Introduction to Bioinformatics

Ernane Rosa Martins
(Organizador)

Atena Editora
2019

Ernane Rosa Martins
(Organizador)

Introduction to Bioinformatics

Atena Editora
2019

Editora Chefe: Profª Drª Antonella Carvalho de Oliveira
Diagramação e Edição de Arte: Geraldo Alves e Natália Sandrini
Revisão: Os autores

Dados Internacionais de Catalogação na Publicação (CIP)
(eDOC BRASIL, Belo Horizonte/MG)

I61 Introduction to bioinformatics [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa (PR): Atena Editora, 2019.

Formato: PDF
Requisitos de sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Inclui bibliografia.
ISBN 978-85-7247-113-8
DOI 10.22533/at.ed.138191202

1. Bioinformática. 2. Inteligência artificial. I. Martins, Ernane Rosa.

CDD 570.285

Elaborado por Maurício Amormino Júnior – CRB6/2422

APRESENTAÇÃO

A bioinformática é um campo interdisciplinar, que busca analisar, interpretar e processar dados biológicos, com foco na aplicação de técnicas computacionais intensivas, tais como: métodos computacionais, teoria de grafos, inteligência artificial, algoritmos matemáticos, reconhecimento de padrões, mineração de dados, algoritmos de aprendizado de máquina, processamento de imagens e simulação computacional. Como um campo interdisciplinar, a bioinformática combina diversas áreas do conhecimento, como: engenharia, matemática, física, química, estatística, ciência da computação e biologia, entre outras.

A coletânea "*Introduction to bioinformatics*" é um livro composto por 6 capítulos que abordam assuntos atuais, tais como: o adenocarcinoma gástrico que é uma malignidade com elevada incidência e mortalidade no mundo; o vírus zika (VZIK) que é um Arbovirus que pertence à família Flaviviridae; As H + -ATPases que são proteínas integrais da membrana plasmática que têm a capacidade de utilizar a energia química da hidrólise de ATP para expulsar os prótons para o ambiente extracelular, atuando na manutenção da homeostase iônica e transporte de solutos; o vírus da família Geminiviridae que tem sido intensamente estudado devido à gravidade das doenças causadas em várias culturas importantes como: feijão, algodão, milho, tomate e mandioca.

Espero que os capítulos deste livro possam contribuir efetivamente na disseminação dos conhecimentos relevantes da bioinformática, proporcionando uma visão ampla sobre este campo de conhecimento.

Assim, desejo a todos uma excelente leitura.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 7

IN-SILICO DETOXIFICATION EVIDENCE OF THE HERBICIDE BISPYRIBAC SODIUM BY A TEORETHICAL MODEL OF GLUTATHIONE S-TRANSFERASE TAU 5 FROM *Oryza sativa* L.

Vinícius Costa Amador
Universidade Federal Rural de Pernambuco, Departamento de Agronomia, Recife-PE.

Ravenna Lins Rodrigues
Universidade Federal de Campina Grande, Centro de Desenvolvimento Sustentável do Semiárido. Sumé-PB.

Felipe de Oliveira França
Universidade Federal de Campina Grande, Centro de Desenvolvimento Sustentável do Semiárido. Sumé-PB.

Rafael Trindade Maia
Universidade Federal de Campina Grande, Centro de Desenvolvimento Sustentável do Semiárido. Sumé-PB.

ABSTRACT: Rice (*Oryza sativa* L.) is one of the most important crops in the world, and it has being the primary source of nutritional layout in developing countries in Asian. Despite this remarkable importance, there are few studies about the development of techniques that minimizes the potential problems inherent of this grain cultivation, such as competition with weeds, making necessary the use of herbicides. The glutathione S-transferases (GSTs) superfamily confers to rice protection against biotic and abiotic stress, and herbicide resistance. However, the three-dimensional structure of a GST Tau class, is unsolved. The objectives of this work were to develop a reliable comparative model for the s-transferase glutathione class Tau 5 from rice, and simulate docking interactions, against herbicides bentazon and metsulfuron. Results showed that the predicted model is reliable and has structural quality. Ramachandran plot set 95,4% of the residues in the most favored regions. All complexes showed negative binding energies values; and bispyribac sodium docked to the glutathione tripeptide, and it represents an in-silico evidence of glutathione conjugation with this herbicide.

KEYWORDS: Molecular Modeling, Docking, OsGST, Herbicide.

1 I INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most importante crops in world, being 90% of the mundial rice cultivation arising from Asia, supplying 60% of the rice mundial demand (DOGARA; JUMARE, 2014). This means an importance against it impact on the agrobusiness, since its nutritional value consists in the primary source of nutritional layout in developing countries in Asian (DOGARA; JUMARE, 2014). According to FAO, and the International Institute of Rice Research (IRRI), rice needs more production to supply the alimentation demand,

since for the year 2050 it is estimated that the demand increases at least 60% in relation to 2016 production level (ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA ALIMENTACIÓN Y LA AGRICULTURA, 2016). However, there is a few studies about development of techniques with the potential to minimize the problems inherent to the cultivation of this cereal, and competition with weeds, (SANTOS, 2006), without despising the damage caused by many herbicides.

Among detoxification enzymes, the most known and studied are glutathione transferases, which conjugates these xenobiotics and turn them into a water soluble complex (WILCE e PARKER, 1994; KREUZ, TOMMASINI, MARINOIA, 1992). This enzimes confers to rice catalitic action, and protection agains biotic and abiotic stress including herbicides resistence/tolerance (FROVA, 2003, 2006), and its main reaction consists in the conjugation of the tripeptide glutathione to a hydrophobic compound, making it more soluble and less (MARRS, 1996), maintaining cellular homeostasis. Tau class GSTs are extremely important as they are involved in the metabolization to a broad spectrum of important commercial herbicides (EDWARDS, DIXON, WALBOT, 2000). A study made by (Lajmanovich et al., 2013) suggest that Bispyribac Sodium has a correlation with Glutathion S-trasnferase expression. The aim of this study was to construct a theoretical model for a tau 5 *Oryza sativa* glutathione S-transferase (*OsGSTU5*) and perform docking simulations against Bispyribac Sodium herbicide.

2 I MATERIAL AND METHODS

2.1 Data mining

At first, the primare *OsGSTU5* (*Oryza sativa* L. glutathiona S-transferase tau 5) sequence were obtained in *.fasta* extension in NCBI data bank, a good quality structure model were searched for template using the “BLASTp” (Basic Local Alignment Search Tool for proteins - http://blast.ncbi.nlm.nih.gov/Blast.cgi) in *.pdb* extension from PDB (Protein Data Bank*, http://www.rcsb.org/)* data bank.

2.2 Homology modeling and model validation

The *OsGSTU5* primary sequence was obtained from NCBI database (https://www.ncbi.nlm.nih.gov/). The SWISS-MODEL server Automated Mode tool (http://swissmodel.expasy.org/) was used to generate the model (ARNOLD et al., 2006; KIEFER et al., 2009) The model validation process used the Ramachandran graph analysis through the PROCHECK program (LASKOWSKI et al., 1993) to verify the stereochemical quality of the structure. Local quality was accessed by ANOLEA (MELO, 1998) and GROMOS force fields (van GUNSTEREN et al., 1998). All the generated docked complexes were visualized with Visual Molecular Dynamics software (SURHONE et al., 2010).

2.3 Protein pockets identification, docking simulation and anchor residues

identification

The protein pockets identification where using the ghecom 1.0 finder server used to find multi-scale pockets on protein surfaces using mathematical morphology (KAWABATA 2007, 2010).

The as ligands structure (Table 1.0) were obtained from the ZINC database (http://zinc.docking.org/) in *.mol2* extension files. These files were converted to *.pdbqt* in Autodock 4.2.1 (MORRIS et al., 2009) (https://www.chpc.utah.edu/documentation/software/autodock.php), polar hydrogens were removed and their molecules were assigned with the Gasteiger parameters (GASTAIGER, 1980).

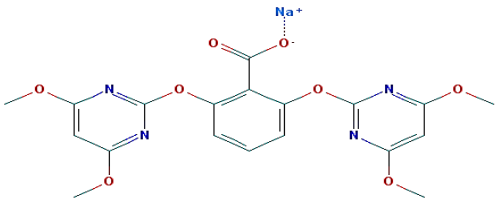
Name	Zinc ID	Class	HRAC	2D
Bispyribac sodium	ZINC 04098944	Ácido Pirimidiniloxibenzóico	B	

Table 1. Herbicide used in docking simulation

The *OsGSTU5* theoretical model, was converted to *.pdbqt* file in Autodock, hydrogens and Kollman parameters were added (WEINER et al., 1984). The GTX was treated as a cofactor. Docking simulations were run on the Autodock 1.5.6 program (MORRIS et al., 2009) and the Lamarckian genetic algorithm (LGA) whas choosen. The simulations had the following parameters: 10,000 replicates, energy analyzes per 1,500,000 and 27,000 generations, population size of 150 and mutation rates and crossing-over of 0.02 and 0.08 respectively. The 10 conformations were generated that were ranked based on the lowest energy and important residues interaction were analyzed in the VMD (SURHONE et al., 2010) (http://www.ks.uiuc.edu/Research/vmd/).

3 I RESULTS AND DISCUSSION

3.1 Data mining

The template used was a *crystal structure of a glutathione S-transferase PtGSTU30* from *Populus Trichocarpa,* in complex with GSH (PDB ID: 5J4U) presented 1.249 Å resolution (THOM et al., 2002), and high 55.09% identity value, revealing a homology between the *OsGSTU5* and *PtGSTU30* proteins appropriate for the modeling what was perceived by the results with other Tau class proteins used in researches (KILILI et al., 2004).

Target Protein		Template data				
Protein	ID	Protein	(PDB-ID)	Metod	Identy	Resolution Å
OsGSTU5	AAG32470.1	PtGSTU30	5j4u	XRD	55.09	1.249

Table 2. Target and template proteins data description

3.2 Sequences alignment and conserved regions identification

The Figure 1 shows conserved and semiconserved regions between *OsGSTU5* and PtGSTU30 (template) sequences (labeled), and some important anchor residue in the catalitic pocket HIS 51 (green arrow) other important residues like LYS53, LYS 111 and LYS 112 (red arrows) were not conserved residues.

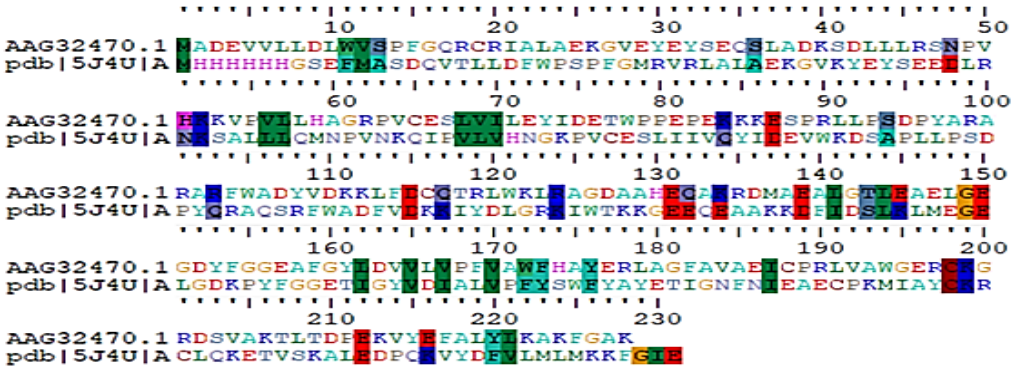


Figure 1. Sequences alignment, identification of conserved and semiconserved regions (labeled) and achor residues (arrows)

3.3 Theoretical model construction, quality evaluation and validation

According to Laskowski (1993, 2012), a reliable predicted model are supposed to display over 90% of residues in core regions of Ramachandran plot (A, B and L), for Ho (2005) some residues as glycine and proline has predictable and distinct distribution on the Ramachandran plot, as they present different stereochemical patterns. The stereochemical quality was accessed considering the Laskowski (2012) critters, the Ramachandran plot showed 95.4% of the residues (black squares and triangles) in regions that were more favorable (red), 4.1% in allowed regions (yellow), 0.5% in generously allowed regions (cream) and 0.0% in regions (white) as shown in Figure 2, a model validation results by Maia and Nadvorny (2014) had 100.0% of the residues in allowed regions.

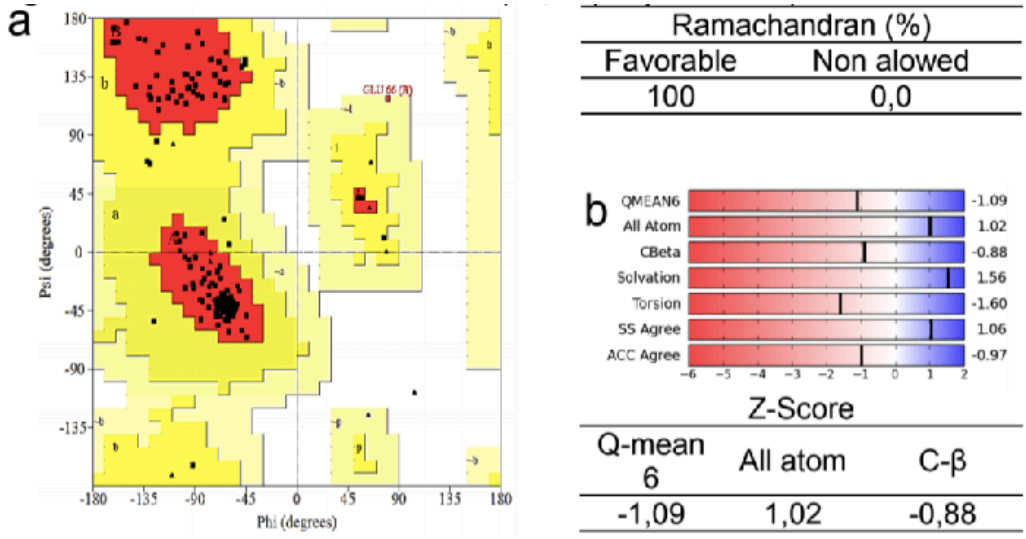


Figure 2. Model validation a. Ramachandran plot, b. quality mean for 6 paramethers.

Quality mean for 6 parameters Z-score (Qmean6) was between 1 and 2 and considered appropriated, meaning that the evaluation of the theoretical native structure protein in comparison with experimental models of the similar size (residues) in databanks had it stereochemical and atom parameters average considered good, showed by the red arrow/star situed in Figure 3.

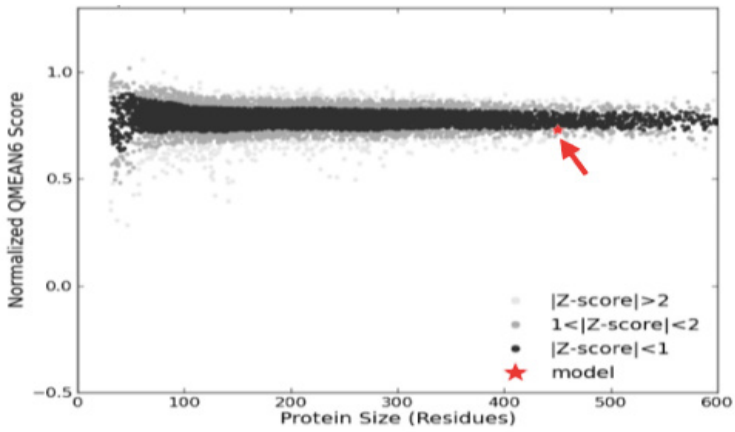


Figure 3. Theoretical model quality comparison among experimental models quality.

The results from ANOLEA (MELO, 1998), and for GROMOS (van GUNSTEREN; BERENDSEN, 1990), were generally negative values, revealing a model with stable energy, and negative values for majority of the residues, corroborating with the results showed by Hamid (2013), Maia and Nadvorny (2014). The generated and validated theoretical model had the atomic coordinate considerated satisfactory for it native structure and appropriated for Docking test.

3.4 Docking and anchor residues and protein pockets identification

The protein pockets were predicted (Figure 4. and Figure 4. A) by a multi-scale, mathematical morphology algorithm (KAWABATA 2007, 2010). The docking anchor

residues were compared throug the Ghecom results, and was perceived that they belonged to the best pocket (in red), the catalytic cleft (Figure 4.B), near the detoxifying subunit.

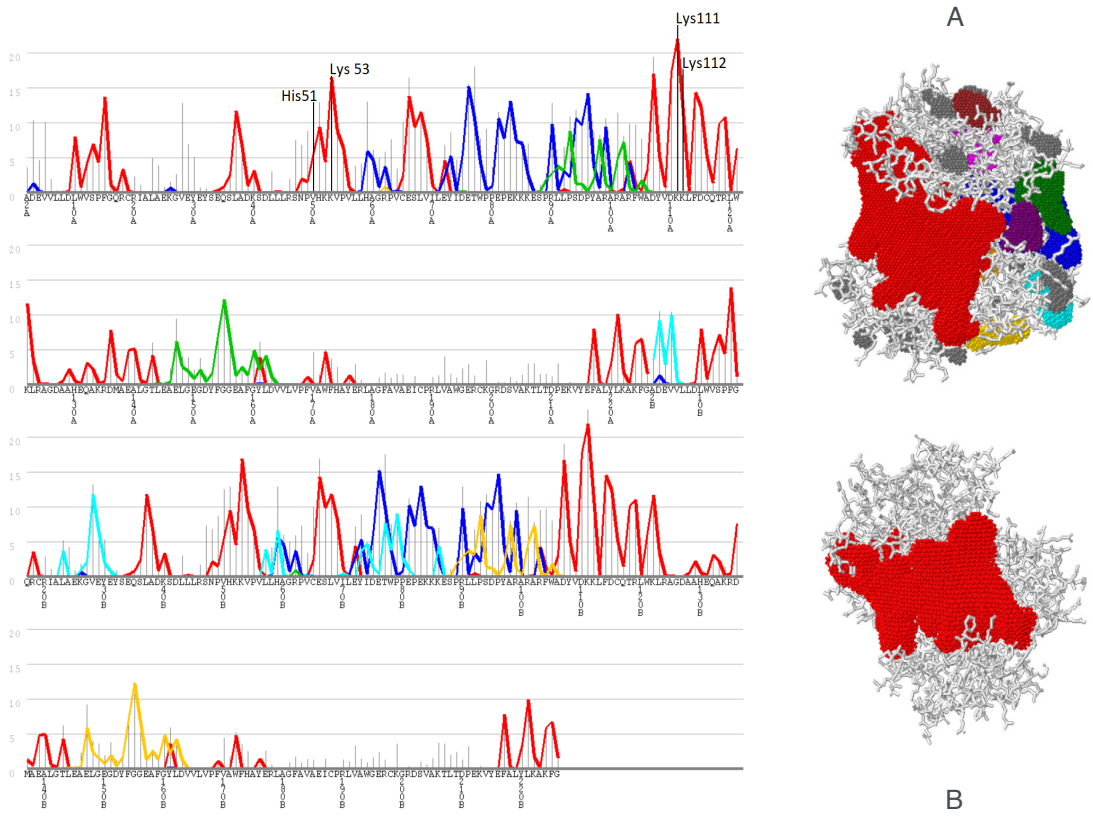


Figure 4. Pockets and aminoacids sequence description, and anchor resides described
Figure4A All ranked pokets, Figure4 B Best pocket described (red)

The docking results for the *Bispyribac Sodium* "2,6-bis[(4,6-dimethoxypyrimidin-2-yl)oxy]benzoic acid", processed by the autodock program ranked ten possible complexes based on intermolecular energy scores, binding energy, and H-bonds that reveals atom (and residues) of the docked region binding of the ligand and the protein (MORRIS *et al*. 2009) that present favorable interation information of the model described in Table 3.

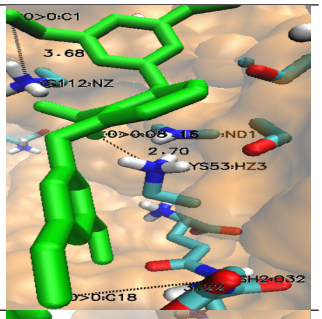
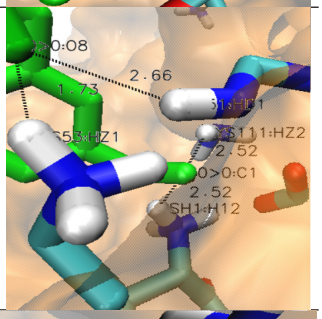
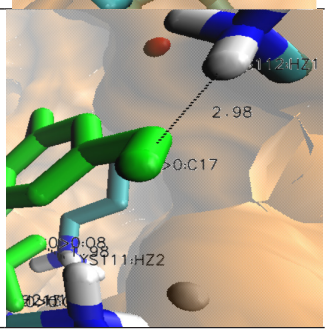
Bispyribac Sodium Herbicide “2,6-bis[(4,6-dimethoxypyrimidin-2-yl)oxy]benzoic acid”					
Protein	Ranking	Sistem Energy	Inhibition Constant (Ki)	Binding Atoms Ligand_Protein residue _distance Å	Image
U5	1	-4,71	354,22	<0>O:C1_ LYS112:NZ1_3.68 <0>O:O8_ LYS53:HZ3_2,70 <0>O:O8_ HIS51:ND1_3,16	
	2	-4,01	1,16	<0>O:O8_ HIS51:HD1_2,66 <0>O:O8_ LYS53:HZ1_1,73 **<0>O:C1_ GSH1:H12_2,52** <0>O:C1_ LYS111:HZ2_2,52	
	3	-3,62	2,21	<0>0:C17_ LYS112:HZ1_2,98 <0>0:O8_ LYS111:HZ2_1,98 **<0>0:O3_ GSH2:H11_2.13**	

Table 3. Best three dockings ranked by the autodock, with descriptions of the interaction.

For clarification the non GSH ligand atoms distribution on the protein topology is described in the Figure 5, and can be perceived theyr belonging to the catalitic cleft of the protein (in White).

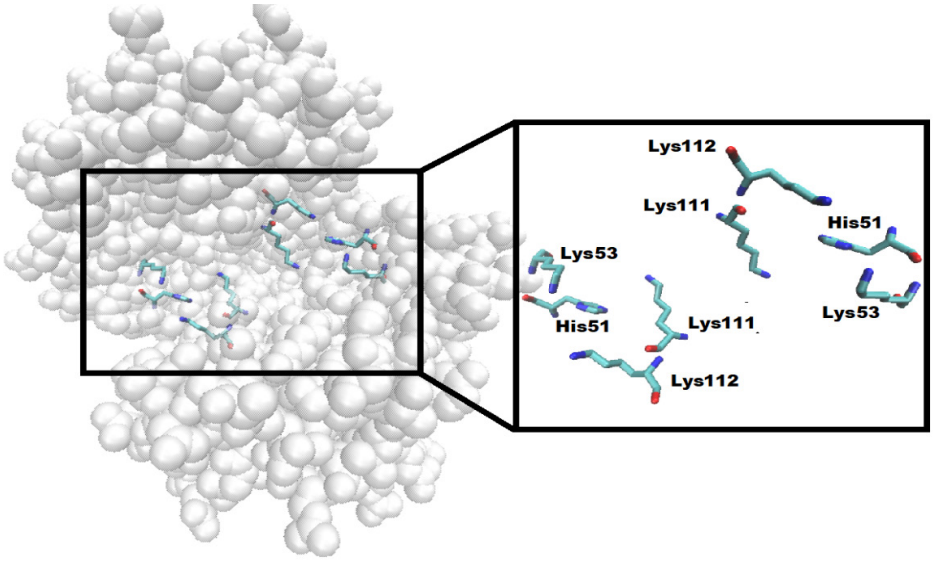


Figure 5. Description of anchor residues location.

The docking results for herbicide Bispyribac Sodium and *OsGSTU5* are displayed on the Table 3. and suggest that Lysins (LYS) situed on the catalitic pocket are importante residues for the interaction, representing anchor residues. The anchors residues seems to be three lysines at positions 53, 111 and 112 and a histidine at the 51 position, located on the catalytic H-site of the enzyme (Figure 5) and a GSH bound listed in Table 3, the results of the docking still reveals that the herbicide can dock to U5 protein, in the tripeptide glutathione (GSH) subunit showed in Figure 3, suggesting a possible detoxifying process (KILILI, 2004). It is probably that *OsGSTU5* metabolizes bispyribac sodium herbicide through GSH-conjugation (YAMAMOTO et al., 2012). A work done by Lajmanovich and Junges (2013) and another done by Lu (2013) showed a correlattion between Bispyribac Sodium and the superfamily of GST expression, its possible that *OsGSTU5* is one of the superfamily members that suffer the same expression effect indicating the correlation between the herbicide metsulfuron and *OsGSTU5*.

4 I CONCLUSIONS

The results lead to conclude that the theoretical model developed (OsGSTU5), presentes quality and is a representative native model suitable to the docking test. Our insights could be used appoint possible molecular markers, for futher marker-assisted selection tolerance/resistance to herbicide in plant studies.

ACKNOWLEDGMENT

BCT (Biologia Computacional e Teórica) CDSA (Centro de Desenvolvimento Sustentável do Semiárido CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) NemPe (Núcleo de estudos em melhoramento genético de plantas de

Pernambuco).

REFERENCES

ARNOLD, Konstantin, et al. **The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling**. Bioinformatics, v. 22, n. 2, p. 195-201, 2006.

SANTOS, HG dos et al. **Cultivo do arroz de terras altas no estado de Mato Grosso**. Embrapa Arroz e feijão, 2006.

DOGARA, Abdulrahman Mahmoud; JUMARE, Aisha Ishaq. **Origin, distribution and heading date in cultivated rice**. Int J Plant Biol Res, v. 2, n. 1, p. 1008, 2014.

EDWARDS, Robert; DIXON, David P.; WALBOT, Virginia. **Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health**. Trends in plant science, v. 5, n. 5, p. 193-198, 2000.

FROVA, Carla. **Glutathione transferases in the genomics era: new insights and perspectives.** Biomolecular engineering, v. 23, n. 4, p. 149-169, 2006.

FROVA, Carla. **The plant glutathione transferase gene family: genomic structure, functions, expression and evolution.** Physiologia plantarum, v. 119, n. 4, p. 469-479, 2003.

GASTEIGER, Johann; MARSILI, Mario. **Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges.** Tetrahedron, v. 36, n. 22, p. 3219-3228, 1980.

HAMID, Azzmer Azzar Abdul et al. **Molecular modelling and functional studies of the non-stereospecific α-haloalkanoic acid Dehalogenase (DehE) from Rhizobium sp. RC1 and its association with 3-chloropropionic acid (β-chlorinated aliphatic acid).** Biotechnology & Biotechnological Equipment, v. 27, n. 2, p. 3725-3736, 2013.

HO, Bosco K.; BRASSEUR, Robert. **The Ramachandran plots of glycine and pre-proline**. BMC structural biology, v. 5, n. 1, p. 14, 2005.

KAWABATA, Takeshi; GO, Nobuhiro. **Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites**. Proteins: Structure, Function, and Bioinformatics, v. 68, n. 2, p. 516-529, 2007.

KAWABATA, Takeshi. **Detection of multiscale pockets on protein surfaces using mathematical morphology**. Proteins: Structure, Function, and Bioinformatics, v. 78, n. 5, p. 1195-1211, 2010.

KIEFER, Florian et al. **The SWISS-MODEL Repository and associated resources.** Nucleic acids research, v. 37, n. suppl_1, p. D387-D392, 2008.

KILILI, Kimiti G. et al. **Differential roles of tau-class glutathione S-transferases in oxidative stress.** Journal of Biological Chemistry, 2004.

KREUZ, Klaus; TOMMASINI, Roberto; MARTINOIA, Enrico. **Old enzymes for a new job (herbicide detoxification in plants).** Plant Physiology, v. 111, n. 2, p. 349, 1996.

LAJMANOVICH, Rafael C. et al. **Individual and mixture toxicity of commercial formulations containing glyphosate, metsulfuron-methyl, bispyribac-sodium, and picloram on Rhinella arenarum tadpoles.** Water, Air, & Soil Pollution, v. 224, n. 3, p. 1404, 2013.

LASKOWSKI, Roman A. et al. **PROCHECK: a program to check the stereochemical quality of protein structures.** Journal of applied crystallography, v. 26, n. 2, p. 283-291, 1993.

LASKOWSKI, R. A.; MACARTHUR, M. W.; THORNTON, J. M. **PROCHECK: validation of protein-structure coordinates.** 2006.

MAIA, Rafael Trindade; NADVORNY, Daniela. **Molecular docking of Anopheles gambiae and Aedes aegypti glutathione S-Transferases Epsilon 2 (GSTE2) against usnic acid: an evidence of glutathione conjugation.** Brazilian Archives of Biology and Technology, v. 57, n. 5, p. 689-694, 2014.

MARRS, Kathleen A. **The functions and regulation of glutathione S-transferases in plants.** Annual review of plant biology, v. 47, n. 1, p. 127-158, 1996.

MELO, Francisco; FEYTMANS, Ernest. **Assessing protein structures with a non-local atomic interaction energy1.** Journal of molecular biology, v. 277, n. 5, p. 1141-1152, 1998.

MORRIS, Garrett M. et al. **AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.** Journal of computational chemistry, v. 30, n. 16, p. 2785-2791, 2009.

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA ALIMENTACIÓN Y LA AGRICULTURA. **El estado mundial de la agricultura y la alimentación. Organización de las Naciones Unidas para la Agricultura y la Alimentanción** - (FAO), p. 192, 2016.

Surhone LM, Timpledon MT, Marseken SF. **Visual Molecular Dynamics [J]**. Version 1.9.3 [software] 2010. Betascript Publishing. 2010. Available: https://www.ks.uiuc.edu/Research/vmd/.

THOM, Russell et al. **Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification.** Biochemistry, v. 41, n. 22, p. 7008-7020, 2002.

VAN GUNSTEREN, Wilfred F.; BERENDSEN, Herman JC. **Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry**. Angewandte Chemie International Edition in English, v. 29, n. 9, p. 992-1023, 1990.

VAN GUNSTEREN, Wilfred F, et al. **GROMOS force field**. Schleyer P.V.R. et al. (Orgs.). Encyclopedia of computational chemistry. 1998. p. 1211–1216.

WEINER, Scott J. et al. **A new force field for molecular mechanical simulation of nucleic acids and proteins**. Journal of the American Chemical Society, v. 106, n. 3, p. 765-784, 1984.

WILCE, Matthew CJ; PARKER, Michael W. **Structure and function of glutathione S-transferases.** Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, v. 1205, n. 1, p. 1-18, 1994.

YAMAMOTO, Kohji et al. **Structural basis for catalytic activity of a silkworm Delta-class glutathione transferase.** Biochimica et Biophysica Acta (BBA)-General Subjects, v. 1820, n. 10, p. 1469-1474, 2012.

SOBRE O ORGANIZADOR

ERNANE ROSA MARTINS Doutorado em andamento em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa, em Porto/Portugal. Mestre em Engenharia de Produção e Sistemas pela PUC-Goiás, possui Pós-Graduação em Tecnologia em Gestão da Informação pela Anhanguera, Graduação em Ciência da Computação pela Anhanguera e Graduação em Sistemas de Informação pela Uni Evangélica. Atualmente é Professor de Informática do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia), ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador do Núcleo de Inovação, Tecnologia e Educação (NITE).

Agência Brasileira do ISBN
ISBN 978-85-7247-113-8