Buzatto *et al. BMC Bioinformatics* (2016) 17:325
DOI 10.1186/s12859-016-1207-2

BMC Bioinformatics

SOFTWARE Open Access

CryGetter: a tool to automate retrieval and analysis of Cry protein data

David Buzatto[1*], Suzelei de Castro França[2] and Sônia Marli Zingaretti[2]

Abstract

Background: For many years, the use of chemical agents to control crop pests has been degrading the environment, bringing problems to humans and all living things. An alternative to deal with the pests is the use of biopesticides, biological agents capable of controlling these harmful organisms. One kind of biopesticide is *Bacillus thuringiensis*, a Gram-positive bacterium that synthesizes a protein that, when ingested by the pests, kills them and does not harm other species.

Results: Since the economical importance of *Bacillus thuringiensis* and its proteins significance, this work presents a software tool, called CryGetter, that is capable of retrieving data related to these proteins, store it and present it in a user friendly manner. The tool also aims to align the protein sequences and generate reports containing some statistical data concerning the alignments that were made.

Conclusions: CryGetter was created to help researchers of *Bacillus thuringiensis* and its proteins to speed up their data retrieval and analysis, allowing them to generate more accurate results. In this sense, the tool circumvents the error prone task of manually getting all the necessary data and processing them in various software systems to get the same result as CryGetter gets in a unique semiautomatic environment.

Keywords: Cry protein, Protein analysis, Sequence alignment, Automatic data retrieval

Abbreviations: ADT, Abstract data type; ALP, Alkaline phosphatase; APN, Aminopeptidase-N; BBMV, Brush border membrane vesicles; Bt, Bacillus thuringiensis; CADR, cadherin-like protein; DTD, Document Type definition; GPI, Glycosylphosphatidyl-inositol; HTML, HyperText markup language; HTTP, Hypertext transfer protocol; JAXB, Java architecture for xml binding; JDK, Java development kit; JRE, Java runtime environment; MEGA, Molecular evolutionary genetics analysis; MSA, Multiple sequence alignment; NCBI, National center for biotechnology information; PDB, Protein data bank; PMDB, Protein model database; SE, Standard edition; UML, Unified modeling language; URL, Uniform resource locator; VMD, Visual molecular dynamics; XLSX, Office open XML SpreadsheetML file format; XML, Extensible markup language

Background

The biopesticides produced by the *Bacillus thuringiensis* (*Bt*) bacterium are a viable alternative for crop pest control using chemical pesticides [1, 2] without the collateral effects of environment contamination, since the toxins synthesized by *Bt* have little effect on non-target insects and vertebrates like birds and mammals [3–7]. The *Bt* is a bacterium present in the soil and produces a protein called Crystal protein (Cry protein) during the sporulation phase. Such protein is lethal to various insect orders [8], including *Coleoptera*, *Lepidoptera* and *Diptera*.

To date, about 600 genes that encode the Crystal protein from different isolates were identified [1]. These genes are sorted according to the insect order to which the protein is toxic to [9–11]. Once ingested (Fig. 1, sections 1 and 2), the Cry protein acts in the insect gut by opening pores in the intestinal membrane (Fig. 1, section 3), which causes the death of larvae due to starvation and/or bacterial infection (Fig. 1, section 4). There are two hypotheses regarding how these proteins act [12, 13]. Both propose the interaction of protein receptors present

*Correspondence: davidbuzatto@ifsp.edu.br
[1] Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP, Câmpus São João da Boa Vista, Acesso Dr. João Batista Merlin, s/n, Jardim Itália, 13872-551 São João da Boa Vista, SP, Brazil
Full list of author information is available at the end of the article

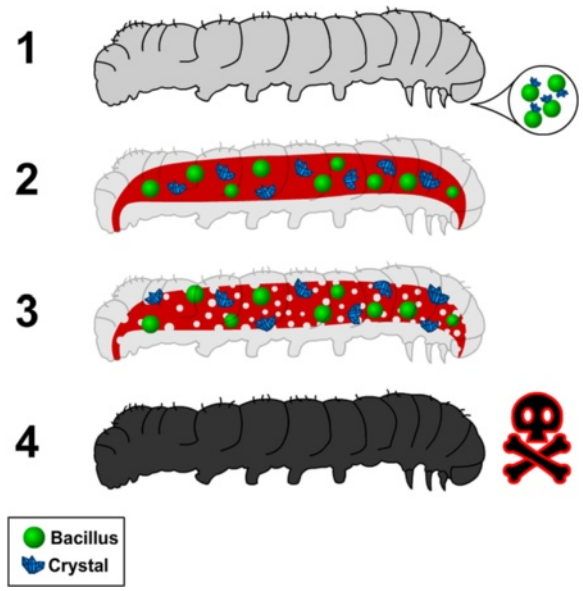

Fig. 1 Cry protein mode of action (adapted from [50])

in the insect's gut; on both models, the cadherin proteins (CADR) are the first receptors to be activated. This interaction between the Cry proteins and their receptors occurs in some regions of the gene known as Domains I, II and III. Domain I is involved in the process of membrane insertion and pore formation. Domains II and III are both related to receptor recognition and binding. The third domain is also related to the role of pore formation [11].

Regarding *Lepidoptera*, at least four types of protein receptors are involved in the binding process of proteins to Brush Border Membrane Vesicles (BBMV), beginning the process of pore formation in different *Lepidoptera* larvae: a cadherin-like protein (CADR), a glycosylphosphatidylinositol (GPI)-anchored aminopeptidase-N (APN), a GPI-anchored alkaline phosphatase (ALP) and a 270kDa glycoconjugate [14]. The protein encoded by the cry genes from some strains of *Bt* are specific to certain insect orders, while others do not exhibit this specificity and can act on two or more different orders. The reason for this specificity is still unknown, but amino acid changes in this region have been associated with toxicity. Tiewsiri and Angsuthanasombat [15] made the substitution by an Alanine (A) in four highly conserved aromatic residues (^{242}W^{244}, ^{245}F^{247}, ^{248}Y^{250} and ^{263}F^{265}) in Cry4B gene, that has a toxin that attacks insects of the *Diptera* order. This resulted in a decrease of toxicity to the mosquito *Stegomyia aegypti*. In recent decades, several plants of different species have been transformed with genes of *Bt* to be commercialized (*Zea mays*, *Gossypium hirsutum*, *Glycine canescens*, *Oryza sativa*, etc.). The pattern recognition in the amino acid sequence of these proteins, that may be associated with specificity, could facilitate the use of these genes in the generation of new transgenic plants resistant to different crop pests, as well as the construction of *Bt* pyramided plants [16, 17], which is a reality nowadays.

There are many software tools capable of analyzing protein sequences like MEGA 6 (Molecular Evolutionary Genetics Analysis) [18, 19], MacClade [20] and Geneious [21–23]. MEGA 6 is capable of storing and processing protein sequences in order to align them, generate phylogeny trees and perform other calculations. It can also work with DNA sequences. The tool can even search online databases for data of interest that can be downloaded free of charge. Like MEGA 6, MacClade is also capable of analyzing DNA sequences to construct their phylogeny trees, allowing the user to identify molecular evolution among the tested gens, but it is only available for Mac OS. Geneious is a more complete tool, since it supports the features of MEGA 6 and MacClade in addition to having more options for protein and DNA analysis. It can also be extended by the use of plug-ins [24, 25]; however it is a paid tool.

In order to support the study of Cry proteins, we created a specialized and open-source software tool that is capable of compile the data of Cry proteins in a central repository, allowing the retrieval of data related to each protein such as their primary structures, three-dimensional models (PDB files), related works, etc., allowing the manipulation of such data in alignment algorithms and generation of reports related to the alignments. This tool, called CryGetter, is presented and detailed in this work.

Implementation

To implement CryGetter, we used the Java SE (Standard Edition) Platform, version 8 [26] and other third party tools and libraries to perform some tasks like protein sequence alignment, alignment visualization (using MView [27] tool) and molecule rendering (using BioJava [28] and Jmol [29] libraries). In addition, CryGetter uses the data of the full toxin list of *Bacillus thuringiensis* Toxin Nomenclature website [30] and Entrez [31] (Global Query Cross-Database Search System) service of NCBI (National Center for Biotechnology Information) for automatic retrieval of protein data and Cry protein models available at the PDB [32, 33] (Protein Data Bank) and the PMDB [34, 35] (Protein Model Database). The application architecture is shown in Fig. 2 and each section of this architecture is explained below.

The "Extraction Module", highlighted in Fig. 2 using the number 1, represents the data extraction feature of CryGetter that is performed by this module of the tool and can be executed pressing the button "Extract" highlighted in section A1 of Fig. 3.

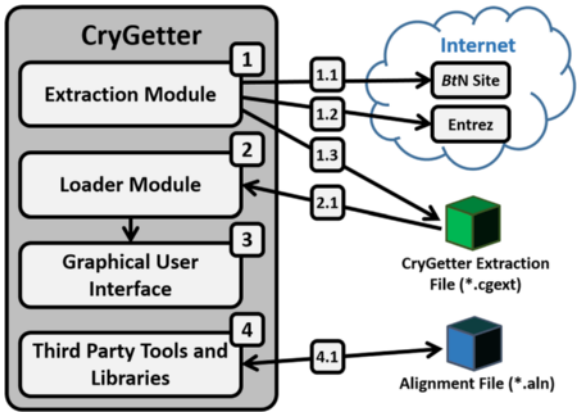

Fig. 2 CryGetter application architecture

When pressed, this button triggers the execution of the extraction module that executes four steps:

1. **BtNomenclature site data extraction:** First, the extraction module retrieves the data of a particular HTML (HyperText Markup Language) file [1] of the BtNomenclature website (arrow 1.1 in Fig. 2). This file contains a HTML table with all Cry proteins that already studied and cataloged by the website owner;
2. **Cry protein data preprocessing:** The data from the raw HTML file gathered in the previous step is then processed using the jsoup [36] library to create a linked list of an ADT (Abstract Data Type) called "CryToxin", which contains only Cry proteins entries (from BtNomenclature website) that has a NCBI hyperlink related to a protein sequence. This ADT and its composition are presented in the UML (Unified Modeling Language) class diagram showed in Fig. 4. This class composition was modeled this way because we need to serialize this data from Java objects to XML (Extensible Markup Language) and deserialize it from XML to Java objects;
3. **NCBI data extraction:** Using the Entrez service (arrow 1.2 in Fig. 2), all proteins GI numbers of each Cry protein entry that was acquired in the previous steps are used to retrieve all data related to each protein. In this case, the URL (Uniform Resource Locator) used to access the Entrez service is https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi and some URL parameters need to be passed:

- **tool:** the name of the tool that is accessing the Entrez service. In this case, "crygetter";
- **email:** e-mail of the responsible for the tool. In this case, "davidbuzatto@ifsp.edu.br";
- **db:** the database that is being accessed. In this case "protein";
- **retmode:** the type of data that will be returned by the service. In this case "xml";
- **id:** the set of ids, separated by commas, that represents the proteins. In this case, all Cry proteins GI that were retrieved in the previous steps;
- **Obs:** more details about the Entrez HTTP (Hypertext Transfer Protocol) interface can be found in its documentation [31].

The result of the request to the Entrez service, in XML format, is processed by the tool, generating a set of temporary files that will later be used in the next step. To process these generated files, the tool

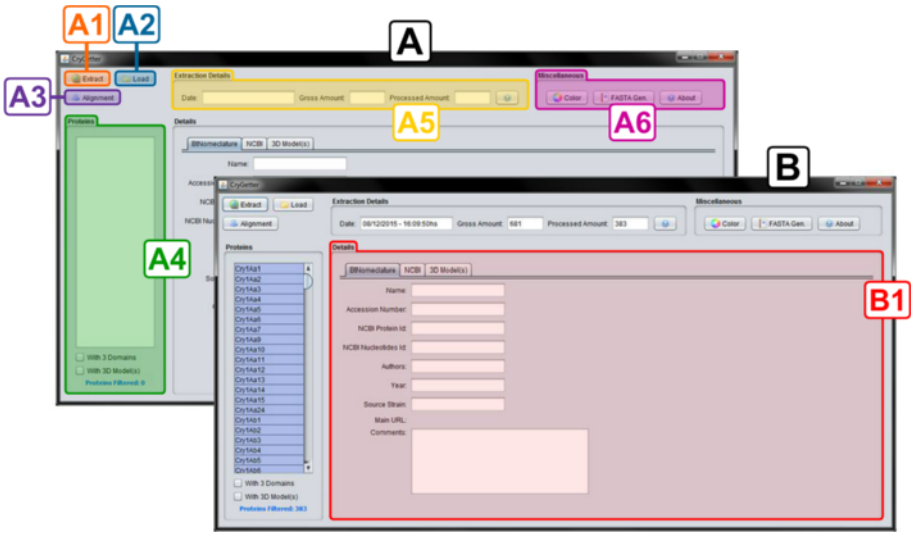

Fig. 3 CryGetter main interface and its features

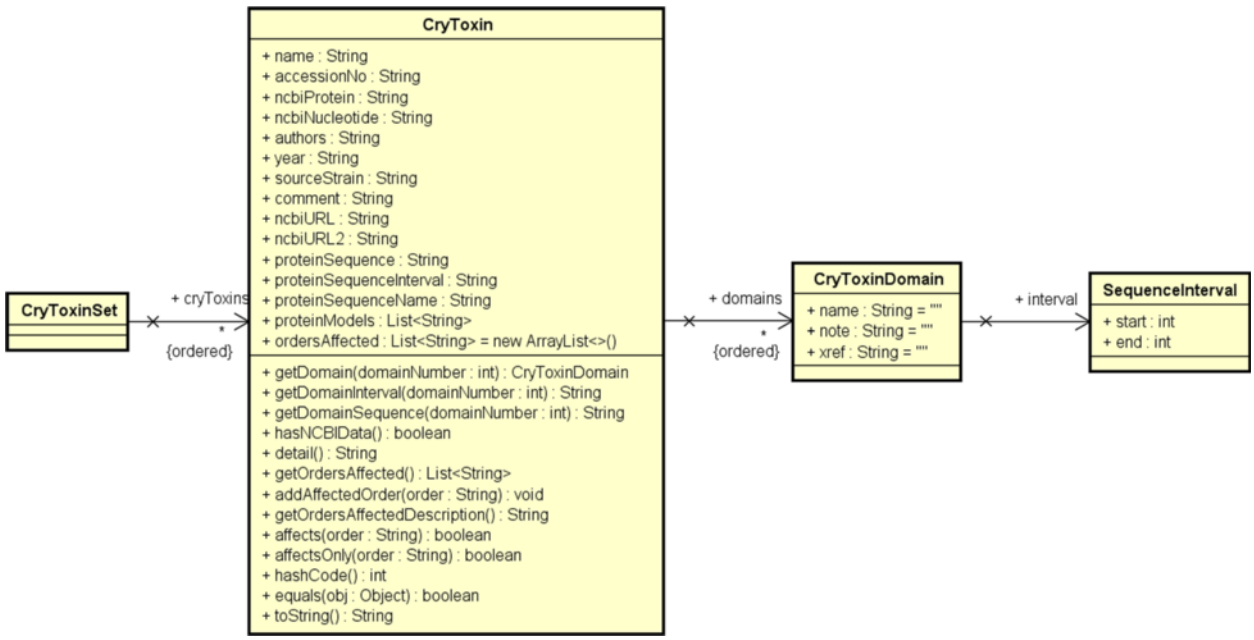

Fig. 4 Class diagram for CryToxin ADT and its composition

needs to parse the returned XML and generate a object composition in memory. To do this, a class composition that reflects the XML structure was created and is shown in Fig. 5. For the sake of simplicity, in this diagram we do not show the class compartments for attributes and operations. This composition is used by the tool to store the result of the XML deserialization, allowing the data to be processed and used to further complement the CryToxin ADT data;

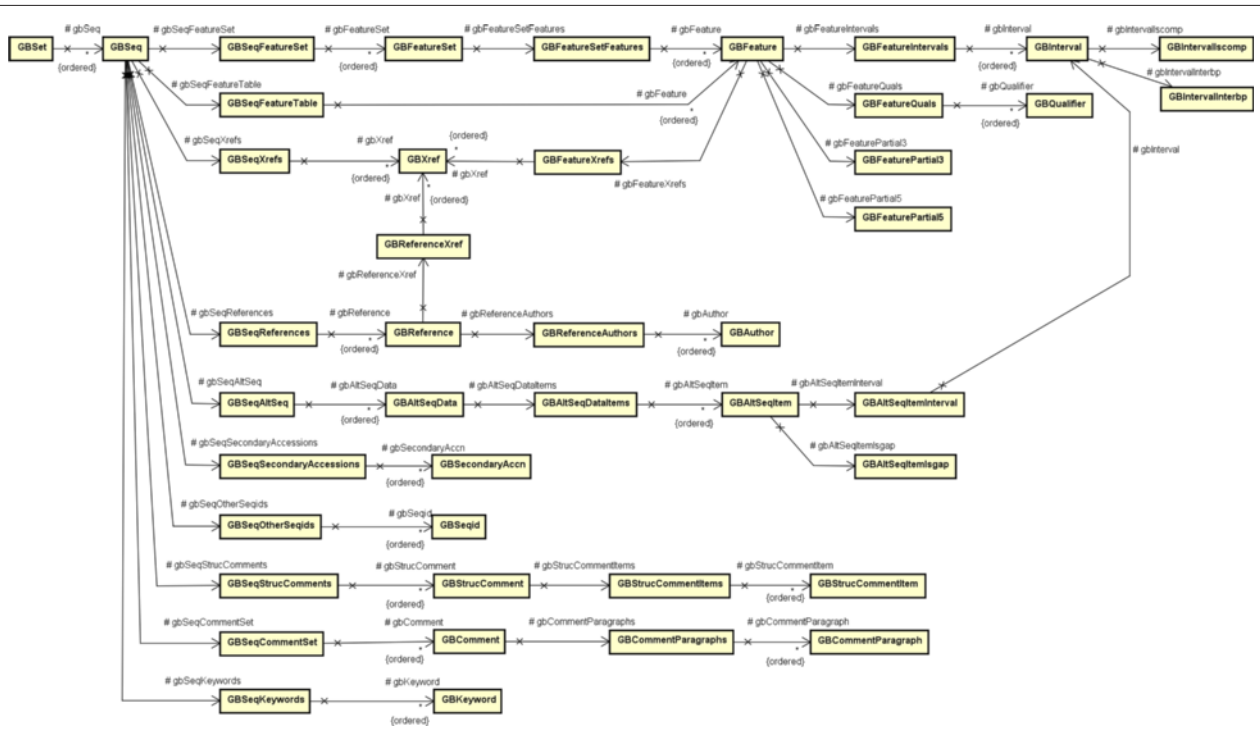

Fig. 5 Class diagram for the XML structure defined by Entrez service return

4. **CryGetter extraction data file generation:** The last step comprises the creation of a file that will be saved by the user (arrow 1.3 in Fig. 2). This file will contain all the data that was obtained and it will be able to be opened by anyone who has the tool. To process the stored data we used two XML parsing libraries: SimpleXML [37] for serialization and deserialization of CryToxin ADT and JAXB [38] (Java Architecture for XML Binding) for serialization and deserialization of Entrez XML data. We chose SimpleXML to process the tool specific XML data and JAXB to manipulate the Entrez return, since the class composition for Entrez was automatically generated by the JAXB compiler, available by default in any JDK (Java Development Kit), using the Entrez DTD [2] (Document Type Definition).
   To open an extraction file, the user must click on the "Load" button highlighted in section A2 of Fig. 3. By clicking on this button, an open dialog box will appear and the user will choose the file that s/he wants to open. When a file is chosen, the "Loader Module" (highlighted in Fig. 2 using the number 2) will execute, unpacking the file (arrow 2.1 in Fig. 2) and presenting its contents in the main interface (represented in Fig. 2 by the "Graphical User Interface" section (highlighted by the number 3) of the tool as showed in section B of Fig. 3.

As result, all Cry protein data is loaded into the tool and shown in the left list (the empty list is highlighted in section A4 of Fig. 3 and the filled list is presented in section B of the same Figure), allowing the user to select the protein that s/he wants to perform tasks like data visualization (section B1 of Fig. 3), protein alignment (accessed by the button "Alignment", presented in section A3 of Fig. 3), protein alignment visualization, protein analysis and FASTA file generation (section A6 of Fig. 3). All these tasks use one or more third party libraries or third party tools (represented in Fig. 2 by the "Third Party Tools and Libraries" section, highlighted by the number 4) and each task is explained hereafter.

Cry protein data visualization

The main functionality related to Cry proteins is data visualization. When the user selects a protein in the list on the left, like Cry1Aa1, all of its related data are shown in tabs located on the right, represented in Fig. 3 by section B1. There are some tabs organized in a hierarchical mode:

- **BtNomenclature:** It presents the data collected in the BtNomenclature website. This tab is shown in Fig. 6;
- **NCBI:** It presents the data collected in NCBI, through Entrez service. This tab is shown in Fig. 7 and the details are divided into three tabs:
  - **Main:** The main data related to the selected protein (section A of Fig. 7);
  - **References:** The set of references related to that protein (section B of Fig. 7);
  - **Sequence:** The amino acid sequence of the selected protein (primary structure), divided into three domains (if these domains are present in the NCBI data) and a diagram showing the position of the domains inside the entire sequence. This tab is then divided in four tabs:
    – **Complete Protein:** The complete amino acid sequence of the protein (section C of Fig. 7);
    – **Domain 1, 2 and 3:** Three tabs containing the data of each domain of the protein (Domain 1 shown in section D of Fig. 7).
- **3D Model(s):** In this tab, the 3D models (if they exist) of the protein are listed (Fig. 8), so the user can use them in some third party visualization tool like VMD [39, 40] (Visual Molecular Dynamics), Swiss PDB Viewer [41, 42] and/or PyMol [43] to visualize and/or process them. It's important to emphasize that, nowadays, there are 22 experimental models, related to 18 different Cry proteins, deposited in PDB and PMDB and all these models are available in CryGetter. The Table 1 shows the proteins that have models available in PDB and PMDB along with their related works. In this table we also list some proteins that have papers reporting the creation of their models, but these models are note available neither in PDB nor in PMDB. It is also important to note that CryGetter isn't able to generate new models by using some protein structure prediction software.

Other features of the tool's main interface are the extraction details (the date of extraction, gross amount of gathered proteins and the processed amount of the proteins, highlighted in the section A5 of Fig. 3), the color change dialog (to change the color of the presented proteins inside the protein list) and the "FASTA Gen." button, used to generate a set of FASTA files of all proteins (section A6 of Fig. 3). For each protein, four files are created, one with the complete sequence and one for each of the three domains. In addition of these features, the tool can perform protein alignment for further analysis. This functionality and its sub features are presented in the next section.

Cry protein alignment

The "Cry Protein Alignment" interface can be accessed through the "Alignment" button located on the top left corner of the CryGetter main interface (section A3 of Fig. 3) and it is shown in Fig. 9.

Fig. 6 BtNomenclature details tab

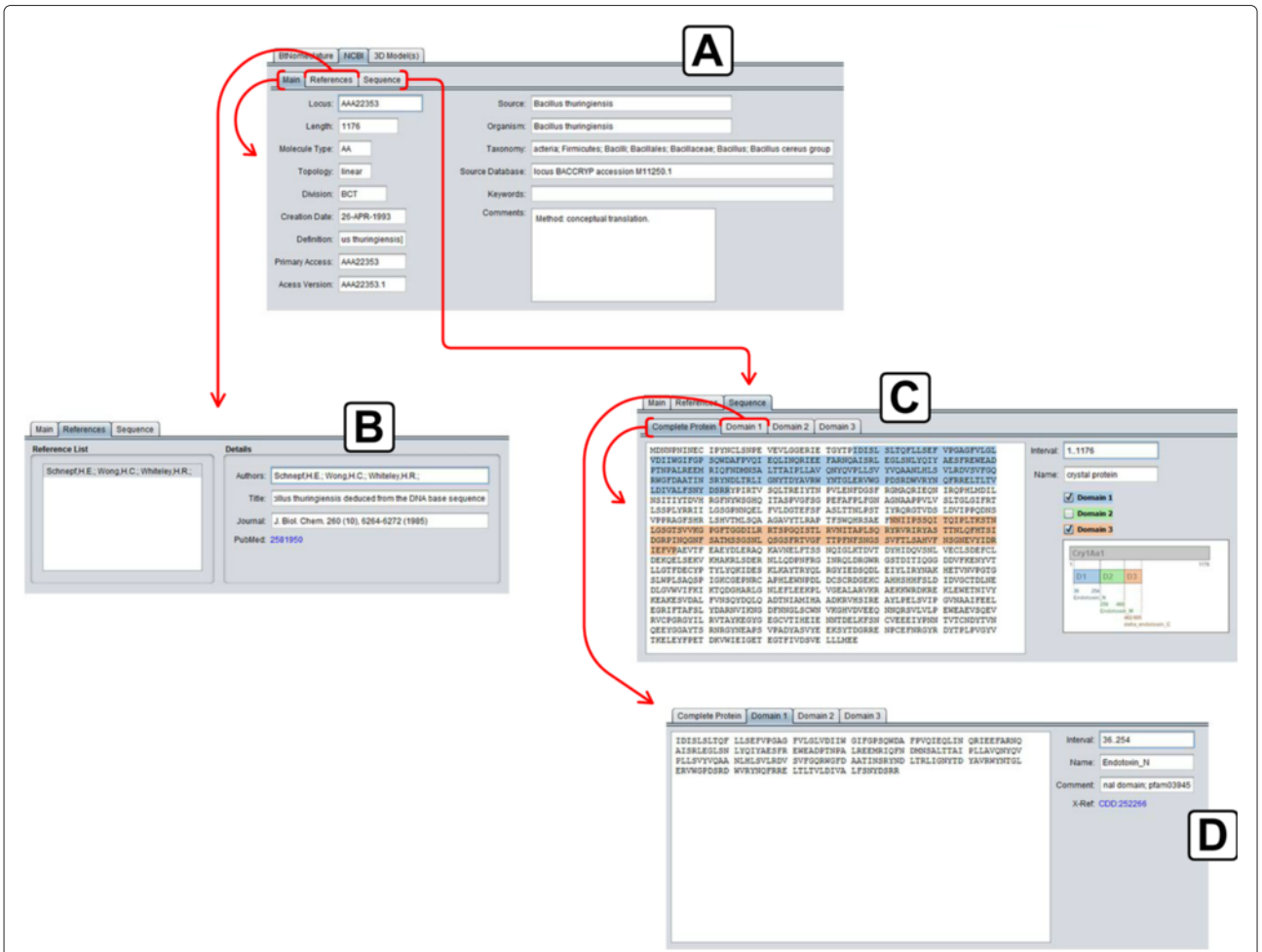

Fig. 7 NCBI details tab. **a**) Main tab of NCBI details tab; **b**) References tab of NCBI details tab; **c**) Sequence tab of NCBI details tab; **d**) Domain 1 tab of Sequence tab

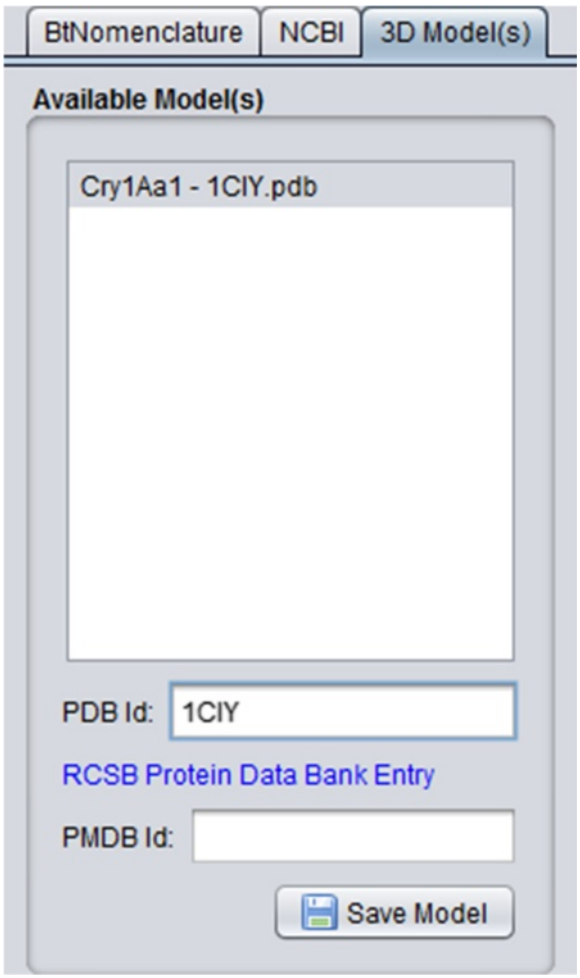

Fig. 8 3D models details tab

In this interface, the user can perform a set of protein alignment tasks with the Cry proteins, generating and/or loading alignment data files (arrow 4.1 in Fig. 2). There are some filters that can be used, such as selecting the proteins by the affected orders. The orders affected by each protein are stored in a XLSX (Office Open XML SpreadsheetML File Format) file that is processed by the Apache POI [44] library. The user can define what sections of the selected proteins s/he wants to align and the MSA (Multiple Sequence Alignment) algorithm that should perform the alignment task. In the current version of the tool, three algorithms are available: Clustal Omega [45], Clustal W [45, 46] and MUSCLE [47]. These three algorithms, executed by external tools, can also be parametrized in CryGetter. As an example, in Fig. 10 we show the result of the alignment of Cry1Aa1, Cry1Aa2 and Cry1Ab1 after executing the Clustal Omega algorithm. When the algorithm finishes its execution, a dialog box is shown to the user, allowing s/he to save the alignment data, which can now be visualized or analyzed.

Alignment results processing

When two or more Cry proteins are aligned, CryGetter can perform an analysis on the alignment result. When clicking the "Analysis" button (Fig. 10), a dialog appears and the user is asked to choose the alignment data file that was generated. By doing this, the program reads the data and shows the "Protein Analysis" interface, as shown in Fig. 11.

In this interface, the user can analyze the alignment by choosing two different proteins and clicking on the "Analyse" button. Doing so, the program will identify the differences between the selected proteins and show them in the differences list. In the example shown in Fig. 12, the differences between the proteins at location 206 is presented. In this case, Cry1Ab1 has an $^{205}H^{207}$ (Histidine) amino acid and Cry1Aa1 has a $^{205}Y^{207}$ (Tyrosine) amino acid. The user can even generate a report that summarizes the alignment data. This feature is presented in the next section, where an experiment comprising the analysis of two different Cry proteins was conducted.

Results and discussion

CryGetter is currently being used as a support tool for two works related to Cry proteins structure analysis. The tool is proving to be useful, since it can automatize the data extraction task and perform a sort of functions related to proteins, drastically reducing possible errors in manual data collection of online sources, since if such task was being done manually, the chance of human error would be enormous, because the process of copy/paste large quantities of data, for example, of protein sequences available through a NCBI web-page to a specific tool, would probably add undesirable and difficult to detect errors like, for example, data truncation, unless all sequences are constantly reviewed. In the next subsection we present a study case of the use of the tool.

Using CryGetter to Analyse two different order specific cry proteins

For this study, two Cry proteins were chosen to be analyzed based in their order specificity:

- **Cry1Aa1:** this Cry protein affects mainly the *Diptera* order;
- **Cry3Aa1:** the *Coleoptera* order is affected by this type of Cry protein.

After loading a generated data package in CryGetter, the "Alignment" button (presented in section A3 of Fig. 3) was clicked, aiming to show the "Alignment Interface". Since we wanted to analyse Cry1Aa1 and Cry3Aa1, these

Table 1 Cry proteins models available in PDB and PMDB and others just described in literature

Protein	Affected order(s)	Model Id[a]	Reference(s)	Obs.
Cry1Aa1	*Diptera*, *Lepdoptera* and *Gastropoda*	1CIY	[51, 52]	
Cry1Ab16	*Diptera*, *Lepdoptera* and *Gastropoda*	unavailable	[53]	There isn't a deposited model
Cry1Ab19	*Diptera*, *Lepdoptera* and *Gastropoda*	unavailable	[54]	There isn't a deposited model
Cry1Ac1	*Diptera*, *Lepdoptera* and *Gastropoda*	4ARX 4ARY 4W8J	Not yet published Not yet published [55]	
Cry1Ld	*Lepdoptera*	unavailable	[56]	There isn't a deposited model
Cry2Aa1	*Diptera*, *Hemiptera* and *Lepdoptera*	1I5P	[57]	
Cry3A	*Coleoptera*, *Hemiptera* and *Hymenoptera*	1DLC	[58]	
Cry3Aa1	*Coleoptera*, *Hemiptera* and *Hymenoptera*	4QX0 4QX1 4QX2	[59]	
Cry3Bb1	*Coleoptera*	1JI6	[60]	
Cry4Aa1	*Diptera*	2C9K	[61, 62]	
Cry4Ba1	Diptera	1W99 4MOA	[63] Not yet published	
Cry5Aa1	*Hymenoptera* and *Rhabditida*	PM0074964	[64]	
Cry5B	*Rhabditida*	4D8M	[65]	
Cry5Ba1	*Rhabditida*	PM0075036	[66]	
Cry6Aa	*Rhabditida*	5J66	Not yet published	Pore formation
Cry23Aa1 Cry37Aa1	*Coleoptera*	4RHZ	Not yet published	Binary protein complex
Cry8Ea1	*Coleoptera*	3EB7	[67]	
Cry11Bb1	*Diptera*	unavailable	[68]	There isn't a deposited model
Cry30Ca2	*Diptera*	unavailable	[69]	There isn't a deposited model
Cry34Ab1	*Coleoptera*	4JOX	[70]	
Cry35Ab1	*Coleoptera*	4JP0	[70]	
Cry51Aa1	*Coleoptera* and *Hemiptera*	4PKM	[71]	

[a]The model ids with 4 characters are from PDB while the model ids with 9 characters are from PMDB

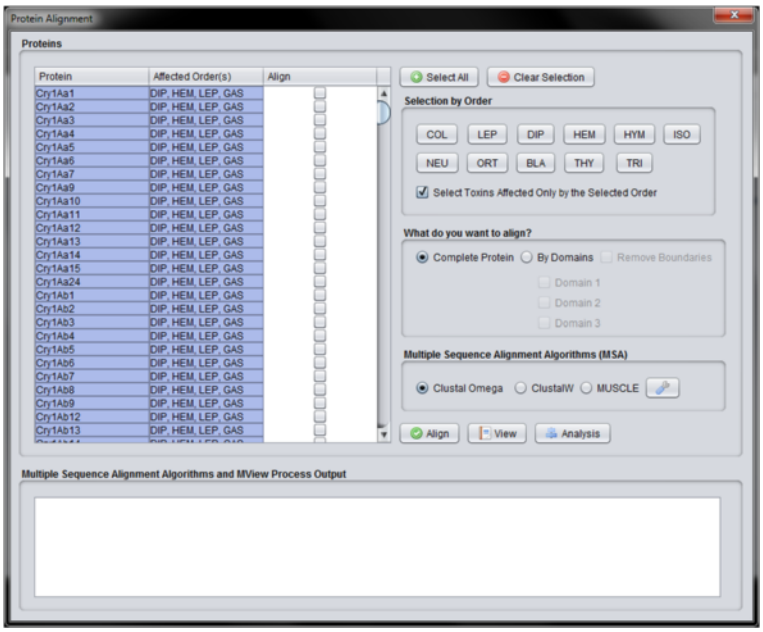

Fig. 9 Protein alignment interface

two protein were selected (clicking on the check-box of the "Align" column) in the protein list (highlighted by the section A of Fig. 13). After the selection, in section B of Fig. 13 we chose to align only the Domain 2 of both proteins, since this domain is mainly responsible for receptor recognition [11] and we wanted to study what is the difference between these two proteins that made their specificity occur. In section C of Fig. 13, the Clustal Omega MSA was selected and finally the "Align" button (section D of Fig. 13) was clicked on. Doing this, the Clustal Omega executable performed the alignment of the selected sequences and a file with the result was saved on disk. The execution output of the selected MSA is shown in section E of Fig. 13.

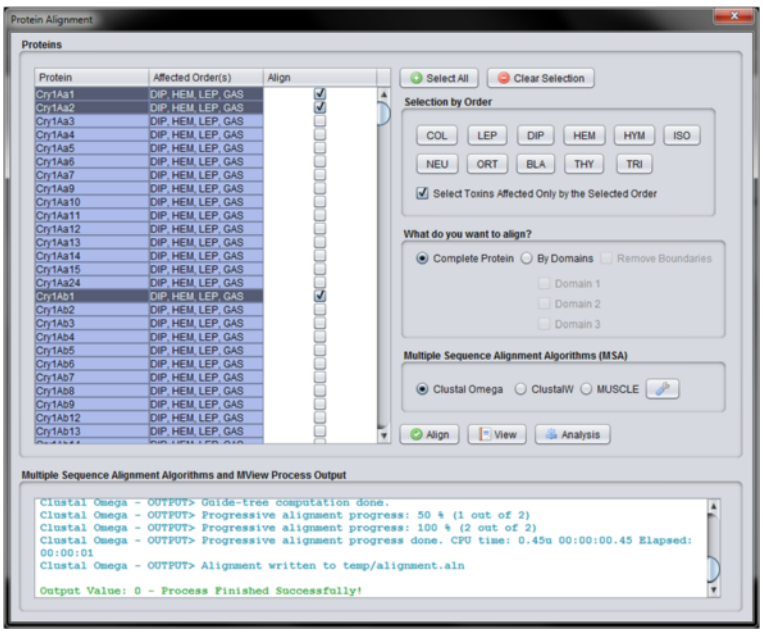

Fig. 10 Alignment result of Cry1Aa1, Cry1Aa2 and Cry1Ab1

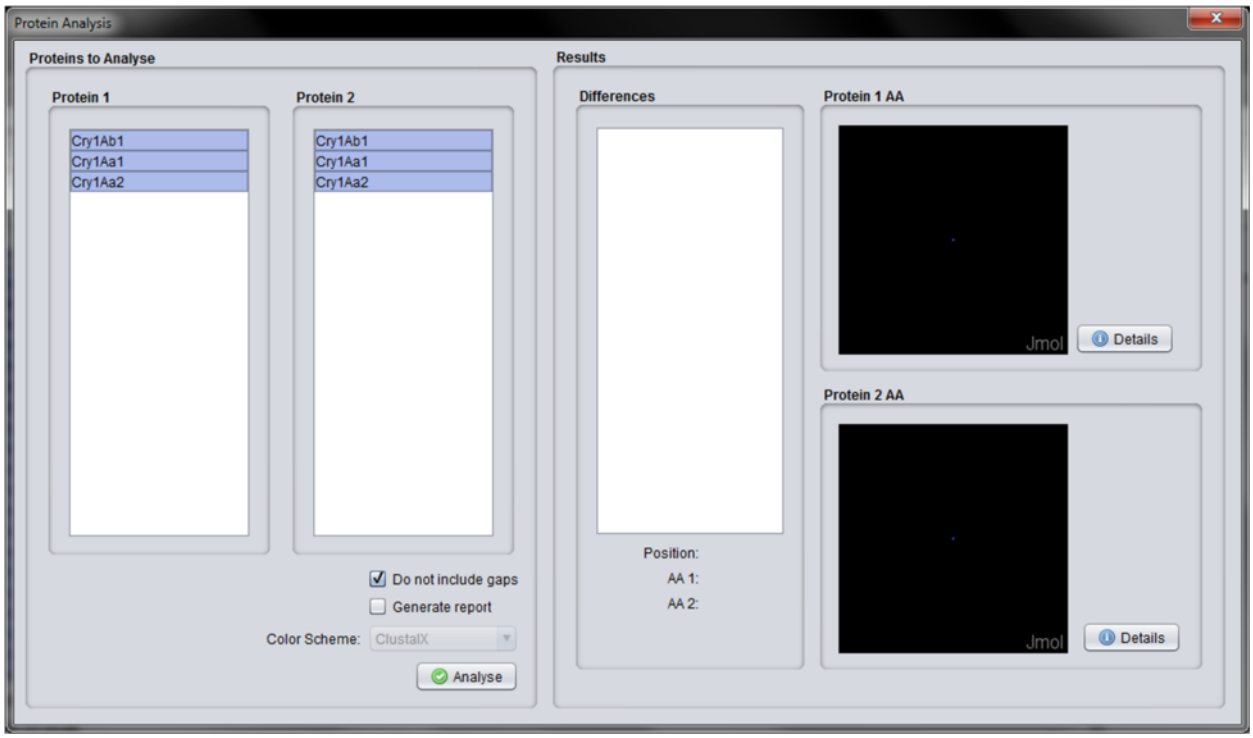

Fig. 11 Protein analysis interface with the result of the alignment of Cry1Aa1, Cry1Aa2 and Cry1Ab1

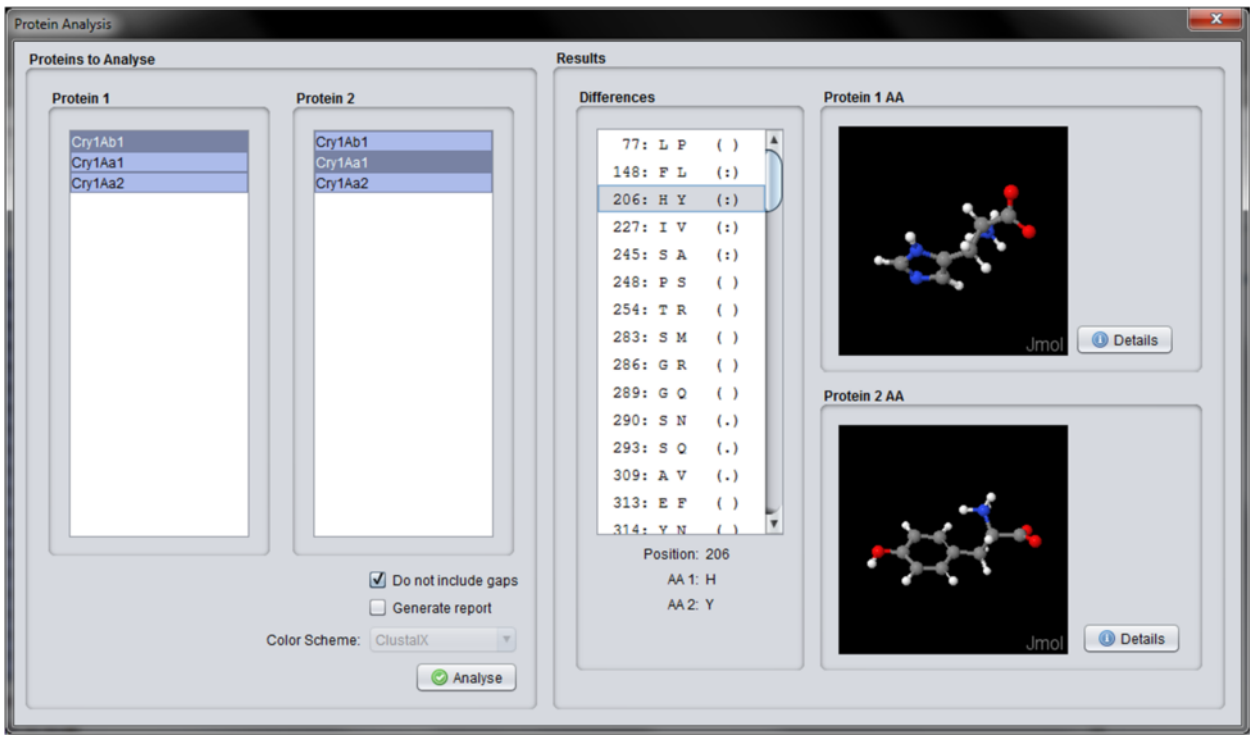

Fig. 12 Comparison between Cry1Ab1 and Cry1Aa1

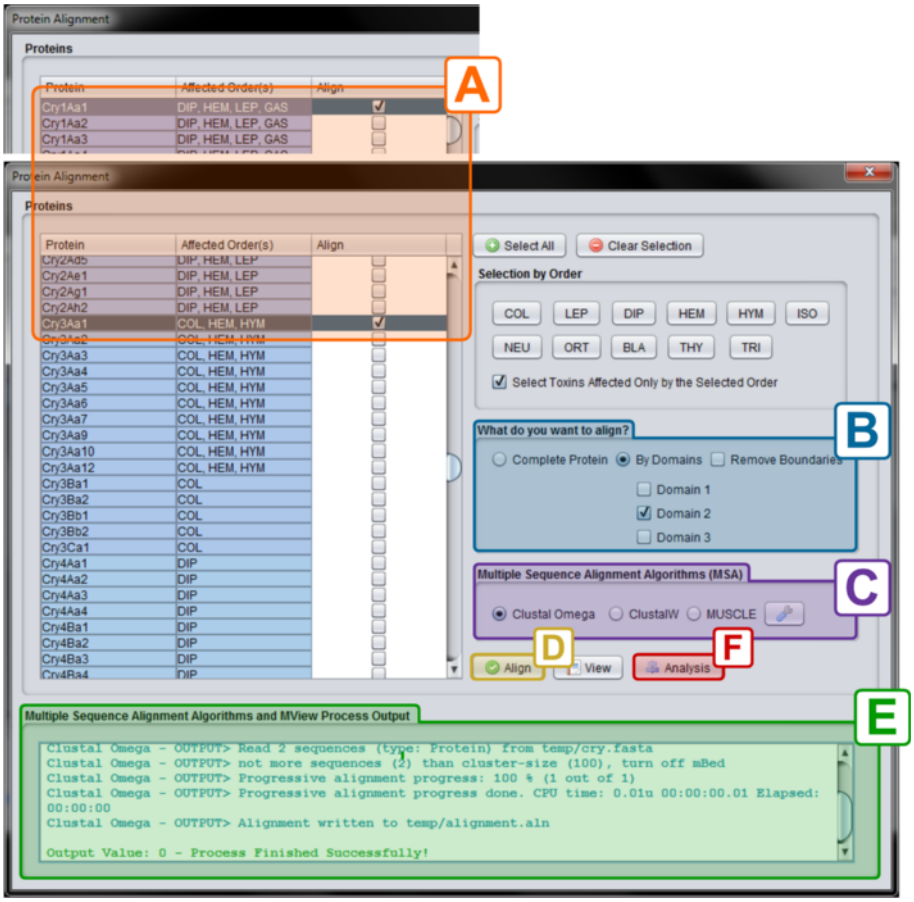

Fig. 13 Preparing the alignment of Cry1Aa1 and Cry3Aa1. **a**) Selection of Cry proteins to align; **b**) Alignment constraints; **c**) Selection of MSA algorithm; **d**) Align button; **e**) MSA algorithms output; **f**) Analysis button

Now, we wanted to analyze the alignment. To do this, we clicked on the button "Analysis" (section F of Fig. 13). After clicking on it, CryGetter required an alignment file. We chose the previously generated alignment file. By doing this, the "Protein Analysis Interface" was shown. In this interface, we chose the two proteins that we wanted to analyze. This selection is shown in section A of Fig. 14. CryGetter was able to infer some results based on the alignment. To do this, we selected the "Generate report" check-box (section B of Fig. 14). Finally, we hit the "Analyse" button (section C of of Fig. 14). A dialog (section D of Fig. 14) was presented, allowing the user to insert some textual observation. In this case, we wrote that these results were related to the alignment of the Domain 2 of both proteins. Clicking on OK, a report was then generated.

This report, created using the JasperReports library [48] and presented in Fig. 15, contained the data of both proteins, presented in section A of Fig. 15. These details comprised the name, accession number, protein id and nucleotide id of the protein, besides the total length of the protein sequence and the size of the three domains. In section B of Fig. 15. a structure diagram of both proteins is presented, showing the disposition of the domains inside the complete sequence, their boundaries and names. In section C of Fig. 15, some statistics are shown:

- C: the amount of conserved residues;
- CM: the amount of conserved mutations;
- SCM: how many semi-conserved mutations;
- C+CM: the sum of C and CM;
- NC: the amount of non conserved residues.

The percentages shown in columns Protein 1 and Protein 2 are related to the amount of each item versus the total number of residues. For example, C is equal to 52, that is 25.74 % of 202 residues of Domain 2 in Protein 1 and 25.37 % of 205 residues of Domain 2 in Protein 2. The total number of residues that are used in these calculations is equal to the size of the sequence that it was aligned to. In this example, we chose to align only Domain 2, that has 202 and 205 residues in Cry1Aa1 and Cry3Aa1 respectively. Finally, in section D of Fig. 15,

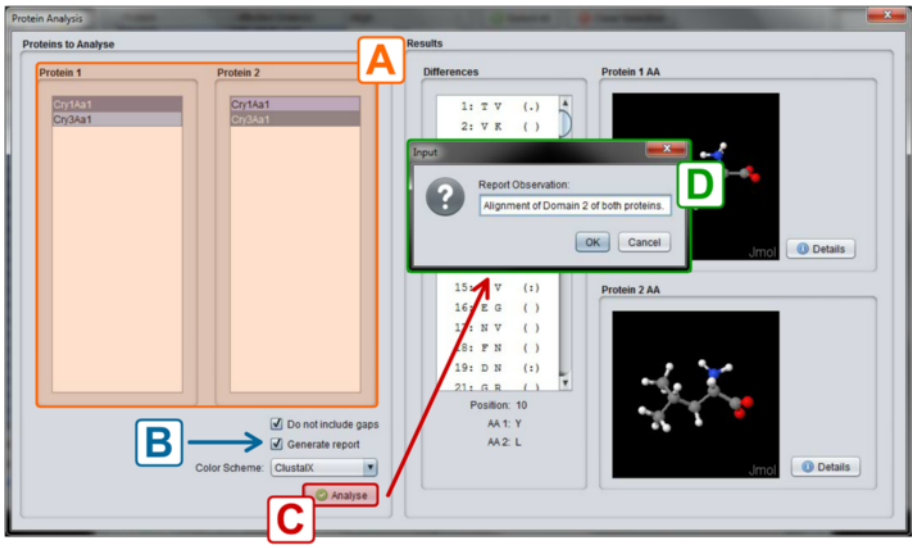

Fig. 14 Analysing the alignment of Domain 2 Cry1Aa1 and Cry3Aa1. **a**) Lists of proteins to analyse; **b**) Analysis options; **c**) Analysis button; **d**) Report Observation dialog

a complete alignment diagram of the two proteins is shown.

The analysis performed by the tool can be used as a starting point to further analyze two Cry proteins. The importance of the tool can be noted, since it simplifies all the process of getting the sequences of interest, aligning them and performing the preliminary inspections in the alignment result.

Conclusions

In this work we presented CryGetter, a tool that aggregates Cry protein data, helping researchers of *Bacillus thuringiensis* and its Crystal proteins to deal with this data and allowing them to get all the relevant information for their work in a faster way compared to a manual protein data collection. Since the tool executes data retrieval and can perform automatic analysis os the protein alignments,

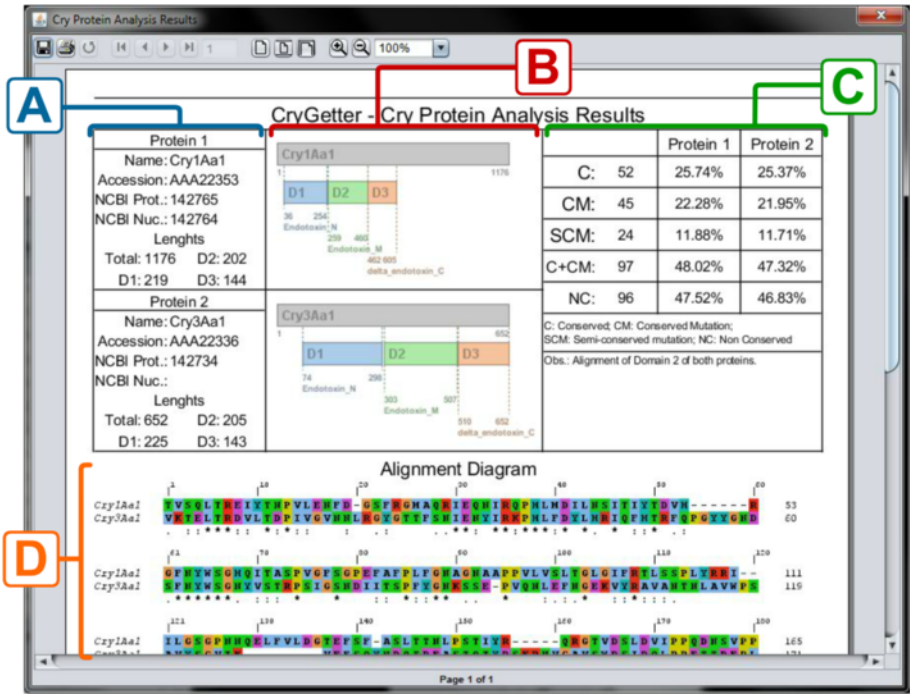

Fig. 15 Analysis report. **a**) Protein data; **b**) Structure diagram; **c**) Report statistics; **d**) Alignment diagram

it allows their users to generate more accurate results, since using it may prevent the error prone task of manually getting all the necessary data and processing them in various software to get the same result that the tool can generate in a unique automatic environment. The development of the tool is also important since these proteins play a significant role in the agro-industry. We hope CryGetter can help the researcher community to improve and accelerate their work with Cry proteins, getting preliminary results faster. As a future work, we will work in the generalization of the tool, allowing the users to extrapolate its functionality related to the data retrieval, enabling them to get data from different online data-sources.

Availability and requirements

Project name: CryGetter
Project home page: http://davidbuzatto.github.io/CryGetter
Operating system(s): Microsoft Windows® 7 or above
Programming language: Java
Other requirements: Java Runtime Environment (JRE) 8 [26] and Perl runtime [49] for MView [27]
License: GNU General Public License v3.0
Any restrictions to use by non-academics: None

Acknowledgements
We thank IFSP (Instituto Federal de Educação, Ciência e Tecnologia de São Paulo) and UNAERP (Universidade de Ribeirão Preto) for partial financial support to this research.

Funding
No funding was obtained for this study.

Authors' contributions
Conceived and designed the study: DB, SCF and SMZ. Performed the study: DB, SCF and SMZ. Implementation: DB. Analyzed and interpreted the data: DB, SCF and SMZ. Wrote the paper and the attached vignettes: DB, SCF and SMZ. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable since our work presents the development of a software tool that deals with publicly available data.

Author details
[1]Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP, Câmpus São João da Boa Vista, Acesso Dr. João Batista Merlin, s/n, Jardim Itália, 13872-551 São João da Boa Vista, SP, Brazil. [2]Universidade de Ribeirão Preto – UNAERP, Av. Costábile Romano, 2201, Ribeirânia, 14096-000 Ribeirão Preto, SP, Brazil.

Received: 6 April 2016 Accepted: 24 August 2016
Published online: 30 August 2016

References
1. Bacillus Thuringiensis Toxin Nomenclature. http://www.btnomenclature.info/. Accessed 23 Mar 2016.
2. Pardo-Lopez L, Gomez I, Rausell C, Sanchez J, Soberon M, Bravo A. Structural changes of the Cry1Ac oligomeric pre-pore from bacillus thuringiensis induced by N-acetylgalactosamine facilitates toxin membrane insertion. Biochemistry. 2006;45(34):10329–36.
3. Siegel JP. The mammalian safety of Bacillus thuringiensis-based insecticides. J Invertebr Pathol. 2001;77(1):13–21.
4. Roh JY, Choi JY, Li MS, Jin BR, Je YH. Bacillus thuringiensis as a specific, safe, and effective tool for insect pest control. J Microbiol Biotechnol. 2007;17(4):547–59.
5. Koch MS, Ward JM, Levine SL, Baum JA, Vicini JL, Hammond BG. The food and environmental safety of Bt crops. Front Plant Sci. 2015;6:283.
6. Masri L, Branca A, Sheppard AE, Papkou A, Laehnemann D, Guenther PS, Prahl S, Saebelfeld M, Hollensteiner J, Liesegang H, Brzuszkiewicz E, Daniel R, Michiels NK, Schulte RD, Kurtz J, Rosenstiel P, Telschow A, Bornberg-Bauer E, Schulenburg H. Host-pathogen coevolution: the selective advantage of bacillus thuringiensis virulence and its cry toxin genes. PLoS Biol. 2015;13(6):1002169.
7. Whiting SA, Strain KE, Campbell LA, Young BG, Lydy MJ. A multi-year field study to evaluate the environmental fate and agronomic effects of insecticide mixtures. Sci Total Environ. 2014;497-498:534–42.
8. van Frankenhuyzen K. Insecticidal activity of Bacillus thuringiensis crystal proteins. J Invertebr Pathol. 2009;101(1):1–16.
9. Crickmore N, Zeigler DR, Feitelson J, Schnepf E, Van Rie J, Lereclus D, Baum J, Dean DH. Revision of the nomenclature for the Bacillus thuringiensis pesticidal crystal proteins. Microbiol Mol Biol Rev. 1998;62(3):807–13.
10. Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH. Bacillus thuringiensis and its pesticidal crystal proteins. Microbiol Mol Biol Rev. 1998;62(3):775–806.
11. de Maagd RA, Bravo A, Crickmore N. How Bacillus thuringiensis has evolved specific toxins to colonize the insect world. Trends Genet. 2001;17(4):193–9.
12. Knowles BH, Dow JAT. The crystal delta-endotoxins of bacillus thuringiensis: models for their mechanism of action on the insect gut. Bioessays. 1993;15(7):469–76.
13. Zhang X, Candas M, Griko NB, Taussig R, Bulla JLA. A mechanism of cell death involving an adenylyl cyclase/PKA signaling pathway is induced by the Cry1Ab toxin of Bacillus thuringiensis. Proc Natl Acad Sci U S A. 2006;103(26):9897–02.
14. Bravo A, Gill SS, Soberon M. Mode of action of Bacillus thuringiensis Cry and Cyt toxins and their potential for insect control. Toxicon. 2007;49(4):423–35.
15. Tiewsiri K, Angsuthanasombat C. Structurally conserved aromaticity of Tyr249 and Phe264 in helix 7 is important for toxicity of the Bacillus thuringiensis Cry4Ba toxin. J Biochem Mol Biol. 2007;40(2):163–71.
16. Reay-Jones FP, Bessin RT, Brewer MJ, Buntin DG, Catchot AL, Cook DR, Flanders KL, Kerns DL, Porter RP, Reisig DD, Stewart SD, Rice ME. Impact of lepidoptera (crambidae, noctuidae, and pyralidae) pests on corn containing pyramided bt traits and a blended refuge in the Southern United States. J Econ Entomol. 2016;109:1859–71.
17. Yang F, Kerns DL, Brown S, Kurtz R, Dennehy T, Braxton B, Head G, Huang F. Performance and cross-crop resistance of Cry1F-maize selected Spodoptera frugiperda on transgenic Bt cotton: implications for resistance management. Sci Rep. 2016;6:28059.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. Mega6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
19. MEGA :: Molecular Evolutionary Genetics Analysis. http://www.megasoftware.net/. Accessed 23 Mar 2016.
20. MacClade Home Page. http://macclade.org/. Accessed 23 Mar 2016.
21. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
22. Geneious - Home Page. http://www.geneious.com/. Accessed 23 Mar 2016.
23. Ripma LA, Simpson MG, Hasenstab-Lehman K. Geneious! simplified genome skimming methods for phylogenetic systematic studies: A case study in oreocarya (boraginaceae). Appl Plant Sci. 2014;2(12):1–12.

24. Ma C, Gunther S, Cooke B, Coppel RL. Geneious plugins for the access of plasmodb and piroplasmadb databases. Parasitol Int. 2013;62(2):134–6.
25. Masters BC, Fan V, Ross HA. Species delimitation–a geneious plugin for the exploration of species boundaries. Mol Ecol Resour. 2011;11(1):154–7.
26. Java.com: Java + You. http://www.java.com/en/. Accessed 23 Mar 2016.
27. MView - MView. http://desmid.github.io/mview/. Accessed 23 Mar 2016.
28. BioJava. http://biojava.org/. . Accessed 23 Mar 2016.
29. Jmol: an Open-source Java Viewer for Chemical Structures in 3D. http://jmol.sourceforge.net/. Accessed 23 Mar 2016.
30. Www.btnomenclature.info. http://www.btnomenclature.info/. Accessed 23 Mar 2016.
31. MLA CE Course Manual: Molecular Biology Information Resources (Entrez). http://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Entrez/. Accessed 23 Mar 2016.
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res. 2000;28(1):235–42.
33. RCSB Protein Data Bank - RCSB PDB. http://www.rcsb.org/. Accessed 23 Mar 2016.
34. Castrignano T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A. The PMDB protein model database. Nucleic Acids Res. 2006;34(Database issue):306–9.
35. PMDB - Protein Model DataBase. https://bioinformatics.cineca.it/PMDB/. Accessed 23 Mar 2016.
36. Jsoup: Java HTML Parser. http://jsoup.org/. Accessed 23 Mar 2016.
37. Simple. http://simple.sourceforge.net/. Accessed 23 Mar 2016.
38. JAXB Reference Implementation. https://jaxb.java.net/. Accessed 23 Mar 2016.
39. Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. J Mol Graphics. 1996;14:33–8.
40. VMD - Visual Molecular Dynamics. http://www.ks.uiuc.edu/Research/vmd/. Accessed 23 Mar 2016.
41. Guex N, Peitsch MC. Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. Electrophoresis. 1997;18(15):2714–3.
42. Swiss PDB Viewer - Home. http://www.expasy.org/spdbv/. Accessed 23 Mar 2016.
43. PyMOL - Www.pymol.org. https://www.pymol.org/. Accessed 23 Mar 2016.
44. Apache POI - the Java API for Microsoft Documents. https://poi.apache.org/. Accessed 23 Mar 2016.
45. Clustal Omega, ClustalW and ClustalX Multiple Sequence Alignment. http://www.clustal.org/. Accessed 23 Mar 2016.
46. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673–680.
47. MUSCLE. http://www.drive5.com/muscle/. Accessed 23 Mar 2016.
48. JasperReports Library. http://community.jaspersoft.com/project/jasperreports-library. Accessed 23 Mar 2016.
49. The Perl Programming Language. https://www.perl.org/. Accessed 23 Mar 2016.
50. How Does Bt Work. http://www.bt.ucsd.edu/how_bt_work.html. Accessed 23 Mar 2016.
51. Knowles BH, Ellar DJ. Colloid-osmotic lysis is a general feature of the mechanism of action of Bacillus thuringiensis d-endotoxins with different insect specificity. Biochimica et Biophysica Acta (BBA) - General Subjects. 1987;924(3):509–18.
52. Grochulski P, Masson L, Borisova S, Pusztai-Carey M, Schwartz JL, Brousseau R, Cygler M. Bacillus thuringiensis CryIA(a) insecticidal toxin: crystal structure and channel formation. J Mol Biol. 1995;254(3):447–64.
53. Kashyap S. Computational Modeling Deduced Three Dimensional Structure of Cry1Ab16 Toxin from Bacillus thuringiensis AC11. Indian J Microbiol. 2012;52(2):263–9.
54. Kashyap S, Singh BD, Amla DV. Computational tridimensional protein modeling of Cry1Ab19 toxin from Bacillus thuringiensis BtX-2. J Microbiol Biotechnol. 2012;22(6):788–92.
55. Derbyshire DJ, Ellar DJ, Li J. Crystallization of the Bacillus thuringiensis toxin Cry1Ac and its complex with the receptor ligand N-acetyl-D-galactosamine. Acta Crystallogr D Biol Crystallogr. 2001;57(Pt 12):1938–44.
56. Dehury B, Sahu M, Sahu J, Sarma K, Sen P, Modi MK, Barooah M, Choudhury MD. Structural analysis and molecular dynamics simulations of novel delta-endotoxin Cry1Id from Bacillus thuringiensis to pave the way for development of novel fusion proteins against insect pests of crops. J Mol Model. 2013;19(12):5301–316.
57. Morse RJ, Yamamoto T, Stroud RM. Structure of Cry2Aa suggests an unexpected receptor binding epitope. Structure. 2001;9(5):409–17.
58. Li JD, Carroll J, Ellar DJ. Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 A resolution. Nature. 1991;353(6347):815–21.
59. Sawaya MR, Cascio D, Gingery M, Rodriguez J, Goldschmidt L, Colletier JP, Messerschmidt MM, Boutet S, Koglin JE, Williams GJ, Brewster AS, Nass K, Hattne J, Botha S, Doak RB, Shoeman RL, DePonte DP, Park HW, Federici BA, Sauter NK, Schlichting I, Eisenberg DS. Protein crystal structure obtained at 2.9 A resolution from injecting bacterial cells into an X-ray free-electron laser beam. Proc Natl Acad Sci U S A. 2014;111(35):12769–12774.
60. Galitsky N, Cody V, Wojtczak A, Ghosh D, Luft JR, Pangborn W, English L. Structure of the insecticidal bacterial delta-endotoxin Cry3Bb1 of Bacillus thuringiensis. Acta Crystallogr D Biol Crystallogr. 2001;57(Pt 8):1101–9.
61. Boonserm P, Angsuthanasombat C, Lescar J. Crystallization and preliminary crystallographic study of the functional form of the Bacillus thuringiensis mosquito-larvicidal Cry4Aa mutant toxin. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 7):1315–8.
62. Boonserm P, Mo M, Angsuthanasombat C, Lescar J. Structure of the functional form of the mosquito larvicidal Cry4Aa toxin from Bacillus thuringiensis at a 2.8-angstrom resolution. J Bacteriol. 2006;188(9):3391–401.
63. Boonserm P, Davis P, Ellar DJ, Li J. Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications. J Mol Biol. 2005;348(2):363–82.
64. Xin-Min Z, Li-Qiu X, Xue-Zhi D, Fa-Xiang W. The theoretical three-dimensional structure of Bacillus thuringiensis Cry5Aa and its biological implications. Protein J. 2009;28(2):104–10.
65. Hui F, Scheib U, Hu Y, Sommer RJ, Aroian RV, Ghosh P. Structure and glycolipid binding properties of the nematicidal protein Cry5B. Biochemistry. 2012;51(49):9911–21.
66. Xia LQ, Zhao XM, Ding XZ, Wang FX, Sun YJ. The theoretical 3D structure of Bacillus thuringiensis Cry5Ba. J Mol Model. 2008;14(9):843–8.
67. Guo S, Ye S, Liu Y, Wei L, Xue J, Wu H, Song F, Zhang J, Wu X, Huang D, Rao Z. Crystal structure of Bacillus thuringiensis Cry8Ea1: An insecticidal toxin toxic to underground pests, the larvae of Holotrichia parallela. J Struct Biol. 2009;168(2):259–66.
68. Gutierrez P, Alzate O, Orduz S. A theoretical model of the tridimensional structure of Bacillus thuringiensis subsp. medellin Cry 11Bb toxin deduced by homology modelling. Mem Inst Oswaldo Cruz. 2001;96(3):357–64.
69. Zhao XM, Zhou PD, Xia LQ. Homology modeling of mosquitocidal Cry30Ca2 of Bacillus thuringiensis and its molecular docking with N-acetylgalactosamine. Biomed Environ Sci. 2012;25(5):590–6.
70. Kelker MS, Berry C, Evans SL, Pai R, McCaskill DG, Wang NX, Russell JC, Baker MD, Yang C, Pflugrath JW, Wade M, Wess TJ, Narva KE. Structural and biophysical characterization of Bacillus thuringiensis insecticidal proteins Cry34Ab1 and Cry35Ab1. PLoS ONE. 2014;9(11):112555.
71. Xu C, Chinte U, Chen L, Yao Q, Meng Y, Zhou D, Bi LJ, Rose J, Adang MJ, Wang BC, Yu Z, Sun M. Crystal structure of Cry51Aa1: A potential novel insecticidal aerolysin-type β-pore-forming toxin from Bacillus thuringiensis. Biochem Biophys Res Commun. 2015;462(3):184–9.