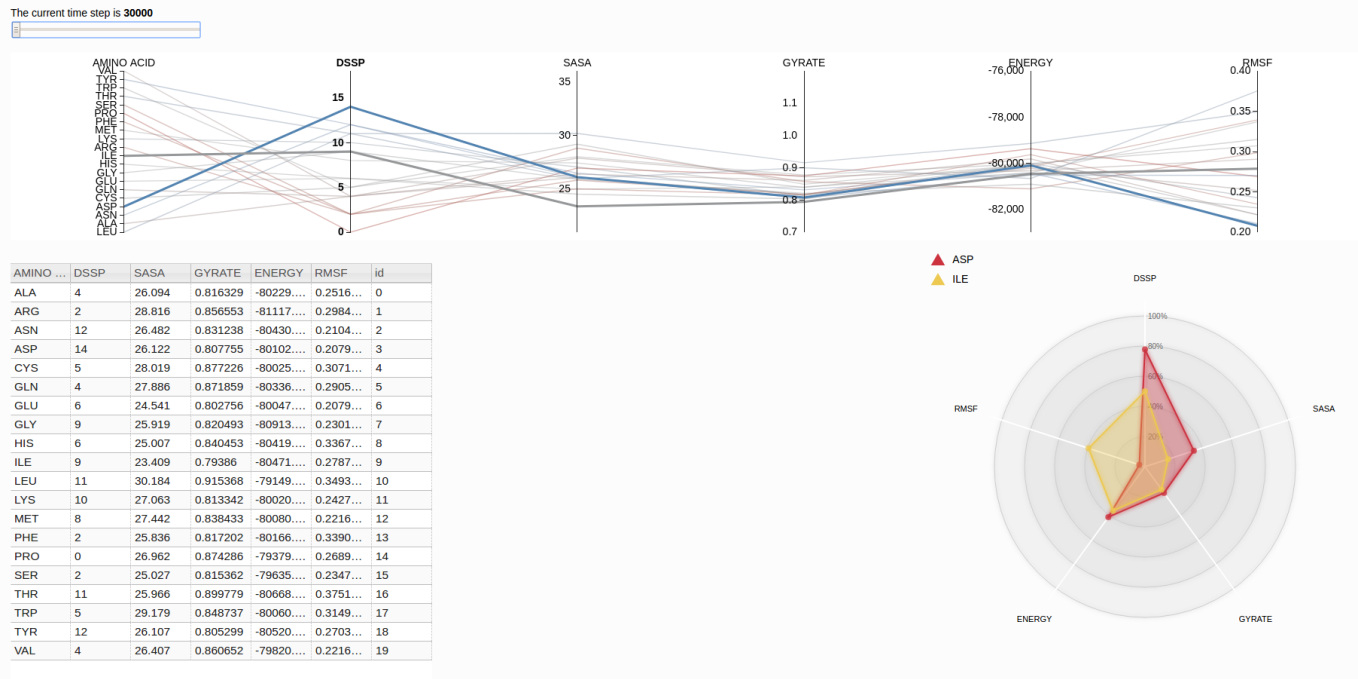
Visualization tool for comparison of a single amino acid change in protein simulation

Bruno Iochins Grisci
Instituto de Informatica
UFRGS
Porto Alegre, Brazil
Email: bigrisci@inf.ufrgs.br



AMINO ...	DSSP	SASA	GYRATE	ENERGY	RMSF	id
ALA	4	26.094	0.816329	-80229....	0.2516...	0
ARG	2	28.816	0.856553	-81117....	0.2984...	1
ASN	12	26.482	0.831238	-80430....	0.2104...	2
ASP	14	26.122	0.807755	-80102....	0.2079...	3
CYS	5	28.019	0.877226	-80025....	0.3071...	4
GLN	4	27.886	0.871859	-80336....	0.2905...	5
GLU	6	24.541	0.802756	-80047....	0.2079...	6
GLY	9	25.919	0.820493	-80913....	0.2301...	7
HIS	6	25.007	0.840453	-80419....	0.3367...	8
ILE	9	23.409	0.79386	-80471....	0.2787...	9
LEU	11	30.184	0.915368	-79149....	0.3493...	10
LYS	10	27.063	0.813342	-80020....	0.2427...	11
MET	8	27.442	0.838433	-80080....	0.2216...	12
PHE	2	25.836	0.817202	-80166....	0.3390...	13
PRO	0	26.962	0.874286	-79379....	0.2689...	14
SER	2	25.027	0.815362	-79635....	0.2347...	15
THR	11	25.966	0.899779	-80668....	0.3751...	16
TRP	5	29.179	0.848737	-80060....	0.3149...	17
TYR	12	26.107	0.805299	-80520....	0.2703...	18
VAL	4	26.407	0.860652	-79820....	0.2216...	19

Fig. 1. Image of the tool with all elements presented.

***Abstract*—**

***Keywords*-one or two words; separated by semicolon; from specific; to generic fields;**

I. Introduction

Proteins are polymers formed by a sequence of 20 different possible amino acids that under physiological conditions fold into a precise shape known as its native state [1]. Each amino acid has an alpha carbon (CA) with bonds to amino (NH2) and carboxyl (COOH) groups and a variable side-chain (R) that determines the particular physicochemical properties of each residue. A peptide is a molecule composed of two or more amino acids chained by a chemical bond called the *peptide bond*. This bond is formed when the carboxyl group of one residue reacts with the amino group of the other residue, releasing a water molecule. The interaction between amino acids in a protein causes the polypeptide chain to fold, usually in a proper configuration, as α-helix, β-sheet, coil or turn. These local folding patterns represent the secondary structure of a protein. The topology or fold is given by the succession of secondary structures connected in a 3D space. The specific characteristics of the peptide bond have significant implications for the 3D fold that can be adopted by polypeptides. The peptide bond (C-N) has a double bond, and it is not allowed rotation of the molecule around this bond. The rotation is only permitted around the bonds N-Cα and C$_\alpha$-C. The analysis of experimental protein structures (X-ray data) reveals that amino acid residues can assume many conformations in proteins [2]. Each amino acid has a set of physiochemical properties which contributes to its intrinsic conformational preference [3]. Amino acids in a secondary structure usually adopt a particular set of backbone torsion angles [2].

The amino acids sequence of a protein is directly related to its structure in three dimensional space, which is defining of the protein biological function. The change of only one

amino acid in the chain is capable of great modification of the structure and function of a protein because of the differences in size and physical-chemical properties among amino acids [4].

A. Related work

II. Data characterization

The data for testing this visualization tool comes from the Kappa-conotoxin PVIIA, with amino acids sequence: $CRIPNQKCFQHLDDCCSRKCNRFNKCV$. Kappa-conotoxins are neurotoxical proteins extracted from sea slugs poison. They are inhibitors of the potassium channel and when injected in different organisms some alterations in their effect can be observed, from hyperactivity in fish to death when combined with other variants of conotoxins. The substitution of the asparagine (N) in the fifth position in the amino acids chain of the Kappa-conotoxin PVIIA by an alanine (A) is know to cause a reduction of 100% in the toxicity of this protein [5] [6] [7].

For this study, this asparagine was changed by all possible 20 amino acids, and for each resulting structure a molecular dynamics simulation was performed in order to evaluate its behavior during a predetermined period of time under chosen conditions. The simulations were performed with GROMACS [8], a software for biomolecules simulation, with the simulated time of 50ns. These simulations mimic the behavior of biomolecules in a solvent under controled temperature. From the data created versus time, the following attributes were studied.

A. Data dimensions

DSSP: Analysis of the secondary structure of the protein for each frame of the simulation. The secondary structure is a local structural conformation of a region in the amino acids sequence which follows specific patterns that, once folded, will originate the three dimensional functional structure of the protein.

SASA: Solvent-accessible surface area is the surface area of the protein that is accessible to the solvent. By analysing the area exposed to the solvent, i.e., what is around the protein, it is possible to infer how denatured is the protein, so the greater the SASA value, the greater the denaturation.

GYRATE: The radius of gyration can be used as a measurement of the compactness of a protein structure. It describes the overall spread of the molecule and is calculated taking the root mean square distance of the atoms from their common centre of gravity. Greater values of radius of gyration means a less compact protein.

ENERGY: This attribute measures the total energy of the system. The energy is a stability indicator, usually comparable, in which systems with less energy will be more stable (less entropy).

RMSF: The root mean square fluctuation of atomic positions is the measure of the average distance between the atoms of the simulated protein and a well-defined average position.

AMINO …	DSSP	SASA	GYRATE	ENERGY	RMSF	id
ALA	4	26.094	0.816329	-80229.…	0.2516…	0
ARG	2	28.816	0.856553	-81117.…	0.2984…	1
ASN	12	26.482	0.831238	-80430.…	0.2104…	2
ASP	14	26.122	0.807755	-80102.…	0.2079…	3
CYS	5	28.019	0.877226	-80025.…	0.3071…	4
GLN	4	27.886	0.871859	-80336.…	0.2905…	5
GLU	6	24.541	0.802756	-80047.…	0.2079…	6
GLY	9	25.919	0.820493	-80913.…	0.2301…	7
HIS	6	25.007	0.840453	-80419.…	0.3367…	8
ILE	9	23.409	0.79386	-80471.…	0.2787…	9
LEU	11	30.184	0.915368	-79149.…	0.3493…	10
LYS	10	27.063	0.813342	-80020.…	0.2427…	11
MET	8	27.442	0.838433	-80080.…	0.2216…	12
PHE	2	25.836	0.817202	-80166.…	0.3390…	13
PRO	0	26.962	0.874286	-79379.…	0.2689…	14
SER	2	25.027	0.815362	-79635.…	0.2347…	15
THR	11	25.966	0.899779	-80668.…	0.3751…	16
TRP	5	29.179	0.848737	-80060.…	0.3149…	17
TYR	12	26.107	0.805299	-80520.…	0.2703…	18
VAL	4	26.407	0.860652	-79820.…	0.2216…	19

Fig. 2. Overview of the used grid.

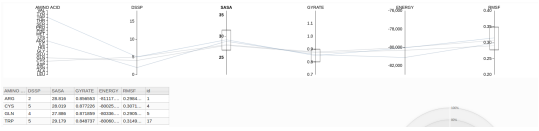


Fig. 3. Example of the use of brush.

All the dimensions are continuous numerical values. As can be seen, the data of a single simulation is multidimensional and changes over the simulated time.

III. Technical background

IV. Technique overview

A. Grid

Item highlighting:
Item selection:

B. Parallel Coordinates

Axis brushing: Fig. 3
Axis sorting:
Axis reordering:
Recoloring by axis selection:

C. Radar Plot

Item highlighting:

D. Slider

V. Experiments

VI. Results and Discussion

A. Future work

VII. Conclusion

A demo can be tested at: http://inf.ufrgs.br/ bigrisci/parallel-coordinates/

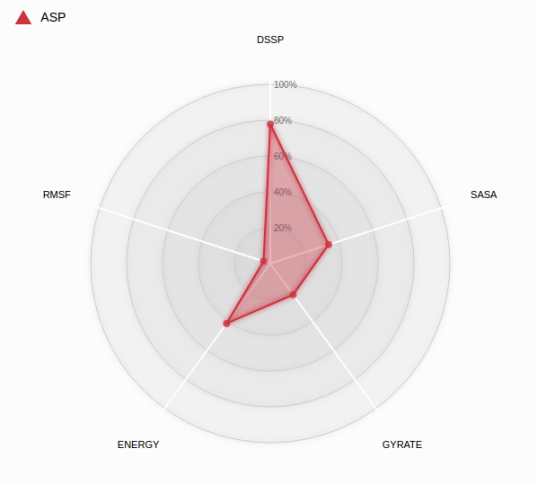


Fig. 4. Example of one amino acid being visualized in the radar plot.

ACKNOWLEDGMENT

REFERENCES

[1] C. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 96, pp. 223–230, 1973.

[2] B. Borguesan, M. Barbachan e Silva, B. Grisci, M. Inostroza-Ponta, and M. Dorn, "Apl: an angle probability list to improve knowledge-based metaheuristics for the three-dimensional protein structure prediction," *Comput. Biol. Chem.*, vol. 59, no. A, pp. 142–157, 2015.

[3] V. Mathura and D. Kolippakkam, "Apdbase: Amino acid physicochemical properties database," *Bioinformation*, vol. 1, no. 1, pp. 2–4, 2005.

[4] D. L. Nelson, A. L. Lehninger, and M. M. Cox, *Lehninger principles of biochemistry*. Macmillan, 2008.

[5] R. B. Jacobsen, E. D. Koch, B. Lange-Malecki, M. Stocker, J. Verhey, R. M. Van Wagoner, A. Vyazovkina, B. M. Olivera, and H. Terlau, "Single amino acid substitutions in κ-conotoxin pviia disrupt interaction with the shaker k+ channel," *Journal of Biological Chemistry*, vol. 275, no. 32, pp. 24 639–24 644, 2000.

[6] R. Mir, S. Karim, M. Amjad Kamal, C. M Wilson, and Z. Mirza, "Conotoxins: structure, therapeutic potential and pharmacological applications," *Current pharmaceutical design*, vol. 22, no. 5, pp. 582–589, 2016.

[7] D. T. Akey, X. Zhu, M. Dyer, A. Li, A. Sorensen, S. Blackshaw, T. Fukuda-Kamitani, S. P. Daiger, C. M. Craft, T. Kamitani *et al.*, "The inherited blindness associated protein aipl1 interacts with the cell cycle regulator protein nub1," *Human molecular genetics*, vol. 11, no. 22, pp. 2723–2733, 2002.

[8] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, "Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of chemical theory and computation*, vol. 4, no. 3, pp. 435–447, 2008.