

**Югозападен университет „Неофит Рилски”
Природо-математически факултет
Катедра „Информатика”
Докторска програма: „Информатика”**

**АВТОРЕФЕРАТ
НА
ДИСЕРТАЦИЯ**

**на Иван Благоев Тодорин
за придобиване на образователната и научна
степен „доктор”
на тема:**

**IN SILICO ПРЕДСКАЗВАНЕ НА ТРИИЗМЕРНАТА
СТРУКТУРА НА ПРОТЕИНИТЕ ЧРЕЗ ЕВРИСТИЧНИ
АЛГОРИТМИ**

Научен ръководител: доц. д-р Иван Тренчев

Благоевград, 2019

Дисертацията съдържа 137 страници, от които 130 основен научен текст и 7 страници приложения. В дисертационния труд са включени 1 таблица и 11 фигури. Библиографията съдържа 50 заглавия, от които 47 на латиница, 3 на български език.

Докторската дисертация е обсъдена на заседание на катедра „Информатика“ на 25.06.2019 година.

Официалната защита на дисертационния труд ще се състои на 17.09.2019 година от 11 часа по Заповед на Ректора на ЮЗУ „Н. Рилски“ №1346/дата 23.07.2019

Научното жури е в състав:

проф. д-р Нели Димитрова

проф. д-р Румен Андонов

проф. д-р Аврам Ескенази

проф. д-р Никола Янев

доц. д-р Татяна Дзимбова

Рецензенти:

1. проф. д-р Никола Янев

2. проф. д-р Румен Андонов

Автори на становища:

1. доц. д-р Татяна Дзимбова

2. проф. д-р Нели Димитрова

3. проф. д-р Аврам Ескенази

Съдържание на дисертационния труд:

1. Увод.....	4
2. Цели и задачи.....	6
3. Състояние на проблема за предсказване триизмерната структура на протеините според водещите световни изследвания.....	8
4. Създадени модели, техни модификации, алгоритми и работа с програмирания по тях софтуер	38
4.1 Модел за предсказване на триизмерната форма на белтъците с целочислени координати, с променливо ограничение за разпростиране в пространството	38
4.2 Модел за предсказване на триизмерната форма на белтъците с нецелочислени координати, с променливо разпростиране и трикомпонентна оценъчна функция.....	61
4.3. Модификация на модела с разполагане на аминокиселините чрез два псевдослучайни ъгъла	93
4.4. Модел за предсказване на триизмерната форма на белтъците с начално намиране на псевдослучайни структури и последващо целенасочено нагъване	104
5. Анализ на областта на приложимост на модела и алгоритъма, показали най-добри резултати при изследването върху един и същ протеин чрез изследването на други протеини.....	116
6. Заключение	119
7. Приноси	122
8. Литература	125
9. Приложения	131

1. Увод

Структурната биоинформатика е един от ключовите изследователски области в полето на изчислителната биология. Структурната биоинформатика е посветена на анализ и прогнозиране на триизмерната (3-D) структура на биологичните макромолекули като протеини, РНК и ДНК. Реалната структура се получава чрез кристалография (рентгенова дифракция), електронна микроскопия или ядрено-магнитен резонанс (NMR). Един от основните проблеми в структурните изследвания в биоинформатиката е предвиждането на триизмерни протеинови структури.

Протеините са дълги последователности, образувани от 20 различни аминокиселини, които в полярната клетъчна среда приемат уникална 3D структура. Познаването на структурата на протеина позволява изследването на биологични процеси и анализ на биологичните свойства, тъй като формата на протеина в клетъчната среда определя биологичната му активност – молекули с комплементарна форма се свързват с протеините. Само една малка част от протеинови последователности са с известни триизмерни структури. През последните 20 години няколко изчислителни методики, системи и алгоритми бяха предложени като решение за прогнозиране на триизмерната структура.

В моделите, предмет на това изследване, се използва само първоначалната структура на протеина, т.е. първичната структура – аминокиселинната последователност, и с различни математически модели се получава 3D структурата. Най-разпространеният математически модел, или по-скоро основен модел за множество от модели, е HP модела. При него се въвеждат структурни опростявания в описанието на протеина, с цел да се намали изчислителната сложност и да се постигнат резултати за приемлив период от време. Приема се, че аминокиселините са единен обект в пространството и се

разделят само на два вида – хидрофобни (Н) и полярни (Р). За по-добра получена структура се смята тази, която има най-много съседни в пространството Н аминокиселини. При разработването в този труд модели е намалена степента на опростяване, по различни начини, с цел да се установи доколко това е целесъобразно и кои точно модификации са най-подходящи. Докато в класическите НР модели размера на решетката е фиксиран, в разработените модели ограничението на пространството се променя в зависимост от степента на трудност за нагъване на молекулата при досегашното ограничение, като изменението може да става при различни параметри – това е едно от основните изследвани положения – как точно е най-добре да се изменя то. Оценъчната функция също отчита повече енергийни влияния, като за нея са изследвани различни варианти. В главите, посветени на отделните модели, са разгледани показалите по-добри резултати функции.

За да се направи съпоставителен анализ между разработваните модели в различните им модификации и при различни параметри, са поместени получените данни отначало за един и същи протеин, както и данните за структурата му, получени чрез кристалографски анализ, а след това са изследвани и други протеини с вече усъвършенствания алгоритъм.

2. Цели и задачи

Изложените в увода данни оправдават необходимостта от разработването на различни математически модели за предсказване на пространствената структура на протеините.
Ето защо целите на настоящата разработка са:

1. Създаването на модели и алгоритми, които да подобрят скоростта на компютърното предсказване на структурата и да позволят работата с дълги

последователности – чрез иновативни алгоритми за процеса на формиране на структурите;

2. Да се доближи степента на подобие на създаваните структури до възможните природни структури – чрез по-правдоподобни модели за представяне.

За постигането на тези цели си поставихме следните задачи:

1. Да се изследва надеждността на НР модела за предсказване на пространствената структура на протеините и изследване на възможностите за модифицирането му, с цел постигането на по-добри резултати;
2. Да се разработят математически модели за предсказване 3D структурата на протеините, насочени към постигането на целите на изследването;
3. Да се създадат компютърни програми, на базата на разработените модели, и да се извършат *in silico* експерименти с тях;
4. Да се направи сравнителен анализ между реалната структура, получена чрез ЯМР, и получените чрез *in silico* експерименти 3D структури, с цел валидиране на моделите и утвърждаване на по-добрите им модификации.
5. Да се направи сравнителен анализ на получените резултати от разработените модели и получените резултати от други автори по критериите определени в целите на изследването – степен на подобие, работа с дълги последователности и скорост на предсказване.

3. Състояние на проблема за предсказване триизмерната структура на протеините според водещите световни изследвания

Въведение в теорията на проблема и в методите за решаването му

Универсалността на всички живи организми се изразява във факта, че кодът който трансформира нуклеотидните последователности в ядрото в последователности от аминокиселини в протеините, е един и същ във всички организми: бактериите; растенията; животните, човека.

Белтъците се конструират от 20 различни аминокиселини. Те имат първична, вторична и третична структура. Първичната структура е линейната последователност от аминокиселините. Вторичната структура е създаването на алфа спирали и бета листове чрез водородни връзки между аминокиселини от различни части на веригата (несъседни в първичната структура), третичната е пространствената ориентация на страничните вериги на хидрофобните аминокиселини към пространствено съседство в структурата на протеина, което по-долу ще дефинираме като контакт между хидрофобните аминокиселини. Много от протеините имат и четвъртична структура, която се получава от дисулфидни мостове между аминокиселини Цистеин, които могат да бъдат част от една и съща пептидна верига или от друга, когато протеинът е многоверижен. 3D структурата на белтъка определя биологичните му свойства, затова предсказването на пространствената форма на протеините е много важна задача. В световен мащаб тази задача не е решена.

Един от методите за определяне на пространствената структура е синтезиране на протеина в химична лаборатория и след това определяне на 3D структурата с кристалографски анализ. Това изследване е много скъпо, затова се разработват математически подходи.

Целта на нагъването на протеините е да се предвиди компактната тримерна структура на даден протеин, базирана на неговата аминокиселинна последователност.

В литературата се представя опростен модел на нагъването на протеините в пространството (HP модел). Този модел е бил предложен от Ken Dill през 1985 г.

В модела всички видове аминокиселини са класифицирани като хидрофобни (H) или полярни (P). Нагъването на една протеинова секвенция е дефинирана като self-avoiding walk в 2D или 3D решетка. HP моделът може да бъде представен в двумерно и тримерно пространство, обикновено чрез квадратични и кубични решетки.

Идеята е получена от наблюдението, че хидрофобните взаимодействия между аминокиселинните остатъци са главната сила, която предизвиква сгъването на протеините в тяхното природно състояние.

Методът на *HP* нагъването на протеини в решетка се основава на наблюдението, че в полярна среда пептидите се нагъват по начин, в който има повече хидрофобни аминокиселини в ядро – при пространствено съседство – във физически **контакт** между тях – и повече полярни аминокиселини в контакт с полярната среда – формирайки водородни връзки с нея. Такава форма на сгънатия пептид е с минимална енергия и е по-стабилен, така че ние можем да очакваме този случай на 3D структура да бъде реалния случай .

HP нагъването в решетка се различава от другите подходи за нагъване в две положения:

1. докато повечето подходи разчитат на цялостната азбука на аминокиселините (20 букви), *HP* нагъването ползва симплифицирана двусимволна азбука, в която всяка аминокиселина е или хидрофобна „h” или полярна „p” ;
2. пространството, в което секвенцията ще се нагъва, е дискретизирано в 2D или 3D решетка.

4. Създадени модели, техни модификации, алгоритми и работа с програмирания по тях софтуер

4.1. Модел за предсказване на триизмерната форма на белтъците с целочислени координати, с променливо ограничение за разпростиране в пространството

Описание на разработения математически модел

В тази глава ще разгледам разработения модел за предсказване на триизмерната форма на белтъците, разновидност на НР модела, при който:

1. Аминокиселините се разполагат с триизмерни целочислени координати, започвайки от $(0,0,0)$ и $(0,0,1)$, следвайки номерата в аминокиселинната последователност, като на случаен принцип се избира позицията на всяка следваща от възможните 5 посоки – по x, y, z в отрицателна и в положителна посока, без обратно в предишната позиция. Използва се генераторът на случайни числа на средата за програмиране на C++, като се взима остатъкът от делене на 6 и нека това бъде числото r и тогава:

Ако $r=0$, то $x_{i+1}=x_i+1$, $y_{i+1}=y_i$, $z_{i+1}=z_i$

Ако $r=1$, то $x_{i+1}=x_i-1$, $y_{i+1}=y_i$, $z_{i+1}=z_i$

Ако $r=2$, то $x_{i+1}=x_i$, $y_{i+1}=y_i+1$, $z_{i+1}=z_i$

Ако $r=3$, то $x_{i+1}=x_i$, $y_{i+1}=y_i-1$, $z_{i+1}=z_i$

Ако $r=4$, то $x_{i+1}=x_i$, $y_{i+1}=y_i$, $z_{i+1}=z_i+1$

Ако $r=5$, то $x_{i+1}=x_i$, $y_{i+1}=y_i$, $z_{i+1}=z_i-1$

където $(x_{i+1}, y_{i+1}, z_{i+1})$ са координатите на новопоставената аминокиселина, а (x_i, y_i, z_i) са координатите на предишната поставена аминокиселина

Разстоянието между съседни аминокиселини е единица нормирано разстояние, а формата на протеина се представя опростено единствено чрез тези координати и под разполагане на аминокиселините ще се разбира разполагането на гръбначната структура на пептидната верига.

2. При разполагането на аминокиселините се спазват следните ограничения:

- не може да се разполага на координати, които вече са заети от друга аминокиселина:

$$(x_i=x_j) \cap (y_i=y_j) \cap (z_i=z_j)$$

където (x_i, y_i, z_i) са координатите на новопоставената аминокиселина, а (x_j, y_j, z_j) са координатите на всяка от поставените до момента аминокиселини

- не може да се разполага по-далеч от центъра на молекулата от разстояние, зададено от променливия параметър за допустимо разпростиране, по всяка от трите ортогонални посоки:

$$(x_i-x_c)^2+(y_i-y_c)^2+(z_i-z_c)^2 < s.l$$

където (x_i, y_i, z_i) са координатите на поставяната аминокиселина, (x_c, y_c, z_c) са координатите на центъра на молекулата построена до момента, l е дължината на пептидната верига, s е променливият параметър за допустимо разпростиране

- при получаване на недопустима позиция, се задава нова посока и така докато се изчерпят всичките 5 възможности. При невъзможност за разполагане на следващата аминокиселина, се прекратява построяването на текущата конформация, актуализира се броячът на провалените конформации и се започва следващата. Това се осъществява в цикъл с определен брой изпълнения, вложен в друг цикъл с определен брой изпълнения, като процентът на провалените конформации (възможни триизмерни форми) при изпълнение на вътрешния цикъл определя новата стойност на променливия параметър за допустимо разпростиране за следващата итерация на външния цикъл.

3. За да се формира стойността на оценъчната функция, следва да се определи кои аминокиселини

са в контакт – ако имат съседни координати. Проверката за близост се дефинира с формулата:

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 = 1$$

където (x_i, y_i, z_i) и (x_j, y_j, z_j) са координатите на двете аминокиселини

Оценъчната функция е сума на стойностите на хидрофобност на аминокиселините във всеки контакт плюс коефициент за отчитане на взаимодействие между водните молекули:

$$SF = \sum_{i,j \in C}^n H_i + H_j + w$$

където SF е стойността на оценъчната функция, w е коефициент за отчитане на взаимодействие между водните молекули H_i и H_j са стойностите на хидрофобност на аминокиселините с номера i и j , а C е множество на двойките в контакт.

В сравнение с класическия НР модел, този има възможност за гъвкаво изменение на ограничението за разпростиране в пространството, вместо фиксиран размер на решетката, което позволява да се намери по-добър баланс между процентът провалени конформации, поради недостатъчно място, и построяването на огромен брой конформации с ниска оценъчна функция, поради прекалено много място. Също така, оценъчната функция позволява да се отчита ролята на всички контакти, като водещата роля остава на хидрофобните.

Анализ на получените данни

Извършените експериментални симулации имат за цел, от една страна, да се намерят най-подходящите параметри за намиране на конформация с най-висока оценъчна функция и брой повторно намиране на същата, а от друга страна да се провери доколко получените „най-добри“ конформации съответстват на действителната триизмерна форма (верификация на модела) – за целта са намерени съседните номерирани двойки аминокиселини, които са в близост

(контакт) за същия изследван протеин - 1UUB с 56 аминокиселини – избран е като подходящ за изследване, тъй като е сравнително къс едноверижен пептид, активен глобуларен протеин, за който има данни от ядреномагнитен резонанс., чиито координати са взети от PDB файла в базата данни.

Експерименталните данни започват с построяването на конформациите при начална стойност на променливия параметър за допустимо разпростиране $s=0,5$ и стъпка на увеличението му $st=0,1$ – увеличението настъпва при повисок процент на провалените конформации в рамките на вътрешния цикъл от 1000000 конформации, като тази проверка във външния цикъл се извършва 1000 пъти – т.е. стойността на параметъра бързо ще достигне и ще се задържи на тази, при която процентът на провалените конформации е колкото зададеният.

При $s=0,5$, $st=0,1$ и 90% провалени конформации се получава следната структура:

- стойност на оценъчната функция – 125.8
- брой пъти намерена същата – 1
- общ брой контакти и списък на двойките аминокиселини в контакт:13

При $s=0,5$, $st=0,1$ и 92% провалени конформации се получава следната структура:

- стойност на оценъчната функция – 168
- брой пъти намерена същата – 3
- общ брой контакти и списък на двойките аминокиселини в контакт:15

При $s=0,5$, $st=0,1$ и 94% провалени конформации се получава следната структура:

- стойност на оценъчната функция – 197.1
- брой пъти намерена същата – 4

- общ брой контакти и списък на двойките аминокиселини в контакт:18

При $s=0,5$, $st=0,1$ и **96%** провалени конформации се получава следната структура:

- стойност на оценъчната функция – 213.6

- брой пъти намерена същата – 6

- общ брой контакти и списък на двойките аминокиселини в контакт:2

При $s=0,5$, $st=0,1$ и **98%** провалени конформации се получава следната структура:

- стойност на оценъчната функция – 285.9

- брой пъти намерена същата – 10

- общ брой контакти и списък на двойките аминокиселини в контакт:26

При $s=0,5$, $st=0,1$ и **99%** провалени конформации се получава следната структура:

- стойност на оценъчната функция – 325.7

- брой пъти намерена същата – 30

- общ брой контакти и списък на двойките аминокиселини в контакт:35

4.2. Модел за предсказване на триизмерната форма на белтъците с нецелочислени координати, с променливо разпростиране и трикомпонентна оценъчна функция

Описание на създадения математически модел

В тази глава ще разгледам разработения модел за предсказване на триизмерната форма на белтъците, разновидност на НР модела, при който:

1. Алфа въглеродните атоми се разполагат с триизмерни нецелочислени координати, започвайки от $(0,0,0)$ и $(0,0,1)$, следвайки номерата на

аминокиселините в аминокиселинната последователност, като на случаен принцип се избира позицията на всеки следващ алфа въглероден атом от възможните усуквания на пептидната верига, взети през 10 градуса точност. Разстоянието между алфа въглеродните атоми на съседни аминокиселини е единица нормирано разстояние, а формата на протеина се представя опростено единствено чрез тези координати и под разполагане на аминокиселините ще се разбира разполагането на тези атоми, определящи гръбначната структура на пептидната верига. Използва се генераторът на случайни числа на средата за програмиране на C++, като се взима остатъкът от делене на 36 и нека това бъде числото r и тогава:

Ако $x_{i-1} - x_{i-2} = \max(x_{i-1} - x_{i-2}, y_{i-1} - y_{i-2}, z_{i-1} - z_{i-2})$, то

$$x_i = x_{i-1} + 0.5(x_{i-1} - x_{i-2}) + 0.5, \quad y_i = y_{i-1} + \sin(10r), \quad z_i = z_{i-1} + \cos(10r)$$

Ако $y_{i-1} - y_{i-2} = \max(x_{i-1} - x_{i-2}, y_{i-1} - y_{i-2}, z_{i-1} - z_{i-2})$, то

$$y_i = y_{i-1} + 0.5(y_{i-1} - y_{i-2}) + 0.5, \quad x_i = x_{i-1} + \sin(10r), \quad z_i = z_{i-1} + \cos(10r)$$

Ако $z_{i-1} - z_{i-2} = \max(x_{i-1} - x_{i-2}, y_{i-1} - y_{i-2}, z_{i-1} - z_{i-2})$, то

$$z_i = z_{i-1} + 0.5(z_{i-1} - z_{i-2}) + 0.5, \quad y_i = y_{i-1} + \sin(10r), \quad x_i = x_{i-1} + \cos(10r)$$

където (x_i, y_i, z_i) са координатите на новопоставената аминокиселина, а $(x_{i-1}, y_{i-1}, z_{i-1})$ и $(x_{i-2}, y_{i-2}, z_{i-2})$ са координатите на предходните аминокиселини

2. При разполагането на аминокиселините се спазват следните ограничения, основани на структурата на протеините:

- не може да се разполага по-близо от 0,7 нормирано квадратично разстояние по всяка от трите ортогонални посоки до друга аминокиселина:

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 > 0,7$$

където (x_i, y_i, z_i) и (x_j, y_j, z_j) са координатите на двете аминокиселини

- не може да се разполага по-далеч от центъра на молекулата от разстояние, зададено от променливия параметър за допустимо разпростиране, по всяка от трите ортогонални посоки:

$$(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2 < s \cdot l$$

където (x_i, y_i, z_i) са координатите на поставяната аминокиселина, (x_c, y_c, z_c) са координатите на центъра на молекулата построена до момента, l е дължината на пептидната верига, s е променливият параметър за допустимо разпростиране

- при получаване на недопустима позиция, се задава нов случаен ъгъл на усукване и така докато се изчерпят всичките 36 възможности. При невъзможност за разполагане на следващата аминокиселина, се прекратява построяването на текущата конформация, актуализира се броячът на провалените конформации и се започва следващата. Това се осъществява в цикъл с определен брой изпълнения, вложен в друг цикъл с определен брой изпълнения, като процентът на провалените конформации (възможни триизмерни форми) при изпълнение на вътрешния цикъл определя новата стойност на променливия параметър за допустимо разпростиране за следващата итерация на външния цикъл.

3. За да се формира стойността на оценъчната функция, следва да се определи кои аминокиселини са в контакт – тъй като страничните радикали могат

да бъдат между двата участъка на пептидната верига, то за контакт се счита близост на алфа въглеродните им атоми до два пъти нормираното разстояние между съседни алфа въглеродните атоми – така между тях не може и да се помести друга аминокиселина. Проверката за близост се дефинира с формулата:

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 < 4$$

където (x_i, y_i, z_i) и (x_j, y_j, z_j) са координатите на двете аминокиселини

Оценъчната функция е произведение на три подфункции, които отчитат степента на изграждане на:

1) вторичната структура – съседство на аминокиселините и изграждане на водородни връзки при гръбначната структура – стойността е броят им:

$$SF = \sum_{i,j \in C} a_{ij}$$

където SF е стойността на оценъчната подфункция, a_{ij} са аминокиселините с номера i и j , а C е множество на двойките в контакт.

2) контакти между хидрофобни аминокиселини – стойността е сума от стойностите на хидрофобност на съседните хидрофобни аминокиселини:

$$SF = \sum_{i,j \in C} H_i + H_j,$$

където SF е стойността на оценъчната подфункция, H_i и H_j са стойностите на хидрофобност на аминокиселините с номера i и j , а C е множество на двойките в контакт.

3) дисулфидни мостове – стойността е броят плюс едно, тъй като може да няма такива:

$$SF = 1 + \sum_{i,j \in C} a_{ij}$$

където SF е стойността на оценъчната подфункция, a_{ij} са аминокиселините с номера i и j , а C е множество на двойките аминокиселини цистеин в контакт.

В сравнение с класическия HP модел, този има по-висока изчислителна сложност, но дава по-реалистично построяване на пептидната верига в пространството – разполагането на аминокиселините следва природния принцип на усукване на пептидната връзка, подобно на движението на кончето в шахмата, но триизмерно. Така има вероятност да се построят триизмерни форми (конформации), които да приличат на реалните форми в природата.

Анализ на получените данни

Параметрите на най-добрата получена структура са:

- стойност на оценъчната функция – 34452

-s=0.5

- брой хидрофобни контакти 21, брой дисулфидни мостове 2 и: общ брой контакти и списък на двойките аминокиселини в контакт: 115

Имайки предвид, че формата на протеина е динамична – движи се, леко се изменя при различни условия в клетъчната среда и особено при замразяване за изследване чрез ЯМР, то съседство на по-голямата част от аминокиселините дава много близка форма, която би могла да проявява същите биохимични свойства. Ако изследването се проведе с оптимизирана оценъчна функция и с по-голям изчислителен ресурс, то най-вероятно ще се постигне и по-голямо геометрично подобие.

4.3. Модификация на модела с разполагане на аминокиселините чрез два псевдослучайни ъгъла

Използва се генераторът на случайни числа на средата за програмиране на C++, като се взима остатъкът от делене на 72 и нека това бъде числото r , а $r1=r/12$ и $r2=r \bmod 12$, тогава:

$$\begin{aligned}x_i &= x_{i-1} + \sin(90 - 30r1), \\y_i &= y_{i-1} + \sin(30r2)\cos(90 - 30r1), \\z_i &= z_{i-1} + \cos(30r2)\cos(90 - 30r1)\end{aligned}$$

където (x_i, y_i, z_i) са координатите на новопоставената аминокиселина, а $(x_{i-1}, y_{i-1}, z_{i-1})$ и $(x_{i-2}, y_{i-2}, z_{i-2})$ са координатите на предходните аминокиселини.

Всичко останало в модела и последователността на работа се запазва, както е описано в началото на главата.

Анализ на получените данни

Разликата при анализа с предишната модификация е, че по-голямата свобода на завиване дава по-голям брой възможни структури, така че вече параметър брой намиране на структурата с най-висока оценъчна функция няма да бъде следен. Поради наблюдаваното при досегашните опити преобладаващо значение на параметъра процент провалени пред стъпката и началната стойност на коефициента на разпростиране s , то той ще се изменя. Също така ще има два варианта на оценъчната подфункция за хидрофобните контакти – с параметър за влиянието на водата $wtw = 8$ и $wtw = 0$ – в първия случай контактите между полярни аминокиселини носят положителна стойност, макар и значително по-малка от хидрофобните, докато във втория вариант носи отрицателна стойност. При 40% провалени и $wtw = 0$ се получи най-добра структура с 123 контакти и 40 съвпадащи със сниманата структура със 125 контакти.

4.4. Модел за предсказване на триизмерната форма на белтъците с начално намиране на псевдослучайни структури и последващо целенасочено нагъване

Описание на създадения математически модел

Основна цел на разработения алгоритъм е да се създаде целенасочено структура с ниска потенциална енергия, като се започне от случайно нагъната форма, вместо да търсим най-добрата структура сред изцяло случайно генерирани форми. Всяка аминокиселина е представена като позиция на алфа въглеродния атом и на центъра на радикала с триизмерни координати.

Първият етап е да се генерира триизмерна форма чрез случайно завиване на пептидната верига през 90 градуса и разстояние 1 между алфа въглеродните атоми и радикалите, като сред получените псевдослучайни структури се избират само тези, при които има съседство между радикалите на аминокиселините цистеин, така че да се създадат възможните дисулфидни мостове.

Вторият етап е да се промени целенасочено пространствената форма, с цел минимизиране на енергията, като се премине към координати дробни числа.

Дефиниция на структурата

- Ако x_i , y_i и z_i принадлежат на \mathbb{R} и са координатите на алфа въглеродните атоми, а x_{r_i} , y_{r_i} , z_{r_i} принадлежат на \mathbb{R} и са координатите на центровете на радикалите и r_i са нормираните радиуси на радикалите при условно разстояние 1 между алфа въглеродните атоми на съседните аминокиселини, то:

$$\sqrt{((x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2)} \leq 1.1$$

$$\sqrt{((x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2)} \geq 0.9$$

$$\sqrt{((x_i - x_{r_i})^2 + (y_i - y_{r_i})^2 + (z_i - z_{r_i})^2)} \leq 1.1 (r_i + 0.3)$$

$$0,6 \leq \sqrt{((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2)}$$

$$0,9(r_i + r_j) \leq \sqrt{((x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2)}$$

$$0,9(r_j + 0.3) \leq \sqrt{((x_i - x_{r_j})^2 + (y_i - y_{r_j})^2 + (z_i - z_{r_j})^2)}$$

Дефиниция на търсения резултат

- Дефиниция на контакт:
 - между несъседни алфа въглеродни атоми: $\sqrt{((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2)} \leq 0.8$
 - между радикали: $\sqrt{((x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2)} \leq 1.2 (r_i + r_j)$
- Цел на целенасоченото нагъване е максимизиране на контактите между радикалите на хидрофобните аминокиселини и на контактите между алфа въглеродните атоми на всички аминокиселини, като задължително има максималния възможен брой контакти между радикалите на цистеините, а радикалите с електрически заряд не са в контакт.
- Оценъчната функция има вида:

$$F(\text{fold}) = \sum_{i \text{ contact } j} (H_i + H_j + wtw) + \sum_{i \text{ contact } j} wtw ,$$

където H_i и H_j са стойностите на хидрофобност на аминокиселините, wtw е параметър за влияние на водородните връзки.

Процес на целенасочено нагъване

- Процесът се извършва циклично, като на всяка итерация се извършват следните стъпки:
 1. Намира се условния център, като средно-аритметични координати по трите направления:

$$x_c = \sum_{i=1 \text{ to } n} x_{r_i} / n$$

$$y_c = \sum_{i=1 \text{ to } n} y_{r_i} / n$$

$$z_c = \sum_{i=1 \text{ to } n} z_{r_i} / n$$
 2. Координатите на радикалите се преместват пропорционално на стойността им на хидрофобност H_i , като ако $H_i > 0$ посоката е към центъра на молекулата, а при $H_i < 0$ е обратно:

$$x_{ni} = x_i + 0.02h_i, \quad x_c > x_i$$

$$x_{ni} = x_i - 0.02h_i, \quad x_c < x_i$$

$$y_{ni} = y_i + 0.02h_i, \quad y_c > y_i$$

$$y_{ni}=y_i-0.02h_i, y_c < y_i$$

$$z_{ni}=z_i+0.02h_i, z_c > z_i$$

$$z_{ni}=x_i-0.02h_i, z_c < z_i$$

3. Координатите на алфа въглеродните атоми се преместват към центъра на молекулата:

$$x_{ni}=x_i+0.01, x_c > x_i$$

$$x_{ni}=x_i-0.01, x_c < x_i$$

$$y_{ni}=y_i+0.01, y_c > y_i$$

$$y_{ni}=y_i-0.01, y_c < y_i$$

$$z_{ni}=z_i+0.01, z_c > z_i$$

$$z_{ni}=x_i-0.01, z_c < z_i$$

4. Извършва се корекция на позициите на всички алфа въглеродни атоми и радикали, с цел запазване на пептидната верига, като алфа карбон i се доближава до $i+1$ и радикал i се приближава към алфа въглероден атом i , ако $\sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2} \leq 1.1$:

$$x_{in}=x_i+(x_{i+1}-x_i)/10$$

$$y_{in}=y_i+(y_{i+1}-y_i)/10$$

$$z_{in}=z_i+(z_{i+1}-z_i)/10$$

$$x_{r_{in}}=x_{r_i}+(x_i-x_{r_i})/10$$

$$y_{r_{in}}=y_{r_i}+(y_i-y_{r_i})/10$$

$$z_{r_{in}}=z_{r_i}+(z_i-z_{r_i})/10$$

5. Координатите на радикалите на цистините от дисулфидните мостове се преместват един към друг, ако $\sqrt{(x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2} \leq 1.1$ (r_i+r_j):

$$x_{r_{in}}=x_{r_i}+(x_{r_j}-x_{r_i})/10$$

$$y_{r_{in}}=y_{r_i}+(y_{r_j}-y_{r_i})/10$$

$$z_{r_{in}}=z_{r_i}+(z_{r_j}-z_{r_i})/10$$

$$x_{r_{jn}}=x_{r_j}+(x_{r_i}-x_{r_j})/10$$

$$y_{r_{jn}} = y_{r_j} + (y_{r_i} - y_{r_j}) / 10$$

$$z_{r_{jn}} = z_{r_j} + (z_{r_i} - z_{r_j}) / 10$$

6. Координатите на радикалите на електрически заредените аминокиселини се раздалечават един от друг,

$$\text{ако } \sqrt{((x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2)} \leq 2:$$

$$x_{r_{in}} = x_{r_i} - (x_{r_j} - x_{r_i}) / 10$$

$$y_{r_{in}} = y_{r_i} - (y_{r_j} - y_{r_i}) / 10$$

$$z_{r_{in}} = z_{r_i} - (z_{r_j} - z_{r_i}) / 10$$

$$x_{r_{jn}} = x_{r_j} - (x_{r_i} - x_{r_j}) / 10$$

$$y_{r_{jn}} = y_{r_j} - (y_{r_i} - y_{r_j}) / 10$$

$$z_{r_{jn}} = z_{r_j} - (z_{r_i} - z_{r_j}) / 10$$

7. Координатите на радикалите на близко разположените хидрофобни аминокиселините се преместват един към друг, ако $\sqrt{((x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2)} \leq 2:$

$$x_{r_{in}} = x_{r_i} + (x_{r_j} - x_{r_i}) / 20$$

$$y_{r_{in}} = y_{r_i} + (y_{r_j} - y_{r_i}) / 20$$

$$z_{r_{in}} = z_{r_i} + (z_{r_j} - z_{r_i}) / 20$$

$$x_{r_{jn}} = x_{r_j} + (x_{r_i} - x_{r_j}) / 20$$

$$y_{r_{jn}} = y_{r_j} + (y_{r_i} - y_{r_j}) / 20$$

$$z_{r_{jn}} = z_{r_j} + (z_{r_i} - z_{r_j}) / 20$$

8. Координатите на алфа въглеродните атоми на близко разположените аминокиселините се преместват един към друг,

$$\text{ако } \sqrt{((x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2)} \leq 2:$$

$$x_{r_{in}} = x_{r_i} + (x_{r_j} - x_{r_i}) / 20$$

$$y_{r_{in}} = y_{r_i} + (y_{r_j} - y_{r_i}) / 20$$

$$z_{r_{in}} = z_{r_i} + (z_{r_j} - z_{r_i}) / 20$$

$$x_{r_{jn}} = x_{r_j} + (x_{r_i} - x_{r_j}) / 20$$

$$y_{r_{jn}} = y_{r_j} + (y_{r_i} - y_{r_j}) / 20$$

$$z_{r_{jn}} = z_{r_j} + (z_{r_i} - z_{r_j}) / 20$$

9. Извършва се корекция на позициите на всички алфа въглеродни атоми, с цел избягване на застъпване на позициите, ако $0,6 > \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ при $j \neq i+1$, и $0,9 > \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ при $j = i+1$:

$$x_{in} = x_i - 0.1(x_j - x_i) / ((x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2)$$

$$y_{in} = y_i - 0.1(y_j - y_i) / ((x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2)$$

$$z_{in} = z_i - 0.1(z_j - z_i) / ((x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2)$$

10. Извършва се корекция на позициите на всички радикали, с цел избягване на застъпване на позициите, ако $0,9(r_i + r_j) > \sqrt{(x_{r_i} - x_{r_j})^2 + (y_{r_i} - y_{r_j})^2 + (z_{r_i} - z_{r_j})^2}$:

$$x_{r_{in}} = x_{r_i} - 0.1(x_{r_j} - x_{r_i}) / ((x_{r_j} - x_{r_i})^2 + (y_{r_j} - y_{r_i})^2 + (z_{r_j} - z_{r_i})^2)$$

$$y_{r_{in}} = y_{r_i} - 0.1(y_{r_j} - y_{r_i}) / ((x_{r_j} - x_{r_i})^2 + (y_{r_j} - y_{r_i})^2 + (z_{r_j} - z_{r_i})^2)$$

$$z_{r_{in}} = z_{r_i} - 0.1(z_{r_j} - z_{r_i}) / ((x_{r_j} - x_{r_i})^2 + (y_{r_j} - y_{r_i})^2 + (z_{r_j} - z_{r_i})^2)$$

11. Извършва се корекция на позициите, с цел избягване на застъпване на между алфа въглероден атом и радикал, ако $0,9(r_j + 0.3) > \sqrt{(x_i - x_{r_j})^2 + (y_i - y_{r_j})^2 + (z_i - z_{r_j})^2}$:

$$x_{r_{in}} = x_{r_i} - 0.1(x_{r_j} - x_i) / ((x_{r_j} - x_i)^2 + (y_{r_j} - y_i)^2 + (z_{r_j} - z_i)^2)$$

$$y_{r_{in}} = y_{r_i} - 0.1(y_{r_j} - y_i) / ((x_{r_j} - x_i)^2 + (y_{r_j} - y_i)^2 + (z_{r_j} - z_i)^2)$$

$$z_{r_{in}} = z_{r_i} - 0.1(z_{r_j} - z_i) / ((x_{r_j} - x_i)^2 + (y_{r_j} - y_i)^2 + (z_{r_j} - z_i)^2)$$

Анализ на получените данни

Изследванията по този модел и алгоритъм позволиха да се сравнят много по-голям брой структури с ниска енергия, в сравнение с предишните модели и алгоритми, тъй като той работи много по-бързо и етапът на целенасочено сгъване на протеина дава винаги структура с ниска енергия. Оценка и подбор на най-добра структура бе направено по два начина – чрез оценъчна функция, както ще работи в реални условия, и чрез процент съвпадащи контакти, което изследване показва потенциала на модела при оптимизиране на оценъчната функция.

С най-добра оценъчна функция се получи най-добра структура със **163** контакта и **69** съвпадащи със сниманата структура със 125 контакта.

С най-голям процент съвпадащи контакти се получи следната най-добра структура със **120** контакта и **62** съвпадащи със сниманата структура със 125 контакта.

Тези резултати превъзхождат получените по предходните модели по степен на молекулярно геометрично подобие, а основното предимство е десетки пъти по-бързото получаване на предсказаната структура, което ще позволи създаването на модел с много по-детайлно и много близко до реалното представяне на протеиновата структура, който, с помощта на създадения алгоритъм за целенасочено формиране, може да работи със задоволителна скорост и може да даде възможност да се предскажат значим брой структури, което със скоростта на досега използваните в света методи на предсказване е изчислително непосилно.

5. Анализ на областта на приложимост на модела и алгоритъма, показали най-добри резултати при изследването върху един и същ протеин чрез изследването на други протеини

Проведеният експеримент с използването на модел с триизмерни нецелочислени координати при разполагане на алфа въглеродния атом и центъра на страничната верига с двуетапен алгоритъм с първоначално случайно разполагане в пространството и последващо целенасочено променяне на формата към такава с по-ниска свободна енергия даде следните резултати:

- протеин 1yua има намерена форма с най-ниска енергия според оценъчната функция с 891 контакта – това е протеин с глобуларна форма, 121 аминокиселини;

- протеин 1adn има намерена форма с най-ниска енергия според оценъчната функция с 86 контакта – това е протеин с множество алфа спирали, 91 аминокиселини;

- протеин 1xjh има намерена форма с най-ниска енергия според оценъчната функция с 265 контакта – това е протеин с алфа спирала, 61 аминокиселини;

- протеин 1cq0 има намерена форма с най-ниска енергия според оценъчната функция с 94 контакта – това е протеин с алфа спирала, 27 аминокиселини;

- протеин 1d0r има намерена форма с най-ниска енергия според оценъчната функция с 102 контакта – това е протеин с алфа спирала, 29 аминокиселини;

- протеин 1mxl има намерена форма с най-ниска енергия според оценъчната функция с 382 контакта – това е протеин с алфа спирала, 89 аминокиселини;

- протеин 1euf има намерена форма с най-ниска енергия според оценъчната функция с 370 контакта – това е протеин с алфа спирала, 92 аминокиселини;

- протеин 1fjr има намерена форма с най-ниска енергия според оценъчната функция с 162 контакта – това е протеин с алфа спирала, 52 аминокиселини;

- протеин 1ry3 има намерена форма с най-ниска енергия според оценъчната функция с 153 контакта – това е протеин с алфа спирала, 46 аминокиселини;

- протеин 1aml има намерена форма с най-ниска енергия според оценъчната функция с 126 контакта – това е протеин с алфа спирала, 40 аминокиселини;

- протеин 1ho9 има намерена форма с най-ниска енергия според оценъчната функция с 82 контакта – това е протеин с алфа спирала, 32 аминокиселини;

- протеин 1hrh има намерена форма с най-ниска енергия според оценъчната функция с 90 контакта – това е протеин с алфа спирала, 37 аминокиселини;

- протеин 1jdm има намерена форма с най-ниска енергия според оценъчната функция с 95 контакта – това е протеин с алфа спирала, 31 аминокиселини;

- протеин 1esx има намерена форма с най-ниска енергия според оценъчната функция с 502 контакта – това е протеин с алфа спирала, 96 аминокиселини;

- протеин 1jfw има намерена форма с най-ниска енергия според оценъчната функция с 351 контакта – това е протеин с алфа спирала, 86 аминокиселини;

Данните за протеините след първия с малко съвпадение на контакти ясно показват, че запазването на вторичната структура от алфа спирали, при формирането на третична структура, ако изобщо се формира такава, пречи да се образуват голям брой хидрофобни контакти между страничните вериги и реалната форма няма добро сходство с формата с най-ниска енергия, която би се получила при директното образуване на третична структура и форма на статистическо кълбо. При протеина 1uua има съвпадение на над една пета от контактите, въпреки наличието на алфа спирали, тъй като има значително формиране и на третична структура – това е глобуларен протеин, но не е изцяло от тип статистическо кълбо – случаят, в който третичната структура изцяло определя формата и свободната енергия е най-ниска и която се получава при преобладаващи големи странични вериги на аминокиселините и заредени такива [46]. Анализът полказва, че използването на методи за предсказване на триизмерната структура на протеините, базирани на търсене на структура с минимална енергия, е приложимо при глобуларни протеини, като постижимото сходство е толкова по-голямо, колкото повече преобладава третичната структура и се доближава до статистическо кълбо, а в случаите на единична голяма алфа спирала напълно отсъства третична структура и няма добро сходство на реалната с предсказаната структура.

6. Заключение

При модела с целочислени координати има тенденция с увеличаване на процента провалени конформации да се увеличава стойността на оценъчната функция, както и броят контакти и броят повторения на конформацията, но няма висок процент съвпадения с

реалната структура и той не расте винаги при по-висока оценъчна функция. Това се дължи на по-ограниченото пространство за нагъване – получават се по-компактни структури, но реалната структура не успява да заеме много компактна форма, и ортогоналното завиване на веригата не позволява да се получи много близка форма до реалната, където веригата завива по съвсем различен начин, а и аминокиселините могат да взаимодействат от по-голямо разстояние от това, на което могат да се доближават и на което се намират съседните във веригата – в този модел, както и в стандартния НР модел това е единица нормирано разстояние.

При модела с координати реални числа се получават най-показателни данни във втората му комбинация от параметри и при 90% провалени конформации се получиха следващите две конформации, като разглеждам и втората по стойност на оценъчната функция конформация, защото тя се оказва с изключително близка форма до тази, получена чрез ЯМР и по това и по разликите в типовете контакти между аминокиселините, може да се направи важен извод: съвпаденията в първата структура са 18 от общо 119, а във втората 49 от 115 плюс 12 много близо до контакт. Имайки предвид, че формата на протеина е динамична – движи се, леко се изменя при различни условия в клетъчната среда и особено при замразяване за изследване чрез ЯМР, то съседство на по-голямата част от аминокиселините дава много близка форма, която би могла да проявява същите биохимични свойства. При това е получена при сравнение на 100 милиона форми (от 1 милиард изпробвани случайни конформации 10% са успешно построени) при възможни 36^{56} случайни форми по този модел, макар повечето от тях да не отговарят на ограниченията за разполагане в пространството. Ако изследването се проведе с оптимизирана оценъчна функция и с по-голям изчислителен ресурс, то най-вероятно ще се постигне и по-голямо геометрично подобие. Това оправдава продължаването на експериментите в тази посока, с цел достигането на

достатъчно бързо и точно работещ модел, което да позволи реалното му приложение за целите на биохимията.

При модела с целенасочено дооформяне на структурата, резултатите превъзхождат получените по предходните модели по степен на молекулярно геометрично подобие, а основното предимство е десетки пъти по-бързото получаване на предсказаната структура, което ще позволи създаването на модел с много по-детайлно и много близко до реалното представяне на протеиновата структура, който, с помощта на създадения алгоритъм за целенасочено формиране, може да работи със задоволителна скорост и може да даде възможност да се предскажат значим брой структури, което със скоростта на досега използваните в света методи на предсказване е изчислително непосилно.

Друго основно предимство е, че могат да се предсказват формите на протеини с толкова дълги аминокиселинни последователности, колкото съществуват в природата – повечето глобуларни биологичноактивни протеини имат стотици и до над 1000 аминокиселини, за разлика от досега използваните алгоритми, които работят с по-малко от 100, а често и по-малко от 40 аминокиселини.

Относно областта на приложимост на използваните методи за предсказване, която засега обхваща основно глобуларни протеини и най-вече тези, при които преобладава третичната структура и няма преобладаваща алфа спирала, то тя може да бъде разширена с предварително предсказване на алфа спиралите – формирането им зависи от преобладаващите по вид аминокиселини в последователност и край им от наличието на аминокиселината Пролин [46]. Ако бъдат създадени предполагаемите алфа спирали като предварителни геометрични обекти и към тях се разполагат останалите части от протеина, представени както досега чрез гръбначна и странични вериги, то би могло да се създадат модели и алгоритми, които са напълно адекватни за останалите видове протеини.

7. Приноси

1. Разработен е математически модел и базирани на него авторски компютърни програми за предсказване на триизмерната структура на протеини, модификация на HP модела, който използва цели координати за разполагане на аминокиселините, опростени до разположението им по върховете на кубична решетка, като вместо размер на решетката има ограничение за отдалечаване от центъра на молекулата, което ограничение се изменя динамично, съобразно степента на трудност за построяването на структурите, а оценъчната функция отчита и контактите между всички аминокиселини, но в по-малка степен, отколкото на хидрофобните, като се използват стойностите на хидрофобност на всяка аминокиселина.
2. Разработен е математически модел и базирани на него авторски компютърни програми за предсказване на триизмерната структура на протеини, модификация на HP модела, който използва дробни координати за разполагане на аминокиселините, опростени до разположението на гръбначната структура на пептидната верига, като има ограничение за отдалечаване от центъра на молекулата, което ограничение се изменя динамично, съобразно степента на трудност за построяването на структурите, а оценъчната функция е трикомпонентна и отчита формирането на вторична структура, хидрофобни контакти и дисулфидни мостове.
3. Разработен е математически модел и базирани на него авторски компютърни програми за предсказване на триизмерната структура на протеини, който използва дробни координати за разполагане на аминокиселините, опростени до разположението на алфа въглеродните атоми и

центровете на страничните вериги – въглеродоридни остатъци (радикали), като има два етапа. Първият етап е създаването на псевдослучайна структура, използвайки алгоритъмът от предходния модел, а вторият етап е допълнително оформяне на структурата чрез постъпково преместване на координатите на обектите (двете характерни точки на всяка аминокиселина), което е целенасочено към минимизиране на енергията, чрез формиране на близост по критериите за енергомимимизиращи контакти. Оценъчната функция е трикомпонентна и отчита формирането на вторична структура, хидрофобни контакти и дисулфидни мостове.

4. Направен е сравнителен анализ между получените *in silico* структури чрез всеки от моделите и всяка тяхна модификация и получената чрез ЯМР реална структура. Получените резултати показват, че разработените модели, алгоритми и програмираните на тяхна база компютърни софтуери генерират 3D структури, близки до реалната.
5. Чрез последния модел и с алгоритъма с целенасочено доформиране на структурата са постигнати всички основни цели на изследването – многократно увеличена скорост на компютърното предсказване, работа с в пъти по-дълги последователности, възможност за по-висока степен на геометрично подобие.
6. Извършени са експерименти с различни по тип структура протеини и от анализа на получените резултати са направени важни изводи за приложимостта на използваните методи на предсказване, а именно че те са най-адекватни за глобуларни протеини и са най-ефективни при преобладаваща структура от тип статистическо кълбо – резултат на третичната структура на протеините, докато при преобладаваща алфа

спирала на вторичната структура е необходимо да се създаде предварително алфа спиралата като геометричен обект и нагъването да продължи при запазването ѝ.

Публикации на автора по темата на дисертационния труд:

1. K. Gashteovski, I. Trenchev, N. Borisova, G. Iliev, I. Todorin „Neural networks – basic principles for construction, their application in drug design and genetic algorithms”, FMNS 2011, Благоевград 2011
2. R. Stoev, K. Gashteovski, I. Trenchev, N. Borisova, I. Todorin, M. Traykov "Bioinformatics research of 3D structure of the proteins", FMNS 2011, Благоевград 2011
3. I. Todorin „A Model for HP Folding Prediction using Variable Size of Lattice”, FMNS 2013, Благоевград 2013
4. I. Todorin, I. Trenchev, A. Stoilov, R. Mavrevski, M. Traykov “A Model for HP Folding Prediction Using Increasing Constrains for Spreading in the Process of Making Conformation”, Proceedings of BIOMATH 2015, Благоевград, 2015: 88
5. I. Todorin, M. Traykov, R. Mavrevski, A. Stoilov, I. Trenchev „3D Visualization of the Biological Structures”, Proceedings of BIOMATH 2015, Благоевград, 2015: 89
6. I. Todorin „Protein Folding Prediction using two stages over random made structure – “Bridge Building” and “Willpower Folding”, „Студентска и докторантска научна сесия 2017” на Техническият факултет на ЮЗУ „Неофит Рилски”, Благоевград, 2017
7. I. Todorin „Protein Folding Prediction using “Willpower Folding” stage over random made structure”, „Студентска и докторантска научна сесия 2017” на Техническият факултет на ЮЗУ „Неофит Рилски”, Благоевград, 2017

8. I. Todorin „Protein Folding Using three componential score function and one angle of curving”, „Seventh International Conference of FMNS (FMNS-2017)”, Благоевград, 2017
9. I. Todorin „Protein Folding Using two angles for curving in space curving”, „Seventh International Conference of FMNS (FMNS-2017)”, Благоевград, 2017
10. I. Todorin,, N. Yanev, *M. Traykov*, B. Yurukov „A new off-lattice HP model with side-chains for protein folding problem”, „2017 International Conference Applied Mathematics, Computational Science and Systems Engineering ([AMCSE 2017](#))” Athens, Greece, 2017
11. I. Todorin, *I. Trenchev* „An Off-Lattice HP Model for Protein folding with three-componential evaluation function”, „CIIT 2018”, Mavrovo, Macedonia, 2018
12. I. Todorin,, N. Yanev, *M. Traykov*, B. Yurukov “An off-lattice HP model with side-chains and S-S bridges impact for protein folding problem” WSEAS TRANSACTIONS on BIOLOGY and BIOMEDICINE, 2018, 15,16-23, E-ISSN:2224