An off-lattice HP model with side-chains and S-S bridges impact for protein folding problem

IVAN TODORIN
Department of Communication and Computer Engineering
University Center for Advanced Bioinformatics Research
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA
itodorin@gmail.com

NICOLA YANEV
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. Georgi Bonchev Str., Block 8, 1113 Sofia
BULGARIA
choby@math.bas.bg

METODI TRAYKOV
Department of Electrical Engineering, Electronics and Automatics
University Center for Advanced Bioinformatics Research
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA
metodi.gt@gmail.com

BORISLAV YURUKOV
Department of Informatics
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA
bobyur@swu.bg

Abstract: - The proposed model takes into account highest degree of complexity – different size of radiuses of side-chains, mentioned below as radicals, and two stages of algorithm, including random structure initially and subsequent purposeful folding process according to the physical forces and energies. The objective function is three-componential – one member corresponding to the impact of the hydrophobic contacts of side-chains, next one to the impact of hydrogen bonds between peptide chain parts and third to the impact of S-S bridges.

Key-Words: - Protein folding prediction, Bioinformatics, Drug Design, Willpower folding, HP model, Amino acid descriptors, S-S bridges

1 Introduction

3D structure of proteins is the major factor that determines their biological activity. The synthesis of new proteins and the crystallographic analysis of their 3D structure is very slow and very expensive process. If we can predict the 3D structure of many proteins, than only proteins with expected properties have to be synthesized. That will increase the number of known structures in the databases for proteins, and they can be used for drug design. The prediction of the 3D structure of proteins, if we know only the primary structure – the amino-acid sequence, is a protein folding problem. The reason for this process of folding in water environment is the interaction between water molecules and between amino-acids and water molecules. As water molecule has higher polarity than amino-acids, there is a minimum of energy when the protein is folded, not to spoil water to water interconnections. The way of folding is determined by the polarity or the hydrophobicity of different amino-acids, so the 3D structure with minimum energy is the real case. [1,2,3] There is less energy when more hydrophobic (H) amino-acids (the hydrophobic type of amino acid depends on the middle nucleotide of the codon [4]) are in contact in the core of the folded 3D structure and more polar (P) amino-acids are in contact with water. As we know the amino-acid sequence and the hydrophobicity of every amino-acid, we can predict the 3D structure – this method is called HP folding [5]. The closest to our model type of HP model is the off-lattice one, following the approach know as HP folding.

2 Problem Formulation

2.1 Description of the set of feasible solutions

2.1.1 Structure and contact defining constraints

Let x_i, y_i, $z_i \in R$ be the unknown coordinates of the alpha carbon atoms, xr_i, yr_i, $zr_i \in R$ be the coordinates of the centers of the radicals and r_i be the normalized radiuses of the radicals, for the i^{th}(j^{th}) amino acid in the peptide chain, where n is the number of amino acids:

$$0.9 \le \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2} \le 1.1$$

$$\sqrt{(x_i - xr_{i+1})^2 + (y_i - yr_{i+1})^2 + (z_i - zr_{i+1})^2} \le 1.1(r_i + 0.3)$$

$$(1) \quad 0.6 \le \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

$$0.9(r_i + r_j) \le \sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2}$$

$$0.9(r_j + 0.3) \le \sqrt{(x_i - xr_j)^2 + (y_i - yr_j)^2 + (z_i - zr_j)^2} \le 1.1(r\_j + 0.3)$$

The constraints below are on corresponding euclidean distances between i^{th} and j^{th} alpha carbon atoms and centers of radicals:

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \le 0.8$$

$$(2) \quad \sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2} \le 1.2(r_i + r_j)$$

The goal of the willpower folding is to maximize the contacts between the radicals of hydrophobic amino acids, the contacts between the radicals of cysteines and the contacts between the alpha carbon atoms of all amino acids and the radicals with electrical charge not in contact – to find the maximum of the following objective function:

$$(3) \quad F(fold) = \sum_{i\ contact\ j\ redicals} (h_i + h_j + wtw + s_i s_j - c_i c_j) + \sum_{i\ contact\ j\ \alpha\ carbons} wtw,$$

where h_i is the hydrophobicity value of the amino acid, *wtw* is a parameter for the influence of hydrogen bonds.

2.2 Model description

The main purpose of the developed algorithm is to create a structure with low potential energy, starting from randomly folded form, instead of looking for the best structure among wholly randomly generated forms. Each amino acid is represented as the position of the alpha carbon atom and the position of the center of the radical with three-dimensional coordinates. The first stage is to generate a three-dimensional shape by randomly turning the peptide chain in 90 degrees and a distance 1 between the alpha carbon atoms in the peptide chain and between each alpha carbon and the radical center of the same amino-acid. For prevention of making rather spread or rather tight forms, there is used variable constraint of spreading [6]. Also, the center of the radicals (side-chains) of

every next amino acid are not placed at closer distance than 2 (at least 2).

The second stage, which is called "Willpower folding" [7], is to purposefully modify the 3D structure in order to minimize the energy while using now non-integer coordinates. In order to build a more nature like 3D form, we use four descriptors of each type of amino acid, which are taken from their different native (physical, chemical, geometrical) properties – 'h' for hydrophobicity, 'r' is already mentioned for size of the side-chain, 's' for ability to form S-S bridge, 'c' for electrical charge. These descriptors are given in the following table:

name	h	r	s	c
alanine	1.8	0.2	0	0
glycine	-0.4	0.1	0	0
methionine	1.9	0.6	0	0
aspartic acid	-3.5	0.9	0	10
asparagine	-3.5	0.6	0	0
valine	4.2	0.5	0	0
leucine	3.8	0.7	0	0
isoleucine	1	0.7	0	0
glutamic acid	-3.5	0.9	0	10
glutamine	-3.5	0.7	0	0
arginine	-1	0.9	0	0
serine	-0.8	0.4	0	0
threonine	-0.7	0.5	0	0
cysteine	2.5	0.4	10	0
lysine	-3.9	0.8	0	0
histidine	-3.2	0.8	0	0
phenylalanine	2.8	0.7	0	0
tyrosine	-1.3	0.7	0	0
proline	-1.6	0.5	0	0
tryptophan	-0.9	0.9	0	0

Table 1 Descriptors of amino acids

2.3 Willpower folding

The steps of the algorithm for maximizing (3) subject to (1) and (2) are given below:

1. Locate the conditional center as the arithmetic mean of the three directions:
$$x_c = \sum_{i=1}^{n} \frac{xr_i}{n};\ y_c = \sum_{i=1}^{n} \frac{yr_i}{n}; z_c = \sum_{i=1\ldots n} \frac{zr_i}{n}$$
2. Displace the coordinates of the radicals in proportion to their hydrophobicity value h_i, if the $h_i > 0$ direction is to the center of the molecule, and if $h_i < 0$ is opposite: $x_{ni} = x_i+0.02h_i$, $x_c > x_i$; $x_{i\_new} = x_i-0.02h_i$, $x_c<x_i$; $y_{i\_new} = y_i+0.02h_i$, $y_c > y_i$; $y_{i\_new} = y_i-0.02h_i$, $y_c < y_i$; $z_{i\_new} = z_i+0.02h_i$, $z_c > z_i$; $z_{i\_new} = x_i-0.02h_i$, $z_c < z_i$;
3. Displace the coordinates of the alpha carbon atoms to the center of the molecule: $x_{i\_new} = x_i+0.01$, $x_c > x_i$; $x_{i\_new} = x_i-0.01$, $x_c < x_i$; $y_{i\_new} = y_i+0.01$, $y_c > y_i$; $y_{i\_new} = y_i-0.01$, $y_c<y_i$; $z_{i\_new}=z_i+0.01$, $z_c > z_i$; $z_{i\_new} = x_i-0.01$, $z_c<z_i$;
4. Correct the positions of all alpha carbon atoms and radicals in order to preserve the peptide chain, such as alpha-carbon i approaching $i + 1$ and radical i approaching the alpha carbon atom i if $\sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2} \leq 1.1$, $x_{i\_new} = x_i + (x_{i+1}-x_i)/10$; $y_{i\_new} = y_i + (y_{i+1}-y_i)/10$; $z_{i\_new} = z_i + (z_{i+1}-z_i)/10$; $xr_{i\_new} = xr_i + (x_i-xr_i)/10$; $yr_{i\_new} = yr_i + (y_i-yr_i)/10$; $zr_{i\_new} = zr_i + (z_i-zr_i)/10$;
5. Separate the coordinates of the radicals of electrically charged amino acids from one another if $\sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2} \leq 2$, $xr_{i\_new} = xr_i-(xr_j-xr_i)/10$; $yr_{i\_new} = yr_i-(yr_j-yr_i)/10$; $zr_{i\_new} = zr_i-(zr_j-zr_i)/10$; $xr_{j\_new} = xr_j-(xr_i-xr_j)/10$; $yr_{j\_new} = yr_j-(yr_i-yr_j)/10$; $zr_{j\_new} = zr_j-(zr_i-zr_j)/10$;
6. Reduce the distance radical coordinates of closely spaced hydrophobic amino acids if $\sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2} \leq 2$, $xr_{i\_new} = xr_i+(xr_j-xr_i)/20$; $yr_{i\_new} = yr_i+(yr_j-yr_i)/20$; $zr_{i\_new} = zr_i+(zr_j-zr_i)/20$; $xr_{j\_new} = xr_j+(xr_i-xr_j)/20$; $yr_{j\_new} = yr_j+(yr_i-yr_j)/20$; $zr_{j\_new} = zr_j+(zr_i-zr_j)/20$;
7. Reduce the distance radical coordinates of closely spaced cysteine amino acids if $\sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2} \leq 2$, $xr_{i\_new} = xr_i+(xr_j-xr_i)/20$; $yr_{i\_new} = yr_i+(yr_j-yr_i)/20$; $zr_{i\_new} = zr_i+(zr_j-zr_i)/20$; $xr_{j\_new} = xr_j+(xr_i-xr_j)/20$; $yr_{j\_new} = yr_j+(yr_i-yr_j)/20$; $zr_{j\_new} = zr_j+(zr_i-zr_j)/20$;
8. Reduce the distance the coordinates of the alpha carbon atoms of closely spaced amino acids if $\sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2} \leq 2$, $xr_{i\_new} = xr_i+(xr_j-xr_i)/20$; $yr_{i\_new} = yr_i+(yr_j-yr_i)/20$; $zr_{i\_new} = zr_i+(zr_j-zr_i)/20$; $xr_{j\_new} = xr_j+(xr_i-xr_j)/20$; $yr_{j\_new} = yr_j+(yr_i-yr_j)/20$; $zr_{j\_new} = zr_j+(zr_i-zr_j)/20$;
9. Adjust the positions of all alpha carbon atoms to avoid overlapping if $0.6 > \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ as $j \neq i+1$,

and

$$0.9 > \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

as $j = i+1$, $x_{i\_new} = x_i\text{-}0.1(x_j\text{-}x_i)/((x_j\text{-}x_i)^2 + (y_j\text{-}y_i)^2 + (z_j\text{-}z_i)^2)$; $y_{i\_new} = y_i\text{-}0.1(y_j\text{-}y_i)/((x_j\text{-}x_i)^2 + (y_j\text{-}y_i)^2 + (z_j\text{-}z_i)^2)$; $z_{i\_new} = z_i\text{-}0.1(z_j\text{-}z_i)/((x_j\text{-}x_i)^2 + (y_j\text{-}y_i)^2 + (z_j\text{-}z_i)^2)$;

10. Correct the positions of all radicals to avoid overlapping if $0.9(r_i + r_j) > \sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2}$, $xr_{i\_new} = xr_i\text{-}0.1(xr_j\text{-}xr_i)/((xr_j\text{-}xr_i)^2 + (yr_j\text{-}yr_i)^2 + (zr_j\text{-}zr_i)^2)$; $yr_{i\_new} = yr_i\text{-}0.1(yr_j\text{-}yr_i)/((xr_j\text{-}xr_i)^2 + (yr_j\text{-}yr_i)^2 + (\mathrm{zr_j}\text{-}\mathrm{zr_i})^2)$; $zr_{i\_new} = zr_i\text{-}0.1(zr_j\text{-}zr_i)/((xr_j\text{-}xr_i)^2 + (yr_j\text{-}yr_i)^2 + (zr_j\text{-}zr_i)^2)$;
11. Adjust positions to avoid overlapping between the alpha carbon atom and the radical, if $0.9(r_j + 0.3) > \sqrt{(xr_i - xr_j)^2 + (yr_i - yr_j)^2 + (zr_i - zr_j)^2}$, $xr_{i\_new} = xr_i\text{-}0.1(xr_j\text{-}x_i)/((xr_j\text{-}r_i)^2 + (yr_j\text{-}y_i)^2 + (zr_j\text{-}z_i)^2)$; $yr_{i\_new} = yr_i\text{-}0.1(yr_j\text{-}y_i)/((xr_j\text{-}x_i)^2 + (yr_j\text{-}y_i)^2 + (\mathrm{zr_j}\text{-}\mathrm{z_i})^2)$; $zr_{i\_new} = zr_i\text{-}0.1(zr_j\text{-}z_i)/((xr_j\text{-}r_i)^2 + (yr_j\text{-}r_i)^2 + (zr_j\text{-}z_i)^2)$;

2.4 Thermo effect

What we call "Thermo effect" is to make a small random move in every step of Willpower folding, which corresponds to the real environment. This prevents building equal structures if the initial random fold is the same and gives more chance to find better structure. This random value is limited to the half of the main value in each step.

3 Problem Solution

First, using our definition of contact (2), we find the contacts in the real structure of protein 1UUB, using the coordinates of alpha carbons in PDB file of the protein data bank (Figure 1).

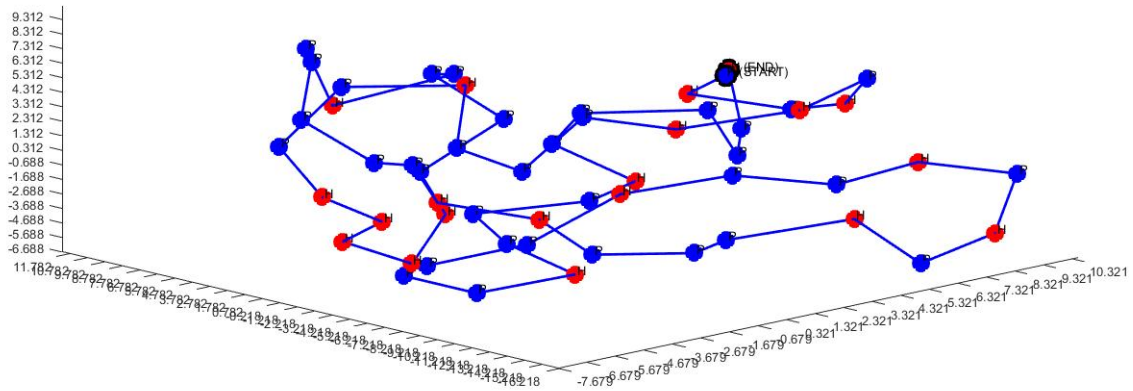

Fig. 1 Real structure of protein 1UUB

There are different 3D structures from crystallographic analysis and for one of them the coordinates of alpha carbons are extracted in the list below:

index	x	y	z
1	-2.218	6.755	3.480
2	-4.621	3.954	4.056
3	-7.988	5.520	3.295
4	-7.352	7.803	2.541
5	-4.113	10.556	2.176
6	-1.323	9.910	-0.509
7	-0.515	6.072	-0.433
8	3.197	5.086	-0.245
9	4.212	3.553	-3.604
10	8.005	3.596	-3.100
11	9.443	4.804	0.307
12	12.187	2.112	0.559
13	12.068	-0.142	-2.577
14	11.132	0.806	-6.186
15	7.324	0.560	-6.688
16	4.874	-2.320	-6.159
17	1.040	-2.273	-6.595
18	-2.213	-3.720	-5.226
19	-5.330	-2.113	-3.563
20	-5.601	0.989	-1.332
21	-8.565	3.100	-0.080
22	-9.103	6.424	-1.965
23	-12.241	7.358	0.083
24	-12.673	10.573	-1.937
25	-12.668	9.795	-5.706
26	-14.112	6.302	-5.914
27	-12.562	4.930	-2.720
28	-10.782	1.519	-3.242
29	-7.707	2.290	-5.333
30	-4.103	0.917	-5.879
31	-0.392	1.344	-4.693
32	0.410	-1.734	-2.470
33	3.871	-0.480	-1.399
34	7.387	0.313	-2.544
35	9.598	-0.882	0.283
36	9.562	-0.541	4.116
37	7.032	-2.287	6.484
38	4.723	-2.774	3.431
39	3.016	0.453	4.721
40	0.454	-1.906	6.433
41	-2.162	-0.992	3.702
42	-2.112	-3.909	1.253
43	-5.415	-5.029	-0.246
44	-6.364	-7.075	-3.349
45	-10.212	-6.890	-3.527
46	-13.385	-5.372	-1.978
47	-13.218	-7.679	1.064
48	-9.590	-6.631	1.648
49	-10.227	-2.931	1.145
50	-13.241	-3.183	3.471
51	-11.616	-5.095	6.386
52	-8.886	-2.430	6.608
53	-10.830	0.858	6.007
54	-14.064	-0.108	4.206
55	-16.218	-1.270	7.116
56	-14.212	-0.494	10.258

There are 125 contacts according to the constraints (2) above.

The test run of the program, realizing the algorithm on 1UUB obtains the following results: among only 100000 randomly generated structures, the value of the evaluation function is 1578.4 and there are 209 contacts – 60 matches with the real one (Figure 2).. The time needed was 70 min on PC, Intel i5, 2.26GHz. The list of the coordinates of alpha carbons carbons are given in the list below:

index(0 for first AA in chain),x,y,z
0,0.984,1.479,-0.860
1,0.595,1.471,-1.802
2,0.132,0.683,-2.314
3,0.428,0.089,-1.874
4,0.859,0.869,-0.388
5,1.751,0.360,-0.755
6,0.546,-0.147,-1.145
7,1.158,0.411,-1.710
8,0.299,0.753,-1.483
9,1.012,0.689,-0.897
10,1.743,0.867,-0.268
11,1.425,0.584,0.857
12,1.940,1.214,0.382
13,2.582,2.052,0.265
14,3.114,2.684,0.912
15,3.834,2.074,1.510
16,3.090,1.806,0.987
17,2.136,1.809,0.899
18,2.464,0.875,0.926
19,2.524,-0.213,1.317
20,1.938,0.432,1.326
21,1.208,0.543,2.303
22,0.676,-0.110,1.448
23,-0.132,0.385,2.137
24,-0.399,1.309,2.425
25,-1.056,0.870,1.722
26,-0.300,0.413,1.33
27,-1.027,1.131,0.888
28,-1.044,1.885,1.511
29,-0.290,1.261,1.690
30,0.586,0.623,1.648
31,1.477,1.044,1.654
32,0.752,0.960,0.939
33,0.755,1.319,1.811
34,1.268,1.558,1.044
35,0.319,1.700,1.189
36,-0.309,2.236,0.649
37,-1.160,1.891,0.653
38,-1.404,2.169,-0.187

39,-1.637,1.267,-0.477
40,-1.726,0.275,-0.720
41,-1.740,-0.790,-0.716
42,-1.974,-1.355,-1.362
43,-2.695,-2.118,-1.634
44,-3.332,-2.900,-1.226
45,-2.590,-3.598,-0.945
46,-2.282,-2.706,-1.170
47,-2.441,-1.751,-0.830
48,-1.733,-1.921,-0.078
49,-1.711,-3.015,-0.185
50,-1.774,-2.299,-0.780
51,-1.506,-2.045,-1.585
52,-0.997,-2.679,-1.009
53,-0.0354,-2.635,-1.633
54,-0.647,-1.702,-1.267
55,-0.871,-2.453,-1.855

The other experiment we done, is to find out how much better structure could be built, classifying fold not by evaluation function, but by ratio contacts matching ratio. In this way we obtain the following structure with 62 contacts matches out of 120 (Figure 3). The list of the coordinates of alpha carbons carbons are given in the list below:

index(0 for first AA in chain),x,y,z
0,-0.928,-0.754,-0.314
1,-1.907,-0.841,0.149
2,-1.690,-1.128,1.196
3,-1.002,-1.843,1.631
4,-1.113,-2.201,0.609
5,-0.643,-3.032,0.072
6,0.193,-3.319,0.710
7,0.373,-4.382,0.479
8,0.799,-5.388,0.504
9,1.345,-5.157,1.162
10,1.250,-4.519,1.965
11,0.779,-4.836,2.897
12,1.105,-5.664,2.807
13,1.990,-6.288,2.973
14,2.144,-6.985,3.815
15,2.017,-8.015,4.178
16,1.943,-9.046,4.533
17,2.067,-9.92,4.720
18,2.412,-10.694,5.496
19,2.331,-12.359,6.122
20,1.371,-12.539,6.237
21,0.373,-12.491,5.327
22,0.481,-11.828,4.463
23,0.726,-12.599,4.078
24,1.492,-13.371,4.144
25,2.192,-13.415,4.979
26,1.543,-12.748,5.496
27,1.025,-11.796,5.323
28,1.837,-11.080,5.451
29,2.691,-11.313,5.256
30,3.177,-12.009,4.658
31,2.369,-12.077,4.267
32,1.818,-11.218,3.885
33,1.333,-10.919,2.623
34,0.247,-10.731,0.678
35,-0.674,-10.534,0.938
36,-1.684,-10.816,0.609
37,-2.431,-10.0376,0.476
38,-2.578,-9.671,-0.536
39,-2.916,-9.502,-1.575
40,-2.751,-8.954,-2.508
41,-2.453,-9.365,-3.249
42,-2.155,-10.230,-3.648
43,-2.500,-10.539,-4.535
44,-3.422,-11.087,-4.778
45,-4.116,-11.888,-4.507
46,-4.582,-12.808,-4.904
47,-3.998,-13.677,-5.224
48,-3.159,-13.512,-4.923
49,-3.025,-12.638,-5.225
50,-2.631,-12.076,-6.079
51,-3.242,-12.582,-6.384
52,-3.202,-13.470,-7.010
53,-2.528,-14.188,-6.569
54,-1.799,-13.595,-5.997
55,-1.215,-13.086,-6.457

Remark: for the source code in C++ and the list of contacts' couples, mail to itodorin@gmail.com.

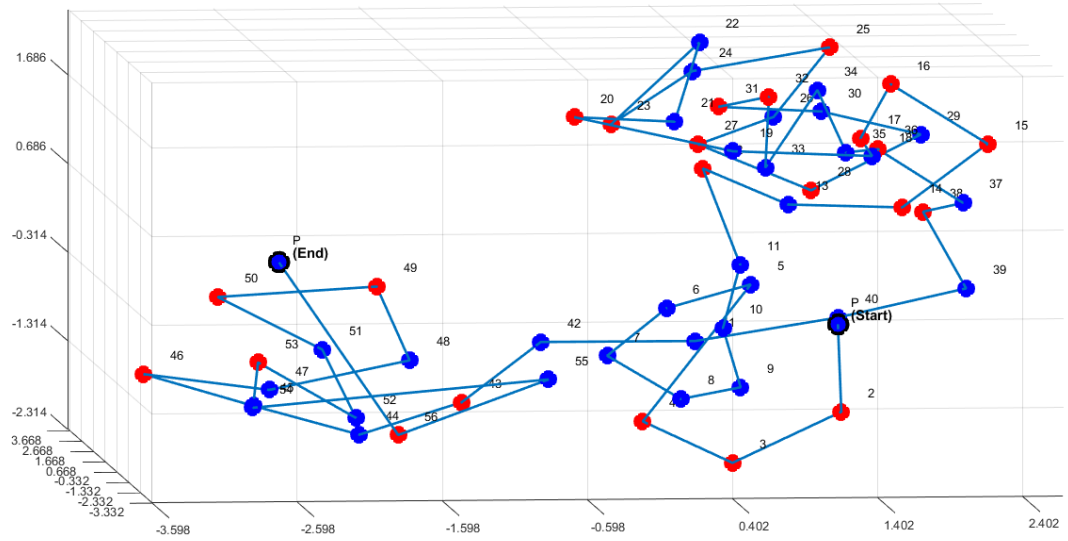

Fig. 2 The structure with 60 matches

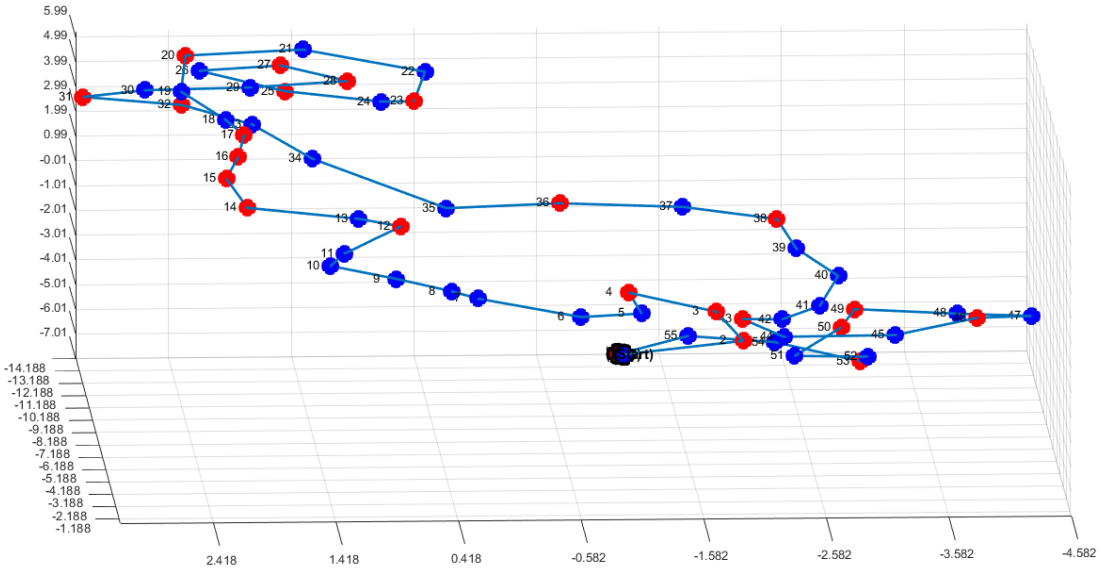

Fig. 3. The structure with 62 contacts matches

4 Conclusions

This approach for protein folding prediction, using “Willpower folding” with “Thermo effect”, appear to be faster than models for finding best structure among randomly built structures. Therefore it is possible rather more structures of proteins to be processed for the same time. Even more, the model is at age under year and might be further developed in direction of granularity – placing every atom, and the evaluation function could take into account the real physical values of the impact of the distribution of electronic density – such degree of granularity is hard to be executed considering computational time through models for finding best structure among randomly built structures.

5. Acknowledgements

This work is partially supported by the project of the Bulgarian National Science Fund, entitled: "Bioinformatics research: protein folding, docking and prediction of biological activity", code NSF I02/16, 12.12.14.

Reference

1. A. Kolinski and J. Skolnick, Monte Carlo simulations of protein folding. i. lattice model and interactionscheme, *Proteins*, vol. 18, 1994, pp. 338-352.
2. B. Berger, and T. Leighton, Protein folding in the hydropho-bichydrophilic (HP) model is NP-complete, *Journal of Computational Biology*, vol. 5, 1998, pp. 27-40.
3. H. Greenberg, W. Hart, G. Lancia, Opportunities for combinatorial optimization in computational biology, *INFORMS Journal on Computing*, vol. 16, 2004, pp. 211-231.
4. P. Milanov, I. Trenchev, N. Pencheva, Explicit description of the set of all theoretical genetic codes, Mathematica Balkanica, vol. 17, 2004, pp. 157-165.
5. S. Istrail, and F. Lam, Combinatorial algorithms for protein folding in lattice models: A survey of mathematical results, *Commun. Inf. Syst.*, vol. 9, 2009, pp. 303-346.
6. I. Todorin, A Model for HP folding prediction using variable size of lattice, *Proceedings of the Fifth International Scientific Conference – FMNS2013*, vol. 1, 2013, pp. 188-192.
7. I. Todorin, Protein Folding Prediction using two stages over random made structure – “Bridge Building” and “Willpower Folding”, *SDNS*, vol. 1, 2017, pp. 1-6 (in printing).