



## Research article

## Genome-wide identification and domain organization of lectin domains in cucumber



Liuyi Dang, Els J.M. Van Damme\*

Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

## ARTICLE INFO

## Article history:

Received 7 May 2016

Received in revised form

4 July 2016

Accepted 9 July 2016

Available online 10 July 2016

## Keywords:

Lectin domain

Cucumber

Domain architecture

Evolution

Biological function

## ABSTRACT

Lectins are ubiquitous proteins in plants and play important roles in a diverse set of biological processes, such as plant defense and cell signaling. Despite the availability of the *Cucumis sativus* L. genome sequence since 2009, little is known with respect to the occurrence of lectins in cucumber. In this study, a total of 146 putative lectin genes belonging to 10 different lectin families were identified and localized in the cucumber genome. Domain architecture analysis revealed that most of these lectin gene sequences contain multiple domains, where lectin domains are linked with other domains, as such creating chimeric lectin sequences encoding proteins with dual activities. This study provides an overview of lectin motifs in cucumber and will help to understand their potential biological role(s).

© 2016 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Most proteins consist of evolutionary units referred as domains. The concept of a 'domain' was first proposed in the 1960s when spatially distinct structural units were identified in the enzyme structures of lysozyme and ribonuclease (Blake et al., 1965; Kartha et al., 1967). Over the last few decades the concept of domains has been used to refer to many aspects of biology although the descriptions vary in different fields. In biochemistry, a domain is usually described as a protein region with assigned experimental function(s). From the viewpoint of protein structure, domains are considered as compact, spatially distinct units. When it comes to sequence comparison, domains are designed in an evolutionary perspective as conserved sequence segments of proteins (Ponting and Russell, 2002). In general, domains can be described as evolutionary conserved segments of proteins with generally independent structural and functional properties (Marsh and Teichmann, 2010).

The study of protein domain evolution has been greatly facilitated by the availability of numerous protein sequences. Nowadays,

domains are considered as modular building units of proteins. Although domains may occur alone in single-domain proteins, the majority of proteins consist of at least two domains (Vogel et al., 2004). In fact, it has been shown that the number of proteins with multi-domain architectures is still increasing rapidly while the number of single domain proteins is saturating (Levitt, 2009). Obviously, not only the creation of new domains but also the rearrangement of existing domains resulted in numerous proteins with diverse functions. During evolution, proteins with new functions or specificities have been created by domain rearrangements or through the differentiation of existing domains (Bjorklund et al., 2005). This is particularly important for plants, since they have to adapt to various unfavorable circumstances which they cannot escape unlike many other species (Kersting et al., 2012). Due to their critical role in protein evolution, the distribution of protein domains and their organization often provides vital clues for the evolutionary relations between different protein families.

Lectins, also known as glycan-binding proteins, are a class of proteins that play important roles in diverse biological processes (Peumans and Van Damme, 1995). The analyses of lectin sequences revealed the existence of several carbohydrate-recognition domains, often referred as lectin domains, which bind selectively and reversibly to specific mono- or oligosaccharides (Sharon and Lis, 2004). It is generally accepted that lectin domains are widespread in all kingdoms. Unfortunately, the name of a protein often reflects

\* Corresponding author.

E-mail addresses: [Liuyi.Dang@UGent.be](mailto:Liuyi.Dang@UGent.be) (L. Dang), [Elsjm.VanDamme@UGent.be](mailto:Elsjm.VanDamme@UGent.be) (E.J.M. Van Damme).

only its main biological function but does not necessarily indicate the presence of lectin domains (Taylor and Drickamer, 2014). For instance, many plant pathogen recognition receptors actually carry lectin domains to recognize the carbohydrate structures from microbial organisms. Consequently, the importance of protein-carbohydrate interactions and the occurrence of lectin motifs can be underestimated. Therefore, it is fascinating to study the distribution of lectin domains in genome data.

In plants, lectin domains are ubiquitously distributed and play crucial roles in plant defense (Lannoo and Van Damme, 2015). Tandem and segmental duplication have been considered as major factors that drive lectin expansion in plants. They may be regarded as drivers to help plants to adapt to various environmental stresses through duplication followed by expression divergence (Jiang et al., 2010). Despite the huge amount of research on lectins and lectin domains in several plant species, little information is available on lectins in cucumber. The species *Cucumis* represents an economically important and widely cultivated crop. In this study, lectin domains from cucumber were identified and their domain architectures analyzed. Finally, possible functions of these lectin domains and their importance for plant growth and development are discussed.

## 2. Results

### 2.1. Identification and distribution of lectin genes in the cucumber genome

To identify the sequences with lectin domains in cucumber, BLAST searches were performed against the complete *Cucumis sativus* genome sequence available from the Cucurbit Genomics Database. Hidden Markov models and multiple sequence alignments were applied to validate the presence of lectin domains in all the collected sequences. In total, 146 putative lectin genes were identified from the cucumber genome, representing homologs belonging to 10 different lectin families (Table 1). Proteins with lectin domains homologous to the *Agaricus bisporus* agglutinin and the cyanovirin family were not found in the cucumber genome database. All lectin sequence information is available in Table S1.

All genes containing at least one lectin domain were mapped on the chromosomes according to the cucumber genome version 2 in Cucurbit Genomics Database (Fig. 1). In general, the lectin genes are spread all over the seven cucumber chromosomes. The large chromosomes such as chromosomes 1, 3 and 6 contain a high number of lectin genes while the small chromosomes 2 and 7 accommodate less lectin genes. Furthermore, chromosomes 1, 3 and 6 also have a high lectin gene density (0.89–1.27 gene/Mbp) compared to e.g. chromosome 5 which shows the lowest density (0.25 gene/Mbp).

**Table 1**  
Predicted lectin genes and their distribution on chromosomes.

Lectin family	Genes	Chromosome
ABA domain	0	/
Amaranthin domain	16	6
CRA domain	4	1, 2
Cyanovirin domain	0	/
EUL domain	1	4
GNA domain	45	All except chr 7
Hevein domain	4	1, 6
Jacalin domain	8	2, 3
Legume domain	29	All except chr 4
LysM domain	10	2, 3, 5, 6, 7
Nictaba domain	20	1, 2, 3, 4, 7
Ricin B domain	9	3, 5, 6

### 2.2. Structural analysis of lectin genes of cucumber

Sequence analyses of the putative lectin genes retrieved from the cucumber genome revealed that most lectin domains are linked to other protein domains. The combination of different domains can give rise to multi-domain proteins with multiple functions. Therefore, the domain architecture of these sequences was analyzed to illustrate the potential role of the lectin genes. All CRA and EUL homologs from cucumber are characterized by a single domain architecture. These sequences contain a single CRA or EUL domain and therefore will not be discussed in more detail (Van Holle and Van Damme, 2015). In addition to protein domains also the presence of a signal peptide and/or a transmembrane region was analyzed (Table 2). Schematic overviews of the lectin domain architecture for the lectin genes within each family are drawn to scale.

#### 2.2.1. Amaranthin-like lectins

The amaranthin domain refers to the GalNAc-specific seed lectins from *Amaranthus* species. However, sequences with amaranthin domains have also been identified in other plant species, such as wheat and flax (Faruque et al., 2015; Puthoff et al., 2005a,b). Searches within the cucumber genome yielded 16 amaranthin homologs, all of which are located on chromosome 6. All amaranthin homologs in cucumber possess a similar domain architecture with two tandem arrayed amaranthin domains linked to one aerolysin domain, a pore-forming toxin that was mainly reported in bacteria (Fig. 2a). The two amaranthin domains correspond to one subunit of the dimeric protein amaranthin and are likely the result of a gene duplication (Transue et al., 1997). Aerolysin is known as a cytolytic toxin that binds to eukaryotic cells and aggregates to form holes in the cell surface. Therefore, the presence of this aerolysin domain might lead to an increase of cell membrane permeability, resulting in cellular lysis. The absence of a signal peptide or a transmembrane region suggests that the amaranthin-like proteins are soluble proteins that reside in the cytoplasm of the cell.

#### 2.2.2. Jacalin-related lectins

The family of jacalin-related lectins is widely distributed in plants and groups all proteins with one or more domains similar to the *Artocarpus integer* lectin, also referred to as jacalin. Searches in the cucumber genome yielded 8 genes containing jacalin domains with various domain architectures (Fig. 2b), including sequences with one jacalin domain (1 gene), two jacalin domains (1 gene) and three jacalin domains (3 genes). Three jacalin-related sequences with a single lectin domain were linked at the N-terminus with DC1 and C1-related domains. None of these sequences contains a signal peptide or a transmembrane sequence. Both DC1 and C1-related domains are similar to the C1 domain (also known as the phorbol ester/diacylglycerol binding domain), which corresponds to the N-terminal region of protein kinase C. Jacalin domains have been reported in wheat, rice, *Arabidopsis thaliana*, *Brachypodium distachyon*, maize, sorghum and soybean (Song et al., 2014). Although different protein domains have been identified in jacalin-related lectin sequences, this is the first time that a chimeric protein is composed of jacalin domain linked to multiple C1-like domains.

#### 2.2.3. Nictaba homologs

The family of Nictaba-related lectins was named after the *Nicotiana tabacum* agglutinin (Nictaba), a 19 kDa lectin first discovered in tobacco leaves (Chen et al., 2002). Screening of Nictaba domains in different plant genomes demonstrated that Nictaba orthologs are widely distributed in the plant kingdom but only a few orthologs consist of single Nictaba domains. In most sequences, the Nictaba domains are fused to unrelated protein domains (Delporte et al.,

Download English Version:

<https://daneshyari.com/en/article/2014704>

Download Persian Version:

<https://daneshyari.com/article/2014704>

[Daneshyari.com](https://daneshyari.com)