Journal of Invertebrate Pathology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

journal homepage: www.elsevier.com/locate/jip

Structural classification of insecticidal proteins – Towards an *in silico* characterisation of novel toxins

Colin Berry [a,*], Neil Crickmore [b]

[a] Cardiff School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3AT, UK
[b] School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

ARTICLE INFO

Article history:
Received 26 February 2016
Revised 1 June 2016
Accepted 28 July 2016
Available online xxxx

Keywords:
Bacillus thuringiensis
Lysinibacillus sphaericus
Cry toxin

ABSTRACT

The increasing rate of discovery of new toxins with potential for the control of invertebrate pests through next generation sequencing, presents challenges for the identification of the best candidates for further development. A consideration of structural similarities between the different toxins suggest that they may be functionally less diverse than their low sequence similarities might predict. This is encouraging from the prospective of being able to use computational tools to predict toxin targets from their sequences, however more structure/function data are still required to reliably inform such predictions.

1. Introduction

The insecticidal toxins of *Bacillus thuringiensis, Lysinibacillus sphaericus, Photorhabdus* spp. and other bacteria represent a rich resource for the control of pest insects. The increasing rate of discovery of new toxins, driven by next-generation sequencing, will expand our arsenal of potential biocontrol agents but this, in itself, presents new challenges. Even with past rates of toxin discovery, toxins have rarely been tested against more than a few species of insects (van Frankenhuyzen, 2009) and, in the future, toxicity testing of large numbers of new toxins against a wide range of insects will not be feasible. To facilitate the selection of toxins for study, different criteria may be applied, including identification of the toxin in a strain known (from a previous screening) to have interesting biocidal activity or relatedness to known toxins. Here we consider the prospects for a further, selective method through the prediction of activity. We highlight some of the challenges that may be encountered and propose steps that will bring us closer to this goal. Useful predictions would not only assist in the selection of toxins for development but would also have value in support of the regulatory process of biopesticide product registration, where the potential to predict off-target activities would be valuable.

The *B. thuringiensis* nomenclature system (Crickmore et al., 1998) currently contains several hundred individual sequences, divided between 74 classes of Cry toxin, 3 classes of Cyt toxin, 4 classes of Vip toxin and one SIP toxin. *L. sphaericus* strains may produce the BinA/B toxin, Mtx1, Mtx2, Mtx3, Mtx4, sphaericolysin, Cry48 and Cry49 (reviewed in (Berry, 2012)) and *Photorhabdus* strains can produce Tc toxins, PirA/B and Mcf toxins (ffrench-Constant et al., 2007). This represents a great diversity of toxins but some simplification can be achieved by considering these proteins in terms of their structural characteristics (known or predicted). Table 1 shows the toxin classes, colour-coded by sequence homology groups. As can be seen, the 3-domain Cry toxins represent the largest structural family (and also encompass the PirA/B toxin, recently shown to be equivalent to a 3-domain toxin with a dissociated domain III (Lee et al., 2015)). There is also a large group of toxins that is rich in beta-sheets with general structural similarity to aerolysin. This group includes Cry46 and toxins identified by Pfam (Bateman et al., 1999) to be members of either the Etx/Mtx2 family or the Toxin_10 family. Other groups include the Cyt toxins, the ADP-ribosyl transferase toxins Mtx1 and Vip1/2 (along with the Vip1-like Vip4 protein). Cry34 is an aegerolysin like protein and with Cry35 is part of a two-component toxin (Kelker et al., 2014). Cry37, which itself is part of a two-component toxin with Cry23, which shows structural homology with Cry34 (Rydel et al., 2001). Other toxins, which appear unrelated and have no published structures, are Cry6, Cry22, Cry55, Vip3 and Mcf. Our knowledge of the structure and function of toxins within these groups varies and it will be useful to consider the major groups separately.

* Corresponding author.
E-mail address: Berry@cf.ac.uk (C. Berry).

http://dx.doi.org/10.1016/j.jip.2016.07.015
0022-2011/

Table 1
Toxins and their homology groups.

Cry1	Cry21	Cry41	Cry61	Vip3
Cry2	Cry22	Cry42	Cry62	Vip4
Cry3	Cry23	Cry43	Cry63	BinA
Cry4	Cry24	Cry44	Cry64	BinB
Cry5	Cry25	Cry45	Cry65	Mtx1
Cry6	Cry26	Cry46	Cry66	Mtx2
Cry7	Cry27	Cry47	Cry67	Mtx3
Cry8	Cry28	Cry48	Cry68	Mtx4
Cry9	Cry29	Cry49	Cry69	Sphaericolysin
Cry10	Cry30	Cry50	Cry70	PirA
Cry11	Cry31	Cry51	Cry71	PirB
Cry12	Cry32	Cry52	Cry72	Mcf
Cry13	Cry33	Cry53	Cry73	
Cry14	Cry34	Cry54	Cry74	
Cry15	Cry35	Cry55	Cyt1	
Cry16	Cry36	Cry56	Cyt2	
Cry17	Cry37	Cry57	Cyt3	
Cry18	Cry38	Cry58	Sip	
Cry19	Cry39	Cry59	Vip1	
Cry20	Cry40	Cry60	Vip2	

Toxins of *B. thuringiensis*, *L. sphaericus* and *Photorhabdus* spp. are shown with colouring to indicate homology groups: light blue = 3-domain toxins; peach = Etx/Mtx2 toxins; pink = Toxin_10 family proteins; violet = Cyt toxins; khaki = aegerolysin toxins; grey = ADP ribosyl transferase-related proteins; toxins not falling into these groups are coloured differently. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. The 3-domain toxins

These toxins are the best-characterised, with the first structure published in 1991 (Li et al., 1991) and with several decades of studies on the specificity and mode of action of members of this family. The steps leading to toxicity for this family are well-known and involve ingestion by the invertebrate target, solubilisation of toxin crystals in the gut, proteolytic activation by gut enzymes, one or more receptor binding step, followed by membrane insertion (Pardo-Lopez et al., 2012). Insect specificity could be mediated by any of the above steps, for example changes in proteinase activity (Loseva et al., 2002) but the most important determinants of specificity are the binding to and specificity for receptors on the surfaces of target cells.

As suggested by the name of this family, the structure of the active toxin is composed of 3 distinct structural domains. Domain I is formed from a bundle of alpha helices and is involved in pore formation by the toxin. Domain II has a beta prism structure that appears to be related to carbohydrate binding proteins and Domain III has a beta sandwich fold. Domains II and III appear to have roles in receptor binding and specificity of the toxins as demonstrated by domain swapping experiments that have altered target specificity (Lee et al., 1995; Pigott and Ellar, 2007). Bioinformatic analysis suggests that the 3 toxin domains evolve at different rates (Bravo, 1997) and this may have implications for target specificity.

Within the 3-domain toxin family, we find toxins with activity against insects in several orders, principally amongst the Lepidoptera with fewer active against the orders Diptera and Coleoptera, and with small numbers active against Hymenoptera and Hemiptera as well as toxins affecting nematodes and gastropods (reviewed recently (Palma et al., 2014a)). Members of this family active against human cancer cells have also been reported (Ohba et al., 2009), although it is clearly unlikely that they have co-evolved with this host. However, correlation between sequence identity and target range is generally poor even when analysis is carried out at the level of the individual domains (de Maagd et al., 2001). This highlights the need for analysis at a level below that of the domains themselves. Within domain II, several exposed loops (the $\propto$8 loop, and loops 1, 2 and 3) have been identified as potentially important for receptor binding. The variability of these

Please cite this article in press as: Berry, C., Crickmore, N. Structural classification of insecticidal proteins – Towards an *in silico* characterisation of novel toxins. J. Invertebr. Pathol. (2016), http://dx.doi.org/10.1016/j.jip.2016.07.015