



UNIVERSITATEA „POLITEHNICA” din BUCUREȘTI
ȘCOALA DOCTORALĂ CHIMIE APLICATĂ ȘI ȘTIINȚA MATERIALELOR

Nr. Decizie din

REZUMATUL TEZEI DE DOCTORAT

*EXPRESIA PROTEINEI DIN CAPSIDA ssCCMV ȘI SIMULARE DE DINAMICĂ
MOLECULARĂ A CAPSIDULUI ssCCMV*

*EXPRESSION OF THE ssCCMV CAPSID COAT PROTEIN AND MOLECULAR
DYNAMICS SIMULATIONS ON THE ssCCMV CAPSID*

Autor: János SZÖVÉRFI

Conducător de doctorat: Prof. dr. ing. Szabolcs LÁNYI

Cuvinte cheie: heterologous expression, protein purification, molecular dynamics,
CCMV capsid, docking, mutation, protein dimers

COMISIA DE DOCTORAT

Președinte	Prof. dr. ing. Ileana RĂU	de la	Universitatea „Politehnica” din București
Conducător de doctorat	Prof. dr. ing. Szabolcs LÁNYI	de la	Universitatea „Politehnica” din București
Referent	Prof. dr. ing. Cristina COSTACHE	de la	Universitatea „Politehnica” din București
Referent	Prof. Habil. dr. ing. Csaba PAIZS	de la	Universitatea „Babes Bolyai”, Cluj Napoca
Referent	Prof. univ. dr. habil László SZILÁGYI	de la	Universitatea Sapientia din Cluj-Napoca

BUCUREȘTI

2022

Contents

1.	Introducere și obiective	5
3.	Metode	7
3.1.	Clonarea, exprimarea și purificarea proteinei din capsida CCMV	7
3.1.1.	Transformare bacteriană	7
3.1.2.	Clonarea plasmidei prin digestia enzimelor de restricție (subclonare)	7
3.1.3.	Expresie de test în shake flask	7
3.1.4.	Optimizarea expresiei proteinelor	7
3.1.6.	Electroforeză cu gel de agaroză	7
3.1.7.	Electroforeză cu gel de SDS polyacrylamide	8
3.1.8.	Purificarea proteinelor	8
3.1.9.	Digerarea proteinei de fuziune a pUBK_CCMV	8
3.2.	Generarea structurilor dimer cu ZDOCK	8
3.2.6.	Minimizarea energiei cu CUDAGMIN	8
3.3.	Sisteme studiate	9
3.4.	Simulări de dinamică moleculară	9
3.4.1.	MD de lungă durată	10
3.4.2.	MD la temperaturi diferite	10
3.4.3.	Simularea dinamicii moleculare cu schimb de replici	10
3.4.4.	Simulări accelerate ale dinamicii moleculare	10
3.5.	Mutații punctuale pentru dimererele T1	11
3.6.	Analiza traiectoriilor	11
3.6.1.	Analiza MM-GBSA	11
4.	Rezultatele	12
4.1.	Clonarea, exprimarea și purificarea proteinei din capsida CCMV	12
4.1.1.	Construcția plasmidei	12
4.1.2.	Transformare și clonare	12
4.1.3.	Digestia plasmidelor	13
4.1.4.	Ligare	13
4.1.5.	Expresie de test în shake flask	14
4.1.6.	Optimizarea expresiei proteinelor	14
4.1.7.	Purificarea proteinei pUBK_CCMV	14
4.1.8.	Deubiquitinarea cu hidrolază C-terminală ubiquitin 1 (YUH1)	15
4.2.	Minimizarea energetică a structurilor dimerice generate cu ZDOCK	16

4.3.	Compararea simulărilor MD scurtă durată pentru dimerile T1, T2 și T3.....	20
4.4.	Dimere selectate pentru investigarea dinamicii.....	21
4.5.	Simulare MD de lungă durată în solvent explicit pentru dimerile de proteine T1, T2, T3 și TX	21
4.6.	Simulări de dinamică moleculară la temperaturi diferite	23
4.7.	Mutații dimerului T1	26
4.7.1.	Mutații de deleție a dimerelor T1 și T2.....	26
4.7.2.	Mutații punctuale ale dimerului T1	28
4.8.	Simulări de dinamică moleculară cu schimb de replici.....	30
4.9.	Dinamica moleculară accelerată pe T1, T2, T3 timp de 500 ns în ansamblul NPT	31
4.10.	Pentamer ale dimerelor.....	32
5.	Concluziile.....	34
5.1.	Rezultate și contribuții originale.....	34
5.2.	Lista publicațiilor originale	35
5.2.1.	Publicații.....	35
5.2.2.	Conferințe	36
5.3.	Perspective pentru dezvoltări ulterioare.....	36
	Bibliografie	37

Mulțumiri

În primul rând, aș dori să-i mulțumesc prof. dr. Lányi Szabolcs pentru că mi-a permis să particip la studii doctorale sub conducerea sa și pentru sprijinul acordat, care mi-a permis să-mi completez teza de doctorat.

Aș dori să mulțumesc lui prof. dr. Albert Beáta și dr. Orbán Csongor pentru că m-au introdus în lumea cercetării în domeniul biotehnologiei și pentru că m-au ajutat și sprijinit întotdeauna atunci când a fost nevoie.

Aș dori să le mulțumesc colegilor mei doctoranzi, în special lui Molnár-Nagy Katalin, Salamon Pál și Balázs-Bécsi Ildikó pentru că m-au ajutat cu munca mea de laborator și au făcut serile lungi din laborator suportabile cu buna lor dispoziție.

Aș dori să-i mulțumesc dr. Fejér Szilárd pentru că m-a luat în echipa sa și m-a ajutat altruist pe tot parcursul studiilor doctorale. Am petrecut mulți ani frumoși în echipa de cercetare a Fundației Provitam și am învățat de la el cum este un adevărat lider.

Mulțumesc tuturor celor care m-au ajutat în vreun fel și m-au motivat pentru a putea ajunge la finalul studiilor.

În cele din urmă, cele mai mari mulțumiri ale mele se îndreaptă către soția mea, care a fost alături de mine de-a lungul anilor de cercetare, m-a susținut și mi-a îndurat absența cu răbdare. Îi sunt recunoscătoare fiului meu că în pauzele orelor stresante a reușit mereu să mă facă să mă simt mai bine.

1. Introducere și obiective

În zilele noastre, chimia computațională are o importanță din ce în ce mai mare în științele vieții. Multe fenomene ale proceselor biologice sunt examinate pe scară largă in vivo și in vitro, dar în ultimele decenii cercetarea in silico a avut o evoluție spectaculoasă.

Virusurile sunt cele mai simple organisme, prin urmare sunt investigate pentru înțelegerea proprietăților fundamentale și a interacțiunilor proteinelor, acizilor nucleici și a altor componente [1][2][3]. Prin studierea formării virușilor, putem obține o perspectivă valoroasă asupra auto-asamblării, un fenomen întâlnit aproape peste tot la diferite dimensiuni. Capsidul virusurilor este de obicei compus din proteine de înveliș, înmulțite pentru a forma o capsidă simetrică în jurul genomului. Folosind instrumente de chimie computațională, căutăm să înțelegem cum și de ce aceste proteine sunt aranjate pentru a forma aceste capside de structură bine definită.

Virusul cowpea chlorotic mottle (CCMV) a fost investigat de zeci de ani datorită capacității sale de a forma capside goale complet funcționale pornind de la monomeri în diferite condiții [5]. Studiile pentru înțelegerea procesului de auto-asamblare în detaliu au fost efectuate în silico, in vitro și in vivo, cu rezultate semnificative, cu toate acestea multe întrebări rămân deschise.

Primul obiectiv a fost expresia in vitro a proteinei capsidice CCMV și studiul dimerizării proteinei. Proteina capsidică ss-CCMV va fi exprimată în *Escherichia coli* folosind o plasmidă ubiquitinată și purificată pentru investigații suplimentare.

Gena proiectată este achiziționată de la agenți comerciali și clonată într-un vector bazat pe ubiquitin-pET 19b. Plasmida va fi transformată în linia celulară *E. coli* BL21(DE3) Rosetta pentru expresie heterologă la scară mică.

Eticheta His-tag încorporată face posibilă purificarea proteinelor cu cromatografie de afinitate de nichel și cromatografie cu permeație pe gel pentru a obține o proteină pură.

În ultima parte a experimentelor in vitro se va efectua digestia proteinei de fuziune cu enzima ubiquitin C-terminal hydrolase 1 care va rezulta proteina capsidică ss-CCMV.

Scopul principal al cercetării noastre este de a înțelege mai profund procesele și efectele care afectează dimerizarea proteinelor capsidice și formarea capsidului complet.

Studiile anterioare in silico au arătat rezultate cu privire la primii pași de formare a dimerului pe niveluri atomistice. Diverse simulări de dinamică moleculară pe diferite intervale de timp și temperaturi au avut rezultate în ceea ce privește stabilitatea dimerelor proteice [7].

Pornim de la structura experimentală 3D a proteinei capsidice pentru a simula formarea subunităților. Interacțiunile dintre proteine sunt examinate pentru a prezice formarea capsidului. Cele trei catene ale unității asimetrice proteice (A, B, C) sunt separate și combinate în posibile dimere. Perechile de dimere proteice sunt relaxate cu protocolul ClassicRelax din Rosetta și se efectuează un proces de andocare pe serverul ZDOCK pentru a găsi cele mai bune 2000 de structuri dimer prezise. Trei tipuri (BB, BC, CC) din dimerele rezultate sunt supuse minimizării a energiei în pachetul de software AMBER. Scopul este de a găsi dimerul cu cea mai mică energie de legare (cel mai stabil dimer) și de a compara această structură cu diferitele interfețe dimerice cunoscute.

În capitolul 4.1 – 4.10. studiem proprietățile de auto-asamblare ale proteinelor capsidice ale virusului CCMV folosind metode de calcul de ultimă generație. Dimerele proteinei din capsidă sunt investigate pentru a înțelege natura și forțele motrice ale auto-asamblării.

În prima fază a activității, am găsit o corelație între interfețele de legare cunoscute ale dimerelor proteice și cele modelate, respectiv potrivirile structurale pentru toate cele trei tipuri de interfețe. Putem prezice că dimerele T1 CC sunt cele mai importante în procesul de formare a capsidelor, rezultat care este în bună concordanță cu calea determinată experimental.

Simulările de dinamică moleculară sunt efectuate pentru structurile cu cel mai mic RMSD în raport cu interfețele originale (determinate cristalografic) pe diferite scări de timp și temperatură, pentru

a studia stabilitatea dimerelor și pentru a determina posibilitatea disocierii diferiților dimeri în anumite condiții.

Studierea stabilității proteinelor în diferite condiții oferă o mai bună înțelegere a funcției și comportamentului lor. Auto-asamblarea capsidului virusului CCMV poate fi modelată pe intervale de timp computaționale realiste dacă luăm în considerare disocierea complexelor proteice ca un proces invers al agregării.

Scopul studiilor de dinamică moleculară prezentate în capitolul 4.2.-4.6. este de a găsi și înțelege condițiile care conduc la o separare a dimerilor proteice stabili. Rezultatele prezentate în capitolul 4.1 au arătat că interfețele de tip 1 (T1) și de tip 2 (T2) sunt importante în timpul formării capsidului, în timp ce tipul 3 (T3) este un dimer instabil [6]. Am efectuat simulare dinamică pe dimerile de proteine T1, T2 și T3 pentru a obține o imagine mai bună asupra interacțiunilor dintre subunitățile capsidului. 7 mutații ale dimerului de T1 au fost, de asemenea, utilizate în cercetare.

Simulările efectuate au fost: dinamica moleculară de lungă durată, dinamica moleculară clasică (cMD) la diferite temperaturi, dinamica moleculară de schimb de replici (REMD) și dinamica moleculară accelerată (aMD).

O modalitate de a evalua stabilitatea interfeței dimer este de a introduce mutații punctuale și de a evalua modul în care interfața răspunde la acestea. Am generat un număr mare de mutații punctuali și le-am evaluat stabilitatea relativă cu diferite metode.

3. Metode

Acest capitol descrie metodele utilizate pe parcursul acestei lucrări, ale căror rezultate sunt prezentate în capitolul 4.

3.1. Clonarea, exprimarea și purificarea proteinei din capsida CCMV

3.1.1. Transformare bacteriană

Plasmidele proiectate sunt introduse în celulele bacteriene în timpul transformării. Plasmida pUC57 care conține ADN-ul proteinelor CCMV a fost transformată în linie celulară de clonare *E. coli Top10*.

3.1.2. Clonarea plasmidei prin digestia enzimelor de restricție (subclonare)

Amestecul pentru digestie a fost preparat în mod similar pentru PUC57 și pUBK: 2 μl B, 1,5 μl SaCII, 1,5 μl BamHI și 5 μl UP cu 5 μl PUC57, respectiv 10 μl pUBK în 20 μl volum final.

După digestie, ambele probe au fost purificate prin electroforeză cu gel de agaroză. Procesul de ligare s-a efectuat simultan timp de 2 ore pe 25°C și peste noapte pe 16°C. Următorul pas este transformarea noilor plasmide într-o celulă competentă TOP10. 50 μl de celulă competentă cu 5 μl de plasmidă a fost supusă șocului termic la 42 °C timp de 90 sec, apoi a fost agitată la 37 °C. Colonia de celule dorită a fost izolată cu Plasmid Miniprep Kit. O digestie de diagnosticare a fost efectuată cu endonucleaze de restricție SaCII și BamHI, așa cum este descris anterior.

3.1.3. Expresie de test în shake flask

Coloniile selectate au fost transformate într-o linie celulară gazdă *E. coli BL21(DE3) Rosetta*. Celulele au fost cultivate pe plăci de agar LB, conținând kanamicină și o colonie bine izolată a fost inoculată în 10 ml de mediu LB cu 30 μg/ml kanamicină și incubată timp de 4 ore la 37 °C într-un incubator de agitare. Cultura celulară a fost transferată în 200 ml de mediu LB și cultivată la OD₆₀₀ = 0,7. Expresia a fost indusă cu 0,8 mM IPTG și incubată timp de 3 ore la 37 °C. Bulionul de cultură a fost centrifugat timp de 10 minute pe 20000 RPM, iar peletele au fost depozitate la -80°C până la prelucrarea ulterioară.

3.1.4. Optimizarea expresiei proteinelor

Pentru a maximaliza expresia proteică a celulelor bacteriene am planificat un proces de optimizare cu schimbarea concentrației IPTG și temperatura după cum urmează:

- 0.1 mM IPTG la 18°C
- 1 mM IPTG la 18°C
- 0.5 mM IPTG la 37 °C
- 0.5 mM IPTG la 18°C
- 0.1 mM IPTG la 37 °C
- 1 mM IPTG la 37 °C

Experimentul a fost realizat în paralel pentru probele selectate din rezultatele digestiei.

3.1.6. Electroforeză cu gel de agaroză

Separarea ADN-ului genomic, rezultatul digeriției enzimei de restricție, a fost efectuată prin electroforeză cu gel de agaroză. Gelul a fost preparat pentru 1% g/v de agaroză și dizolvat în tampon TAE. Redsafe a fost adăugat la soluția fiartă.

5 μl de probe de ADN au fost amestecate cu 1 μl de colorant 6X și introduse în gel. Electroforeza a fost efectuată la 90 V, iar gelurile au fost vizualizate sub lumina UV.

3.1.7. Electroforeză cu gel de SDS polyacrylamide

De obicei, 25 μ l de probă proteică a fost amestecată cu 25 μ l de tampon de probă SDS 4x, încălzită la 98 °C timp de 5 minute și centrifugată timp de 10 minute pe 14000 rpm. 10 μ l de probă a fost introdus în buznarul gelului. Analiza a fost efectuată cu un gel de stacking de 5% și un gel de rezoluție de 12% la o tensiune constantă de 120 V pe gheață. După electroforeză gelul a fost tratat cu colorant Coomassie Brilliant Blue timp de 40 de minute cu agitare constantă.

3.1.8. Purificarea proteinelor

4 ml de probă proteică au fost încărcate pe o coloană HisTrap HP de 5 ml și s-a utilizat un sistem de purificator Akta. Deoarece proteina noastră conținea 10X His tag, puteam utiliza cromatografia de afinitate de nichel. Margele noi sau regenerate NTA-agaroză au fost utilizate pe o coloană într-un raport 1:10 margele NTA:proteine. Margelele au fost introduse în coloană și spălate cu 10 coloane volum (CV) tampon de spălare. Proba a fost adăugată la margele și laminată timp de 40 de minute la 4 °C. Debitul a fost colectat și s-a adăugat 1 ml tampon de spălare de 8 ori, urmat de eluție de 8 ori. Frațiunile de 1 ml au fost analizate cu SDS-PAGE.

3.1.9. Digerarea proteinei de fuziune a pUBK_CCMV

Plasmida YUH1 a fost transformată în linie celulară gazdă *E. coli BL21(DE3) Rosetta* și inoculată în mediu LB. Amestecul a fost incubat la 37 °C, agitare pe 250 RPM timp de 60 de minute, și cultivate pe plăci de agar care conțin kanamicină pe 37°C. A doua zi, o colonie izolată a fost inoculată în mediu LB de 7,5 ml, cu un stoc de kanamicină de 2,3 μ l (80 μ g/ml) și incubată la 37°C până la OD₆₀₀=0,67. Amestecul a fost indus cu 200 μ l de IPTG (1 mol/l). Bulionul de cultură a fost recoltat prin centrifugare la 4500 RPM 4°C timp de 10 minute. Pentru purificare peletele au fost suspendate în tampon de lizare de 4X2 ml și sonicate timp de 3X20 sec. După 60 de minute de centrifugare la 20000 RPM și 4 °C, supernatantul a fost colectat și purificat cu cromatografie de afinitate a nichelului. Frațiunile de 1 ml au fost colectate și analizate cu SDS-PAGE.

Enzima YUH1 [8] taie ubiquitina de pe proteina purificată, rezultând monomerul proteinei capside CCMV. Două fracții (CCMV1 și CCMV2) colectate din separarea FPLC au fost concentrate cu filtru centrifugal Amicon Ultra. 3,5 ml din ambele probe au fost centrifugate timp de 40 de minute la 4000 RPM, iar concentratul a fost colectat din rezervorul de probă al dispozitivului de filtrare. Concentrația soluțiilor a fost măsurată cu spectrofotometru Nanodrop la 260 nm. Soluțiile au fost amestecate în raport 1:5 enzimă:proteină și digerat timp de 60 de minute la temperatura camerei. Amestecurile au fost separate cu cromatografie de afinitate a nichelului, cum s-a descris anterior.

3.2. Generarea structurilor dimer cu ZDOCK

Proteina studiată este o proteină de capsidă din virusul ss-CCMV [9], rezoluție 2.7 Å. Proteina conține 3 catene: A, B, C. Am generat o simulare de andocare proteină-proteină pe serverul ZDOCK [10] pentru posibilele permutății ale catenei B și ale catenei C: dimerul BB, dimerul BC, dimerul CC. Catenele structurilor monomere au fost relaxate cu protocolul ClassicRelax din Rosetta [11] pentru a elimina orice suprapunere. Cele mai bune 2000 de dimere prezise din aceste structuri, au fost salvate pe baza funcției de notare ZDOCK.

3.2.6. Minimizarea energiei cu CUDAGMIN

Fisierele pdb rezultate din procesul de andocare au fost pregătite pentru minimizarea energiei în software-ul tleap, din pachetul software AmberMD[12]. Câmpul de forță folosit a fost ff03 [13], o versiune modificată a ff99 [14]. Pentru solvatare, am folosit generalized Born solvation model

(GB). Opțiunea $igb=2$ folosită de noi este o versiune modificată a modelului GB[16]. Setul de raze recomandat pentru modelul $igb=2$ este $mbondi2$.

Concentrația de sare a fost setată la 0,1 M. Pentru minimizările inițiale, a fost utilizată un cutoff de 12 Å. Ulterior, pentru rulările CUDA (calculul câmpului de forță optimizat pentru unitățile de procesare grafică - GPU-uri), nu au fost utilizate cutoff-uri pentru evaluarea termenilor non-bonding ai potențialului. Valoarea $rGBmax$ a fost setată la 8.23 Å. Minimizarea energetică a modelelor a fost efectuată folosind software-ul GMIN [17].

Am folosit partea de minimizator L-BFGS din GMIN, nu s-au luat în considerare pașuri de basin-hopping pentru structuri.

Evaluarea energiei pentru sistemul care utilizează un GPU nVidia Tesla K40 este de aproximativ 100-200 de ori mai rapidă decât aceeași evaluare folosind doar un procesor CPU. Setul de dimmeri (2000 de structuri pentru BB, BC și CC) a fost minimizat atât în formă asociată, cât și disociată. Criteriile de convergență (SLOPPYCONV și TIGHTCONV) au fost stabilite la 10^{-2} . Parametrul MAXERISE este setat la 10^{-4} .

Energia de legare pentru fiecare dimer este diferența dintre energiile potențiale pentru forma asociată și disociată a dimerelor, rezultată din rularea minimizării. Cea mai bună aliniere a fost căutată pe baza a trei tipuri de interfețe găsite în capsida virală icosaedrală:

- tip 1 (T1): CC, BA
- tip 2 (T2): BC1, BC2, AA1, AA2
- tip 3 (T3): AC, AB

Am folosit alfa-carbonii reziduurilor de la 40 la 190 în fiecare structură pentru calculul RMSD. Am folosit rutina PERMOPT în GMIN pentru evaluarea RMSD-urilor ale catenei de bază.

3.3. Sisteme studiate

Patru structuri au fost folosite pentru a studia comportamentul dimerelor CCMV:

- Cele mai bune structuri din optimizarea globală au fost utilizate pentru tipul 1 (T1), tipul 2 (T2) și tipul 3 (T3)
- Structura cu cel mai mare RMSD comparat cu interfața T3, numit TX
- Mutații de deleție, unde 10 reziduuri din terminalele N ale ambelor lanțuri din dimerele T1 (T1DM) și T2 (T2DM) au fost eliminate
- cinci mutații cu mutații punctuale de T1 și doi mutații pe baza profilului de sensibilitate la mutație generat, cu cel mai mare $\Delta\Delta G$ și unul cu cel mai mic $\Delta\Delta G$
- un pentamer de dimere din capsida ss-CCMV

3.4. Simulări de dinamică moleculară

Simulările de dinamică moleculară au fost efectuate cu modulul $pmemd.cuda$ al pachetului software Amber14. Câmpul de forță folosit a fost $ff03$ [13]. Simulările au fost efectuate cu solvent implicit și explicit. Solvarea explicită a fost făcută cu o cutie octaedrală de TIP3P [18], unde moleculele de apă au fost adăugate până la o distanță de 8 Å de la proteină.

Protocolul MD a fost adaptat de la Mafucci *et al.* [19] și conține următorii pași:

1. minimizarea atomilor de hidrogen (1000 de cicluri de steepest descent și 5000 de cicluri de gradient conjugat)
2. minimizarea moleculelor de apă (2000 de cicluri de steepest descent și 5000 de cicluri de gradient conjugat)

3. echilibrarea cutiei de solvent la 300 K cu 100 ps nvt și 100 ps de simulare NPT folosind un termostat Langevin
4. minimizarea catenelor laterale și a apelor cu restricția catenei principale de 25 kcal/mol
5. minimizare totală cu restricții pe catena principală de 10 kcal/mol (2500 de cicluri de steepest descent și 5000 de cicluri de gradient conjugat)
6. încălzirea sistemului la 300 K în 6 trepte de câte 5 ps fiecare (LT = 50 K), unde restricțiile catenei principale au fost reduse de la 10,0 kcal/mol la 5 kcal/mol
7. echilibrarea completă în ansamblul NVT (100 ps, restricția catenei principale = 5.0 kcal/mol) și în ansamblul NPT (1 pas de 200 ps, restricția catenei principale = 5 kcal/mol; 3 pași de 100 ps fiecare, reducerea restricțiilor ale catenei principale de la 5,0 kcal/mol la 1,0 kcal/mol și 1 treaptă de 1 ns cu 1,0 kcal/mol de restricția catenei principale)
8. simulațiile de producție au fost efectuate la 300K

Un cutoff electrostatic de 8,0 Å, un barostat Berendsen, o însumare Particle Mesh Ewald (PME) pentru interacțiuni electrostatice cu rază lungă de acțiune și algoritmul SHAKE au fost aplicate tuturor calculelor cu apă explicită.

3.4.1. MD de lungă durată

Dimerele T1, T2 și T3 au fost simulate în apă explicită timp de 2 μs folosind protocolul de la 4.3.. Simularea a fost efectuată pe 300 K, reglarea temperaturii a fost prevăzută cu termostatul Langevin (NTT = 3) cu o frecvență de coliziune de GAMMA_LN = 2. Condițiile periodic boundary au fost aplicate folosind un cutoff cu 8 Å pentru a rezolva interacțiunile non-bonding cu rază lungă de acțiune. Intervalul de timp folosit a fost de 2 fs (DT=0.002). Metoda Particle Mesh Ewald [20] și algoritmul SHAKE [21] a fost aplicat pentru a limita lungimile legăturilor pentru atomii de hidrogen.

3.4.2. MD la temperaturi diferite

Patru simulări paralele în solvent implicit au fost efectuate pentru T1, T2, T3 și TX timp de 200 ns la 350 K. Nu s-a utilizat nici un cutoff (CUT=999), deoarece nu au existat condiții de limitare periodică. 600 ns de simulare cu solvent explicit au fost efectuate pentru T1, T2, T3 și TX la 350 K cu aceiași parametri ca cei enumerați la 4.3.

3.4.3. Simularea dinamicii moleculare cu schimb de replici

Simulările REMD bazate pe temperatură au fost utilizate pentru dimerele menționate la punctul 3.1. la 8 temperaturi diferite: 300.00 K, 306.65 K, 313.42 K, 320.28 K, 327.25 K, 334.34 K, 341.53 K, 348.81 K, 356.24 K. Temperaturile au fost generate pe serverele web <http://folding.bmc.uu.se/remd/> [22]

Au fost efectuate 8 simulații paralele pentru 25 ns (200 ns în total).

3.4.4. Simulări accelerate ale dinamicii moleculare

aMD a fost efectuat pe dimere T1, T2, T3 pentru 500 ns în ansamblul NPT cu parametri descriși la 4.3., cu adăugarea parametrilor specifici aMD enumerați mai jos. Valorile potențiale ale energiei și ale energiei dihedrale au fost extrase din dinamica moleculară convențională de 2 μs.

Table 4-1 parametrii de simulare specifici aMD pentru trei tipuri de dimere

Dimer	EthreshD (kcal/mol)	alphaD	EthreshP (kcal/mol)	alphaP
T1	3527	231	-145491	8186
T2	4457	184	-152923	799
T3	4438	184	-333461	799

3.5. Mutații punctuale pentru dimerul T1

S-au efectuat mutații *in silico* pe aminoacizii de interfață ale dimerului T1 cu rutina `MUTATE_MODEL` a `MODELLER` [23]. Cei 178 de mutații rezultați au fost minimizați în forme asociate și disociate pentru a calcula energiile de legare. Simulările dinamicii moleculare au fost efectuate timp de 100 ns pe 300 K pe 5 mutații selectați, cu protocolul descris la punctul 3.4.

3.6. Analiza traiectoriilor

Analiza și vizualizarea traiectoriei au fost realizate cu software-ul `cpptraj` din pachetul Amber14 [24], `VMD` [25] și `UCSF Chimera` [26].

`Cpptraj` poate citi fișierele de topologie și coordonate din diferite simulări MD și au fost utilizate următoarele calcule: `rmsd`, `atomicfluct`, `molsurf` [27], `secstruct` [28].

3.6.1. Analiza MM-GBSA

MM-GBSA face parte din pachetul software Amber. Calculul a fost efectuat cu scriptul `MMPBSA.py` [29]. Analiza a fost efectuată pentru 10 ns de simulație MD. Dimerele proteice pentru rularea MM-GBSA au fost selectate din cele 2000 de structuri minimizate cu `GMIN`, cu un `iRMSD` (`RMSD` a mișcării aminoacizilor de pe interfață) mai mic de 3 Å în ceea ce privește interfața originală. Calculul a fost efectuat pentru 7 dimeri de T1 și 5 dimeri de T2. Din moment ce nu am găsit dimer de interfață T3 pentru criteriile `iRMSD`, am folosit dimerul original T3 A-B din capsida completă și dimerul T3 cu cea mai mică valoare `iRMSD`.

4. Rezultatele

4.1. Clonarea, exprimarea și purificarea proteinei din capsida CCMV

4.1.1. Construcția plasmidei

Gena proteinei ss-CCMV, achiziționată de la agenți comerciali, a fost clonată în vectorul de clonare pUC57 cu site-uri de restricționare Xba1 și BamH1.

Gena a fost transformată într-un vector pUBK (pUBK_CCMV).

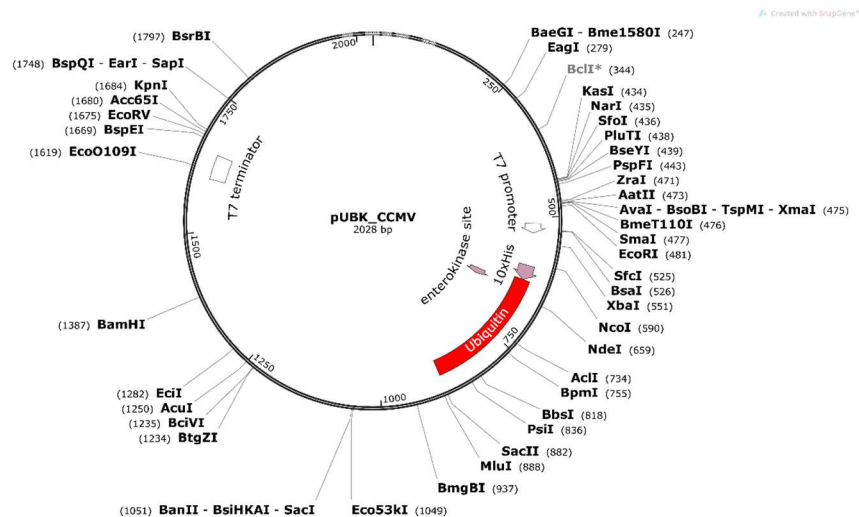


Fig. 4.-1 Harta plasmidei pUBK_CCMV

Plasmida conține o ubiquitină care asigură detectarea și creșterea activității biologice prin afectarea solubilității proteinei [30].

4.1.2. Transformare și clonare

Gena proteinei capsidice ss-CCMV a fost clonată în pUC57 prin strategia de clonare EcoRV (GenScript). Plasmida care conține ADN-ul proteinelor CCMV a fost transformată în linie celulară de clonare *E. coli Top10*.

4.1.3. Digestia plasmidelor

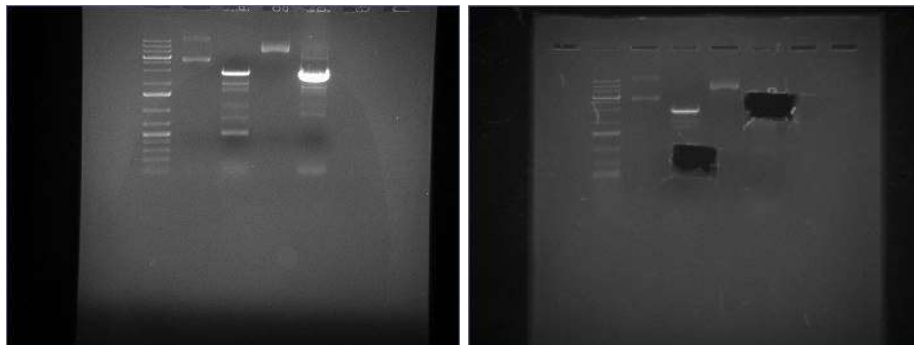


Fig. 4-1 Digestia plasmidei pUBK_CCMV

Plasmidele au fost digerate cu o reacție dublă de digestie folosind două endonucleaze de restricție: BamHI și SacII la 37 °C timp de 60 de minute. După reacție, enzimele au fost inactivate la 80 °C timp de 20 de minute și examinate cu electroforeză cu gel de agaroză 1,2% (Fig.4-2). Fragmentele corespunzătoare au fost tăiate și purificate cu *Gene Jet Gel Extraction Kit*. Concentrația vectorului (pUBK) a fost de 23.869 μg/ml, iar pentru insert (ss-CCMV) de 1,56 μg/ml.

4.1.4. Ligare

În timpul reacției de ligare au fost combinate vectorul și inserția preparată prin reacția de digestie. Reacția a fost promovată de enzima ligazei ADN T4 și efectuată timp de 2 ore la 25 °C și peste noapte la 16 °C. Produsul a fost transformat în linia celulară *E. coli Top10* și cultivat pe plăci de agar care conțineau ampicilină, la 37 °C peste noapte. 4 colonii izolate au fost inoculate în mediu BL și incubate peste noapte la 37 °C într-un incubator de agitare. Plasmidele recombinante au fost identificate cu reacții duble de digestie, așa cum au fost descrise anterior în 13 amestecuri identice. Rezultatele digestiei sunt prezentate pe Fig. 4-3 Fig. 4-23.

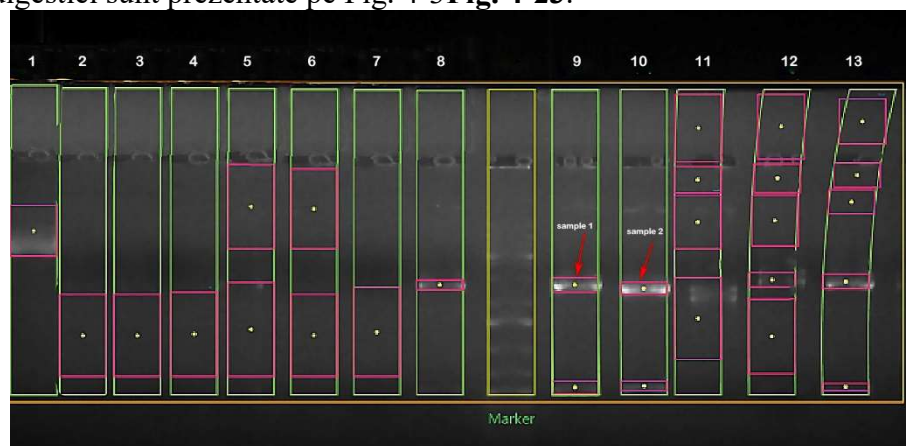


Fig. 4.-3 Reacțiile de dublă digestie pentru 13 probe de plasmidă

Două mostre (benzile 9 și 10) au fost selectate pentru prelucrare ulterioară.

Pentru a confirma selecția, a fost efectuată secvențierea genelor din probe. Mosra 1 a arătat o asemănare crescută cu cea teoretică, astfel încât acest stoc a fost folosit pentru exprimare.

4.1.5. Expresie de test în shake flask

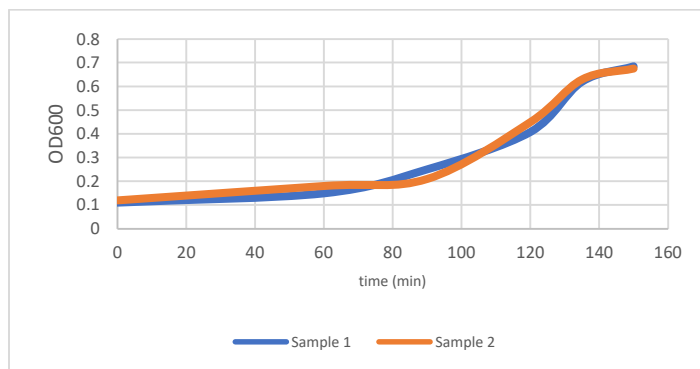


Fig. 4-4 Creșterea *E. coli* în timpul expresiei de testare

Stocurile de celule selectate au fost utilizate pentru expresia proteinelor. În primul rând, plasmidele au fost transformate în linia celulară gazdă *Escherichia coli* BL21(DE3). Fig. 4-4), cultura celulară a fost indusă cu 0,5 mM IPTG și incubată timp de 3 ore la 37 °C. Suspensia celulară a fost centrifugată timp de 10 minute pe 20000 rpm și 4 °C. Peleții au fost depozitați la -80 °C .

4.1.6. Optimizarea expresiei proteinelor

După expresia de test, am planificat o optimizare pentru a maximiza randamentul expresiei. Rezultatele SDS-PAGE au arătat că condițiile optime sunt concentrația IPTG de 0,1 nM și temperatura de incubație de 37 °C (Fig. 4-6, lane 9).

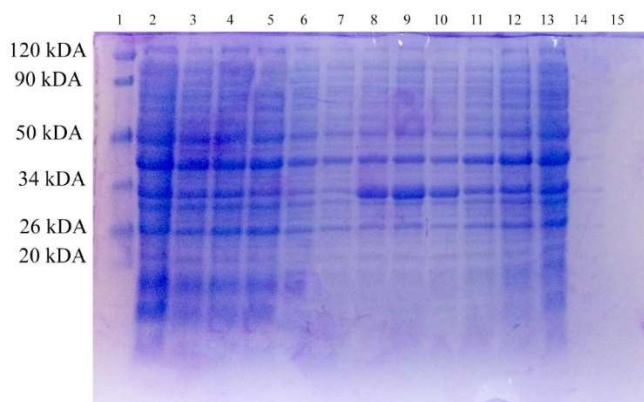


Fig. 4.-6 SDS-PAGE for optimization of pUBK_CCMV expression

4.1.7. Purificarea proteinei pUBK_CCMV

Purificarea pUBK_CCMV a fost efectuată pe un sistem FPLC AKTA (Amersham Biosciences, Uppsala, Suedia). Frațiile au fost colectate separat: verde (CCMVI) și roșu (CCMVII). Frațiile care conțin proteinele au fost dializate și depozitate pentru analize suplimentare.

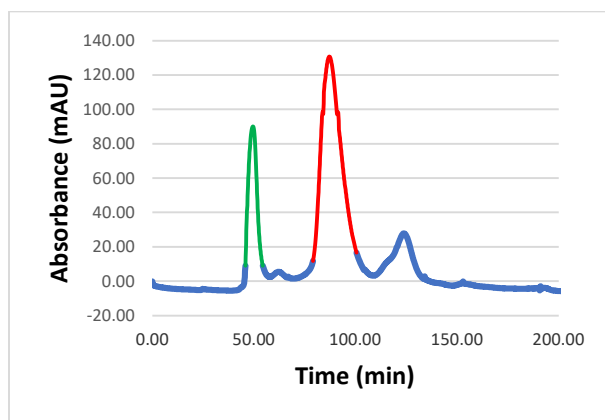


Fig. 4.-7 Cromatogramă pentru purificarea expresiei proteinelor pUBK_CCMV

Rezultatele purificării FPLC au fost verificate prin SDS-PAGE (**Fig. 4-8**). Proteinele au fost separate în mod corespunzător; cu toate acestea, alte proteine au fost legate lângă proteina CCMV.

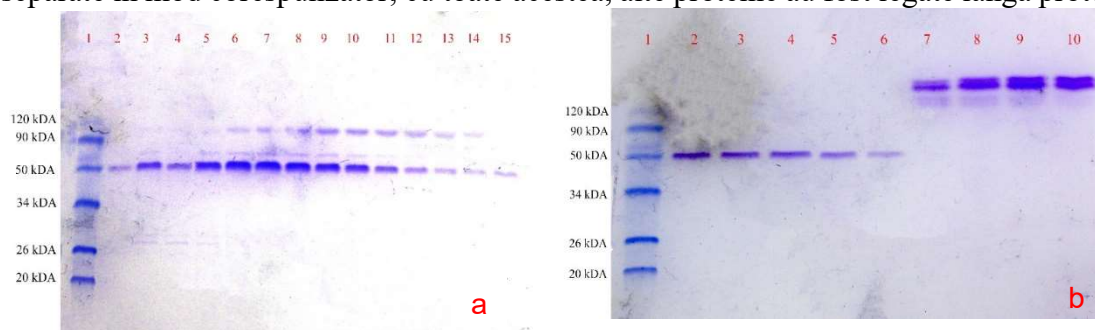


Fig. 4-8 Verificarea SDS-PAGE după purificarea pUBK_CCMV pe FPLC

4.1.8. Deubiquitinarea cu hidrolază C-terminală ubiquitin 1 (YUH1)

Proteina YUH1 a fost folosită pentru a tăia ubiquitina de pe proteina capsidică CCMV. Ambele fracții colectate din purificarea FPLC au fost digerate cu YUH1. 8 ml de soluție CCMV1 a fost concentrată și a rezultat 2,2 ml cu o concentrație de 550 $\mu\text{g/ml}$. Soluțiile au fost amestecate într-un raport proteină: enzimă de 1:5 și după 1 oră de incubație a fost purificat cu cromatografie de afinitate Ni-NTA.

După efectuarea unei analize SDS_PAGE, proteina CCMV a fost detectată în mod clar (**Fig. 4-8**, lane 2).

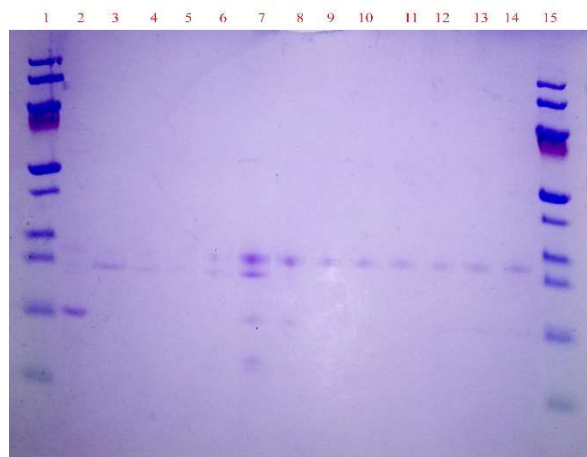


Fig. 4-11 Digestia pUBK_CCMV cu YUH1

Proteina capsidică CCMV poate fi produsă cu construcția plasmidă prezentată, prin expresie heterologă. Deoarece această metodă este comună și eficientă, se poate obține un randament ridicat de proteine.

4.2. Minimizarea energetică a structurilor dimerice generate cu ZDOCK

Unitatea asimetrică din structura de referință (PDB ID 1za7) conține trei catene (**Fig. 4-12**), cu o conformație aproape identică. În capsidul complet al virusului cu simetrie icosaedrală, se pot observa 8 interfețe posibile între aceste catene (AA1, AA2, AB, AC, BA, BC1, BC2, CC). Acestea pot fi clasificate în trei tipuri de interfețe, pe care le numim T1, T2 și T3.

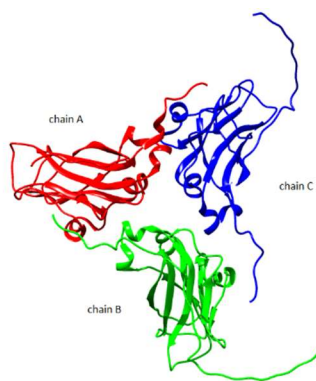


Fig. 4-12 The crystal structure of the asymmetric unit of ss CCMV(PDB ID 1za7)

Pentru generarea structurilor dimer prin serverul ZDOCK, introducem coordonatele lanțului B sau C ca "substrat" și aceleași lanțuri pentru "ligand". Prin urmare, calculările ZDOCK au fost efectuate pe următoarele perechi: BB, BC și CC. Am lăsat coordonatele catenei A, deoarece structura cristalului nu conține reziduurile 26-40 din această structură (aceasta aparținând cozii flexibile N-terminale a proteinei).

Simularea ZDOCK a salvat cele mai bune 2000 de structuri pe baza scorului de andocare, iar aceste structuri au fost procesate în continuare prin minimizarea energiei cu AMBER.

Minimizarea energiei pentru cele trei seturi de 2000 de structuri a fost realizată cu versiunile normale și cuda ale pachetului software GMIN, interconectate la pachetul de software AMBER [17].

Comparam mai jos rezultatele celor patru serii de minimizare pentru fiecare dintre dimere (fiecare pereche într-o formă asociată și disociată a structurilor ZDOCK).

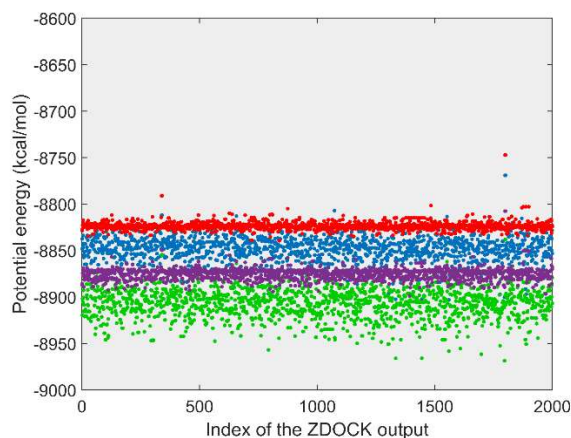


Fig.4-13 Rezultatele minimizării energiei al dimerului BB în formă asociată cu GMIN (albastru), CUDAGMIN (verde) și în formă disociată cu GMIN (roșu), CUDAGMIN (violet)

A treia

și a patra rulare au fost efectuate pe plăci grafice atât în forme asociate, cât și în forme disociate. Putem observa că dimerul disociat are cele mai mari valori energetice din cauza lipsei de interacțiuni între lanțuri, în timp ce în rulările asociate pot apărea ciocniri între catenele dimerului (Fig. 4-13). Energia de legare a fost calculată prin diferența dintre valorile energetice în formele asociate și disociate ale dimerelor.

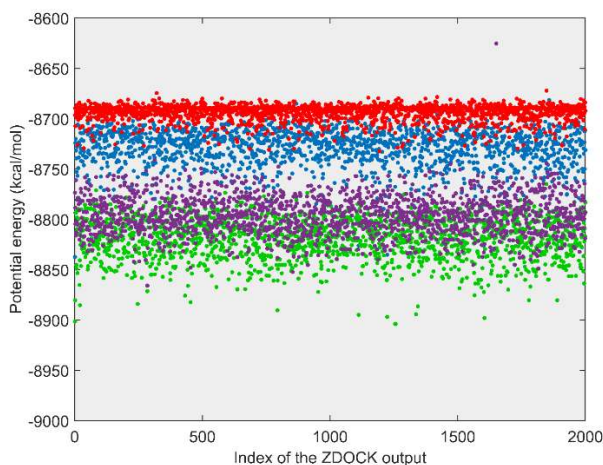


Fig. 4-14 Rezultatele minimizării energiei pentru dimerul BC în formă asociată cu GMIN (albastru), CUDAGMIN (verde) și în formă disociată cu GMIN (roșu), CUDAGMIN (violet)

Minimele energetice ale dimerelor BC (Fig. 4-14) sunt în general mai mari decât cele ale dimerelor BB și CC. Valorile energetice în cazul dimerului BC prezintă, de asemenea, o împrăștiere mai mare, datorită flexibilității mai mari a cozii N-terminale a proteinei. Modificările conformaționale prezintă o varietate mai mare pe o scară mai mare.

Motivul unei variații mai mari a energiilor în cazul configurațiilor asociate este că în forma asociată este posibil un număr mare de interacțiuni suplimentare între cele două lanțuri, care sunt, desigur, inexistente în forma disociată.

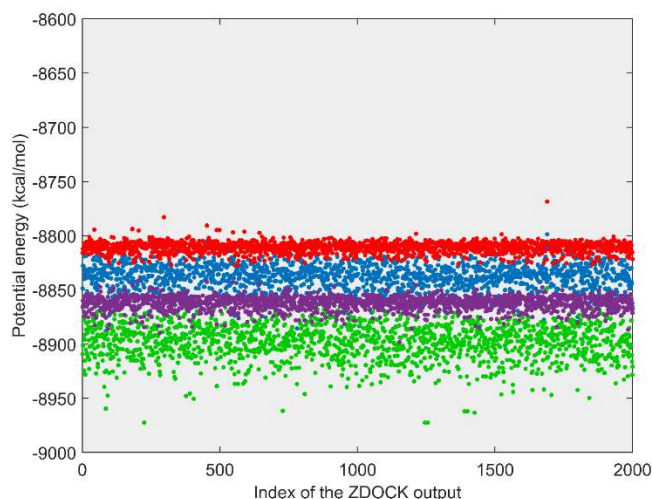


Fig. 4-15 Rezultatele minimizării energiei pentru dimerul CC în formă asociată cu GMIN (albastru), CUDAGMIN (verde) și în disociat pentru GMIN (roșu), CUDAGMIN (violet)

Energiile potențiale calculate cu AMBER, folosind modelul implicit solvent, nu sunt energii potențiale "pure", deoarece modelul implicit de solvent conține schimbări de energie libere datorate solvării/desolvării atomilor de pe suprafață. Prin urmare, modificările de entropie asociate cu legarea/dezbinarea apei de pe suprafață sunt luate implicit în considerare la evaluarea termenului de energie Generalized Born.

Energiile medii calculate pentru dimerile asociate și disociate sunt rezumate în **Table 4-2**.

Table 4-2 Energiile medii ale dimerelor BB, BC și CC

	Energie medie (kca/mol)		Deviația standard	
	Asociat	Disociat	Asociat	Disociat
BB GMIN	-8847.675	-8823.670	44.803	4.593
BB CUDA	-8906.350	-8875.450	45.729	6.021
BC GMIN	-8710.890	-8693.810	186.979	8.566
BC CUDA	-8798.770	-8795.730	200.001	19.208
CC GMIN	-8833.065	-8810.840	102.098	4.845
CC CUDA	-8893.080	-8862.160	102.941	6.240

Energia de legare calculată (**Fig. 4-16**) este reprezentată în funcție de cea mai bună aliniere a celor 6000 de structuri cu dimerul CC (interfața T1). Aici putem observa o topologie asemănătoare pâlniei. Putem observa că pentru mai multe dimere, o valoare scăzută a RMSD corespunde o energie de legare scăzută. Dimerul cu cea mai mică energie este similar cu interfața de tip CC din proteina capsidului ss-CCMV, care are probabil un rol major în primele etape ale asamblării capsidelor.

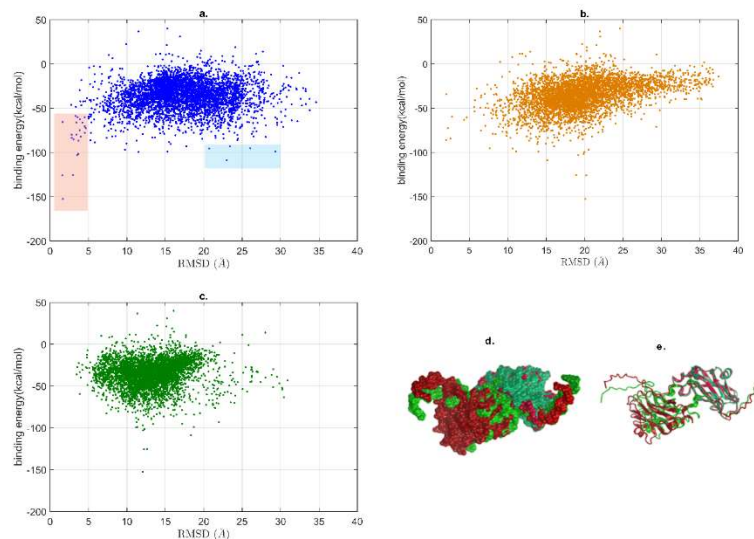


Fig. 4-16 Energii de legare în funcție de RMSD a structurilor din CC-T1 (a.), BC1-T2 (b.), AC- T3 (c.). Structura cu cea mai mică energie de legare suprapusă cu dimerul CC al virusului cu reprezentări de suprafață (d.) și cartoon (e.)

Din cele trei grafice de mai sus putem concluziona că interfața T1 (CC și BA) este cea mai stabilă dintre toate interfețele posibile din capsidul icosaedral. Mai mult decât atât, structura cu cea mai mică energie de legare AMBER este una de interfață T1. Acest rezultat sugerează că optimizarea globală cu potențialul AMBER este utilă pentru predicția a celei mai stabile interfețe dintre proteinele capsidice ale virusului.

În **Fig. 4-16 d.** am reprezentat structura cu cea mai mică energie de legare (roșie) aliniată la dimerul CC al virusului (verde). Alinierea a fost efectuată pe reziduurile 40-190 dintr-un lanț.

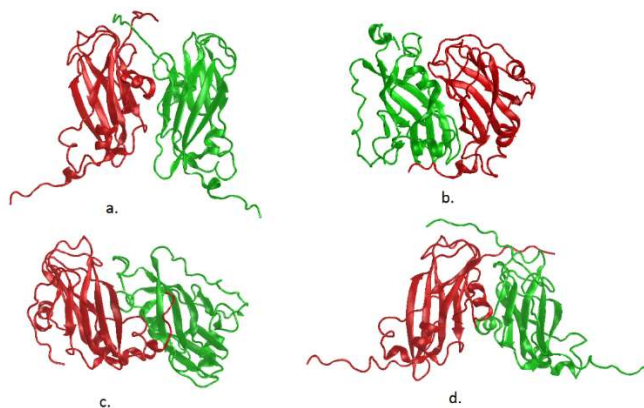


Fig. 4-17 Structuri simetrice: a. dimer BB cu simetrie aproape de C2; b. dimer CC cu simetrie C1 c. dimer CC cu simetrie C2 d. dimer BB cu simetrie C2

Am găsit alte 4 structuri cu energie de legare mică, care nu se aliniază bine cu cele trei tipuri de interfață (Fig. 4-17)

4.3. Compararea simulărilor MD scurtă durată pentru dimerile T1, T2 și T3

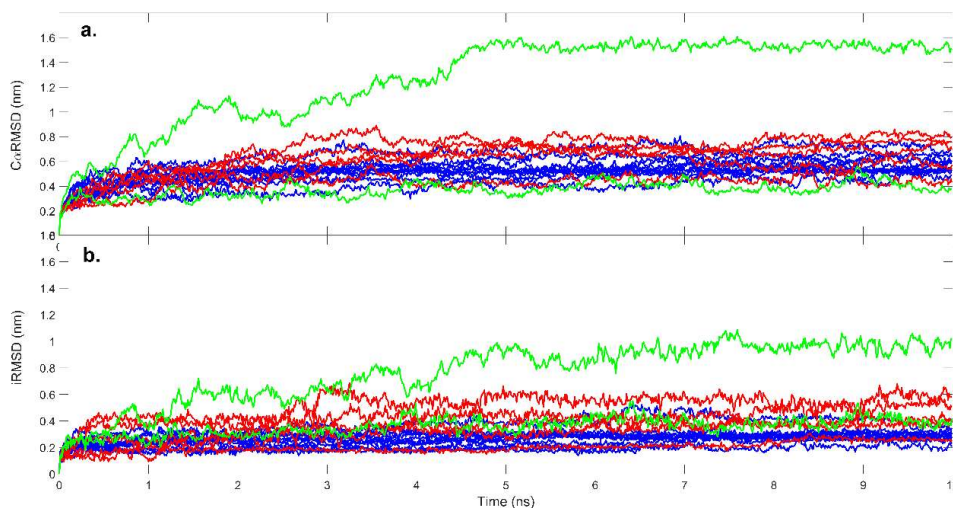


Fig. 4-19 C α RMSD (a) și iRMSD (b) pentru dimerile T1 (albastru), dimerile T2 (roșu) și dimerile T3 (verde)

Am selectat dimere din cele 6000 de structuri ZDock cu un iRMSD mai mic de 0,3 nm la interfața respectivă din capsidul ss-CCMV, ca structuri de pornire pentru simulările noastre MD. Deoarece nu am găsit dimere T3 corespunzătoare criteriilor, am selectat două structuri: dimerul AB original, preluat din capsid, și structura cu cel mai mic iRMSD din rezultatele ZDOCK.

Am rulat simulare MD în solvent implicit timp de 10 ns (Fig. 4-8). Putem observa o separare clară pentru cele trei tipuri de interfețe (Fig. 4-19). Aici este interesant să ne uităm la schimbarea aminoacizilor de pe interfață (regiunile gri). Pentru T1 și T2 nu putem observa o fluctuație semnificativă a reziduurilor de interfață, în timp ce pentru T3 putem observa că cea mai mare mișcare este exact pentru reziduurile de interfață.

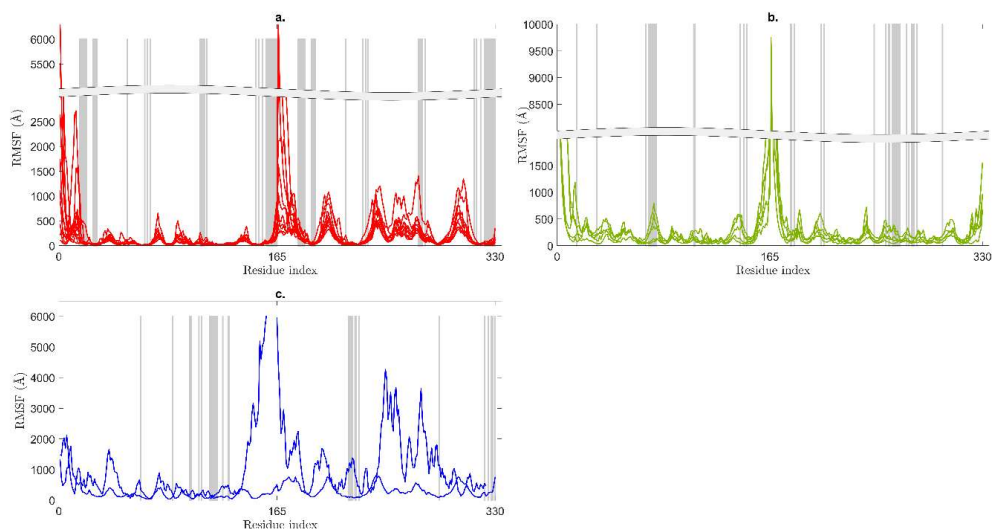


Fig.4-16 RMSF pentru dimerile T1 (a.), T2 (b.) și T3 (c.) cu aminoacizii de interfață evidențiate în gri.

Calcululele MM-GBSA au fost, de asemenea, efectuate pe cele 14 structuri. Rezultatele calcululelor sunt prezentate în **Table 4-3**, împreună cu energiile de legare calculate din rulările de minimizare AMBER și valorile iRMSD. Energia medie de asociere este de $-185,7 \pm 10,4$ kcal/mol pentru T1, $-108,2 \pm 8,9$ kcal/mol pentru T2 și $-76,5 \pm 6,7$ kcal/mol pentru T3.

Rezultatele arată că interfața T2 este cu aproximativ 40% mai puțin stabilă decât interfața T1. IRMSD-ul interfeței T3 a crescut mai mare de 1 nm în timpul rulării 10 ns MD. Acest lucru sugerează că interfața T3 nu joacă un rol în asocierea proteinelor.

Table 4-3 MM-GBSA binding energies

Minimum label	$_{\Delta G_{total}}$ (kcal/mol)	stdev	binding energy	Interface type	iRMS (nm)
BC1	-206.18	10.85	-152.52	1	0.167
BC2	-194.65	10.19	-125.72	1	0.156
BC410	-211.71	10.66	-125.44	1	0.299
BC407	-181.08	9.77	-83.42	1	0.279
BC10	-139.44	11.91	-79.93	1	0.28
BC3	-186.38	10.22	-65.47	1	0.161
BC73	-180.24	11.83	-46.16	1	0.275
BC11	-118.96	8.39	-85.90	2	0.195
BC63	-101.88	8.95	-84.33	2	0.252
BC61	-96.53	8.08	-72.41	2	0.220
BC1273	-111.58	8.59	-66.71	2	0.199
BC187	-112.16	10.15	-59.62	2	0.273
BC1000	-79.79	7.09	-45.40	3	0.716
BB321	-73.18	6.24	-36.13	3	0.341

Structurile cu valori scăzute ale energiei de legare cu iRMSD mică sunt considerate dimere de proteine stabile. După cum se poate observa, printre dimerii T1 trei structuri (BC1, BC2, BC410 – numerotare conform indicelui structurii ZDOCK) sunt sub -100 kcal/mol și valorile iRMSD în raport cu dimerul proteic original sub 0,4 nm. T2 este reprezentat cu două structuri bune (BC11, BC63) cu valori scăzute ale energiei de legare și iRMSD sub 0,3 nm.

Pe baza acestor rezultate au fost selectate trei structuri pentru a efectua simulări de dinamică moleculară și pentru a investiga în detaliu stabilitatea acestora (T1 – BC1, T2 – BC11, T3 – BB321)

4.4. Dimere selectate pentru investigarea dinamicii

Pe baza minimizării energetice am selectat trei structuri dimerice reprezentative cu cel mai mic RMSD la tipurile de interfață (T1, T2, T3). De asemenea, am ales o structură cu cel mai mare RMSD la interfața T3, numită TX, pentru a observa comportamentul dinamic al unui dimer cu similitudine scăzută față de toate interfețele.

4.5. Simulare MD de lungă durată în solvent explicit pentru dimerile de proteine

T1, T2, T3 și TX

Simulările de lungă durată fac posibilă o perspectivă mai profundă cu privire la comportamentul dimerilor. Aici vă prezentăm rezultatele pentru simulările de 2 μ s pornite de la diferitele configurații optimizate (T1, T2, T3 și TX) în apă explicită. Interfața C_{α} RMSD a fost calculată

pentru toate traiectoriile cu referire la prima structură. S-au făcut calcule pentru catena principală a proteinelor și pentru reziduurile de pe interfața de legare (**Fig. 4-17**).

Media valorilor de C_{α} RMSD pentru dimerul T1 este de 0,94 nm, mai mare decât cea pentru T2 (0,83 nm). Structura T1 este mai flexibilă, permițând o mișcare accentuată a reziduurilor. Valorile RMSD pentru T3 sunt cele mai mari, cu o fluctuație mai mare (1,08 nm). Dacă aruncăm o privire la valorile RMSD de interfață (**Fig. 4-23, b.**) putem vedea o separare clară a celor trei tipuri de dimere CCMV. T1 are cea mai mică valoare de medie (0,57 nm).

Valoarea iRMSD crește mai lent în cazul dimerului T2 la începutul simulării, dar are o valoare medie globală mai mare cu 0,3 ns pentru reziduurile de interfață (1,14 nm) decât în cazul T1. De asemenea, valoarea nu este convergentă, prezentând o tendință continuă de creștere. Rezultatul indică faptul că interfața T2 este mai puțin stabilă decât T1, acest lucru fiind în acord cu studiul nostru anterior [6].

Interfața dimerului T3 este mai mică decât cea a T1 și T2, făcând posibile mișcări mai mari ale catenelor. Prin urmare, valorile iRMSD sunt între T1 și T2, cu o medie de 0,87 nm, cu o fluctuație mai mare a valorilor.

Fluctuațiile aminoacizilor au fost calculate pentru a monitoriza mișcarea lor (**Fig. 4-23 c**). Există unele mișcări comune ale părților proteice în toate cele trei proteine. Butoaiele β de 7 unități sunt subunități stabile ale monomerului proteic, după cum putem observa în digrama RMSF. Pe **Fig. 4-23 c.** reziduurile de interfață ale dimerelor sunt reprezentate cu bare de aceeași culoare ca pentru proteina respectivă.

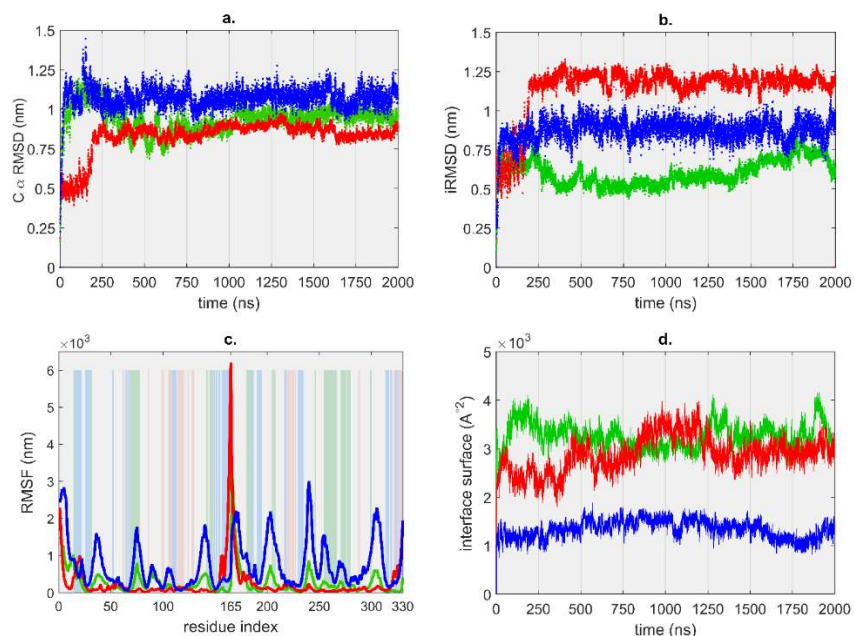


Fig. 4-23 C_{α} RMSD (a.), iRMSD (b.), fluctuația reziduurilor (c.) și suprafața interfeței (d.) pentru T1 (verde), T2 (roșu) și T3 (albastru)

Am calculat suprafața interfeței ca diferența dintre SASA și suma SASA a monomerilor. Suprafața interfeței T3 este cea mai mică și rămâne relativ constantă în timpul simulării cu deplasări temporare minore ($1339.2 \pm 182.2 \text{ \AA}^2$). Suprafața interfeței T1 crește de la 14% din întreaga suprafață la 19%. T2 prezintă o creștere majoră a suprafeței interfeței după 1000 ns de simulări, în timp ce SASA este în scădere. Fluctuațiile mari ale valorilor arată că PPI-ul fiecărui dimer se modifică permanent.

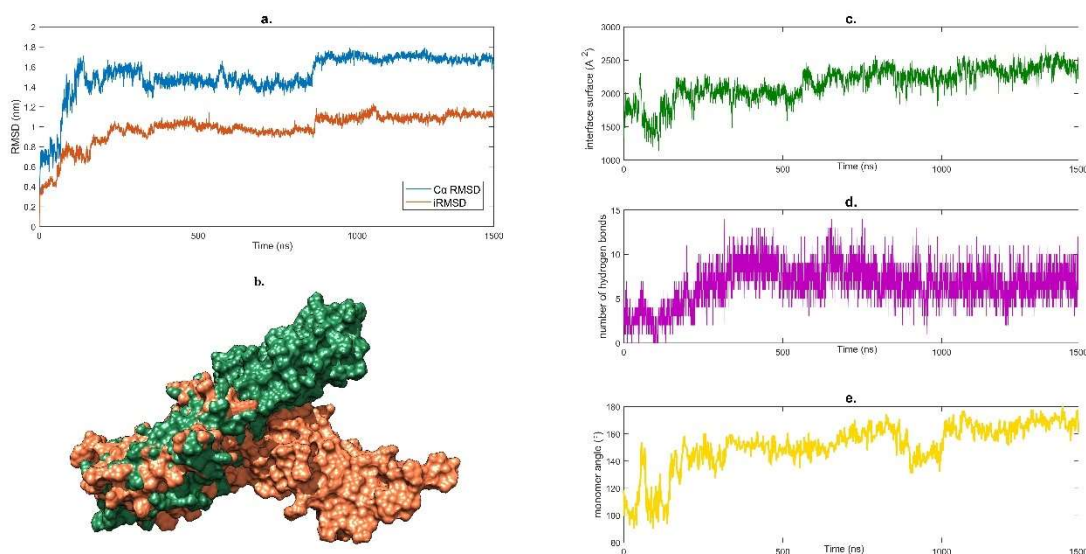


Fig.4-26 Analiza NPT de 1,5 μs, simularea explicită a solvenților pentru interfața TX: a. C_{α} RMSD și iRMSD; b. structurile inițiale (roșu) și finale (verde) ale TX suprapuse, cu reprezentarea suprafeței; c. variația suprafeței interfeței; d) numărul de legături de hidrogen; e. variația unghiului dintre monomerii proteinei.

Simularea de 1,5 μs la presiune constantă, în solvent explicit a fost efectuată pe TX. Investigarea dinamicii unui dimer cu structură diferită față de cele trei tipuri de dimeri poate duce la informații suplimentare despre comportamentul proteinelor în condiții de contacte nefavorabile.

În timpul simulării au avut loc modificări drastice în structura dimerului TX. Inițial, lanțurile proteinei au fost atașate între ele prin coada N-terminală într-o formă asemănătoare V-ului (**Fig. 4-26 b**). În prima fază a simulării, dimerul formează o structură asemănătoare tije. Datorită acestei mișcări, valorile RMSD sunt ridicate, cu o medie de $1,51 \pm 0,2$ nm pentru C_{α} RMSD și $0,98 \pm 0,15$ nm pentru iRMSD (**Fig. 4-26 a**). Suprafața PPI este în creștere pe tot parcursul simulării, sugerând o stabilitate tot mai mare a dimerului. În timpul simulării, numărul legăturilor de hidrogen este, de asemenea, în creștere, cu fluctuații mari.

4.6. Simulări de dinamică moleculară la temperaturi diferite

Simulări MD au fost efectuate pe temperaturi mai ridicate pentru dimerul T1, T2, T3 și TX pentru a studia condițiile de disociere a dimerelor. Scopul principal a fost de a realiza o disociere a dimerilor pentru o mai bună înțelegere a mecanismului când un capsid devine destabilizat.

Patru simulări paralele cu viteze inițiale diferite au fost efectuate în solvent implicit pentru T1, T2, T3 și TX la temperatura de 350 K timp de 200 ns. Cele patru serii ale dimerului T1 prezintă o asemănare între ele, cu o medie de 1,1 nm C_{α} (Fig. 4-19 c.). Modificările interfeței sunt neglijabile pentru T1 la temperaturi mai ridicate. Fluctuațiile valorilor se datorează mișcărilor relative ale catenelor, dar nu se poate observa o variație conformațională mai mare.

Dimerul T2 se comportă diferit față de T1. În trei simulări din patru dimerul se apropie de o stare de disociere; cu toate acestea, cozile N-terminale ale dimerelor rămân atașate și împiedică separarea

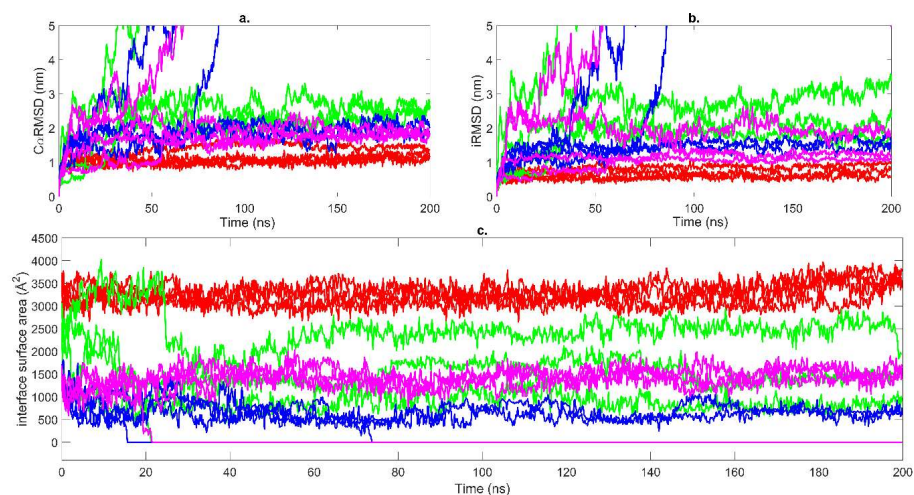


Fig. 4-27 C_{α} RMSD (a), iRMSD (b) și suprafața interfeței (c) pentru T1 (roșu), T2 (verde), T3 (albastru) și TX (violet) pe 350 K, solvent implicit

completă a lanțurilor. Se observă, de asemenea, o pierdere accentuată în suprafața interfeței. În cea de-a patra simulare, disocierea dimerului este completă cu o pierdere treptată a interacțiunilor dintre cozile N-terminale ale monomerilor.

Structura T3 se disociază, de asemenea, în două simulări, în timp ce în celelalte ajunge la o stare aproape de disociere. Interfața este mult mai mică decât pentru celelalte două tipuri, prin urmare mișcarea reziduurilor de pe interfață este accentuată.

Au fost create binding energy landscapes (Fig. 4-28) pentru simulările de 200 ns să observăm modificările numărului de contacte pe PPI, distanța dintre lanțurile de proteine și corelația cu energiile de legare.

Structurile T1 au valori energetice de legare scăzute de-a lungul traiectoriilor, nu pot fi detectate modificări drastice în structura proteinei. La început dimerul T2 are o energie de legare de -103 kJ/mol, care crește în timp cu pierderea contactelor. Energia de legare scade exponențial la 0 pe măsură ce monomerii se îndepărtează unul de celălalt la o pierdere totală a contactelor, ajungând

la o stare disociată. T3 și TX au energii de legare inițiale mai mici, cu mai puține contacte. Graficele sunt similare cu cele pentru T2 pe măsură ce ajung la disociere.

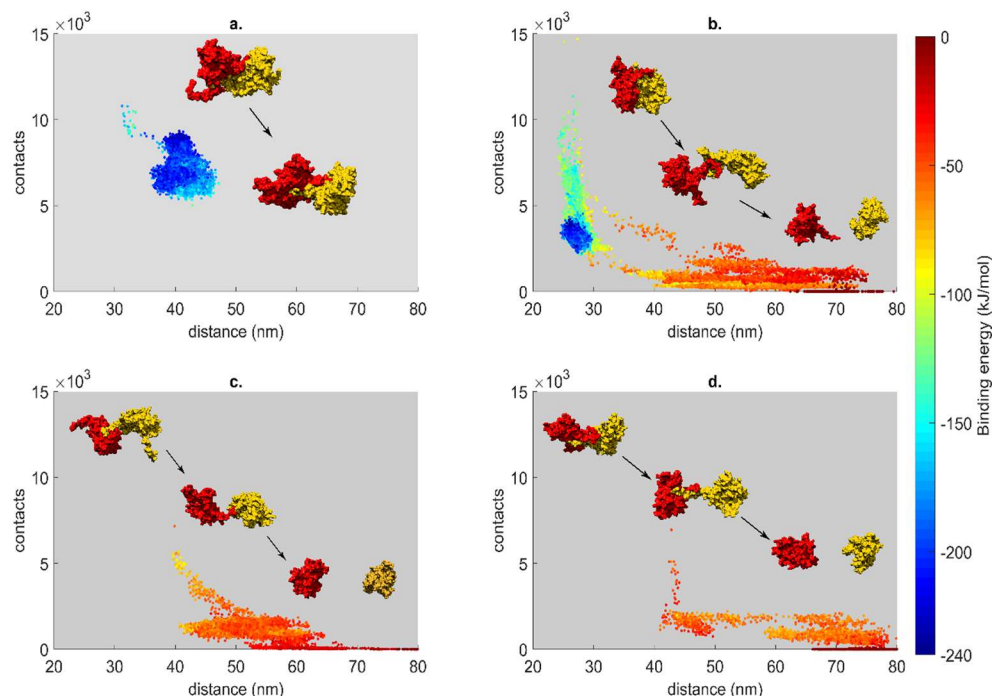


Fig. 4-28 Binding energy landscape în funcție de contactele dintre monomeri și distanțele centrului de masă cu structurile inițiale și finale prezentate, pentru simulările MD în solvent implicit pentru T1 (a), T2 (b), T3 (c) și TX (d) pe 350K

După ce am constatat că, în anumite condiții, se întâmplă disocierea dimerelor, am efectuat simulări MD în solvent explicit la 350 K, presiune constantă timp de 600 ns cu cele patru dimere. Tendințele valorilor C_a (Fig. 4-29 a) pentru T1 ($0,99 \pm 0,15$ nm) și T2 ($0,78 \pm 0,12$ nm) sunt similare cu simulările de 300 K cu fluctuații mai mari. În cazul T3 ($1,57 \pm 0,33$ nm) și TX ($1,52 \pm 0,42$ nm) observăm o creștere majoră a valorilor care poate fi explicată odată cu modificările structurale ale dimerelor. Aminoacizii de pe interfața dintre monomeri prezintă un comportament diferit (Fig. 29 b).

Valorile PCA au fost legate de fiecare atom și vizualizate cu plugin-ul NMWiz în VMD [31]. Mișcările relative ale atomilor sunt afișate ca săgeți proporționale cu valorile lor (Fig. 4-29 d-g).

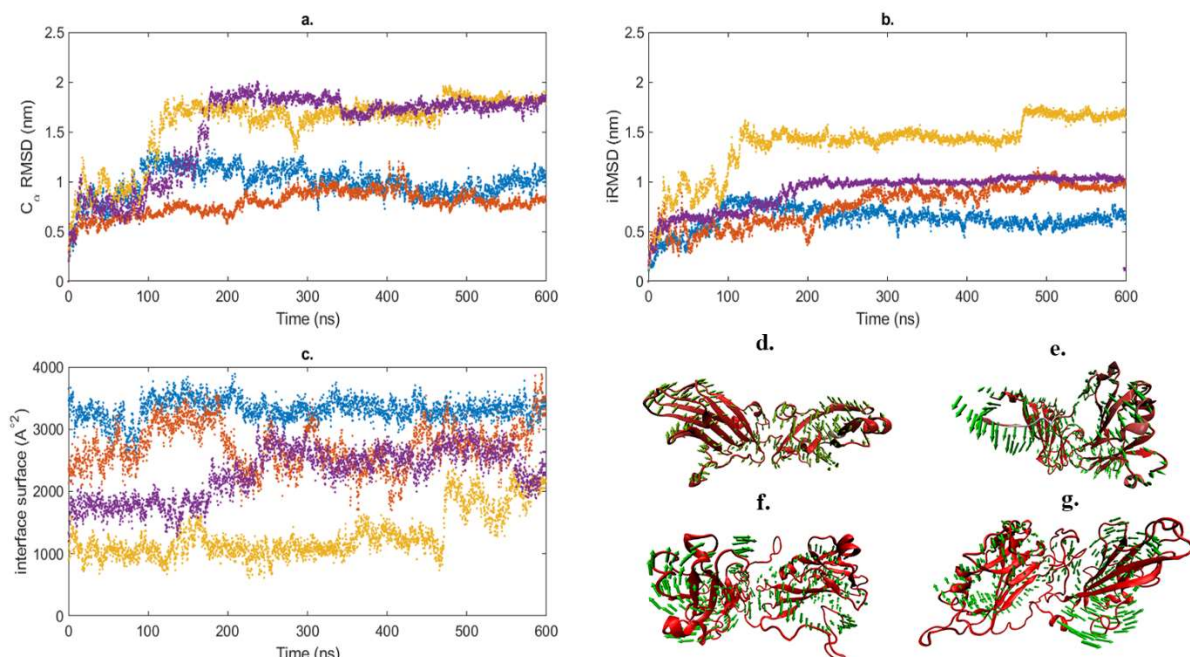


Fig. 4-29 C_{α} RMSD (a), i RMSD(b) și suprafața interfeței (c) pentru T1 (albastru), T2 (roșu), T3 (galben) și TX (violet) în solvent explicit, ansamblu NPT timp de 600 ns. Analiza modului normal pentru T1 (d), T2 (e), T3 (f) și TX (g).

Fluctuația reziduurilor la 350 K în T1 arată o mișcare relativ echilibrată, apar valori mai mari pentru structurile iregulare. Orientarea săgeților sugerează o mișcare de închidere a lanțurilor.

Mișcarea accentuată a reziduurilor N-terminale oferă cea mai mare parte a fluctuației globale a proteinei T2 (Fig. 4-29 e). O mișcare colectivă de răsucire a reziduurilor din primul lanț poate fi observată pentru dimerul T3 (Fig. 4-29 f.), în timp ce N-terminusul celui de-al doilea lanț se îndepărtează de corpul proteinei. Mișcări similare sunt observate în cazul TX (Fig. 4-29 g), totuși cu un impact mai mic asupra interfeței dintre monomeri.

În concluzie, simulările explicite ale solvenților la temperaturi mai ridicate tind să destabilizeze structura terțiară a dimerilor, în timp ce cele din solventul implicit pot destabiliza interfața proteină-proteină.

4.7. Mutanții dimerului T1

4.7.1. Mutanții de deleție a dimerelor T1 și T2

În timpul simulărilor MD la temperaturi ridicate, dimerul T2 a ajuns la o stare aproape de disociere, cu toate acestea, cozile N-terminale ale dimerelor au rămas atașate și au împiedicat separarea completă a lanțurilor. Pentru a depăși acest efect, am eliminat 10 reziduuri din terminalele N ale ambelor lanțuri în dimerul T1 și T2 și am rulat simulări ulterioare pe aceste proteine mutante.

Simularea MD, NPT a fost efectuată pentru mutații de deleție T1 (T1DM) și T2 (T2DM) în solvent explicit cu parametri similari cu cei de la 4.3. În timpul simulării în solvent explicit de 200 ns pe 350 K, ambele dimere au rămas atașate. Mișcarea reziduurilor a fost accentuată, datorită temperaturii crescute. Valorile C_{α} (Fig. 4-31 a), în timp ce pentru proteina inițială în aceleași condiții acestea au fost în jur de $0,99 \pm 0,15$ nm.

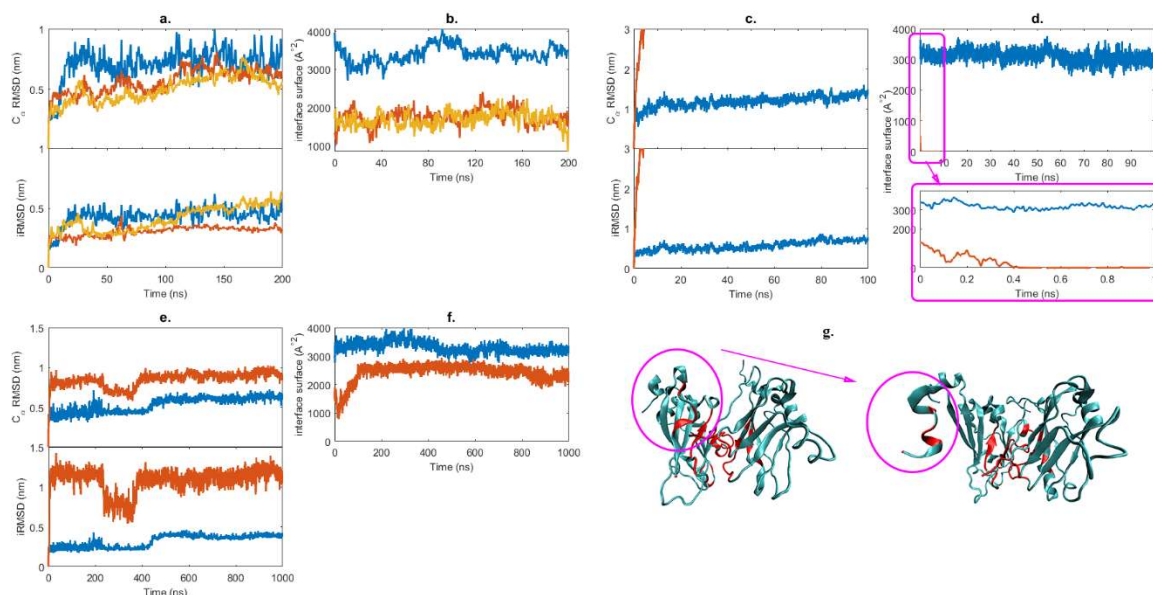


Fig. 4-31 C_{α} RMSD, iRMSD (a) și suprafața interfeței (b) pentru 200 ns NPT, simulare în solvent explicit pe 350K pentru T1DM (albastru), T2DM (roșu), respectiv T2DM pe 400K (galben)

C_{α} RMSD, iRMSD (c) și suprafața interfeței (d) pentru simularea în solvent implicit de 100 ns pe 350K pentru T1DM (albastru) și T2DM (roșu)

C_{α} RMSD, iRMSD (e) și suprafața interfeței (f) pentru simularea în solvent implicit de 1 μ s pe 300K pentru T1DM (albastru) și T2DM (roșu)

Diferențe similare pot fi observate și pentru T2DM cu o medie de $0,61 \pm 0,11$ nm C_{α} RMSD pe 350K și $0,60 \pm 0,10$ nm pe 400K (datele de 400K neprezentate); cu toate acestea, nu s-a observat o convergență. Reziduurile de interfață se comportă în același mod atât la 350K, cât și la 400K. În acest interval de timp nu se poate observa o diferență mare în mișcarea dimerelor.

T1DM are o mișcare mai mare în solvent implicit pe 350K, dar rămâne relativ stabil în timpul simulării ($1,18 \pm 0,15$ nm C_{α} RMSD și $0,57 \pm 0,11$ nm iRMSD). T2DM ajunge la disociere deja în 1 ns de simulare (Fig. 4-31 c). Lipsa terminalului N ar putea provoca disocierea rapidă, deoarece în cazul dimerului T2 de tip sălbatic disocierea a avut loc după 20 ns.

Am efectuat simulare implicită timp de 1 μ s pe 300 K pentru a observa stabilitatea generală a mutațiilor de deleție (Fig. 4-31. e-f). Valorile mai mici pentru T1DM pe 300 K pot indica faptul că reziduurile terminale sunt importante pentru păstrarea compactității dimerului. În cazul T2DM putem detecta un comportament anormal, deoarece C_{α} RMSD este mai mic decât iRMSD.

Aceste mutații sunt intervenții forțate în structura proteinelor, dar rezultatele demonstrează că reziduurile terminale participă la stabilizarea dimerilor, în special pentru dimerul T1.

4.7.2. Mutații punctuale ale dimerului T1

Mutații punctuale pe interfața dimerului T1 au fost generate prin înlocuirea reziduurilor originale, atât în lanțul B, cât și în cel C, pe baza proprietăților după cum urmează:

- Polar la apolar
- apolar la polar
- Acidic la bazic
- bazic la acidic
- Aromatice la polare

Cele 178 de structuri mutante rezultate au fost minimizate cu CUDAGMIN fără cutoff la interacțiunile nonbonded și cu 10^{-4} RMS, criteriu de convergență a forței cu $igb=2$, $saltcon=0.1$. Mai multe structuri mutante au o energie de legare scăzută, dintre care 9 sunt peste -100 kcal/mol. Rezultatele arată că mutațiile punctuale au potențialul de a afecta în mod semnificativ dimerizarea. Pe baza energiei de legare calculate am ales cinci mutații cu energie mare și doi mutații pe baza profilului de sensibilitate la mutație generat, cu cel mai mare $\Delta\Delta G$ și unul cu cel mai mic $\Delta\Delta G$.

Table 4-4 Lista mutațiilor selectate

Mutant	Reziduu mutat	Mutație	Energie de legare (kcal/mol)	$\Delta\Delta G_{pred}$ (kcal/mol)
G16T	Glycine	Threonine	-68.5482	0.590324
S26A	Serine	Alanine	-92.6142	-0.40726
G64S	Glycine	Serine	-98.7593	0.64485
A111S	Alanine	Serine	-95.5852	0.241031
S160P	Serine	Proline	-93.2639	-0.13606
E151C	Glutamic acid	Cysteine		-2.77259
R65D	Arginine	Aspartic acid		2.330864

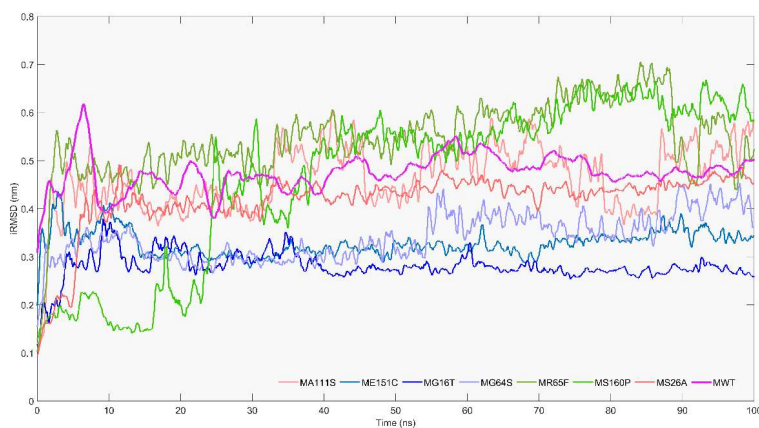


Fig. 4-34 Media mobilă a iRMSD pentru tipul sălbatic (violet) și mutații pentru simulările 100 ns MD pe 300 K

Simulările MD au fost efectuate pentru 100 ns pe 300 K, în scopul de a vedea schimbările într-un interval de timp scurt cauzate de mutațiile introduse.

Deoarece toate mutațiile au fost făcute pe PPI, am analizat comportamentul interfeței. iRMSD-ul de tip sălbatic a fost comparat cu cei 7 mutații (Fig. 4-34). S26A și A111S prezintă un comportament similar cu tipul sălbatic. Diferențe majore sunt observate în mișcările interfeței în ambele direcții. G64S, G16T și E151C au valori mai mici cu aproximativ 0,2 nm, în timp ce R65D și S160P au valori mai mari. Scăderea valorilor iRMSD sugerează o stabilizare a dimerului, cu toate acestea am dobândit o perspectivă suplimentară asupra modificărilor cauzate de mutații printr-o analiză mai profundă.

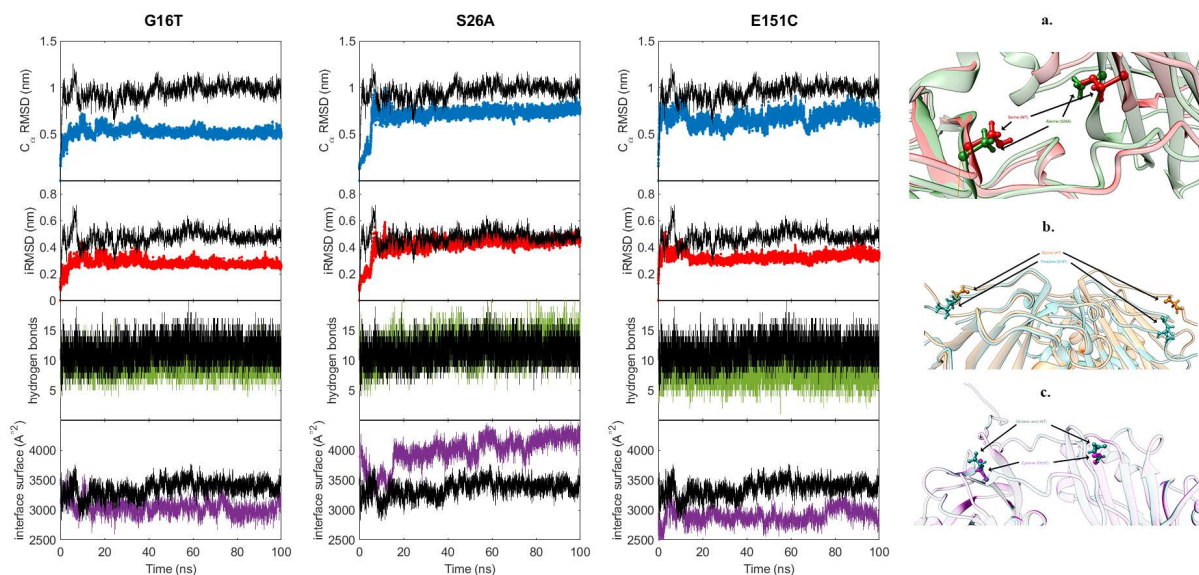


Fig. 4-35 C_{α} RMSD (albastru), iRMSD (roșu), numărul de legături de hidrogen (verde) și suprafața de interfață (violet) a G16T, S26A și E151C comparat cu WT (negru). Aminoacizii mutate la G16T (a.), S26A (b.) și E151C (c.)

Am selectat trei mutații de dimeri și am analizat simulările 100 ns, implicit solvent MD (Fig. 4-35).

În G16T, o glicină a fost înlocuită cu treonină la reziduurile 16 și 181, părți din regiunile N-terminale ale monomerilor (Fig. 4-35 a). Atât C_{α} RMSD ($0,52 \pm 0,05$ nm), cât și iRMSD ($0,28 \pm 0,03$ nm) au scăzut comparativ cu WT (C_{α} RMSD = $0,92 \pm 0,09$ nm, iRMSD = $0,47 \pm 0,05$ nm).

În al doilea mutant, serina situată pe poziția a 26-a este mutată la alanină (S26A). Schimbări semnificative pot fi observate în valorile C_{α} RMSD și suprafața interfeței. Fluctuațiile reziduurilor de interfață sunt similare cu cele ale WT.

Mutația E151C stabilizează proteina, pe baza predicției MAESTRO. Rezultatele simulării arată un comportament similar. C_{α} RMSD și iRMSD este mai mic decât în cazul WT cu fluctuații mai mari. Din datele noastre colectate până în prezent, putem concluziona că sunt necesare calcule suplimentare pentru a estima consecința mutațiilor punct pe dimer de proteine CCMV de tip sălbatic.

4.8. Simulări de dinamică moleculară cu schimb de replici

Simulările REMD cu solvent implicit au fost efectuate pe T1, F161P și T2DM pentru a depăși barierele energiei libere, făcând posibilă o imagine mai precisă a modificărilor în timpul simulărilor. Fiecare simulare a fost efectuată pentru 25 ns, 200 ns în total, pentru fiecare structură. Temperaturile utilizate pentru simulări au fost de 300,00 K, 306,65 K, 313,42 K, 320,28 K, 327,25 K, 334,34 K, 341,53 K, 348,81 K, 356,24 K. Replicile cu cea mai scăzută și cea mai mare distribuție a temperaturii au fost selectate din cele 8 serii de producție REMD.

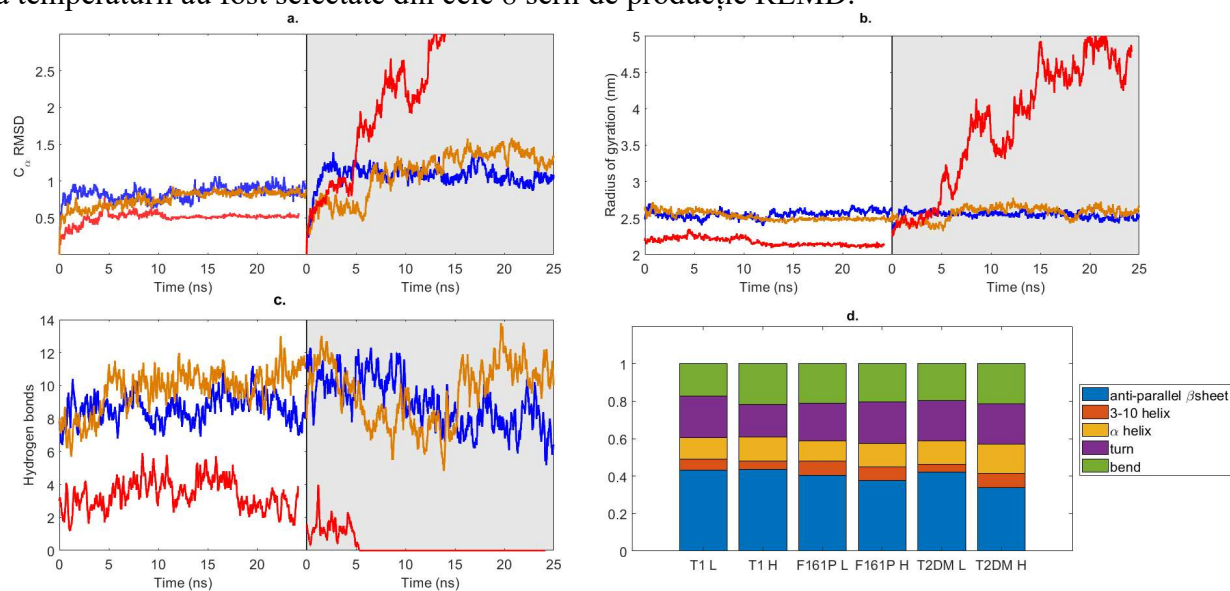


Fig. 4-36 Variația C_{α} RMSD (a.), RoG (b.) și legăturilor hidrogen (c.) în timpul simulării REMD de 25 ns pentru sistemul de distribuție la temperaturi scăzute (fundal alb) și sistemul de distribuție la temperaturi ridicate (fundal gri) de T1 (albastru), F161P(galben) și T2DM(roșu); Structura secundară medie a sistemelor investigate (d.)

Dimerul T1 este stabil în ambele cazuri, cu o fluctuație minoră a RoG (Fig. 4-36 b.). Se observă o schimbare majoră în cazul mutantului de deleție. La o temperatură mai scăzută, RoG este mai mic decât în cazul T1 și F161P, din cauza lipsei reziduurilor N-terminale. La temperaturi mai ridicate, creșterea mare a RoG pentru mutantul de deleție T2 se datorează disocierii. O ușoară schimbare poate fi observată în cazul F161P. La temperaturi mai scăzute, stabilitatea dimerului crește odată cu scăderea RoG, în timp ce la temperaturi mai ridicate se poate observa o creștere lentă.

4.9. Dinamica moleculară accelerată pe T1, T2, T3 timp de 500 ns în ansamblul NPT

Dinamica moleculară accelerată (aMD) este o versiune îmbunătățită a MD-ului clasic, unde potențialul este modificat pentru a reduce înălțimea barierelor energetice locale, prin urmare simularea poate ajunge la rezultate similare într-un timp mai scurt.

aMD a fost efectuat pe dimerale T1, T2 și T3 pentru 500 ns pe 300 K în solvent explicit. Rezultatele sunt aproximativ similare cu simulările cMD de 2 μ s. Media C_{α} RMSD a T1 a fost de 0,92 (\pm 0,15) nm. Putem observa o creștere atât a valorilor C_{α} RMSD, cât și a iRMSD după 400 ns, ajungând la valorile simulării cMD. Nu putem vedea această schimbare pe suprafața interfeței; astfel, schimbarea nu are loc la suprafața dintre monomeri. Schimbarea este cauzată de mișcarea de deschidere a monomerilor, unghiul dintre monomeri fiind în creștere.

Valorile mai mici ale RMSD pentru dimerul T2 se datorează atașării cozilor terminale la corpul proteinei în prima fază a simulării. Prin urmare, mișcarea mare a regiunilor cozii este împiedicată. T3 se comportă ca în simularea cMD, cu modificări conformaționale mai mari la începutul simulării și fluctuația accentuată a suprafeței de interfață, cauzate de structura flexibilă a dimerului.

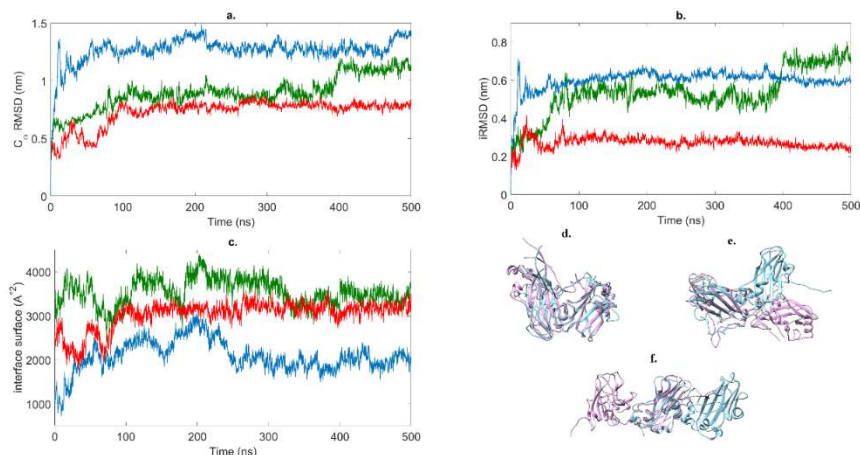


Fig. 4-38 C_{α} RMSD (a.), iRMSD (b.) și suprafața interfeței (c.) pentru simulări aMD pentru T1 (verde), T2 (roșu) și T3(albastru). Structurile inițiale și finale ale simulărilor de 500 ns suprapuse sunt reprezentate pentru T1 (d.), T2(e.) și T3 (f.)

După calculele PCA, primele două eigenvectoare au fost proiectate pentru a genera un peisaj de energie liberă (Fig. 4-38), unde energia liberă inspectează direcția fluctuației. Culoarea albastră închisă indică stări conformaționale stabile ale moleculelor. Sunt prezentate, de asemenea, structuri proteice reprezentative ale bazinelor. Analiza detaliată a simulărilor cMD este discutată în capitolul 4.5..

Dimerul T1 are un singur bazin cu o configurație stabilă (structura 1), deoarece simularea a convergent în prima fază. În timpul a 500 ns de aMD două minime de energie pot fi observate (structura 2 și 3) în alte locații, decât la cMD. Schimbarea se datorează mișcării de îndoire discutate mai sus. Cu toate acestea, proteina a explorat un spațiu conformațional mai mare în timpul simulării aMD.

Expresia proteinei din capsida ssCCMV și simulare de dinamică moleculară a capsidului ssCCMV

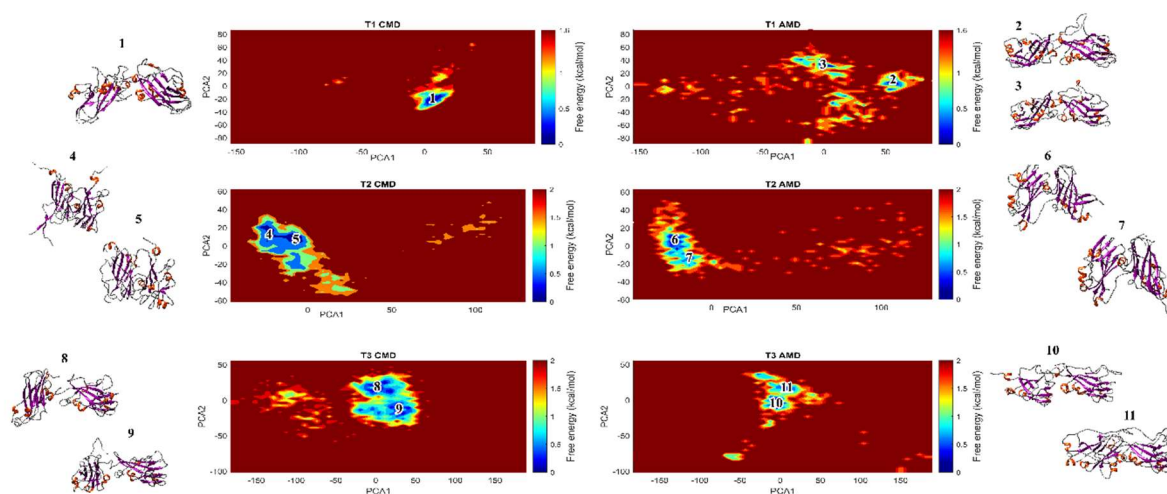


Fig. 4-39 Compararea peisajelor energiei libere pentru simulările cmd (prima coloană) și amd (a doua coloană) cu structurile reprezentative

Dimerul T2 are minimele în aceleași locații, structurile proteice corespunzătoare sunt, de asemenea, foarte asemănătoare (structurile 4 până la 7). Se poate observa un comportament diferit în cazul dimerului T3. Proteina are o mobilitate mai mare în timpul simulării convenționale. Bazinele se găsesc în aceleași locuri cu bariere scăzute între ele, astfel încât tranzițiile sunt susceptibile să apară cu ușurință. Dimerul T3 este relativ instabil, cu mai puține interacțiuni între monomeri, rezultând o flexibilitate ridicată a proteinei.

4.10. Pentamer ale dimerelor

Un pentamer ale dimerelor (PD) a fost selectat din capsida de ss-CCMV (PDB ID 1za7). Partea interioară a PD este formată din cinci catene C conectate cu catene exterioare prin interfețe T1, în timp ce între ele sunt prezente interfețele T2 (Fig. 4-40 .c).

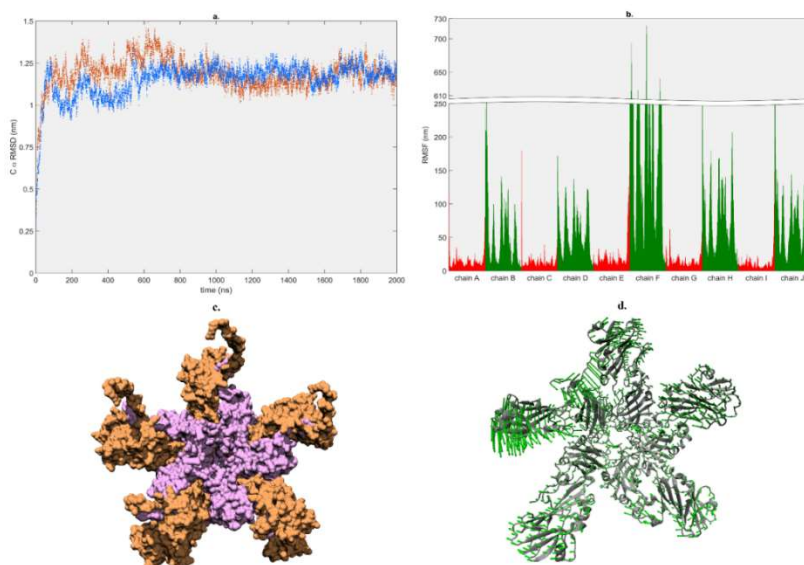


Fig.4-40 C_{α} RMSD pentru cele 2 simulări paralele ale PD; b. Fluctuația reziduurilor din PD: catene interioare - roșu, catene exterioare - verde; c. Structura PD cu reprezentare a suprafeței; d. Mișcarea reziduurilor în timpul simulării pe baza calculului PCA

Simulările de lungă durată au fost efectuate în solvent explicit pentru PD și analizate pentru o mai bună vizualizare a modificărilor conformaționale ale capsidului CCMV. Au fost efectuate două simulări ale dinamicii moleculare NTP pentru 2 microsecunde în apă explicită cu viteze inițiale diferite.

Pentru o mai bună înțelegere a modificărilor în timpul simulării am separat dimerii după tipurile interfețelor, prin urmare am obținut 5 dimere de interfață T1 și 5 dimere de interfață T2

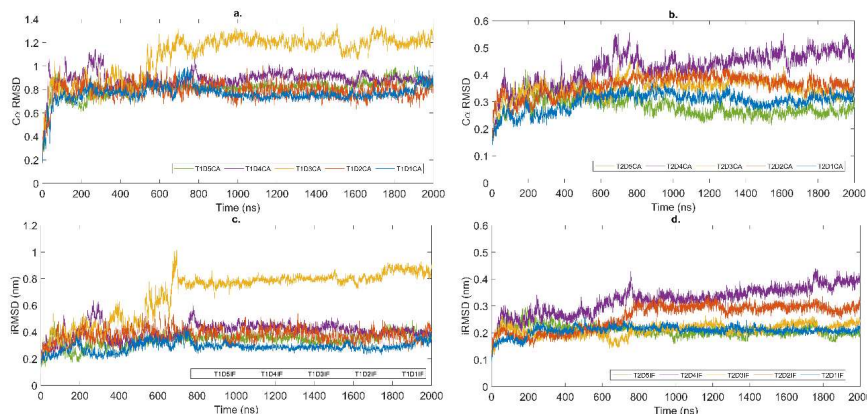


Fig.4-41 C_{α} RMSD pentru dimerii T1 (a) și T2 (b). iRMSD pentru dimerii T1 (c) și T2 (d)

Printre dimerii T1, al treilea (T1D3) arată un comportament diferit de celelalte. După simularea 500 ns se poate observa o creștere de 0,4 nm atât pentru C_{α} RMSD, cât și pentru iRMSD. Monomerul exterior se deplasează în sus din planul PD. Schimbarea este similară cu cea văzută cu un dimer T1 independent. Regiunile N-terminale ale monomerilor interioare sunt aranjate într-o conformație specifică, formând legături de hidrogen cu regiunile terminale a două lanțuri învecinate pe partea dreaptă. Aranjamentul este conservat în timpul simulării, astfel se subliniază rolul regiunii N-terminal în formarea și stabilizarea pentamerului.

Luând în considerare rezultatele, putem afirma că dimerul T1 D3 de la decamer se comportă similar cu dimerul T1 independent.

S-a calculat schimbarea numărului de legături de hidrogen de pe interfețe în timpul simulării. O mică scădere a numărului de legături de hidrogen poate fi observată pentru cele două dimere (T1 D1, T1 D3). Celelalte dimere T1 au avut valori similare cu cele ale dimerilor independente.

T2 D3 și T2 D4 se comportă similar cu dimerul independent. Deși mișcarea dimerilor T2 în decamer este împiedicată, comportamentul analog cu dimerul independent poate duce la concluzia că interfața dintre monomeri determină considerabil comportamentul decamerului.

5. Concluziile

5.1. Rezultate și contribuții originale

În partea experimentală a lucrării, am studiat expresia proteinei capsidice CCMV în celula gazdă *E. coli*.

Plasmida care conține ADN-ul proteinei a fost achiziționată din resurse comerciale și transformată în celule Rosetta *E. coli* BL21(DE3). Anterior, ADN-ul proteinei ss-CCMV a fost clonat într-un vector care conținea un his-tag și un partener de fuziune ubiquitin.

Clonarea plasmidei și exprimarea pUBK_CCMV a avut succes, confirmată prin secvențiere. Expresia a fost optimizată cu o serie de experimente cu schimbarea temperaturii expresiei și a concentrației agentului de inducere. Proteina a fost purificată cu cromatografie de afinitate nichel și ubiquitina a fost îndepărtat cu enzima YUH1. Sistemul de expresie prezentat poate fi folosit pentru expresia proteinei capsidice ss-CCMV cu un randament ridicat.

În timpul studiilor am studiat anumite aspecte ale autoasamblării capsidice virusului ss-CCMV, prin modelarea și investigarea interacțiunilor dintre catenele proteice care la rândul lor formează capsomeri, constituenți ai capsidului viral. Am examinat corelația dintre interfețele cunoscute și dimerul proteic modelat și am constatat potriviri structurale pentru toate cele trei tipuri de interfețe. Pe baza energiei de legare și a calculului de RMSD, putem prezice că dimerul CC (T1) sunt cele mai importante în formarea capsidului. Această concluzie este confirmată de calea de asamblare inițială, care se desfășoară prin dimerizare și adăugarea cooperativă a dimerelor. Predicțiile făcute de PISA webserver arată, de asemenea, că această interfață este cea mai importantă pentru formarea unității biologice. De asemenea, este important de notat faptul că este posibil construirea întregului capsid prin adăugarea ulterioară a dimerilor de T1 la un pentamer de dimere. Pentamerul dimerelor este nucleul pentru asamblarea CCMV determinat experimental [33].

Cele mai bune structuri generate de ZDOCK, organizate în ordinea scorului de andocare, corespund destul de bine cu datele noastre simulate. Pentru un anumit dimer (BC), primele structuri din rezultatele ZDOCK se numără printre dimerul cu cele mai mici energii de legare calculate cu AMBER. Acest lucru ne conduce la concluzia că câmpul de forță AMBER poate fi utilizat pentru modelarea interfeței proteină-proteină cu o bună capacitate predictivă. Cu toate acestea, selecția configurațiilor inițiale este o problemă importantă. De asemenea, trebuie luată în considerare posibila flexibilitate conformațională a proteinei; prin urmare, este recomandat inițializarea calculului ZDOCK de la conformații proteice ușor diferite. Orice predicție este la fel de bună ca eșantionarea de bază, prin urmare este foarte important să pornim de la cât mai multe configurații de diferite dimere pentru a efectua andocarea oarbă.

După studiile de andocare, am investigat comportamentul celor mai bune structuri dimerice prin efectuarea de simulări MD în diferite circumstanțe.

Studierea în detaliu a interacțiunilor de pe interfețele de legare a dimerelor proteice poate ajuta la înțelegerea procesului general de formare a capsidului. Simularea formării complexului proteic este dificilă și durează foarte mult, de aceea am ales să analizăm procesul invers, disocierea. Temperatura este un factor care influențează stabilitatea complexelor proteice. Am efectuat calcule la diferite temperaturi și intervale de timp.

Simulările MD de scurtă durată efectuate pe structurile cu cea mai bună aliniere la interfețele originale au condus la selectarea a trei dimere pentru investigații suplimentare.

În timpul simulărilor scurte pentru dimerul T1 din capsida ss-CCMV la temperaturi ridicate, proteina a început să se deplieze și nu am observat disocierea în acest interval de timp. Am constatat stabilitatea structurilor secundare, structurile β -butoi par a fi mai stabile decât α -helix.

S-au efectuat simulări de 2 μ s MD clasic în solvent explicit și s-a dovedit stabilitatea crescută a T1 în comparație cu dimerul T2. Deși ultima structură nu s-a disociat, interfața s-a schimbat continuu pe tot parcursul simulării. T3 s-a dovedit a fi cel mai puțin stabil dintre diferitele tipuri de interfețe. Dimerele au fost simulate la temperaturi diferite (300 și 350 K). În cazul configurației T1 s-au observat doar mici modificări cauzate de temperatura ridicată, a început un proces lent de depliere. În schimb, T2, așa cum era de așteptat pe baza stabilității scăzută, a ajuns la o stare aproape disociată pe 350 K, doar cozile N-terminale ale lanțurilor monomerice au împiedicat separarea. Pentru a reduce acest efect, am eliminat zece reziduuri N-terminale din ambele lanțuri, iar simularea repetată a dus la o disociere completă a lanțurilor proteice.

Simulările de schimb de replici efectuate în solvent implicit la un interval de temperatură de 300 până la 356 K susțin rezultatele noastre anterioare rezumate mai jos.

Dimerul T1 este stabil în cazul simulărilor lungi și la temperaturi mai ridicate, de asemenea, este de preferat deplierea în loc de disociere la creșterea temperaturii, arătând stabilitatea extrem de ridicată a interfeței T1. T2 este mai puțin stabil și își pierde interfața originală la temperaturi mai ridicate, dar disocierea totală este împiedicată de coada N-terminală în unele cazuri. Mutantul de deleție, provenit din dimerul T2, prezintă o disociere rapidă la temperaturi mai ridicate. Cu toate acestea, disocierea totală a fost observată pentru configurația originală T2 în unele simulări de MD în solvent implicit la temperaturi ridicate. Disocierea rapidă a fost observată și pentru interfața T3 în aceste condiții. Coada flexibilă N-terminal poate juca un rol în formarea dimerului acționând ca o ancoră, facilitând eventual contactul inițial între doi monomeri.

Mutațiile punctuale au fost generate pentru aminoacizii de pe interfață, iar mutații au fost minimizați. Au fost selectate cinci structuri aleatorii și au fost efectuate simulări MD pentru a prezice orice modificare care afectează stabilitatea proteinelor. Nu se pot observa modificări semnificative și sunt necesare calcule suplimentare pentru a găsi o mutație care are un impact mai mare asupra stabilității proteinelor.

După ce am dobândit o perspectivă mai profundă asupra comportamentului dimerelor proteice, am ales un pentamer de dimeri din capsidul original ss-CCMV, deoarece acesta este următorul nivel în procesul de auto-asamblare. Pentamerul dimerelor, care conținea interfețe T1 și T2, a fost stabil în timpul simulării în solvent explicit de 2 μ s. Dimerele separate pentru analiză au prezentat comportamente diferite în timpul simulării, unele dintre ele au arătat asemănare cu simularea dimerelor independente.

5.2. Lista publicațiilor originale

5.2.1. Publicații

1. Z. Antal, **J. Szoverfi**, and S. N. Fejer, Predicting the Initial Steps of Salt-Stable Cowpea Chlorotic Mottle Virus Capsid Assembly with Atomistic Force Fields, *J. Chem. Inf. Model.*, vol. 57, no. 4, 2017, doi: 10.1021/acs.jcim.7b00078, **IF₂₀₁₇= 3.804**
2. **J. Szövérfi**, Cs. K. Orbán, B. Albert, K. Nagy, P. Salamon, Sz. Lányi, In Vitro Study Of The CCMV Capsid Protein: Cloning, Expression, And Purification, *U.P.B. Sci. Bull., Series B*, Vol. 83, Iss. 1, 2021, pp. 135-142

3. **Szoverfi, J.**, Fejer, S.N. Dynamic stability of salt stable cowpea chlorotic mottle virus capsid protein dimers and pentamers of dimers. *Sci Rep* 12, 14251 (2022).
<https://doi.org/10.1038/s41598-022-18019-9> , **IF₂₀₂₁= 4.996**

5.2.2. Conferințe

1. Molecular modeling in chemistry and biochemistry MOLMOD 2016, November 2016, Cluj Napoca, Romania “Modelling the dimerization of the CCMV capsid protein”, oral presentation
2. 22nd International Conference on Chemistry, November 2016, Timisoara, Romania, “Modelling the dimerization of the CCMV capsid protein”, oral presentation
3. 20th Romanian International Conference on Chemistry and Chemical Engineering, September 2017, Poiana Brasov, Romania, “Modelling the thermal stability of wild-type and mutant dimers of the CCMV capsid protein”, poster
4. 23rd International Conference on Chemistry, October, 2017, Deva, Romania, “Modelling the thermal stability of wild-type and mutant dimers of the CCMV capsid protein”, oral presentation
5. 24th International Conference on Chemistry, October, 2018, Sovata, Romania, “Cloning, Heterologous Expression and Molecular Dynamics Simulation of the CCMV Capsid Protein”, oral presentation
6. Molecular modeling in chemistry and biochemistry MOLMOD 2018, October 2019, Cluj Napoca, Romania, “Molecular Dynamics Studies of CCMV Capsid Protein oligomers”, oral presentation

5.3. Perspective pentru dezvoltări ulterioare

Auto-asamblarea CCMV este studiată pe scară largă atât *in vitro*, cât și *in silico* din anii '60, cu toate acestea multe întrebări au rămas fără răspuns. Capsidele goale sunt utilizate ca nanocarriers, astfel încât numeroase aplicații sunt posibile.

Pentru a continua investigarea formării a capsidelor, se poate genera o suprafață energetică potențială pentru două unități proteice rigide. O funcție analitică poate fi montată pe suprafața energetică potențială.

Odată cu coarse-graining a structurilor cu interfețe dimerice stabile, simularea auto-asamblării a unităților proteice coarse-grained poate fi simplificată, realizând simularea întregului capsid.

În ceea ce privește partea experimentală pot fi generați mutanți, care perturbă cea mai bună interfață și se poate face o comparație a cineticii de asamblare pentru tipul sălbatic și proteina mutantă.

Digestia proteinelor și procesul de purificare pot fi optimizate în continuare, pentru un randament mai mare al proteinei capside CCMV.

Expimarea pe scară largă a proteinelor poate fi efectuată în bioreactor.

Bibliografie

- [1] A. Zeltins, "Construction and characterization of virus-like particles: A review," *Mol. Biotechnol.*, vol. 53, no. 1, pp. 92–107, 2013, doi: 10.1007/s12033-012-9598-4.
- [2] R. F. Bruinsma, G. J. L. Wuite, and W. H. Roos, "Physics of viral dynamics," *Nature Reviews Physics*, vol. 3, no. 2. Springer Nature, pp. 76–91, Feb. 01, 2021, doi: 10.1038/s42254-020-00267-1.
- [3] E. C. Dykeman, P. G. Stockley, and R. Twarock, "Building a viral capsid in the presence of genomic RNA," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 87, no. 2, Feb. 2013, doi: 10.1103/PhysRevE.87.022717.
- [4] E. R. May, K. Arora, R. V. Mannige, H. D. Nguyen, and C. L. Brooks, "Multiscale Modeling of Virus Structure, Assembly, and Dynamics," pp. 167–189, 2012, doi: 10.1007/978-1-4614-2146-7_7.
- [5] L. Lavelle, J. P. Michel, and M. Gingery, "The disassembly, reassembly and stability of CCMV protein capsids," *J. Virol. Methods*, vol. 146, no. 1–2, pp. 311–316, 2007, doi: 10.1016/j.jviromet.2007.07.020.
- [6] Z. Antal, J. Szoverfi, and S. N. Fejer, "Predicting the Initial Steps of Salt-Stable Cowpea Chlorotic Mottle Virus Capsid Assembly with Atomistic Force Fields," *J. Chem. Inf. Model.*, vol. 57, no. 4, 2017, doi: 10.1021/acs.jcim.7b00078.
- [7] D. Law-Hine *et al.*, "Reconstruction of the disassembly pathway of an icosahedral viral capsid and shape determination of two successive intermediates," *J. Phys. Chem. Lett.*, vol. 6, no. 13, pp. 3471–3476, 2015, doi: 10.1021/acs.jpcllett.5b01478.
- [8] H. A. Yu *et al.*, "Characterization of ubiquitin C-terminal hydrolase 1 (YUH1) from *Saccharomyces cerevisiae* expressed in recombinant *Escherichia coli*," *Protein Expr. Purif.*, vol. 56, no. 1, pp. 20–26, 2007, doi: 10.1016/j.pep.2007.07.005.
- [9] "RCSB PDB - 1ZA7: The crystal structure of salt stable cowpea chlorotic mottle virus at 2.7 angstroms resolution." <https://www.rcsb.org/structure/1ZA7> (accessed May 29, 2021).
- [10] "ZDOCK Server: An automatic protein docking server." <https://zdock.umassmed.edu/> (accessed May 29, 2021).
- [11] S. Chaudhury, M. Berrondo, B. D. Weitzner, P. Muthu, H. Bergman, and J. J. Gray, "Benchmarking and analysis of protein docking performance in Rosetta v3.2," *PLoS One*, vol. 6, no. 8, p. 22477, 2011, doi: 10.1371/journal.pone.0022477.
- [12] C. Schafmeister, W. S. Ross, and V. Romanovski, "LEaP," *Univ. California, San Fr.*, 1995.
- [13] Y. Duan *et al.*, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations," *J. Comput. Chem.*, vol. 24, no. 16, pp. 1999–2012, 2003, doi: 10.1002/jcc.10349.
- [14] "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? - Wang - 2000 - Journal of Computational Chemistry - Wiley Online Library." <https://onlinelibrary.wiley.com/doi/10.1002/1096-987X%28200009%2921%3A12%3C1049%3A%3AAID-JCC3%3E3.0.CO%3B2-F> (accessed May 29, 2021).
- [15] P. Alam and N. Van Hieu, "Protein unfolding versus β -sheet separation in spider silk nanocrystals," *Adv. Nat. Sci. Nanosci. Nanotechnol.*, vol. 5, no. 1, p., 2014, doi: 10.1088/2043-6262/5/1/015015.
- [16] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the generalized born model suitable for macromolecules," *J. Phys. Chem. B*, vol. 104, no. 15, pp. 3712–3720, Apr. 2000, doi: 10.1021/jp994072s.
- [17] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms," *J. Phys. Chem. A*, vol. 101, no. 28, pp. 5111–5116, Jul. 1997, doi: 10.1021/jp970984n.
- [18] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein,

- “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, p. 926, 1983, doi: 10.1063/1.445869.
- [19] I. Maffucci and A. Contini, “Improved Computation of Protein-Protein Relative Binding Energies with the Nwat-MMGBSA Method,” *J. Chem. Inf. Model.*, vol. 56, no. 9, pp. 1692–1704, 2016, doi: 10.1021/acs.jcim.6b00196.
- [20] T. Darden, D. York, and L. Pedersen, “Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems,” *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993, doi: 10.1063/1.464397.
- [21] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes,” *J. Comput. Phys.*, vol. 23, no. 3, pp. 327–341, 1977, doi: 10.1016/0021-9991(77)90098-5.
- [22] A. Patriksson, D. van der Spoel, H. J. C. Berendsen, B. M. Rode, R. Huber, and H. J. C. Berendsen, “A temperature predictor for parallel tempering simulations,” *Phys. Chem. Chem. Phys.*, vol. 10, no. 15, p. 2073, Apr. 2008, doi: 10.1039/b716554d.
- [23] E. Feyfant, A. Sali, and A. Fiser, “Modeling mutations in protein structures,” *Protein Sci.*, vol. 16, no. 9, pp. 2030–41, 2007, doi: 10.1110/ps.072855507.
- [24] D. R. Roe and T. E. Cheatham, “PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data,” *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, 2013, doi: 10.1021/ct400341p.
- [25] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996, doi: 10.1016/0263-7855(96)00018-5.
- [26] E. F. Pettersen *et al.*, “UCSF Chimera—A Visualization System for Exploratory Research and Analysis,” *J. Comput. Chem.*, vol. 25, pp. 1605–1612, 2004, doi: 10.1002/jcc.20084.
- [27] M. L. Connolly, “Analytical molecular surface calculation,” *J. Appl. Crystallogr.*, vol. 16, pp. 548–558, 1983, doi: 10.1107/S0021889883010985.
- [28] W. Kabsch and C. Sander, “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features,” *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983, doi: 10.1002/bip.360221211.
- [29] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, “MMPBSA.py: An efficient program for end-state free energy calculations,” *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 3314–3321, 2012, doi: 10.1021/ct300418h.
- [30] A.-M. Catanzariti, T. A. Soboleva, D. A. Jans, P. G. Board, and R. T. Baker, “An efficient system for high-level expression and easy purification of authentic recombinant proteins,” *Protein Sci.*, vol. 13, no. 5, pp. 1331–9, May 2004, doi: 10.1110/ps.04618904.
- [31] A. Bakan, L. M. Meireles, and I. Bahar, “ProDy: Protein dynamics inferred from theory and experiments,” *Bioinformatics*, vol. 27, no. 11, pp. 1575–1577, 2011, doi: 10.1093/bioinformatics/btr168.
- [32] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins,” *Nature*, vol. 267, no. 5612, pp. 585–90, Jun. 1977, Accessed: Jan. 19, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/301613>.
- [33] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, “Mechanism of capsid assembly for an icosahedral plant virus,” *Virology*, vol. 277, no. 2, pp. 450–456, 2000, doi: 10.1006/viro.2000.0619.