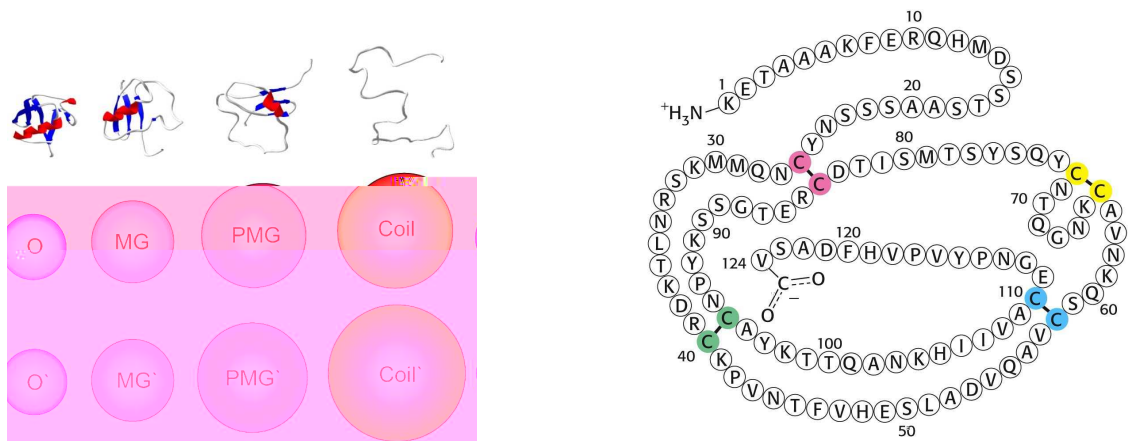


Primena tehnika istraživanja podataka u cilju uspostavljanja korelacije između antigenih regiona i nuređenih delova proteina

Broj neuređenih proteina za koje je poznato da imaju ulogu u signaliziranju i regulaciji, naglo raste. Takođe je utvrđeno da su upravo neuređeni proteini ili neuređeni regioni u proteinu odgovorni za visoko specifične interakcije niskog afiniteta, višestruko vezivanje jednog proteina za više partnera, ili obrnuto. Utvrđeno je i da je većina sistemskih autoantigena ekstremno neuređena. Kako nije poznato zašto neki proteini postaju autoantigeni, ovo svojstvo je jako bitno. Time se nameće pitanje da li postoji korelacija između neuređenih regiona i antigenih regiona u proteinu. U ovom radu smo tehnikama istraživanja podataka utvrdili da su epitopi (delovi antigena) više zastupljeni u uređenim regionima proteina, čak i kada su u pitanju izuzetno neuređeni proteini, ali i da epitopa ima i u neuređenim regionima. Tvđenje važi za sve molekule histokompatibilnog kompleksa (MHC klase I i II). Za pronalaženje neuređenih regiona i epitopa smo koristili programe za predviđanje: VSL2, NetMhcPan i NetMhciiPan respektivno.

Mnogi proteinski regioni ili neki celi proteini nemaju definisanu tercijalnu (3D) strukturu, kao što pokazuju eksperimentalni rezultati dobijeni u *in vitro* uslovima. Neuređeni regioni se eksperimentalno identifikuju na osnovu tercijalne strukture proteina metodama: Difrakcione kristalografije X-zracima; Neuklearno magnetne rezonantne spektroskopije (NMR) i Cirkularnog dihiroizma (CD). Zbog vremenske zahtevnosti ovih metoda kao i niza drugih ograničenja razvijeno je preko 60 programa za predviđanje neuređenih regiona na osnovu primarne strukture proteina (niske uzastopnih amino kiselina u proteinskoj rečenici). Programi za predviđanje neuređenih regiona se mogu podeliti u dve grupe: prvu čine programi zasnovani na fizikohemijskim osobinama amino kiselina u proteinu (PONDR, FoldUnFold, PreLink, IUPred, GlobProt, FoldIndex, VSL2); Drugu grupu čine programi zasnovani na metodama poravnanja homologih proteinskih rečenica (RONN, DISOPRED).

Postoje različite strukture proteina: najuređenija struktura je nasumično klupko (engl. „random coil“) a predstavlja najviše razvijeno stanje globularnih proteina, pretopljiva globula (eng. „pre-molten globule“) izdužena delimično struktuisana forma, topljiva globula (eng. „molten globule“) kompaktna neuređena struktura koja može sadržati značajne delove uređene strukture. Navedene strukture proteina su prikazane na slici 1 :



Eksperimenti izvedeni 1950 – tih godina su dokazali da redosled amino kiselina u proteinskoj rečenici sadrži sve informacije za pravljenje aktivne tercijalne strukture. (Christian Anfisen na enzimima).

Za predviđanje neuređenih regiona smo u ovom radu koristili VSL2 program , koji podjednako dobro predviđe neuređene regione nezavisno od dužine[1].

Antigen (skr. Ag, od prvobitnog eng. „antibody generator“) je molekul koji prepoznaje imuni sistem organizma, dok je epitop region ili deo antigenakoji se vezuje za receptore na Ag – vezujućim ćelijama imunog sistema. Imuni sistem čine organizovana tkiva koja brane organizam od stranih molekula, infektivnih mikroorganizama i njihovih toksina. Imuni dogovor čini prepoznavanje antigena, aktivacija limfocita i efektorna faza eliminacije antigena. Adaptivni imuni sistem sisara je evolirao tako da izlaže delove (epitope) proteina, kao i sopstvene proteine, ćelijama imunog sistema. Delovi proteina (epitopi) su peptidi (niz uzastopnih amino kiselina u proteinskoj rečenici) veličine do 25 amino kiselina. Oni se transportuju na površinu ćelija u kompleksu sa proteinima glavnog histokompatibilnog kompleksa organizma, da bi ih prepoznale efektorne ćelije imunog sistema. Molekuli glavnog histokompatibilnog kompleksa (eng. „major histocompatibility complex“) klase I i II (skr. MHC I i MHC II) kod čoveka nose naziv HLA I i HLA II (eng. „human leukocyte antigens“), jer su prvobitno imunološki definisani kao antigeni na na leukocitima prilikom transfuzije krvi. Individualna genska i imunološka specifičnost organizma je genski definisana alelima klasa MHC I i MHC II. Do danas je poznato 1469 alela MHC klase I i 517 alela MHC klase II za čoveka.

Iako imuni sistem ima ulogu u odbrani od bolesti, virusa, infekcija, itd... i on nekada zakaže. Jedan pristup u proveru zašto i kada se to dešava je da se simulira unapređen model imunog sistema i analizira veza između patogena i domaćina. Najvažniji zadatak ovakvog modela je pronalaženje delova proteina – epitopa, koje imuni sistem prepoznaje, i na taj način indukuje imuno odgovor. Prepoznavanje epitopa je važno za razvoj vakcina zasnovanih na peptid – koktelima, i daje bolji uvid u prirodu različitih oboljenja: alergija, kancer i autoimunih oboljenja.

Grupa danskih imunologa bioinformatičara je razvila niz programa i metoda za identifikovanje epitopa [2]. U ovom radu koristimo njihove programe NetMhcPan i NetMhciiPan koji predviđaju epitope koji se vezuju redom za MHC klase I i II.

Za ovo istraživanje je prikupljeno 654 proteina iz različitih funkcionalnih grupa. Najveću grupu čine proteini iz disprot baze (483) za koje je utvrđeno eksperimentalnim putem da poseduju neuređene regione[3]; 21 protein iz PDB baze: 16 sa 90% uređenom strukturom i 5 sa 90% neuređenom strukturom; 19 proteina iz grupe kancer testis antigena; 4 proteina iz EBNA grupe [4]; Bakterijski proteini (131): 81 protein je kompletno uređen po VSL2 programu za predviđanje neuređenih regiona, a 50 proteina su kompletno neuređeni po istom programu.

Za obradu prikupljenih proteina, propuštanje kroz programe za predviđanje neuređenih i antigenih regiona, vizuelni i uporedni prikaz navedenih regiona, automatizovano izvršavanje programa kao i skladištenje dobijenih rezultata u svrhu daljeg istraživanja, je razvijena aplikacija nazvana EPDIS. EPDIS aplikacija je napisana u programskom jeziku Java, verzija 6. Podaci se skladište u relacionu bazu podataka (IBM DB2, nekomercijalna verzija Express-C 9.7).

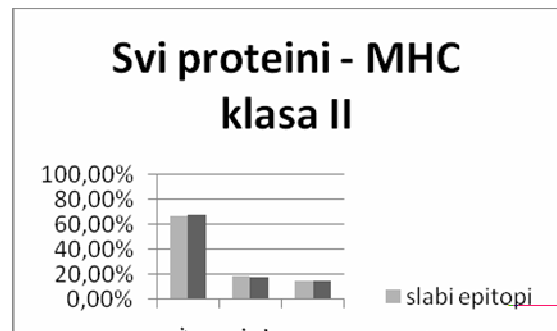
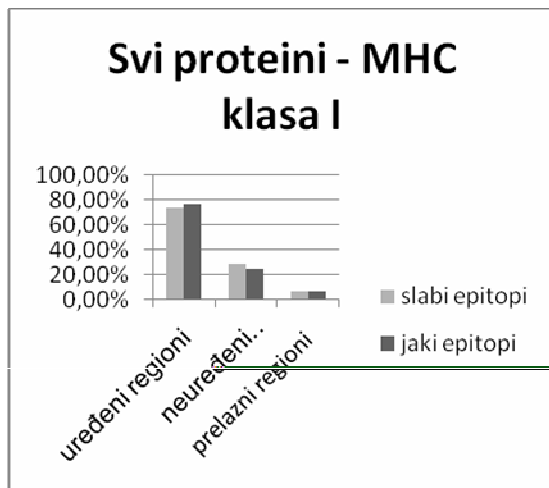
Kada su programi za predviđanje završili sa radom, ustanovljeno je da 49.13% svih amino kiselina (iz svih prikupljenih proteina) pripada neuređenim regionima. Broj razmatranih i uskladištenih peptida je preko 200 miliona, a o svakom je sačuvan i podatak o afinitetu vezivanja za obe MHC klase, da li je prepoznat kao epitop ili ne, hidrofobna vrednost, protein kome pripada i funkcionalna grupa iz koje je izdvojen protein. Ukupan broj prepoznatih epitopa (jakih i slabih) kao i raspodela po uređenim / neuređenim regionima za obe MHC klase je dat tabelom 1:

MHC I	
ukupan broj epitopa	1073894
ukupan broj slabih epitopa	822261
ukupan broj jakih epitopa	251633
ukupan broj epitopa	215457
broj slabih epitopa	169510
broj jakih epitopa	45947
ukupan broj epitopa	792991
broj slabih epitopa	601216
broj jakih epitopa	191775
ukupan broj epitopa	65446
broj slabih epitopa	51535
broj jakih epitopa	13911

MHC II	
ukupan broj epitopa	3608750
ukupan broj slabih epitopa	3409297
ukupan broj jakih epitopa	199453
ukupan broj epitopa	678708
broj slabih epitopa	643001
broj jakih epitopa	35707
ukupan broj epitopa	2403925
broj slabih epitopa	2269419
broj jakih epitopa	134506
ukupan broj epitopa	526117
broj slabih epitopa	496877
broj jakih epitopa	29240

Programom VSL2 za predviđanje neuređenih regiona je pronađeno 2446 neuređenih regiona u razmatranim proteinima. Dužine neuređenih regiona su od 1 do 799, prosečna dužina neuređenih regiona je 30 amino kiselina. Indeks hidropatije izračunat prema Kajt Doolitl skali (engl. "Kytte Doolittle") za epitope se kreće od -3.5 do 3.5. Izračunavanja su sprovedena za sve ljudske alele MHC klase I (ukupno 1649) i MHC II (517).

Ukupan broj pronađenih epitopa, gore navedenim programima, je 4 682 644. Od kojih se 1 073 894 vezuje za molekule MHC klase I, a 3 608 750 za molekule MHC klase II. Epitopi koji se vezuju za molekule MHC klase II su brojniji, ali tvrdjenje da se više epitopa nalazi u uređenim regionima važi za obe klase. Sledećim slikama je prikazan odnos epitopa (jakih i slabih) u uređenim, neuređenim i prelaznim regionima. Grafici su dati za sve analizirane proteine, kao i za izdvojenu funkcionalnu grupu kancer - testis antigenih proteina:





43.58%	Skroz uređena	Slabi epitopi	[-1.5, 2]	ITTQSTLPY	HLA*A0276
22.38%	Skroz neuređena	Slabi epitopi	[-2, 1.5]	TSFESMIEY	HLA*A0269
15.39%	Prelazni regioni	Svi epitopi	[-2, 1.5]	MSLPMNSLY	HLA*A0269
12.55%	Skroz uređena	Jaki epitopi	[-2, 3.5]	IWEEGTFNI	HLA*A0269
6.10%	Skroz neuređena	Jaki epitopi	[-2, 1.5]	SSNSSFLSF	HLA*A0211

Najdominantniji aleli su isti za sve utvrđene grupe. Što znači da se za MHC klasu I ne izdvajaju različite grupe alela koje dobro prepoznaju epitope u zavisnosti od strukture proteina, ili tipa epitopa (jaki i slabi) već da su aleli koji prepoznaju najviše epitopa najfrekventniji u svim strukturama izabranih proteina. Sledećom tabelom su prikazani najdominantniji aleli za proteine iz grupa: DisProt, PDB i kancer:

Category	Cluster	Whole population
HLA-A0276		572
HLA-A0211		707
HLA-A0269		708
HLA-A0214		482
HLA-A0248		376
HLA-A0272		354
HLA-A0247		378
HLA-A0228		377
HLA-A0244		407
HLA-A0206		362
HLA-A0213		448
HLA-A0279		352
HLA-A0221		350
HLA-A0257		270
HLA-A0216		436

ALLELE CODE

Rezultati koji su do sada prikazani se odnose na sve analizirane proteine osim bakterijskih (131 protein). Epitope u bakterijskim proteinima prepoznaje druga grupa alela, koja nema preseka sa prethodnom, a prikazana je sledećom tablicom:

Category	Cluster	Whole population
HLA-B1503		1,612
HLA-B1516		1,476
HLA-B1517		1,144
HLA-B1529		1,032
HLA-B1525		901
HLA-B0813		867
HLA-B1528		722
HLA-B1501		722
HLA-B1505		671
HLA-B1524		659
HLA-B1506		650
HLA-B1502		467
HLA-B0750		453
HLA-B1507		448
HLA-B0737		441

Primenom tehnike pravila pridruživanja , izdvojena su pravila sa najvećom podrškom i nivoom poverenja 100%. Značajno pravilo dobijeno na ovaj način je da slabi epitopi koji se nalaze u neuređenim regionima imaju hidrofobnost iz intervala [-1.9, 1.1].

Kod bakterijskih proteina i jaki i slabi epitopi uzimaju vrednosti iz intervala [-1.9, 1.1] u neuređenim regionima. U uređenim regionim iz intervala [-0.6, 2.4], sa retkim izuzecima preko 2.4.

Izdvojeni su i epitopi koji se zajedno javljaju u neuređenim regionima, u tabeli 6 je prikazan deo rezultata:

ID	Rule	Support	Confidence	Lift
6,887	[WQAPMFDAI] ==> [WQIPHSYAI]	26.6571%	100.0000%	2.58
6,872	[YQVFSVLSL] ==> [WQIPHSYAI]	25.4323%	100.0000%	2.58
6,821	[YQVFSVLSL] + [RQFMWTTLL] ==> [WQIPHSYAI]	23.9914%	100.0000%	2.58
6,816	[LPLPLIPL] ==> [LPMSLPLPL]	23.7752%	100.0000%	3.48
6,810	[RQWFLDLPL] + [WQAPMFDAI] ==> [WQIPHSYAI]	23.6311%	100.0000%	2.58
6,802	[RQWFLDLPL] + [YQVFSVLSL] ==> [WQIPHSYAI]	23.4870%	100.0000%	2.58
6,673	[MQQAYAAPM] + [YQVFSVLSL] ==> [WQIPHSYAI]	22.4784%	100.0000%	2.58
6,657	[WQAPMFDAI] + [RQFMWTTLL] ==> [WQIPHSYAI]	22.4784%	100.0000%	2.58
6,670	[YQVFSVLSL] + [WQAPMFDAI] ==> [WQIPHSYAI]	22.4784%	100.0000%	2.58
6,676	[RQWFLDLPL] + [YQVFSVLSL] + [RQFMWTTLL] ==> [WQIPHSYAI]	22.4784%	100.0000%	2.58
6,637	[RQWFLDLPL] + [LQFWLNILL] ==> [WQIPHSYAI]	22.4063%	100.0000%	2.58
6,635	[MQQAYAAPM] + [WQAPMFDAI] ==> [WQIPHSYAI]	22.4063%	100.0000%	2.58
6,598	[YQVFSVLSL] + [RQYDRVAEL] ==> [WQIPHSYAI]	22.2622%	100.0000%	2.58
6,583	[LPTTMNYPL] + [LPYALRIEL] ==> [LPMSLPLPL]	22.1902%	100.0000%	3.48

Analogno su pronađeni aleli koji u neuređenim regionima prepoznaju iste epitope. Sledećom tabelom je preikazan deo rezultata:

Rule	Support	Confidence
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B1598]+[ALLELE_CODE=HLA-B9503] ==> [ALLELE_CODE=HLA-B1503]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1598] ==> [ALLELE_CODE=HLA-B1503]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B9527]+[ALLELE_CODE=HLA-B1503]+[ALLELE_CODE=HLA-B9503] ==> [ALLELE_CODE=HLA-B9532]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9532]+[ALLELE_CODE=HLA-B1598]+[ALLELE_CODE=HLA-B1503] ==> [ALLELE_CODE=HLA-B1569]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9532]+[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B1503] ==> [ALLELE_CODE=HLA-B9527]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B9527] ==> [ALLELE_CODE=HLA-B9532]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9527]+[ALLELE_CODE=HLA-B1503]+[ALLELE_CODE=HLA-B9503] ==> [ALLELE_CODE=HLA-B1569]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B1598]+[ALLELE_CODE=HLA-B9532]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9532] ==> [ALLELE_CODE=HLA-B1598]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9527] ==> [ALLELE_CODE=HLA-B1503]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B9527] ==> [ALLELE_CODE=HLA-B1598]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9532]+[ALLELE_CODE=HLA-B1503]+[ALLELE_CODE=HLA-B9503] ==> [ALLELE_CODE=HLA-B1569]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B9527]+[ALLELE_CODE=HLA-B1503] ==> [ALLELE_CODE=HLA-B9532]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B9503] ==> [ALLELE_CODE=HLA-B1598]	19.5395%	100.0000%
[ALLELE_CODE=HLA-B9532]+[ALLELE_CODE=HLA-B1569]+[ALLELE_CODE=HLA-B1598]+[ALLELE_CODE=HLA-B1503] ==> [ALLELE_CODE=HLA-B9503]	19.5395%	100.0000%

Za epitope MHC klase II demografskim klasterovanjem dobijamo osnovne statistike o raspodeli epitopa po strukturnim regionima proteina. Rezultati su prikazani tabelom:

62.89%	Skroz uređena	Slabi epitopi	[-1.5, 4]	LQSMRALDF	DRB1*0101
17.82%	Skroz neuređena	Slabi epitopi	[-1.5, 1.5]	FPRMSNLRL	DRB1*0101
13.77%	Prelazni regioni	Slabi epitopi	[-1.5, 2]	MNKLKMMAL	DRB1*0101
3.73%	Skroz uređena	Jaki epitopi	[-1.5, 3.5]	VEVLQSMRA	DRB1*0101
0.99%	Skroz neuređena	Jaki epitopi	[-1.5, 2]	FPRMSNLRL	DRB1*0101
0.81%	Prelazni regioni	Jaki epitopi	[-1.5, 2.5]	FKMIDTDNS	DRB1*0101

Demografskim klasterovanjem se epitopi grupišu u zavisnosti od toga da li pripadaju uređenim / neuređenim regionima i tipu epitopa (jaki ili slabi). Neuronskim klasterovanjem se dobija bolja podela koja uzima u obzir peptid koji predstavlja epitop, hidrofobnu vrednost epitopa i alele koje ga prepoznaju. Dobijeni model grupiše epitope nešto drugačije po strukturnim regionima proteina. Model je tačnosti 0.967, sličnost slogova u klasteru je veća kao i „udaljenost“ klastera .

32.77%	Skroz uređena	Slabi epitopi	[0, 4]	VSYLVRYMG	DRB1*1527
30.11%	Skroz uređena	Slabi epitopi	[-1.5, 1]	LQSMRALDF	DRB1*0101
18.81%	Skroz neuređena	Svi epitopi	[-1.5, 2]	FPRMSNLRL	DRB1*0101
14.58%	Prelazni regioni	Svi epitopi	[-1.5, 2]	MNKLKMMAL	DRB1*0101
3.73%	Skroz uređena	Jaki epitopi	[-1.5, 3.5]	VEVLQSMRA	DRB1*0101

Epitopi koji se nalaze u neuređenim regionima se ponašaju isto, nezavisno od toga da li su jaki ili slabi. Uzimaju vrednosti za hidrofobnost iz intervala $[-1.5, 2]$. Najdominantniji alel je DRB1*0101, koji je i najčešći ljudski alel.

Aleli MHC klase II koji se ponašaju slično, prepoznaju iste epitope u većini slučajeva, su dati tabelom 10. koja uključuje nivo podrške (učestalost pojavljivanja pravila), i nivo poverenja sa kojim se aleli javljaju zajedno.

ID	Rule	Support	Confidence	Lift
ILW10033	ILW10033	0.00000	1.00000	1.00000
ILW10034	ILW10034	0.00000	1.00000	1.00000
ILW10035	ILW10035	0.00000	1.00000	1.00000
ILW10036	ILW10036	0.00000	1.00000	1.00000
ILW10037	ILW10037	0.00000	1.00000	1.00000
ILW10038	ILW10038	0.00000	1.00000	1.00000
ILW10039	ILW10039	0.00000	1.00000	1.00000
ILW10040	ILW10040	0.00000	1.00000	1.00000
ILW10041	ILW10041	0.00000	1.00000	1.00000
ILW10042	ILW10042	0.00000	1.00000	1.00000
ILW10043	ILW10043	0.00000	1.00000	1.00000
ILW10044	ILW10044	0.00000	1.00000	1.00000
ILW10045	ILW10045	0.00000	1.00000	1.00000
ILW10046	ILW10046	0.00000	1.00000	1.00000
ILW10047	ILW10047	0.00000	1.00000	1.00000
ILW10048	ILW10048	0.00000	1.00000	1.00000
ILW10049	ILW10049	0.00000	1.00000	1.00000
ILW10050	ILW10050	0.00000	1.00000	1.00000
ILW10051	ILW10051	0.00000	1.00000	1.00000
ILW10052	ILW10052	0.00000	1.00000	1.00000
ILW10053	ILW10053	0.00000	1.00000	1.00000
ILW10054	ILW10054	0.00000	1.00000	1.00000
ILW10055	ILW10055	0.00000	1.00000	1.00000
ILW10056	ILW10056	0.00000	1.00000	1.00000
ILW10057	ILW10057	0.00000	1.00000	1.00000
ILW10058	ILW10058	0.00000	1.00000	1.00000
ILW10059	ILW10059	0.00000	1.00000	1.00000
ILW10060	ILW10060	0.00000	1.00000	1.00000
ILW10061	ILW10061	0.00000	1.00000	1.00000
ILW10062	ILW10062	0.00000	1.00000	1.00000
ILW10063	ILW10063	0.00000	1.00000	1.00000
ILW10064	ILW10064	0.00000	1.00000	1.00000
ILW10065	ILW10065	0.00000	1.00000	1.00000
ILW10066	ILW10066	0.00000	1.00000	1.00000
ILW10067	ILW10067	0.00000	1.00000	1.00000
ILW10068	ILW10068	0.00000	1.00000	1.00000
ILW10069	ILW10069	0.00000	1.00000	1.00000
ILW10070	ILW10070	0.00000	1.00000	1.00000
ILW10071	ILW10071	0.00000	1.00000	1.00000
ILW10072	ILW10072	0.00000	1.00000	1.00000
ILW10073	ILW10073	0.00000	1.00000	1.00000
ILW10074	ILW10074	0.00000	1.00000	1.00000
ILW10075	ILW10075	0.00000	1.00000	1.00000
ILW10076	ILW10076	0.00000	1.00000	1.00000
ILW10077	ILW10077	0.00000	1.00000	1.00000
ILW10078	ILW10078	0.00000	1.00000	1.00000
ILW10079	ILW10079	0.00000	1.00000	1.00000
ILW10080	ILW10080	0.00000	1.00000	1.00000
ILW10081	ILW10081	0.00000	1.00000	1.00000
ILW10082	ILW10082	0.00000	1.00000	1.00000
ILW10083	ILW10083	0.00000	1.00000	1.00000
ILW10084	ILW10084	0.00000	1.00000	1.00000
ILW10085	ILW10085	0.00000	1.00000	1.00000
ILW10086	ILW10086	0.00000	1.00000	1.00000
ILW10087	ILW10087	0.00000	1.00000	1.00000
ILW10088	ILW10088	0.00000	1.00000	1.00000
ILW10089	ILW10089	0.00000	1.00000	1.00000
ILW10090	ILW10090	0.00000	1.00000	1.00000
ILW10091	ILW10091	0.00000	1.00000	1.00000
ILW10092	ILW10092	0.00000	1.00000	1.00000
ILW10093	ILW10093	0.00000	1.00000	1.00000
ILW10094	ILW10094	0.00000	1.00000	1.00000
ILW10095	ILW10095	0.00000	1.00000	1.00000
ILW10096	ILW10096	0.00000	1.00000	1.00000
ILW10097	ILW10097	0.00000	1.00000	1.00000
ILW10098	ILW10098	0.00000	1.00000	1.00000
ILW10099	ILW10099	0.00000	1.00000	1.00000
ILW10100	ILW10100	0.00000	1.00000	1.00000

U ovom radu pronalaze se neuređeni i antigeni regioni u proteinu, analizira veza i učestalost pojavljivanja antigenih regiona u različitim strukturama proteina za sve, do danas, poznate ljudske alele. U obzir je uzeta i hidrofobna vrednost antigenih regiona – epitopa i ustanovljen je interval hidropatije u neuređenim regionima proteina. Pronalaženje navedenih regiona je izvedeno programima za predviđanje VSL2, NetMhcPan i NetMhciiPan. Nad dobijenim rezultatima su sprovedene napredne tehnike obrade i istraživanja podataka (klasterovanje i pravila pridruživanja). Relativna jednostavnost tih tehnika leži u činjenici da oni podatke nad kojima se primenjuju tretiraju u potpunosti sintaksno i bez potrebe za ekspertskom pomoći. One su u ovom radu primenjene na oblast imunologije, sa ciljem klasterovanja antigenih regiona (epitopa). Takvi ciljevi imaju svoj teorijski i praktičan značaj. Veoma je važno grupisati epitope i alele koje ih prepoznaju prema strukturi i funkcionalnoj grupi proteina kojoj pripadaju, jer se tako može ustanoviti ponašanje imunog sistema. Ustanovljena pravila i rezultati su važni za pravljenje i poboljšanje vakcina zasnovanih na peptid koktelima za različite vrste bolesti: alergije, autoimune bolesti, kancerogena oboljenja, itd. Rezultati klasterovanja epitopa prema strukturi proteina kojoj pripadaju, hidrofobnoj vrednosti i alelama koje ga prepoznaju dobijeni tehnikama prikazanim u ovom radu su veoma dobri. Za potrebe obrade podataka dobijenih navedenim programima je napisana aplikacija EPDIS. Aplikacija je napisana u programskom jeziku Java, verzija 6. EPDIS aplikacija je omogućila grafički prikaz rezultata navedenih programa, a time i analizu rezultata za

pojedinačne proteine i alele, skladištenje dobijenih epitopa i neuređenih regiona u relacionu bazu podataka DB2, preklapanje intervala koji čine epitop sa uređenim – neuređenim delovima proteina i pridruživanje hidrofobne vrednosti svakom epitopu.

Pretpostavlja se da je broj epitopa u uređenim regionima veći nego u neuređenim, ali su sve pretpostavke o raspodeli epitopima po uređenim – neuređenim regiona zasnovane na analizi pojedinačnog proteina za jedan ili mali podskup postojećih ljudskih alela.

U ovom radu je analizirano 654 proteina iz različitih funkcionalnih grupa i 1986 alela. Rezultati su sledeći:

- 1) Broj epitopa u uređenim regionima je oko 3 puta veći nego u neuređenim. Tvđenje važi za obe klase MHC I i II.
- 2) Broj prepoznatih epitopa za MHC klasu II je 3.36 puta veći nego za MHC klasu I i pored činjenice da MHC klasa II ima skoro 3 puta manje alela.
- 3) Aleli koji prepoznaju najviše epitopa MHC klase I, za proteine iz DisProt baze, PDB baze i kancer – testis antigene su jednako dominantni u svim strukturama proteina. Izdvojeni aleli su ujedno i najčešći ljudski aleli.
- 4) Aleli koji prepoznaju najviše epitopa MHC klase I za bakterijske proteine čine nepreklapajući skup alela sa navedenim u rezultatu 3). Ne zavise od strukture proteina (uređeni – neuređeni delovi u proteinu).
- 5) Hidrofobnost epitopa u neuređenim regionima je u intervalu [-2, 1.5]
- 6) Aleli koji prepoznaju najviše epitopa MHC klase II su isti za sve funkcionalne grupe analiziranih proteina uključujući i bakterijske. Najbrojniji su u svim strukturama proteina i ujedno predstavljaju najčešće ljudske alele.