

Наставно-научном већу
Математичког факултета
Универзитета у Београду

Одлуком Наставно-научног већа Математичког факултета донетом на 372 седници одржаној 03.07.2020. године, именовани смо у комисију за преглед и оцену докторске дисертације под називом **Анализа заступљености поновака са применом у предвиђању Т-ћелијских епитопа** који је предат као докторска дисертација кандидата, Ане Јеловић. Након прегледа рукописа подносимо Наставно-научном већу следећи

Извештај

1 Биографија кандидата

Ана Јеловић рођена је 30.јула 1973. године у Београду. Основну школу је започела у Великој Британији, а завршила у Београду, где је завршила и гимназију. Школске 1997/1998 године дипломирала је на Математичком факултету у Београду (смер Информатика и Рачунарство) са просечном оценом 8,70. Школске 1999/2000. године на Математичком факултету уписала је магистарске студије на којима је положила све испите предвиђене планом и програмом са просечном оценом 10,00. Школске 2010/2011. године је уписала докторске академске студије на Математичком факултету, смер Информатика. Положила је све испите предвиђене планом и програмом докторских студија са просечном оценом 10,00.

Од 1999. године запослена је на Саобраћајном факултету Универзитета у Београду као асистент приправник, а од 2013. године запослена је на Саобраћајном факултету као асистент на Катедри за општу и примењену математику. До сада је учествовала у извођењу наставе из следећих предмета: Математика 1, Математика 2, Математика 3, Програмски пакети у Математици, Математички програмски пакети у саобраћају и транспорту, као и припремној настави за полагање пријемног испита на Саобраћајном факултету.

Основне области интересовања су јој истраживање података у биоинформатици.

2 Списак научних радова

Ана Јеловић је до сада била аутор два рада у међународним часописима на СЦИ листи, од којих је један самосталан, и једног рада у међународном часопису који

није на СЦИ листи. Имала је пет саопштења на научним скуповима, и коаутор је две збирке задатака.

Радови у међународним часописима са СЦИ листе

1. Ana M. Jelovic: RepeatsPlus - Program for Finding Motifs and Repeats in Data Sequences, *Journal of Bioinformatics and Computational Biology*, Vol.19, No 3, 2020, <https://doi.org/10.1142/S0219720021500104> (M23, ИФ=1.261)
2. Ana M. Jelovic, Nenad S. Mitic, Samira Eshafah, Milos V. Beljanski: Finding Statistically Significant Repeats in Nucleic Acids and Proteins, *Journal of Computational Biology*, Vol.25, No 4, 2018, <https://doi.org/10.1089/cmb.2017.0046> (M21, ИФ=3.118)

Радови у међународним часописима

1. Slavica Dozic, Ana Jelovic, Milica Kalic, Mirjana Cangalovic: Variable Neighborhood Search to solve an airline fleet sizing and fleet assignment problem, *Transportation Research Procedia*, Volume 37:258-265, 2019, <https://doi.org/10.1016/j.trpro.2018.12.191>

Саопштења на научним скуповима

1. Ana Jelović RepeatPlus-program for finding motifs and repeats in data sequences. *Bioinformatics Conference 2021*. 21. - 25. Jun, 2021, Belgrade
2. Nenad Mitic, Gordana Pavlovic-Lazetic, Sasa Malkov, Milos Beljanski, Mirjana Maljkovic, Ana Jelovic Accuracy of disorder predictors results-comparasion on DisProt DB data. *Book of Abstracts: International Multiconference on Bioinformatics of Genome Regulation and Structure / Systems Biology - BGRS/SB-2020*. Novosibirsk, 06.-10 Jul,2020
3. Ana Jelović, Nenad Mitić, Miloš Beljanski RepeatPlus-program for finding repeats in nucleic acids and proteins. *Book of Abstracts Belgrade Bioinformatics Conference 2018*. *Biologia Serbica*. 2018;40(1):110 – 110. isbn: 2334-6590, 18. - 22. Jun, 2018, Belgrade
4. Positional Biases of the Experimentally Characterized T-cell Epitopes, Mirjana D. Pavlović, Ana Jelović, Davorka Jandrlić, Nenad Mitić Positional Biases of the Experimentally Characterized T-cell Epitopes *Book of Abstracts Belgrade Bioinformatics Conference 2018*. *Biologia Serbica*. 2018;40(1):125 – 125. isbn: 2334-6590, 18. - 22. Jun, 2018, Belgrade
5. Ana Jelović, Miloš Beljanski, Nenad Mitić Filtering of repeat sequences in genomes, *Book of Abstracts Belgrade Bioinformatics Conference 2016*:62- 62. isbn: 978-86-7589-108-6, 20. – 24. Jun, 2016, Belgrade

Објављене збирке задатака

1. Т. Левајковић, К. Кукић, Д. Илић, А. Јеловић, М. Борисављевић, Н. Тирић, А. Перовић: Збирка решених задатака из Математике 1, Саобраћајни факултет Универзитета у Београду ISBN 978-86-7395-333-5, 2015.

2. С. Милорадовић, Н. Ђирић, З. Шама, М. Борисављевић, С. Младеновић, В. Стојановић, А. Јеловић, Т. Левајковић: Збирка решених испитних задатака из Математике I, Саобраћајни факултет Универзитета у Београду, ISBN 86-7395-178, 2004.

3 Предмет и садржај дисертације

Предмет докторске дисертације је анализа биолошких аминокиселинских секвенци и налажење веза између статистички значајних поновака различитих типова и позиција епитопа у циљу прецизније локализације позиције епитопа у протеинским нискама.

Централна област примене информатичких метода у имунологији је анализа континуалних кратких ниски аминокиселина које настају биолошком деградацијом протеина самог организма или страних протеина. Имуни систем организма има способност да препозна и разликује материјал из сопствених ћелија, као и да разликује да ли је то опасно или неопасно за организам. Стране супстанце које се препознају, углавном протеини, и на које имуни систем реагује да би заштитио организам, се називају антигени (Аг). Најселективнија карика имуног система је специфични имуни систем који може да се прилагођава и чија су кључна компонента Т лимфоцити (Т-ћелије), од којих потиче назив ћелијски имунитет. Т-ћелије се деле у две основне групе: Т-помоћне (Тх, од енг. Т-helper) ћелије, које помажу осталим белим крвним ћелијама у имуним процесима, и Т-цитотоксичне лимфоците (ћелије) (ЦТЛ), које директно могу да униште ћелије организма које имају једру инфициране вирусима и туморске или трансплантиране ћелије.

Ћелијски имуни одговор започиње тако што у антиген-презентујућим ћелијама (АПЦ, од енг. antigen presenting cells) долази до разградње (протеолизе) протеина на фрагменте (пептиде) и њиховог везивања за две класе МХЦ (од енг. Major Histocompatibility Complex) молекула који транспортују ове пептиде на површину АПЦ и представљају их, у виду МХЦ-пептид комплекса, Т-ћелијским рецепторима. Пептиди који буду препознати се називају епитопи. Молекули МХЦ класе 1 представљају пептиде ЦТЛ, а МХЦ класе 2, Тх ћелијама, и њихова велика полиморфност развијена је да би се препознао велики број пептидних епитопа. Са друге стране полиморфност МХЦ има за последицу разлике у имуном одговору сваког појединца на исти Аг, као и трансплантационе реакције. Познавање оба основна типа Т ћелијског имуног одговора неопходно је за успешну имунизацију и добијање вакцина против заразних болести, тумора или сузбијање алергија и аутоимуних болести.

Експериментална истраживања Т ћелијских епитопа у циљу дијагностике и терапије заразних болести и тумора, спадају у најдуготрајнија и најскупља биолошка и клиничка испитивања. Рачунарске методе, засноване на афинитету везивања континуалних ниски пептида за МХЦ молекуле, што је кључни догађај селекције Т ћелијских епитопа, су данас веома бројне и значајане су за дизајнирање синтетичких вакцина

против различитих болести. Рачунарске и математичке методе су у овом подручју кључне како би омогућиле систематично истраживање и идентификовање Т-ћелијских епитопа на великом скупу података. Познато је, међутим, да ове методе у великој мери прецењују или греше и у броју и у врсти предвиђених епитопа, делом због полиморфности МХЦ молекула (алела), за које не постоји довољан број експерименталних података који се користе у тренирању алата за предвиђање. Стога се често уводе друге врсте анализа које би имале за циљ да дефинишу молекуларну структуру интеракције између епитопа и МХЦ алела или способност епитопа да доведе до активације имуног одговора. Често се унутар протеинских ниски налазе краће подниске, настале унутрашњим дуплирањем гена, које се понављају два или више пута. Овакве подниске су означене као поновци (повнављајуће ниске или енг. "repeats"). Повнављајуће ниске представљају делове секвенци (подсеквенце) који се јављају два или више пута. У зависности од тога да ли је подсеквенца копије идентична оригиналној, или заједно са оригиналном чини палиндром, поновци могу бити директни или обрнути. Статистички значајни поновци представљају подскуп поновака секвенце за које није очекивано да ће се појавити у насумичној секвенци исте дужине. Установљавање евентуалне повезаности статистички значајних поновака и епитопа може олакшати одређивање позиције епитопа епитопа у протеинским нискама.

4 Приказ дисертације

Рукопис се састоји од 459 страница (viii+451) и има следећу структуру:

1. Увод
2. Методе за одређивање заступљености поновака
3. Базе експериментално потврђених епитопа и линеарних мотива
4. Методе истраживања података
5. Заступљеност поновака и предвиђање Т-ћелијских епитопа
6. Заступљеност поновака и предвиђање Б-ћелијских епитопа
7. Заступљеност поновака и предвиђање кратких линеарних мотива
8. Закључак

уз Резиме (на енглеском и српском језику), Садржај, Списак слика, Списак табела, Додатак, Списак литературе који се састоји од 35 библиографских јединица и Биографију кандидата.

У уводном поглављу је дат приказ основних појмова, описан предмет истраживања и циљ дисертације.

Друго поглавље садржи опис метода за одређивање заступљености поновака и мотива. Дата је формална дефиниција поновака, и приказана математичка основа одређивања

статистички значајних поновака у улазној секвенци. Приказано је филтрирање поновака и провера тачности статистички заснованих поновака на конкретним улазним подацима. У поглављу је такође дата формална дефиниција мотива, и описано налажење статистички значајних мотива. На крају поглавља је описана програмска имплементација предложених метода за одређивање (статистички значајних) поновака и мотива.

Треће поглавље садржи кратак опис база биолошких података из којих је узет материјал обрађиван у дисертацији. Описане су *IEDB* - база експериментално утврђених података о антителима и епитопима, како хуманих тако и животињских врста, и *ELM* - база кратких линеарних мотива који представљају посебну класу протеинских функционалних модула који учествују у интеракцијама протеин-протеин. За обе базе поред кратког описа наведене су и датотеке које су преузете за потребе дисертације и дат опис њиховог садржаја.

Четврто поглавље садржи кратак опис основних појмова везаних за правила придруживања (као методу истраживања података) која се користи у раду за одређивање веза епитопа и поновака.

Петом, шестом и седмом поглављу чине централни део рада. У петом поглављу описана је метода за утврђивање повезаности позиција поновака и Т-ћелијских епитопа. У протеинима који садрже експериментално потврђене епитопе одређени су статистички значајни поновци најмање дужине 2 и 3 и њихове позиције. Сваком поновку су придружене карактеристике које се односе на постојање пресека њихове позиције са позицијама епитопа, индикатори да ли позиције леве или десне компоненте поновка секу позицију епитопа, као и да ли поновак комплетно припада епитопу или епитоп припада поновку. На овај материјал су примењена правила придруживања којима је одређен скуп поновака који са поузданошћу већом од 50% пресеца позицију епитопа. Показало се да поновци најмање дужине 2 не дају задовољавајуће резултате јер је поузданост њиховог предвиђања око 20%, што је оцењено као недовољно. Из тог разлога су у приказу резултата коришћени само поновци минималне дужине 3.

Резултати су показали да постоји велики број поновака који са поузданошћу 100% пресецају позицију епитопа или припадају епитопима. Конструисан модел омогућује, уколико се такви поновци јаве у протеину, брзо и ефикасно одређивање позиције епитопа у том протеину. На скуп поновака који имају поузданост већу од 50% примењена су поново правила придруживања при чему је епитопима придружен проширени скуп карактеристика који укључује идентификацију алеле, тип епитопа, тип ћелије и хидропатију епитопа. На тај начин су добијене детаљне везе поновака (који предвиђају позиције епитопа) и карактеристика (тако предвиђених) епитопа.

Ови резултати су добијени и за директне некомплементарне и за индиректне некомплементарне (палиндромске) поновке. Урађена је и анализа покривености улазног скупа епитопа предложеном методом. Добијени резултати су, због обима, само делимично наведени у додатку на крају рада, уз могућност добијања комплетних резултата на захтев. Поглавље се завршава приказом мотив маске које обједињују пронађене поновке.

У шестом поглављу је метода описана у петом поглављу примењена на Б-ћелијске епитопе. Иако то није било покривено иницијалним предлогом теме, поступак је примењен и на ову групу епитопа због њихове важности (делови антигена за које се везује имуноглобулин или антители). Б-ћелијски епитопи су у највећем броју конформациони због чега су обрађене само њихове линеарне компоненте за оба типа поновака (директне некомплементарне и индиректне некомплементарне). Добијени резултати који омогућују предвиђање (линеарних компоненти) Б-ћелијских епитопа. У овом случају није спроведен поступак додатне карактеризације епитопа због непостојања додатних карактеристика (нпр. овај тип епитопа нема идентификацију алеле, тип, итд.). Као и у претходном поглављу, урађена је и анализа покривености улазног скупа епитопа предложеном методом, док су добијени резултати, такође због обима, само делимично наведени у додатку на крају рада, уз могућност добијања комплетних резултат на захтев.

Седмо поглавље обрађује заступљеност поновака и предвиђање кратких линеарних мотива на основу њихових позиција. Примењен је исти начин налажења веза који је детаљно описан у петом поглављу. На крају поглавља приказане су мотив маске које обједињују пронађене поновке за предвиђање кратких линеарних мотива. Комплетни резултати добијени у овом поглављу због обима нису наведени у тексту рада ни у додатку, али се могу добити на захтев од аутора.

У осмом поглављу *Закључак* је дат сумарни приказ садржаја дисертације и дискутован квалитет добијених резултата. Подршка за добијене поновке и за Т-ћелијске и за Б-ћелијске епитопе је очекивано мала, али поузданост добијених правила и мера лифт говоре о врло виском квалитету добијених резултата.

У Додатку су приказане табеле које садрже резултате за различите типове поновака и Т- и Б-ћелијске епитопе. Због обима резултата, и сам додатак је велики (преко 350 страница). Наведени су најзначајнији резултати (правила са највећим процентом поузданости) који могу бити интересанти за имунологе. У додатку Е је дат списак резултата који нису наведени у раду а који се могу добити од аутора на захтев.

5 Закључак

Рукопис **Анализа заступљености поновака са применом у предвиђању Т-ћелијских епитопа** садржи вредан научни допринос у области примене истраживања података у имунологији и биоинформатици. У раду је разматран проблем одређивања повезаности понављајућих секвенци и епитопа. Предложена је нова метода за одређивање оквирне позиције епитопа у протеинима употребом правила придруживања. На основу скупа експериментално потврђених епитопа конструисан је модел који на основу појаве конкретних поновака у протеину предвиђа локације епитопа. Добијени резултати су од великог интереса за имунологе. Конструисање оваквог модела ће омогућити истраживачима да идентификују регион у протеинској секвенци у којој се са великом поузданошћу може очекивати Т-ћелијски епитоп који је потребно експериментално

потврдити. Имајући у виду цену и дужину трајања експерименталног одређивања Т ћелијских епитопа, сужавање простора претраге улазне секвенце би могло донети значајне користи. На сличан начин, са сличним резултатима је цео поступак примењен и код Б-ћелијских епитопа и кратких линеарних мотива.

Имајући у виду претходно наведено предлажемо Наставно-научном већу Математичког факултета да рукопис **Анализа заступљености поновака са применом у предвиђању Т-ћелијских епитопа** кандидата Ане Јеловић прихвати као докторску дисертацију и одреди комисију за њену одбрану.

У Београду, 26.07.2022.

Чланови комисије за преглед и оцену

(проф. др Ненад Митић, редовни проф.)

(проф. др Саша Малков, ванредни проф.)

(др Јована Ковачевић, доцент)

(др Зоран Огњановић, научни саветник)