



ACADEMIA ROMÂNĂ
INSTITUTUL DE BIOCHIMIE

REZUMATUL TEZEI DE DOCTORAT

**Metode de calcul cu utilizări în modelarea și
simularea biomoleculară**

Coordonator științific

Dr. Andrei-José Petrescu

Doctorand

Marius Surleac

BUCUREȘTI

2017

CUPRINS (TEZA IN EXTENSO)

CUPRINS	i
CUPRINS FIGURI	iii
CUPRINS TABELE	v
Mulțumiri	vi
Lista de abrevieri	vii
Scopul lucrării	x
I. INTRODUCERE GENERALĂ	1
I.1 Structura proteinelor	2
I.1.1 Domeniile	3
I.1.2 Structura secundară și torsiunea lanțului polipeptidic	4
I.1.3 Motivele structurale	5
I.2 Structura acizilor nucleici - Structura ADN-ului	5
I.2.1 Unghiurile de torsiune de pe lanțul de fosfați al ADN-ului	9
I.2.2 Geometria perechilor de baze	10
I.2.3 Îndoirea și topologia ADN-ului	11
I.3 Metode de determinare a structurii macromoleculare	12
II. METODE COMPUTAȚIONALE DE MODELARE ȘI SIMULARE MOLECULARĂ	14
II.1 Modelarea prin omologie a structurilor de proteine	15
II.1.1 Identificarea de proteine șablon	18
II.1.2 Predicțiile făcute pe baza secvenței țintă	19
II.1.2.1 Predicțiile de structură secundară	19
II.1.2.2 Predicțiile de dezordine	20
II.1.2.3 Predicțiile de domenii, motive	20
II.1.2.4 Predicțiile de modificări posttranslaționale	21
II.1.2.5 Predicțiile de hidrofobicitate și pliere	21
II.1.2.6 Profilurile de sarcină și variabilitate	22
II.1.3 Identificarea de similarități între secvențele țintă și șablon	22
II.1.4 Generarea modelelor structurale 3D pentru secvențele țintă ale proteinelor studiate	23
II.1.5 Validarea unui model de omologie	24
II.2 Metode de simulare de dinamică moleculară a proteinelor și acizilor nucleici	25
II.3 Metodă inovativă - Generarea modelelor de <i>coarse-grain</i> de îndoire a fragmentelor de ADN - studiu de caz: ADN-urile 12RSS și 23RSS din recombinarea V(D)J	27
II.4 Simulările de docking molecular	28
III. RECOMBINAREA V(D)J. COMPLEXUL PROTEIC RAG1/2	30
III.1 Introducere în mecanismul de acțiune al V(D)J-RAG1/2	31
III.2 Rezultate	33
III.2.1 Modelarea fragmentelor de ADN 12RSS & 23RSS	34
III.2.1.1 Modelul structural de ADN 23RSS îndoit în PC	40

III.2.1.2	Modelul structural de ADN 12RSS îndoit în PC și SC	45
III.2.2	Modelarea structurii de β -propeller din RAG2	49
III.2.3	Modelarea domeniului catalitic RNase-H din RAG1	63
III.3	Discuții	69
III.3.1	Comparația structurilor tridimensionale ale modelelor și structurilor cristalizate de RAG1 și RAG2	69
III.3.2	Punerea la punct a ansamblului RAG1/2-12/23RSS	73
III.3.3	Comparația ansamblului RAG1/2-12/23RSS modelat, cu structura cristalizată a aceluiași complex	78
III.4	Perspective de viitor	79
IV.	TOPOIZOMERAZELE II α ȘI II β	80
IV.1	Introducere în mecanismele de acțiune ale topoisomerazelor II α și II β	81
IV.2	Rezultate	84
IV.2.1	Modelarea domeniului de legare al ADN-ului din topoisomeraza II β	84
IV.2.2	Modelarea domeniului de legare al ADN-ului din topoisomeraza II α	87
IV.3	Discuții	90
IV.3.1	Generarea modelelor de dimer pentru izoformele topo II α și II β	91
IV.3.2	Analiza diferențelor dintre izoformele topo II α și II β în domeniul DNAbd	92
IV.3.3	Analiza modificărilor posttranzlaționale prezise pentru domeniul DNAbd din ambele izoforme de topo II	94
IV.3.4	Analiza regiunilor din imediata vecinătate a Y β 656/Y α 640	96
IV.3.5	Analiza domeniilor CTD ale ambelor izoforme	99
IV.3.6	Model de acțiune al celor două izoforme în suprarăsucirile (+)(-)	101
IV.4	Perspective de viitor	103
V.	ENZIMA "DECAPPING SCAVENGER" (DcpS)	104
V.1	Introducere în mecanismele de degradare a ARN-ului mesager	105
V.2	Rezultate	106
V.2.1	Modelarea structurii de dimer a enzimei DcpS_Ce	107
V.2.2	Docking de compuși în situsul activ al enzimei DcpS_Ce	109
V.3	Discuții	114
V.3.1	Analiza situsului de legare al m ⁷ GpppG în DcpS_Ce	114
V.4	Perspective de viitor	116
	Concluzii finale	117
	Listă de lucrări și participări la conferințe	119
	Referințe	124

Mulțumiri

Această lucrare a fost realizată cu sprijinul financiar al următoarelor proiecte:

UEFISCDI: PN-II-ID-PCE-2011-3-0342;

UEFISCDI: PN-II-PT-PCCA-2013-4-1407;

H2020: HIVERA-INinRAGI;

Aș dori să mulțumesc în mod special:

- Familiei, pentru tot sprijinul acordat de-a lungul timpului
- Coordonatorului de doctorat Dr. Andrei-José Petrescu și doamnei Director Dr. Ștefana Petrescu, pentru îndrumare și ajutorul acordat pentru punerea la punct a prezentei teze.
- Dr. Laurențiu Spiridon
- Dr. Adina Milac
- Dr. Marius Micluță
- Dr. Mihai Ciubotaru din cadrul departamentului de Imunologie, Universitatea Yale, New Haven, Connecticut, SUA și IFIN-HH, București, România
- Profesorului David G. Schatz, șeful departamentului de Imunobiologie, Universitatea Yale, New Haven, Connecticut, SUA
- Dr. Mahrukh Ganapathi și Dr. Ram Ganapathi din cadrul Levine Cancer Institute, Carolinas Healthcare System, Carolina de Nord, SUA
- Dr. Anna Wypijewska del Nogal și Dr. Elżbieta Bojarska din cadrul Facultății de Fizică, Universitatea din Varșovia, Polonia
- Colegilor din Institutul de Biochimie

Scopul lucrării

Scopul studiilor cuprinse în această teză de doctorat a fost în primul rând acela de a dezvolta și implementa tehnici speciale de modelare și simulare biomoleculară adaptate studierii interacțiilor și efectelor induse de proteine asupra acizilor nucleici, în procese biologice vitale precum: recombinarea somatică, decatenarea și degradarea ARN-mesager; procese aflate

printre temele cele mai importante, și intens studiate de nenumărate grupuri de cercetare de top, din universități de prestigiu și care se întind pe perioade de zeci de ani.

- Recombinarea somatică sau V(D)J - Variable, Diverse, Joining - este un proces extrem de important pentru răspunsul imun al organismului la factori externi (ex. antigeni) care pot afecta integritatea sa și are rol în codificarea zonelor hipervariabile din imunoglobuline și în receptorii de celule T. Investigațiile noastre s-au concentrat aici, în special pe modelarea structurală a principalelor proteine și fragmente genice implicate în acest mecanism, în speță proteinele RAG1 și RAG2 și fragmentele 12RSS și 23RSS, și a avut ca scop înțelegerea mecanismului de acțiune a acestor tipuri de proteine în interacția cu structurile ADN. Mutațiile, sau alți factori care afectează buna conduită a acestor proteine, cauzează o serie de boli autoimune care, de cele mai multe ori sunt letale.
- Decatenarea și relaxarea ADN-ului cromozomial suprarăsucit, este un mecanism molecular esențial întâlnit într-o serie de procese precum replicare, transcripție, reparare de ADN, ciclul celular. Studiul s-a concentrat pe modelarea structurilor și analiza mecanismului de acțiune al celor două izoforme de topoizomerază umană de tip II (topo II α și topo II β) în controlul punctului de control de decatenare dar și pentru a înțelege implicarea lor în suprarăsucirea ADN-ului. Aceste două proteine sunt vitale în funcționarea în condiții normale a celulei iar mutațiile sau factorii care le afectează funcția, duc la implicarea acestora într-o serie de boli precum cancerul, fiind astfel intens studiate pentru dezvoltarea de medicamente.
- Reglarea expresiei genice prin procesul de degradare a ARN-ului mesager. Studiul în această direcție s-a concentrat pe modelarea și analiza mecanismului de acțiune al proteinei DcpS, cu rol în fragmentarea capetelor de ARN mesager, ulterioră îndepărtării cozii poly(A) prin procesul de deadenilare. În același timp a fost studiată capacitatea de legare a diverși compuși (posibil terapeutici) analogi structurii nucleotidei metilate de la capătul 5' al moleculei de ARN mesager, raportat la situsul activ al proteinei DcpS.

În plus, investigarea acestor importante sisteme a fost un bun prilej de a dezvolta sau de a îmbunătăți o serie de metode de calcul mai generale, care să fie, de altfel, de ajutor în viitoarele modelări de structură de proteine și acizi nucleici. Astfel, am pus la punct o metodă de îndoirea a fragmentelor de ADN prin intermediul simulărilor de dinamică moleculară. Am îmbunătățit, de asemenea, metodele de modelare a proteinelor prin înglobarea de informații multiple ce țin de predicții (de structură secundară, de dezordine, de modificări posttranslaționale, etc.), de

punerea la punct a unor profile de sarcină și variabilitate între specii. Am îmbunătățit analiza modelelor dar și a simulărilor de dinamică moleculară a acestora, prin dezvoltarea de programe/script-uri scrise în limbaje de programare precum AWK, Tcl, care să calculeze deviații standard între structuri de proteine, distanțe între tipuri de atomi, care să creeze distribuții de valori ale unghiurilor de torsiune în proteine, care să analizeze baze de date de secvențe, etc. În esență, dezvoltarea tuturor acestor programe are rolul de a simplifica analize de date care, în lipsa lor, ar putea dura ore sau chiar nenumărate zile de lucru; și de asemenea acoperă segmente de cercetare pentru care nu s-au dezvoltat astfel de instrumente. Toate acestea sunt incluse în lucrările științifice și capitolele de carte publicate.

INTRODUCERE GENERALĂ ȘI METODE

Metabolismul celular depinde de mii de interacții și reacții coordonate între ele în spațiu și timp, dependente atât de instrucțiunile genice cât și de mediul de operare. Efectorii acestui întreg complex sistem sunt proteinele, de a căror structură și dinamică intricată depinde întreaga funcționalitate a sistemului biologic.

În scurta introducere de mai jos sunt prezentate sumar principalele concepte și tehnici structurale ce stau la baza investigării sistemelor proteină-ADN ce fac obiectul prezentei lucrări.

Structura proteinelor

Proteinele sunt lanțuri polimerice lineare formate din 20 tipuri de aminoacizii (aa) înșiruiți în secvențe de zeci, sute sau mii de unități. Informația referitoare la secvență este codificată în genom; iar în funcție de nevoi aceasta este transcrisă și tradusă în lanț proteic pe ribozomi în timpul procesului de biosinteză (Alberts et al., 2008; Berg et al., 2011). Concomitent cu biosinteza, chiar în timpul elongării, lanțul proteic începe să se organizeze spațial sub un control strict celular intermediat de proteine de asistență numite chaperoane (Hartl et al., 2011). Procesul de împachetare 3D a lanțului proteic și aducerea sa la forma nativă, funcțională, poartă numele de pliere, și în funcție de complexitatea lanțului poate dura chiar ore după terminarea biosintezei (Dill și MacCallum, 2012; Rognoni et al., 2014). Forma finală, funcțională a unei proteine poate conține zone ordonate, cu structură 3D bine precizată precum și zone intrinsec dezordonate în care lanțul polipeptidic adoptă configurații multiple. Experimental, structura domeniilor pliabile, cu organizare 3D bine precizată poate fi determinată prin cristalografie de raze X (Van Benschoten et al., 2016), rezonanță magnetică nucleară (Ma et al., 2015) sau mai

recent criomicroscopie electronică (Fernandez-Leiro și Scheres, 2016). Aceste tipuri de experimente structurale sunt extrem de complexe, costisitoare și consumatoare de timp, motiv pentru care numărul structurilor determinate până în prezent este sub 150.000, cu aproximativ trei ordine de mărime mai mic decât cel al secvențelor cunoscute (Finn et al., 2017; Dawson et al., 2017). Datorită importanței cruciale pe care cunoașterea structurii 3D o are în înțelegerea bazelor moleculare ale vieții în ultimii 20-30 de ani, eforturi susținute au fost dedicate analizei și organizării datelor structurale consacrate prin apariția bioinformaticii structurale (Samish et al., 2015). Dacă secvența polipeptidică este în genere cunoscută și sub numele de *structură primară* a proteinei, analiza împachetării lanțului proteic mai identifică încă trei nivele structurale, specifice organizării spațiale: structura *secundară*, *terțiară* și *cuaternară*. Structura *secundară* se referă la conformația locală a lanțului polipeptidic - care poate fi repetitivă, stabilizată prin punți de hidrogen (HB) sau nonrepetitivă, stabilizată sau nu de punți de hidrogen (Nelson și Cox, 2004). Modul în care elemente de structură secundară se grupează în spațiu pe porțiuni de secvență mai lungi formează așa-numitele *motive* structurale sau *structura super-secundară*; iar aranjarea generală a tuturor atomilor din proteină, incluzând și contacte între aminoacizi care se află la distanță mare unul față de altul în secvență, reprezintă al treilea nivel de organizare, cel de *structură terțiară*. Al patrulea nivel de organizare este dat de *structura cuaternară*. Structura cuaternară constă în aranjarea într-un singur complex a mai multor lanțuri polipeptidice / subunități care pot fi similare (homo-oligomeri) sau diferite (hetero-oligomeri). Subunitățile interacționează între ele și pot forma de exemplu situsuri active în proteină, pot fi implicate în interacțiunea cu alte proteine. Ținând cont de toate aceste niveluri de organizare, proteinele pot fi clasificate în două mari grupe: - proteine fibrilare (lanțurile polipeptidice sunt aranjate sub formă de fâșii lungi); - proteine globulare (lanțurile polipeptidice sunt pliate în formă globulară sau sferică) (Nelson și Cox, 2004).

Structura acizilor nucleici - Structura ADN-ului

Într-o celulă, moleculele care poartă informația sunt de două tipuri: ADN (acid deoxiribonucleic) și ARN (acid ribonucleic), construite pe baza unor unități esențiale numite *nucleotide* (Lodish et al., 2003). ADN-ul este molecula de bază a fiecărei celule, se găsește în nucleul celulei la eucariote, iar pentru a-și executa funcția purtătoare de informație, trebuie să facă mai mult decât să se copieze pe sine, fiind astfel implicat în ghidarea sintezei celorlalte molecule din celulă, în primul rând ARN și proteine (Lukacs et al., 2000). Deteriorarea ADN-ului (sub acțiunea radiațiilor X, UV sau a anumitor agenți chimici precum speciile reactive de

oxigen) sau mutațiile (care pot să apară și să ducă la aberații cromozomiale în timpul separării cromozomiale) pot cauza diverse tipuri de cancer, moarte celulară sau îmbătrânire (Hoeijmakers, 2001).

Structura tridimensională a ADN-ului constă în două *catene*, răsucite împreună pentru a forma așa-numita structură de *elice dublă* (Alberts et al., 2008).

Scheletul laturilor ADN-ului este dat de grupările fosfat-glucid (aflate spre exteriorul dublei elice) ale nucleotidelor adiacente legate împreună, iar laturile sunt ținute împreună cu ajutorul legăturilor de H care se formează între bazele complementare, aflate în interiorul dublei elice (Berg et al., 2011).

Stabilitatea structurii de elice dublă a ADN-ului se datorează în principal influenței a doi factori: asocierea între bazele din catenele complementare (prin intermediul legăturilor de H și interacțiilor hidrofobe între baze) și interacției de stacking dintre bazele adiacente, care la rândul lor sunt influențate de temperatura de topire și de concentrațiile de săruri (Yakovchuk et al., 2006) mai ales datorită cationilor, care pot neutraliza sarcinile negative ale grupărilor fosfat (Zhang et al., 2015).

Carbonul C3' din unitatea glucidică este conectat printr-o grupare fosfat la carbonul C5' a următoarei unități glucidice. Legătura astfel formată se numește legătură fosfodiester 3'-5' (Pratt și Cornely, 2013). Toate catenele de ADN sunt citite de la capătul 5' la capătul 3', unde capătul 5' se termină cu o grupare fosfat iar capătul 3' se termină cu o unitate glucidică (Koolman și Roehm, 2005). Fiecare unitate este legată covalent prin atomul C1' la una din cele 4 baze posibile (la atomul N1 din pirimidine sau la atomul N9 din purine) printr-o legătură N-β-glicozidică (Nelson și Cox, 2004). Bazele sunt orientate perpendicular la axa elicei; sunt hidrofobe în direcția perpendiculară la planul bazelor (astfel nu pot forma legături de H cu apa) (Kuriyan et al., 2013) în timp ce exteriorul ADN-ului este încărcat negativ. Datorită formei sale de dublă elice, de-a lungul lungimii moleculei de ADN se formează două cavități, numite *canelura minoră* și *canelura majoră* (Pratt și Cornely, 2013) care sunt implicate în interacția cu o serie de proteine printre care și factori de transcripție (Privalov et al., 2007). Catena 5'-3' se numește *catenă sens*, în schimb catena 3'-5' se numește *catenă anti-sens* (Koolman și Roehm, 2005).

Există trei forme de ADN (A-, B- și Z-) din care cea mai des întâlnită structură este cea B-ADN - care, în funcție de anumiți factori, poate adopta formele A-ADN (în condiții de hidratare scăzută) sau Z-ADN (având concentrație crescută de regiuni GC în secvență) (Koolman și Roehm, 2005).

Molecula de ADN nu este o structură rigidă - în realitate, fiecare dintre aceste tipuri de structuri se află într-o continuă fluctuație termică, interacții cu alte molecule (ex. proteine, molecule de apă), interacții cu ioni - toate acestea duc la răsuciri locale, întinderi, îndoiri sau desfaceri ale catenelor, etc (Westman, 2006).

Metodă inovativă - Generarea modelelor de *coarse-grain* de îndoire a fragmentelor de ADN - studiu de caz: ADN-urile 12RSS și 23RSS din recombinarea V(D)J

Distanțele măsurate prin FRET (între fluorofori atașati de diverse baze de ADN) de către colaboratorii noștri de la Yale School of Medicine, sugerează că în cazul formării complexului PC (23+12RSS) din cadrul recombinării V(D)J, fragmentele de ADN 12/23RSS sunt puternic îndoite de către complexul proteic RAG1/2+HMGB1/2. Pentru a modela curbura ADN-ului în conformitate cu constrângerile FRET observate experimental este nevoie de impunerea unor constrângeri inegale pe două părți opuse ale ADN-ului. În ADN-ul liniar de formă B, periodicitatea este de 10.5 baze per tură iar distanțele medii dintre atomii echivalenți ale celor două catene antiparalele este de 20Å și de 14Å de-a lungul canelurilor majore, și respectiv minore (Ciubotaru et al., 2013). Modelul de *coarse-grain* pe care l-am dezvoltat în cadrul acestei teze împreună cu colegii mei, presupune îndoirea ușoară a ADN-ului pe baza unor serii de constrângeri de distanță armonice, impuse gradual și inegal pe canelurile majore și minore, astfel încât să afecteze parametrii *tilt* și *roll*, discutați în capitolul I de Introducere generală. Pentru ca structura ADN-ului să nu devieze de la forma B într-o formă neregulată, am pus constrângeri de distanță și pe parametrii σ , ω și κ (corespunzători deschiderii perechilor de baze, parametrului propeller twist și respectiv a îndoirii perechilor de baze). Astfel, pe o parte a elicei dublu-catenare am impus o serie de constrângeri de distanță mai mari pe ambele tipuri de caneluri (cu aproximativ 10% între reziduurile n , $n+10$, etc.) pentru a forța o ușoară depărtare între bazele de pe partea convexă a îndoirii, iar pe cealaltă parte a ADN-ului am impus constrângeri de distanță cu aproximativ 10% mai mici pe ambele tipuri de caneluri (între reziduurile $n+5$, $n+15$, etc.).

Această metodă o voi discuta mai în detaliu în capitolul de rezultate dedicat recombinării V(D)J.

REZULTATE

RECOMBINAREA V(D)J. COMPLEXUL PROTEIC RAG1/2

Până de curând nu se știau prea multe despre structurile fragmentelor de ADN recombinant 12RSS și 23RSS în complex cu proteinele RAG, și mecanismul în detaliu al modului lor de acțiune. Pentru a adresa această problemă, primul pas în rezolvarea acestui "puzzle" a fost să generăm modele structurale teoretice pentru cele două RSS-uri pe bază de date experimentale FRET primite de la colaboratorii noștri de la universitatea Yale, dar și modele pentru situsul catalitic din RAG1 și domeniul de tip β -propeller al RAG2.

ADN-ul 23RSS de formă B, pe care l-am folosit pentru modelare, conține o secvență de nucleotide de 65 de perechi de baze, fiind format dintr-o zonă de 16 perechi de baze ce codifică pentru zonele hipervariabile din receptorii de antigene, un heptamer de 7 perechi de baze, un "spacer" de 23 perechi de baze, un nonamer de 9 perechi de baze și un scurt fragment de 10 perechi de baze ce nu sunt implicate în codificare. ADN-ul 12RSS de formă B, pe care l-am folosit, conține o secvență de 59 de perechi de baze, cu mențiunea că singurele deosebiri față de ADN-ul 23RSS se regăsesc în zona spacer-ului (acesta având o lungime de 12 perechi de baze) și în regiunea care nu codifică (această zonă are 15 perechi de baze), pe când celelalte trei zone sunt identice cu cele din 23RSS.

Primul pas în modelarea structurilor de ADN 23RSS și 12RSS a fost generarea unei structuri 3D, pe baza secvențelor celor două RSS-uri, extinse pentru fiecare din cele două, folosind programul NAB (Macke și Case, 1998) care face parte din suita software AMBER (Salomon-Ferrer et al., 2013).

Pe baza acestor structuri extinse de RSS, următorul pas a fost cel în care am pregătit fișierele de input necesare proceselor de minimizare de energie și dinamică moleculară (MD) de "călire" simulată Generalized Born (GB), asemenea celor descrise la capitolul de Metode. Pentru început, în pregătirea fișierelor de input am folosit programul xLEaP din Amber9 - am generat fișierele de topologie și coordonate, am adăugat atomii de H și ionii de Mg^{2+} la structura ADN-ului, folosind parametrii pentru proteine și acizi nucleici din câmpul de forțe ff99SB (Hornak et al., 2006). Apoi am folosit programul *ambpdb* din Amber9 (Case et al., 2006) pentru a face conversia fișierelor obținute din Amber la formatul de tip *.pdb*.

În pasul următor am pregătit fișierele de constrângeri. Pentru păstrarea caracterului de ADN de formă B am folosit: a) constrângeri pe unghiuri diedre - două tipuri de constrângeri unghiulare aplicate pe perechile de baze, ca să fie păstrată planaritatea acestora; b) constrângeri de distanță, de tip Watson-Crick, pe legăturile de H dintre bazele de ADN - pentru a nu permite separarea

catenelor; c) constrângeri de distanță pe ionii de Mg^{2+} - în care am inclus distanțele dintre ioni și atomii de P din scheletul de fosfați ai ADN-ului, astfel încât ionii de Mg^{2+} să rămână la o distanță optimă față de ADN, și să nu perturbe structura sa locală, dar nici să nu fie prea departe de ADN în așa fel încât să nu exercite nici o influență asupra acestuia.

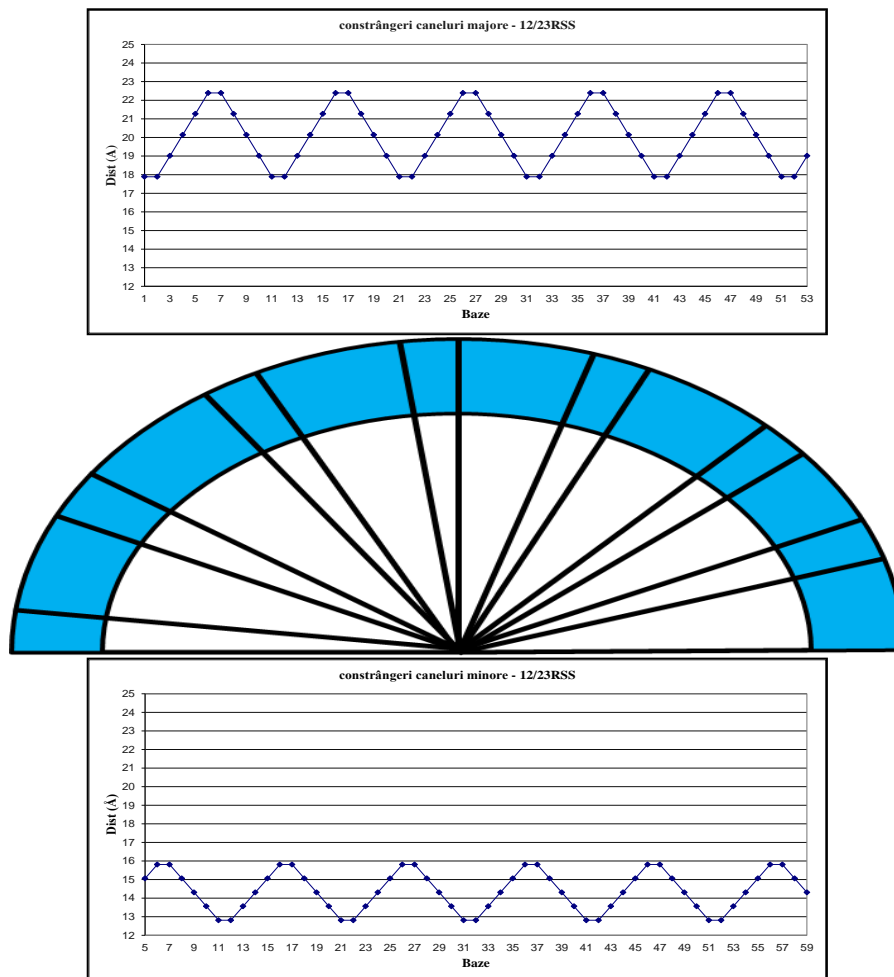


Fig. III.3 Constrângerile de distanță pe canelurile majore și minore în ADN-urile 23RSS și 12RSS

Pe lângă aceste constrângeri am mai aplicat constrângeri de distanță inegale și periodice (Fig. III.3) pe canelurile minoră și majoră, de altfel foarte importante pentru generarea modelului coarse-grain de îndoire a ADN-urilor. Aceste constrângeri fac parte dintr-o metodă inovativă dezvoltată în cadrul laboratorului, discutată și în capitolul de Metode.

Simulările au fost efectuate pe un sistem de calcul de înaltă performanță (HPC) Bull NovaScale R422/R423 HPC-cluster, 32 nuclee CPU. Nu în ultimul rând am efectuat și simulări ale modelelor finale în solvent explicit, folosind programul NAMD (Phillips et al., 2005).

Modelul rezultat pentru ADN-ul 23RSS în PC arată că molecula de ADN este puternic îndoită sub forma literei "U", cu "spacer"-ul ocupând baza lui "U" iar celelalte zone din ADN (ex. heptamer, nonamer) aflându-se pe brațele lui "U" (Fig. III.5.A), plecând de la o distanță de

aproximativ 220Å între capetele ADN-ului (în forma extinsă a acestuia) și ajungând la aproximativ 77Å între capetele sale (în forma îndoită). În cadrul modelului se observă mai multe îndoituri distincte, dintre care 4 cele mai evidente se găsesc la interfața dintre porțiunea ce codifică și heptamer (51°, unghi exterior); interfața dintre heptamer și spacer (44°); centrul spacer-ului (49°) și interfața dintre spacer și nonamer (55°) (Fig. III.5.B) (Ciubotaru et al., 2013). Mai mult decât atât, este interesant că atunci când sunt reprezentate pe model situsuri de interacție (observate experimental prin procedura de *interferență de etilare*) între RAG și ADN în complex SC, majoritatea contactelor se găsesc pe suprafața interioară (concavă) a ADN-ului. În contrast, situsurile cu sensibilitate la Deoxiribonucleaza I în complex SEC, se regăsesc predominant pe suprafața convexă (spre exterior) a ADN-ului. E de ținut cont că aceste din urmă situsuri se găsesc în imediata apropiere a îndoiturilor cele mai evidente din spacer și îndoiturilor de la interfața dintre spacer și nonamer (Fig. III.5.C).

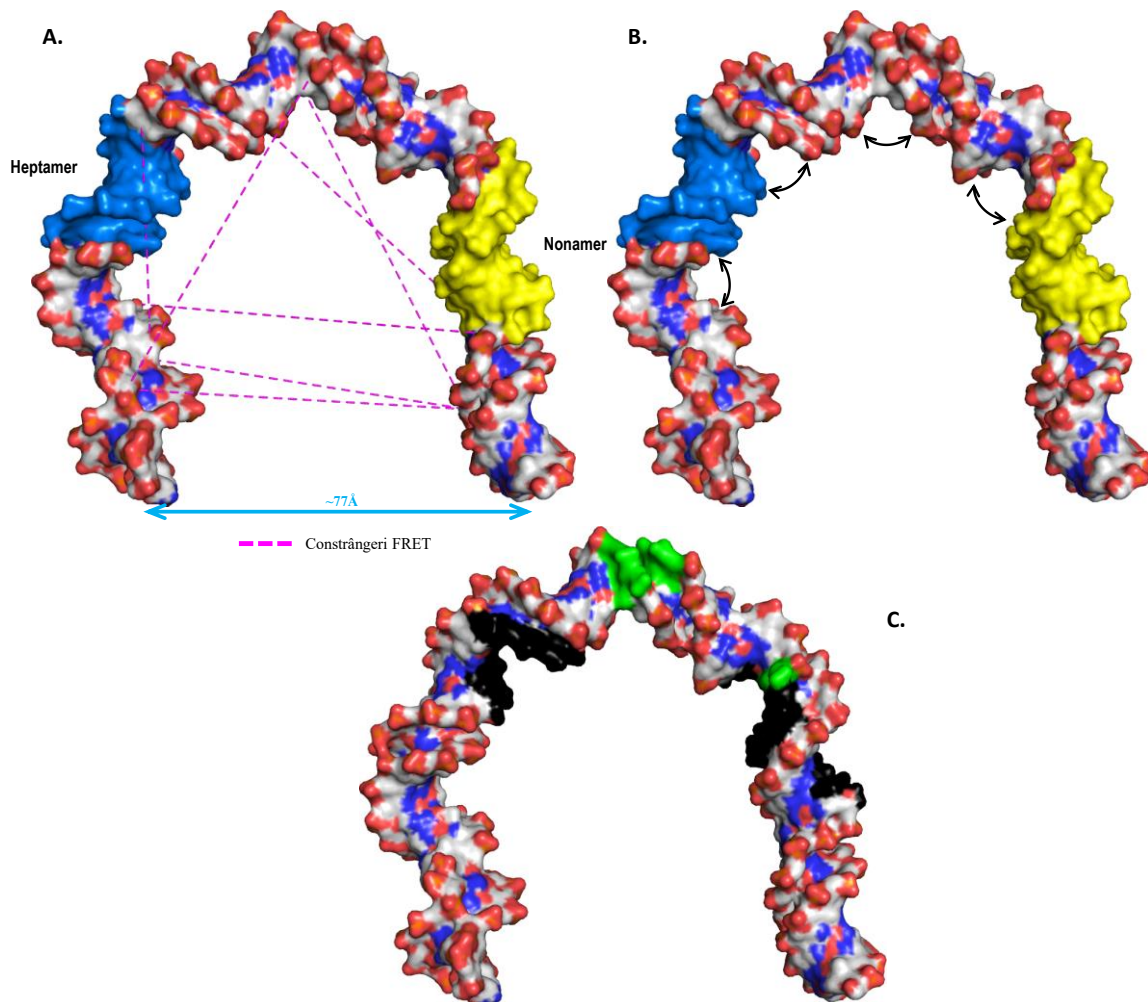


Fig. III.5 Model final de îndoire a ADN-ului 23RSS
A. Constrângeri FRET; **B.** Îndoiturile principale din 23RSS; **C.** Situsurile de etilare (negru) și hipersensibilitate (verde)

Analiza modelului de ADN-23RSS îndoit arată că majoritatea distanțelor din model sunt în conformitate cu cele măsurate în FRET. În plus, modelul a fost validat ulterior prin noi experimente de FRET.

Metoda de îndoire a ADN-ului 12RSS a fost similară cu cea de îndoirea a ADN-ului 23RSS.

Proteina RAG2, alături de RAG1, este o componentă a complexului proteic RAG, complex ce mediază faza de tăiere a ADN-ului în timpul recombinării somatice (sau V(D)J). În cadrul complexului, RAG2 nu reprezintă componenta catalitică, acest rol fiind jucat de către RAG1, însă RAG2 este necesară pentru activitățile catalitice mediate de RAG1.

RAG2 are o lungime de 527 de aminoacizi și este împărțită în două mari regiuni: un domeniu central N-terminal (1-383aa) care este necesar pentru reacția de recombinare, și un al doilea domeniu C-terminal (384-527) care nu este esențial (Elkin et al., 2005). Se știe că domeniul central adoptă o structură de β -propeller, ce reprezintă un motiv structural foarte întâlnit la multe proteine.

Inițial colaboratorii de la Yale ne-au pus la dispoziție secvența proteinei RAG2 umane (*Homo sapiens*) pentru care am generat primul model de omologie iar apoi, pe baza acestuia, am generat un model pentru proteina RAG2 de la șoarece (*Mus musculus*). La momentul generării modelului de RAG2 nu existau informații referitoare la structura tridimensională (ex. date de cristalografie de raze X, RMN) a acestui domeniu.

Pentru generarea modelului de RAG2 am luat în considerare primii 351aa din secvență. Pe baza acestei secvențe s-au efectuat predicții de structură secundară, predicții de dezordine, profiluri de sarcină dar și căutări de proteine similare, ”șablon”, ale căror structură este cunoscută, folosind metodologia prezentată detaliat în capitolul de Metode.

RAG1 este cea de-a doua proteină din cadrul complexului RAG ce mediază recombinarea V(D)J, cea care assemblează regiunile variabile ale genelor imunoglobulinice și cele ale receptorilor de celule T, în cadrul dezvoltării limfocitelor B și T (Zhang et al., 2015). RAG1 este principala componentă pentru legarea și tăierea ADN-ului. În prezentul studiu am lucrat pe secvența RAG1 de la șoarece (*Mus musculus*) care are o lungime de 1040 de aminoacizi. Această proteină este compusă dintr-o serie de domenii ”zinc finger” (ZF) în zona N-terminală conectate printr-o serie de legături (linker), urmate de un domeniu de dimerizare, un domeniu NBD (nonamer binding domain) care se leagă de nonamerii din 12RSS și 23RSS, un al doilea domeniu de dimerizare și legare de ADN (DDBD - dimerization and DNA binding domain), urmând ca spre zona C-terminală să fie continuat de un domeniu RNase-H. Domeniul RNase-

H este împărțit în două zone separate de un alt domeniu de inserție (ID - insertion domain) de aproximativ 230aa. Porțiunea centrală, zona minimă din RAG1 care este necesară pentru activitate, cuprinde aa384-1008, începând cu domeniul NBD. Pe lângă capacitatea RAG1 de a se lega direct de nonamerul din RSS-uri, RAG1 este capabilă de a interacționa și cu proteina RAG2 dar și de a recunoaște heptamerul din fiecare RSS (Shinkai et al., 1992; Zhang et al., 2015).

Domeniul RNase-H reprezintă centrul activ al proteinei RAG1 și conține motivul DDE care formează situsul catalitic al RAG1 și coordonează doi ioni de Mg^{2+} . Acest motiv DDE conține doi acizi aspartici și un acid glutamic, deși în unele proteine acidul glutamic este înlocuit tot cu un acid aspartic. Datorită prezenței acestui domeniu, RAG1 dispune de o serie de similarități cu alte proteine ce conțin același domeniu, printre care transpozazele *Tn5*, *Hermes* sau chiar integrezele HIV-1, HIV-2, PFV, care funcționează prin intermediul unor mecanisme similare de tăiere a ADN-ului. La fel ca și în cazul RAG2 (la momentul efectuării acestui studiu) prea puține se știau despre structura tridimensională a RAG1, în afară de cristalul domeniului NBD (fiind singura regiune cu structură cunoscută).

În prezentul studiu am generat un model de omologie pentru domeniul RNase-H al RAG1, folosind aceeași metodologie ca și în cazul RAG2.

Din păcate programele folosite pentru identificarea de proteine șablon nu au putut găsi o proteină șablon globală pentru întreaga proteină RAG1. Astfel, prin folosirea metodelor de modelare prin omologie îndepărtată a fost posibilă construirea modelului pentru acest domeniu RNase-H și pentru încă două zone (extensii) de la capetele acestui domeniu, corespunzând intervalelor aa538-593 și aa997-1010 (Zhang et al., 2015). Pentru metoda de modelare prin omologie îndepărtată am ținut cont de informațiile provenite din alinierea secvențelor domeniilor RNase-H de la mai multe proteine, efectuând astfel întâi o aliniere multiplă de secvențe care a fost ulterior rafinată pe baza alinierii elementelor de structură secundară după care e recunoscut acest tip de domeniu, dar în același timp am "blocat" în aliniere aminoacizii din motivul DDE (D600, D708 și E962). Domeniul RNase-H constă dintr-o serie de 9 elemente de structură secundară în următoarea ordine: $\beta 1-\beta 2-\beta 3-\alpha 1-\beta 4-\alpha 2-\beta 5-\alpha 3-\alpha 4$ și corespunde intervalelor aa594-732 și aa960-996 din secvența de RAG1.

Pentru construirea modelului am folosit drept șablon următoarele proteine care conțin domeniul RNase-H: transpozaza *Hermes* (cod PDB - 4d1q.pdb) (Hickman et al., 2014), metnază (cod PDB - 3k9k.db) (Goodwin et al., 2010), integraza HIV2 (cod PDB - 3f9k.pdb) (Hare et al., 2009-B), integraza HIV1 (cod PDB - 4ovl.pdb) (Peat et al., 2014), integraza de la virusul *Visna*

(cod PDB - 3hpg.pdb) (Hare et al., 2009-A), integrează PFV (Prototype Foamy Virus) (cod PDB - 3oya.pdb) (Hare et al., 2010).

Regula 12/23 în care se încadrează ADN-urile 12RSS și 23RSS în PC sugerează o preferință de către proteinele RAG pentru o asimetrie față de substrat, însă nu se știe motivul exact pentru care există această preferință (Ciubotaru et al., 2015). Motivația din spatele acestei cercetări ghidată atât pe fragmentele de ADN RSS cât și pe proteinele RAG a fost aceea de a înțelege pe cât posibil mecanismul de acțiune al recombinării somatice V(D)J și dacă nu cumva această asimetrie de substrat ar putea fi reflectată prin intermediul diferențelor în structura ADN-urilor 12RSS și 23RSS. Din rezultatele expuse în cadrul acestui capitol reiese că forma ambelor structuri de ADN îndoite este destul de similară, iar îndoirea puternică a ambelor substraturi este datorată tocmai proteinelor RAG și HMGB, rolul critic fiind jucat de proteina RAG1.

TOPOIZOMERAZELE II α ȘI II β

În studiul de față am generat modele de omologie pentru domeniile DNAbd (DNA binding domain) ale topoisomerazelor II de la om, am generat modele ale dimerilor celor două izoforme, am studiat implicațiile pe care două mutante Y α 640F și Y β 656F le au asupra funcției și am analizat secvențele domeniului C-terminal din fiecare izoformă împreună cu implicațiile pe care le pot avea. Cele două izoforme au o identitate de aproximativ 68% și o similaritate de 77% pe întreaga secvență. Ambele sunt exprimate similar în celulele care se divid, inclusiv în tumori, însă topo II β este mai abundentă în celulele care nu depind de ciclul celular (Shapiro și Austin, 2014). Topo II α este esențială pentru supraviețuirea celulelor *in vitro* (Akimitsu et al., 2003), în schimb topo II β nu este esențială pentru celule *in vitro* dar este necesară, de exemplu, pentru diferențierea neuronală (Lyu și Wang, 2003).

La momentul efectuării acestui studiu, nu existau prea multe informații cu privire la structura tridimensională a celor două izoforme, în special date de cristalografie de raze X pentru fiecare dintre ele. Dintre cele trei domenii se știa doar structura domeniului ATPbd obținută cu ajutorul metodei de cristalografie de raze X, însă nu și pentru celelalte două domenii. Secvențele de topoisomerase II α (1531aa) și II β (cea de 1621aa), necesare pentru a genera modelul domeniului DNAbd din fiecare din cele două izoforme, au fost puse la dispoziție de către colaboratorii noștri Ram Ganapathi și Mahrukh Ganapathi, de la Departamentul de Farmacologie a cancerului, din cadrul Levine Cancer Institute, Carolinas Healthcare System, din Carolina de Nord, SUA.

Domeniul DNAbd este la rândul său împărțit în alte câteva subdomenii: TOPRIM (aa445-731), WHD (aa731-906), Tower și C-gate (aa906-1201); dintre care TOPRIM conține triada acidă ce chelează Mg^{2+} iar WHD conține tirozina catalitică (Wu et al., 2011). Ținând cont de organizarea topo II de la drojdie, acestea la rândul lor se mai pot împărți în 2 mari subdomenii A' și B', conectate între ele printr-o "balama" care permite un grad de libertate în plus a mișcării unui domeniu în raport cu celălalt și în raport cu ADN-ul (Classen et al., 2003).

Pentru început am generat un model de omologie pentru domeniul DNAbd folosind secvența de aminoacizi din fereastra aa449-1202 a întregii secvențe de topo II β (Grozav et al., 2011). Pentru generarea modelului am folosit aceeași procedură ca și în cazul proteinelor RAG1/2. Pe scurt am generat predicții de structură secundară, de dezordine, profiluri de sarcină, folosind programele descrise anterior. Cu ajutorul programului Phyre (Kelley et al., 2009) am găsit 3 secvențe șablon de topoizomeraza II de la drojdie (*Saccharomyces cerevisiae*), având următoarele coduri PDB: 3l4j.pdb (Schmidt et al., 2010), 1bgw.pdb (Berger et al., 1996), 1bjt.pdb (Fass et al., 1999). Alinierea secvențelor și studiul structurilor tridimensionale ale celor 3 structuri, au arătat că secvența structurii 3l4j.pdb este cea mai similară cu secvența de topo II β (47% identitate și 75% similaritate), are cea mai bună rezoluție (2.48Å) și, spre deosebire de celelalte două structuri, nu prezintă la fel de multe zone cu coordonate lipsă în structura sa 3D.

Topoizomerazele II α și II β fac parte dintr-o serie de factori esențiali în tratarea a diverse tipuri de cancer, prin punerea la punct și dezvoltarea de medicamente inhibitori (sau "otrăvuri") de topoizomeraza la om, dar pot fi folosite și ca ținte pentru medicamente antibacteriene (Nitiss, 2009-B). Inhibitorii de topoizomeraze II umane sunt de obicei cei care interferă cu segmentul-G de ADN din tumori, acolo unde expresia topoizomerazelor este crescută (Durbecq et al., 2004; Skotheim et al., 2003). Datorită faptului că cele două izoforme sunt implicate și în procese diferite una față de cealaltă, ele sunt intens studiate pentru a înțelege mecanismul diferit de acțiune.

Colaboratorii noștri au descoperit următoarele, studiind punctul de control de decatenare: i) două mutante: Y α 640F, Y β 656F, din DNAbd afectează negativ eficiența procesului de decatenare; ii) au observat un mod de acțiune diferit în procesul de decatenare - atunci când domeniul CTD lipsește din topo II α , această izoformă decatenează ADN-ul mai eficient, în schimb eficiența decatenării în topo II β scade ușor în lipsa domeniului CTD, comparativ cu formele wild type ale celor două izoforme (Kozuki et al., 2017); iii) comparativ cu forma topo II α , topo II β wild type este mai eficientă în decatenare, însă decatenarea este cea mai eficientă atunci când ambele izoforme sunt prezente. În formele cu CTD lipsă, au fost păstrate doar

porțiunile de secvență de localizare nucleară din CTD-ul fiecărei izoforme, fiind eliminate zonele K1201-Q1453 și F1498-F1531 din topo II α , și zonele K1219-S1521 și Q1574-N1621 din topo II β (Kozuki et al., 2017).

Prin intermediul acestui studiu am încercat găsirea unor răspunsuri care să explice aceste diferențe observate experimental. Pentru început, având modelele monomerilor de topoizomerază II α și II β în domeniul DNAbd am încercat o abordare a problemei mutantelor iar pentru asta e necesară cunoașterea pozițiilor lor în spațiu, în cadrul dimerului de topoizomerază.

Modelele de dimer pentru cele două izoforme au fost generate suprapunând modelele de monomer pe fiecare monomer din structuri de dimer cristalizate, de la bacterii, iar interfața primară de dimerizare a fost remodelată folosind ca șablon cristalul 2wl2.pdb (fiind singura structură cu coordonate în acea zonă, comparativ cu lipsa coordonatelor din cristalul 3l4j.pdb de la drojdie).

La scurt timp după ce am trimis colaboratorilor noștri modelele de dimer, pentru cele două izoforme, au fost publicate structurile cristalizate ale domeniului DNAbd sub formă de dimer pentru ambele izoforme de la om - întâi structura topo II β (cod PDB - 3qx3.pdb) (Wu et al., 2011) și apoi structura topo II α (cod PDB - 4fm9.pdb) (Wendorff et al., 2012).

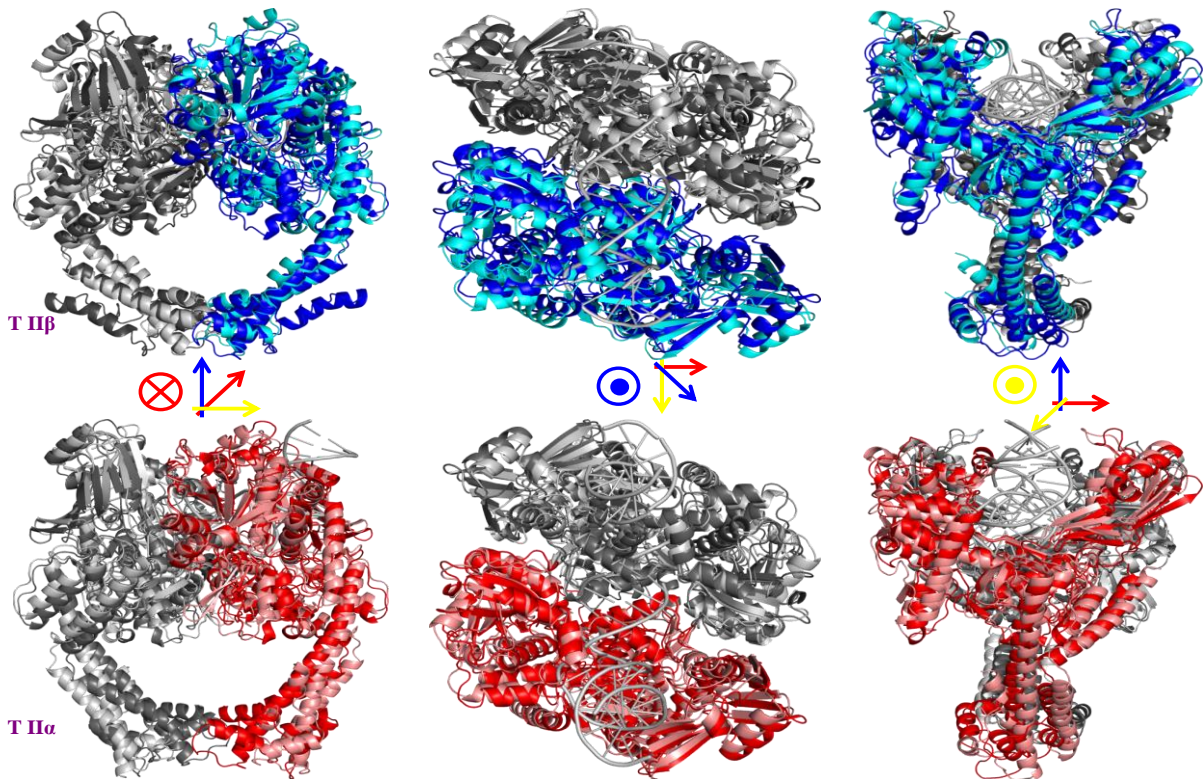


Fig. IV.4 Suprapunerea cristalelor de topo II α și topo II β cu modelele de omologie ale dimerilor

Din comparația cristalelor (culori cu nuanțe deschise în Fig. IV.4) cu modelele de omologie (culori cu nuanțe închise în Fig. IV.4) pentru dimerii ambelor izoforme rezultă o suprapunere foarte bună pe cei aproximativ 1500aa (câți sunt în componența unui dimer), astfel RMSD-ul dintre cristal și model în cazul topo II β este de 3.5Å, iar în cazul topo II α RMSD-ul este de 2.8Å. Aceste rezultate denotă faptul că modelele de dimer generate au o calitate foarte bună și prin urmare crește gradul de încredere în metodologia de modelare dezvoltată în cadrul laboratorului nostru.

De vreme ce reglarea punctului de control de decatenare în faza G2 a ciclului celular este influențată de mutarea tirozinelor Y β 656/Y α 640 la fenilalanine și pentru a încerca să explicăm răspunsul diferit al acestora, am analizat mai în detaliu subdomeniul B' din toate cristalele domeniului DNAbd publicate până în prezent.

Structurile analizate sunt următoarele: - în cazul lui topo II α există un singur cristal publicat (cod PDB - 4fm9.pdb) (Wendorff et al., 2012) în timp ce pentru topo II β există 5 structuri cristalizate (coduri PDB - 3qx3.pdb, 4g0u.pdb, 4gov.pdb, 4g0w.pdb, 4j3n.pdb) (Wu et al., 2011; Wu et al., 2013). Din analiza cristalelor rezultă, la prima vedere, că în cristalele de topo II β există două regiuni în secvența domeniului TOPRIM care nu au coordonate (acestea lipsesc în densitatea electronică de sarcină): - "R-1" (în intervalul 593-IVKA...GTST-636, care corespunde unui motiv de pliere de tip "Greek key" în topo II α) și "R-2" (în intervalul 696-LPEQ...YGTA-705, care corespunde unei bucle nestructurate în topo II α , în vecinătatea tirozinei Y α 640).

Absența acestor regiuni în densitatea electronică de sarcină a cristalelor de topo II β sugerează că ele se pot găsi în configurații multiple în această izoformă (Kozuki et al., 2017). E posibil ca această situație să apară datorită lipsei unor contacte de "ancorare" critice, în raport cu restul proteinei și/sau poate datorită unor schimbări configuraționale induse de o interacție mai puternică cu segmentul-T de ADN, în cazul lui topo II β .

O dovadă clară a lipsei unor astfel de "ancore" în topo II β este faptul că în cazul lui topo II α regiunile R-1 și R-2 sunt stabilizate printr-o legătură de H între E α 597-Y α 684. Această legătură se pierde în cazul lui topo II β , de vreme ce Y α 684 este înlocuit de F β 700. O altă posibilă ancoră în topo II α este reprezentată de legătura de H dintre S α 621 (din capătul C-terminal al regiunii R-1) cu D α 963 (din domeniul DNAbd de pe celălalt monomer), legătură ce se pierde în cazul lui topo II β deoarece serina este înlocuită cu alanina A β 637.

Folosind structura cristalizată de topo II α am măsurat toate distanțele $\leq 4.5\text{\AA}$ între toți atomii din regiunile R-1 și R-2 cu toți atomii din restul proteinei (inclusiv cu cei de pe monomerul

vecin), pe baza unui program pe care l-am scris în limbajul de programare AWK și pe care l-am dezvoltat în cadrul prezentului studiu (Kozuki et al., 2017).

Pentru a efectua aceleași măsurători și în cazul lui topo II β , am modelat prin omologie cele două regiuni R-1 și R-2. Din măsurătorile de distanță pentru cele două izoforme s-a observat o pierdere de contacte favorabile în cazul lui topo II β , comparativ cu topo II α . Astfel, în topo II α există mai multe interacții care contribuie la stabilitatea generală a regiunilor R-1 și R-2 (Kozuki et al., 2017).

Este interesant de menționat că în jurul acestor regiuni ”flexibile”, numărul de aminoacizi bazici în domeniul TOPRIM al lui topo II β este crescut în comparație cu topo II α (Fig. IV.8).

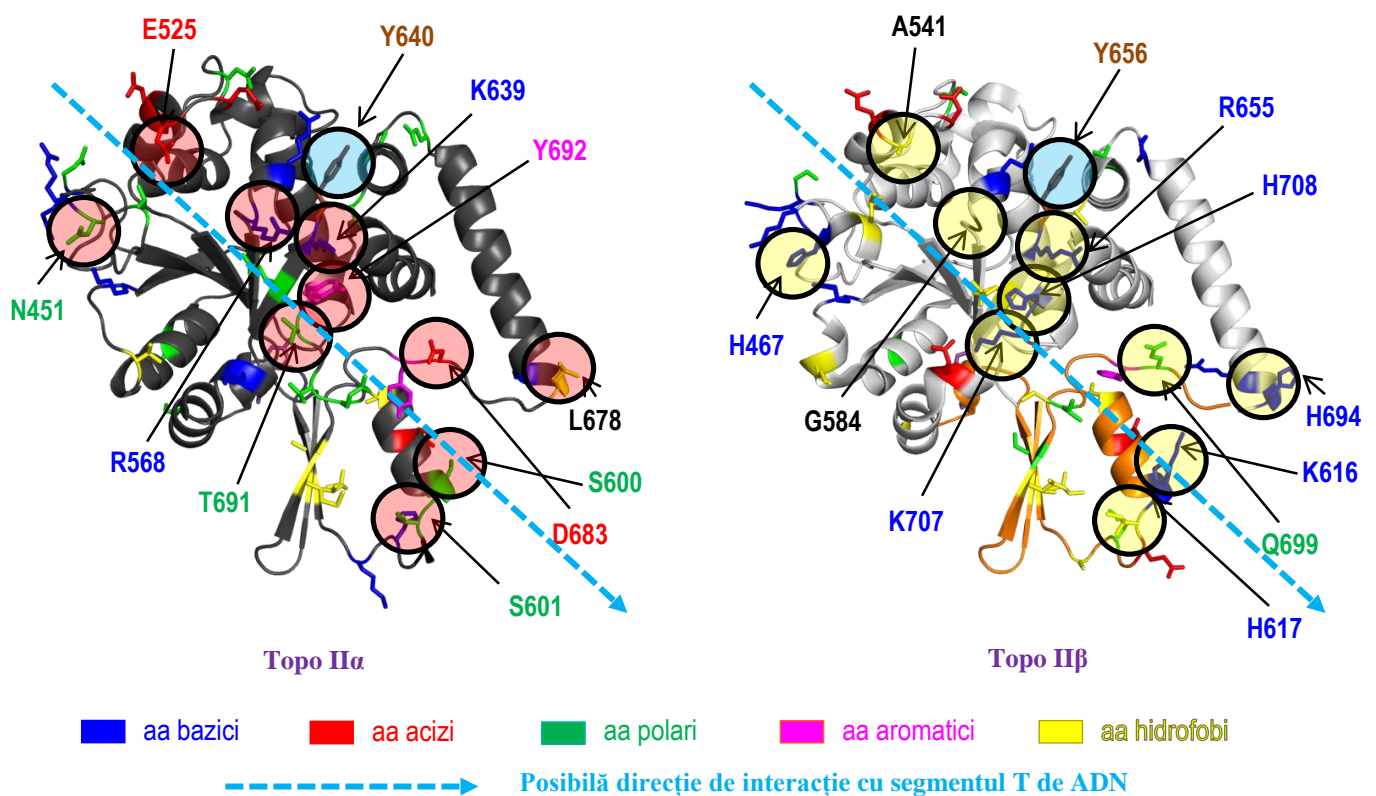


Fig. IV.8 Diferențele dintre topo II α și topo II β de la suprafața domeniului TOPRIM
(adaptare după (Kozuki et al., 2017))

Incluzând cele două regiuni R-1 și R-2 (colorate cu portocaliu în Fig. IV.8), în domeniul TOPRIM din topo II β sunt prezente 5 sarcini pozitive în plus comparativ cu topo II α , în timp ce D α 683 din topo II α este înlocuit cu Q β 699 în topo II β , după cum se poate observa în Fig. IV.8 (au fost adăugate etichete doar la aminoacizii de la suprafață cei mai diferiți între cele două izoforme). O posibilă interpretare a acestor observații este că segmentul-T de ADN ar putea interacționa și ar putea fi aliniat mai bine la domeniul TOPRIM din topo II β , în comparație cu

topo II α unde această interacție nu ar mai fi la fel de semnificativă (Fig. IV.8) (Kozuki et al., 2017). Toate aceste observații rezultate din analiza regiunii învecinate tirozinelor Y $_{\beta}$ 656/Y $_{\alpha}$ 640 ar arăta că interacția segmentului-T de ADN cu domeniul TOPRIM, prin urmare și decatenarea, ar putea fi afectate de mutațiile Y $_{\alpha}$ 640F și Y $_{\beta}$ 656F, ceea ce este în concordanță cu rezultatele experimentale (Kozuki et al., 2017).

Din analiza secvențelor domeniului CTD, analiza subdomeniului TOPRIM și a studiilor recente privind afinitatea de legare a topo II β la ADN (Gilroy și Austin, 2011) dar și a faptului că topo II α decatenează suprarăsuciri negative mult mai rapid decât pozitive în absența domeniului CTD (Seol et al., 2013) (implicând poate favorizarea capturii segmentului-T în suprarăsucire pozitivă atunci când CTD este prezent) se poate trage concluzia că segmentul-T ar putea interacționa cu regiunea TOPRIM în imediata apropiere a tirozinelor Y $_{\beta}$ 656/Y $_{\alpha}$ 640 iar mutațiile acestora la fenilalanină și/sau diferențele specifice din domeniul CTD ar putea afecta procesul de decatenare.

Motivația prezentului studiu a fost aceea de a înțelege cât mai bine mecanismul de acțiune al izoformelor de topoizomerază II de la om, și am încercat totodată identificarea proprietăților structurale ce determină acțiunea diferită a acestora în diverse condiții (ex. mutații, lipsa domeniilor CTD) prin analiza secvențelor, modelarea domeniilor DNAbd și predicțiile efectuate.

ENZIMA "DECAPPING SCAVENGER" (DcpS)

Nu în ultimul rând, în cadrul acestei teze am studiat și enzima DcpS de la *Caenorhabditis elegans* (DcpS_Ce). Enzimele DcpS fac parte din familia HIT de pirofosfataze ce conțin motivul HIT, crucial pentru hidroliza legăturii trifosfat-5'-5' din capătul ARNm.

Motivul pentru care a fost aleasă proteina de la *C. elegans* (în ciuda faptului că se cunoaște structura omologului uman) pentru a o studia, este acela de a dezvolta o serie de inhibitori de DcpS care să fie folosiți pentru tratarea unor boli în care o scădere a activității DcpS duce la reducerea patologiei (ex. atrofie musculară spinală) (Singh et al., 2008) dar și ca agenți terapeutici într-o serie de cancere. În plus, *C.elegans* este unul dintre cele mai studiate, și mai ușor de testat, organisme însă prea puține se știau despre structura DcpS de la acest organism, dar și despre afinitatea sa pentru substraturile native sau pentru alți compuși derivați.

Secvența DcpS_Ce a fost pusă la dispoziție de către colaboratorii noștri de la Departamentul de Biofizică al Institutului de Fizică experimentală, Facultatea de Fizică, Universitatea din Varșovia, Polonia; care au testat specificitatea de substrat a unei serii de compuși analogi

capetelor dinucleotidice, cu modificări în prima $m^7\text{Guo}$ (N-metilguanozină) sau în a doua nucleozidă (Wypijewska del Nogal et al, 2013). Ei au observat experimental că aceste modificări nu afectează degradarea structurilor de pe capete și mai mult decât atât au observat impactul, pe care diverse grupări funcționale, îl au asupra specificității de substrat și vitezei de hidroliză. Bazele structurale și mecanismele ce determină selectivitatea acestor enzime pentru diverși compuși erau însă necunoscute, aceasta fiind motivația prezentului studiu. Analogii studiați au la bază structura $m^7\text{GpppG}$ care a suferit o serie de modificări în special în zona punții de fosfați; și diferă astfel prin dimensiune, electronegativitate a substituentului, etc. Pentru fiecare analog au fost măsurate experimental, de către colaboratorii noștri, constante de asociere la echilibru (K_{AS}) și energia liberă Gibbs de legare (ΔG^0) pentru enzima *wild type* (WT), fără a-i afecta afinitatea nativă de legare (Wypijewska et al., 2010).

Pentru a înțelege mai bine modul de legare al analogilor la situsul activ dar și pentru a scoate în evidență mecanismul prin care unii compuși interacționează cu aminoacizii din situsul activ, am generat un model de omologie al enzimei DcpS_Ce și ulterior am efectuat experimente de "docking" molecular folosind o serie de compuși. Enzima DcpS este activă sub formă de homodimer.

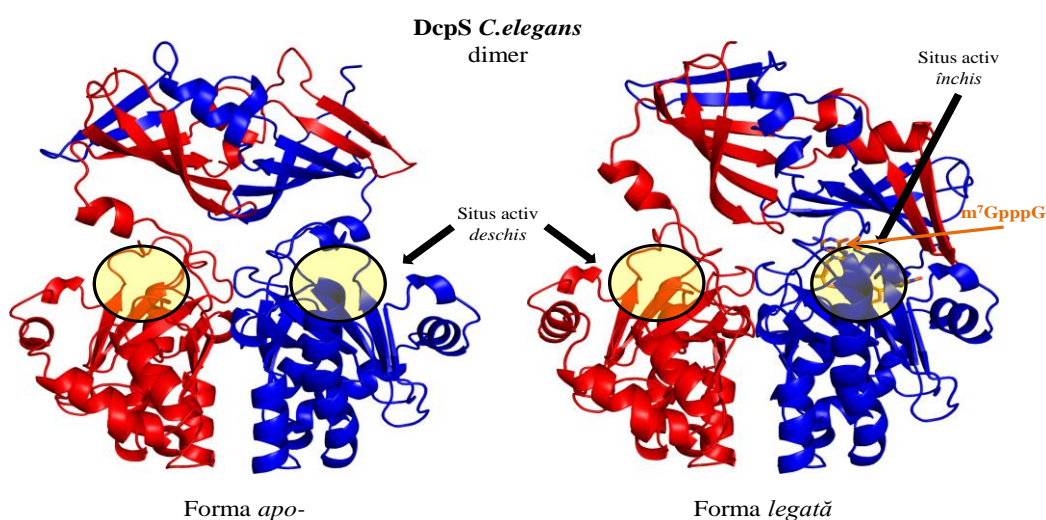


Fig. V.2 Modele DcpS_Ce

Am generat modele tridimensionale ale enzimei DcpS_Ce, în formă *legată* (conformație asimetrică a monomerilor, având un situs ocupat și unul liber pe părțile opuse ale dimerului) de structura de capăt a ARNm și în formă *apo* (inactivă sau ne-legată, cu simetrie între monomeri), deoarece apare o schimbare a conformației domeniilor.

Pentru a înțelege mai bine modul de legare a ligandului și conformațiile din situsul activ al DcpS de la *C. elegans*, am efectuat studii de docking al DcpS (Ce) cu câteva clase de liganzi

derivați din m^7GpppG : - mononucleotide cu număr variat de grupări fosfat (m^7GMP , m^7GDP și m^7GTP); - compuși de tip ARCA ($m_2^{7,2'-O}GpppG$ și $m_2^{7,3'-O}GpppG$); - analogi conținând modificări de tip punte în situsul de tăiere al fosfaților (m^7GpCH_2ppG și $m^7GpNHppG$) (Fig. V.3) (Wypijewska del Nogal et al, 2013).

Experimentele de docking între DcpS și compușii analogi au fost efectuate folosind programul AUTODOCK 4.2 din cadrul suitei AutoDock Tools 1.5.4 (Morris et al., 2009). Pentru simplitate, am luat în considerare doar situsul activ în conformație închisă din enzima DcpS de la *C. elegans*.

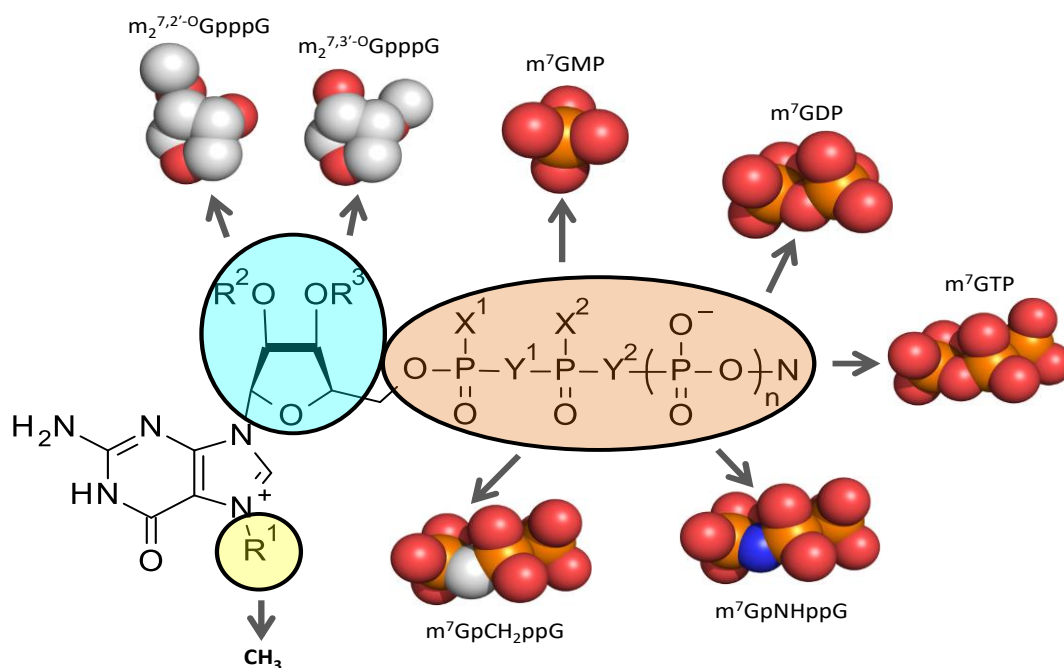


Fig. V.3 Structurile liganzilor analogi m^7GpppG folosiți în experimentele de docking, cu modificări în zona ribozei și a fosfaților (adaptare după (Wypijewska del Nogal et al., 2013))

Structurile analoage le-am generat pornind de la structura m^7GpppG din situsul în formă închisă al cristalului 1st0.pdb de la *H. sapiens*, folosind modulul *Builder* din cadrul programului PyMol iar geometria fiecărui ligand a fost optimizată cu ajutorul programului Avogadro 1.0.0 (Hanwell et al., 2012). Din clasificarea conformațiilor, obținute prin docking pentru fiecare ligand, în funcție de valorile cele mai mici ale energiei libere a rezultat că ierarhia din experimentele de docking este similară cu cea din măsurătorile experimentale.

Concluzii Finale

Pe tema Recombinării V(D)J, studii realizate de DBSB în colaborare cu Departamentul de Imunobiologie al Universității din Yale, se pot trage următoarele concluzii:

- Prin modelarea proteinelor RAG1/2 și a fragmentelor de ADN 12/23RSS am reușit explorarea mai în detaliu și o bună înțelegere a mecanismelor structurale de acțiune a recombinării somatice (V(D)J).
- Pe baza modelării fragmentelor de ADN 12/23RSS am reușit punerea la punct a unei metode de tip *coarse-grain* de îndoire a ADN-ului, metodă ce poate fi folosită în viitoarele proiecte de modelare.
- Procedura de îndoire a ADN-ului este una de încredere, ceea ce rezultă și din validarea experimentală (ex. noi experimente de FRET pe baza structurii prezise).
- În cadrul modelării proteinei RAG2 am dezvoltat un program (scris în limbajul de programare AWK) care ajută la analiza mai ușoară a unui volum mare de date de traiectorie de simulare de dinamică moleculară, efectuând o analiză detaliată a unghiurilor Phi și Psi și obținând distribuții ale acestor unghiuri de torsiune.
- Procedurile de docking care stau la baza formării complexului RAG1/2-12/23RSS, dar și informațiile obținute din analiza și modelarea tuturor acestor structuri, au dus la asamblarea ADN-ului (raportat la heterotetramerul RAG1/2) foarte similară cu cea observată prin cristalografie de raze X.

Pe tema rolului topoizomerazelor II α /II β în decatenare, studii realizate de DBSB în colaborare cu Institutul de Cancer al Carolina Healthcare System, se pot trage următoarele concluzii:

- Modelarea celor două izoforme de topoizomerază II de la om a ajutat la o mai bună înțelegere a mecanismului de acțiune al acestora în decatenarea ADN-ului.
- Analiza în detaliu a secvențelor celor două izoforme, predicțiile de modificări posttranslaționale și analiza intensă a literaturii au dus la obținerea de modele foarte similare cu cele obținute experimental.
- Pe baza unor programe de calcul de distanțe (scrise în limbajul de programare AWK) am reușit analiza în detaliu a structurilor de topoizomeraze, și a interacțiilor cheie care au rol în stabilitatea unor regiuni din domeniul DNAbd al acestora.
- Modelele de acțiune sugerate vin în sprijinul rezultatelor experimentale obținute de colaboratorii noștri.

Pe tema interacțiilor inhibitorilor enzimei DcpS de la *C. elegans*, studii realizate de DBSB în colaborare cu Facultatea de Fizică a Universității din Varșovia, se pot trage următoarele concluzii:

- Pe baza modelului de DcpS de la *C. elegans* am reușit punerea la punct a unei analize de compuși prin metode de docking.
- Rezultatele, susținute de datele obținute experimental de către colaboratorii noștri, ajută la înțelegerea mecanismelor funcționale ale degradării ARNm, precum și la dezvoltarea de noi agenți terapeutici.

În urma tuturor acestor colaborări în care am reușit punerea la punct a unor metode sau îmbunătățirea altora, în care am dezvoltat instrumente informatice, se pot trage următoarele concluzii:

- Procedurile de modelare prin omologie și omologie îndepărtată descrise în cadrul acestei teze de doctorat au dus la obținerea de modele de foarte bună calitate care ulterior au fost validate de structuri tridimensionale obținute prin metode experimentale precum cristalografia de raze X.
- Pentru a ușura analiza structurilor tridimensionale ale biomoleculelor, a contactelor între acestea sau a contactelor între aminoacizii dintr-o anumită regiune dintr-o proteină, a traiectoriilor de dinamică moleculară, a selecției celor mai bune frame-uri dintr-o simulare, pe tot parcursul desfășurării prezentei teze de doctorat am dezvoltat o serie de mici programe/scripturi care să fie utile și în viitoarele proiecte desfășurate în cadrul laboratorului
- Tot în perioada desfășurării prezentei teze de doctorat am dezvoltat programe de analiză a bazelor de date de secvențe (de la diverse specii); în speță, am efectuat diverse analize statistice ale posibilelor situsuri de N-glicozilare.
- Prezenta teză de doctorat, pe lângă rezultatele observate pe fiecare direcție studiată, a dus la o mai bună înțelegere a interacțiunilor dintre proteine și acizi nucleici, și a implicațiilor pe care acestea le au. Mai mult decât atât, toate aceste informații dobândite mă vor ajuta în viitoarele proiecte de cercetare.

Listă de lucrări și participări la conferințe

Lucrări publicate:

Articole și capitole de carte publicate în calitate de autor principal:

1. Kozuki T*, Chikamori K*, Surleac MD*, Micluta MA, Petrescu AJ, Norris EJ, Elson P, Hoeltge GA, Grabowski DR, Porter ACG, Ganapathi RN, Ganapathi MK. "*Roles of the C-*

*terminal domains of topoisomerase II α and topoisomerase II β in regulation of the decatenation checkpoint." **Nucleic Acids Res.** **45(10)**, 5995-6010 (2017) [[PMID: 28472494](#)]
IF: 10.162; AI: 3.5*

Kozuki T*, Chikamori K*, **Surleac MD***: autori cu contribuții egale

2. **Surleac MD**, Spiridon LN, Tacutu R, Milac AL, Petrescu SM, Petrescu AJ, "The Structural Assessment of Glycosylation Sites Database – SAGS – An Overall View on N-Glycosylation.", **Glycosylation**, Chap. 1, 1-22 (2012). [[DOI: 10.5772/51690](#)]

Articole publicate în calitate de autor colaborator:

3. Ciubotaru M, **Surleac MD**, Metskas LA, Koo P, Rhoades E, Petrescu AJ, Schatz DG, "The architecture of the 12RSS in V(D)J recombination signal and synaptic complexes", **Nucleic Acids Research**; 43(2):917-931, 2015. [[PMID: 25550426](#)]

IF: 9.202; AI: 3.5

4. Ciubotaru M, Trexler AJ, Spiridon LN, **Surleac MD**, Rhoades E, Petrescu AJ and Schatz DG, "RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complex", **Nucleic Acids Research**; 41(4):2437-2454, 2013. [[PMID: 23293004](#)]

IF: 8.808; AI: 3.4

5. Zhang YH, Shetty K, **Surleac MD**, Petrescu AJ, Schatz DG, "Mapping and Quantitation of the Interaction Between the Recombination Activating Gene Proteins RAG1 and RAG2", **J Biol Chem**; 290(19):11802-17, 2015. [[PMID: 25745109](#)]

IF: 4.258; AI: 1.7

6. Anna Wypijewska del Nogal, **Marius D. Surleac**, Joanna Kowalska, Maciej Lukaszewicz, Jacek Jemielity, Martin Bisailon, Edward Darzynkiewicz, Adina L. Milac, Elzbieta Bojarska, "Analysis of Decapping Scavenger (DcpS)-cap complex using modified cap analogs reveals molecular determinants for efficient cap binding", **FEBS Journal**; Volume 280, Issue 24, pages 6508–6527, 2013. [[PMID: 24119043](#)]

IF – 3.986; AI - 1.3

7. Mihai Ciubotaru, **Marius Surleac**, Mihaela G. Mușat, Andreea M. Rusu, Elena Ioniță, Paul C. C. Albu, „DNA bending in the synaptic complex in V(D)J recombination: turning an

ancestral transpososome upside down”, **DISCOVERIES**, Vol. 2, No. 1, P.1-15, 2014. [DOI: [10.15190/d.2014.5](https://doi.org/10.15190/d.2014.5)]

Prezentări la conferințe:

Expuneri orale și premii

- ”*Mass Spectrometry interactomics of topoisomerase II α and II β involved in human carcinoma cell lines*”, IXth Symposium "Acad. Nicolae Cajal", 13-14 Mai 2014, București, România - (**Marius D. Surleac**, Cristian Munteanu, Ram Ganapathi, Mahrukh Ganapathi, Andrei J. Petrescu). Distins cu Premiul Herbert Berler-Barbu pentru cea mai bună lucrare a unui tânăr cercetător, 2014, din partea Fundației “Academician Nicolae Cajal”.

- ”*Modeling protein-DNA interactions with experimental constraints. A study case on RAG proteins and the paired complex formation*”, XIth Symposium "Acad. Nicolae Cajal", 17-19 Martie 2016, București, România - (Petrescu A-J, Surleac MD, Spiridon LN, Ciubotaru M, Schatz D).

- ”*Computationally guided research in molecular life science at IBAR*”, 25 Years of Promoting Molecular Life Sciences in Romania - International Conference of the Romanian Society of Biochemistry and Molecular Biology, 17-18 Septembrie 2015, București, România - (Andrei J. Petrescu, Adina Milac, Marius Micluța, Laurențiu Spiridon, Marius Surleac, O. Căldăraru, Cristian V. Munteanu).

- ”*Modelling Structures in the Twilight Zone and Beyond. Lessons from Resistance and Effector Gene Families*”, COST FA 1208 Workshop: Structure-guided Investigation of Effector Function, Action and Recognition, 10-12 Septembrie 2014, București, România - (Spiridon LN, Milac AL, Surleac MD, Căldăraru O, Petrescu AJ).

- ”*Identifying interactors of human topoisomerase II α and II β through combined bioinformatics and Mass Spectrometry*”, The Annual International Conference of the RSBMB, 5-6 Iunie 2014, Băile Felix, Romania - (Marius D. Surleac, Cristian Munteanu, Ram Ganapathi, Mahrukh Ganapathi, Andrei J. Petrescu).

Postere și premii

- ”*Structural insights into the functional divergence of human Topoisomerase II α and II β on the decatenation checkpoint*”, The Annual International Conference of the Romanian Society for Biochemistry & Molecular Biology, 8-9 Iunie 2017, Timișoara, România - (**Surleac D. Marius**, Ganapathi Mahrukh, Ganapathi Ram, Petrescu J. Andrei). Distins cu Premiul I,

Secțiunea Correlation Structure–Properties in Biological Models – Drug Design, din partea Societății Române de Biochimie și Biologie Moleculară (SRBBM).

- ”Modeling the bent PC-23RSS DNA based on FRET data and molecular dynamics simulations” și ”Fluorescent quantification of Gibbs free energy of cap binding to DcpS reveals DcpS-cap interactions”, 9th EBSA European Biophysics Congress, 13-17 Iulie 2013, Lisabona, Portugalia - (**M. D. Surleac**, L. N. Spiridon, M. Ciubotaru, D. G. Schatz, A.-J. Petrescu) și respectiv (A. Wypijewska, M. D. Surleac, J. Kowalska, M. Lukaszewicz, J. Jemielity, M. Bisailon, R. E. Davis, E. Darzynkiewicz, A. L. Milac, E. Bojarska). Pentru participarea la această conferință am obținut o bursă de student din partea Societății Europene de Biofizică (EBSA).

- ”Modelling of the human Topoisomerase II dimer”, 2011 International Conference of RSBMB, 28-30 Septembrie 2011, Craiova, România - (**Marius D. Surleac**, Laurențiu N. Spiridon, Adrian G. Grozav, Ram Ganapathi, Andrei-Jose Petrescu). Distins cu Premiul pentru cel mai bun poster, din partea Societății Române de Biochimie și Biologie Moleculară (SRBBM).

- ”De novo Peptide Design for Enhanced Heavy Metal Accumulation”, The Annual International Conference of the Romanian Society for Biochemistry & Molecular Biology, 8-9 Iunie 2017, Timișoara, România - (Martin Eliza. C., Caldararu Octav, Ruta L. Lavinia, Ghenea Simona, Surleac D. Marius, Spiridon Laurentiu, Milac Adina, Farcasanu C. Ileana, Petrescu J. Andrei).

- ”Modelling protein-DNA complexes using FRET constraints”, COST FA 1208 Workshop: Structure-guided Investigation of Effector Function, Action and Recognition, 10-12 Septembrie 2014, București, România - (Surleac MD, Spiridon LN, Ciubotaru M, Schatz DG, Petrescu AJ).

- “The new model of mRNA degradation based on the recent advances in the DcpS enzyme specificity towards m⁷GDP”, 1st Congress of the Polish Biochemistry, Cell Biology, Biophysics and Bioinformatics BIO 2014, 9-12 Septembrie 2014, Varșovia, Polonia - (Anna Wypijewska del Nogal, Marius D. Surleac, Joanna Kowalska, Maciej Lukaszewicz, Janusz Stepinski, Richard E. Davis, Martin Bisailon, Jacek Jemielity, Edward Darzynkiewicz, Adina L. Milac and Elzbieta Bojarska).

- ”Structure-affinity relationship (SAFIR) for the mRNA cap analogs binding to C. elegans DcpS enzyme”, XIVth Congress of the Spanish Biophysical Society (SBE 2014), 11-13 Iunie 2014, Alcalá de Henares, Spania - (Anna Wypijewska del Nogal, Marius D. Surleac, Joanna Kowalska, Maciej Lukaszewicz, Jacek Jemielity, Martin Bisailon, Edward Darzynkiewicz, Adina L. Milac and Elzbieta Bojarska).

- "Role of EDEM3 in ERAD. one step at a time", The Annual International Conference of the RSBMB, 5-6 Iunie 2014, Băile Felix, România - (Cristian Marian Butnaru, Mărioara Chirițoiu, Marius Surleac, Andrei J. Petrescu, Ștefana M. Petrescu).
- "Molecular modelling and docking offer structural insights into mRNA cap binding by *C. elegans* DcpS scavenger decapping enzyme" și "Structural Analysis of Wheat Pm3 Disease Resistance Protein", 2012 International Conference of RSBMB, 13-14 Septembrie 2012, București, România - (Marius Surleac, Anna Wypijewska, Jacek Jemielity, Elzbieta Bojarska, Edward Darzynkiewicz, Andrei-Jose Petrescu, Adina-Luminița Milac) și respectiv (Laurențiu N. Spiridon, Marius Surleac, Andrei J. Petrescu).
- "An update on SAGS, the Structural Assessment of Glycosylation Sites database", 21st International Symposium on Glycoconjugates, 21-26 August 2011, Viena, Austria - (Petrescu A-J., Spiridon L., Tăcutu R., Milac A., Surleac M.).
- "Modelling the bend of DNA 23- & 12-RSS with experimental constraints", "Modelling, MD simulation and in-silico mutation of cecropin P for optimizing the interaction with tumor cell membranes" și "Modelling and MD simulation of CC domains of some R proteins", Annual International Conference of the RSBMB, 23-24 Septembrie 2010, București, România - (Marius D. Surleac, Laurențiu N. Spiridon, Mihai Ciubotaru, David Schatz, A-J Petrescu), (Marius A. Micluță, Cătălina A. Nenu, Marius D. Surleac, Laurențiu N. Spiridon, A-J Petrescu) și respectiv (Laurențiu N. Spiridon, Marius A. Micluță, Marius D. Surleac, A-J Petrescu).

Referințe selectate:

- CHEN, N., WALSH, M.A., LIU, Y., PARKER, R., SONG, H. 2005. Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J Mol Biol*, 347(4):707-18.
- CIUBOTARU, M., KRIATCHKO, A.N., SWANSON, P.C., BRIGHT, F.V., SCHATZ, D.G. 2007. Fluorescence resonance energy transfer analysis of recombination signal sequence configuration in the RAG1/2 synaptic complex. *Mol Cell Biol*, 27(13):4745-58.
- CORNEO, B., MOSHOUS, D., CALLEBAUT, I., DE CHASSEVAL, R., FISCHER, A., DE VILLARTAY, J.P. 2000. Three-dimensional clustering of human RAG2 gene mutations in severe combined immune deficiency. *J Biol Chem*, 275(17):12672-5.
- GILROY, K.L., AUSTIN, C.A. 2011. The impact of the C-terminal domain on the interaction of human DNA topoisomerase II α and β with DNA. *PLoS One*, 6(2):e14693.

GROZAV, A.G., WILLARD, B.B., KOZUKI, T., CHIKAMORI, K., MICLUTA, M.A., PETRESCU, A.J., KINTER, M., GANAPATHI, R., GANAPATHI, M.K. 2011. Tyrosine 656 in topoisomerase II β is important for the catalytic activity of the enzyme: Identification based on artifactual +80-Da modification at this site. *Proteomics*, 11(5):829-42.

GRUNDY, G.J., RAMÓN-MAIQUES, S., DIMITRIADIS, E.K., KOTOVA, S., BIERTÜMPFEL, C., HEYMANN, J.B., STEVEN, A.C., GELLERT, M., YANG, W. 2009. Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. *Mol Cell*, 35(2):217-27.

HUANG, S., TAO, X., YUAN, S., ZHANG, Y., LI, P., BEILINSON, H.A., ZHANG, Y., YU, W., PONTAROTTI, P., ESCRIVA, H., LE PETILLON, Y., LIU, X., CHEN, S., SCHATZ, D.G., XU, A. 2016. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell*, 166(1):102-14.

KAWANO, S., KATO, Y., OKADA, N., SANO, K., TSUTSUI, K., TSUTSUI, K.M., IKEDA, S. 2016. DNA-binding activity of rat DNA topoisomerase II α C-terminal domain contributes to efficient DNA catenation in vitro. *J Biochem*, 159(3):363-9.

KIM, M.S., LAPKOUSKI, M., YANG, W., GELLERT, M. 2015. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*, 518(7540):507-11.

MCCLENDON, A.K., GENTRY, A.C., DICKEY, J.S., BRINCH, M., BENDSEN, S., ANDERSEN, A.H. & OSHEROFF, N. 2008. Bimodal recognition of DNA geometry by human topoisomerase II alpha: preferential relaxation of positively supercoiled DNA requires elements in the C-terminal domain. *Biochemistry*, 47, 13169-13178.

MILAC, A.L., PETRESCU, A.J. 2001. Protein Structure Prediction. (I) Homology Modelling. *Rom. J. Biochem.*, 38 (2), 249 -275.

NITISS, J.L. 2009-A. DNA topoisomerase II and its growing repertoire of biological functions. *Nat Rev Cancer*, 9(5):327-37.

PAPILLON, J., MENETRET, J.F., BATISSE, C., HELYE, R., SCHULTZ, P., POTIER, N. AND LAMOUR, V. 2013. Structural insight into negative DNA supercoiling by DNA gyrase, a bacterial type 2A DNA topoisomerase. *Nucleic acids research*, 41, 7815-7827.

PAWLOWSKI, M., BOGDANOWICZ, A. & BUJNICKI, J.M. 2013. QA-RecombineIT: a server for quality assessment and recombination of protein models. *Nucleic Acids Res*, 41, W389–W397.

PETRESCU, A.J. 1999. Computer simulation in molecular biology. (I). basic concepts and algorithms. *Rev. Roum. Biochim.*, 36, 1-2, 115-142.

RU, H., CHAMBERS, M.G., FU, T.M., TONG, A.B., LIAO, M., WU, H. 2015. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*, 163(5):1138-52.

SATHYAPRIYA, R., VIJAYABASKAR, M.S. & VISHVESHWARA, S. 2008. Insights into protein-DNA interactions through structure network analysis. *PLoS computational biology*, 4, e1000170.

SCHMIDT, B.H., BURGIN, A.B., DEWEESE, J.E., OSHEROFF, N., BERGER, J.M. 2010. A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. *Nature*, 465(7298):641-4.

SINGH, J., SALCIUS, M., LIU, S.W., STAKER, B.L., MISHRA, R., THURMOND, J., MICHAUD, G., MATTOON, D.R., PRINTEN, J., CHRISTENSEN, J., BJORNSSON, J.M., POLLOK, B.A., KILEDJIAN, M., STEWART, L., JARECKI, J., GURNEY, M.E. 2008. DcpS as a therapeutic target for spinal muscular atrophy. *ACS Chem Biol*, 3(11):711-22.

SWANSON, P.C., KUMAR, S., RAVAL, P. 2009. Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes. *Adv Exp Med Biol*, 650:1-15.

VAN DIJK, E., LE HIR, H., SÉRAPHIN, B. 2003. DcpS can act in the 5'-3' mRNA decay pathway in addition to the 3'-5' pathway. *Proc Natl Acad Sci*, 100(21):12081-6.

WENDORFF, T.J., SCHMIDT, B.H., HESLOP, P., AUSTIN, C.A., BERGER, J.M. 2012. The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. *J Mol Biol*, 424(3-4):109-24.

WYPIJEWSKA, A., BOJARSKA, E., STEPINSKI, J., JANKOWSKA-ANYSZKA, M., JEMIELITY, J., DAVIS, R.E., DARZYNKIEWICZ, E. 2010. Structural requirements for *Caenorhabditis elegans* DcpS substrates based on fluorescence and HPLC enzyme kinetic studies. *FEBS J*, 277(14):3003-13.

YIN, F.F., BAILEY, S., INNIS, C.A., CIUBOTARU, M., KAMTEKAR, S., STEITZ, T.A., SCHATZ, D.G. 2009. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol*, 16(5):499-508.