STRUCTURAL COMPUTATIONAL BIOLOGY STUDIES OF A NON-SPECIFIC DNA-BINDING PEPTIDE

Democritus University of Thrace
Department of Molecular Biology and Genetics
Msc TRANSLATIONAL RESEARCH IN BIOMEDICINE: MOLECULAR DIAGNOSTICS, BIOMARKERS AND TARGETED THERAPIES

Dimitra Alexiadou

under the supervision of Professor Nicholas M. Glykos, to whom I am sincerely grateful

Structural computational biology studies of a non-specific DNA-binding peptide

1. INTRODUCTION

1.1. Backround

Several theories have been developed over the years for predicting the three-dimensional structure of proteins. This need originated from the importance the three-dimensional structure plays for the function of a protein, its interaction with other molecules, its biological importance in health and disease. Understanding this procedure would provide insight on how little changes lead to different properties, on an otherwise identical sequence, and what impact that had on evolution. Additionally, it would allow us to manage usefully data emerging from genome sequencing and, more importantly even design proteins with on-demand properties.

Early in the history, it was simply assumed that proteins fold through distinct intermediate states in a distinct pathway. It was Anfinsen who demonstrated that proteins can fold spontaneously, without any outside intervention and also reversibly. Then, Levinthal perceived that the native structure could not be reached by random searching through the vast number of structural options. He placed the protein folding problem as two mutually excluding goals, either this happens after thorough undergoing of all different structures to find the most thermodynamically stable one, that takes a lot of time to achieve, or after quick transition on a predetermined pathway that may lead to local optima.[1,2,3] Baldwin took up the challenge and tried to experimentally define kinetic folding intermediates and pathways, combined with theory. He, thus, helped to establish the multipath funneled energy landscape.

The energy landscape is a theory by which proteins obtain their native structure. This theory suggests that folding occurs through an ensemble of trajectories rather than through only a few uniquely defined structural intermediates. The energy landscape represents the possible trajectories as a funnel-shaped landscape, each point symbolizing the thermodynamical state of the protein in regard with time, biased down toward the native, most thermodynamically stable, structure.[4] Through small random changes made repeatedly, the system moves to conformations of lower energy. The folding trajectory is therefore formed by many moves, carried out in succession, driven by the principal of the protein being more thermodynamically stable.

Defining how the three dimensional structure is encoded in the sequence of the protein is a great challenge, mainly because the basis is the weak-local preference for one structure over

another. During the first stages of folding, the weak-local preferences of secondary structures along the sequence, together with the free-energy preferences that tend to bury hydrophobic amino-acids in the center of the peptide, while exposing the polar ones on the surface, lead the peptide to conformation. [5]

As far as forces are concerned, the interactions stabilizing a protein structure also guide the protein in attaining that structure. The folding of polypeptide chains is determined by many noncovalent interactions i.e. canonical forces: the hydrophobic effect, conventional hydrogen bonding, Coulombic interactions, and van der Waals interactions.

The challenges that continue to arise in the field of protein structure prediction suggest that there is an incomplete understanding of the topic. Apparently, canonical forces alone seem to be insufficient when it comes to protein biophysics. Therefore, researchers suggested that additional interactions are to be taken into consideration, contributing to protein folding.[1] These are the secondary forces and consist of $n \rightarrow \pi^*$ Interactions, C–H···O Hydrogen bonds, π–π Interactions, C5 Hydrogen bonds, Cation–π interactions, Sulfur–arene interactions, Anion–π interactions, Chalcogen bonds, X–H···π Interactions, but describing them in more detail is not within the aim of this paper. [6]

The folding rate for small proteins has been linked to the contact order. That is measure of the locality of the contacts between the amino acids in tertiary structure. It is equal to the average sequence distance between residues in the folded protein divided by the total length of the protein. Higher contact orders indicate longer folding times and low contact order has been suggested as a predictor of folding that occurs without a free energy barrier.

Since folding is a reaction that is rate-limited by an entropic barrier, increasing the proportion of contacts that are short-distance in the native state and thus during the transition state, accelerates folding through two mechanisms: by locating the transition state with diffuse search and by lowering the entropic barrier.

Generally, a-helices fold faster, because they have a larger proportion of sequentially short-range contacts. In other words, the transition states of a-helices include many of the native interactions, so that the transition state has a conformation close to that of the native state.

For small proteins, there is usually a single folding nucleus, around which the entire folding procedure occurs, meanwhile, peripherally, the protein is able to explore different conformations . For larger proteins, multiple sites serving as folding nucleus are likely to exist and lead to more complex procedures, including a range of intermediates. For proteins

with multiple nucleation sites, this step is often the slowest one in the entire folding process.

Point mutations affect significantly the folding rate, because the interactions involved in the folding nucleus depend on specific features of the side chains. This is the basis for protein engineering experiments, which show that folding rates exhibit an approximately exponential dependence on changes in free energy, changing with regard to these point mutations.

The encorporation of computers in the field gave rise to molecular dynamics simulation, that is, the in silico prediction of the atom position and movement of a biomolecular system, e.g a protein in water, by calculating the forces each atom pose on each other, given an original position. As far as proteins are concerned, molecular dynamics simulation is the trajectory a peptide would follow to end up having its final atomic-level configuration. The ability of in silico simulation over exprerimental laboratory work is more powerful, in the sense of it being able to presicely control the conditions and also being able to capture the position and motion of every atom at every moment. [5, 7]

1.2. Protein folding motifs

Each protein is a unique sequence of amino acids and this sequence is the fingerprint of the protein, meaning that it defines its structure and, thus, its function. The sequence is called the primary structure of the protein, in other words, the linear way the amino acids are bound to one another with peptide bonds. It is worth mentioning that a peptide bond has the character of a double bond, meaning it cannot freely rotate.

The next step to protein folding is secondary structure. In this step, proteins form folding patterns locally, as a result from its backbone interactions. This means that these motifs are the result of interactions between components of amino acids that are common to all amino acids, more specifically, through hydrogen bonds between the amino hydrogen and carboxyl oxygen atoms in the peptide backbone.

The secondary structure includes several motifs that have been observed in protein folding. The most common motifs are alpha helices and beta sheets. Other motifs, that are, however, more rare to encounter, are 3 10 helices, pi helices, α/β/γ/δ/π turns, polyproline helices, alpha sheets. Finally, a structure called coil is observed, but is not strictly a motif, rather than a conformation not fitting to the aforementioned structures.

-the alpha helix:

this is a conformation of the backbone rotating around an axis clockwise as a helix, forming a cylinder. In this pattern, a hydrogen bond is formed between every fourth peptide bond (i → i + 4 hydrogen bond), linking the C=O of one peptide bond to the N–H of the fourth in line, taking 3,6 amino acids to complete a full turn. Alpha helices can fold around each other forming coiled coils.

-the beta sheets

this is, the backbone of the polypeptide chain forms a zigzag (β conformation) rather than helical structure. The arrangement of several zigzags side by side is called a β sheet. The individual polypeptide segments that have a zigzag form lead to a pleated appearance of the overall shape. Hydrogen bonds form between adjacent chains within the sheet. The individual zigzags that unite to form a β sheet are usually close together on the polypeptide chain but can also be distant from each other in the linear sequence of the polypeptide; they may even be in different polypeptide chains.

-the 3 10 -helix

It is the fourth most common type of secondary structure in proteins, commonly found as N- or C-terminal extensions to an α-helix and have been proposed to be intermediates in the folding/unfolding of α-helices. It is characterized by an i → i + 3 hydrogen bond, in contrast to alpha helix.

- pi helices

This is a helix with i → i + 5 hydrogen bond. [8,9,10,11]

1.3. The peptide

The peptide that will be used for the purpose of this essay is a 12-mer peptide with SVSVGMKPSPRP sequence.

This peptide has been prove to adhere to DNA after an experiment conducted by Wolcke and Weinhold . For this purpose ,Wolcke and Weinhold selected this peptide from a random peptide phage display with competitive elusion using DNA methytransferase M TaqI and then ELISA. [12]

The SVSVGMKPSPR peptide provided the base for many studies and patents based on its affinity to organic and inorganic targets.

Some of the properties this peptide was proven to have, is that it recognized tumor-related neovasculature, but not normal blood vessels in mice bearing human tumors. Also, it increased the survival rates of xenografted mice with human lung and oral neoplasms

when it was linked to the liposome-carrying doxorubicin. In other works, the peptide was founded to have affinity for a monoclonal antibody against the envelope protein of Japanese encephalitis virus and for monoclonal antibody 2G12, which neutralizes human immunodeficient virus 1 (HIV-1). Additionally, it was found to bind specifically to carboxypeptidase B from pig pancreas and to endothelial cells, to the cationic amino acid transporter 1, the human prostate cancer cell line DU145, Torpedo acetylcholinesterase, and HIV-1 Virion infectivity factor protein. The SVSVGMKPSPRP peptide was found to bind to the Eph family of receptor tyrosine kinases, but its affinity was not evaluated.

Numerous publications have also suggested the adhesion of SVSVGMKPSPRP onto inorganic targets. The SVSVGMKPSPRP peptide was identified for binding on GaN and hydroxyapatite with strong affinity and specificity and was able to function as a template for the synthesis of cobalt-platinum nanoparticles. It was isolated as single- walled carbon nanotubes (SWNTs) binding peptide and was also identified as a pigment-binding peptide and proposed as a seed for metal nucleation. Furthermore, the peptide was identified for binding to various semiconductor materials like gallium arsenide, gallium antimonide, zinc telluride, zinc selenide and cadmium selenide. Small polypeptides as such, with specific identities, could be incorporated into materials science and engineering. [13]

1.4. Principal Component Analysis

Principal Component Analysis (PCA) is is a statistical technique used to identify patterns in big data sets. It finds use in determining similarities and differences. Especially when when high dimensions of data are concerned, PCA can surpass the difficulty of representing the data in graphics and compress them without significant loss of information.

The first step is calculating PCA1, which is the line that passes through the data in the direction of the biggest variance, given that the data are represented on a Cartesian coordinate system. Then PCA2, which is perpendicular to PCA1, describes in a linear form what is not described in PCA1. If a third dimension is to be taken into consideration, that would be perpendicular to the previous two PCAs and so on.

The first step to calculating PCA1, is to find the minimum of the sum of squared distances the data have from the beginning of the system. The sum of squared distances is called the eigenvalue of PCA1 and the unit vector that represents PCA1 (a vector is parallel to the line that best describes the data and has a length of 1 unit of measurement) is called

the eigenvector of PCA1. The higher the eigenvalue, the higher is the variance of the data in that direction, eg how much the data are spread on the line. [14]

1.5. Root Mean Square Deviation (RMSD)

Root Mean Square Deviation (RMSD) or Standard deviation of the residuals or Root Mean Square Error is RMSD is a standard measure of structural distance between coordinate sets. That is, a statistical technique, used broadly in bioinformatics, to calculate the distance ie the disagreement, between the atoms of a molecule and a linear regression model or a molecule and a superimposed structure. This is a useful tool, when it comes to grouping structures that have similar structures, as by calculating the values, we can quantify the similarity of the structures. [15]

2. PURPOSE

The essay is a hands-on application of molecular dynamics simulation. The purpose of this essay is to explore the possible three-dimensional states that the peptide SVSVGMKPSPRP can acquire, using molecular dynamics simulation. Then an attempt to depict the three-dimensional states with graphics was made. After the determination of the thermodynamically most stable forms of the peptide, the ability of each of these forms to attach to DNA was assessed.

3. METHOD

For the completion of this essay, several steps where involved. The first step was the acquisition of the full sequence of the peptide, including the coordinates of the atoms involved in the formation of the peptide, in a .dcd form and two different initial trajectories where chosen. Along with the sequence, information about forces, velocities regarding the peptide was acquired, in a .psf form. All the subsequent analyses were performed twice, once for each of the two initial trajectories, with the fisrt one being *99sb* and the second one *star.*

Then, with the use of *grcarma*[16], the graphical user interface of a software which encodes automated techniques for the analysis of of biomolecular simulations, the secondary structure of the peptide was depicted in relation with time. The conclusion out of

these diagrams was that amino acids 2-10 had the most stable structure and thus, were used for the next step, again with the use of grcarma. This step included the determination of the RMSD matrix for each of the two initial forms of the peptide, showcasing in clusters, the structures which the peptide tend to form more often.

After RMSD matrix determination, dPCA was performed for the 2-10 amino acids of both the initial forms, and thus, isolating on a per cluster basis, average structures, representative structures, corresponding to the atomic root mean root fluctuations.

With the use of *RasMol*[17], the representative forms were visualized as three dimensional structures and a discussion was made, as to how the three-dimensional structure would affect the binding of the peptide on DNA.

The next step evaluated the ability of the peptide to dock onto DNA with the use of the program *HDOCK*[18] and the visual result was produced with the use of *RasMol* and *PyMOL*[19]. All dynamic simulations were performed with the use of the NAMD program.[20]

4. RESULTS

4.1. THREE-DIMENSIONAL STRUCTURE

4.1.1. Secondary structure

The .dcd and .psf files of the peptide were used as an input to the grcarma software. Two different initial trajectories were used for the peptide, the *99sb* trajectory and the *star* trajectory and all calculations were performed twice, once for each trajectory, in order to eliminate false positive results and compare the results. An estimation of secondary structure was calculated and the results are depicted on Figure1 for *99sb* and Figure2 for *star*, showing the participation of each residue of the peptide in the different patterns the secondary structure can acquire, in regard with time. Appendix1 can be used as a tool for interpretation of the diagrams.

A profound observation is that the peptide is mostly comprised of B/G turns but also, in a descending order, of B sheets, 3-10 helices, A-helices.

What is also clear from figures 1& 2, is that the terminal residues 1, 11, 12 are mostly involved in coils rather than fixed patterns, meaning they are mobile, thus they do not play a significant role in the formation of the definite three-dimensional structure of the peptide

and this is the case for both of the initial formations of the peptide. Therefore, they are not included in the simulation experiments that follow.

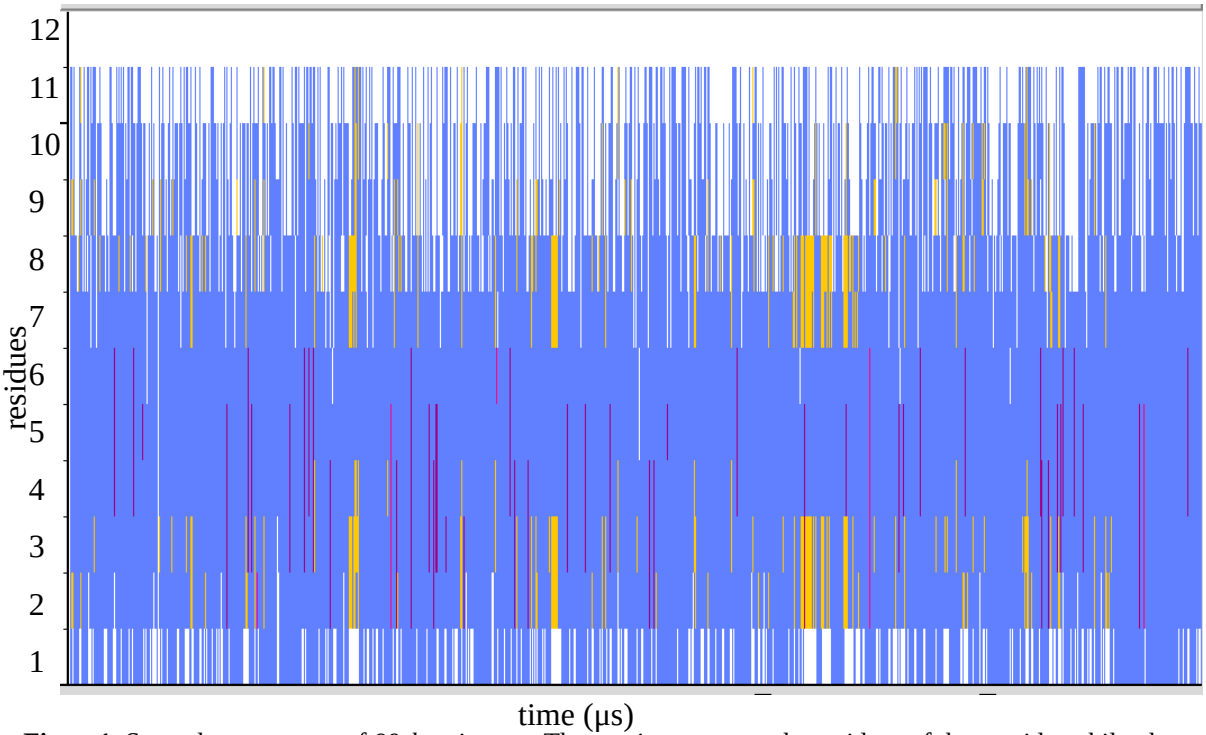


Figure1. Secondary structure of *99sb* trajectory. The y axis represents the residues of the peptide, while the x axis is time. With white color representing coils, it is easily noticed that residues 1, 11 and 12 are not involved, for a significant amount of time, in the formation of the secondary structure and therefore, do not play an important role in the three-dimensional shape of the peptide.

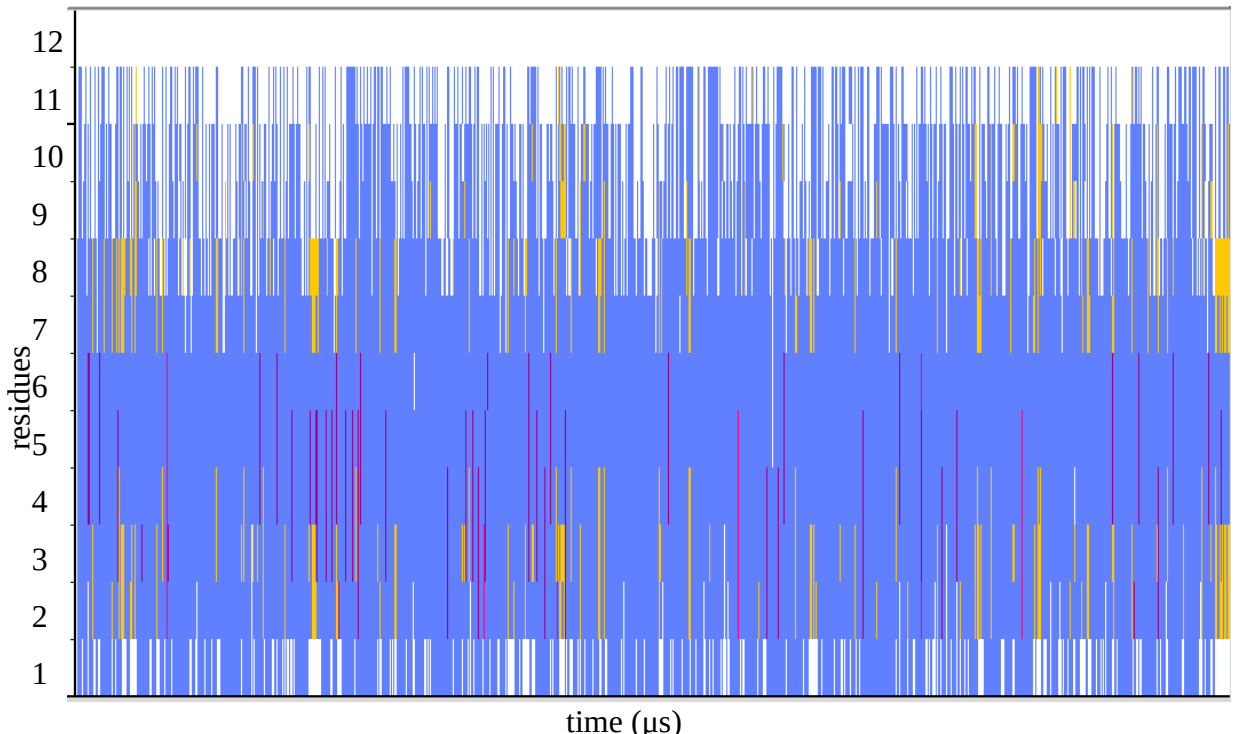
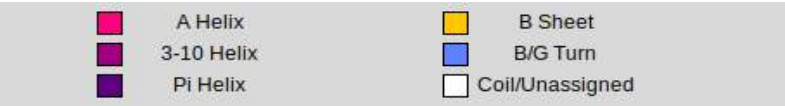


Figure2. Secondary structure of *star* trajectory. The same observations as with 99sb trajectory can be induced from the diagram. Residues 1, 11 and 12 are mostly involved in coils and therefore do not play an important role in the in the three-dimensional shape of the peptide.



Appendix1. Different patterns of secondary structure and their color correspondence on the diagrams shown in Figure1 and Figure2.

4.1.2. RMSD matrix

With the use of grcarma and the input of .psf and .dcd files, the RMSD matrices of the two trajectories were calculated.

The matrices are a frame-to-frame comparison of determined structures, with a user selected number and types of atoms and residues.

A step of 1000 between the frames was chosen, i.e . every other 1000th frame was taken into consideration and the result was a 12458x12458 matrix for the *99sb* trajectory and a 12400x12400 matrix for the *star* trajectory. The RMSD matrices are depicted on Figure3 and Figure4 respectively.

A color-coded scale, visible in Appendix2, represents the frequency by which the different structures are met. Areas with dark blue color represent structures that the peptide conforms into more often, while brown areas represent structures being met less often.

It is easily observed that the matrices are symmetrical to the diagonal that joins the left upper corner with the right lower corner, since both the vertical and the horizontal axes are the same values compared with each other.

Appendix2. The color coded scale used in RMSD matrices to describe the RMSD value of the frames compared. Colors placed from left to right correspond to a declining order of RMSD values, i.e. increased stability.

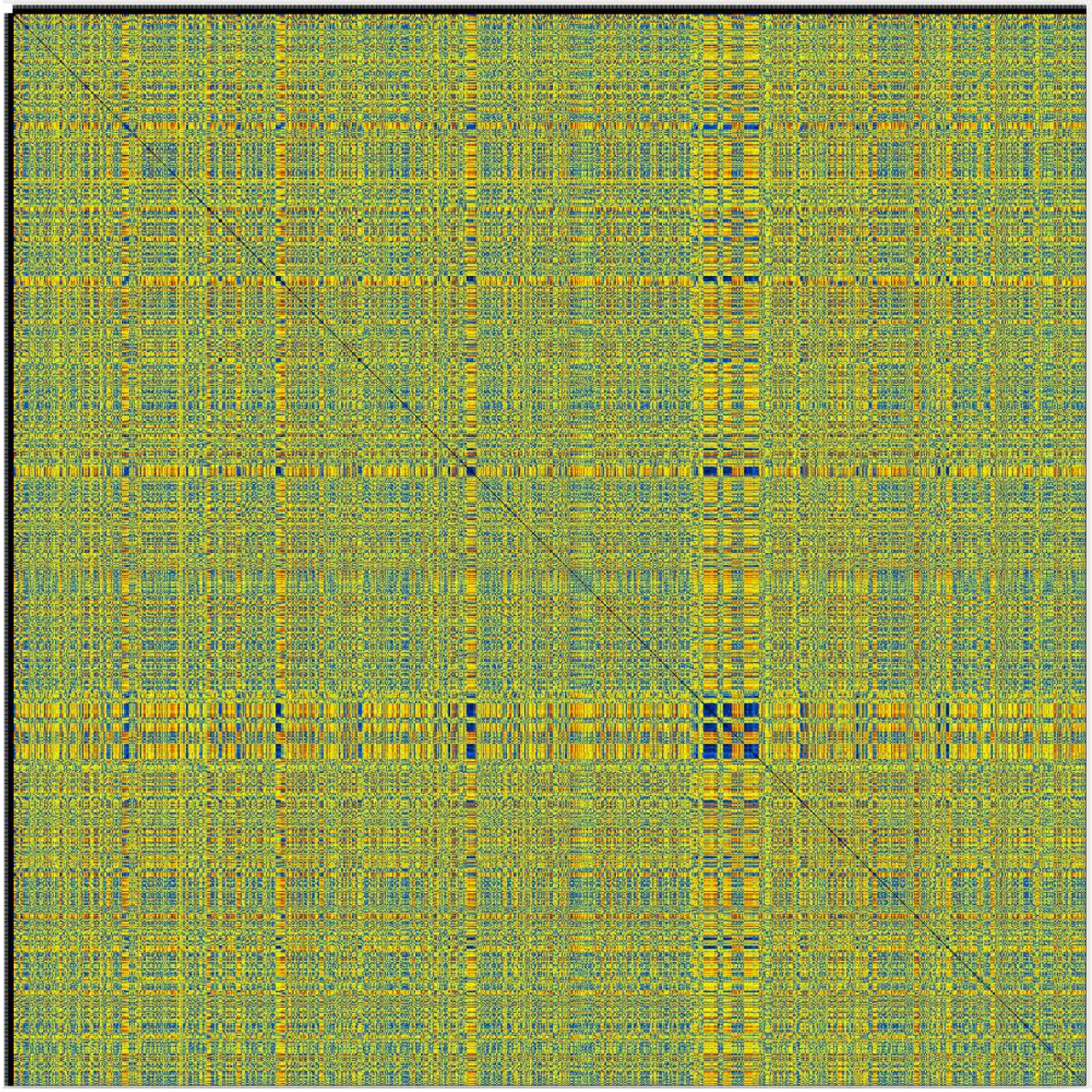


Figure3. RMSD matrix of 99sb trajectory. 12458 frames were compared with each other and the result is a matrix symmetrical to the diagonal passing from left upper corner to right lower corner. Blue areas indicate low RMSD values between successive structures, and thus indicate that a structurally stable conformer has been located.

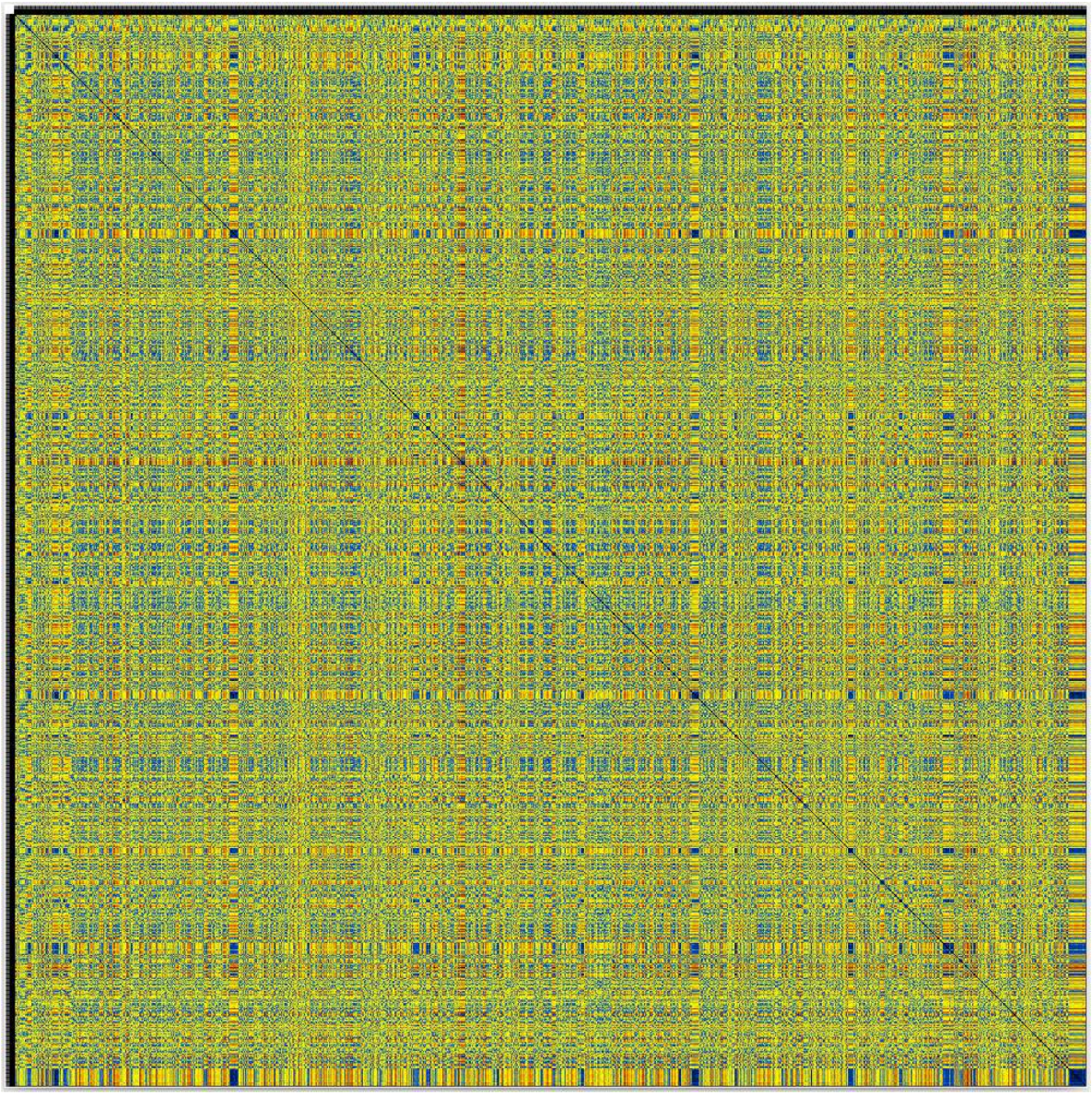


Figure4. RMSD matrix of star trajectory. 12400 frames were compared with each other and the result is a matrix symmetrical to the diagonal passing from left upper corner to right lower corner. Blue areas indicate low RMSD values between successive structures, and thus indicate that a structurally stable conformer has been located.

4.1.3. dPCA without Ch1

During this step the .psf and .dcd files were used again as an input to grcarma and the dihedral Principal Component Analysis (dPCA) of the trajectories was calculated. For both of the initial trajectories, the selected step between the frames was 500 and only residues 2-10 was taken into consideration, in accordance with the findings of the secondary structure on the first step.

For this calculation, only the backbone structure was used, without the side chains (Ch1). The results were 10 clusters for each intitial trajectory, meaning that the calculation formed 10 different groups with structures similar within the same group, but different between the

groups for *99sb* and 10 groups for *star.*

Further analysis isolated the frames corresponding to each cluster and gathered ,on a per cluster basis, the average structures, representative structures and a superposition of 500 equally spaced structures from each cluster.

A selection of the results is depicted on Figure5, Figure6, Figure7, Figure8 for the *99sb* trajectory and on Figure9, Figure10, Figure11, Figure12 for the *star* trajectory.

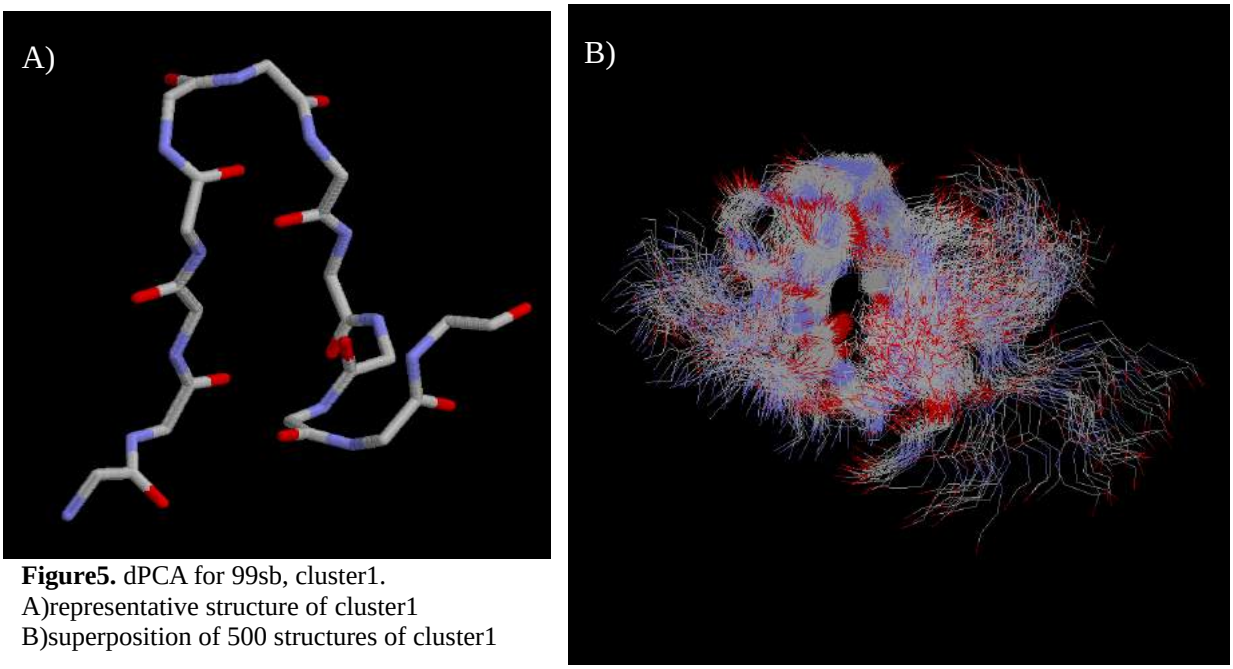
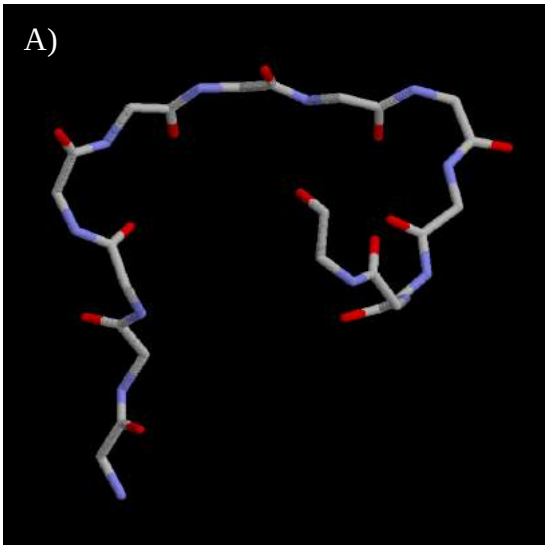


Figure5. dPCA for 99sb, cluster1.
A)representative structure of cluster1
B)superposition of 500 structures of cluster1



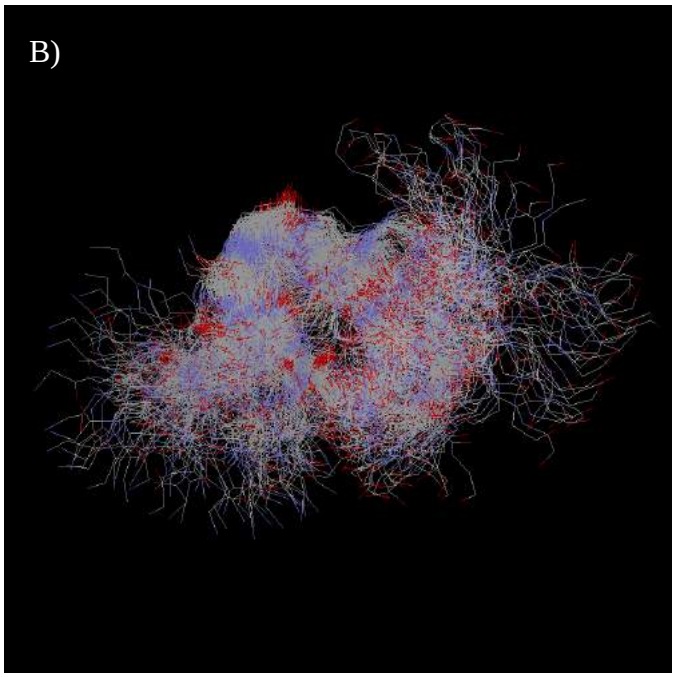
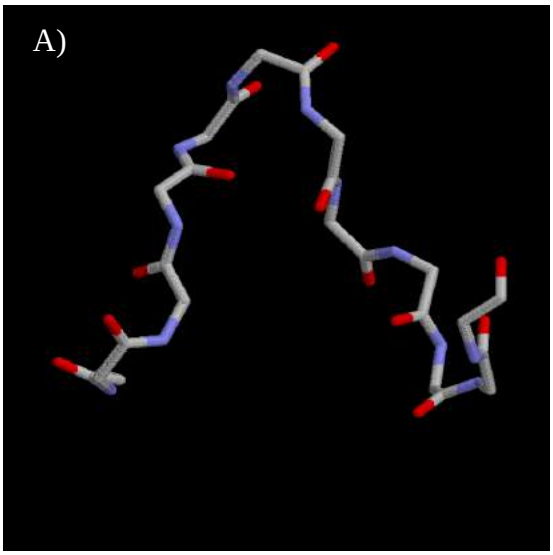


Figure6. dPCA for 99sb, cluster3.
A)representative structure of cluster3
B)superposition of 500 structures of cluster3



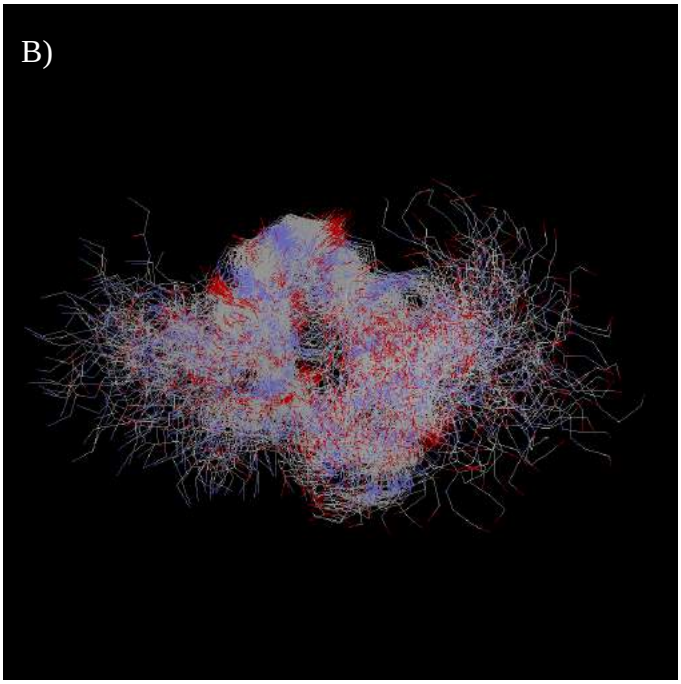
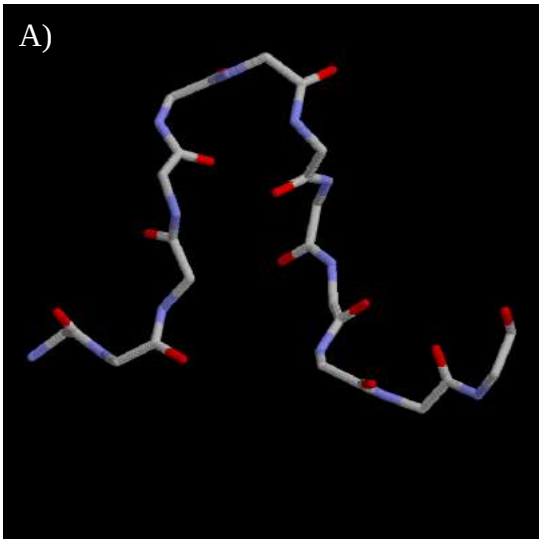


Figure7. dPCA for 99sb, cluster5.
A)representative structure of cluster5
B)superposition of 500 structures of cluster5



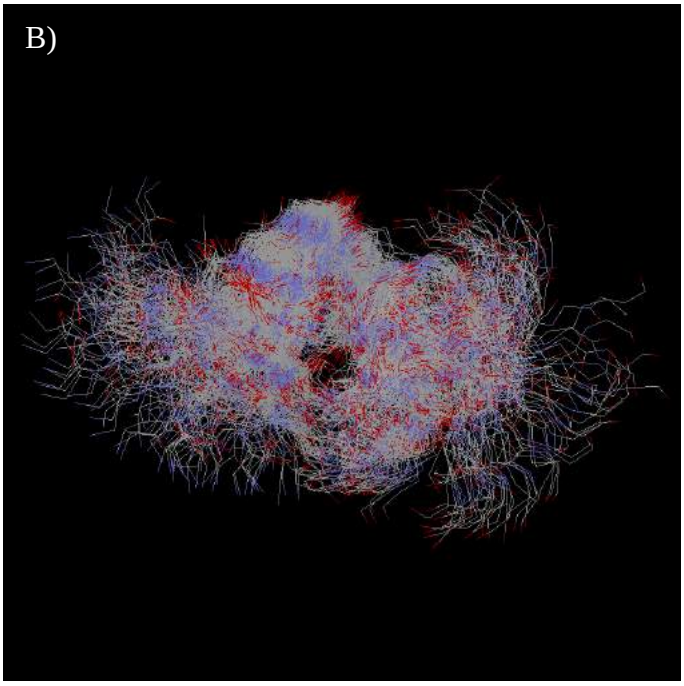
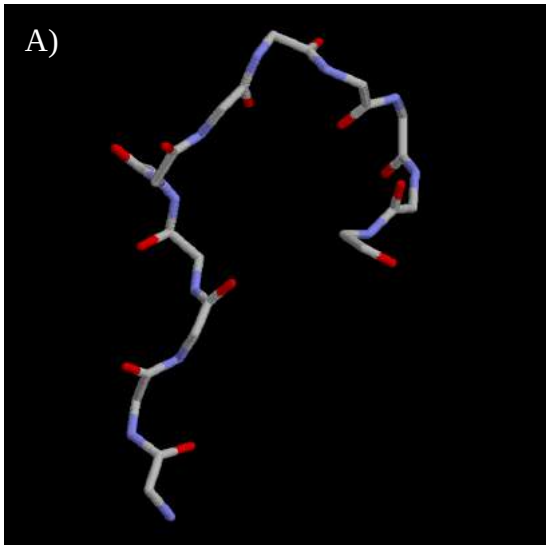


Figure8. dPCA for 99sb, cluster9.

A)representative structure of cluster9

B)superposition of 500 structures of cluster9



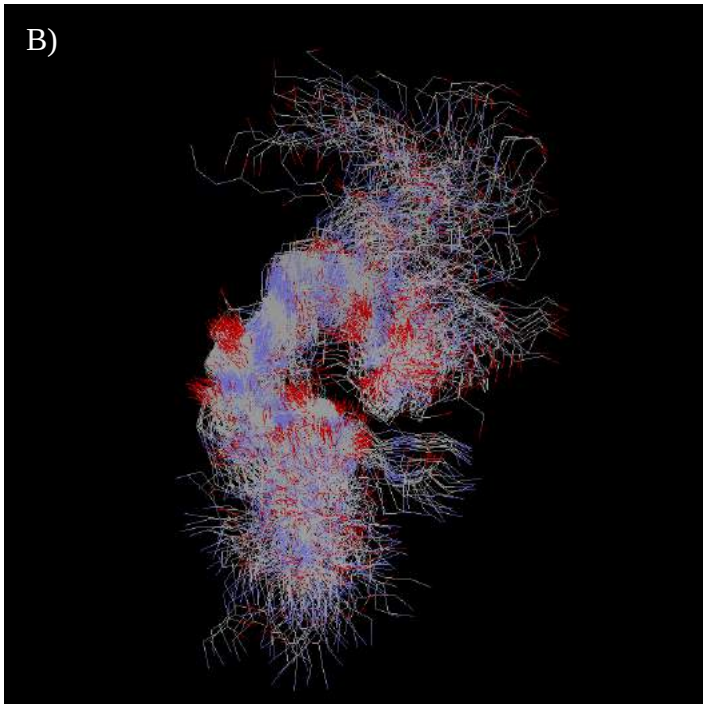
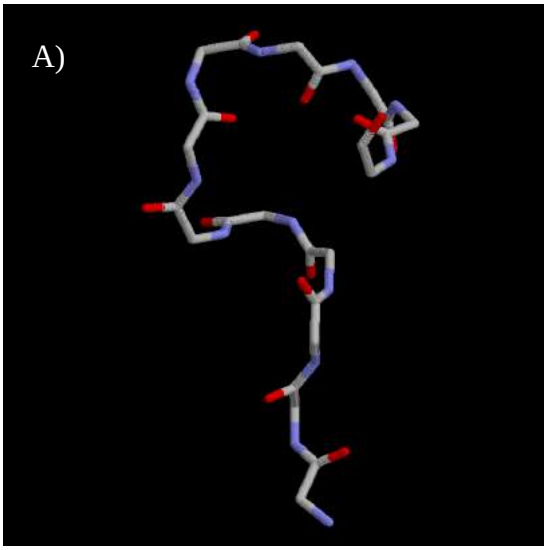


Figure9. dPCA for star, cluster2.
A)representative structure of cluster2
B)superposition of 500 structures of cluster2



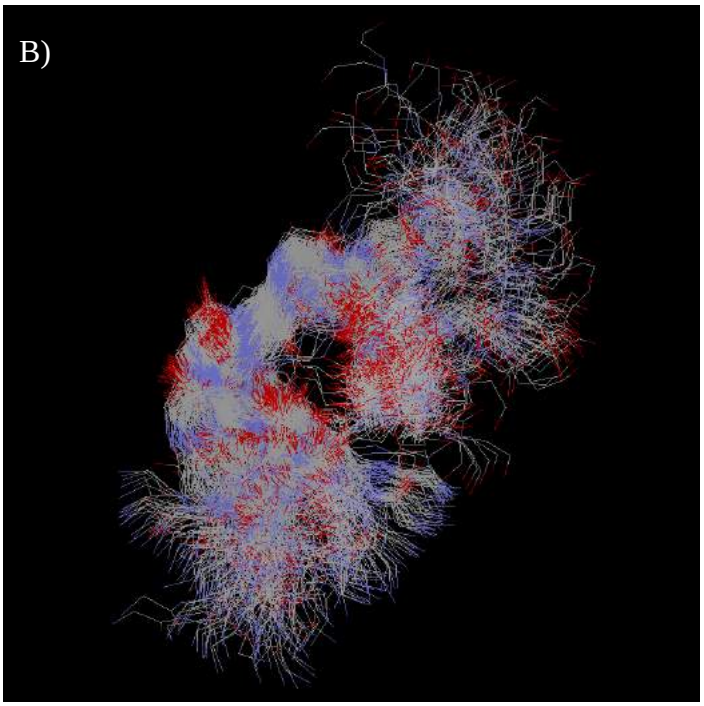
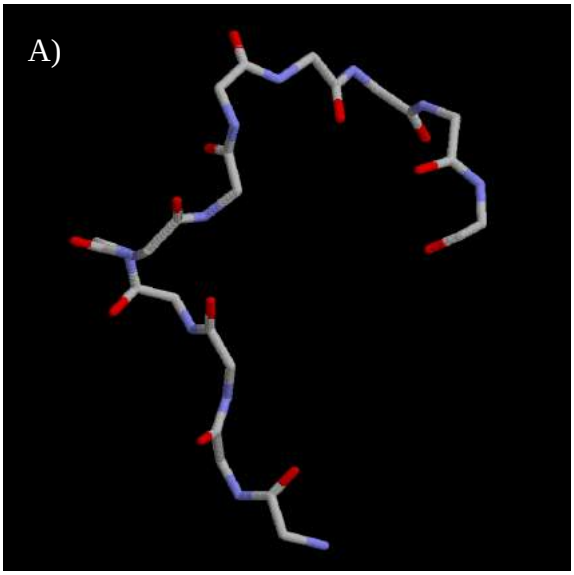


Figure10. dPCA for star, cluster3.
A)representative structure of cluster3
B)superposition of 500 structures of cluster3



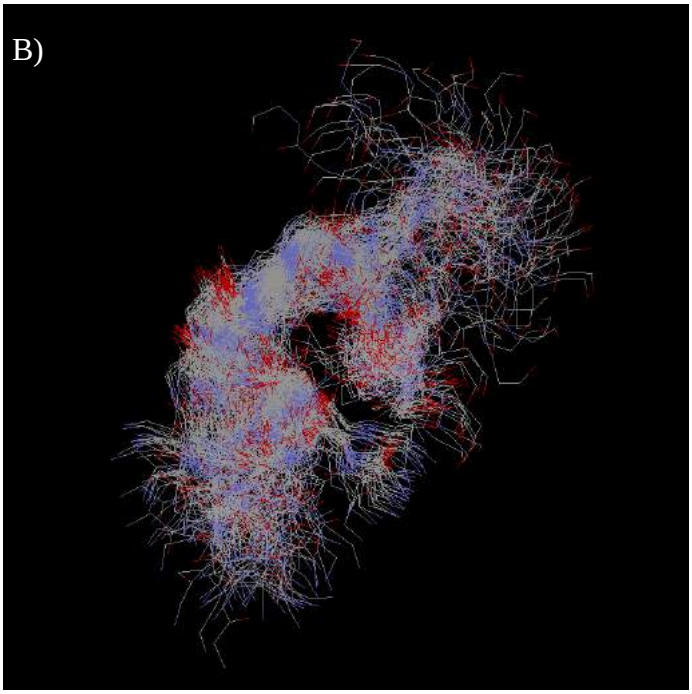


Figure11. dPCA for star, cluster8.
A)representative structure of cluster8
B)superposition of 500 structures of cluster8

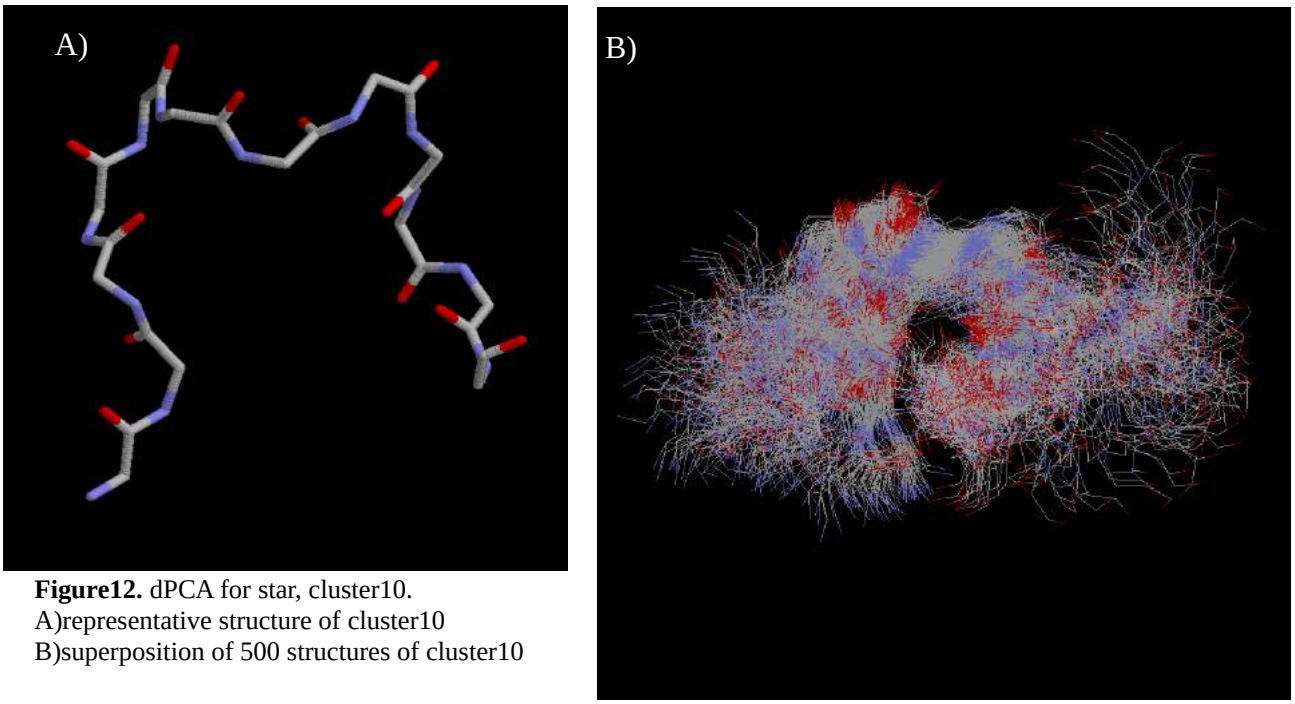


Figure12. dPCA for star, cluster10.
A)representative structure of cluster10
B)superposition of 500 structures of cluster10

The diagrams were made with the use of *RasMol* and the colors correspond to different molecular parts of the residues. Thus, grey represents the carbon groups, red the hydrogen groups and blue the nitrogen groups.

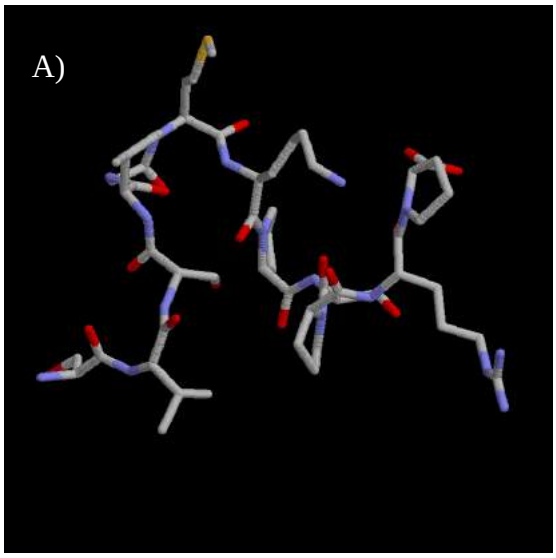
The peptide shows a preference in "c" shaped structures, rich in sheets rather than helices, which explains its fitting with the niches of the DNA, in the chapters that follow.

4.1.4 dPCA with Ch1

For this calculation, the backbone structure was used along with the side chains (Ch1). The results were 3 clusters for the 99sb trajectory and 4 clusters for the star trajectory. Again, the frames corresponding to each cluster were isolated and on a per cluster basis, the average structures, representative structures and a superposition of 500 equally spaced structures from each cluster were gathered.

A selection of the results is depicted on Figure13 for the *99sb* trajectory and on Figure14 for the *star* trajectory.

Again, the graphs were made with the use of *RasMol* and the colors correspond to different molecular parts of the residues. Grey represents the carbon groups, red the hydrogen groups, blue the nitrogen groups and yellow the sulfur groups.



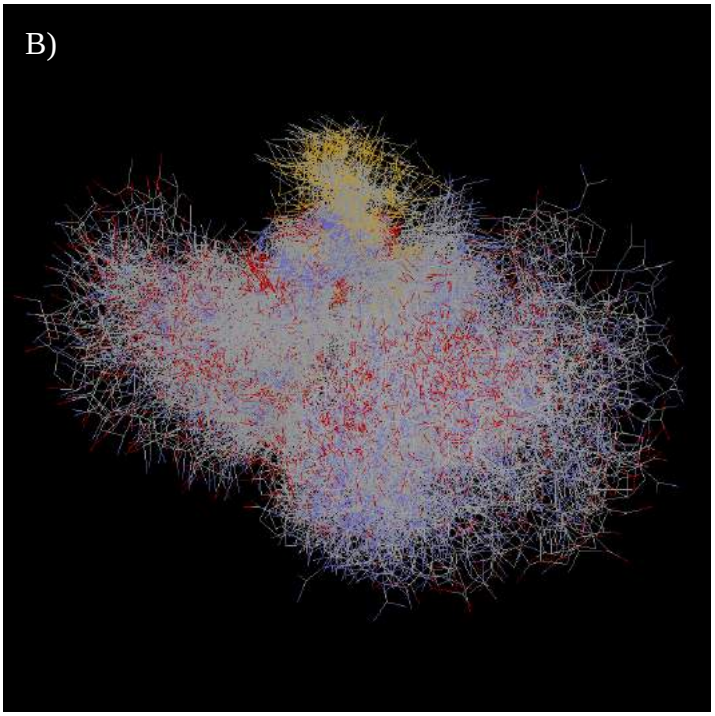
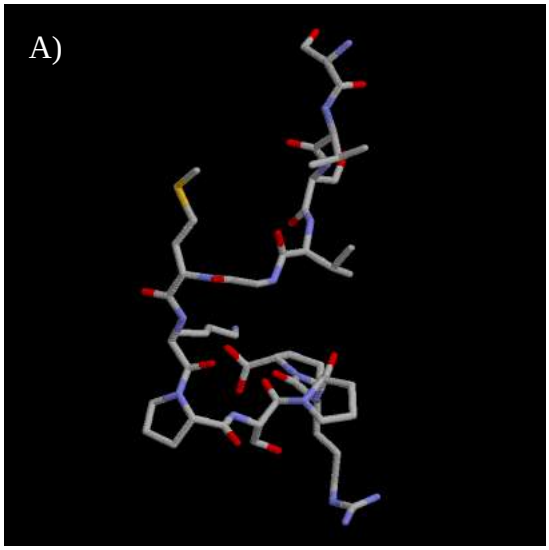


Figure13. dPCA with Ch1 for 99sb cluster2.
A)representative structure of cluster2
B)superposition of 500 structures of cluster2



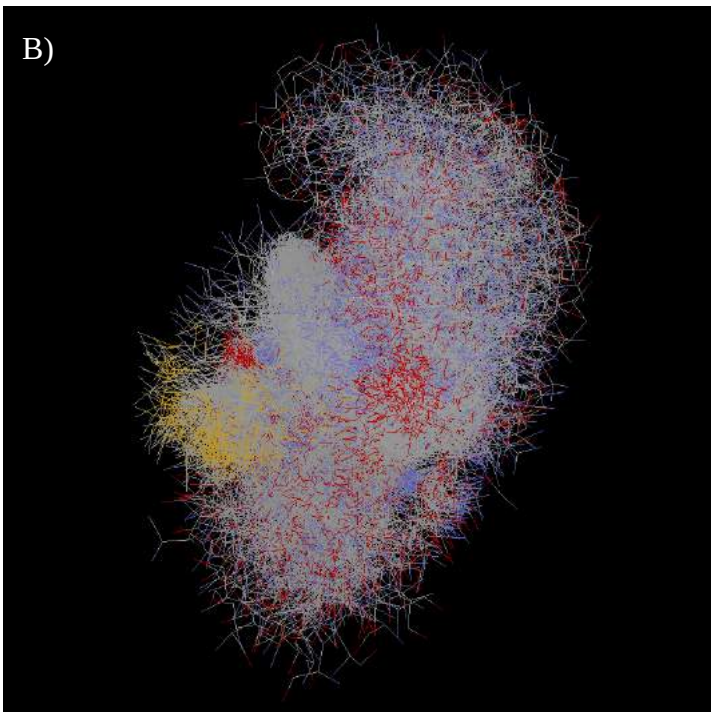


Figure14. dPCA with Ch1 for star cluster1.
A)representative structure of cluster1
B)superposition of 500 structures of cluster1

4.2. DOCKING ON DNA

After calculating the dihedral PCA of the two trajectories, the results were used as an input for the *HDOCK* program. For this step, only the structures without the side chains were used. The below shown docking corresponds to the clusters that were chosen during the previous step.

In order to visualize the ability of the peptide for docking, a part of B-DNA available in .pdb form and the *HDOCK* program were used. The representative structure of each cluster was chosen and *HDOCK* was configured to produce 10 output poses per representative structure.

The program *PyMOL* was used for the visualization of the docking, which is depicted on Figure15, Figure16, Figure17, Figure18 for *99sb* and Figure19, Figure20, Figure21, Figure22 for *star.*

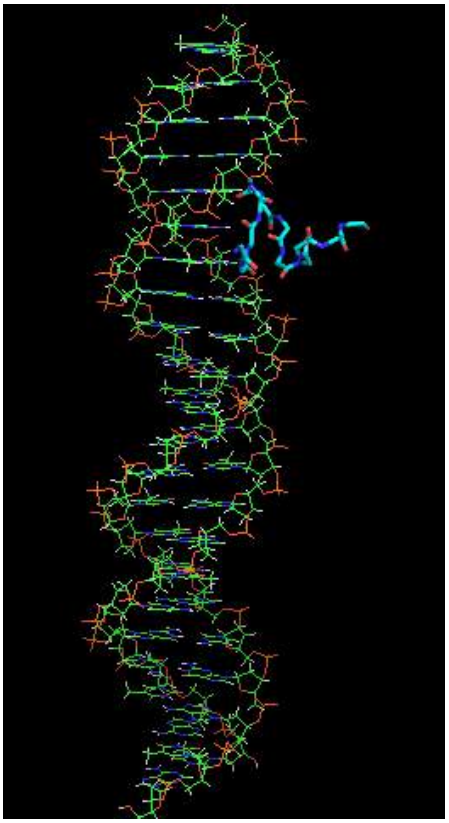


Figure15. Docking of the representative structure of cluster1 from the 99sb trajectory – model1.

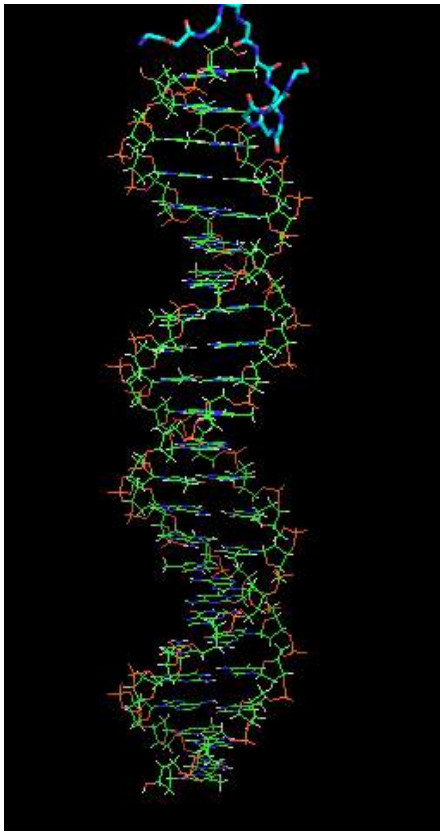


Figure16. Docking of the representative structure of cluster3 from the 99sb trajectory – model1.

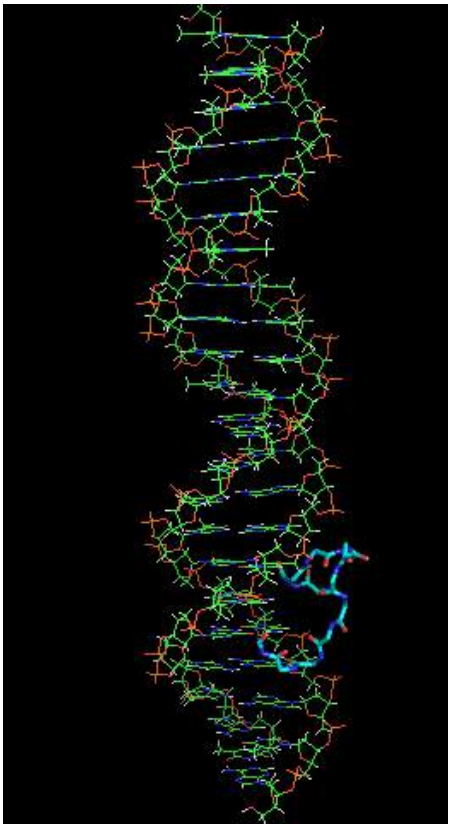


Figure17. Docking of the representative structure of cluster5 from the 99sb trajectory – model1.

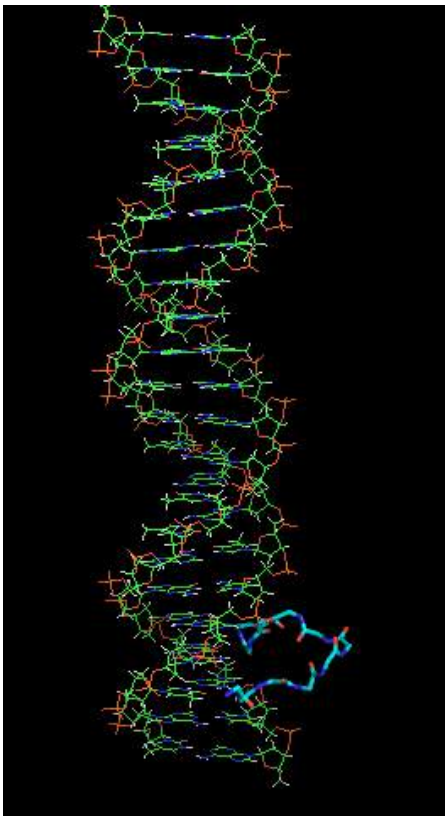


Figure18. Docking of the representatives tructure of cluster9 from the 99sb trajectory – model1.

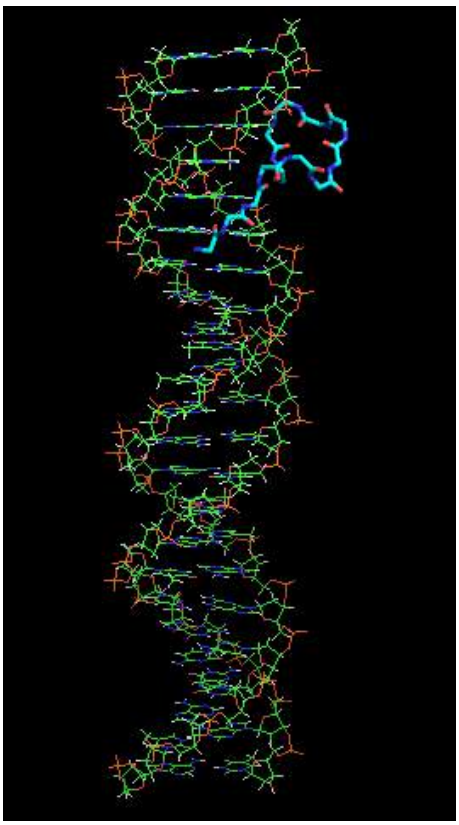


Figure19. Docking of the representative structure of cluster2 from the star trajectory – model1.

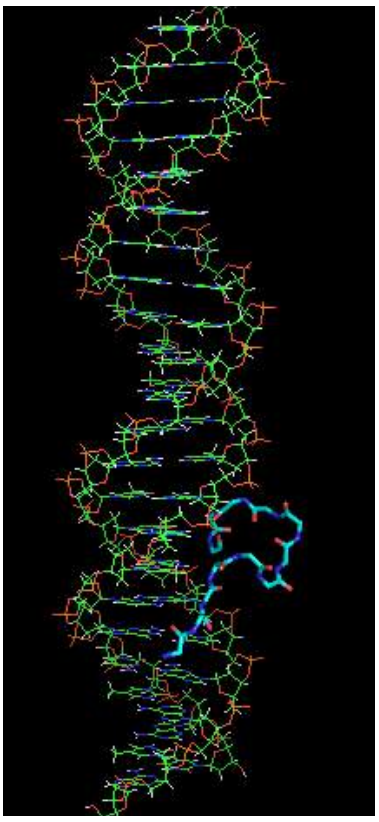


Figure20. Docking of the representative structure of cluster3 from the star trajectory – model1.

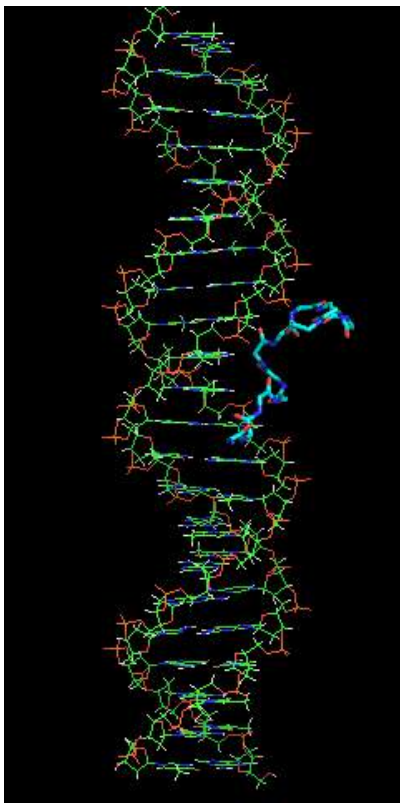


Figure21. Docking of the representative structure of cluster8 from the star trajectory – model1.

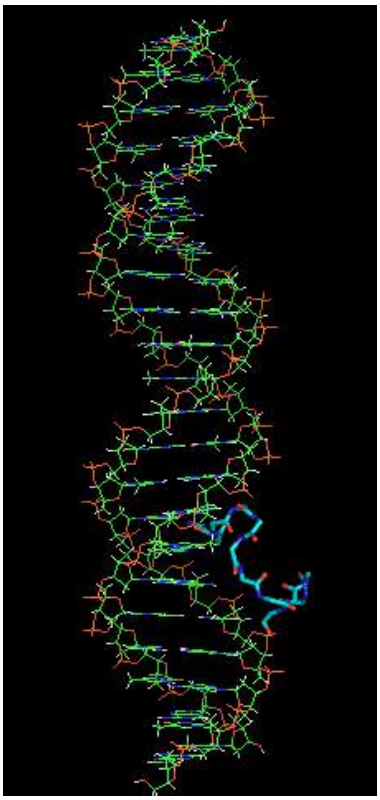


Figure22. Docking of the representative structure of cluster10 from the star trajectory – model1.

The peptide has the ability to dock onto the selected DNA, a property that alters in regard with the structure of the peptide itself. Even though all of the representative structures could bind with the DNA, not the same DNA position accounted for every structure. In fact, the binding positions could be on opposite directions.

As discussed previously, the “c” shape of the peptide enables the binding with the in between natural gaps of the DNA. This can be an interpretation of the multiple binding areas, as the gaps of the DNA repeat themselves throughout the structure.

5. SUMMARY AND CONCLUSIONS

The main objective of this paper was to present a hands-on experiment of bioinformatics, applied to a small peptide of multiple scientific interest. Such an analysis, was performed through automated steps for the complex mathematical calculations and the results were used as an input to broadly used programs in bioinformatics, in order to visualize them.

More specifically, the peptide used in this paper, SVSVGMKPSPRP, has been experimentaly proven to bind to DNA by Wolcke and Weinhold.

With the use of grcarma, the secondary structure of the peptide was calculated and residues 1, 11, 12 were excluded from the following simulations, as they were mobile and therefore, not interfering with the three-dimensional structure. All simulations were performed twice, once for each of the two initial trajectories that were used as an input.
Through the next step, the RMSD matrices were calculated and multiple clusters of structures with similar trajectories were visibly formed.

Among those, ten clusters for each initial trajectory were used and their dPCA was calculated, both with and without the side chains of the residues, which seemed to alter the results, by providing less structures when the side chains were added.

With the use of *RasMol* and the results from the previous step, the three-dimensional structure of the peptide was visualized in two ways, by presenting the representative structure of the cluster and by presenting a superposition of 500 structures of that cluster. A selection of those were included in the paper.

Finally, the ability of the peptide to bind onto DNA was assessed with the help of HDOCK for the calculations and *PyMOL* for the visualization of the docking. Again, a selection corresponding to the previous selection was presented in the paper.

The simulation suggests that the peptide tends to form a “c” shaped structure that probably enables the affinity of the peptide for the DNA.

Further analyses can provide information about the electrostatics of the peptide and their role in docking, which will help model a more thorough explanation for the mechanisms behind.

F. REFERENCES

1. K A Dill, H S Chan, Nat Struct Biol. **1997** 4(1) 10-9

2. S. Walter Englander, Leland Mayne, PNAS, **2014** 111 (45) 15873-15880

3. Martin Karplus, Folding and Design, **1997** 2 (1) S69-S75

4. A R Dinner, A Sali, L J Smith, C M Dobson, M Karplus, Trends Biochem Sci, **2000** 25(7) 331-9

5. Scott A. Hollingsworth, Ron O. Dror, Neuron. **2018** 99(6) 1129–1143

6. Robert W. Newberry, Ronald T. Raines, ACS Chem Biol. **2019** 14(8) 1677–1686

7. Yiwen Chen, Feng Ding, Huifen Nie et al, Arch Biochem Biophys. **2008** 469(1) 4–19

8. Roger Armen, Darwin O.V. Alonso, Valerie Daggett, Protein Sci. **2003** 12(6) 1145–1157.

9. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell, 4th edition. **2002** 123-143

10. David L. Nelson, Michael M. Cox, Lehninger Principles of Biochemistry, 7th Edition, **2017** 345-377

11.K A Dill, Justin L MacCallum, Science, **2012** 338(6110) 1042-6

12. J Wölcke, E Weinhold, Nucleosides Nucleotides Nucleic Acids, **2001** 20(4-7) 1239-41

13. Elias Estephan, Jérôme Dao, Marie-Belle Saab, Biomed Tech (Berl), **2012** 57(6):481-9

14. L I Smith, **2002,** A tutorial on Principal Components Analysis

15. A I Jewett, Conrad C Huang, Thomas E Ferrin, Bioinformatics, **2003** 19(5) 625–634

16. P. I. Koukos, N.M. Glykos. J. Comput. Chem. **2013**, DOI: 10.1002/jcc.23381

17. Herbert J. Bernstein, **2009**, *RasMol*

18.Yumeng Yan, Di Zhang, Pei Zhou, Botong Li, Sheng-You Huang, Nucleic Acids Res. **2017** 45(Web Server issue): W365–W373.

19. Schrödinger, L. & DeLano, W., **2020**. *PyMOL*

20. James C. Phillips, David J. Hardy, Julio D. C. Maia, John E. Stone, Joao V. Ribeiro, Rafael C. Bernardi, Ronak Buch, Giacomo Fiorin, Jerome Henin, Wei Jiang, Ryan McGreevy, Marcelo C. R. Melo, Brian K. Radak, Robert D. Skeel, Abhishek Singharoy, Yi Wang, Benoit Roux, Aleksei Aksimentiev, Zaida Luthey-Schulten, Laxmikant V. Kale, Klaus Schulten, Christophe Chipot, and Emad Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD, *Journal of Chemical Physics*, 153:044130, **2020**. DOI:10.1063/5.0014475