SVEUČILIŠTE U RIJECI

**TEHNIČKI FAKULTET**

Diplomski sveučilišni studij računarstva

Diplomski rad

**ANALIZA SLIČNOSTI PEPTIDA TEMELJENA NA LATENTNOM PROSTORU AUTOENKODERA**

Rijeka, rujan 2022. Bruno Novosel   
 0069078199

SVEUČILIŠTE U RIJECI

**TEHNIČKI FAKULTET**

Diplomski sveučilišni studij računarstva

Diplomski rad

**ANALIZA SLIČNOSTI PEPTIDA TEMELJENA NA LATENTNOM PROSTORU AUTOENKODERA**

Mentor: doc. dr. sc. Goran Mauša

Rijeka, srpanj 2020. Bruno Novosel   
 0069078199

IZJAVA

Izjavljujem da sam ovaj rad izradio samostalno, uz vodstvo i stručnu pomoć mentora doc. dr. sc. Gorana Mauše.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Bruno Novosel

ZAHVALA

Zahvaljujem mentoru doc. dr. sc. Goranu Mauši na pomoći, stručnom vođenju i velikom strpljenju kojim me vodio kroz izradu ovog rada.

Sadržaj

[1. UVOD 1](#_Toc110008986)

[2. PEPTIDI 2](#_Toc110008987)

[2.1. Što su peptidi? 2](#_Toc110008988)

[2.2. Antimikrobna svojstva 4](#_Toc110008989)

[2.3. Antiviralna svojstva 5](#_Toc110008990)

[2.4. Sekundarne strukture 6](#_Toc110008991)

[2.5. Predviđanje sekundarne strukture 8](#_Toc110008992)

[3. AUTOENKODERI 9](#_Toc110008993)

[3.1. Varijacijski autoenkoderi 9](#_Toc110008994)

[4. MODEL 10](#_Toc110008995)

[5. ZAKLJUČAK 11](#_Toc110008996)

[LITERATURA 12](#_Toc110008997)

[POPIS KRATICA 14](#_Toc110008998)

[Sažetak 15](#_Toc110008999)

[Abstract 15](#_Toc110009000)

# UVOD

UREDIT SVE

Zahvaljujući napretku medicine i znanosti, otkrivena su i proizvedena mnoga cjepiva i lijekovi koji su značajno doprinijeli suzbijanju raznih bolesti. Unatoč tome, mnoge bolesti i dalje predstavljaju veliku prijetnju ljudskom zdravlju te se neprestano traže nova rješenja za njihovo suzbijanje. U tom kontekstu, peptidi su se istaknuli kao obećavajuće novo rješenje pri liječenju. Iako se još uvijek smatraju pomalo nekonvencionalnim pristupom liječenju raznih bolesti i infekcija i relativno su brojčano ograničeni, peptidi su pokazali obećavajuće rezultate u mnogim istraživanjima [1]. Međutim, predviđanje svojstava novo sintetiziranih peptida i poboljšanje postojećih metoda predviđanja još uvijek ima prostora za napredak.

Brojna fizikalno-kemijska svojstva AVP, poput naboja, gustoće naboja, indeksa nestabilnosti, alifatskog indeksa, hidrofobnog omjera i drugih, pokazala su se kao potencijalni indikatori antiviralne učinkovitosti [2]. Iako AVP spadaju u širi spektar antimikrobnih peptida (AMP) koji se bore protiv bakterija, gljivica, parazita i virusa, dosadašnja predviđanja koja su razvijena za AMP nisu se pokazala učinkovitima u predviđanju AVP svojstava. Stoga se ulaže napor u razvoj boljih modela koji bi se specifično usredotočili na predviđanje svojstava AVP. Jedno takvo istraživanje proveli su Kuan Y. Chang i Je-Ruei Yang na Odjelu za računarstvo Sveučilišta Tajvanski Otok [1]. Njihovo istraživanje je pokazalo da njihov model predviđanja temeljen na metodi slučajnih šuma (random forest) značajno nadmašuje dosadašnje modele, što je glavna motivacija za ovaj rad.

Osim toga, područje istraživanja peptida preklapa se s tehnikama strojnog učenja, pružajući uzbudljive mogućnosti za dizajn i otkrivanje novih peptida. Prirodna evolucija je stvorila bogat repertoar biomolekula, uključujući peptide i proteine, s jedinstvenim funkcijama, ali često s zajedničkim evolucijskim porijeklom. Međutim, ograničena raznolikost sekvenci unutar ovih molekula može otežati istraživanje novih varijanti. Kako bismo se nosili s tim izazovom, predlažemo iskorištavanje snage strojnog učenja i generativnih algoritama.

Temeljeno na uspjesima strojnog učenja u različitim područjima, koristimo generativni duboki model za učenje na temelju opsežne baze podataka poznatih antimikrobnih peptida. Konkretno, koristimo varijabilni autoenkoder za generiranje novih sekvenci peptida s antiviralnom aktivnošću. Istražujući latentni prostor, možemo identificirati peptide s željenim svojstvima i interpolirati ih duž prediktivnog vektora, čime otkrivamo nove peptide koji pokazuju doza-odgovornu antiviralnu aktivnost. Ovaj pristup predstavlja obećavajući koncept za primjenu umjetne inteligencije u dizajniranju peptida, zaobilazeći iscrpno pretraživanje sekvenci u knjižnicama.

Ovaj rad prezentira rezultate našeg istraživanja, koji prikazuju potencijal spajanja strojnog učenja i istraživanja peptida. Korištenjem inovativnih računalnih metoda, cilj nam je ubrzati otkrivanje i dizajn antimikrobnih peptida s poboljšanom učinkovitošću. Integracija umjetne inteligencije u inženjering peptida i proteina pruža izvanredne mogućnosti za razvoj novih terapijskih strategija.

1. PEPTIDI

## Što su peptidi?

Peptidi su biomolekule prisutne u svim živim organizmima. Sastavljeni su od aminokiselina povezanih peptidnim vezama. Imaju raznolike uloge u biološkim sustavima, kao što su signalizacija, obrana od patogena, regulacija enzimskih aktivnosti i strukturna podrška.

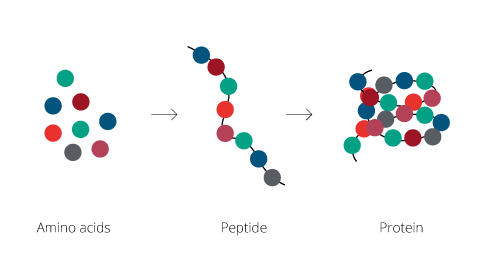
U prirodi se pojavljuje dvadeset standardnih aminokiselina. Svaka aminokiselina ima jedinstvenu strukturu i svojstva te one među sobom mogu stvarati peptidne veze i tako se povezivati u lance.

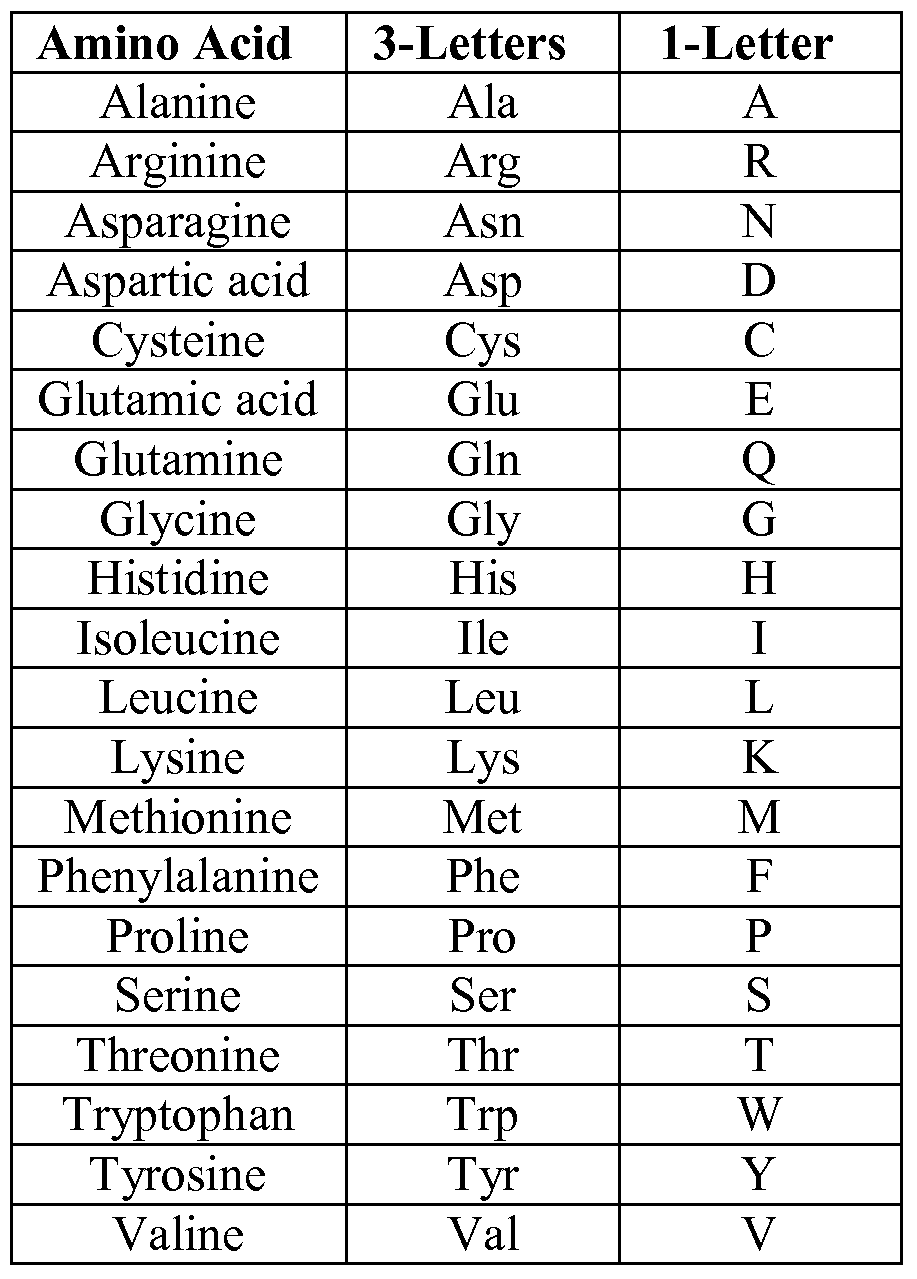
Peptidi se mogu sastojati od samo nekoliko aminokiselina do nekoliko stotina aminokiselina. Ovisno o broju aminokiselina, peptidi se mogu klasificirati kao dipeptidi, tripeptidi, oligopeptidi ili polipeptidi. Peptidi također mogu biti linearni ili ciklički, ovisno o načinu povezivanja aminokiselina.

Peptidi mogu imati antimikrobnu aktivnost, boreći se protiv bakterija, gljivica, parazita i virusa. Takvi peptidi, poznati kao antimikrobni peptidi (AMP), imaju široki spektar djelovanja i mogu biti važni u razvoju novih lijekova protiv infekcija. Identificiranje peptida koji se vežu na specifične receptore ili proteine može pomoći u razumijevanju bioloških mehanizama i razvoju ciljanih terapija.

Peptidi se mogu sintetizirati na različite načine. Klasična metoda sinteze peptida je tekućinska sinteza, koja uključuje postupno dodavanje aminokiselina na rastući peptidni lanac. Napredak u tehnologiji omogućio je razvoj automatskih sintetičkih aparata koji ubrzavaju proces sinteze.

Peptidi imaju raznolike primjene u različitim područjima. U medicini, peptidi se koriste kao terapijski agensi za liječenje različitih bolesti, uključujući rak, dijabetes, i neurološke poremećaje. Peptidi se također koriste u kozmetičkoj industriji zbog svojih svojstava pomlađivanja kože i stimulacije kolagena.

  
*Slika 2.1. Pojednostavljeni prikaz aminokiselina, peptida i proteina (preuzeto iz [4])*



*Slika 2.2. Popis svih prirodnih aminokiselina i njihove oznake (preuzeto iz [5])*

## Antimikrobna svojstva

Antimikrobni peptidi (AMP) su prirodni peptidi koji igraju ključnu ulogu u obrani organizama od mikrobnih infekcija. Oni su dio prirodnog imunološkog sustava prisutnog kod širokog spektra živih organizama, uključujući ljude, životinje, biljke i mikroorganizme. AMP se ističu svojom sposobnošću da selektivno ciljaju i ubiju mikrobe, uključujući bakterije, gljivice, parazite i viruse.

Antimikrobni peptidi imaju raznoliku strukturu koja može uključivati linearno raspoređene aminokiseline, cikličke strukture, kao i kombinaciju različitih strukturnih motiva. Ovi peptidi često sadrže hidrofobne i hidrofilne regije, što im omogućuje interakciju s membranama mikroba. Mnogi antimikrobni peptidi imaju pozitivno nabijene aminokiseline, kao što su lizin i arginin, što im pomaže u vezivanju za negativno nabijene komponente mikrobnih membrana.

Antimikrobni peptidi djeluju na mikrobe na različite načine, ciljajući njihove vitalne strukture i funkcije. Neki peptidi djeluju na mikrobne membrane, oštećujući njihovu strukturu i uzrokujući curenje unutarstaničnih tvari. Drugi peptidi prodiru u unutrašnjost mikroba i ciljaju vitalne procese kao što su sinteza proteina i DNA replikacija. Također, neki peptidi imaju imunomodulatorna svojstva, potičući imunološki odgovor organizma i potencijalno regulirajući upalne procese.

Jedna od ključnih prednosti antimikrobnih peptida je njihova manja sklonost razvoju otpornosti kod mikroorganizama. Otpornost na antimikrobne lijekove postala je globalni problem, ali mnogi antimikrobni peptidi djeluju na mikrobe putem mehanizama koji se razlikuju od konvencionalnih antibiotika, što smanjuje vjerojatnost razvoja otpornosti.

## Antiviralna svojstva

STARI DIO  
 Podvrsta antimikrobnih peptida – antiviralni peptidi (AVP), je poznata po tome što se bori protiv raznih virusa. Postoji više mehanizama koje AVP koriste u borbi protiv virusa, npr. blokiranje vezanja virusa, prevencija spajanja virusa sa stanicama domaćina, prekid procesa signaliziranja virusa ili sprječavanje razmnožavanja virusa u stanicama domaćina [1].

Kao i ostali peptidi, mogu biti prirodni (dobiveni iz biljaka, insekata, vodozemaca, podmorskih organizama ili sisavaca) ili umjetno sintetizirani. Postoje tri glavne metode za racionalno stvaranje novih peptida [2]:

* dizajn temeljen na predlošku (*template-based design)* – pokušava dodati selektivnost i poboljšati aktivnost već postojeće sekvence peptida dodavanjem neke aminokiseline ili premještanjem već postojeće; ovakve modifikacije mogu dovesti do stvaranja novog AVP čak i od neaktivnih peptida
* fizikalno-kemijska metoda – također koristi već poznate sekvence kako bi stvorili sličnu sekvencu s drugačijim fizikalno-kemijskim svojstvima
* ‘de novo’ metoda – stvara nove peptide koristeći uzorke i učestalosti peptida

NOVI DIO – UREDIT

Antiviralni peptidi (AVP) su specifična podvrsta antimikrobnih peptida koja ima sposobnost suzbijanja virusnih infekcija. Virusne infekcije predstavljaju značajan zdravstveni izazov diljem svijeta, a antiviralni peptidi pružaju obećavajuće alate za suzbijanje ovih infekcija.

Antiviralni peptidi imaju specifična svojstva koja im omogućuju borbu protiv virusa. Oni mogu ciljati vitalne faze virusnog životnog ciklusa, kao što su ulazak virusa u stanice domaćina, replikacija virusne RNA ili DNA, te montaža i oslobađanje novih virusnih čestica. Osim toga, antiviralni peptidi mogu modulirati imunološki odgovor domaćina i poticati stvaranje protuvirusnih citokina.

Antiviralni peptidi djeluju na viruse na različite načine. Kao i kod AMP, neki se mogu vezati za virusne čestice i spriječiti njihov ulazak u stanice domaćina, a drugi mogu prodrijeti u unutrašnjost zaraženih stanica i inhibirati virusnu replikaciju.

Predviđanje antiviralnih svojstava peptida izuzetno je važno za razvoj novih lijekova i prevenciju virusnih infekcija. Mnoge metode strojnog učenja i računalnog modeliranja koriste se za predviđanje antiviralnih svojstava peptida na temelju njihovih fizičko-kemijskih svojstava i strukture. Ova područja istraživanja doprinose razvoju boljih modela za predviđanje antiviralne aktivnosti peptida i optimizaciju njihovih svojstava.

# AUTOENKODERI

## Varijacijski autoenkoderi

# MODEL

# ZAKLJUČAK

Peptidi s antiviralnim svojstvima sve više obećavaju u liječenju bolesti izazvanih virusima, a da bi se mogli koristiti i primijeniti treba ih prvo otkriti. Kako ne bi morali testirati svaki peptid za svaki virus, što je vremenski zahtjevno, skupo i neefikasno, tu je vrlo obećavajuće postalo predviđanje antiviralne aktivnosti peptida prema njihovim fizikalno-kemijskim svojstvima.

Takvo predviđanje može biti vrlo učinkovito, ako se dobro izvede. Tako je i ovim jednostavnim modelom strojnog učenja postignuta točnost od preko 80% što je vrlo dobro za neke jednostavne potrebe. Ugađanjem nekih parametara i uključivanjem nekih dodatnih svojstava prema kojima bi se mogla predviđati antiviralna aktivnost, npr. sekundarnih struktura, koje su bile zamišljene kao dio ovog modela, ali zbog određenih problema na kraju nisu uključene u model, mogla bi se i povećati točnost predviđanja i tako bi ono moglo imati i širu primjenu i to čak za neka veća i važnija znanstvena istraživanja. Predviđanje antiviralne aktivnosti peptida predstavlja jedan težak, kompleksan, ali i zanimljiv zadatak te kada bi se shvatio i iskoristio potencijal koji nudi, to bi moglo vrlo pozitivno utjecati na budućnost i ostvariti velike korake u medicini, a samim time i ljudskim životima.

# LITERATURA

[1] Kuan Y. Chang, Je-Ruei Yang, s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3734225/>, 25. lipnja 2020.

[2] Liana Costa Pereira Vilas Boas, Marcelo Lattarulo Campos, Rhayfa Lorrayne Araujo Berlanda, Natan de Carvalho Neves, Octávio Luiz Franco, s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7079787/>, 26. lipnja 2020.

[3] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Peptide>, 25. lipnja 2020.

[4] World of Peptides, s Interneta, [world-of-peptides.com](http://www.world-of-peptides.com/what-is-the-difference-between-a-peptide-and-a-protein/), 5. srpnja 2020.

[5] Courtney Simons, s Interneta, <https://cwsimons.com/structure-of-amino-acids-and-proteins/>, 5. srpnja 2020.

[6] [Ali Adem Bahar](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bahar%20AA%5BAuthor%5D&cauthor=true&cauthor_uid=24287494),[Dacheng Ren](https://www.ncbi.nlm.nih.gov/pubmed/?term=Ren%20D%5BAuthor%5D&cauthor=true&cauthor_uid=24287494), s Interneta, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3873676/>, 26. lipnja 2020.

[7] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Protein_secondary_structure>, 27. lipnja 2020.

[8] ThoughtCo., s Interneta, <https://www.thoughtco.com/protein-structure-373563>, 5. srpnja 2020.

[9] Golifescience, s Interneta, <https://www.golifescience.com/secondary-structure-of-proteins/>, 27. lipnja 2020

[10] C.Mirabello, G.Pollastri, s Interneta, <http://distillf.ucd.ie/porterpaleale/>, 29. lipnja 2020.

[11] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Relative_accessible_surface_area>, 6. srpnja 2020.

[12] Claus AF Andersen & Burkhard Rost, s Interneta, <https://rostlab.org/papers/2003_rev_assign/paper_slow.html>, 6. srpnja 2020.

[13] Guy Yachdav, Peter Hoenigschmid, Laszlo Kajan, s Interneta, <https://github.com/Rostlab/reprof>, 29. lipnja 2020.

[14] RPBS Web portal, s Interneta, <https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::psipred>, 29. lipnja 2020.

[15] J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8 (2015),

A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5: 725-738 (2010),

Y Zhang. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, vol 9, 40 (2008), s Interneta, <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>, 29. lipnja 2020.

[16] Niklas Donges, s Interneta, <https://builtin.com/data-science/random-forest-algorithm>, 30. lipnja 2020.

[17] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Decision_tree>, 5. srpnja 2020.

[18] Tutorialspoint, s Interneta, <https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm>, 30. lipnja 2020.

[19] Tony Yiu, s Interneta, <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>, 6. srpnja 2020.

[20] Abilash R., s Interneta, <https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57>, 5. srpnja 2020.

[21] Wikipedia, s Interneta, <https://en.wikipedia.org/wiki/Weka_(machine_learning)>, 30. lipnja 2020.

[22] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[23] Thakur N, Qureshi A, Kumar M (2012) [AVPpred: collection and prediction of highly effective antiviral peptides.](http://nar.oxfordjournals.org/content/40/W1/W199.long) Nucleic Acids Res. 2012 Jul;40(Web Server issue):W199-204. doi: 10.1093/nar/gks450.  
PMID: [22638580](http://www.ncbi.nlm.nih.gov/pubmed/22638580), s Interneta, <http://crdd.osdd.net/servers/avppred/>, 30. lipnja 2020.

[24] Modlamp dokumentacija, s Interneta, <https://modlamp.org/modlamp.html#module-modlamp.descriptors>, 6. srpnja 2020.

[25] Gufosowa - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/wiki/File:K-fold_cross_validation_EN.svg>, 1. srpnja 2020.

# POPIS KRATICA

AVP – antiviralni peptidi

AMP – antimikrobni peptidi

MSA – *multiple sequence alignment*

SVM – *support vector machines*

WEKA - *Waikato Environment for Knowledge Analysis*

# Sažetak

Virusi i dalje predstavljaju glavni uzrok bolesti kod ljudi usprkos masovnoj proizvodnji cjepiva i antivirusnih lijekova, a velik potencijal za nove pomake u tom području predstavljaju peptidi s antiviralnom aktivnošću. Kako bi efikasno otkrivali nove takve peptide potrebno je koristiti neku vrstu predviđanja, a pokazano je da je to moguće na osnovi njihovih fizikalno-kemijskih svojstava. Takvo predviđanje, ako se dobro izvede, može biti vrlo učinkovito. Posebno se istaknulo strojno učenje metodom *random forest*, kakvo je implementirano i u ovom radu. Daljnjim trudom i ulaganjem u ovakva istraživanja može se ostvariti velik napredak u medicini i poboljšati kvaliteta ljudskog života.

Ključne riječi: antiviralni peptidi, predviđanje sekundarne strukture, algoritam *random forest*, strojno učenje

# Abstract

Viruses are still the main cause of human diseases despite the mass production of vaccines and antiviral drugs. Antiviral peptides have a great potential to change things for the better in that particular field. For efficient discovery of such new peptides, some kind of prediction should be used and it was shown that this was possible based on their physicochemical properties. Such prediction, if done right, can be very effective. Machine learning using the random forest method was proven especially efficient, which was also implemented in this study. With further effort and investments in such research, a great progress in medicine can be achieved and the quality of human lives improved.

Keywords: antiviral peptides, secondary structure prediction, random forest algorithm, machine learning