SCIENTIFIC REPORT - Stage II – 2022 -

" *Method Development (Part II), Software Development (Part II), User Accessibility. In Silico Experiment Assistance (Part I).*"

Project: PN-III-P4-ID-PCE-2020-2444

Scientific Description

Following objectives were realized:
(a2) Non-equilibrium Hamiltonian Monte Carlo
(b1) Molecular configuration externalization
(b2) Robosample refactorization for speed.
(b3 Python interface for user accesibility
(c1.1) NS2-NS3 precleavage sampling.
(c2.1) NBARC ATP/ADP molecular simualtion
(c2.2) NB-ARC switch mechanism study.
(3.1) Proposals for enhancing the antigenicity of YMD and EEY peptides in melanoma cells.

Deliverables Realization

Deliverables realized:
D-a2 (2.1). Mean First Passage times following non-equilibrium HMC sampling in alanine dipeptide and user accessibility.
D-b1 (2.2) Externalization of molecular configurations for other programs - Part II.
(a 2.3) Quantification of fidelity and data transmission speed following the use of externalization modules.
D-b3 (2.4.) Python Interface for Robosample.
D-c1.1 (2.5.) Free energy surface of HCV NS2-NS3.
D-c2.1 (2.6.) Free energy surface of NBARC proteins.
D-c3.1 (2.7.) Proposals for YMD and EEY mutations.

Indicatori de rezultat: 0 articole publicate

Disemination:

- Sulea TA, Martin EC, Ungureanu VG, Petrescu AJ, Spiridon L, „*Highly Efficient Exploration of Conformational Spaces using Robot Mechanics - A Helical Bundle Case Study*”, Joint 25th IUBMB, 46th FEBS, and 15th PABMB Congress, ***Talks. FEBS Open Bio,*** Volume 12, pg 32-32, 9-14 iulie, Lisabona, Portugalia;
- Șulea TA, Ungureanu VG, Martin EC, Petrescu AJ, Spiridon L, *Robosample: A Molecular Simulation Program That Combines Enhanced Sampling with High-Speed Robotics Algorithms*, ***New Trends And Strategies In The Chemistry Of Advanced Materials With Relevance In Biological Systems, Technique And Environmental Protection*** October 20-21, 2022, Timișoara, România

Act 2.1. Implementation of the Non-Equilibrium Hamiltonian Monte Carlo Method (NEHMC) - Part II.

Subactivity 2.1.1 Testing Non-Equilibrium Hamiltonian Monte Carlo with Robosample using passage times between the alpha and beta basins of alanine dipeptide.

In the theoretical context established in activity 1.2, where overcoming energy barriers through mechanical work was partially implemented, this stage focuses on directed deformation of internal coordinates for surpassing energy barriers. Unlike the previous stochastic angle perturbation, this method introduces deterministic perturbation of coordinates based on targeted transitions. The alanine dipeptide system is studied, known for its energy barrier between conformational basins: C7eq and αL. The proposed method, involving deterministic changes through a function $\mathscr{M}$, was successfully implemented and tested with three 1.5 ns simulations, showing correct probability distributions in the dipeptide's Ramachandran space (Table 1.1).

	C5	PP2	C7eq	aL
C5	4.82	10.43	11.17	2227.40
PP2	17.99	4.59	8.54	2226.65
C7eq	25.46	15.27	1.82	2219.56
aL	84.69	76.37	62.56	37.08

	C5	PP2	C7eq	aL
C5	4.59	10.21	11.56	3149.46
PP2	17.44	4.41	8.91	3146.34
C7eq	24.68	14.80	1.84	3138.91
aL	66.60	56.25	42.93	75.09

	C5	PP2	C7eq	aL
C5	4.70	10.63	11.05	2442.40
PP2	17.86	4.61	8.34	2441.62
C7eq	24.82	14.90	1.84	2434.72
aL	87.00	79.00	65.54	38.76

	C5	PP2	C7eq	aL
C5	4.63	10.52	11.23	3292.66
PP2	17.81	4.63	8.45	3290.10
C7eq	24.97	14.90	1.81	3284.24
aL	71.50	61.64	49.34	69.20

Tabel 1.1. Mean First Passage Times of alanine dipeptide in equilibrium regimes (left) and non-equilibrium regimes (right) during the simulation (top) and at the end (bottom).

Act. 2.2 Externalization of molecular configurations for other programs - Part II.

Subactivity 2.2.1 Testing externalization modules with OpenMM by quantifying fidelity and data transmission speed.In cadrul acestei activitati a fost realizat complet obiectivul (b1) externalizarea configuratiilor moleculare catre alte module de calcul.

Functional tests ensured calculation accuracy, behavior in corner cases, and assessed computational cost variations and speed on different platforms

Act. 2.3 Refactoring Robosample for increased execution speed.

Subactivity 2.3.1 Refactoring Robosample for increased execution speed.

In the previous stage, to enhance the execution speed of the potential energy and force calculation for each type of interaction, the calculation of bonded energy terms was outsourced to the

OpenMM library. Additionally, various functionalities were implemented for better parameter customization. In this stage, several new optimizations were introduced, resulting in significant progress in reducing effective execution time: RAM resource usage optimizations, compiler optimizations, and improvements in the motion integration stage.

A. RAM Resource Usage Optimizations

RAM resource usage optimizations resulted in a reduction of approximately 66%.

B. Compiler Optimizations

The GCC compiler used allows for various compilation optimizations. To generate the optimized program at the highest possible level, the previously described options must be executed in a specific order. Thus, a subroutine was created to automate all compilation tasks, streamlining the process for both users and developers. Compiler optimization improvements led to a program that performs the same calculation task 16 times faster than previous versions. Additionally, through implemented code parallelizations, the program fully utilizes all available resources on a computing machine.

C. Motion Integration Process Optimizations

Further optimizations were made within the molecular motion integration process, a stage with significant computational costs. Starting from the mechanism of outsourcing to the OpenMM library developed in the previous stage, functionalities were implemented to calculate molecular motion integration using GPU acceleration, using either the OpenCL or CUDA platform. Parallelizing the integration stage is feasible only in the unconstrained regime, a regime used in every simulation type, regardless of additional constraint types added.

As a result of all the refactoring levels, a significant improvement in effective execution time was achieved. For a medium-sized system (MLA - 1967 atoms), conducting a 100,000-step integration simulation took 324 seconds in the initial prototype form at the beginning of the project; after all optimizations, the time is reduced to just 5 seconds:

	Base prototype	+ Eneergy & Forces module optimisations	+ Motion integration optimisations
Energy & Forces	-	✓	✓
Motion integration	-	-	✓
Execution time - 100.000 steps	**2191 sec**	**324 sec**	**5 sec**
		x 6.7 faster	**x 438 faster**

Act. 2.4 Simplifying the use of Robosample through a Python interface.

Subactivity 2.4.1 Developing a Python interface for Robosample.

To streamline the use of Robosample, a Python interface, RoboPy, has been developed. RoboPy facilitates the generation of flexibility files for Robosample through a 'Flexor' class and utilizes databases tailored to each type of biomolecule (amino acids, monosaccharides, etc.). Additionally, users can add new molecular types to the corresponding database for the studied systems.

Act. 2.5 Sampling the configurational space of pre-cleaved HCV NS2-NS3 (Part I).

Subactivity 2.5.1 Estimating the conformational basins of pre-cleaved HCV NS2-NS3 and proposing key mutations for their demonstration (Part I).In cadrul acestei activitati a fost realizat complet obiectivul (c1.1) esantionarea conformatiilor proteinei NS2-NS3 precleavage.

Biological Context

Within the structure of the HCV polyprotein, ten domains are distinguished: capsid protein, envelope proteins E1 and E2, p7 protein, and the non-structural proteins (NS) NS2, NS3, NS4A, NS4B, NS5A, and NS5B. Through cleavage by host proteases (of structural domains, also called entry factors) or through self-cleavage (of non-structural proteins, also called replication factors), the HCV polyprotein ensures its entry into cells and subsequent replication.

The autocleavage of NS2/3 is essential for the stability and replication of HCV viral RNA [2]. Structurally, it forms a dimer [3], and the cleavage site, consisting of H952, E972, and C993, contains residues from both protein chains. The conformational space of the pre-cleavage NS2/3 system is challenging to explore due to the system's size and abrupt energy barriers.

Approach

The structural model of NS2 was generated based on the crystalized structure with the RCSB code 2HD0 for the cytosolic portion. At the N-terminal end, the last transmembrane domain was added using the structure with the RCSB code 2KWZ. The first 11 amino acids of NS3 were also added to this model.

For the efficient sampling of the NS2/3 dimer, a set of 4 worlds (robotic joints) was constructed. In addition to the Cartesian and torsional worlds (which include the entire NS2/3 domain), two torsional worlds were used—one allowing torsions of all regions exposed to the solvent (exposure calculated using the Shrake-Rupley algorithm, implemented in MDTraj) and one allowing torsions of the C-terminal "tail" of the system (E1009-T1036). This set of worlds enabled the capture of conformational basins shown in Figures 5.1 and 5.2.

Results

Following the described sampling, multiple conformational basins were observed for both the cleaved NS2/3 region (Fig. 5.1) and the domains around the cleavage site. Additionally, two distinct basins of the W986 amino acids from the two chains were observed, rarely involved in π–π stacking interactions.

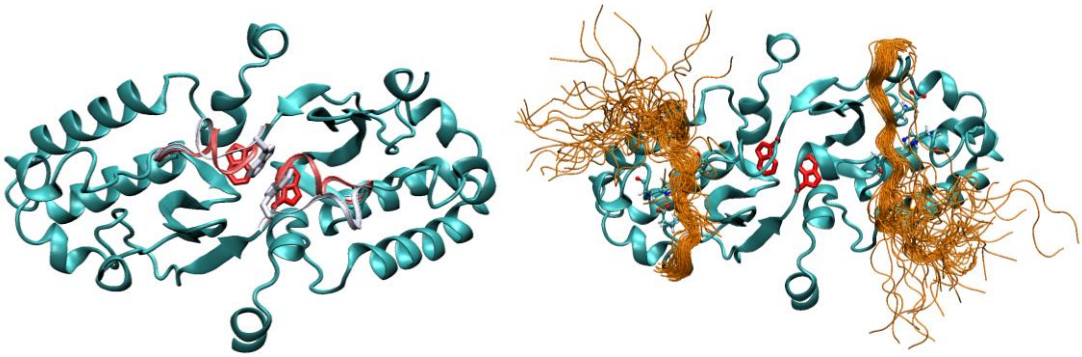

Figura. 5.1 - Structură 3D a sistemului NS2/3, cu bazinele conformaționale ale cozii C-terminale evidențiate prin reprezentare tip cilindru.

Figura. 5.2 - Structură 3D a sistemului NS2/3, cu bazinele conformaționale ale reziduurilor W986 reprezentate cu linii groase (alb - în interacție π–π stacking; roșu - fără π–π stacking.

Mutant Proposals

After analysis, four mutations were proposed for validation in activity 3.2.

Act. 2.6 Sampling the NBARC configurational space to identify basins allowing the formation of the ADP/ATP binding pocket (Part I).
Subactivity 2.6.1 Sampling the NBARC configurational space to identify basins allowing the formation of the ADP/ATP binding pocket in economically significant plant resistance proteins like tomato (Part I).

Biological Context
Common to all classes of NLR receptors present in innate plant resistance is the presence of a 'switch' domain, NBS/NBARC, consisting of 3 subdomains: NBD, ARC1, and ARC2. It functions as an 'ON/OFF' switch, with its native form in the 'closed' state, having ADP as a ligand at the intersection of the 3 subdomains. Upon detection of pathogen-associated molecules, the switch domain is activated, transitioning to the 'ON' active form. Activation involves drastic conformational changes: i) the NBD subdomain opens, allowing ii) the release of the ADP molecule, iii) the ARC2 subdomain rotates approximately 160 degrees relative to the rest of the domain, and iv) the three subdomains close in the activated 'ON' conformation, binding an ATP molecule at their intersection.

In 2019, the first 3D structure of a plant NLR receptor - the ZAR1 receptor from thale cress (Arabidopsis thaliana) - functioning as an adaptor-type NLR receptor and mediating the detection of a wide range of pathogens through a diverse set of kinase proteins that can activate this receptor, was experimentally obtained. The three-dimensional structures of this protein, in the inactive, active, and intermediate forms, were resolved through cryo-electron microscopy.

So far, this transition could not be observed in-silico due to the size of the system and the rough potential energy surface. In other words, through 'traditional' molecular sampling methods, it is unlikely to observe this transition within a reasonable simulation time.

Results

In this stage, the switch domain NBS of the ZAR1 immune receptor from Arabidopsis thaliana - a model organism in the Brassicales taxonomic order - and from tomato (Solanum lycopersicum) in the Solanales taxonomic order was studied.

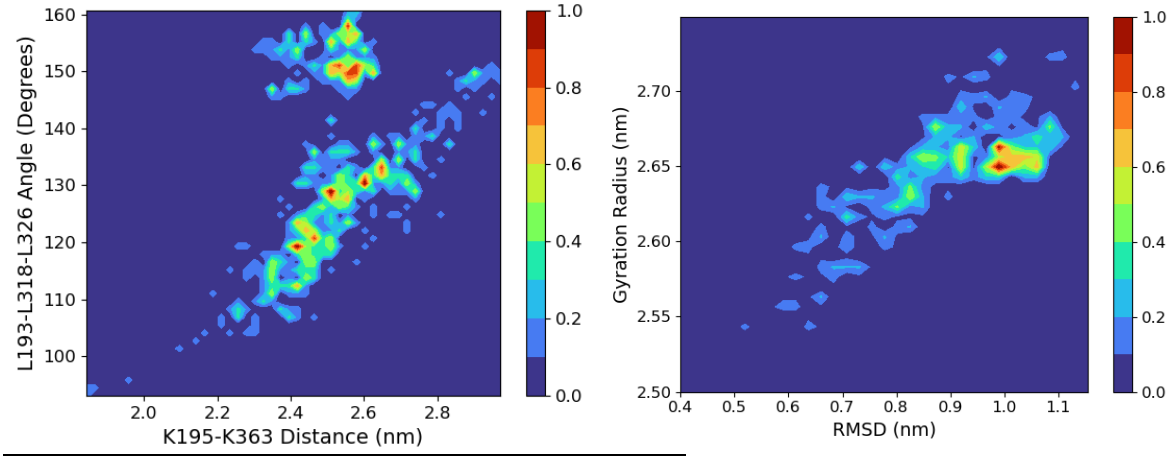

Figura. 6.7 - Distribuția de probabilitate a sistemului NB-ARC (A.Thaliana), fără primul domeniu de legătură (secvența N145-G165), în funcție de două metrici geometrice ale zonei de interacție dintre NB-ARC și ADP

.

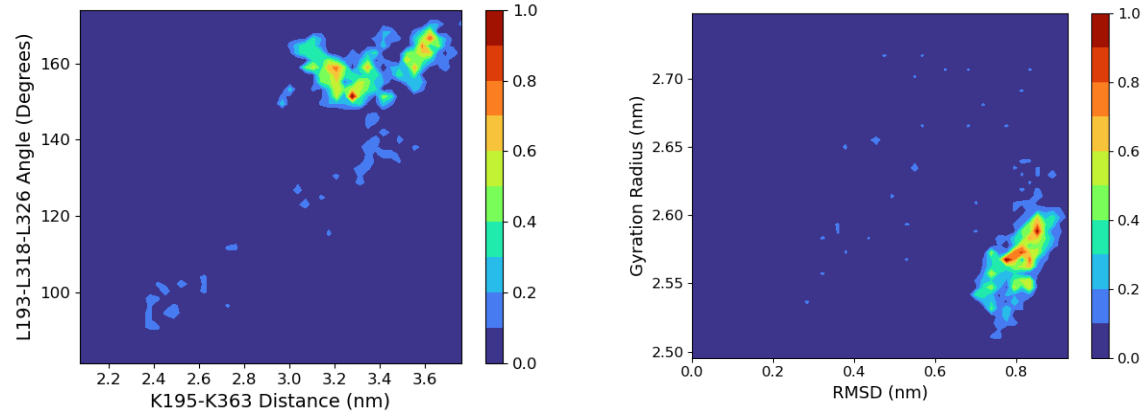

Figura. 6.9 - Distribuția de probabilitate a sistemului NB-ARC (S. Lycopersicum), fără primul domeniu de legătură (secvența N145-G165), în funcție de două metrici geometrice ale zonei de interacție dintre NB-ARC și ADP

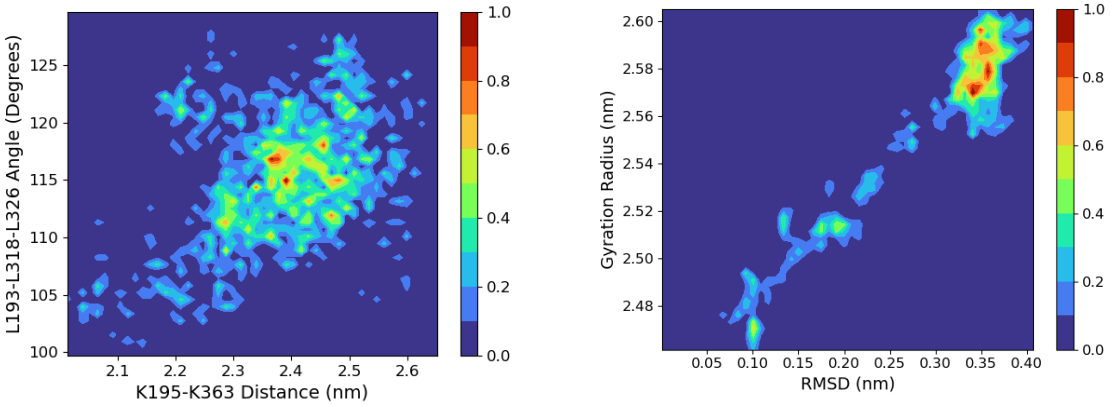

Figura. 6.10 - Distribuția de probabilitate a sistemului NB-ARC (S. Lycopersicum) în funcție de două metrici geometrice ale zonei de interacție dintre NB-ARC și ADP. Se observă o valoare mult mai mică a RMSD-ului și a distanței K195-K363.

Act. 2.7 Enhancing the Antigenicity of YMD and EEY Peptides from Tyrosinase in HLA-A0201 and HLA-B40:01 Positive Melanoma Cells (Part I).

Subactivity 2.7.1 Mutation Proposals for Increasing the Antigenicity of YMD and EEY from Tyrosinase in HLA-A0201 and HLA-B40:01 Positive Melanoma Cells (Part I).In cadrul acestei activitati a fost realizat complet obiectivul (3.1) propuneri de crestere a antigenicitatii peptidelor YMD si EEY din celulele de melanom.

Biology Context

Major Histocompatibility Complex I (MHC I) proteins are heterodimers composed of two protein chains, the α chain and the β-microglobulin chain. The α chain, with its three segments, plays distinct roles in T cell receptor interaction. The transmembrane segment α3 is responsible for non-covalent binding to the β-microglobulin chain, which interacts with the CD8 receptor on T cells. This interaction anchors MHC while the T cell evaluates the antigenicity of the peptide bound to the α1 and α2 chains, forming a groove where the peptide is captured. In tumor or virus-infected cells, mutant forms of endogenous proteins compete with host proteins for surface exposure. As 'non-self' (exogenous) peptides accumulate on the cell surface, the probability of immune system activation increases. Once activated, the cytotoxic CD8+ T cell lysates the target cell by releasing perforin and granzyme, triggering cell death. Tyrosinase, an oxidase enzyme, is a rate-limiting factor in melanin pigment synthesis and serves as a biomarker for melanoma. Investigating ways to enhance the presentation of tyrosinase epitopes on the tumor cell surface is crucial for triggering an autoimmune response against the tumor tissue. The peptide 'YMNGTMSQV' is a tyrosinase epitope corresponding to amino acids 369-377 of tyrosinase. In melanoma, poorly folded tyrosinase is degraded and highly presented on the cell surface. Mutating the constituent amino acids, such as in the 'YMDGTMSQV' (referred to as 'YMD') epitope, suggests that chemical modifications to the peptide can have immunological effects.

Approach

To study the conformational space of the YMD peptide, Hamiltonian Monte Carlo simulations were conducted using the Robosample program. The system was built with a linear structure, subjected to minimization to convergence using OpenMM v.7.7.0, and then equilibrated using Robosample. The following peptides were proposed:

YMNGTMSQV: Native form of the tyrosinase epitope Tyr369-377.
YMDGTMSQV: Modified form of the tyrosinase epitope found expressed on melanoma cell surfaces, obtained by N371 deamidation.
YMDGVMSQV: Mentioned in several articles, this form has a similar affinity to threonine homologs, but changing T373V destroys the N371 glycosylation site. Its presence on cell surfaces suggests an alternative N371 deamidation route independent of glycosylation.
YMQGTMSQV: A mutation preserving the character of N371 but preventing deamidation, as the native form is not as strongly expressed as its deamidated counterpart.
EEYNSHQSL: Native tyrosinase epitope (Tyr280-288) with affinity for HLA-B*40:01. Antigenicity relative to Tyr369-377 is unknown.
Results

Hamiltonian Monte Carlo simulations with two 'worlds'—one Cartesian and one with cylindrical joints—revealed multiple conformational transitions.

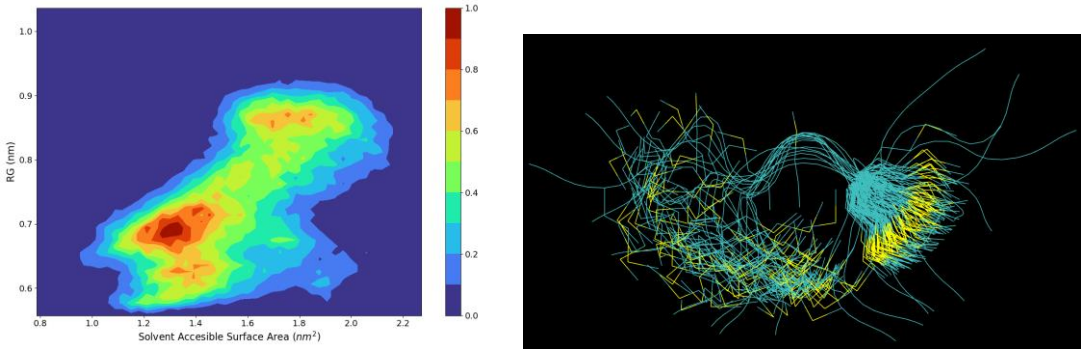

Figure 7.2 - Probability distribution of the solvent-accessible surface area (SASA) for amino acid M6 in the YMD peptide versus the overall peptide gyration radius. Distinct conformational transitions can be identified in the overlay of YMD peptide conformations.

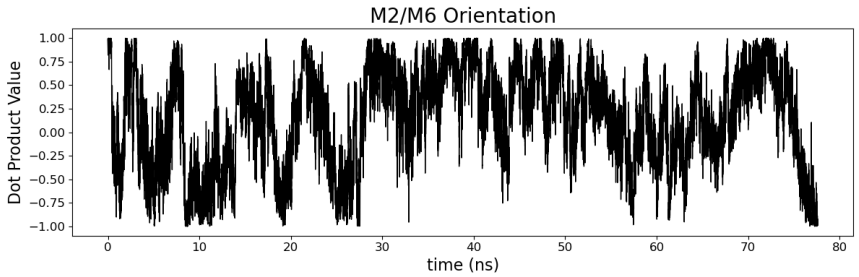

Figura 7.3 - Dinamica orientării lanțului lateral al aminoacidului M2 față de M6. Se evidențiază un număr semnificativ de tranziții între orientări paralele și antiparalele ale lanțurilor laterale.

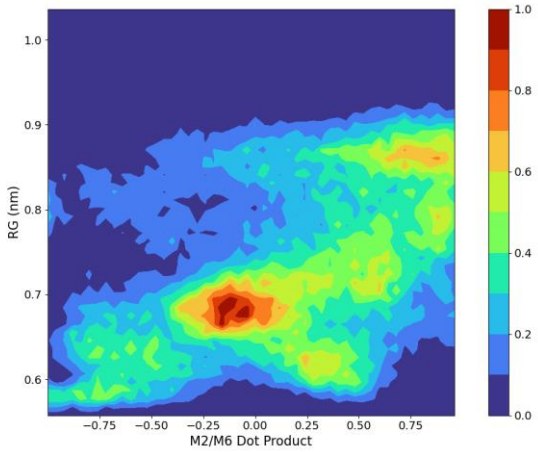

Figura 7.4 - Distribuția de probabilitate a orientării lanțurilor laterale M2/M6 în funcție de raza de girație a peptidei. Lanțurile sunt preponderent paralele în forma extinsă, dar adoptă orientări perpendiculare sau antiparalele când raza de girație scade**.**

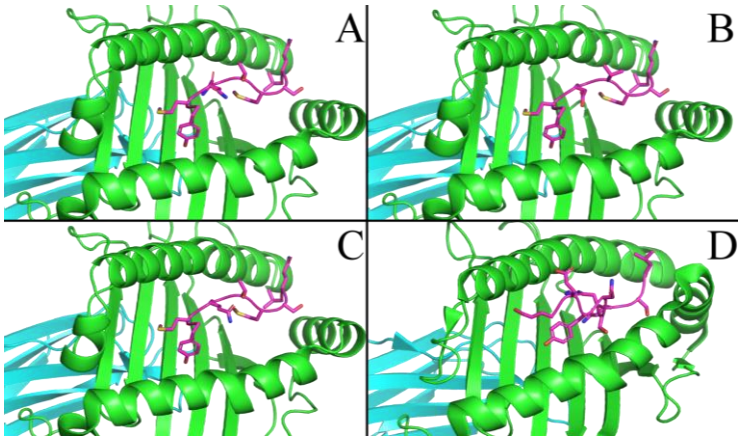

Figura 7.5 - Mutații ale epitopului YMD în situsul de legare al HLA-A:0201.
A: YMNGTMSQV; B: YMDGVMSQV; C: YMQGTMSQV; D: EEYNSHQSL

Presentation

- Sulea TA, Martin EC, Ungureanu VG, Petrescu AJ, Spiridon L, „*Highly Efficient Exploration of Conformational Spaces using Robot Mechanics - A Helical Bundle Case Study*", Joint 25th IUBMB, 46th FEBS, and 15th PABMB Congress, ***Talks. FEBS Open Bio,*** Volume 12, pg 32-32, 9-14 iulie, Lisabona, Portugalia;

- Şulea TA, Ungureanu VG, Martin EC, Petrescu AJ, Spiridon L*, Robosample: A Molecular Simulation Program That Combines Enhanced Sampling with High-Speed Robotics Algorithms,* ***New Trends And Strategies In The Chemistry Of Advanced Materials***

References

[1] – McGibbon R. T. et al., "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories", **Biophys. J.**, ***vol. 109***, issue 8, pp. 1528–1532, 2015, doi: 10.1016/j.bpj.2015.08.015
[2] – Welbourn S. et al., "Hepatitis C virus NS2/3 processing is required for NS3 stability and viral RNA replication", **J. Biol. Chem.**, ***vol. 280***, issue 33, pp29604-11, 2005, doi: 10.1074/jbc.M505019200
[3] – Lorenz I. C., Marcotrigiano J., Dentzer T. G., Rice C. M., "Structure of the catalytic domain of the hepatitis C virus NS2-3 protease", **Nature**, ***vol. 442***, pp. 831-835, 2006, doi: 10.1038/nature04975
[4] – Eastman P. et al., "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics", **PLoS Comput. Biol.**, ***vol. 13***, issue 7, pp. 1–17, 2017, doi: 10.1371/journal.pcbi.1005659
[5] – Hewitt E. W., "The MHC class I antigen presentation pathway: Strategies for viral immune evasion", **Immunology.**, ***vol. 110***, issue 2, pp. 163-9, 2003, doi: 10.1046/j.1365-2567.2003.01738.x
[7] – Boyle J. L., Haupt H. M., Stern J. B., Multhaupt H. A. B., "Tyrosinase expression in malignant melanoma, desmoplastic melanoma, and peripheral nerve tumors: An immunohistochemical study", **Arch. Pathol. Lab. Med.**, ***vol.126***, issue 7, pp. 816-22, 2002, doi: 10.5858/2002-126-0816-TEIMMD.
[8] - Altrich-VanLith M. L. et al., "Processing of a Class I-Restricted Epitope from Tyrosinase Requires Peptide N -Glycanase and the Cooperative Action of Endoplasmic Reticulum Aminopeptidase 1 and
Cytosolic Proteases", **J. Immunol.**, ***vol.177***, issue 8, pp. 5440-50, 2006, doi: 10.4049/jimmunol.177.8.5440
[9] – Ostankovitch M., Altrich-VanLith M., Robila V., Engelhard V. H., "N -Glycosylation Enhances Presentation of a MHC Class I-Restricted Epitope from Tyrosinase", **J. Immunol.**, ***vol.182***, issue 8, pp.4830-5, 2009, doi: 10.4049/jimmunol.0802902
[10] – Gloger A., Ritz D., Fugmann T., Neri D., "Mass spectrometric analysis of the HLA class I peptidome of melanoma cell lines as a promising tool for the identification of putative tumor-associated
HLA epitopes", **Cancer Immunol. Immunother.**, ***vol. 65***, issue 11, pp 1377-93, 2016, doi: 10.1007/s00262-016-1897-3.
[11] Spiridon L, Minh DDL. "*Hamiltonian Monte Carlo with Constrained Molecular Dynamics as Gibbs Sampling.*" ***J Chem Theory Comput.*** 2017 Oct 10;**13**(10):4649-4659.
[12] Spiridon L, Şulea TA, Minh DDL, Petrescu AJ. Robosample: "*Robosample: A rigid-body molecular simulation program based on robot mechanics.*" ***Biochim Biophys Acta Gen Subj.*** 2020 Aug;1864(8):129616
[13] Nilmeier JP, Crooks GE, Minh DD, Chodera JD. *Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation*. ***Proc Natl Acad Sci*** U S A. 2011 Nov 8;**108**(45):E1009-18. doi: 10.1073/pnas.1106094108. Epub 2011 Oct 24. Erratum in: Proc Natl Acad Sci U S A. 2012 Jun 12;109(24):9665. PMID: 22025687; PMCID: PMC3215031.