TD având cele mai bune rezultate. Dinamicile Ramachandran depășesc dinamicile all-bonds.

From\To	C5	PPII	C7 _{eq}	α_L
Complet Flexibilă				
C5	3.2 ± 0.2	9.1 ± 0.5	14.3 ± 1.7	5380.0 ± 285.0
PPII	9.6 ± 0.9	5.7 ± 0.4	10.9 ± 1.8	5380.0 ± 285.0
C7_{eq}	16.5 ± 1.1	12.6 ± 0.6	2.8 ± 0.2	5370.0 ± 285.0
α_L	502.0 ± 195.0	499.0 ± 194.0	491.0 ± 194.0	14.1 ± 4.4
Mixed-RamaTD				
C5	1.8 ± 0.03	4.0 ± 0.1	2.8 ± 0.1	631.0 ± 92.2
PPII	2.5 ± 0.1	3.2 ± 0.1	2.6 ± 0.1	631.0 ± 92.2
C7_{eq}	3.2 ± 0.2	4.7 ± 0.2	1.5 ± 0.02	630.0 ± 92.2
α_L	50.9 ± 1.1	52.4 ± 1.2	49.4 ± 1.1	8.3 ± 1.3
Mixed-RamaCYL				
C5	2.3 ± 0.03	5.2 ± 0.1	3.5 ± 0.02	615.0 ± 33.6
PPII	3.2 ± 0.1	4.1 ± 0.1	3.3 ± 0.005	615.0 ± 33.4
C7_{eq}	4.1 ± 0.1	5.9 ± 0.1	1.9 ± 0.01	614.0 ± 33.6
α_L	46.4 ± 4.5	48.4 ± 4.6	44.7 ± 4.7	12.0 ± 1.4
Mixed-RamaBall				
C5	2.4 ± 0.01	5.2 ± 0.01	3.1 ± 0.1	518.0 ± 18.6
PPII	3.1 ± 0.01	4.1 ± 0.04	3.0 ± 0.1	518.0 ± 18.5
C7_{eq}	3.7 ± 0.01	5.6 ± 0.02	1.9 ± 0.02	518.0 ± 18.4
α_L	37.0 ± 1.4	39.1 ± 1.5	35.8 ± 1.6	12.8 ± 1.0

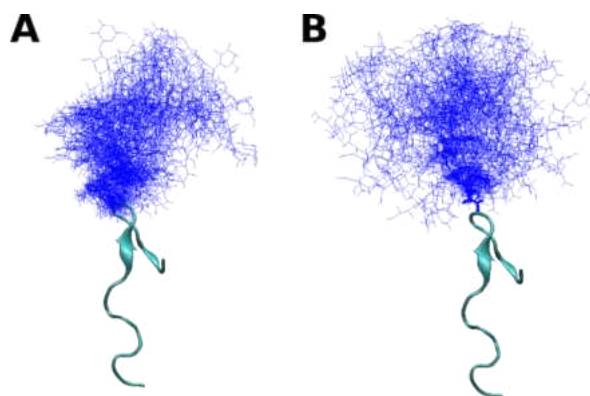
Tabel 3-1 Timpurile medii de trecere ale primei treceri (Mean First Passage Time - MFPT) ale simulărilor complet flexibile și Ramachandran, exprimate în pași de dinamică moleculară. Notă: în lucrarea originală, acestea au fost împărțite la 200 pentru a fi comparabile cu lucrarea care descrie inițial metoda CDHMC..

3. Rezultate și discuții – Glycoproteina E2N1

Figură 3-3 arată că, în aceeași perioadă de timp, valoarea medie și fluctuațiile RMSD sunt mult mai mari pentru simulările CDHMC decât pentru simulările HMC complet flexibile. O suprapunere a unui subset de cadre evidențiază o diferență mai clară între cele două metode. (Figură 3-4).



Figură 3-3 RMSD obținută din simulări de tip MD (sus) și de tip HMC (jos)



Figură 3-4 Cadre suprapuse din simulări de MD (A) și HMC (B). Pentru claritate, au fost incluse doar structurile polizaharidei, iar polipeptidul a fost păstrat din primul cadru.

4. Concluzii

Rezultatele prezentate în această lucrare sugerează că Robosample oferă un cadru robust pentru realizarea simulărilor moleculare constrânse, asigurându-se în același timp ergodicitatea prin eșantionare Gibbs. Dincolo de mobilitățile tradiționale de rotație și unghi/torsiune utilizate frecvent în simulările constrânse (Vaidehi and Jain, 2015), , Robosample extinde tipurile de articulații utilizabile prin încorporarea unor tipuri suplimentare de mobilitate robotică, folosind articulațiile mecanice disponibile în Simbody. Acest lucru permite explorarea unor cuplaje arbitrare ale gradelor de libertate, inclusiv interacțiuni slabe de legături/torsiune întâlnite în articulațiile cilindrice.

S-a demonstrat că, prin optimizarea parametrilor și utilizarea diferitelor tipuri de articulații, software-ul poate reproduce suprafața de energie liberă a ALA2 într-un mod mai eficient decât este posibil prin simulări MD clasice. În plus, dinamica Ramachandran utilizând articulații sferice, care nu au fost utilizate în alte software-uri MD cu DoF interne, are cel mai scurt MFPT pentru tranziția cea mai rară.

Folosind un sistem real (glicoproteina E2N1), am demonstrat că explorarea spațiului conformațional al structurilor foarte flexibile poate fi mai eficientă dacă se utilizează Robosample. Faptul că utilizarea alternativă a constrângerilor accelerează explorarea spațiului sugerează că există, în teorie, o combinație optimă de blocuri Gibbs care poate

explora orice spațiu mai rapid decât MD simplu. În prezent, explorăm utilizarea Replica Exchange în tandem cu algoritmul HMC, pentru a depăși barierele de energie potențial mai mari.

PARTEA IV-A – APLICAREA MODELĂRII ȘI SIMULĂRII PENTRU STUDIEREA INTERACȚIUNILOR MOLECULARE COMPLEXE

CAPITOLUL 4 – MODELAREA COMPLEXĂRII RECEPTORULUI TROMBOPOIETINEI DE CĂTRE CALRETICULINA MUTANTĂ

1. Introducere

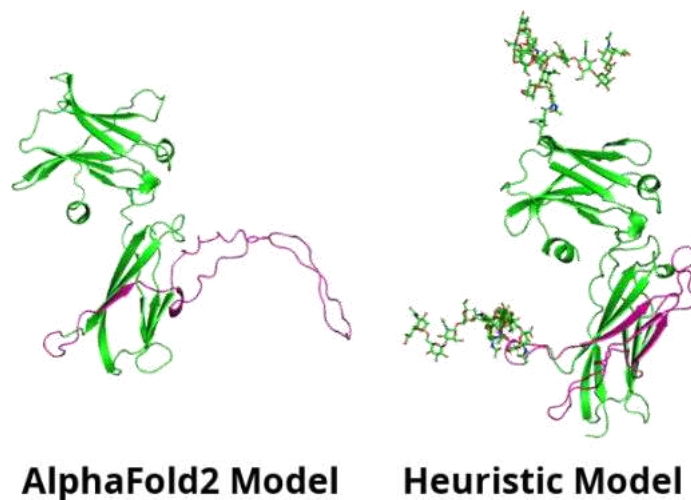
În capitolul actual, am folosit o strategie integrativă pentru a descompune mecanismul molecular prin care mutantele calreticulin del52/ins5 (mCALR) se leagă în mod specific și persistent de TpoR și pentru a face lumină asupra modului în care această interacțiune determină dimerizarea și activarea TpoR. Înțelegerea modului în care mutațiile de tip frameshift într-un chaperon cheie duc la noi capacități de legare este esențială atât dintr-o perspectivă conceptuală, cât și pentru dezvoltarea terapeutică. Pentru a dezvălui baza moleculară a formării unui complex permanent TpoR-mCALR, procedăm la modelarea precisă a celor două componente, în încercarea de a localiza posibile puncte de interes responsabile pentru interacțiunea lor, așa cum este prezentat mai jos.

2. Modelarea prin omologie a TpoR

În prima etapă, secvența TpoR a fost analizată pentru a genera multiple profile ale structurii secundare, profile de accesibilitate la solvent și profile de dezordine intrinsecă. Aceste profile au fost utilizate pentru a rafina alinierea între secvența țintă și modelele folosite.

Următoarea etapă a constat în căutarea structurilor omoloage care să fie utilizate în modelarea prin omologie. Astfel, secvența TpoR a fost supusă unor căutări prin threading molecular, utilizând PHYRE2 (Kelley *et al.*, 2016). Primul rezultat (cod RCSB: 1ERN (Livnah *et al.*, 1999)) is a crystal of the Erythropoietin Receptor (EpoR). este un cristal al Receptorului Eritropoietinei (EpoR). Au fost adunate 20 de structuri EpoR din baza de date RCSB, iar structura 1CN4 (Syed *et al.*, 1998) a fost aleasă pentru a fi utilizată ca template pentru modulului de recepție al citokinelor (CRM) distal al membranei. Aceasta prezenta aproximativ 25% identitate cu secvența TpoR. Chiar dacă secvența țintă poate fi modelată prin omologie folosind EpoR, aceasta prezenta o inserție mare de 65 de aminoacizi în al

doilea său domeniu, pentru care nu au fost găsite structuri omoloage. Interesant este că modelele automate generate cu platforma de ultimă oră AlphaFold2 erau nerealiste. Soluția propusă pentru această inserție de 65aa era un coil random, deplasându-se separat de corpul principal pliat al TpoR, fiind complet accesibil solventului, în ciuda predicțiilor sale de structură secundară și accesibilitate.



Figură 4-1 Comparația dintre modelul AlphaFold2 și modelul nostru euristic. Partea violet reprezintă inserția de 65 de aminoacizi. Glicanii sunt reprezentați cu o reprezentare de tip "sticks".

Prin urmare, pentru a modela mai realist această inserție, au fost efectuate căutări pentru arhitecturi moleculare mai adecvate. Structura de referință, EpoR, prezintă o arhitectură tip β -sandwich, în care una dintre fețe are 3 structuri extinse, iar cealaltă are 4 (β -sandwich 3/4). Această arhitectură face parte din superfamilia CATH 2.60.40.10, care include și arhitecturi β -sandwich 4/5. Presupunerea că TpoR adoptă o astfel de arhitectură a pornit de la observația că 2 structuri extinse au fost prezise în inserția mare a secvenței TpoR. Astfel, a fost propusă o arhitectură β -sandwich 4/5 pentru TpoR. Grupul imunoglobulinic al superfamiliei 2.60.40.10 este cunoscut pentru adoptarea acestei arhitecturi, cu aceeași topologie ca cea propusă. Reprezentantul CATH pentru grupul imunoglobulinic este anticorpul 2E8 pentru receptorul LDL, cu codul RCSB: 12E8. (Trakhanov *et al.*, 1999).

CRM-ul membranei-proximal al domeniului extra-celular (ECD) TpoR a fost modelat folosind software-ul AlphaFold2, care a fost rulat pe clusterul nostru local de calcul. Modelul generat a fost atașat la modelul CRM 1, utilizând MODELLER v.9.12 (Webb and Sali, 2018). Au fost adăugate 4 legături disulfidice de cisteină pe baza distanței dintre reziduurile de cisteină adiacente.

Domeniul transmembranar (TMD) a fost modelat pe baza lanțului E al structurii cristaline cu ID-ul PDB: 6S1K (Cassidy *et al.*, 2020). Această structură cristalină a fost găsită utilizând BLAST pe secvența TMD-ului TpoR. Acest lucru a fost realizat pentru a genera o

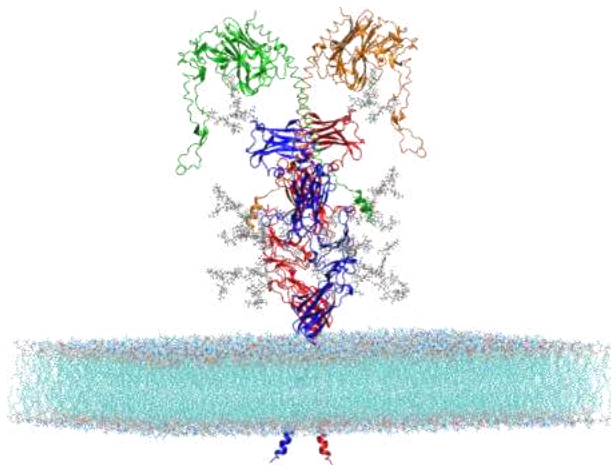
structură helicoidală realistă. TMD-ul a fost atașat la ECD folosind MODELLER. Unghiul de înclinație al TMD-ului, relativ la membrana ipotetică pe care trebuie să o traverseze, a fost ales pentru a fi consistent cu datele experimentale furnizate (adică distanța dintre reziduurile adiacente L508 ar trebui să fie de cel mult 6Å) și pentru a forma un dimer în formă de cruce similar cu cel găsit de (Defour *et al.*, 2013).

Monomerii CALR del52/ins5 au fost generați utilizând AlphaFold2, iar dimerul CALR a fost generat folosind RosettaDock (Lyskov and Gray, 2008).

3. Evaluarea interacțiunii CRM1 a TpoR cu calreticulina mutantă

Analiza secvenței TpoR D1D2 arată un exces de încărcături negative. În schimb, C-terminalul CALR del52 este puternic pozitiv. Astfel, am formulat o ipoteză conform căreia interacțiunea detectată între CALR și TpoR ar putea fi de natură electrostatică.

Prin intermediul unui flux de lucru extins care a implicat multiple rulări de docking molecular, dinamica moleculară și simulări de reciere simulată (simulated annealing), au fost determinate 3 poziții de legare pentru dimerul TpoR/mCALR C-terminal. Prin utilizarea Prodigy (Vangone and A. M. J. J. Bonvin, 2015; Xue *et al.*, 2016) și MM-GBSA (Kollman *et al.*, 2000), acestea au fost clasificate, iar cea mai bine clasată a fost utilizată pentru a genera modelul complet de tetramer, care a fost ulterior integrat într-o membrană de POPC și complet hidratat.



Figură 4-2 Sistem TpoR – mCALR complet hidratat, compus din aproximativ 1,5 milioane de atomi.

Mutantul ins5 al CALR a fost generat folosind aceeași procedură.

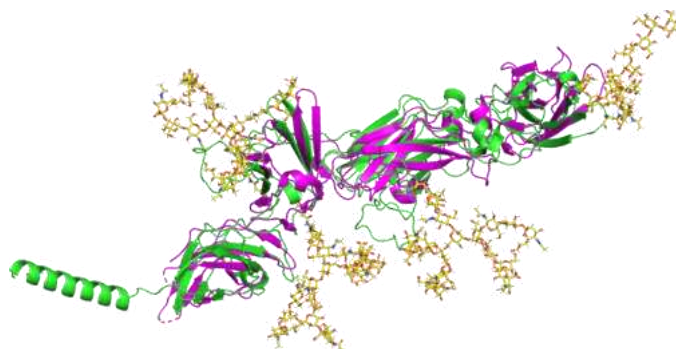
Stabilitatea sistemului a fost verificată prin rularea unei simulări de MD, timp de 100 de nanosecunde. Numarul de puncte de contact dintre glicanii de pe suprafața TpoR și CALR a fost conservat.

4. Concluzii

Lucrarea prezentată oferă perspective mecanistice asupra recunoașterii și activării ulterioare a receptorului pentru trombopoietină (TpoR) de către mutantele patogene ale calreticulinei, declanșând neoplasme mieloproliferative. Modelul structural pe care l-am creat arată că legarea se realizează pe două zone distincte: o legare fiziologică a CALR/N-glicanului, care nu este specifică perechii TpoR/CALR, și o interacțiune electrostatică, mult mai stabilă, care duce la patologii sanguine. Modul de activare pe care îl propunem poate fi folosit în studii suplimentare pentru dezvoltarea de terapii țintite, care ar putea viza, de exemplu, N-terminalul moleculei CALR.

Metoda pe care am folosit-o pentru descoperirea conformației complexului TpoR/CALR del52 evidențiază un efect interesant de tip „anaconda” al moleculei CALR. C-terminalul, fiind încărcat pozitiv, execută o mișcare asemănătoare unei alunecări în jurul CRM-ului distale de membrană al TpoR, până când se „blochează” pe una dintre cele două zone acide. Este tentant să presupunem că acest tip de legare nu este unic pentru această pereche de proteine și ar putea fi folosit ca o ipoteză de plecare atunci când studiem alte interacțiuni proteică/proteică cu semnificație biologică.

Ca observație finală, am dori să menționăm că o structură experimentală de cryo-EM a TpoR, publicată la câteva luni după publicarea modelului nostru tetrameric 2x(TpoR-mCALR) (Tsutsumi *et al.*, 2023b) a confirmat structura TpoR prezisă inclusă în modelul nostru mai complex de glicoproteină heteromoleculară globală, așa cum se poate observa în Figură 4-3. Mai mult, din structura cryo-EM lipsește structura glicanilor care este inclusă în model. Chiar mai important – nucleul proteic al TpoR din modelul nostru euristic se potrivește mai bine cu structura cryo-EM decât soluția structurală automată a TpoR propusă în acel moment de AlphaFold2, așa cum am discutat anterior (Figură 4-1).



Figură 4-3 Alinierea structurii între modelul generat (verde și galben) și structura rezolvată (magenta – RCSB Code: 8G04) a ECD-ului TpoR.

CAPITOLUL 5 – OXIDAREA EPITOPULUI TIROZINAZEI ȘI EFECTUL ACESTEIA ASUPRA REACTIVITĂȚII CELULELOR T ÎN MELANOM

1. Introducere

Proceduri complexe de bioinformatică și biocomputare pot fi, de asemenea, utilizate în eforturi legate de o înțelegere profundă a interacțiunilor proteină-ligand atunci când liganzii sunt peptide flexibile complexe.

Acest capitol prezintă efortul nostru de a dezvălui mecanismele moleculare care duc la creșterea energiei libere de legare a complexului ternar de imunitate HLA-YMD-TCR după oxidarea YMD - unde YMD este peptida 369-YMDGTMSQV-377 derivată din tirozinază. Această prezentare este precedată de o scurtă introducere privind cele trei obiecte moleculare și contextual biologic al acestui studiu.

Proteinele HLA, MHC I sunt heterodimere, compuse din două lanțuri proteice, α și β -microglobulină. Lanțul α este format din trei module, fiecare având un rol distinct în interacțiunea sa cu receptori speciali aflați pe suprafața celulelor T. Lanțul $\alpha 3$ este singurul lanț transmembranar din structura MHC I și este, de asemenea, responsabil pentru legătura non-covalentă cu lanțul β -microglobulină. În plus, regiunea $\alpha 3$ interacționează cu receptorul CD8 găsit pe suprafața celulelor T. Această interacțiune blochează temporar complexul MHC în timp ce celula T verifică antigenicitatea peptidei legate în groapa care se deschide între domeniile $\alpha 1$ și $\alpha 2$ (Hewitt, 2003).

Receptorul T-cell (TCR) este un heterodimer compus din două lanțuri, α și β . În cazuri foarte rare, acesta poate consta din lanțurile γ și δ în loc de α și β . Fiecare dintre lanțurile α și β are două regiuni distincte:

- O regiune constantă, situată mai aproape de membrana celulei T (dar încă extracelulară)
- O regiune variabilă, care conține trei segmente hiper-variabile ce determină complementaritatea peptidei (CDR) și sunt responsabile pentru legarea peptidei antigenice.

2. Metode – Modelarea complexului ternar, generarea de parametrii pentru YM2M6D, HREX

În linii mari, munca computațională a fost realizată în trei etape:

- Modelarea sistemului binar HLA-A02:01/YMD, atât în forma nativă (WT), cât și în forma M2SOM6SO

- Modelarea sistemului ternar HLA-A02:01/YMD/TCR, în ambele forme
- Calcularea energiei de legare liberă a peptidei YMD la complexul ternar

Pentru complexul HLA/YMD, procesul de selecție a identificat YLSPIASPL (Y9L) și MLIYSMWGK (A14) ca cele mai bune structuri de referință:

- Y9L (PDB: 5F9J) este legat în mod natural de HLA-A02:01, ceea ce îl face un candidat puternic pentru haplotipul experimental. În plus, se aliniază bine cu profilele de hidrofobicitate și volumetric ale YMD și prezintă o tirozina pe poziția 1, care formează o interacțiune de tip stacking cu HLA, la fel ca YMD

- A14 (PDB: 4N8V) provine din proteina membranară a virionului A14 și a fost ales datorită prezenței metioninei pe poziția 6, care reflectă YMD. Cu toate acestea, această structură este asociată cu HLA-A11, care diferă de haplotipul experimental.

Ambele șabloane conțin reziduuri hidrofobice pe pozițiile 2 și 6, care sunt implicate în oxidare și interacționează cu HLA. Totuși, aranjamentele lor spațiale în cadrul spațiului de legare HLA sunt distincte:

- În Y9L, lanțul lateral de pe poziția 6 se îndreaptă direct către situsul de legare al HLA.
- În A14, lanțul lateral de pe poziția 6 este orientat paralel cu suprafața.

Aceste diferențe oferă o oportunitate de a evalua modul în care configurațiile inițiale de legare influențează estimările energiei libere.

Având în vedere că din toate formele posibile oxidate ale peptidei YMD, a fost demonstrat prin experimente HPLC că M2SOM6SO este forma predominantă, au fost construite patru modele ale complexului binar:

- HLA/YMD nativ cu M6 îndreptându-se către situsul HLA/spre solvent
- HLA/YMD oxidat, cu M6 îndreptându-se către situsul HLA/spre solvent

3. Modelarea complexului ternar HLA/YMD/TCR

Construirea modelelor complexe ternare HLA-YMD-TCR a fost realizată în două etape secvențiale:

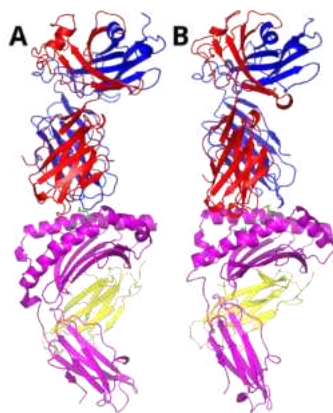
- **Asamblarea TCR:** Modele care încorporează buclele CDR ale clonelor specifice pacientului 3 (TCRc3) au fost generate utilizând o abordare comună bazată pe fragmente de modelare prin omologie
- **Docking:** Modelul TCRc3 a fost adăugat (docked) la cele patru complexe binare HP create în faza inițială.

Pentru a construi modelul TCRc3, baza de date HP/HPT-DB a fost scanată pentru a identifica candidați structurați similar, corespunzători regiunilor CDR specifice clonelor pacientului. Analiza secvenței a arătat că diferitele bucle CDR și regiunile constante au avut potriviri diferite în ceea ce privește omologia secvenței. Drept urmare, a fost implementată o strategie comună de modelare prin omologie bazată pe fragmente, urmând metodologii descrise anterior (Slootweg *et al.*, 2013; Rajaraman *et al.*, 2016).

Patru structuri de referință (ID-uri PDB: 5HHM, 3PWP, 2PYE și 2BNQ) au fost selectate:

- 5HHM și 3PWP au fost utilizate pentru a construi cadrele conservate ale lanțurilor α și β , respectiv.
- 2PYE și 2BNQ u servit drept șabloane pentru asamblarea buclelor CDR ale TCRc3

Toate modelele HLA-A*02:01 și TCRc3 au fost generate folosind MODELLER v9.21. Pe scurt, buclele variabile care leagă regiunile conservate ale secvenței (SCRs) au fost optimizate iterativ prin protocolul de rafinare a buclelor al MODELLER. Regiunile variabile ale secvenței au fost supuse unui proces repetat de annealing simulativ și minimizare a energiei, iar modelele finale au fost validate utilizând serverul MolProbity, obținând scoruri de 0,59 pentru HLA și 1,00 pentru TCR, confirmând calitatea lor structurală.



Figură 5-1 Modelele HLA/YMD/TCR generate, cu M6 îndreptate spre TCR (configurația $\uparrow$); A – Stare nativă; B – Stare mutantă

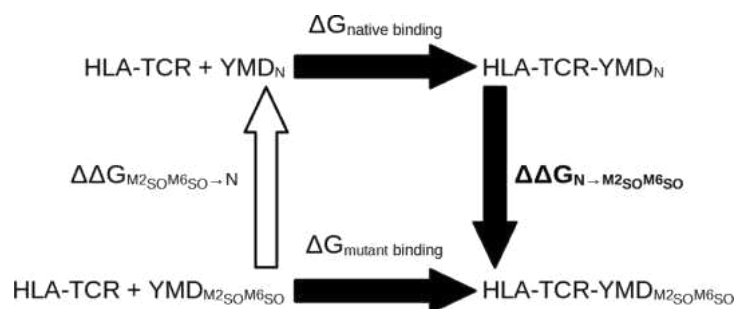
4. Calculul energiei libere de legare

În această fază, estimările energiei libere de legare absolute au fost obținute folosind Hamiltonian Replica Exchange (HRE), unde stările termodinamice au fost progresiv scalate între o stare complet cuplată și o stare complet decuplată a ligandului. Acest proces implică o transformare alchemicală, în care interacțiunile exercitate de HLA și TCR asupra YMD sunt slăbite treptat până când sunt complet eliminate. Cele două modele de complex HLA/YMD/TCR (cu YMD nativ și YMD oxidat) au fost supuse unui total de 472,5 ns de

simulare, pe parcursul a 135 de replici. O schimbare de replici urma să fie încercată la fiecare 10^{-3} ns. Traietoriile complet cuplate au fost ulterior folosite pentru analiza structurală diferențială, inclusiv analiza stabilității, analiza numărului de clustere și analiza accesibilității solvatului.

Calculul energiei libere de legare a fost realizat utilizând YANK (Wang *et al.*, 2013), utilizând Hamiltonian Replica Exchange (HRE) pentru un eșantionaj îmbunătățit. Multistate Bennett Acceptance Ratio (MBAR), implementat prin pachetul PyMBAR a fost utilizat pentru a calcula energiile libere de legare. Numărul de replici și parametrii lor de scalare au fost determinați automat folosind modulul inovator YANK. Peptida a fost păstrată în interiorul site-ului de legare adăugând restricții armonice, care apăreau înainte ca parametrii nonbondiți să dispară.

Diferențele de energie liberă de legare ($\Delta\Delta G$) au fost estimate folosind un ciclu termodinamic care implică energiile libere de legare ale ambelor forme. Diferența de energie de pe săgeata din dreapta reprezintă diferența în energia de legare dată de oxidarea reziduurilor de metionină



Figură 5-2 Ciclu termodinamic care stă la baza calculului efectului oxidării peptidei.

Rezultatele MBAR sunt prezentate în Tabel 5-1

ΔG N $\uparrow$	ΔG M2SO M6SO $\uparrow$	ΔG N $\downarrow$	ΔG M2SO M6SO $\downarrow$
-43.070 +/- 1.085	-43.277 +/- 0.595	-17.097 $\pm$ 0.669	-21.05 $\pm$ 0.557
$\Delta\Delta G \uparrow$		$\Delta\Delta G \downarrow$	
~ 0.0 KCAL/MOL		~ -4.0 kcal/mol	

Tabel 5-1 Energii libere de legare absolute ale celor patru sisteme descrise.

Analiza Tabel 5-1 relevă faptul că doar diferența de energie liberă de legare în configurația „down” ($\downarrow$) ($\Delta\Delta G \downarrow$) este semnificativă din punct de vedere statistic. Acest lucru sugerează că deplasarea observată experimental a echilibrului către forma oxidată în complexul ternar este în principal determinată de configurația YMD $\downarrow$. Având în vedere că oxidarea la M6 (M6SO) se așteaptă să influențeze direct legarea de TCR, este oarecum neașteptat ca afinitatea crescută pentru forma oxidată să fie atribuită în principal configurației „down” ($\downarrow$), mai

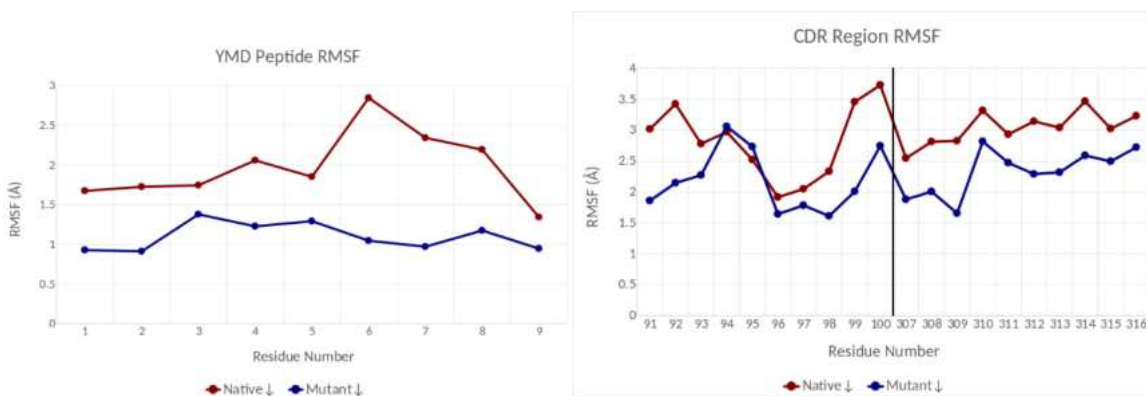
degrabă decât interacțiunii directe cu TCR. Totuși, variațiile energiei libere de legare provin din echilibre termodinamice complexe, care includ atât contribuții entalpice, cât și entropice, ce rezultă dintr-o gamă largă de interacțiuni moleculare dinamice la temperatura camerei.

Prin analiza replicii „complet cuplate”, s-a constatat că forma mutantă a peptidei a crescut stabilitatea locală a buzunarului de legare. În primul rând, stabilitatea moleculelor de apă a crescut în forma mutantă, așa cum a fost măsurată prin timpul de rezidență al apei și numărul mediu de molecule de apă din site pentru cele două sisteme.

	N ↓	M2 _{so} M6 _{so} ↓
Average	11.5	15.4
St. Dev	5.8	5.1

Tabel 5-2 Numarul mediu de molecule de apă din situs, pentru fiecare formă a complexului

În al doilea rând, RMSF (fluctuațiile medii ale distanțelor la fiecare pas de simulare) ale peptidei YMD și ale buclelor CDR ale TCR sunt mai mici în forma mutantă:



Figură 5-3 RMSF al peptidei YMD și ale buclelor CDR din TCR pentru complexe nativ și mutant. Linia verticală separă cele două regiuni CDR.

5. Concluzii

În lucrarea descrisă în capitolul curent, am generat mai multe modele ale complexului binar HLA: A0201/YMD și capacitatea acestora de a se lega la o clonă specifică a unui receptor de celule T. Nu numai că rezultatele noastre computaționale au fost de acord cu constatările experimentale (descrise mai detaliat în lucrare), dar prin modelele noastre am putut să adăugăm informații asupra contribuției fragmentului sulfoxid.

Este important de menționat că un număr mare de aproximări au fost făcute în timpul construirii modelului. În primul rând, faptul că obiectul molecular prezentat constă din mai multe molecule care au fost construite prin modelare prin omologie, nepreluate din nicio structură cristalină, adaugă un grad de incertitudine. Acest lucru este exacerbat de faptul că buclele CDR sunt foarte variabile și sunt, de asemenea, direct implicate în legare. Mai mult, proteina CD8, care este implicată în stabilizarea complexului ternar, nu a fost inclusă în model, având în vedere lipsa de informații structurale privind orientarea relativă a acesteia.

În concluzie, datele computaționale sugerează că oxidarea crește legarea printr-o creștere a ordinii locale a interfeței, mai degrabă decât printr-o interacțiune directă cu TCR.

CAPITOLUL 6 – INTERACȚIUNEA CANALULUI VOLTAJ-DEPENDENT NAV1.5 CU CENOBAMATUL

1. Introducere

Lucrarea de față prezintă o combinație de tehnici de modelare moleculară și simulare pentru a investiga efectul pe care opt mutații punctuale ale canalului NaV1.5 îl au asupra interacțiunii sale cu cenobamatul, un nou medicament antiepileptic, mutații cunoscute pentru a produce distorsiuni QT. Mai precis, descrie modelarea domeniilor NaV1.5 α , precum și opt mutații punctuale semnificative, modelarea precisă a mediului și echilibrarea acestuia prin simulări MD din ce în ce mai flexibile. Apoi descrie metoda prin care am folosit MD și MM-PBSA pentru a evalua dacă mutațiile descrise pot afecta legarea cenobamatului.

Din punct de vedere structural, canalele Na⁺ dependente de tensiune (Nav) constau din patru domenii α omoloage care formează pori, fiecare constând din șase elice transmembranare. Între al 3-lea și al 4-lea domeniu α există o buclă intracelulară scurtă, care ca o poartă de inactivare, blocând porul din interior în timpul depolarizării susținute a membranei.

2. Generarea modelului

Modelul pentru forma de tip nativ a Na(v)1.5 a fost generat folosind AlphaFold2. Proteina a fost încorporată în membrana DPPC folosind PACKMOL-Memgen (Schott-Verdugo and Gohlke, 2019). A fost utilizată apă explicită de tip TIP3P cu o tărie ionică de 150 mM (compusă din Na⁺ și Cl⁻). A fost generat o membrană lipidică de aproximativ 200 Å x 200 Å, cu un tampon de apă de 17,5Å deasupra și sub proteină, pentru a asigura o celulă adecvată în care domeniul citosolic nu ar comunica cu domeniul extracelular peste granița periodică.

Formele mutante ale proteinelor au fost generate după echilibrarea formei de tip nativ. Ligandul cenobamat a fost parametrizat folosind GAFF2.

Docking-ul moleculei de cenobamat la diferiți mutanți a fost realizată folosind Autodock Vina 1.2.5 (Trott and Olson, 2009)

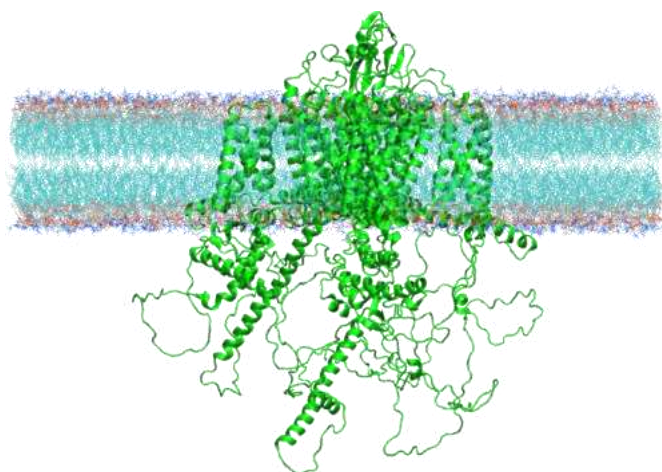
Minimizarea globală inițială a fost efectuată folosind OpenMM. Toate simulările moleculare au fost efectuate folosind NAMD3 (Phillips *et al.*, 2020). Fiecare dintre cele 9 forme ale complexelor receptor-ligand Na(v)-1,5 a fost supusă la 100ns de simulare MD.

Având în vedere dimensiunea sistemului, a fost utilizat un interval de timp de 2fs și s-a aplicat SHAKE pe legăturile de hidrogen.

Pentru a calcula energia liberă de legare, au fost efectuate calcule MM-PBSA. Parametrii de calcul au fost derivați din cazuri de utilizare similare în alte lucrări publicate (Wang *et al.*, 2022). Au fost folosite 100 de cadre din fiecare simulare.

3. Rezultate – Modelare și simulare

Domeniul α al NaV1.5 uman a fost modelat cu succes folosind AlphaFold2. Modelul a fost încorporat într-un plasture de 200Åx200Å de DPPC, sistemul rezultat a fost solvatat într-o cutie de apă explicită, cu un tampon de 17,5Å deasupra și sub proteină. Ioni Na⁺/Cl⁻ au fost adăugați pentru a neutraliza sistemul și până la o concentrație de 0,15 M. Cei 8 mutanți au fost generați după minimizarea și echilibrarea formei de tip nativ. După generare, mutații au fost minimizați pentru a se asigura că nu vor fi prezente ciocniri după ce mutațiile au fost făcute.



Figură 6-1 NaV1.5 (reprezentare tip cartoon) încorporat în membrana lipidică (reprezentare folosind linii). Apa și ionii au fost omiși pentru claritate

Pe parcursul simulărilor explicite de solvenți de 100 ns, cele 9 modele de NaV1.5 nu au suferit modificări conformaționale majore, măsurate de RMSD.

Analiza RMSD a moleculei de cenobamat, atașată la diferitele forme ale proteinei NaV1.5, arată că numai mutația N1463K, în timpul perioadei de echilibrare, permite ligandului să fie mai flexibil, toate celelalte mutații îl fac mai rigid decât forma wildtype. Legarea mai strânsă prezentată cu formele mutante ar putea indica faptul că există influențe entalpice care conduc la valorile K_d estimate mai mici (Tabel 6-1).

4. Rezultate – Calculul de energie liberă de legare

Efectul mutațiilor punctuale asupra legării cenobamatului de NaV1.5 a fost estimat atât prin andocare moleculară, cât și prin MM-PBSA. După cum este prezentat în Tabel 6-1, experimentele de docking sugerează că N927S, N1463K, N1463Y și M1766R leagă cenobamatul mai puternic decât varianta wildtype. Cu toate acestea, calculele MM-PBSA prezic că toți mutanții leagă molecula de cenobamat mai puternic decât varianta wildtype, mutantul N932S fiind cel mai bun candidat pentru legare.

VARIANTĂ	SCOR MEDIU DE LA DOCKING (KCAL/MOL)	ΔG_{LEGARE} MM-PBSA (KCAL/MOL)
WILDTYPE	-5.50	-9.0828+/-3.0573
N927S	-5.62	-11.6598+/-2.3896
N932K	-5.00	-11.6168+/-2.6316
N932S	-5.03	-14.394+/-2.0402
L935V	-4.83	-11.3162+/-2.7667
S1458Y	-5.11	-11.8996+/-2.2859
N1463K	-6.20	-11.8148+/-2.1851
N1463Y	-6.01	-11.0899+/-1.7401
M1766R	-6.03	-10.1092+/-2.0133

Tabel 6-1 Afinitatea de legare estimată a cenobamatului la diferite forme de NaV1.5 uman

5. Concluzii

Cenobamatul, un nou medicament anti-epileptic, prezintă efecte de legare asupra canalului NaV1.5 la concentrații relevante clinic. Lucrarea prezentată arată că compusul se leagă în cavitatea centrală a NaV1.5 wildtype și mai puternic la variantele mutante selectate. Ca atare, mutațiile în canalul cardiac NaV1.5 trebuie luate în considerare atunci când se prescrie cenobamat, deoarece mutațiile în canalul menționat ar putea duce la aritmii ventriculare potențial fatale.

Concluzii generale

Prezenta teză descrie munca întreprinsă în timpul pregătirii mele de doctorat în cadrul Departamentului de Bioinformatică și Biochimie Structurală a Dr. Petrescu de la IBAR. Aceasta a fost o perioadă interesantă care a concurat cu progresul uluitor în bioinformatică și biocomputing. Acest lucru mi-a permis să înțeleg tehnicile de ultimă generație în modelarea și simularea moleculară și să le folosesc pentru a obține o perspectivă aprofundată asupra comportamentului mai multor sisteme biomoleculare complexe relevante în medicina moleculară.

Prima parte a muncii mele este dedicată explorării capacităților și limitelor noii generații de platforme de predicție a structurii conduse de DL în modelarea sistemelor de

proteine mai complexe cu mai multe stări și mai multe domenii - luând ca studiu de caz o familie de receptori CNL NOD-Like de la *Arabidopsis thaliana* asupra căreia departamentul nostru are expertiză structurală semnificativă și rezultate în ultimul deceniu. Rezultatele noastre indică faptul că, în timp ce astfel de platforme de modelare automată arată o precizie ridicată în prezicerea structurii modulelor de secvență care se pliază în topologii bine-cunoscute și validate, ele nu prevăd configurația regiunilor cu arhitectură tip „Coiled-Coil”, a omologilor la distanță și a regiunilor lipsite de șabloane - tehnicile euristice manuale mai flexibile depășesc modelarea automată.

Am dezvoltat fluxuri de lucru de filtrare a informațiilor AlphaFold2 menite să conducă procesul de modelare către configurații specifice în instanțe cu mai multe stări. Acest lucru dezvăluie două căi majore de cercetare, în adaptarea datelor de intrare (secvențe și structuri) bazat pe familia din care aparțin proteinele de interes și în utilizarea șabloanelor structurale multimer pentru a genera structuri monomerice cu conformație multimer, care este mult mai puțin consumatoare de timp și consumatoare de resurse.

Ambele căi descrise vor fi explorate în continuare după susținerea tezei mele, mai ales prin scrierea flux complet de lucru care eficientizează selecția secvențelor și structurilor. Dincolo de modelare, care returnează structuri unice „înghețate”, a doua parte a acestei teze prezintă contribuțiile mele la dezvoltarea Robosample – o platformă de simulare extrem de eficientă a eșantionării conformaționale moleculare bazată pe algoritmi robotici dezvoltați și coordonați în Departamentul nostru de Dr. Laurențiu Spiridon.

Robosample – care este prima platformă românească de simulare moleculară – este o realizare absolută care se află în prezent în plină dezvoltare și extindere. Prin urmare, intenționez să fac din Robosample prima și principală prioritate a mea de lucru după susținerea tezei, continuând să optimizez experiența utilizatorului final și colaborând cu alții în cercetarea de grup asupra noilor selecții de blocuri Gibbs. În plus, intenționez să implementez Hamiltonianul Monte Carlo pentru a completa implementarea existentă a T-REX și pentru a explora eficiența software-ului pentru simulările de andocare moleculară. A treia parte a lucrării mele de doctorat se concentrează pe utilizarea modelării și simulării moleculare pentru a evalua interacțiunile moleculare în sisteme biomoleculare complexe relevante în medicina moleculară.

De exemplu, lucrările privind receptorul de trombopoietină și peptida YMD au oferit ambele perspective practice asupra mecanismelor moleculare relevante în oncologie, abordând aspecte legate de neoplasmele mieloproliferative și, respectiv, melanom. Primul identifică mecanismele moleculare prin care o schimbare a cadrului C-terminal în gena

Calreticulină, care exprimă o însoțitoare de lectină în ER, induce activarea constitutivă a unui receptor de citokine, în special receptorul de trombopoietină, în timp ce al doilea dezvăluie detalii mecaniciste legate de efectul oxidării asupra tirozinazei YMD și de ce are un efect mai mare de recunoaștere a tirozinazei T și de ce are un efect mai mare de recunoaștere a celulei T. Lucrarea asupra acestor sisteme a implicat în primul rând generarea de modele computaționale de mare încredere ale complexelor tetramerice/ternare, având în vedere că nu au fost rezolvate structuri experimentale adecvate pentru niciunul dintre sisteme la momentul în care a fost efectuată cercetarea. Mai mult, aceste modele extrem de precise au permis efectuarea calculelor de energie liberă de legare pentru ambele sisteme prin diferite tehnici bazate pe simularea dinamicii moleculare.

În cele din urmă, al treilea studiu evidențiază modelarea moleculară și aplicarea simulării în farmacogenetică prin descrierea muncii noastre asupra canalului de sodiu dependent de tensiune și efectul diferitelor mutații asupra legării cenobamatului, un nou medicament anti-convulsii. Calculele noastre de legare a energiei libere au arătat că cenobamatul prezintă o afinitate crescută pentru toți mutanții în comparație cu canalele NaV1.5 de tip nativ și se leagă în mod special de mutația N932S. Astfel, cel mai bine ar fi să fiți atenți la prezența acestei mutații atunci când prescrieți cenobamat pacienților epileptici.

Per total, rezultatele prezentate aici dezvăluie, pe de o parte, tendințele actuale în dezvoltarea metodelor de biocomputing și, pe de altă parte, puterea tehnicilor de calcul în cercetarea în științele moleculare ale vieții, mai ales atunci când acestea sunt utilizate în fluxuri de lucru care încorporează rezultate experimentale.

Contribuții Personale

1. Spiridon L, **Șulea TA**, Minh DDL, Petrescu AJ. "*Robosample: A Rigid-Body Molecular Simulation Program Based on Robot Mechanics*", *Biochimica Biophysica Acta General Subjects* 1864(8): 129616, (2020) DOI: 10.1016/j.bbagen.2020.129616;
AI: 1.40(Q2) IF: 3.68
2. Chiritoiu G., Munteanu CVA., **Șulea TA.**, Spiridon L., Petrescu AJ., Jandus C., Romero P., Petrescu SM. "Methionine oxidation selectively enhances T cell reactivity against a melanoma antigen", *iScience*(107205), (2023) DOI: 10.1016/j.isci.2023.107205
AI: 1.63(Q1) IF: 6.10
3. Papadopoulos N, Nédélec A, Derenne A, **Șulea TA**, Pecquet C, Chachoua I, Vertenoeil G, Tilmant T, Petrescu AJ, Mazzucchelli G, Iorga BI, Vertommen D, Constantinescu SN "Oncogenic CALR mutant C-terminus mediates dual binding to the thrombopoietin receptor triggering complex dimerization and activation", *Nature communications* 14(1): 1881, (2023), DOI: 10.1038/s41467-023-37277-3

- AI: 5.61(Q1) IF: 17.69**
4. **Şulea TA**, Draga S, Mernea M, Corlan AD, Radu BM, Petrescu AJ, Amuzescu B . "Differential Inhibition by Cenobamate of Canonical Human Nav1.5 Ion Channels and Several Point Mutants", *International journal of molecular sciences* 26(1), (2025), DOI: 10.3390/ijms26010358
AI: 1.05(Q1) IF: 4.90
 5. **Şulea TA**, Martin EC, Bugeac CA, Bectaş FS, Iacob AL, Spiridon L, Petrescu AJ . "Lessons from Deep Learning Structural Prediction of Multistate Multidomain Proteins-The Case Study of Coiled-Coil NOD-like Receptors", *International journal of molecular sciences* 26(2), (2025) DOI: 10.3390/ijms26020500
AI: 1.05(Q1) IF: 4.90

Referințe

1. Abramson, J. *et al.* (2024) 'Accurate structure prediction of biomolecular interactions with AlphaFold 3', *Nature*, 630(8016), pp. 493–500. Available at: <https://doi.org/10.1038/s41586-024-07487-w>.
2. Ahdriz, G. *et al.* (2024) 'OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization', *Nature Methods*, 21(8), pp. 1514–1524. Available at: <https://doi.org/10.1038/s41592-024-02272-z>.
3. Berman, H.M., Vallat, B. and Lawson, C.L. (2020) 'The data universe of structural biology', *IUCrJ*, 7(4), pp. 630–638. Available at: <https://doi.org/10.1107/S205225252000562X>.
4. Carugo, O. and Djinoić-Carugo, K. (2023) 'Structural biology: A golden era', *PLOS Biology*, 21(6), p. e3002187. Available at: <https://doi.org/10.1371/journal.pbio.3002187>.
5. Cassidy, C.K. *et al.* (2020) 'Structure and dynamics of the E. coli chemotaxis core signaling complex by cryo-electron tomography and molecular simulations', *Communications Biology*, 3(1), p. 24. Available at: <https://doi.org/10.1038/s42003-019-0748-0>.
6. Defour, J.P. *et al.* (2013) 'Tryptophan at the transmembrane-cytosolic junction modulates thrombopoietin receptor dimerization and activation', *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), pp. 2540–2545. Available at: <https://doi.org/10.1073/pnas.1211560110>.
7. Eastman, P. *et al.* (2024) 'OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials', *Journal of Physical Chemistry B*, 128(1), pp. 109–116. Available at: <https://doi.org/10.1021/acs.jpcc.3c06662>.
8. Hewitt, E.W. (2003) 'The MHC class I antigen presentation pathway: strategies for viral immune evasion', *Immunology*, 110(2), pp. 163–169. Available at: <https://doi.org/10.1046/j.1365-2567.2003.01738.x>.
9. Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: <https://doi.org/10.1038/s41586-021-03819-2>.
10. Kelley, L.A. *et al.* (2016) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*, 10(6), pp. 845–858. Available at: <https://doi.org/10.1038/nprot.2015-053>.
11. Kollman, P.A. *et al.* (2000) 'Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models', *Accounts of Chemical Research*, 33(12), pp. 889–897. Available at: <https://doi.org/10.1021/ar000033j>.
12. Krishna, R. *et al.* (2024) 'Generalized biomolecular modeling and design with

- RoseTTAFold All-Atom', *Science*, 384(6693), p. eadl2528. Available at: <https://doi.org/10.1126/science.adl2528>.
13. Livnah, O. *et al.* (1999) 'Crystallographic Evidence for Preformed Dimers of Erythropoietin Receptor Before Ligand Activation', *Science*, 283(5404), pp. 987–990. Available at: <https://doi.org/10.1126/science.283.5404.987>.
 14. Lyskov, S. and Gray, J.J. (2008) 'The RosettaDock server for local protein-protein docking', *Nucleic Acids Research*, 36(Web Server), pp. W233–W238. Available at: <https://doi.org/10.1093/nar/gkn216>.
 15. Martin, E.C. *et al.* (2023) 'NLRscape: an atlas of plant NLR proteins', *Nucleic Acids Research*, 51(D1), pp. D1470–D1482. Available at: <https://doi.org/10.1093/nar/gkac1014>.
 16. Montgomery Pettitt, B. and Karplus, M. (1985) 'The potential of mean force surface for the alanine dipeptide in aqueous solution: a theoretical approach', *Chemical Physics Letters*, 121(3), pp. 194–201. Available at: [https://doi.org/10.1016/0009-2614\(85\)85509-3](https://doi.org/10.1016/0009-2614(85)85509-3).
 17. Phillips, J.C. *et al.* (2020) 'Scalable molecular dynamics on CPU and GPU architectures with NAMD', *Journal of Chemical Physics*, 153(4). Available at: <https://doi.org/10.1063/5.0014475>.
 18. Prentoe, J. *et al.* (2019) 'Hypervariable region 1 and N-linked glycans of hepatitis C regulate virion neutralization by modulating envelope conformations', *Proceedings of the National Academy of Sciences*, 116(20), pp. 10039–10047. Available at: <https://doi.org/10.1073/pnas.1822002116>.
 19. Rajaraman, J. *et al.* (2016) 'An LRR/Malectin Receptor-Like Kinase Mediates Resistance to Non-adapted and Adapted Powdery Mildew Fungi in Barley and Wheat', *Frontiers in Plant Science*, 7. Available at: <https://doi.org/10.3389/fpls.2016.01836>.
 20. Sarson-Lawrence, K.T.G. *et al.* (2024) 'Cryo-EM structure of the extracellular domain of murine Thrombopoietin Receptor in complex with Thrombopoietin', *Nature Communications*, 15(1), p. 1135. Available at: <https://doi.org/10.1038/s41467-024-45356-2>.
 21. Schott-Verdugo, S. and Gohlke, H. (2019) 'PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building', *Journal of Chemical Information and Modeling*, 59(6), pp. 2522–2528. Available at: <https://doi.org/10.1021/acs.jcim.9b00269>.
 22. Shen, W., Zhou, T. and Shi, X. (2023) 'Enhanced sampling in molecular dynamics simulations and their latest applications—A review', *Nano Research*, 16(12), pp. 13474–13497. Available at: <https://doi.org/10.1007/s12274-023-6311-9>.
 23. Sliotweg, E.J. *et al.* (2013) 'Structural Determinants at the Interface of the ARC2 and Leucine-Rich Repeat Domains Control the Activation of the Plant Immune Receptors Rx1 and Gpa2 1 [C][W][OA]', 162(July), pp. 1510–1528. Available at: <https://doi.org/10.1104/pp.113.218842>.
 24. Spiridon, L. *et al.* (2020) 'Robosample: A rigid-body molecular simulation program based on robot mechanics', *Biochimica et Biophysica Acta - General Subjects*, 1864(8). Available at: <https://doi.org/10.1016/j.bbagen.2020.129616>.
 25. Spiridon, L. and Minh, D.D.L. (2017) 'Hamiltonian Monte Carlo with Constrained Molecular Dynamics as Gibbs Sampling', *Journal of Chemical Theory and Computation*, 13(10), pp. 4649–4659. Available at: <https://doi.org/10.1021/acs.jctc.7b00570>.
 26. Syed, R.S. *et al.* (1998) 'Efficiency of signalling through cytokine receptors depends critically on receptor orientation', *Nature*, 395(6701), pp. 511–516. Available at:

- <https://doi.org/10.1038/26773>.
27. Trakhanov, S. *et al.* (1999) 'Structure of a monoclonal 2E8 Fab antibody fragment specific for the low-density lipoprotein-receptor binding region of apolipoprotein E refined at 1.9 Å', *Acta Crystallographica Section D Biological Crystallography*, 55(1), pp. 122–128. Available at: <https://doi.org/10.1107/S090744499800938X>.
 28. Trott, O. and Olson, A.J. (2009) 'AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *Journal of Computational Chemistry*, 31(2), p. NA-NA. Available at: <https://doi.org/10.1002/jcc.21334>.
 29. Tsutsumi, N. *et al.* (2023a) 'Structure of the thrombopoietin-MPL receptor complex is a blueprint for biasing hematopoiesis', *Cell*, 186(19), pp. 4189-4203.e22. Available at: <https://doi.org/10.1016/j.cell.2023.07.037>.
 30. Tsutsumi, N. *et al.* (2023b) 'Structure of the thrombopoietin-MPL receptor complex is a blueprint for biasing hematopoiesis', *Cell*, 186(19), pp. 4189-4203.e22. Available at: <https://doi.org/10.1016/j.cell.2023.07.037>.
 31. Vaidehi, N. and Jain, A. (2015) 'Internal Coordinate Molecular Dynamics: A Foundation for Multiscale Dynamics', *The Journal of Physical Chemistry B*, 119(4), pp. 1233–1242. Available at: <https://doi.org/10.1021/jp509136y>.
 32. Vangone, A. and Bonvin, A.M. (2015) 'Contacts-based prediction of binding affinity in protein–protein complexes', *eLife*, 4. Available at: <https://doi.org/10.7554/eLife.07454>.
 33. Vangone, A. and Bonvin, A.M.J.J. (2015) 'Contacts-based prediction of binding affinity in protein–protein complexes', *eLife*, 4(JULY2015), pp. 1–15. Available at: <https://doi.org/10.7554/eLife.07454>.
 34. Varadi, M. *et al.* (2024) 'AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences', *Nucleic Acids Research*, 52(D1), pp. D368–D375. Available at: <https://doi.org/10.1093/nar/gkad1011>.
 35. Waman, V.P. *et al.* (2024) 'CATH 2024 : CATH-AlphaFlow Doubles the Number of Structures in CATH and Reveals Nearly 200 New Folds', (xxxx). Available at: <https://doi.org/10.1016/j.jmb.2024.168551>.
 36. Wang, K. *et al.* (2013) 'Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics', *Journal of Computer-Aided Molecular Design*, 27(12), pp. 989–1007. Available at: <https://doi.org/10.1007/s10822-013-9689-8>.
 37. Wang, S. *et al.* (2022) 'MM/PB(GB)SA benchmarks on soluble proteins and membrane proteins', *Frontiers in Pharmacology*, 13. Available at: <https://doi.org/10.3389/fphar.2022.1018351>.
 38. Webb, B. and Sali, A. (2018) 'Comparative Protein Structure Modeling Using MODELLER', *Physiology & behavior*, 176(1), pp. 139–148. Available at: <https://doi.org/10.1117/12.2549369.Hyperspectral>.
 39. Williams, C.J. *et al.* (2018) 'MolProbity: More and better reference data for improved all-atom structure validation', *Protein Science*, 27(1), pp. 293–315. Available at: <https://doi.org/10.1002/pro.3330>.
 40. Wu, R. *et al.* (2022) 'High-resolution de novo structure prediction from primary sequence', *bioRxiv* [Preprint]. Available at: <https://doi.org/10.1101/2022.07.21.500999>.
 41. Xue, L.C. *et al.* (2016) 'PRODIGY: A web server for predicting the binding affinity of protein-protein complexes', *Bioinformatics*, 32(23), pp. 3676–3678. Available at: <https://doi.org/10.1093/bioinformatics/btw514>.