

Универзитет у Београду
Математички факултет

Јована Поповић

Анализа повезаности секундарне структуре протеина и различитих
начина класификација аминокиселина

Мастер рад

Београд
2016
Универзитет у Београду – Математички факултет
Мастер рад

Аутор: Јована Поповић

Наслов: Анализа повезаности секундарне структуре протеина и различитих начина класификација аминокеселина.

Ментор: проф. др Саша Малков, Универзитет у Београду – Математички факултет

Чланови комисије:

проф. др Миодраг Живковић, Универзитет у Београду – Математички факултет

проф. др Ненад Митић, Универзитет у Београду – Математички факултет

Датум:

Садржај

1	Увод.....	4
2	Опис проблема.....	8
2.1	Учесталост парова аминокиселина у секундарним структурама.....	8
2.2	Класификација аминокиселина.....	8
2.3	Рачунање зависности.....	10
2.4	Проналажење значајних вредности.....	11
3	Решење проблема.....	14
3.1	Израчунавање статистике.....	15
3.2	Класификација аминокиселина.....	15
3.3	Корелације.....	16
3.4	Обрада резултата.....	17
4	Резултати.....	18
5	Дискусија.....	37
6	Закључак.....	41
7	Прилози.....	42
8	Референце.....	46

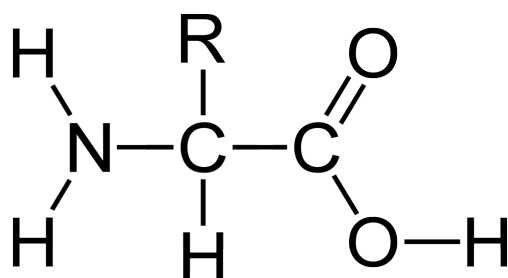
1 Увод

Протеини су макромолекули који имају вишеструку улогу у функционисању живих организама. Добијају се процесом синтезе протеина, где ћелија гради протеин потребан за њену изградњу и функционисање, али и за обављање функција које нису стриктно везане за ћелију која их производи. Улога протеина је одређена просторним обликом који протеин прави и који се дефинише као структура протеина. Протеин је полимер аминокиселина - садржи један или више ланаца састављених од аминокиселина. Називи и једнословне ознаке свих аминокиселина приказане су у Табели 1.

Ознаке аминокиселина и њихови називи			
A	Аланин	L	Леуцин
R	Аргинин	K	Лизин
N	Аспарагин	M	Метионин
D	Аспарагинска киселина	F	Фениланин
C	Цистеин	P	Пролин
Q	Глутамин	S	Серин
E	Глутаминска киселина	T	Треонин
G	Глицин	W	Триптофан
H	Хистидин	Y	Тирозин
I	Изолеуцин	V	Валин

Табела 1. Списак свих 20 аминокиселина и њихових једнословних ознака.

Аминокиселина је сачињена од централног атома угљеника (α угљеника) који је повезан са амино ($-\text{NH}_2$) и карбоксилном ($-\text{COOH}$) групом (слика 1.1) Преостале две везе угљеник гради са једним атомом водоника и бочним ланцем, названим R-остатак, специфичним за сваку аминокиселину.



Слика 1.1: Шема аминокиселине

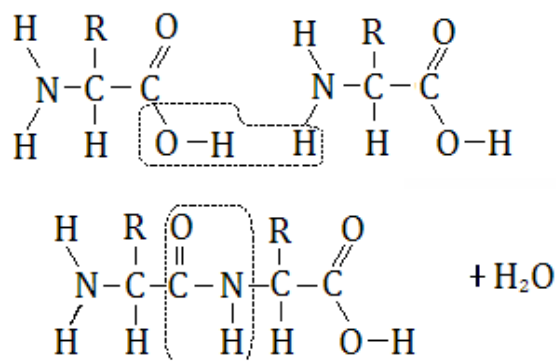
Две аминокиселине се повезују **пептидним везама**. Оваква веза се гради када се атом угљеника из карбоксилне групе једне аминокиселине веже за атом азота амино групе друге аминокиселине при чему се ослобађа један молекул воде [1] (слика 1.2). Тако се добијају дугачки ланци где са једне стране имамо неповезану карбоксилну, а са друге стране неповезану амино групу.

Постоје четири нивоа структуре протеина:

- примарна структура
- секундарна структура
- терцијарна структура
- кватенарна структура

Примарну структуру протеина представља низ аминокиселина које га граде.

Између α -амино групе (N - H) и α -карбоксилне групе (C = O) различитих аминокиселина може доћи до водоничних веза које производе увијање ланца и на тај начин могу да утичу на просторни облик протеина. Ако се може утврдити образац по коме су везане, он се дефинише као **секундарна структура**.



Слика 1.2: Грађа пептидне везе

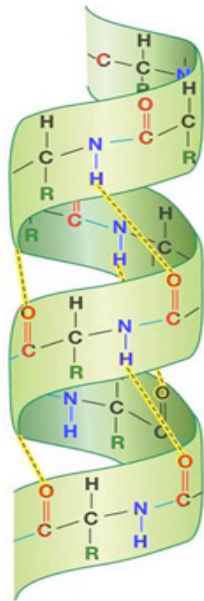
α хеликс је једна од најзаступљенијих секундарних структура. Јавља се када до водоничних веза долази између аминокиселина са размаком од четири остатка. То доводи до хеликоидног увијања пептидног ланца. Растојање између аминокиселина које граде водоничне везе у ланцу може бити и три и пет места, што такође производи хеликоидно увијање. Илустрација оваквог понашања у ланцу аминокиселина налази се на слици 1.3.

β трака је такође честа структура где су два ланца повезана водоничним везама при чему се добија наборана плоча. Илустрација β траке је на слици 1.4.

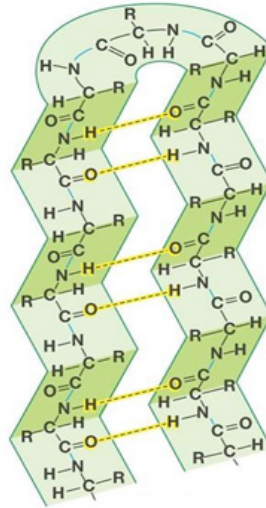
Уколико постоји тачна просторна структура протеина, применом алгоритма ДССП (*Define Secondary Structure of Protein*) одређује се која аминокиселина гради коју секундарну структуру[2]. Издваја се осам секундарних структура. Следи приказ њихових уобичајених једнословних ознака [3]:

- G - 3_{10} хеликс - са три заокрета
- H - α хеликс - са четири заокрета
- I - π хеликс - са пет заокрета

- S – кривина (једини елемент који се не базира на водоничним везама)
- T – кривина премошћена водоничном везом
- E – β трака у паралелним и антипаралелним елементима ланца.
- V – остатак у изолованом β мосту.
- C – све остало, без препознативе структуре.



Слика 1.3: α хеликс



Слика 1.4: β раван

Терцијарна структура протеина је тачан просторни распоред језгара атома који чине протеин. У протеинима може доћи до веза код бочних ланаца протеина где су аминокиселине удаљеније у ланцу али су просторно близу услед савијања протеина. Оне дефинишу терцијарну структуру [4].

Кватернарну структуру чине међусобни односи више полипептидних ланаца у протеину.

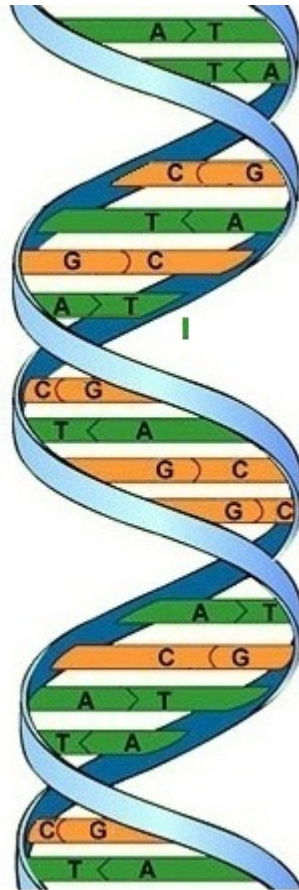
Протеини настају у ћелији, у процесу **синтезе протеина**. Први корак синтезе протеина је **транскрипција**; процес где се од молекула ДНК добија РНК. Једно ћелије садржи молекул дезоксирибонукленске киселине (**ДНК**) упакован у хромозом. **ДНК** се састоји од два полимерна ланца сачињена од четири врсте нуклеотида: Аденин (A), тимин (T), гуанин (G) и цитозин (C). Два ланца у ДНК спојена су водоничним везама међу паровима нуклеотида што доводи до увијања оба ланца. Парови који се спајају су аденин и тимин (две водоничне везе) као и цитозин и гуанин (три водоничне везе), што омогућава да се анализом секвенце једног ланца предвиди секвенца другог. На слици 1.5 је приказан један део молекула ДНК.

Дакле, са слике се може прочитати секвенца левог ланца као и његов комплементарни десни ланац.

A T G A C T C T G G T G T A C T
T A C T G A G A C C A C A T G A

Рибонуклеинска киселина (**РНК**) има један ланац а у процесу транскрипције добија се од једног ланца ДНК са његовим комплементарним нуклеотидама при чему се тимин замењује са урацилом (U). Уколико се РНК синтетише из левог ланца на слици 1.5, одговарајући РНК ланац је:

A U G A C U C U G G U G U A C U



Слика 1.5: Део молекула ДНК

Други корак синтезе протеина је процес **транслације** у ком се на основу ланца РНК синтетише протеин. Транслација се врши на основу скупа правила који се назива *генетички код*. РНК ланац се чита по триплетима нуклеотида, где сваки триплет одговара једној аминокиселини. Ови триплети су названи кодони.

Транслација секвенце РНК почиње од стартног кодона, AUG – кодона за метионин. Затим се линеарно остатак секвенце преводи у аминокиселине све док се не дође до једног од завршних кодона: UAA, UAG и UGA. Према примеру са слике почетак одговарајуће секвенце протеина је:

AUG ACU CUG GUG UAC U...
M T L V Y ...

2 Опис проблема

Секундарне структуре које граде аминокиселине могу бити у одређеним случајевима зависне од класе самих аминокиселина. Анализа склоности појединих класа аминокиселина да граде одређене секундарне структуре може бити значајна за више аспеката анализе протеина. Предвиђање примарне структуре познавајући класификацију аминокиселина као и склоност те класе ка одређеним структурама може бити прецизније, а такође, ове зависности су важне и за анализу саме функције протеина.

Утврђивање и анализа ових зависности обухватају:

- Анализу учесталости парова аминокиселина у секундарним структурама;
- Анализу припадности парова аминокиселина у одређеној класи;
- Рачунање зависности између класа аминокиселина и врста секундарних структура и
- Препознавање и истицање значајних међузависности.

2.1 Учесталост парова аминокиселина у секундарним структурама

Рачунање статистике појављивања биграма састоји се од обраде података добијених из базе података PDB, из којих су, применом алгоритма ДССП, добијени подаци о примарним и секундарним структурама протеина. Ови подаци садрже следеће елементе:

- PDBID – јединствени идентификатор у бази PDB
- SeqID – ознака секвенце протеина
- PrimaryStruct – примарна структура протеина.
- SecondaryStruct – секундарна структура протеина.

Циљ је, за сваки биграма аминокиселина, избројати појављивање у свим паровима секундарних структура, као и целокупно појављивање у узорку.

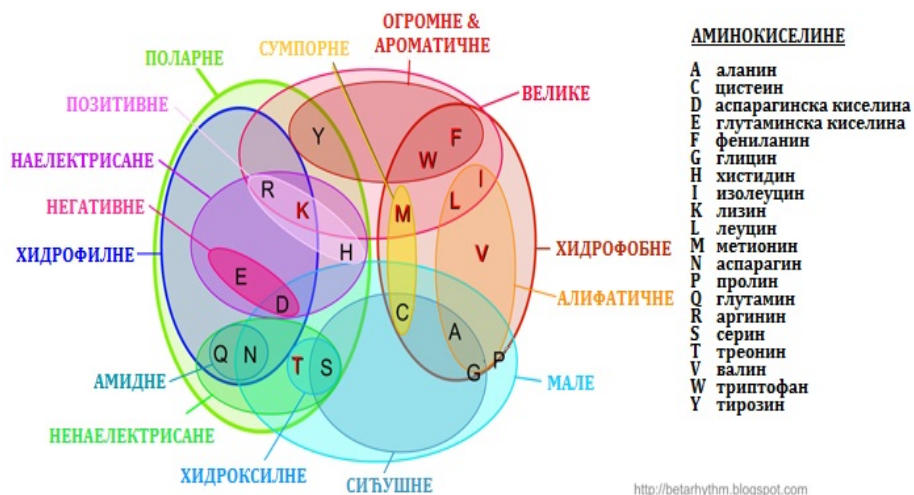
Добијени резултати ће се користити у наредним корацима.

2.2 Класификација аминокиселина

Аминокиселине се сврставају у одређене класе на основу специфичних особина које зависе искључиво од њиховог хемијског састава. R-остатак се може састојати од само једног атома као што је случај код *глицина* (где је тај остатак атом водоника) а може бити и релативно сложен [5]. Може бити алифатични или ароматични, може бити позитивно или

негативно наелектрисан и слично. Метионин и цистеин имају бочни ланац који садржи сумпор па спадају у класу *сумпорних* аминокиселина. Сумпор је, за разлику од кисеоника, у мањој мери електронегативан, па је ова аминокиселина мање хидрофобна. Улога цистеина је, на пример, значајна због његове склоности ка дисулфидним везама којима формира “мостове” у различитим деловима ланца [6].

Склоност ка контакту са водом је једна од битних особина аминокиселина, па оне могу бити хидрофилне или хидрофобне. Хидрофобне аминокиселине се често налазе у унутрашњости протеина како би биле изоловане од молекула воде.



Слика 2.1 Класе аминокиселина

Аминокиселине *поларне* класе су већински хидрофилне. Бочни ланци наелектрисаних аминокиселина се налазе ван површине ланца протеина. Могу се спајати јонским везама (супротна наелектрисања) и тако утицати на структуру протеина [7].

На слици 2.1 је приказ класификације аминокиселина коришћених у овој анализи. Свака аминокиселина припада некој класи (+) или не (-) што је таблично приказано на слици 2.2.

Поред наведених класа разматране су и класе аминокиселина за које је раније препознато да имају склоност ка некој секундарној структури: 'склоне хеликсу', 'склоне равни', 'склоне заокрету' и 'без склоности' [8].

Класификација парова аминокиселина у односу на ове класе захтева разматрање припадности класи сваке од аминокиселина у пару. Тако добијамо четири односа биграма и класе:

- ++ - обе аминокиселине из биграма припадају класи,
- +- - прва аминокиселина припада а друга не припада класи
- -+ - прва аминокиселина не припада али друга припада класи
- -- - ниједна аминокиселина из биграма не припада класи.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Поларне	-	+	+	+	-	+	+	-	+	-	-	+	-	-	-	+	+	-	+	-
Позитивне	-	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-
Наелектрисане	-	+	-	+	-	-	+	-	+	-	-	+	-	-	-	-	-	-	-	-
Негативне	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Хидрофилне	-	+	+	+	-	+	+	-	-	-	-	+	-	-	-	-	-	-	-	-
Амидне	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ненаелектрисане	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-
Хидроксилне	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
Сићушне	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-
Мале	+	-	+	+	+	-	-	+	-	-	-	-	-	-	+	+	+	-	-	-
Алифатичне	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	+
Хидрофобне	+	-	-	-	+	-	-	-	-	+	+	-	+	+	-	-	-	+	-	+
Велике	-	+	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	+	+	-
Огромне и ароматичне	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	-
Сумпорне	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
Склоне хеликсу	+	+	-	-	-	+	+	-	-	-	+	+	+	-	-	-	-	-	-	-
Склоне равни	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	+	+	+
Склоне заокрету	-	-	+	+	-	-	-	+	-	-	-	-	-	-	+	+	-	-	-	-
Без склоности	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-

Слика 2.2. Класификација аминокиселина

2.3 Рачунање зависности

Зависности две посматране карактеристике могу да се изразе помоћу коефицијената корелације. Добијају се на више различитих начина, а који ће се метод користити зависи од стања самих променљивих чију корелацију тражимо. Најчешћи је Пирсонов коефицијент корелације који се примењује на променљиве које су линеарно повезане.

Корелације се рачунају у односу на класу. Како постоји четири типа односа парова аминокиселина и класа, потребно је, за сваки од њих, утврдити у којој мери аминокиселине из биграма које су у том односу теже да граде секундарне структуре за које се рачуна корелација. За сваку класу се рачунају следеће корелације:

	++	+-	-+	--
$S_{11}S_{12}$	X_{1++}	X_{1+-}	X_{1-+}	X_{1--}
$S_{21}S_{22}$	X_{2++}	X_{2+-}	X_{2-+}	X_{2--}
...				

X_{1++} је мера склоности пара аминокиселина које припадају класи да се нађу у структури $S_{11}S_{12}$. X_{1+-} је мера склоности пара аминокиселина, где прва припада класи а друга не, да граде исти пар $S_{11}S_{12}$. Ове мере се рачунају Пирсоновим коефицијентом корелације над бинарним променљивим. За пар секундарних структура $S_{11}S_{12}$ и један од односа класе и биграма аминокиселина, на пример "+-" рачунају се следеће вредности:

	Биграми класификовани као "+-"	
Биграми граде пар $S_{11}S_{12}$	✓	x
✓	A	B
x	C	D

- **A** : број појављивања биграма аминокиселина које граде структуре $S_{11}S_{12}$ а притом са класом за коју се рачуна имају однос "+-".
- **B**: број појављивања биграма који граде структуре $S_{11}S_{12}$ али са класом имају неки од преосталих односа.
- **C**: број појављивања свих биграма који су за дату класу класификовани као "+-" али граде неке друге структуре.
- **D**: остатак (нису класификовани као "+-" нити граде структуру $S_{11}S_{12}$).

Коефицијент корелације рачуна се по следећој формули [9]:

$$R = \frac{AD - BC}{\sqrt{(A+B)(A+C)(B+D)(C+D)}} \quad (*)$$

Тако добијамо коефицијент корелације X_{1+-} . На исти начин се рачунају корелације за друге односе са класом, као и за друге парове секундарних структура $S_{11}S_{12}$.

2.4 Проналажење значајних вредности

Из резултата корелација добијених у претходном кораку потребно је издвојити оне које су значајне за даљу анализу. Резултати се добијају у табличном облику. Корелације добијене у анализи помножене су са 10000 како би резултати у табелама били прегледнији.

Општа таблица корелација за класу "Наелектрисане" приказана је на слици 2.3. У сваком реду имамо пар секундарних структура а затим резултате корелација за сваки од могућих односа биграма секундарних структура и класе. Табела садржи вредности корелација за све комбинације парова секундарних структура.

Издвајање значајних вредности у табелама заснива се на израчунавању критичне вредности r . Корелације које су веће од r означавају тежњу парова аминокиселина које су са класом у једном од четири могућа односа да граде пар секундарних структура S_1S_2 (означен у првој колони реда). Овакве корелације се у табелама обележавају зеленом бојом и подебљаним фонтом. Корелације које су мање од $-r$ обележавају се црвеном бојом и курзивним фонтом а представљају одбијање парова аминокиселина које су у датом односу са класом да граде секундарне структуре S_1S_2 .

Над овако обрађеним таблицама, лако се могу уочити одређене зависности и приступити анализи резултата. Тако се на примеру класе "Наелектрисане" (слика 2.3) може уочити да уколико бар једна аминокиселина из биграма припада класи онда је

велика вероватноћа да ће секундарна струкура изграђена од њих бити α хеликс. Такође, може се уочити и да уколико ниједна аминокиселина из биграма није у класи, онда је велика вероватноћа је да неће уопште бити у α хеликсу. Потпуно супротна ситуација је у случају β траке.

	++	+-	-+	--
HH	477	165	186	-522
HE	-28	15	-52	44
EH	-3	21	-30	9
HT	108	66	-33	-82
TH	-24	-7	-33	45
HG	-43	-49	19	46
GH	-25	56	-16	-19
HS	-17	50	-86	37
SH	9	-9	47	-34
HB	-21	-17	-22	42
BH	-39	-13	-41	62
HI	-5	-8	0	9
IH	-6	-8	-4	12
HC	-8	21	-44	22
CH	52	91	24	-118
EE	-725	-340	-332	904
ET	103	52	49	-133
TE	21	39	15	-53
EG	-1	43	-14	-22
GE	-61	139	-114	11
ES	-88	-106	74	71
SE	5	196	-118	-65
EB	-35	-7	-18	38
BE	-44	12	-39	44
EI	-2	-3	-3	6
IE	-2	-3	-3	6
EC	-163	-67	25	117
CE	-106	44	-63	70
TT	228	98	66	-247
TG	-19	8	-9	10
GT	-31	12	9	-1
TS	-5	187	-93	-72
ST	26	0	7	-19

	++	+-	-+	--
TB	-18	-9	-26	37
BT	0	1	-6	5
TI	3	-6	-12	13
IT	-2	3	1	-2
TC	-50	3	-53	65
CT	123	131	-18	-153
GG	258	8	147	-255
GS	-33	106	-71	-11
SG	-3	27	-20	-4
GB	-31	27	-31	20
BG	-10	36	-29	0
GI	0	0	0	0
IG	-4	-7	5	4
GC	-23	132	-83	-27
CG	10	131	-63	-59
SS	197	33	43	-161
SB	-32	6	-27	33
BS	-57	-49	9	60
SI	3	1	-4	0
IS	-6	-3	5	1
SC	16	-30	0	16
CS	-13	-154	206	-35
BB	-20	-6	-24	34
BI	-2	-3	-3	6
IB	-2	-3	-3	6
BC	-91	-39	23	60
CB	-74	-23	-29	79
II	9	-1	-5	0
IC	-7	-8	4	7
CI	5	2	-1	-3
CC	-127	-160	63	142

Слика 2.3: Таблица резултата

Вредности у табелама са груписаним подацима по једној од секундарних струкура (Слика 2.4), показују јасније зависности секундарне структуре на месту једне од аминокиселина са односом пара аминокиселина и класе. За случај класе "Наелектрисане" уочава се зависност α хеликса од класе и у случају када само једна од аминокиселина припада класи. Супротан закључак важи за траку.

Случајеви Z_{10} хеликса и заокрета премошћеним водоничним везама се у овом групном приказу јасније издвајају. Издваја се особина да се, у случају када ниједна аминокиселина не припада овој класи, ове струкуре вероватно неће ни формирати.

	++	+-	-+	--
HX	488	186	153	-519
XH	474	189	185	-539
EX	-729	-345	-279	868
XE	-712	-238	-365	843
TX	138	152	-19	-175
XT	297	174	43	-324
GX	176	115	58	-227
XG	209	69	91	-234
SX	138	64	-3	-119
XS	72	-2	103	-116
BX	-114	-44	-6	98
XB	-95	-18	-59	110
IX	4	-4	-3	3
XI	9	-4	-8	5
CX	-93	-75	96	31
XC	-192	-143	20	195

Слика 2.4 Приказ груписаних резултата

Резултати из табеле 2.3 могу да се представе и на начин приказан на слици 2.5. У редовима и колонама су секундарне структуре, а њихови пресеци су вредности корелација по припадности класи (“++”, “+-” горњи ред, “-+”, “--” доњи ред). Дакле, за пар структура S_1S_2 , довољно је пронаћи ћелију у пресеку реда чија је вредност S_1 и колоне чија је вредност S_2 тачније ћелију:

	S ₂	
S ₁	++	+-
	-+	--

Вредности корелација су исте као у случају са слике 2.3, међутим, овако формирана таблица је прегледнија и олакшава проналазак резултата корелација па ће се користити уместо таблице 2.3.

S ₁ \S ₂	H		E		T		G		S		B		I		C	
H	477	165	-28	15	108	66	-43	-49	-17	50	-21	-17	-5	-8	-8	21
	186	-522	-52	44	-33	-82	19	46	-86	37	-22	42	0	9	-44	22
E	-3	21	-725	-340	103	52	-1	43	-88	-106	-35	-7	-2	-3	-163	-67
	-30	9	-332	904	49	-133	-14	-22	74	71	-18	38	-3	6	25	117
T	-24	-7	21	39	228	98	-19	8	-5	187	-18	-9	3	-6	-50	3
	-33	45	15	-53	66	-247	-9	10	-93	-72	-26	37	-12	13	-53	65
G	-25	56	-61	139	-31	12	258	8	-33	106	-31	27	0	0	-23	132
	-16	-19	-114	11	9	-1	147	-255	-71	-11	-31	20	0	0	-83	-27
S	9	-9	5	196	26	0	-3	27	197	33	-32	6	3	1	16	-30
	47	-34	-118	-65	7	-19	-20	-4	43	-161	-27	33	-4	0	0	16
B	-39	-13	-44	12	0	1	-10	36	-57	-49	-20	-6	-2	-3	-91	-39
	-41	62	-39	44	-6	5	-29	0	9	60	-24	34	-3	6	23	60
I	-6	-8	-2	-3	-2	3	-4	-7	-6	-3	-2	-3	9	-1	-7	-8
	-4	12	-3	6	1	-2	5	4	5	1	-3	6	-5	0	4	7
C	52	91	-106	44	123	131	10	131	-13	-154	-74	-23	5	2	-127	-160
	24	-118	-63	70	-18	-153	-63	-59	206	-35	-29	79	-1	-3	63	142

Слика 2.5. Матрични приказ резултата

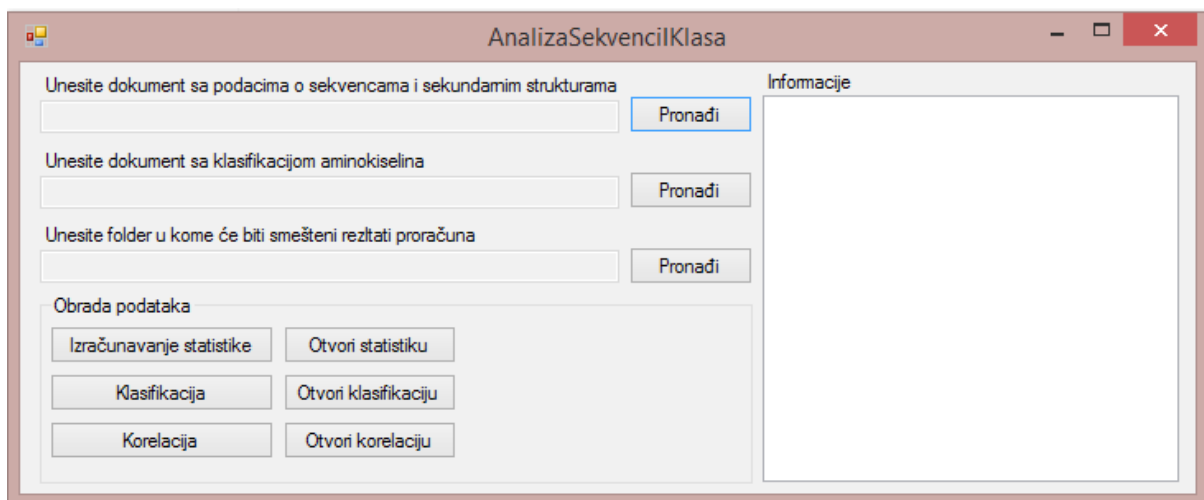
3 Решење проблема

Написан је програм који рачуна потребне статистичке вредности из три корака:

- рачуна статистику појављивања биграма аминокиселина у секундарним структурама и укупно;
- класификује биграме аминокиселина у једну од четири групе односа парова аминокиселина за сваку класу
- рачуна корелације класификације и секундарних структура

Затим се добијени подаци обрађују за даљу анализу.

Приликом покретања програма отвара се прозор за унос основних датотека и података за обраду (слика 3.1). Уноси се фајл у коме су секвенце аминокиселина и њихових секундарних структура добијених применом алогоритма ДССП (прилог 1), затим текстуални фајл са подацима о класификацији аминокиселина (прилог 2) и путања до директоријума у коме ће бити смештени резултати .



Слика 3.1: Улазни програм за обраду података

Приликом покретања се иницијализују и основне листе аминокиселина, тј.

- `ListOfAA` , основна листа аминокиселина, тачније њихових једнословних ознака (метод описан у класи `PomocneMetode`). Прави се на основу фајла `aminokiseline.txt` у коме је списак свих аминокиселина (прилог 3).
- `BigramsOfAA`, листа биграма аминокиселина – биграми једнословних ознака аминокиселина.

3.1 Израчунавање статистике

Израчунавање статистике појављивања биграма аминокиселина у секундарним структурама отпочиње кликом на дугме *Израчунавање статистике*. Покреће се иницијализација листе секвенци и речника појављивања потребних за рачун саме статистике.

- Листа секвенци, `primarySecondaryStructList`, добија се из фајла унетог на почетку (ДССП фајл). За сваки ред из документа у листу се додаје нови објекат класе `PDBSequence` у који се уписују подаци из ДССП документа (`PdbId`, `SeqId`, `PrimaryStruct`, `SecondaryStruct`).
- Речник `Dictionary<string, Dictionary<string, int>> dictSS` у коме ће се рачунати и чувати број појављивања иницијализује се као празан. Кључеви `dictSS` су биграми секундарних структура а вредност за сваки пар структура је нови речник где су кључеви парови аминокиселина а вредности бројеви појављивања парова аминокиселина у секундарним структурама из кључа главног речника. Прва вредност кључа у речнику `dictSS` је “*SUM*”, а његов одговарајући подречник је речних са паровима аминокиселина и укупним бројем појављивања тих парова у узорку за сваки кључ подречника.

Потом се, за сваки елемент из листе `primarySecondaryStructList`, приступа рачунању статистике по следећим правилима:

- У секвенци секундарних структура празан карактер се замењује са “C”. Примарне структуре где постоји аминокиселина означена као “непозната” аминокиселина “Z” се не обрађују.
- За сваку аминокиселину у примарној структури креира се биграма *aa* од ње и њој следеће, као и одговарајући биграма секундарних структура *S₁S₂* на истим позицијама као и биграма аминокиселина.
- Биграми који садрже аминокиселину означену са *X* се прескачу.
- `dictSS["SUM"][aa]` се увећава за један
- `dictSS[ss][aa]` се увећава за један..

На крају је речник попуњен подацима појављивања биграма аминокиселина из ДССП фајла за сваки пар секундарних структура и укупно. Резултати се уписују у csv фајл *statistikaPojavljivanja.csv*, у директоријуму чија је путања унета у поље за резултат при покретању програма. Код који рачуна статистику приказан је у прилогу 4.

3.2 Класификација аминокиселина

Кликом на дугме *Класификација* почиње разврставање биграма аминокиселина по класама. Разврставање се врши у класи `KlaseAminokiselina`, методом `BigramInClass(string aa)` која за биграма аминокиселина враћа једну од четири могућа односа биграма аминокиселина и изабране класе (“++”, “+-”, “-+”, “--”) у зависности од припадности

аминокиселине из биграма класи. Метода је приказана у прилогу 5.

Резултати се чувају у речнику `Dictionary<string, Dictionary<string, string>>` `dictKlasifikacija` где су кључеви називи класа аминокиселина а вредности нови речници са паровима аминокиселина као кључем чија је вредност један од четири могућа односа биграма аминокиселина из кључа и класе.

Подаци из овог речника уписују се у фајл *klasifikacija.csv*.

3.3 Корелације

На почетку рачунања корелације потребно је учитати речнике `dictKlasifikacija` и `dictStatistika` са њиховим вредностима (класификације биграма аминокиселина и статистике појављивања аминокиселина). Иницијализује се објекат `Correlation` `corrFinal` у коме ће се чувати добијене корелације за сваки од четири типа припадности класи.

Рачунање се изводи редом по класи. За сваку класу аминокиселина рачуна се однос парова аминокиселина са датом класом, а затим се за сваки пар секундарних структура рачуна корелација Пирсоновим коефицијентом корелације за бинарне променљиве. Тачније, рачунају се следеће вредности:

- `inStructureAndTypeInit` је број свих појављивања биграма аминокиселина у структури `S1S2` а које имају тражени однос са класом.
- `inTypeWithoutStructureInit` је број појављивања биграма који су у структури `S1S2` али немају тражени однос са класом.
- `inStructureWithoutTypeInit` је број појављивања свих биграма који са датом класом имају тражени однос али нису у структури `S1S2`
- `otherInit` је број појављивања биграма који нису ни у структури `S1S2` ни у траженом односу са класом аминокиселина..

На основу ових вредности рачуна се коефицијент корелације за тренутни тип класе и тај коефицијент се чува у објекту `corrFinal` за ту секундарну структуру а затим се прелази се на рачунање корелације за следећи тип припадности класе.

Други корак се заснива на рачунању података груписаних по структури.

- Иницијализује се речник `Dictionary<string, Dictionary<string, int>>` `dictGrupisano` и попуњава се на основу речника `dictSS`. Прва вредност кључа у речнику `dictGrupisano` је "SUM", а његов подречник се ископира из речника `dictSS`. Остали кључеви у `dictGrupisano` креирају се на основу секундарних структура. За сваку секундарну структуру `S` креирају се кључеви `SX` и `XS`. Одговарајући подречници ових кључева имају парове аминокиселина у кључу, а вредности се рачунају из речника `dictSS`. За пар аминокиселина из кључа броје се појављивања у свим паровима структура фиксираним по једном елементу. Дакле, подречник за кључ `SX` садржи број појављивања парова аминокиселина у паровима секундарних структура таквих да је прва структура `S` док друга није битна. Прилог 6 приказује метод који попуњава овај речник.

- Рачунају се корелације за овако добијене вредности по истом принципу.

Вредности корелација се уписују у одговарајуће табеле резултата које се на крају рачуна уписују у одговарајуће фајлове.

3.4 Обрада резултата

За добијене вредности приказане у таблицама потребно је издвојити границе значајности. Анализа значајних повезаности састоји се од примене статистичког теста значајности. Студентов т-тест је најчешћи метод за тестирање нулте хипотезе. Тест се спроводи формулом

$$t = r * \sqrt{\frac{(n-2)}{(1-r^2)}}$$

тачније, инверзним поступком по формули[9]

$$r = \frac{t}{\sqrt{t^2 + n - 2}}$$

где се добија вредност која се узима као граница за одређивање значајних резултата. Сви резултати који су по апсолутној вредности већи или једнаки од добијене вредности **r** сматрају се значајним.

Критична вредност узета за анализу резултата је **0.015**. Узима се врло мала вредност (много мање од 0.1%) за одбацавање нулте хипотезе како би се за значајне зависности сматрале само оне за које се са врло великом верофватноћом може тврдити да нису добијене случајно. Како су резултати у табелама корелација помножени са 10000 за издвајање значајних зависности узима се вредност од **150** (односно **-150**).

Резултати корелација добијају се у три табеле описане у поглављу 2.4. Над сваком од њих се примењује поступак издвајања значајних резултата на основу узете критичне вредности од 150.

За анализу у раду сви резултати добијени применом алгоритма представљени су у табеларном облику 2.4 и 2.5.

4 Резултати

Програм је примењен на податке о секвенцама аминокиселина и њихових секундарних структура у протеину. На основу класификација аминокиселина израчунате су зависности парова аминокиселина одређене класе и парова секундарних структура. У наставку су резултати корелација за сваку класу:

Алифатичне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	1019	374	51	-19	5	213	80	86	8	-7	32	13	12	1	-73	72
	441	-1253	21	-31	-167	-41	-24	-97	-57	48	-1	-28	-1	-7	-145	101
E	-72	-125	983	535	-178	39	-44	-59	-132	82	-2	-2	-2	-3	-69	292
	139	30	454	-1375	-153	196	81	7	-146	129	23	-16	12	-6	-74	-138
T	-38	-117	-209	-326	-541	-155	-13	-37	-215	-343	-57	-104	7	-3	-374	-539
	128	12	126	284	-660	978	73	-22	-8	410	38	86	3	-4	327	388
G	24	-53	0	-106	-32	9	-214	-73	1	-78	21	-40	0	0	-54	-117
	126	-74	228	-100	-30	36	-167	320	110	-27	60	-29	0	0	146	8
S	-168	-223	-125	-230	-163	-68	-77	-66	-377	-296	-49	-54	-6	0	-417	-102
	55	234	163	127	-127	253	-13	109	-375	765	6	68	5	-2	-308	576
B	-39	-102	42	-27	-74	-25	-26	-43	-44	10	19	-8	-2	-3	-9	106
	170	-33	70	-59	-47	101	82	-17	-34	45	8	-11	-3	6	-33	-54
I	14	7	-2	-3	-2	3	-5	11	-2	0	19	-3	4	-5	-2	14
	3	-17	12	-6	3	-4	4	-10	8	-6	-3	-6	-3	4	-1	-8
C	-347	-487	-122	-177	-320	-255	-170	-188	-359	-136	-58	-27	-6	-8	-543	146
	148	477	319	-46	-171	532	23	232	-315	575	33	29	9	3	-367	493

Табела 4.1: Матрични приказ резултата алифатичних аминокиселина

	++	+-	-+	--
HX	994	441	353	-1222
XH	866	166	525	-1064
EX	815	612	352	-1258
XE	802	307	617	-1217
TX	-739	-640	-254	1157
XT	-638	-143	-715	1069
GX	-200	-147	-8	242
XG	-264	-157	-104	365
SX	-644	-410	-403	1035
XS	-575	-395	-462	1031
BX	-51	41	11	-12
XB	-76	-89	58	69
IX	5	0	1	-4
XI	6	-7	1	1
CX	-871	-299	-291	984
XC	-797	-21	-332	748

Табела 4.2: Груписани резултати алифатичних аминокиселина

Амидне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	15	-32	-6	14	90	0	-9	-19	32	37	-5	-8	-3	15	49	43
	41	-10	-24	9	210	-173	2	15	22	-50	-7	12	-1	-9	65	-90
E	-10	28	-249	-419	30	40	-3	15	-8	-65	-15	-20	-1	-2	-43	-84
	-32	5	-387	641	150	-144	-33	14	71	-2	-17	30	-2	3	32	48
T	-9	63	51	218	160	41	-11	30	75	218	12	46	-3	-2	88	249
	-39	-15	-30	-148	349	-320	-33	5	-22	-159	12	-44	-10	9	-6	-196
G	1	60	-5	123	15	16	34	-55	-6	67	-3	28	0	0	30	104
	-29	-23	-61	-43	7	-20	138	-68	-29	-26	-28	1	0	0	7	-87
S	-1	102	-11	65	-4	49	-14	52	89	34	-9	19	-2	-5	52	81
	-60	-30	-67	4	-43	-3	-39	-6	162	-163	-9	-5	-5	8	63	-116
B	-6	17	-13	9	6	21	-3	14	1	-19	-6	-10	-1	21	-23	-45
	-40	18	-32	20	12	-26	-16	2	14	3	-15	19	-2	-14	19	24
I	-2	-7	-1	-2	-4	-12	-1	-4	-2	0	-1	-2	-4	-14	-2	-7
	-1	6	-2	3	4	7	5	0	0	1	-2	3	-17	23	10	-2
C	10	216	-49	-6	0	202	-26	113	13	-15	-14	-7	-3	13	-30	-75
	-113	-77	-101	89	-120	-59	-112	5	35	-18	-25	26	-5	-5	29	40

Табела 4.3: Матрични приказ резултата амидних аминокиселина

	++	+-	-+	--
HX	51	-20	112	-78
XH	15	59	-11	-38
EX	-244	-421	-309	584
XE	-242	-322	-422	595
TX	202	317	229	-441
XT	166	146	328	-381
GX	40	42	96	-109
XG	11	17	50	-51
SX	81	140	89	-185
XS	97	100	134	-192
BX	-21	-27	5	21
XB	-15	17	-30	12
IX	-6	-20	-12	24
XI	-6	-7	-20	20
CX	-40	95	-104	16
XC	29	62	78	-108

Табела 4.4: Груписани резултати амидних аминокиселина

Наелектрисане аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	477	165	-28	15	108	66	-43	-49	-17	50	-21	-17	-5	-8	-8	21
	186	-522	-52	44	-33	-82	19	46	-86	37	-22	42	0	9	-44	22
E	-3	21	-725	-340	103	52	-1	43	-88	-106	-35	-7	-2	-3	-163	-67
	-30	9	-332	904	49	-133	-14	-22	74	71	-18	38	-3	6	25	117
T	-24	-7	21	39	228	98	-19	8	-5	187	-18	-9	3	-6	-50	3
	-33	45	15	-53	66	-247	-9	10	-93	-72	-26	37	-12	13	-53	65
G	-25	56	-61	139	-31	12	258	8	-33	106	-31	27	0	0	-23	132
	-16	-19	-114	11	9	-1	147	-255	-71	-11	-31	20	0	0	-83	-27
S	9	-9	5	196	26	0	-3	27	197	33	-32	6	3	1	16	-30
	47	-34	-118	-65	7	-19	-20	-4	43	-161	-27	33	-4	0	0	16
B	-39	-13	-44	12	0	1	-10	36	-57	-49	-20	-6	-2	-3	-91	-39
	-41	62	-39	44	-6	5	-29	0	9	60	-24	34	-3	6	23	60
I	-6	-8	-2	-3	-2	3	-4	-7	-6	-3	-2	-3	9	-1	-7	-8
	-4	12	-3	6	1	-2	5	4	5	1	-3	6	-5	0	4	7
C	52	91	-106	44	123	131	10	131	-13	-154	-74	-23	5	2	-127	-160
	24	-118	-63	70	-18	-153	-63	-59	206	-35	-29	79	-1	-3	63	142

Табела 4.5: Матрични приказ резултата наелектрисаних аминокиселина

	++	+-	-+	--
HX	488	186	153	-519
XH	474	189	185	-539
EX	-729	-345	-279	868
XE	-712	-238	-365	843
TX	138	152	-19	-175
XT	297	174	43	-324
GX	176	115	58	-227
XG	209	69	91	-234
SX	138	64	-3	-119
XS	72	-2	103	-116
BX	-114	-44	-6	98
XB	-95	-18	-59	110
IX	4	-4	-3	3
XI	9	-4	-8	5
CX	-93	-75	96	31
XC	-192	-143	20	195

Табела 4.6: Груписани резултати наелектрисаних аминокиселина

Аминокиселине склоне хеликсу

S1\S2	H		E		T		G		S		B		I		C	
H	2052	47	-13	12	385	230	95	-9	30	103	-2	3	-10	-1	-41	161
	212	-1874	-9	8	-150	-379	18	-84	-88	-37	-8	6	-3	11	-156	28
E	-70	-102	-629	35	-195	-43	-35	-66	-187	-44	-34	-20	-3	-3	-252	20
	112	47	-82	545	47	153	76	19	-20	207	-14	57	11	-5	-105	277
T	-36	-103	-189	-150	-360	358	-9	-50	-186	-99	-79	-51	3	-7	-346	-273
	115	19	172	133	-293	230	68	-8	35	206	34	79	3	1	208	334
G	52	-20	-42	61	-5	39	41	47	43	28	-10	4	0	0	-8	40
	47	-65	-44	18	-24	-10	100	-162	2	-61	1	4	0	0	-6	-23
S	-181	-200	-142	-56	-131	-64	-69	-51	-323	29	-77	6	-3	0	-382	81
	199	146	27	140	8	155	15	87	-123	342	-23	77	4	-1	-164	380
B	-62	-72	-41	13	-99	-21	-13	-21	-75	-26	-31	6	-3	-3	-103	26
	77	45	-14	34	19	81	15	15	4	80	-22	39	-3	8	-45	99
I	5	-2	-3	-3	0	1	0	-8	6	-1	-3	11	-9	-5	1	-10
	9	-11	-3	8	9	-9	9	-1	7	-10	-3	-5	-5	17	8	1
C	-375	-391	-161	-33	-294	-205	-153	-117	-339	-124	-90	11	-12	-15	-488	83
	395	297	17	143	129	303	73	162	13	369	-44	100	12	12	-233	524

Табела 4.7: Матрични приказ резултата аминокиселина склони хеликсу

	++	+-	-+	--
HX	2116	149	129	-1943
XH	1868	-117	376	-1726
EX	-761	9	-95	686
XE	-716	-16	-33	618
TX	-577	18	-18	462
XT	-328	250	-211	228
GX	44	76	79	-172
XG	-35	-42	146	-64
SX	-591	-17	-127	601
XS	-524	-84	-79	564
BX	-172	-13	-13	160
XB	-144	-10	-37	157
IX	-6	-7	2	9
XI	-13	-10	-1	21
CX	-838	-220	15	853
XC	-802	43	-249	825

Табела 4.8: Груписани резултати аминокиселина склони хеликсу

Хидрофилне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	540	99	-42	30	199	13	-50	-52	-1	54	-25	-19	-4	3	19	11
	155	-541	-54	45	16	-146	36	44	-92	32	-18	47	-5	4	-36	9
E	-12	36	-958	-373	134	24	-13	50	-97	-127	-47	-12	-2	-3	-202	-83
	-35	6	-346	1183	119	-200	-25	-12	136	52	-21	56	-3	7	69	136
T	-36	33	77	134	376	19	-32	24	56	273	-10	8	2	-11	24	118
	-46	33	-42	-123	154	-374	-18	16	-140	-144	-26	21	-16	21	-100	-30
G	-25	80	-72	209	-40	20	278	-67	-32	137	-38	47	0	0	0	167
	-29	-28	-121	-29	23	-11	157	-245	-87	-22	-38	16	0	0	-81	-71
S	34	34	-19	239	19	6	-16	48	283	-16	-29	11	0	0	50	-10
	1	-50	-152	-60	-8	-10	-32	-3	53	-204	-37	39	-5	3	25	-43
B	-42	6	-55	22	3	6	-16	46	-59	-52	-26	-13	-2	12	-110	-50
	-54	66	-48	55	2	-9	-30	-3	31	54	-25	47	-3	-6	53	66
I	-8	-9	-2	-3	-6	-2	-5	-8	-2	-4	-2	-3	-5	-4	-3	-13
	-1	13	-3	7	4	2	10	1	4	2	-3	7	-8	13	6	8
C	86	194	-160	86	134	202	-15	187	-12	-165	-87	-8	0	5	-209	-148
	-56	-167	-91	103	-93	-173	-104	-59	220	-39	-34	88	0	-4	123	149

Табела 4.9: Матрични приказ резултата хидрофилних аминокиселина

	++	+-	-+	--
HX	581	106	138	-559
XH	547	168	122	-576
EX	-958	-390	-251	1117
XE	-943	-219	-408	1097
TX	320	220	-14	-366
XT	456	120	114	-473
GX	197	90	62	-247
XG	207	37	80	-223
SX	209	75	-10	-182
XS	152	-5	114	-184
BX	-133	-45	23	99
XB	-105	3	-68	118
IX	-9	-9	-4	16
XI	-5	-3	-13	17
CX	-167	16	76	26
XC	-218	-85	82	136

Табела 4.10: Груписани резултати хидрофилних аминокиселина

Хидрофобне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	1118	137	80	-40	-8	203	84	98	25	-12	58	6	14	-1	-67	97
	221	-1136	29	-49	-195	-2	-47	-107	-36	24	1	-49	-4	-6	-119	68
E	-82	-127	1266	412	-212	90	-60	-62	-150	168	4	6	-3	-3	-67	373
	160	30	329	-1584	-171	227	98	12	-164	107	35	-40	11	-5	-60	-229
T	-20	-117	-245	-344	-643	-133	-8	-22	-246	-341	-41	-107	15	-5	-435	-523
	171	-34	205	303	-604	1124	82	-47	38	449	78	56	0	-7	478	359
G	44	-44	16	-125	-14	69	-215	-75	30	-75	53	-44	0	0	-25	-100
	114	-95	254	-126	-26	-27	-106	318	149	-87	74	-65	0	0	184	-56
S	-211	-224	-141	-256	-205	-16	-97	-59	-459	-285	-36	-50	-8	9	-515	0
	87	275	257	102	-128	278	2	122	-329	881	24	50	9	-10	-253	601
B	-44	-108	65	-35	-76	6	-21	-51	-36	55	35	-10	-3	-3	19	158
	196	-45	63	-73	-52	96	89	-18	-44	17	9	-25	11	-5	-9	-146
I	17	12	-3	-3	4	0	1	3	9	2	15	-3	16	-9	-1	11
	-6	-18	11	-5	4	-6	3	-6	-1	-7	-3	-5	-7	3	-3	-6
C	-436	-487	-118	-153	-406	-169	-212	-178	-454	-20	-50	-19	-10	-7	-674	338
	224	552	387	-121	-173	601	66	254	-311	626	87	-23	19	-3	-292	452

Табела 4.11: Матрични приказ резултата хидрофобних аминокиселина

	++	+-	-+	--
HX	1093	210	135	-1106
XH	936	-66	343	-931
EX	1066	554	239	-1483
XE	1058	189	564	-1441
TX	-860	-619	-84	1252
XT	-773	-55	-691	1227
GX	-169	-129	70	176
XG	-286	-149	-31	370
SX	-785	-330	-297	1130
XS	-690	-289	-408	1123
BX	-21	105	31	-104
XB	-46	-82	126	-6
IX	21	-4	-7	-6
XI	18	-11	-1	-3
CX	-1072	-64	-166	989
XC	-952	225	-170	648

Табела 4.12: Груписани резултати хидрофобних аминокиселина

Хидроксилне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	-351	-530	-7	-14	-61	-38	-26	-30	33	55	-7	-11	-4	-8	4	61
	-473	836	-10	20	-13	55	-10	37	19	-64	-1	11	3	5	-31	-23
E	57	199	-117	25	17	32	3	-4	112	153	12	-4	-1	-2	52	-47
	-41	-133	50	-22	-76	27	-5	6	2	-145	36	-27	-2	3	233	-151
T	-19	-17	-20	-105	16	-50	-27	-28	30	54	-6	-21	-4	-1	-27	-71
	-38	46	29	62	-39	60	-14	38	48	-84	-9	24	6	-2	0	60
G	-28	-38	-24	-26	-28	-26	22	1	-17	16	-17	-20	0	0	-22	-17
	-39	64	-38	54	-50	64	18	-21	-12	2	-11	28	0	0	-33	43
S	64	375	24	-6	26	76	3	72	175	166	13	-3	-2	7	78	-17
	-63	-247	26	-21	-43	-31	-32	-31	133	-269	33	-25	1	-5	181	-143
B	-1	245	-18	-9	10	51	3	27	36	46	1	4	-1	-2	5	-38
	-45	-146	-10	19	-33	-16	-24	-3	14	-54	6	-8	-2	3	75	-28
I	-3	-8	-1	-2	-5	0	-2	-5	-3	3	-1	-2	4	-4	-3	2
	-3	9	-2	3	-9	8	-5	8	-8	4	-2	3	-6	7	-3	2
C	132	790	29	55	31	112	2	54	247	305	15	2	-4	12	239	14
	-167	-495	67	-97	-32	-67	-70	11	98	-367	63	-52	-7	-2	431	-395

Табела 4.13: Матрични приказ резултата хидроксилних аминокиселина

	++	+-	-+	--
HX	-358	-514	-469	824
XH	-299	-227	-530	641
EX	-59	57	115	-109
XE	-100	18	78	-41
TX	-4	-90	-14	77
XT	7	24	-91	46
GX	-11	-20	-25	37
XG	11	29	-27	-5
SX	191	208	183	-342
XS	307	369	165	-480
BX	17	63	36	-78
XB	13	-13	65	-42
IX	0	-5	-11	12
XI	0	-3	-5	6
CX	352	466	323	-680
XC	223	-40	498	-400

Табела 4.14: Груписани резултати хидроксилних аминокиселина

Велике аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	924	190	49	-12	60	202	79	73	49	41	54	8	12	-3	-18	162
	297	-1074	37	-55	-113	-120	-23	-100	-37	-38	-3	-43	2	-7	-142	-5
E	-54	-114	739	297	-173	64	-38	-52	-145	55	13	2	-2	-3	-78	283
	136	19	162	-920	-137	185	67	14	-111	151	4	-14	-3	8	-164	-48
T	-6	-98	-149	-231	-477	-77	7	-7	-177	-228	-11	-68	11	6	-327	-372
	138	-30	225	110	-374	728	46	-38	44	285	60	14	3	-15	255	332
G	83	-22	69	-56	68	80	-125	-98	92	-35	81	-2	0	0	45	-19
	75	-104	98	-85	-28	-93	22	154	106	-126	10	-65	0	0	87	-91
S	-181	-211	-79	-168	-162	-11	-77	-50	-357	-187	-26	-26	-7	14	-403	58
	109	216	175	50	-132	238	-21	116	-208	595	31	14	-1	-6	-242	444
B	-31	-83	46	-10	-66	5	-2	-14	-25	22	9	12	-2	-3	38	162
	127	-16	18	-39	-47	83	34	-16	-41	34	-4	-13	11	-5	-83	-96
I	13	5	-2	-3	5	5	1	-8	19	-5	-2	11	8	-3	4	8
	-9	-6	11	-5	7	-13	9	-2	3	-12	-3	-5	-1	-1	-6	-4
C	-374	-453	3	-202	-354	-168	-176	-137	-382	-137	-12	-51	-10	-12	-451	163
	205	479	261	-54	-152	528	29	217	-192	556	102	-35	15	5	-294	432

Табела 4.15: Матрични приказ резултата великих аминокиселина

	++	+-	-+	--
HX	932	279	230	-1099
XH	775	4	403	-899
EX	587	389	57	-802
XE	647	114	337	-849
TX	-623	-419	-31	830
XT	-564	-17	-469	820
GX	-15	-91	98	4
XG	-184	-147	45	218
SX	-607	-195	-238	804
XS	-538	-269	-252	832
BX	1	115	-59	-49
XB	1	-72	114	-37
IX	14	-1	0	-9
XI	9	-4	3	-6
CX	-779	-251	-166	911
XC	-672	172	-326	608

Табела 4.16: Груписани резултати великих аминокиселина

Негативне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	188	221	-19	-8	-48	35	-32	-47	-38	-18	-15	-28	4	-2	-24	-58
	182	-353	-45	45	-89	54	-34	70	-83	86	-27	46	-5	4	-19	64
E	-1	48	-432	-605	75	133	5	55	-53	-80	-28	-26	-1	-2	-103	-126
	-46	-2	-488	935	130	-217	-24	-25	70	22	-29	49	-2	3	65	76
T	-4	21	7	55	169	133	-9	14	5	232	-7	-13	12	-12	9	38
	-42	17	-100	31	227	-316	-16	4	-117	-86	-49	48	-6	10	-25	-12
G	-28	63	-35	51	-29	3	226	142	-25	87	-22	6	0	0	-11	118
	-56	3	-94	42	-42	37	247	-354	-83	5	-36	29	0	0	-75	-29
S	74	81	-18	213	55	19	6	40	146	108	-25	-13	-2	0	33	-46
	21	-97	-131	-55	43	-63	-1	-30	140	-227	-45	51	6	-4	79	-35
B	-27	-9	-24	-28	16	11	-10	8	-35	-61	-18	-19	-1	-2	-64	-121
	-49	51	-49	64	5	-16	-22	14	-15	66	-20	34	-2	3	18	95
I	-3	-8	-1	-2	1	7	-2	-5	-3	2	-1	-2	2	8	-3	-9
	-3	10	-2	3	1	-6	-5	8	2	-2	-2	3	5	-10	-4	10
C	163	308	-112	-10	173	282	31	176	13	-92	-62	-70	5	3	-92	-201
	39	-304	-203	191	44	-293	-19	-125	222	-100	-87	134	3	-6	46	142

Табела 4.17: Матрични приказ резултата негативних аминокиселина

	++	+-	-+	--
HX	158	210	135	-301
XH	235	323	176	-438
EX	-434	-588	-396	855
XE	-448	-505	-585	937
TX	131	223	68	-254
XT	211	279	188	-408
GX	160	201	123	-287
XG	197	204	186	-346
SX	136	131	107	-216
XS	69	87	157	-201
BX	-72	-121	-16	123
XB	-75	-74	-122	167
IX	0	6	3	-7
XI	6	4	3	-7
CX	2	22	62	-62
XC	-108	-223	75	141

Табела 4.18: Груписани резултати негативних аминокиселина

Аминокиселине без склоности

S1\S2	H		E		T		G		S		B		I		C	
H	-134	-220	8	25	-18	-38	3	4	25	31	21	18	-2	-6	35	50
	-178	308	8	-25	45	-1	-18	10	57	-67	-7	-12	5	1	83	-101
E	0	34	-46	75	-13	83	-5	13	-1	40	2	12	0	-1	15	54
	-28	-4	96	-112	-40	-28	-6	-4	-27	-9	32	-32	-1	2	120	-127
T	-12	23	-13	-8	-13	-71	-3	24	10	14	24	15	-2	5	31	13
	-2	-12	18	-5	33	29	-11	-9	27	-31	37	-42	-7	2	105	-89
G	-7	36	-2	37	11	36	14	-43	4	15	12	7	0	0	26	30
	-11	-16	-11	-17	11	-36	47	-6	19	-25	34	-32	0	0	28	-46
S	-23	8	-9	-4	-2	62	1	38	19	-37	-2	23	-1	8	40	28
	-58	40	44	-27	-50	-8	-24	-10	43	-8	27	-36	-3	-3	-6	-23
B	-2	1	-10	14	7	40	1	2	26	38	-2	20	0	-1	48	35
	-11	8	27	-27	-11	-22	-9	5	-5	-29	28	-33	-1	2	104	-108
I	-1	-4	0	-1	-2	-3	-1	-3	-1	-4	0	-1	-1	4	-1	-5
	-4	7	-1	2	1	2	-3	4	-4	6	-1	2	2	-4	4	1
C	-31	6	-21	5	-14	152	-7	51	24	104	-5	24	-2	13	210	16
	-99	72	40	-28	-113	-25	-57	6	-51	-42	39	-44	-6	-4	-2	-51

Табела 4.19: Матрични приказ резултата аминокиселина без склоности

	++	+-	-+	--
HX	-126	-212	-141	274
XH	-145	-204	-213	323
EX	-40	118	119	-160
XE	-55	73	118	-124
TX	6	-33	95	-45
XT	-25	45	-28	-7
GX	23	0	56	-44
XG	9	0	4	-5
SX	28	20	11	-28
XS	39	69	15	-67
BX	47	62	81	-111
XB	10	43	71	-83
IX	-2	0	1	0
XI	-2	7	0	-4
CX	143	126	-82	-59
XC	214	72	124	-180

Табела 4.20: Груписани резултати аминокиселина без склоности

Поларне аминокселине

S1\S2	H		E		T		G		S		B		I		C	
H	76	-151	-26	29	166	-76	-66	-37	66	59	-29	2	-1	-5	76	52
	-42	120	-49	46	-34	-53	37	65	-95	-28	-6	33	3	3	-84	-43
E	59	117	-978	103	162	17	-4	44	2	-33	-33	-4	-3	-4	-109	-86
	-109	-66	131	734	-41	-137	-35	-5	59	-28	17	20	-4	11	199	-6
T	-33	40	66	18	351	-152	-38	32	108	223	-2	4	0	-8	18	17
	-50	43	-16	-68	20	-214	-17	23	-165	-166	-11	9	-8	17	-78	44
G	-26	93	-73	220	-20	40	322	-204	-8	142	-31	63	0	0	18	167
	-48	-19	-124	-25	-9	-12	102	-216	-94	-41	-45	13	0	0	-116	-70
S	132	178	68	138	31	75	-3	100	433	-128	-8	-1	-6	8	101	-73
	-156	-152	-136	-69	-82	-24	-75	-23	-43	-257	-6	15	-1	-1	56	-83
B	-21	163	-54	39	23	35	-15	68	-26	-10	-15	0	-3	10	-80	-37
	-97	-45	-35	50	-47	-10	-45	-8	26	9	-16	31	-4	-3	103	13
I	-13	-10	-3	-4	-8	6	-8	-9	-5	-2	-3	-4	-3	-4	-3	-7
	8	16	-4	11	4	-3	8	9	2	6	-4	11	-10	16	3	7
C	311	455	-89	77	138	229	-42	243	192	-96	-59	-7	-5	21	-38	-192
	-412	-353	-1	12	-216	-150	-149	-53	92	-187	40	25	-3	-13	319	-91

Табела 4.21: Матрични приказ резултата поларних аминокиселина

	++	+-	-+	--
HX	135	-154	-76	99
XH	174	31	-199	-3
EX	-910	73	187	641
XE	-919	176	76	658
TX	314	-4	-104	-203
XT	441	-12	-131	-292
GX	251	-13	-11	-225
XG	233	-39	3	-194
SX	394	-10	-102	-278
XS	437	-20	-55	-358
BX	-84	35	35	12
XB	-68	6	17	44
IX	-10	-6	-5	20
XI	-5	0	-11	17
CX	149	115	50	-315
XC	-15	-176	296	-108

Табела 4.22: Груписани резултати поларних аминокиселина

Позитивне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	211	203	-5	13	117	103	-7	-32	10	77	-4	-2	-5	-12	13	82
	266	-412	-33	16	103	-189	38	-3	-27	-40	-10	10	1	10	-38	-36
E	-2	-23	-252	-124	-2	-13	-6	-2	-45	-83	-7	3	-1	-2	-54	-25
	6	13	-229	340	-14	21	6	-1	-9	82	-7	5	-2	4	-99	108
T	-12	-42	8	4	17	97	-9	-11	-8	-1	-5	-8	-5	4	-45	-58
	-12	44	127	-99	-34	-52	-3	13	-1	4	7	2	-9	6	-66	106
G	1	-2	-21	95	-7	1	13	-19	-7	32	-11	15	0	0	3	32
	26	-18	-73	-10	42	-30	48	-26	-19	-7	-14	3	0	0	-44	9
S	-42	-95	14	37	-18	-13	-9	-6	24	23	-13	10	-3	6	2	6
	32	60	-11	-24	-27	35	-25	26	4	-28	2	-4	-7	1	-79	54
B	-14	-21	-17	24	-17	-9	-2	33	-19	-23	-11	7	-1	-2	-25	34
	-17	32	-19	2	-12	21	-17	-11	3	21	-14	8	-2	4	-26	2
I	-3	-4	-1	-2	-6	-2	-2	-5	-3	-8	-1	-2	2	-3	-3	-4
	-4	6	-2	4	0	3	10	-3	2	6	-2	4	-5	5	5	0
C	-78	-181	9	20	-26	-76	-22	-7	-14	-109	-8	9	4	-1	58	-89
	2	157	78	-75	-25	82	-54	52	26	66	18	-17	-5	3	-56	89

Табела 4.23: Матрични приказ резултата позитивних аминокиселина

	++	+-	-+	--
HX	242	249	280	-466
XH	176	126	265	-344
EX	-263	-144	-252	375
XE	-228	-90	-168	262
TX	-15	32	-27	1
XT	43	75	2	-71
GX	4	19	25	-33
XG	-4	-25	19	5
SX	3	4	-55	36
XS	-11	-54	6	39
BX	-41	21	-36	24
XB	-20	12	9	-10
IX	-2	-7	-3	8
XI	0	-4	-9	10
CX	0	-190	-17	152
XC	5	-69	-161	168

Табела 4.24: Груписани резултати позитивних аминокиселина

Мале аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	-1287	-187	-42	17	-87	-166	-67	-53	-14	-78	-31	-19	-10	-1	21	-181
	-284	1488	-19	37	207	34	50	57	68	20	-3	46	-4	13	256	-86
E	35	204	-1112	-317	150	-60	27	104	185	-68	-11	3	13	-3	-68	-238
	-113	-112	-97	1293	222	-270	-51	-70	177	-251	10	-3	-3	-4	440	-127
T	-9	149	38	470	851	-200	-16	35	378	150	23	113	-11	-2	435	528
	-111	-28	-230	-250	285	-778	-14	-6	-156	-306	-65	-63	0	11	-305	-562
G	-89	34	-83	82	-63	-47	112	118	-96	45	-45	2	0	0	-48	29
	-63	100	-104	88	59	41	-41	-163	-62	95	-32	63	0	0	-30	41
S	234	296	31	301	265	-38	129	27	668	4	24	57	2	-1	525	-87
	-180	-299	-191	-126	76	-252	-1	-131	20	-572	-14	-59	4	-4	315	-640
B	-22	194	-55	3	99	-11	3	49	60	-20	-17	-11	-3	11	-68	-136
	-94	-72	-26	65	47	-114	-41	-9	57	-83	13	12	-3	-4	194	3
I	-7	-6	-3	11	-16	4	-1	9	-11	-1	-3	-3	0	-4	-8	-3
	5	7	-3	-4	6	4	-8	0	-1	12	-3	9	-6	9	8	2
C	529	602	-131	409	646	98	278	146	692	38	-19	116	-3	22	560	-165
	-401	-619	-302	10	-55	-571	-71	-297	18	-621	-46	-48	-5	-13	404	-678

Табела 4.25: Матрични приказ резултата малих аминокиселина

	++	+-	-+	--
HX	-1287	-273	-166	1459
XH	-1086	72	-446	1234
EX	-985	-383	139	1034
XE	-1073	21	-296	1134
TX	997	346	-93	-1050
XT	980	-209	367	-950
GX	3	128	-76	-50
XG	224	184	-69	-289
SX	910	114	132	-973
XS	1019	55	20	-908
BX	-14	-67	158	-71
XB	-12	151	-67	-67
IX	-9	-3	-3	13
XI	-5	1	-7	10
CX	1108	322	28	-1231
XC	791	-106	580	-1083

Табела 4.26: Груписани резултати малих аминокиселина

Аминокиселине склоне равни

S1\S2	H		E		T		G		S		B		I		C	
H	-228	-47	27	-1	-194	-102	-27	54	-31	22	44	13	19	0	-73	111
	-71	214	57	-59	-248	381	-55	15	-10	7	16	-47	5	-14	-118	44
E	3	-41	1760	789	-160	157	-53	-38	-62	223	36	27	-2	-3	98	402
	49	-8	648	-2073	-202	120	29	35	-145	-29	21	-57	-3	6	-66	-319
T	-16	-59	-132	-284	-449	-372	-23	4	-168	-236	-6	-71	22	-4	-296	-392
	78	-6	278	75	-434	879	11	0	-11	285	69	5	-4	-5	191	316
G	-11	-17	14	-101	-24	49	-234	-187	-7	-57	27	-5	0	0	-39	-62
	67	-34	253	-129	5	-30	-151	393	95	-26	59	-57	0	0	95	-6
S	-83	-66	-49	-185	-130	10	-63	-33	-258	-219	-14	-36	-5	1	-305	-61
	-21	113	312	-75	-162	190	-36	89	-156	435	83	-30	-4	5	-191	362
B	68	-8	76	8	-46	38	-1	-27	-2	71	45	12	-2	-3	62	187
	43	-64	72	-104	-63	44	38	-8	-36	-27	26	-54	12	-6	-26	-161
I	5	8	-2	-3	-2	13	-4	5	0	1	-2	-3	16	2	-1	22
	-4	-6	12	-6	-2	-8	-1	-1	-7	5	12	-6	4	-13	-12	-8
C	-208	-174	90	-156	-277	-164	-146	-140	-257	-18	44	-46	0	-7	-267	55
	-16	261	526	-342	-252	477	-48	226	-248	347	148	-105	7	0	-162	226

Табела 4.27: Матрични приказ резултата аминокиселина склоних равни

	++	+-	-+	--
HX	-292	-49	-160	321
XH	-291	-116	-55	289
EX	1620	949	514	-2020
XE	1631	570	937	-2061
TX	-585	-656	-130	934
XT	-618	-334	-636	1098
GX	-215	-199	-13	283
XG	-287	-220	-153	449
SX	-437	-256	-163	565
XS	-407	-177	-279	578
BX	68	183	-22	-164
XB	50	-68	193	-126
IX	14	11	0	-17
XI	24	-1	5	-16
CX	-458	-171	-105	462
XC	-428	61	-175	317

Табела 4.28: Груписани резултати аминокиселина склоних равни

Сумпорне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	29	13	2	15	19	31	13	24	13	1	12	18	-1	-6	13	0
	86	-74	14	-21	4	-28	13	-28	39	-31	16	-26	1	4	23	-18
E	-4	0	38	69	-16	2	-7	-25	-13	25	-3	9	0	-1	17	38
	-4	4	99	-125	-63	45	1	18	-52	20	33	-29	-1	2	82	-88
T	0	-5	-18	-83	-19	-128	-12	-11	-9	-67	10	-14	-1	-6	-18	-93
	21	-11	3	60	-127	184	-9	16	25	32	45	-23	13	-5	112	-10
G	-3	18	13	-11	4	21	-15	-71	8	-4	0	-2	0	0	20	-15
	-5	-9	27	-13	-4	-13	-33	77	57	-38	52	-35	0	0	68	-40
S	-11	-44	-17	-60	-3	-12	-7	-19	-16	-107	0	10	-1	9	-16	-8
	-54	72	23	30	-61	53	-21	29	-66	126	4	-10	9	-13	-53	46
B	-3	-12	-2	0	19	17	1	-9	43	27	26	7	0	-1	57	43
	19	-4	24	-17	-25	3	5	3	-9	-20	33	-32	-1	2	118	-123
I	-1	14	0	-1	-1	-8	-1	-3	-1	-4	0	-1	-4	3	-1	-4
	-4	-7	-1	2	3	4	12	-7	-4	6	-1	2	-5	2	-4	6
C	-31	-67	-5	44	-25	46	-12	-38	-22	46	-2	51	-1	-6	-27	128
	-83	111	14	-41	-129	62	-42	58	-114	51	32	-59	8	-2	-60	-45

Табела 4.29: Матрични приказ резултата сумпорних аминокиселина

	++	+-	-+	--
HX	39	24	95	-90
XH	18	-12	56	-34
EX	35	82	100	-135
XE	27	50	106	-114
TX	-31	-192	-19	155
XT	-19	-58	-184	174
GX	-2	-59	16	32
XG	-20	-82	-48	96
SX	-31	-105	-99	150
XS	-21	-56	-96	111
BX	70	47	94	-110
XB	7	43	65	-77
IX	-5	1	-4	3
XI	-5	0	2	-1
CX	-53	129	-161	30
XC	-12	77	43	-83

Табела 4.30: Груписани резултати сумпорних аминокиселина

Сићушне аминокиселине

S1\S2	H	E	T	G	S	B	I	C								
H	-393	-218	-26	6	126	-65	1	-41	14	-49	-5	-15	-8	-7	40	-124
	-164	476	-3	10	255	-205	86	-35	144	-80	-4	16	-10	16	310	-162
E	3	19	-464	-345	0	-90	-14	-38	92	32	-2	-29	-1	-3	-76	-223
	-13	-7	-323	729	162	-56	16	23	159	-189	29	1	13	-8	164	81
T	42	60	23	416	339	-100	26	-1	215	163	26	117	-4	-5	400	665
	-16	-54	-83	-267	628	-564	41	-42	254	-421	-13	-92	-2	7	-4	-695
G	-32	-34	-40	-10	-6	-61	41	51	-33	-21	-28	-23	0	0	-16	-36
	-20	57	-70	80	103	-29	-20	-43	-32	56	-16	43	0	0	20	20
S	50	224	-6	83	48	90	36	56	282	168	6	64	5	2	164	249
	-96	-122	-23	-44	-35	-65	-21	-44	125	-356	-38	-23	2	-6	-64	-218
B	-21	34	-20	-16	43	-4	-4	-17	38	20	1	-18	-1	-3	-38	-121
	-34	10	-14	32	70	-71	-10	23	52	-73	12	4	-3	5	38	81
I	1	-2	-1	13	-7	3	-4	20	-5	3	-1	-3	-13	-1	1	-7
	6	-3	-3	-8	-4	3	-7	-8	-6	5	-3	5	6	3	9	-2
C	-10	215	-89	41	-18	16	-21	-16	249	184	-33	46	-8	10	201	-66
	-177	-24	-190	156	-57	39	-45	57	95	-331	-85	46	-9	3	-130	57

Табела 4.31: Матрични приказ резултата сићушних аминокиселина

	++	+-	-+	--
HX	-342	-261	-13	370
XH	-377	-113	-229	438
EX	-441	-416	-174	659
XE	-463	-203	-395	675
TX	552	457	535	-1019
XT	312	-106	608	-530
GX	4	0	-9	4
XG	36	33	-15	-30
SX	310	385	-11	-432
XS	438	277	314	-658
BX	-10	-89	63	25
XB	-19	90	-81	1
IX	-15	1	5	2
XI	-16	-2	1	8
CX	205	125	-229	-16
XC	353	188	4	-312

Табела 4.32: Груписани резултати сићушних аминокиселина

Склоне ка заокрету

S1\S2	H		E		T		G		S		B		I		C	
H	-1324	-803	-31	-8	-159	-255	-64	-76	-55	-119	-24	-38	-6	5	-32	-199
	-976	2142	-37	53	256	87	14	84	102	43	-18	57	-3	2	350	-102
E	-7	203	-966	-748	159	36	16	160	108	-47	-17	8	-2	12	-144	-174
	-85	-91	-493	1517	357	-401	-48	-99	326	-282	0	2	-3	-6	397	-99
T	-20	219	33	695	970	3	-15	73	459	308	21	182	-7	-4	375	942
	-137	-54	-220	-398	709	-1098	-43	-15	-65	-445	-59	-109	-9	14	-193	-803
G	-84	38	-71	98	-54	-50	111	195	-73	34	-40	15	0	0	-37	61
	-96	92	-121	57	72	12	61	-265	-93	87	-57	56	0	0	-46	8
S	217	357	-8	423	282	48	144	79	698	205	15	92	-5	6	525	167
	-80	-340	-185	-185	196	-349	41	-175	268	-759	3	-84	11	-11	572	-876
B	-29	119	-43	-8	96	25	1	61	29	-10	-15	-17	-2	12	-92	-116
	-85	-11	-49	69	107	-158	-37	-19	85	-75	-6	26	-3	-6	141	31
I	-7	-8	-2	12	-10	-4	5	5	-7	1	-2	-3	2	-2	-7	-4
	-8	16	-3	-6	0	8	-7	-1	1	2	-3	6	-3	3	7	2
C	466	768	-182	412	674	316	307	274	596	185	-42	109	3	27	545	96
	-175	-728	-338	40	173	-759	30	-411	341	-746	-33	-38	-7	-18	656	-898

Табела 4.33: Матрични приказ резултата аминокиселина склони ка заокрету

	++	+-	-+	--
HX	-1357	-911	-812	2116
XH	-1146	-469	-1050	1838
EX	-894	-729	-186	1219
XE	-963	-305	-675	1308
TX	1081	850	304	-1512
XT	1057	49	874	-1314
GX	18	204	-12	-163
XG	237	324	55	-432
SX	918	490	517	-1305
XS	984	328	450	-1158
BX	-44	-53	144	-48
XB	-33	186	-59	-84
IX	-4	-4	-4	8
XI	-1	7	-6	1
CX	1027	733	515	-1557
XC	715	399	922	-1444

Табела 4.34: Груписани резултати аминокиселина склони ка заокрету

Ненаелектрисане аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	-346	-389	-6	-1	29	-46	-28	-34	68	54	-14	-10	-1	3	55	64
	-289	658	-26	24	126	-72	-2	39	13	-79	-2	15	0	-2	5	-75
E	47	177	-357	-185	40	47	1	7	105	59	-4	-15	-1	-3	19	-103
	-65	-104	-142	395	36	-79	-28	15	31	-111	20	-2	-3	5	212	-92
T	-28	38	58	44	141	-45	-31	4	119	165	18	6	-7	-1	81	88
	-53	23	-23	-40	186	-165	-30	33	-12	-166	-8	-6	-1	4	-36	-73
G	-32	20	-26	72	-18	-7	51	-51	-3	55	-23	10	0	0	15	52
	-45	32	-71	10	-33	37	103	-60	-36	-14	-24	20	0	0	-31	-23
S	82	361	28	32	27	91	7	92	255	103	13	6	-4	3	128	14
	-119	-218	-36	-8	-74	-24	-60	-27	168	-311	17	-23	-2	1	164	-188
B	-3	215	-27	4	23	50	-1	33	37	17	-3	-3	-1	14	-11	-62
	-66	-113	-25	27	-26	-27	-31	-2	14	-38	-5	8	-3	-8	78	-8
I	-5	-10	-1	-3	-9	-7	-3	-6	-5	3	-1	-3	-1	-13	-5	-2
	-2	11	-3	5	-2	11	0	6	-5	3	-3	5	-16	23	6	-1
C	181	759	-12	45	51	221	-27	135	243	196	-2	-1	-7	21	178	-66
	-271	-447	-13	-20	-131	-90	-130	7	60	-295	37	-27	-8	-7	354	-293

Табела 4.35: Матрични приказ резултата ненаелектрисаних аминокиселина

	++	+-	-+	--
HX	-315	-378	-246	604
XH	-280	-91	-392	483
EX	-292	-174	-47	288
XE	-320	-134	-156	352
TX	196	94	94	-223
XT	156	83	118	-217
GX	30	3	37	-43
XG	23	30	8	-38
SX	280	205	147	-383
XS	399	287	143	-491
BX	7	30	31	-49
XB	3	1	32	-26
IX	-6	-16	-16	27
XI	-5	-6	-17	19
CX	303	399	137	-532
XC	242	-39	419	-389

Табела 4.36: Груписани резултати ненаелектрисаних аминокиселина

Огромне и ароматичне аминокиселине

S1\S2	H		E		T		G		S		B		I		C	
H	53	-38	41	10	-22	-46	-5	50	28	23	33	34	9	6	19	79
	-43	46	64	-64	-106	115	-29	-14	42	-54	33	-57	6	-11	8	-68
E	-9	-56	371	576	-40	29	-13	-26	-21	111	16	18	-1	-2	-16	178
	37	16	527	-891	-125	80	0	22	-94	-7	12	-26	-2	3	-79	-68
T	17	17	-28	-127	-96	-213	17	41	-36	-105	30	5	-3	13	-68	-179
	102	-90	72	47	-184	312	28	-55	-47	120	63	-56	9	-15	65	99
G	20	69	28	-5	27	154	-1	-56	34	48	33	73	0	0	48	80
	30	-77	73	-57	38	-147	92	-26	79	-100	31	-84	0	0	50	-106
S	-45	-115	-10	-88	-37	6	-18	-27	-83	-183	-1	-8	-2	2	-98	-72
	-9	101	181	-65	-89	69	-18	37	-125	244	48	-29	-5	3	-173	202
B	-13	-24	31	31	3	19	-5	-11	8	64	10	21	-1	-2	14	132
	27	1	21	-45	-30	7	12	1	-25	-30	14	-28	20	-13	-20	-84
I	15	-1	-1	-2	5	15	-1	-4	16	11	-1	-2	7	7	-2	9
	-1	-2	-2	3	9	-19	-4	7	-1	-11	-2	3	11	-15	-2	-5
C	-97	-267	-10	-54	-85	-125	-45	-92	-96	-86	2	-23	-3	-5	-153	-67
	-4	220	174	-85	-173	237	-14	88	-211	239	97	-55	7	-1	-240	261

Табела 4.37: Матрични приказ резултата огромних и ароматичних аминокиселина

	++	+-	-+	--
HX	55	-29	-64	54
XH	22	-125	-24	102
EX	325	621	419	-834
XE	337	471	613	-869
TX	-117	-311	-69	304
XT	-133	-197	-309	399
GX	42	50	144	-151
XG	-23	-79	70	13
SX	-141	-219	-154	305
XS	-117	-155	-229	308
BX	19	135	-21	-87
XB	27	5	132	-106
IX	12	14	12	-21
XI	7	10	15	-20
CX	-228	-258	-250	425
XC	-180	-34	-260	257

Табела 4.38: Груписани резултати огромних и ароматичних аминокиселина

5 Дискусија

Уређене таблице корелација, на основу критичне вредности, јасно показују одређене зависности. За аминокиселине које се класификују као **алифатичне** уочава се склоност да граде структуре α хеликса и β траке. Тачније, довољно је да бар једна аминокиселина биграма буде алифатична да би постојала тенденција ка изградњи било које од ове две структуре. Оваква зависност се посебно истиче у груписаним подацима приказаним у табели 4.2. Када посматрамо појединачне резултате (табела 4.1), уочава се да припадност обе аминокиселине класи повећава шансе за изградњу парова структура НН и ЕЕ. Уколико ниједна аминокиселина није алифатична показује се одступање од изградње НН и ЕЕ али показује тенденцију ка изградњи заокрета, кривине премошћеној водоничним везама и Z_{10} хеликса. Може се десити и да су обе без препознативе структуре или на граници, тачније у СН, СТ, СГ и СС као и ТС и СС

Парови аминокиселина, где су неке класификоване као **амидне** имају утицаја на изградњу β траке и кривине премошћене водоничним везама. Случај када ниједна аминокиселина није амидна увећава могућност да се нађу у β равни, док случај када је бар једна аминокиселина амидна умањује ту могућност. Супротан сценарио важи за кривину премошћену водоничним везама. Овакве зависности се уочавају се у табели 4.4 са груписаним подацима док су све зависности приказане у табели 4.3.

Код **наелектрисаних** аминокиселина најочљивије су зависности са α хеликсом и β траком. Када погледамо груписане резултате (табела 4.6) види се да уколико је бар једна аминокиселина наелектрисана велика је вероватноћа да ће се наћи у некој од комбинација са α хеликсом, а јако ретко у некој од комбинација са β траком. Анализом појединачних резултата (табела 4.5) читава се тежња ка изградњи НН а одбијање изградње ЕЕ. Комбинације аминокиселина где ниједна није наелектрисана показује зависност ка изградњи ЕЕ, док се може сматрати да ниједна аминокиселина неће изградити Н.

Из појединачних резултата корелације уочљива је и склоност ка изградњи парова структура ТТ, ГГ и СС али само у случају када су обе аминокиселине из биграма наелектрисане, док у случају када ниједна аминокиселина није наелектрисана, до таквих структура неће доћи.

Аминокиселине које су **склоне хеликсу**, показују готово сигурну изградњу НН док је могућност изградње комбинација са структуром β траке, кривином премошћеном водоничним везама, кривином врло мала као и могућност да се нађе у комбинацији без препознативе структуре. Када ниједна није склона хеликсу важе потпуно супротне зависности. Уколико је прва аминокиселина у пару склона хеликсу, обе ће бити у некој структури, док, у случају када је друга, независно од прве, склона хеликсу, може доћи и до комбинација без препознативе структуре. Овакве зависности су уочене у табелама 4.7 и 4.8.

Посматрањем резултата корелација за **хидрофилну** класу аминокиселина (табела 4.9) уочавају се зависности у случајевима када су обе аминокиселине из биграма хидрофилне и оних када ниједна аминокиселина није хидрофилна. За први случај из групних резултата (табела 4.10) се види да постоји тежња ка изградњи α хеликса, кривине премошћене водоничним везама, Z_{10} хеликса и кривине. За случај када ниједна аминокиселина није хидрофилна, вероватноћа да се изграде ове структуре је мала.

Резултати корелација за β траку показују другачије особине хидрофилних, или делимично хидрофилних парова аминокиселина. У случају када је бар једна аминокиселина хидрофилна, β раван се вероватно неће појавити ни у једној комбинацији. Из таблице појединачних резултата, примећује се да је та комбинација управо ЕЕ. Одсуство хидрофилних аминокиселина у биграмама даје значајно висок степен зависности ка изградњи ЕЕ па се за овакав пар аминокиселина изградња ЕЕ може сматрати сигурном. Даљом анализом појединачних резултата уочава се и тежња биграма аминокиселина, где је друга хидрофилна, да гради парове НН, ТТ и GG, као и тежње да се овакви парови секундарних структура не изграде уколико ниједна од аминокиселина није хидрофилна.

За **хидрофобне** парове аминокиселина, тачније, за парове где су обе аминокиселине хидрофобне, из груписаних резултата (табела 4.12) уочава се тежња ка изградњи α хеликса и β равни, док су јако мале шансе за изградњу кривине премошћене водоничним везама, Z_{10} хеликса и кривине. Такође, мале су шансе да ови парови аминокиселина буду без препознативе структуре. У случајевима када ниједна аминокиселина из биграма није хидрофобна, вероватно се неће изградити ни НН ни ЕЕ, док је највероватније да то буду ТТ, SS, TS, CS или SC а у нешто мањој мери и GG. Присуство хидрофобне аминокиселине у биграму значајно повећава могућност изградње неких од парова са β траком а умањује могућност парова са кривином. То се посебно уочава у појединачним резултатима приказаним у табели 4.11.

Код **хидроксилних** аминокиселина једина зависност која се може уочити из табеле груписаних резултата (табела 4.14) је при изградњи парова са α хеликсом, кривином и комбинацијама без препознативе структуре. Биграма аминокиселина, где је бар једна хидроксилна, имају малу шансу за изградњу НН, али већу за изградњу SS. Прецизније, SS и SH имају шансу да се изграде када је друга аминокиселина из биграма хидроксилна. Овакав пар аминокиселина може да изгради и СН што се уочава у табели 4.13. Парови аминокиселина где ниједна није хидроксилна у већој мери граде НН али јако слабо SS и SH. За овакав случај парова аминокиселина мале су шансе да буду без препознативе структуре, док присуство бар једне хидроксилне аминокиселине може повећати шансу за такав случај.

Анализом резултата корелација аминокиселина које спадају у **велике** и парова секундарних структура може се закључити да велике аминокиселине имају склоност ка изградњи парова структура састављених од α хеликса и β траке а немогућност изградње секундарних структура састављених од кривине премошћене водоничним везама и кривине (табела 4.16). Такође, велике аминокиселине смањују шансе да буду без препознативе структуре. У случајевима када је бар једна аминокиселина из биграма велика, постоји велика склоност ка изградњи НН и ЕЕ. Парови аминокиселина где ниједна није велика овакве структуре вероватно неће изградити, али је могуће да се изграде СН, ТС и SC. Појединачни резултати (табела 4.15) показују да овакви парови аминокиселина могу изградити и парове структура XT и XS. Тачније, ET, TT, ST, CT, TS, SS и CS, док су за изградњу ових парова структура мале шансе уколико су обе аминокиселине из биграма велике.

Добијене зависности за **негативне** аминокиселине показују значајне склоности парова аминокиселина. Парови, где ниједна аминокиселина није негативна, имају тежњу ка изградњи парова структура са β траком, али и тежњу да не граде α хеликс, кривину премошћену водоничним везама и Z_{10} хеликс. Из појединачних резултата (табела 4.17) може се уочити да, уколико је бар једна аминокиселина негативна, постоји тенденција ка изградњи НН, а уколико су обе негативне могу се изградити ТТ и GG. За β траку се уочава супротна ситуација. ЕЕ или СЕ могу да се изграде само у случају када ниједна

аминокиселина није негативна. Груписани резултати су у табели 4.18.

Аминокиселине **без склоности** не показују неке значајне зависности. Уочљива је ситуација да парови аминокиселина, где ниједна није у овој класи, повећавају могућност стварања неке од комбинација са α хеликсом (табела 4.20), и то конкретно НН. Табела појединачних резултата (табела 4.19) показује и особину да су, уколико је једна аминокиселина без склоности а друга није, независно од редоследа, мале шансе да се изгради НН.

Поларне аминокиселине показују јасне склоности ка одређеним паровима секундарних структура. Анализом груписаних података (табела 4.22) примећује се да биграма аминокиселина, где су обе аминокиселине поларне, имају склоност да граде парове структура са кривином премошћеном водоничним везама, Z_{10} хеликсом и кривином. Из појединачних резултата (табела 4.21) може се уочити склоност оваквих парова аминокиселина да граде НТ, ЕТ, ТТ и GG. Такође се уочава да, уколико друга аминокиселина није поларна, структуре ТТ и GG се вероватно неће изградити. Биграма где ниједна аминокиселина није поларна показују склоност да не граде парове секундарних структура са Т, G и S, осим у случају β траке, где се вероватноћа за изградњу у овом случају повећава.

Када је бар једна класа у биграму **позитивна** и из појединачних и из груписаних резултата (табеле 4.23 и 4.24) видљива је тенденција ка изградњи НН. Уколико ниједна није позитивна постоји одступање од изградње НН. Парови аминокиселина, где је једна, тачније друга позитивна, показује мале шансе ка изградњи ЕЕ. Аминокиселине, где ниједна није позитивна вероватно ће изградити управо ЕЕ или ће бити без препознативе структуре.

Анализом резултата за класу **малих** аминокиселина уочава се склоност парова аминокиселина, где ниједна није мала, да граде парове структура са α хеликсом и β траком (табела 4.26). Овакви парови аминокиселина ретко граде структуре са кривином премошћеним водоничним везама, кривином и ретко су без препознативе структуре. Уочавају се и мале шансе за изградњу парова структура где је друга Z_{10} хеликс. Табела појединачних резултата (табела 4.25) показује тежњу парова аминокиселина, где је друга мала, да не граде НТ, ЕТ, ТТ и СТ.

Из резултата је уочљиво и да парови аминокиселина, од којих неке могу бити **склоне равни**, имају тенденцију изградње парова структура састављених од β равни. Највеће су шансе да се изгради ЕЕ (табела 4.27). Када је само једна од аминокиселина у пару склона равни да се изграде структуре могуће је да ће се изградити нека од структура ХЕ, али неће градити ХТ, док се овакве структуре граде када ниједна аминокиселина у пару нису у овој групи. Груписани резултати аминокиселина склоних равни су у табели 4.28.

Парови аминокиселина, где ниједна није **сумпорна**, показује тежњу ка изградњи ТТ. Резултати су приказани у табелама 4.29 и 4.30.

Код груписаних резултата парова **сићушних** аминокиселина (табела 4.32), уочава се тежња да се, уколико бар једна аминокиселина припада класи, не изграде структуре са α хеликсом и β траком. До изградње ових парова секундарних структура, тачније НН и ЕЕ може доћи код парова аминокиселина где ниједна није сићушна. Парови где је бар једна аминокиселина сићушна имају тенденцију да граде парове структура са кривином премошћеном водоничним везама и кривином конкретно TS а такође, у одређеним

случајевима може бити без структуре. Појединачни резултати су приказани у табели 4.31.

Из резултата зависности код парова аминокиселина **склоних ка заокрету** јасно се уочава тежња парова аминокиселина, где је бар једна склона ка заокрету да овакве структуре и изгради. Уколико су обе аминокиселине из ове групе постоји велика шанса да буду у некој од комбинација без препознативе структуре, док парови аминокиселина, где ниједна није из ове групе, искључују такву могућност. Појединачни резултати су приказани у табели 4.33 а груписани резултати у табели 4.34.

Код парова аминокиселина где је бар једна **ненаелектрисана** вероватно неће доћи до изградње неке од комбинација структура са α хеликсом и β траком (табела 4.36). За парове аминокиселина где ниједна није из ове групе, врло је вероватно да ће се изградити парови структура са α хеликсом и β траком, највероватније НН и ЕЕ (табела 4.35). Када је прва аминокиселина ненаелектрисана, или у неким случајевима и обе, уочава се вероватноћа да такав пар формира пар структура где је прва С. Ово важи за парове структура СН, СS а уколико су обе ненаелектрисане, могућа је изградња СС.

Групни приказ резултата **огромних и ароматичних аминокиселина** (табела 4.38), показује јасну тежњу парова аминокиселина, где је бар једна из ове групе, да граде секундарне структуре у комбинацији са β траком. Углавном ЕЕ. Парови аминокиселина, где ниједна није из ове групе показује одбојност ка ЕЕ, али се зато повећава шанса за изградњу ТТ, SS, SC и СС. Ово се види у табели појединачних резултата (табела 4.37)

Анализом резултата истичу се изузетно високе зависности парова аминокиселина одређених класа да изграде одређене секундарне структуре. То посебно важи са случајеве када обе аминокиселине у пару граде исту секундарну структуру.

НН ће се, на пример, вероватно изградити од парова аминокиселина где су обе алифатичне, склоне хеликсу и хидрофобне а готово сигурно неће од парова аминокиселина где ниједна није из ових класа. Неће се формирати ни у случају да ниједна аминокиселина није велика. Са друге стране, резултати показују да парови аминокиселина где ниједна није мала или склона заокрету врло вероватно формирају НН. Уочава се и висок степен склоности парова аминокиселина где су обе хидрофобне или склоне равни да се нађу у ЕЕ. Ту ће се наћи и парови аминокиселина где ниједна није хидрофилна, мала или склона заокрету. Са друге стране уочава се да парови аминокиселина, где ниједна није алифатична, хидрофобна или склона равни, вероватно неће градити ЕЕ.

Резултати показују и изузетно висок степен склоности парова аминокиселина где ниједна није хидрофобна да формирају ТТ. У случајевима када је само прва аминокиселина хидроксилна, могуће је да ће се наћи на "почетку" неке структуре, тачије у СН, док у случају да је само прва аминокиселина склона заокрету, велика вероватноћа је да ће се наћи у ТС.

6 Закључак

Дугачки низови аминокиселина, које чине протеин, у простору се извијају и праве одређене облике. Тачан распоред аминокиселина у ланцу и њихов просторни облик одређују функцију протеина. Како би се утврдила функција протеина у организму, потребно је наћи механизме за предвиђање просторног облика протеина.

Аминокиселине, класификоване на основу својих хемијских састава, у ланцу протеина везане су са аминокиселинама истих или других класа, и као такве граде секундарне структуре. Бројањем парова аминокиселина у секундарним структурама које граде и упоређивањем са начином на који су аминокиселине у пару класификоване, добијене су одређене зависности. Резултати ових зависности, са издвојеним значајним вредностима, јасно показују да парови аминокиселина, истих или различитих класа, показују склоност да изградње или не различите комбинације секундарних структура, што може да олакша предвиђање секундарне структуре протеина и његове функције.

- **Прилог 2: Документ класификације аминокиселина**

```
Polar|Y,R,K,H,E,D,Q,N,T,S
Positive (Basic) |R,K,H
Charged|R,K,H,E,D
Negative (Acidic) |E,D
Hydrophillic|R,K,E,D,Q,N
Amide|Q,N
Uncharged|Q,N,T,S
Hydroxylic|T,S
Tiny|C,A,G,S
Small|N,D,T,S,C,A,G,P
Aliphatic|I,V,L,A
Hydrophobic|F,W,I,L,M,V,C,A
Large|Y,R,K,M,W,F,L,I
Very Large & Aromatic|Y,W,F
Sulphur Containing|M,C
Helix-pref|A,L,E,Q,R,M,K
Strand-pref|V,I,Y,F,T,W
Turn-pref|G,N,P,D,S
Noss-pref|C,H
```

Документ који садржи називе класа и аминокиселине које тој класи припадају.

- **Прилог 3: Листа аминокиселина**

```
A|Alanine
R|Arginine
N|Asparagine
D|Aspartic acid
C|Cysteine
Q|Glutamine
E|Glutamine acid
G|Glycine
H|Histidine
I|Isoleucine
L|Leucine
K|Lysine
M|Methionine
F|Phenylalanine
P|Proline
S|Serine
T|Threonine
W|Tryptophan
Y|Tyrosine
V|Valine
```

Документ *aminokiseline.txt* који садржи свих двадесет аминокиселина и њихове једнословне ознака из ког се пуни листа аминокиселина ListOfAA.

- **Прилог 4: Рачунање статистике**

```
private void IzracunavanjeStatistike()
{
    int numOfSequenceData = primarySecondaryStructList.Count;

    for (int rowCount = 0; rowCount < numOfSequenceData; rowCount++)
    {
        PDBSequence sequenceInProgress = primarySecondaryStructList[rowCount];

        List<char> primarySec = sequenceInProgress.PrimaryStruct.ToList<char>();
        List<char> secondarySec = sequenceInProgress.SecondaryStruct.Replace(' ', 'C').ToList<char>();

        if (primarySec.Contains('Z')) continue;
        else
        {
            int lenghtOfStruct = primarySec.Count;
            for (int i = 0; i < lenghtOfStruct - 1; i++)
            {
                string aa = PomocneMetode.KreirajBigramAA(primarySec[i], primarySec[i + 1]);
                string ss = string.Format("{0}{1}", secondarySec[i], secondarySec[i+1]);

                if (!aa.Contains("X"))
                {
                    dictSS["SUM"][aa]++;
                    dictSS[ss][aa]++;
                }
            }
        }
    } //end second else
} //end for
}
```

Метод који рачуна статистику појављивања и резултате бележи у dictSS.

- **Прилог 5: Класификација**

```
public string BigramInClass(string aa)
{
    string bigramContainedInClass = string.Empty;

    if (aa.Count() != 2)
        throw new ArgumentException();
    else
    {
        if (_clanoviKlase.Contains(aa[0].ToString())) bigramContainedInClass += "+";
        else bigramContainedInClass += "-";

        if (_clanoviKlase.Contains(aa[1].ToString())) bigramContainedInClass += "+";
        else bigramContainedInClass += "-";
    }
    return bigramContainedInClass;
}
```

Метод BigramInClass који, за пар аминокиселина aa, класификује исти у један од могућих односта аминокиселина и класе.

- Прилог 6: Корелација – груписано

```
public static Dictionary<string, Dictionary<string, int>> GrupisaniPodaci(Dictionary<string, Dictionary<string, int>> dictStatistika,
                                                                    List<string> BigramsOfAA)
{
    Dictionary<string, Dictionary<string, int>> statistikaGrupisano = new Dictionary<string, Dictionary<string, int>>();

    foreach (string s in secondaryStructures)
    {
        statistikaGrupisano.Add(s + "X", InitDictOfAA(BigramsOfAA).ToDictionary(p => p.Key, p => (int)p.Value));
        statistikaGrupisano.Add("X" + s, InitDictOfAA(BigramsOfAA).ToDictionary(p => p.Key, p => (int)p.Value));
    }

    foreach (var kvss in dictStatistika)
    {
        if (kvss.Key == "SUM") continue;
        char[] sstochar = kvss.Key.ToCharArray();
        foreach (var kvaa in kvss.Value)
        {
            statistikaGrupisano[sstochar[0] + "X"][kvaa.Key] += kvaa.Value;
            statistikaGrupisano["X" + sstochar[1]][kvaa.Key] += kvaa.Value;
        }
    }

    return statistikaGrupisano;
}
```

Метод `GrupisaniPodaci` који иницијализује и попуњава дневник `statistikaGrupisano` на основу ког се рачунају корелације за груписане податке

8 Референце

- [1] Peter Clote, Rolf Backofen, **Computational Molecular Biology: An Introduction**, Wiley 2000, ISBN 978-0471872528
- [2] Wolfgang Kabsch, Christian Sander, **Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features**, *Biopolymers*, 22(12), 2577–2637, December 1983
- [3] Wolfgang Kabsch, Christian Sander, **Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features**, *Biopolymers*, 22(12), 2577–2637, December 1983
- [4] Engelbert Buxbaum, **Fundamentals of Protein Structure and Function**, Springer 2007, ISBN 978-3-319-19920-7.
- [5] David P. Clark, **Molecular Biology**, Academic Press 2005, ISBN 978-0080454214
- [6] John T. Brosnan, Margaret E. Brosnan, **The Sulfur-Containing Amino Acids: An Overview**, *The Journal of Nutrition*, 136(6), 1636S-1640S, Jun 2006
- [7] James Morris, Daniel L. Hartl, Andrew H. Knoll, Robert Lue, **Biology: How Life Works**, W. H. Freeman 2015, ISBN 978-1319056919
- [8] Gregory A. Petsko, Dagmar Ringe, **Protein Structure and Function**, New Science Press 2004, ISBN 978-1405119221
- [9] Matthijs Joost Warrens, **Similarity Coefficients for Binary Data**, 2008, ISBN 978-90-8891-0524
- [10] Margaret Lewis, **Applied Statistics for Economists**, Routledge 2012, ISBN 9781136668715