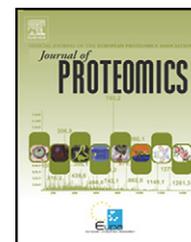


available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/jprot](http://www.elsevier.com/locate/jprot)

## Technical note

# Caititu: A tool to graphically represent peptide sequence coverage and domain distribution

Paulo C. Carvalho<sup>a</sup>, Magno Junqueira<sup>b</sup>, Richard H. Valente<sup>c</sup>, Gilberto B. Domont<sup>d,\*</sup>

<sup>a</sup>Systems Engineering and Computer Science Program, COPPE, Federal University of Rio de Janeiro, Brazil

<sup>b</sup>Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

<sup>c</sup>Laboratory of Toxinology, FIOCRUZ/IOC, Av. Brasil 4365, Rio de Janeiro, CEP 21045-900, Brazil

<sup>d</sup>Laboratory of Protein Chemistry, Institute of Chemistry, Federal University of Rio de Janeiro, Brazil

### ARTICLE INFO

#### Article history:

Received 9 June 2008

Accepted 27 June 2008

#### Keywords:

Proteomics

Software tool

Toxinology

Caititu

Protein coverage

### ABSTRACT

Here we present Caititu, an easy-to-use proteomics software to graphically represent peptide sequence coverage and domain distribution for different correlated samples (e.g. originated from 2D gel spots) relatively to the full-sequence of the known protein they are related to. Although Caititu has a broad applicability, we exemplify its usefulness in Toxinology using snake venom as a model. For example, proteolytic processing may lead to inactivation or loss of domains. Therefore, our proposed graphic representation for peptides identified by two dimensional electrophoresis followed by mass spectrometric identification of excised spots can aid in inferring what kind of processing happened to the toxins, if any. Caititu is freely available to download at: <http://pccarvalho.com/things/caititu>.

© 2008 Elsevier B.V. All rights reserved.

## 1. Technical note

In proteomics, analyzing identified peptides of correlated samples as to sequence coverage and domain distribution in regard to their known related protein is an important follow-up step. An example is to assay protein coverage from two dimensional gel electrophoresis spots identified by mass spectrometry. Compiling such information into a graphical representation can become a dull and laborious task, especially because, as far as we know, there are no specialized tools for this chore.

The present tool was originally created to analyze data generated from snake venom proteomics experiments at our labs [1]. The continuous interest in this field of Toxinology

relies on the fact that a number of snake venom toxins or its derivatives are involved in therapeutic and diagnostic applications [2]. When studying these toxins, one should keep in mind that some of them, such as the multidomain metallo-proteinases, may undergo proteolytic processing leading to their inactivation or to the release of an active domain [3,4]. Other toxins such as phospholipases A<sub>2</sub> may present themselves as molecular isoforms, with high sequence identity but totally different isoelectric points [5]. Finally, venom molecules such as the C-type lectin-like or C-type lectin proteins, are prone to heteromeric or homodimeric oligomerization, respectively [6]. Hence, when analyzing data originated from snake venom bidimensional electrophoresis (under denaturing, reducing and alkylating conditions) followed by spot

\* Corresponding author. Universidade Federal do Rio de Janeiro, Instituto de Química, Departamento de Bioquímica, Av. Athos da Silveira Ramos 149 Centro de Tecnologia, Bloco A-sala 543-Ilha do Fundão, Rio de Janeiro, RJ 21941-909, Brasil. Tel.: +55 21 25627353; fax: +55 21 24955765.

E-mail address: [gilberto@iq.ufrj.br](mailto:gilberto@iq.ufrj.br) (G.B. Domont).



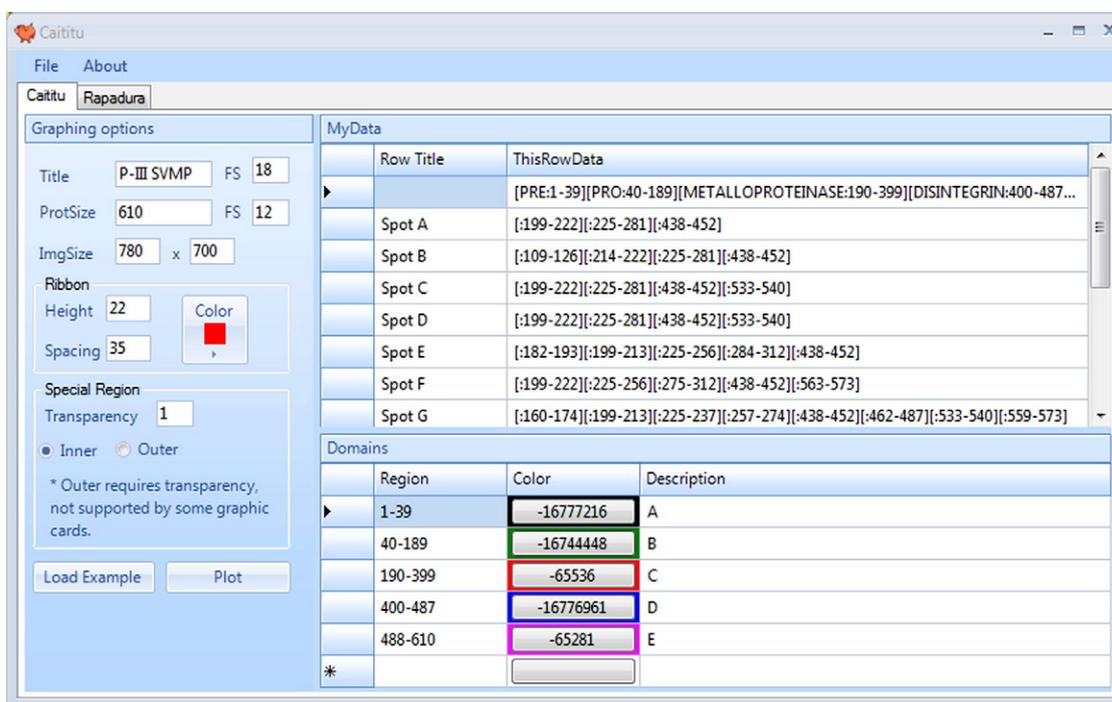


Fig. 2 – Caititu's main interface.

excision and mass spectrometric identification, one would like to have a tool that can graphically represent identified peptides distributed along the multidomain structure they originated from, in order to infer what kind of processing happened to those toxins, if any.

Here we present new software named Caititu that was used to serve this purpose but that has general applicability. As shown in Fig. 1, it has been specifically useful for the analysis of sequenced peptides from 2D-spots having different pI and molecular masses (data not shown) but identified by mass spectrometry as belonging to the same protein (a P-III metalloproteinase). The graphical representation of the data indicates that proteolytic events occurred, to different extents, in these molecules.

Caititu has graphic user interface and was designed to be of simple usage; in fact, it can be installed with one single click of a mouse in a Windows (XP or VISTA) based PC. Caititu's main screen is presented on Fig. 2. The latter is divided into two panels; the graphing options pane (left) and the data pane (right).

The data pane is divided into two sections (i.e., MyData and Domains) where the information from the identified peptides can be entered. The former contains a table with two columns (Row title, ThisRowData). In each row, the user should enter information regarding identification as to its title (i.e., Spot A) in column 1, and the identified regions, in column 2. The region data comprises as to what amino acid numbers the

identified peptide maps to in the reference protein and must abide by the following format:

[InitialAminoAcidNumber<sub>1</sub>–FinalAminoAcidNumber<sub>1</sub>]

[InitialAminoAcidNumber<sub>2</sub>–FinalAminoAcidNumber<sub>2</sub>]...

[InitialAminoAcidNumber<sub>n</sub>–FinalAminoAcidNumber<sub>n</sub>].

For example, suppose that Spot A protein was identified by two peptides; also, let the first peptide be a match to the reference protein within amino acids 200 to 300, and the second, from 350 to 400. This information would be encoded as: Column 1:Spot A; Column 2:[:200–300][:350–400]. Moreover, if one would like to associate further information regarding Spot A, such as experimentally determined pI and M<sub>r</sub>, it could be typed in the RowTitle column.

The lower part of the data pane is used to delimit domain distributions and is composed of three columns: Region, Color, and Description. The region field follows the format InitialAminoAcid–FinalAminoAcid. The Color field is assessed by clicking over the cell; the user will then be prompted with a color box to pick a color to highlight the domain. Finally, the description consists in a text message describing the domain. An example of such table including data from 5 domains is exemplified in Fig. 2.

The left pane offers mostly options to customize the graph. For example, the user may enter the graphs title followed by

**Fig. 1 – Examples of the Inner (A) and Outer (B) Caititu plots. Data represent spots that were assigned as being the same protein (a P-III snake venom metalloproteinase). From the gel we could observe that the spots had different molecular masses and pIs. Indeed, the alignment of the sequenced peptides indicates that spots I through M present only their disintegrin-like and cysteine-rich domains while the other spots still possess the extra metalloproteinase domain.**

its font size (FS) in the Title and FS fields. The image size may be set as to the number of desired pixels in the Image Size field. In output graph, each protein will be represented as a line, and the sequence coverage, as ribbons over that line (Fig. 1). The ribbon color, height and spacing can also be customized by completing the fields in the Ribbon group. Finally, the reference protein size (number of amino acids), should also be entered before plotting the final graph. The user should then opt to one of the different output flavors as to Inner (Fig. 1A), or Outer (Fig. 1B). We note that outdated graphic cards do not support transparency effects, therefore are unable to handle the Outer graphing style.

The plot can be obtained by clicking on the Plot button and then saved to the PNG format by clicking in the file menu and choosing the option “Save Graph”. The data used to generate the graph can be exported to a XML file by accessing the file menu and choosing the option “Save Tables”. Caititu has an automatic update feature to ensure that the user will always experience the latest features. Nevertheless, if one is unsatisfied with new modifications, or if breaking changes have occurred, a rollback function is available to restore to the previous installed version. This feature is made available in the windows control panel and selecting Caititu and the change option. To ease user experience, Caititu has a button entitled “Load Example”; by clicking on it, a complete and pre-configured example will be loaded. Caititu requires a Windows based-PC with Windows XP or VISTA and .NET 3.5 or later. In case the user does not have the latest .NET framework, Caititu can automatically update the computer by interfacing with Microsoft’s website. Taken together, Caititu is an easy-to-use tool for proteomics that allows the user to graphically represent peptide sequence coverage and domain distribution

for different correlated samples (ex. 2D gel spots) relatively to the full-sequence of the known protein they are related to.

---

## Acknowledgments

The authors acknowledge CAPES, CNPq, FAPERJ, the Rio de Janeiro Proteomic Network and the Genesis Molecular Biology Laboratory for financial support.

---

## REFERENCES

- [1] Junqueira MR. Análise proteômica do veneno da serpente *Bothrops insularis*, Cellular and Molecular Biology Master Program. Rio de Janeiro: FIOCRUZ; 2005.
- [2] Fox JW, Serrano SM. Approaching the golden age of natural product pharmaceuticals from venom libraries: an overview of toxins and toxin-derivatives currently involved in therapeutic or diagnostic applications. *Curr Pharm Des* 2007;13:2927–34.
- [3] Fox JW, Serrano SM. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. *Toxicon* 2005;45:969–85.
- [4] Fox JW, Serrano SM. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. *FEBS J* 2008;275:3016–30.
- [5] Kini RM, Iwanaga S. Structure-function relationships of phospholipases. II: Charge density distribution and the myotoxicity of presynaptically neurotoxic phospholipases. *Toxicon* 1986;24:895–905.
- [6] Morita T. Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities. *Toxicon* 2005;45:1099–114.