



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ENGENHARIA DE PESCA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA DE RECURSOS
NATURAIS**

JOSÉ CAETANO DA SILVA FILHO

**BASES MOLECULARES DO PROCESSAMENTO PÓS-TRADUCIONAL NA
ESTRUTURA TRIDIMENSIONAL DE UMA LECTINA DE SEMENTES DE *Dioclea
grandiflora***

FORTALEZA

2017

JOSÉ CAETANO DA SILVA FILHO

BASES MOLECULARES DO PROCESSAMENTO PÓS-TRADUCIONAL NA
ESTRUTURA TRIDIMENSIONAL DE UMA LECTINA DE SEMENTES DE *Dioclea*
grandiflora

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia de Recursos Naturais da Universidade Federal do Ceará como pré-requisito para obtenção do título de Doutor em Biotecnologia de Recursos Naturais. Área de concentração: Biotecnologia de Recursos Naturais.

Orientador: Prof. Dr. Bruno Anderson Matias da Rocha

FORTALEZA

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S58b Silva Filho, José Caetano da.
Bases Moleculares do Processamento Pós-Traducional na Estrutura Tridimensional de uma Lectina de Sementes de *Dioclea grandiflora* / José Caetano da Silva Filho. – 2017.
91 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Biotecnologia de Recursos Naturais, Fortaleza, 2017.
Orientação: Prof. Dr. Bruno Anderson Matias da Rocha.
1. Lectina. 2. Precursor. 3. Manose. 4. Cristalografia. I. Título.

CDD 660.6

JOSÉ CAETANO DA SILVA FILHO

BASES MOLECULARES DO PROCESSAMENTO PÓS-TRADUCIONAL NA
ESTRUTURA TRIDIMENSIONAL DE UMA LECTINA DE SEMENTES DE *Dioclea*
grandiflora

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia de Recursos Naturais da Universidade Federal do Ceará como pré-requisito para obtenção do título de Doutor em Biotecnologia de Recursos Naturais. Área de concentração: Biotecnologia de Recursos Naturais.

Orientador: Prof. Dr. Bruno Anderson Matias da Rocha

Aprovada em: ____ / ____ / ____

BANCA EXAMINADORA

Prof. Dr. Bruno Anderson Matias da Rocha (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Alexandre Holanda Sampaio
Universidade Federal do Ceará (UFC)

Prof. Dr. Celso Shiniti Nagano
Universidade Federal do Ceará (UFC)

Prof. Dr. Claudener Souza Teixeira
Universidade Federal do Maranhão (UFMA)

Prof. Dr. Bruno Lopes de Sousa
Universidade Estadual do Ceará (UECE)

AGRADECIMENTOS

- À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa disponibilizada, permitindo que me dedicasse integralmente ao Programa;
- À Universidade Federal do Ceará (UFPB) e ao Programa de Pós-Graduação em Biotecnologia de Recursos Naturais, pela oportunidade concedida de aprofundar e adquirir novos conhecimentos;
- Ao Prof. Dr. Plínio Delatorre, que me orientou desde o início de minha graduação e que tem me ensinado de forma direta e indireta a me apaixonar mais ainda pelo estudo estrutural de proteínas;
- Ao prof. Dr. Bruno Anderson Matias da Rocha que, de última hora, aceitou me orientar nos momentos finais desse trabalho. Agradeço por suas sugestões e conhecimentos que, direta e indiretamente, me fizeram crescer academicamente;
- Ao prof. Dr. Benildo Sousa Cavada que em muitos momentos me ajudou grandemente durante o decorrer do curso. Agradeço por ter me recebido em seu grupo de pesquisa desde o primeiro dia no programa e me dado condições de me aprofundar nos estudos estruturais de lectinas;
- ao prof. Dr. Bruno Lopes de Sousa que se tornou não apenas um companheiro na Academia, compartilhando seus conhecimentos, mas um grande amigo que me engrandeceu como pessoa. Muito obrigado, boy;
- Aos amigos Monna Ramalho e Gustavo Nobre que me acolheram em Fortaleza, fazendo com que eu me sentisse em casa. Muito obrigado pela força e apoio e pelos momentos de descontração;
- À minha esposa, Adlane Florentino, companheira e amiga incondicional que tem estado ao meu lado desde minha entrada na graduação, dando força e estimulando minha vida acadêmica. A ela todo meu agradecimento;
- À minha “sogra-avó”, D. Maria da Luz, que muito me apoiou para que eu me mudasse para Fortaleza e pudesse desenvolver esse trabalho. Sem seu apoio tudo teria ficado mais difícil. Muito obrigado, moça;
- Aos meus pais, José Caetano e Maria Salomé, por tudo o que fizeram por mim durante todos estes anos e principalmente por tudo o que passaram para que eu pudesse estudar.

RESUMO

Lectinas, em especial aquelas de origem vegetal, são amplamente estudadas e caracterizadas por apresentarem atividades antidepressivas, antitumorais e anti- e pró-inflamatórias. Todas essas atividades levam em consideração a forma madura da lectina, a qual, nas representantes da subtribo *Diocleinae*, é sintetizada como uma glicoproteína precursora inativa que sofre modificações pós-traducionais que incluem deglicosilação, clivagem proteolítica e inversão da cadeia polipeptídica. A deglicosilação torna o precursor capaz de se ligar a carboidratos. O precursor deglicosilado da lectina de *Dioclea grandiflora* (pDGL) não só se liga a monossacarídeos, apresentando atividade hemaglutinante, como tem maior afinidade por manose em relação à sua contra-parte madura. Apesar do processamento pós-traducional ser bem conhecido, as bases moleculares desse mecanismo na estrutura tridimensional de lectinas bem como sua influência sobre a afinidade a carboidratos ainda não estão muito claras. Assim, o presente trabalho teve como objetivo identificar as características estruturais de uma lectina não processada e os fatores que influenciam sua interação com carboidratos, comparando-os com sua contra-parte madura. Para isso, a pDGL recombinante foi co-cristalizada com manose e os cristais obtidos foram difratados e seus dados resultantes tratados computacionalmente. Análises *in silico* foram realizadas para se comparar as características estruturais da pDGL e de sua forma madura, bem como seus modos de interação com a manose e o trimanosídeo 3,6-di-O-(α -D-manopiranosil)- α -D-manopiranosose. A estrutura cristalográfica da pDGL apresentou-se como um tetrâmero, mantendo todas as características estruturais de sua contra-parte madura, incluindo a presença de estruturas secundárias não-repetitivas que moldam o domínio de reconhecimento a carboidrato (CRD). Entretanto, diferente desta, a pDGL parece não ser capaz de formar arranjos tetraméricos estáveis, já que faltam duas importantes interações interdiméricas envolvendo os resíduos His22 e His194. A manose se ligou nas quatro cadeias do modelo cristalográfico sendo estabilizada por diversas interações. As análises *in silico* mostraram que a pDGL, em relação à sua forma madura, se liga com maior afinidade à manose e ao trimanosídeo devido seu CRD possuir menor profundidade. Em conjunto, esses dados sugerem que o processamento pós-traducional permite a formação de tetrâmeros estáveis e gera sutis mudanças na disposição dos resíduos do CRD que fazem a proteína madura ter menor afinidade a carboidratos.

Palavras-chave: Lectina. Precursor. Manose. Cristalografia.

ABSTRACT

Lectins, specially the ones from plants, are widely studied and characterized by present great biotechnological potential such as antidepressant, antitumoral and anti- and pro-inflammatory activities. These properties take into account the mature form of the lectin, which is synthesized as an inactive glycoprotein precursor that is post-translationally modified by deglycosylation, proteolytic breakdown and polypeptide chain inversion. Deglycosylation makes the precursor able to bind carbohydrates. The deglycosylated precursor from *Dioclea grandiflora* lectin (pDGL) do not solely binds to monosaccharides, presenting hemagglutinating activity, but also presents higher affinity to mannose in relation to its mature counterpart. Although the post-translational processing is well understood, the structural basis for this different affinity pattern is less clear. In this way, the present work aimed to elucidate the crystallographic structure of pDGL in complex to mannose to identify the structural characteristics that determine its greater affinity to this hexose. Recombinant pDGL was co-crystallized with mannose and crystals were submitted to X-ray diffraction and the resulting data were computationally analyzed. *In silico* studies were carried out to compare the interaction pattern of mannose and more complex carbohydrates between pDGL and their mature form. The crystallographic structure of pDGL presented as a tetramer, conserving all structural characteristics of its mature counterpart, including the presence of non-repetitive secondary structures that fold the Carbohydrate Recognition Domain (CRD). However, pDGL seems unable to form stable tetramers, once it lacks two important interdimeric contacts orchestrated by histidine residues His22 and His194. Mannose bound at all four polypeptide chains being stabilized by several interactions, including a CH— π ligation. *In silico* analyses showed that pDGL, in relation to its mature form, binds with higher affinity to mannose and the trimannoside 3,6-di-O-(α -D-manopyranosyl)- α -D-mannopyranose due its CRD presents lower depthness. Together, these results show that post-translational processing (i) allows the formation of stable tetramers and (ii) produces subtle changes in the disposal of CRD residues that decreases the affinity of the mature protein for carbohydrate molecules.

Keywords: Lectin. Precursor. Mannose. Crystallography.

SUMÁRIO

1	INTRODUÇÃO.....	10
2	FUNDAMENTAÇÃO TEÓRICA.....	12
2.1	Lectinas.....	12
2.1.1	<i>Lectinas vegetais</i>	13
2.1.1.1	<i>Lectinas de leguminosas</i>	14
2.1.1.1.1	DGL, uma lectina isolada de sementes da <i>Dioclea grandiflora</i>	16
2.1.2	<i>Estrutura de lectinas da subtribo Diocleinae</i>	17
2.1.3	<i>Domínio de Reconhecimento a Carboidrato (CRD)</i>	19
2.2	Maturação de lectinas ConA-like ocorre por permutação circular.....	23
3	OBJETIVOS.....	27
3.1	Objetivo geral.....	27
3.2	Objetivos específicos.....	27
4	MATERIAL E MÉTODOS.....	28
4.1	Obtenção da forma recombinante da pDGL.....	28
4.2	Ensaio de cristalização.....	29
4.3	Coleta de dados e processamento inicial.....	30
4.4	Elucidação das estruturas tridimensionais.....	30
4.4.1	<i>Substituição molecular</i>	30
4.4.2	<i>Refinamento do modelo</i>	31
4.5	Análises <i>in silico</i>	31
4.5.1	<i>Análise do padrão de oligomerização</i>	31
4.5.2	<i>Cálculos de ângulos phi (Φ) e psi (Ψ)</i>	32
4.5.3	<i>Simulações de docking molecular e cálculo de dimensões do CRD</i>	32
5	RESULTADOS.....	34
5.1	Cristalização e Difração de Raios X.....	34
5.2	Estrutura Geral da pDGL.....	34
5.2.1	<i>Tetrâmero da pDGL</i>	43
5.3	Ligação da manose à pDGL.....	46
5.4	Interação da pDGL com trimanosídeo e da DGL com manose.....	46

6	DISCUSSÃO.....	53
6.1	A pDGL é estruturalmente semelhante à sua contra-parte madura.....	53
6.2	A pDGL parece não formar tetrâmeros estáveis.....	57
6.3	Interação CH- - π estabiliza a manose no CRD da pDGL.....	59
6.4	A pDGL tem maior afinidade para manose e trimanosídeo do que a DGL.....	61
7	CONCLUSÕES E PERSPECTIVAS.....	63
	REFERÊNCIAS.....	64
	APÊNDICE A.....	76
	APÊNDICE B.....	77

1 INTRODUÇÃO

A capacidade de se decifrar os glicocódigos, uma terceira forma de se codificar informações biológicas, adicionalmente aos ácidos nucleicos e proteínas, tem sido um dos grandes desafios da biotecnologia, uma vez que glicoconjugados estão envolvidos em inúmeros eventos de reconhecimento celular que são alvos essenciais para o desenvolvimento de produtos e processos de aplicação biomédica (GUPTA; SUROLIA, 2012; ŠEBESTÍK; REINIŠ; JEŽEK, 2012). Tal dificuldade está relacionada com a grande variedade de estruturas ramificadas e modificações que suas unidades monoméricas são capazes de formar e sofrer, respectivamente (PILOBELLO; MAHAL, 2007).

Na busca para entender esses glicocódigos, muitos estudos em biotecnologia tem se concentrado na caracterização molecular e aplicação de lectinas, uma classe de proteínas que apresenta a capacidade de se ligar de forma excepcionalmente específica a carboidratos, sendo, portanto, uma importante ferramenta no campo da glicômica (AMBROSI; CAMERON; DAVIS, 2005; HU; TATENO; HIRABAYASHI, 2015).

Devido a essa habilidade de interagir de maneira específica e reversível a carboidratos, sem contudo modificá-los enzimaticamente (LIS; SHARON, 1998), as lectinas podem ser encontradas em todas as formas de vida conhecida, embora aquelas de origem vegetal, especialmente as oriundas de sementes de legumes, onde atuam como importantes fontes de nitrogênio e carbono durante o desenvolvimento da semente (MILLERD, 1975), sejam os principais alvos de estudo, já que aí apresentam-se em grandes quantidades e com relativa facilidade em ser purificada (DALKIN; BOWLES, 1983; NASCIMENTO et al., 2012).

Nas sementes de legumes, as lectinas são encontradas predominantemente em pequenas organelas chamadas de corpos proteicos, os quais são circundados por uma membrana oriunda do retículo endoplasmático (PERNOLLET, 1978). Aí, além de serem fontes ricas de nitrogênio e carbono para o processo germinativo, vários estudos descrevem a participação dessas (glico)proteínas em atividades de defesa vegetal contra estresses bióticos e abióticos (DANG; VAN DAMME, 2015; MACEDO; OLIVEIRA; OLIVEIRA, 2015; SCHUTTER; VAN DAMME, 2015; VANDENBORRE; SMAGGHE; VAN DAMME, 2011), bem como no processo ativo de germinação, sendo sugerido seu envolvimento no desenvolvimento da radícula (DELATORRE et al., 2013). Adicionalmente, tais proteínas, e em especial aquelas provenientes da subtribo Diocleinae, desenvolvem em modelos animais e in vitro efeitos de

grande interesse biotecnológico, como atividades antidepressivas, antitumorais e anti- e pró-inflamatórias (BARAUNA et al., 2006; ROCHA et al., 2011; ROCHA et al., 2015; SOUSA et al., 2016a; YAU et al., 2015).

Todas essas atividades levam em consideração a forma madura das lectinas, a qual, sendo dependentes dos íons metálicos Ca^{2+} e Mn^{2+} para a estruturação do sítio de ligação a carboidrato (LORIS et al., 1998), origina-se a partir de um complexo, e ao mesmo tempo fascinante, mecanismo pós-traducional de permuta circular, onde uma glicoproteína precursora, inicialmente incapaz de se ligar a carboidratos, é modificada através de etapas de deglicosilação, clivagem proteolítica e rearranjo de sua cadeia principal para dar lugar a uma proteína com sequência invertida, onde antes o que eram N e C-terminais no precursor, torna-se o C- e N-terminais, respectivamente, na forma madura (BOWLES et al., 1986; CHRISPEELS et al., 1986).

Ao mesmo tempo que esse processamento por meio da permuta circular é bem caracterizado, especialmente para as representantes da subtribo Diocleinae, muitos aspectos ainda continuam pouco compreendidos, como por exemplo com relação ao ponto exato onde a lectina se torna capaz de se ligar de forma específica a um determinado carboidrato, sendo reportado que a deglicosilação sem clivagem polipeptídica seria o passo-chave para essa ativação (MIN; DUNN; JONES, 1992; SHELDON; BOWLES, 1992), mas posteriormente sendo considerado insuficiente (RAMIS et al., 2001). Além disso, mais recentemente, foi visto que o precursor deglicosilado apresenta a mesma afinidade à sacarídeos (manose e glicose) que sua forma madura, embora com potência duas vezes maior (SOUSA et al., 2016b), demonstrando, dessa forma, que mais estudos devem ser realizados a fim de se compreender com maior profundidade o processo de maturação dessas proteínas.

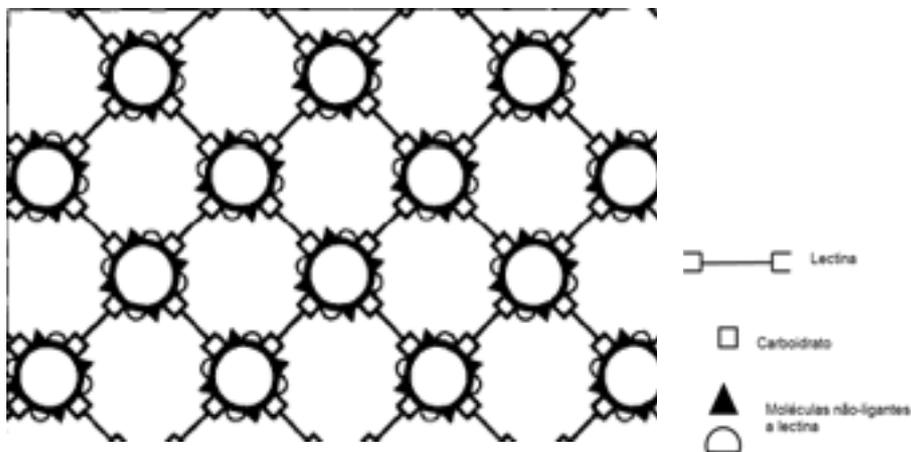
Na tentativa de dar maior esclarecimento acerca desses dados um tanto quanto incongruentes e na perspectiva de que a caracterização estrutural de proteínas revela muito de sua função, o presente trabalho teve por objetivo fazer uma análise detalhada da estrutura cristalográfica do precursor deglicosilado de uma lectina Diocleinae complexada com manose.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Lectinas

Segundo Lis e Sharon (1998), lectinas são proteínas de origem não imune que possuem pelo menos um sítio não-catalítico de ligação reversível a mono- e oligossacarídeos específicos. Apesar dessa ser a definição mais aceita para essas proteínas e das mesmas terem a recebido na segunda metade do século passado, elas vem sendo estudadas desde o final do século XIX, quando Stillmark, em 1888, isolou a primeira dessas proteínas, a qual foi denominada de ricina devido à planta da qual foi extraída, *Ricinus communis*. Devido sua propriedade de aglutinar eritrócitos (FIGURA 1), essa proteína, bem como todas aquelas de origem vegetal que apresentavam a mesma característica, foi denominada de fitohemaglutinina. Entretanto, sabe-se hoje que a ricina apresenta uma atividade hemaglutinante relativamente baixa, sendo caracterizada pela sua alta toxicidade devido à sua capacidade de inibir a síntese proteica (PEUMANS; HAO; VAN DAMME, 2001).

Figura 1 – Ação aglutinante de lectinas.



Fonte: Adaptado de Kennedy *et al.* (1995).

Lectinas com mais de um sítio de ligação podem interagir com os carboidratos na superfície de células adjacentes prendendo-as em uma rede que impede sua livre movimentação (aglutinação).

A partir de seus estudos com o sistema sanguíneo ABO, Boyd e Shapleigh (1954), corroborando diversos trabalhos que demonstraram que aglutininas oriundas de diferentes leguminosas reagiam de modo específico com células sanguíneas (revisado por SHARON; LIS, 2004), passaram a utilizar o termo lectina (do latim *legere* – selecionar, escolher) para nomear essas proteínas que ligavam-se com especificidade aos glicoconjugados da superfície dos eritrócitos que caracterizavam cada um destes no sistema ABO.

As lectinas constituem um grupo de proteínas presente em todas as formas de vida (SCHUTTER; VAN DAMME, 2015). Cada tipo de organismo apresenta lectinas com a propriedade comum de reconhecer carboidratos, embora suas sequências de resíduos de aminoácidos bem como suas estruturas terciárias e quaternárias sejam bastante particulares, criando um espectro de arranjos biológicos altamente diverso.

Devido à sua ampla distribuição, bem como aos seus aspectos estruturais e de especificidade a sacarídeos, as lectinas podem ser classificadas de diversas formas. Neste contexto, Vasconcelos e Oliveira (2004) revisaram essa temática, destacando que as lectinas podem ser agrupadas levando-se em consideração três parâmetros: (1) estrutura tridimensional; (2) relação estrutural e evolutiva e (3) afinidade a monossacarídeos.

Sendo sua principal característica a capacidade de se ligar com especificidade a carboidratos, as lectinas podem ser utilizadas, por exemplo, como ferramentas para a detecção, o isolamento e a caracterização de glicoconjugados, sendo essenciais para o entendimento de mudanças nas superfícies celulares em processos fisiológicos e patológicos, bem como nos estudos imunológicos e oncológicos (ETZLER, 1985; KOMATH; KAVITHA; SWAMY, 2006; SHARON; LIS, 2004).

2.1.1 Lectinas vegetais

Lectinas estão presentes em virtualmente todos os grupos taxonômicos do reino Plantae, desde as mais basais até aquelas caracterizadas pelo desenvolvimento de flores (ETZLER, 1985).

Suas especificidades a carboidratos, bem como suas estruturas, tem sido estudadas para uma grande variedade de espécies vegetais, onde se percebe que lectinas de um determinado grupo taxonômico apresentam características únicas que as distingue de outras categorias (STINISSEN; PEUMANS; CARLIER, 1983). Essas diferentes características, principalmente no que se refere às relações evolutivas e de sequência do domínio de reconhecimento a carboidrato (CRD, do inglês *Carbohydrate-Recognition Domain*), permitem com que, atualmente, as lectinas oriundas de plantas sejam classificadas em 12 famílias (VAN DAMME; LANNOO; PEUMANS, 2008) (TABELA 1), das quais aquelas provenientes de leguminosas são as mais bem estudadas.

Tabela 1 – Famílias de lectinas vegetais.

Família	Carboidrato Ligante Típico	Localização Celular Preditada
Lectina de <i>Agaricus bisporus</i>	GlcNAc/GalNAc	Núcleo, citosol
Amarantina	GalNAc	Núcleo, citosol
Aglutinina relacionada à quitinase	Glicanos N-ligados	Vacúolo, ligada à membrana plasmática
Cianovirina	Manose	Núcleo
Lectina de <i>Eonymus europaeus</i>	Galactosídeos, Glicanos N-ligados ricos em manose	Núcleo, citosol
Lectina de <i>Galanthus nivalis</i>	Manose	Vacúolo, núcleo, citosol ou ligada à membrana plasmática
Heveína	Quitina	Vacúolo
Jacalina	Manose, Galactose	Núcleo, citosol
Leguminosa	Manose	Núcleo, citosol
LysM	Quitina, peptidoglicano	Vacúolo, núcleo, citosol ou ligada à membrana plasmática
Lectina de <i>Nicotiana tabacum</i>	Glicanos N-ligados ricos em manose e complexos	Núcleo, citosol
Ricina-B	Gal/GalNAc conjugados ou não ao ácido siálico	Vacúolo, núcleo, citosol

Fonte: adaptado de Schutter e Van Damme (2015).

¹ GlcNAc = N-acetilglicosamina; GalNAc = N-acetilgalactosamina.

² Oligossacarídeo covalentemente ligado ao nitrogênio amídico de uma asparagina em uma glicoproteína.

2.1.1.1 Lectinas de leguminosas

As sementes dos legumes são particularmente ricas em lectinas, embora estas proteínas possam também ser encontradas em outros órgãos e tecidos (LAM; NG, 2011). A família botânica a qual pertence os legumes, *Fabaceae* (Leguminosae), é muito estudada com relação a essa classe de proteínas, uma vez que diversos estudos indicam grande homologia sequencial e estrutural das lectinas, especialmente se as mesmas são encontradas em espécies de uma mesma tribo ou subtribo (ETZLER, 1985). É interessante destacar que o estudo de lectinas extraídas de leguminosas sempre leva em consideração suas homologias com a Concanavalina A (ConA), extraída de sementes da *Canavalia ensiformis* e que é a lectina vegetal mais bem estudada com relação à sua estrutura e suas atividades biológicas (LIS; SHARON, 1998). A partir de estudos comparativos, constatou-se grande conservação de sequência e estrutura nos domínios de reconhecimento a carboidrato e nos sítios hidrofóbicos (FIGURA 2). Devido a essa grande homologia, as demais lectinas dessa família são denominadas de *ConA-like*.

Figura 2 - Homologia de seqüências de lectinas dos gêneros *Dioclea* e *Canavalia*.

DRL	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTVHISYNSVSKR
DGL	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTVHISYNSVAKR
DWL	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTVHISYNSVAKR
DVL	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTVHISYNSVAKR
DVir	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTAHISYNSVAKR
Dros	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTVHISYNSVAKR
Dguia	ADTIIVAVELDSYPNTDIGDPNYPHIGIDIKSVRS*STARWNMQTGKVGTAHISYNSVAKR
ConGF	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKIAKWNMQDGKVATAHIIYNSVGKR
ConBol	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR
CGL	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR
ConM	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR
ConV	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR
ConBr	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR
ConA	ADTIIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVDKR
	*****:*****:*****:***: *;**** ***,*_* ** ** **
DRL	LSAVSYSGSSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
DGL	LSAVSYSGSSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
DWL	LSAVSYSGSSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
DVL	LSAVSYSGSSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSA
DVir	LSAVSYTGSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
Dros	LSAVSYTGSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
Dguia	LSAVSYTGSSTTVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
ConGF	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
ConBol	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
CGL	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
ConM	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
ConV	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
ConBr	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
ConA	LSAVSYPNADSATVSYDVLNNVLPWVVRVGLSATTGLYKQNTNILSWSFYSKLTNSI
	***** ..:*****:*****:*****:*****:*****:*****:***
DRL	ADENSLHFSFHKFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
DGL	ADENSLHFSFHKFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
DWL	ADENSLHFSFHKFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
DVL	ADENSLHFSFHKFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
DVir	ADANSLHFSFNQFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
Dros	ADANSLHFSFNQFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
Dguia	ADENSLHFSFNQFSQNPKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
ConGF	AETNALHFTFNQFQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
ConBol	HETNALHFMFNQFSKQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
CGL	HETNALHFMFNQFSKQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
ConM	HETNALHFMFNQFSKQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
ConV	HETNALHFMFNQFSKQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
ConBr	HETNALHFMFNQFSKQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
ConA	HETNALHFMFNQFSKQKDLILQGDASTDSGNLQLTQVSSSGDPQGNVGRALFYAPVH
	: *;**** *;:***: ***** ** * .;****:***:*** ** ** *****
DRL	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
DGL	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
DWL	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
DVL	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
DVir	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
Dros	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
Dguia	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
ConGF	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
ConBol	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
CGL	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
ConM	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
ConV	IWEKSAVVASFDATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
ConBr	IWEKSAVVASFEATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
ConA	IWEKSAVVASFEATFTFLIKSPDRPADGITFFFIANTDTSIPSGSGRLLGLFPDAN
	_***:*****:*****:*****:*****:*****:*****

- DRL = *Dioclea reflexa* (81%)
- DGL = *D. grandiflora* (81%)
- DWL = *D. wilsonii* (81%)
- DVL = *D. violacea* (81%)
- DVir = *D. virgata* (85%)
- Dros = *D. rostrata* (83%)
- Dguia = *D. guianensis* (83%)
- ConGF = *Canavalia grandiflora* (91%)
- ConBol = *C. boliviana* (97%)
- CGL = *C. gladiata* (98%)
- ConM = *C. maritima* (98%)
- ConV = *C. virosa* (98%)
- ConBr = *C. brasiliensis* (98%)
- ConA = *C. ensiformis*

Fonte: Elaborada pelo autor.

O alinhamento mostra a pouca variabilidade das seqüências. Os números entre parênteses, à direita, referem-se ao grau de identidade das seqüências em relação à ConA. As cores indicam o grupo químico dos resíduos de aminoácidos: azul = carregados negativamente, rosa = carregados positivamente, verde = polares não carregados e vermelho = hidrofóbicos. * Resíduos conservados; : modificações de conservação; . modificações semi-conservadas.

2.1.1.1.1 DGL, uma lectina isolada de sementes da *Dioclea grandiflora*

Dioclea grandiflora Mart. ex Benth., popularmente conhecida como mucunã, mucunã-de-caroço, mucunã vermelha ou “olho-de-boi”, é uma leguminosa encontrada na caatinga e no cerrado do Nordeste brasileiro, cujas sementes e raízes são utilizadas pela medicina tradicional no tratamento de distúrbios prostáticos e de doenças renais (SILVA *et al.*, 2010), mas que de acordo com diversos outros estudos também apresenta atividades antinociceptivas, antioxidantes e anti-inflamatórias (SÁ; ALMEIDA; BHATTACHARYYA, 2013).

Como várias outras leguminosas, suas sementes são abundantes em lectinas, uma das quais, chamada de DGL e isolada primeiramente por Moreira *et al.* (1983), apresenta características bioquímicas e físico-químicas muito semelhantes à ConA, tendo, portanto, afinidade geral por resíduos de manose e glicose, mas principalmente pelo núcleo trimanosídico [3,6-di-*O*-(α -D-manopiranosil)- α -D-manopiranosídeo] presente em oligossacarídeos *N*-ligados (BREWER; BHATTACHARYYA, 1986; CHERVENAK; TOONE, 1995). Apesar disso, são observadas diferenças relacionadas às suas especificidades finas para grandes oligossacarídeos de manose *N*-ligados e para carboidratos complexos, onde a DGL é praticamente incapaz de se ligar a essas moléculas (GUPTA *et al.*, 1996), provavelmente devido a mudanças conformacionais em *loops* (alças) adjacentes ao seu CRD (ROZWARSKI *et al.*, 1998).

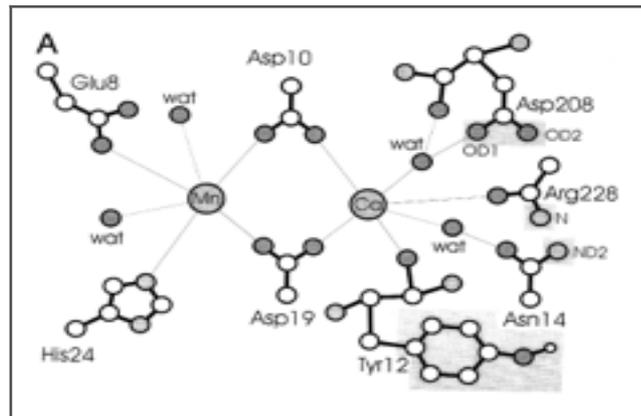
Além de ser muito estudada para se entender como lectinas de diferentes origens, mas com a mesma composição de resíduos de aminoácidos e arquitetura de CRD, apresentam afinidades diferenciadas por carboidratos (DAM *et al.*, 1998a; DAM *et al.*, 1998b; WEATHERMAN *et al.*, 1996), a DGL é empregada como ferramenta de caracterização das potenciais atividades biológicas desenvolvidas pelas lectinas. Neste sentido, alguns trabalhos tem demonstrado que a DGL tem (1) caráter pró-inflamatório (BENTO *et al.*, 1993), (2) capacidade de induzir a produção de óxido nítrico (NO) *in vitro* e *in vivo* (ANDRADE *et al.*, 1999) e (3) capacidade de inibir a formação de biofilmes (TEIXEIRA *et al.*, 2006).

2.1.2 Estrutura de lectinas da subtribo *Diocleinae*

As lectinas isoladas de espécies da subtribo *Diocleinae* são excelentes fontes de estudos para se entender as relações estrutura-função que caracterizam uma proteína, uma vez que apesar da enorme homologia existente entre suas estruturas primárias e terciárias, as mesmas apresentam distintas especificidades por mono- e oligossacarídeos (MANOJ; SUGUNA, 2001) e variações nas associações quaternárias (BRINDA; SUROLIA; VISHVESHWARA, 2005). Seus monômeros apresentam um peso médio de 25kDa, possuindo dois sítios de ligação, S1 e S2, altamente conservados para os íons divalentes Mn^{2+} e Ca^{2+} , respectivamente, os quais são essenciais para a estabilização do CRD e, assim, para a interação das lectinas com moléculas de carboidrato (BECKER *et al.*, 1975; KAUSHIK; MOHANTY; SUROLIA, 2009). Cada íon está coordenado por seis interações de hidrogênio (FIGURA 3), sendo quatro provenientes da proteína e duas de moléculas de água. Interessantemente, a ligação desses íons em seus respectivos sítios parece acontecer de forma sequencial (BOUCKAERT *et al.*, 1996), com o Mn^{2+} , ou outro metal de transição, como Co^{2+} , Ni^{2+} , Zn^{2+} , Cd^{2+} ou Mg^{2+} (SHOHAM; KALB; PECHT, 1973), sendo coordenado inicialmente em S1 para, em seguida, o Ca^{2+} ser coordenado em S2.

Embora essa seja uma característica geral para lectinas de leguminosas, é interessante ressaltar que existem exceções para essa necessidade de sítios de ligação para íons divalentes. Por exemplo, as lectinas α AI (*amylase inhibitor*) e *arcelin* (MIRKOV *et al.*, 1994; OSBORNI *et al.*, 1988), derivadas da leguminosa *Phaseolus vulgaris* (subtribo *Phaseolinae*), não possuem o *loop* correspondente aos resíduos Pro13-Pro23 de outras lectinas de leguminosas, o qual está envolvido na coordenação dos íons. Neste caso, a estabilização das proteínas parece ser efetuada através de interações entre os anéis da Tyr85 e da Phe127, bem como por duas interações de hidrogênio entre as cadeias principais dos resíduos Ala84/Thr205 e Gly207/Ala84 (BOMPARD-GILLES *et al.*, 1996).

Figura 3 – Coordenação dos íons Ca^{2+} e Mn^{2+} .



Fonte: Adaptado de Loris *et al.* (1998).

A coordenação do Ca^{2+} envolve resíduos de aminoácidos que também participam da ligação a moléculas de carboidrato no CRD, como Tyr12 e Asn14. “wat” refere-se a moléculas de água.

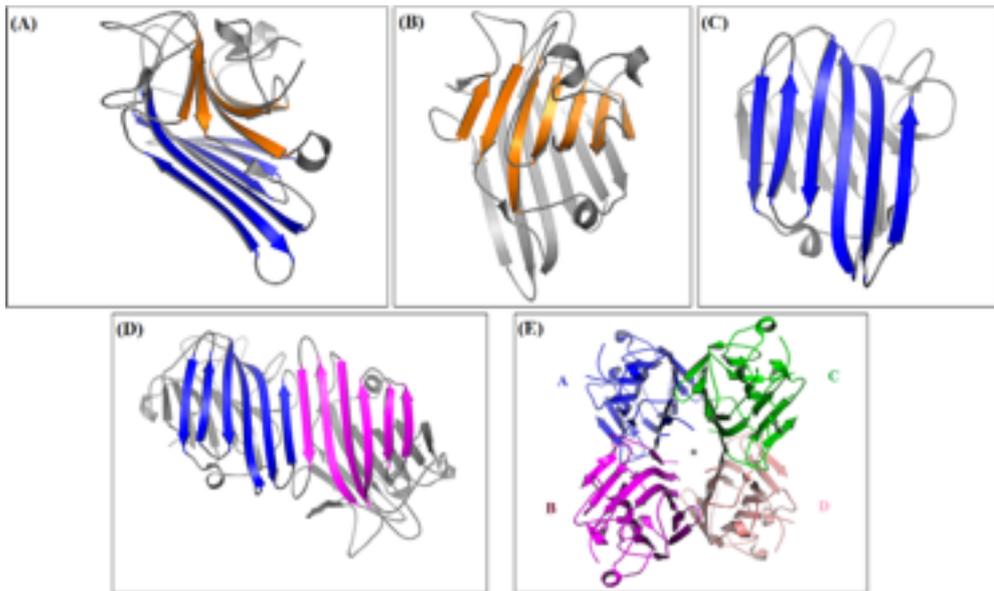
Cada monômero possui uma arquitetura tridimensional denominada de *jelly-roll* (SRINIVAS *et al.*, 2001), onde um “sanduiche” é formado por uma parte frontal contendo sete fitas- β antiparalelas que se sobrepõe a uma porção posterior com seis fitas- β antiparalelas, de modo que estas duas regiões encontram-se conectadas por diversos *loops* e uma pequena folha- β de cinco fitas (FIGURAS 4A, B e C). Entretanto, as formas biologicamente ativas das lectinas dessa subtribo se encontram di- ou tetramerizadas, fato este influenciado pelo pH do meio (NAGANO *et al.*, 2008; OLIVEIRA *et al.*, 2008; SENEAR; TELLER, 1981).

A formação de dímeros e tetrâmeros nas lectinas de leguminosas depende prioritariamente da maneira como as seis fitas posteriores se associam entre as diferentes subunidades, dando origem a sete tipos diferentes de associação ou interface (TABELA 2). Uma destas, conhecida como dímero canônico ou interface tipo II (Figura 4D), foi descrita primeiramente para a ConA (EDELMAN *et al.*, 1972; HARDMAN; AINSWORTH, 1972), sendo caracterizada pela interação lado a lado de dois monômeros de modo a formar uma arquitetura contínua, através da associação das porções posteriores de cada monômero, de 12 fitas- β antiparalelas. As outras associações (tipo X) caracterizam-se por uma disposição face-à-face das folhas- β posteriores, cujas variações (X1-X5) dependem da diferença angular entre as estruturas que estão em contato (SINHA *et al.*, 2007).

O arranjo tetramérico, que apresenta duas interfaces distintas (geralmente uma tipo II e outra tipo X₁₋₅), é formada pela associação entre as partes posteriores de dois dímeros canônicos (Figura 4E), as quais, sendo côncavas, formam uma cavidade central preenchida com água que pode ligar compostos não-glicídicos (DELATORRE *et al.*, 2013; EDELMAN;

WANG, 1978; ROBERTS; GOLDSTEIN, 1983). Entretanto, de acordo com o tipo de interface, esta associação pode ocorrer de diferentes formas, de modo que variações são detectadas nos arranjos tetraméricos (BOUCKAERT *et al.*, 1999; PRABU *et al.*, 1998; VIJAYAN; CHANDRA, 1999), refletindo-se, assim, nas diferenças de atividades biológicas observadas para as lectinas vegetais, como abordado por Sanz-Aparicio *et al.* (1997) para a lectina de *Canavalia brasiliensis* (ConBr).

Figura 4 – Características gerais do arranjo tridimensional das lectinas *Diocleinae*.



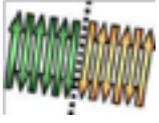
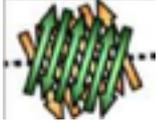
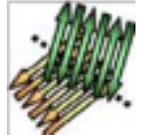
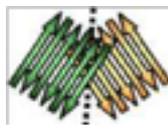
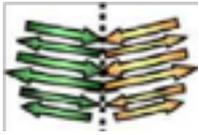
Fonte: Elaborada pelo autor.

(A) Representação do arranjo monomérico conhecido como β -sandwich, destacando-se a sobreposição das sete fitas- β frontais (laranja) em relação às seis posteriores (azul). (B) e (C) Disposição das sete fitas- β frontais e das seis posteriores, respectivamente. (D) Dímero canônico, destacando o arranjo contínuo de 12 fitas- β formado pelas porções posteriores de cada monômero (azul e magenta). (E) Arranjo tetramérico formado a partir da associação entre dois dímeros canônicos. * Cavidade central gerada pela disposição dos dímeros canônicos no tetrâmero.

2.1.3 Domínio de Reconhecimento a Carboidrato (CRD)

Como já mencionado, lectinas apresentam uma ampla variedade de especificidade para mono- e oligossacarídeos, podendo ser classificadas em cinco grupos de acordo com sua afinidade por monossacarídeos (MANOJ; SUGUNA, 2001; VASCONCELOS; OLIVEIRA, 2004): manose/glicose (Man/Glc), galactose/*N*-acetilgalactosamina (Gal/GalNAc), *N*-acetilglicosamina (GlcNAc), fucose e ácido *N*-acetilneuramínico.

Tabela 2 – Tipos de interfaces encontradas em lectinas vegetais.

Tipo de Interface	Ocorrência	Representação
Canônica ou Tipo II	Presente em lectinas diméricas como a PSL, ou tetraméricas como ConA e SBA.	
X1	Observada no dímero DB58 e nas interfaces não-canônicas da PHA e SBA.	
X2	Apenas em tetrâmeros, como na interface não canônica da ConA e DGL.	
X3	Dímeros tipo EcorL.	
X4	Dímeros tipos GS4 e GS1.	
X5	Observada em dímeros do tipo GS1. Assemelha-se a interface X4, entretanto, com uma orientação diferente.	
Interface não-usual da PNA	Lectina de <i>Arachis hypogea</i> (PNA).	

Fonte: adaptado de Sinha *et al.* (2007).

PSL = lectina de *Pisum sativum* (ervilha); SBA = lectina de *Glycine max* (soja); DB58 = lectina vegetativa de *Dolichos biflorus* (feijão-de-corda); PHA = lectina de *Phaseolus vulgaris* (feijão comum); EcorL = lectina de *Erythrina corallodendron* (flor-de-coral); GS1 e GS4 = lectinas I e IV de *Griffonia simplicifolia* (grifonia).

Esse espectro de possibilidades de interação com carboidratos pode ser entendido a partir da análise do Domínio de Reconhecimento a Carboidrato (CRD) (DRICKAMER, 1988), o qual, de maneira geral, possui uma arquitetura muito similar nas lectinas de leguminosas, sendo formado por quatro *loops* presentes na superfície da região frontal de cada monômero (SHARMA; SUROLIA, 1997; YOUNG; OOMEN, 1992). Nesse sentido, ressaltase que lectinas extraídas de plantas filogeneticamente relacionadas e que possuem a mesma afinidade por monossacarídeos, como é o caso daquelas pertencentes à subtribo *Diocleinae*, apresentam seus CRDs sequencialmente conservados (MANOJ; SUGUNA, 2001), embora a especificidade por carboidratos mais complexos seja diferente (LIS; SHARON, 1998). A

partir disso, constata-se que essa variação se reflete nas atividades biológicas desempenhadas por essas proteínas, onde, por exemplo, verifica-se que ConA, ConM e ConBr (*Canavalia brasiliensis*) se diferenciam, dentre outras coisas, quanto aos efeitos pró- e antiinflamatórios (ALENCAR *et al.*, 1999; ASSREUY *et al.*, 1999; BENTO *et al.*, 1993), liberação de histamina (GOMES *et al.*, 1994; FERREIRA *et al.*, 1996), produção de óxido nítrico (ANDRADE *et al.*, 1999), mitogenicidade (BARRAL-NETO *et al.*, 1992) e produção de citocinas (CAVADA *et al.*, 2001).

Mas como uma região que apresenta extrema conservação em sua sequência de resíduos de aminoácidos pode apresentar especificidades tão distintas? Uma resposta para tal questionamento foi dada pelos trabalhos de Gadelha *et al.* (2005) e Delatorre *et al.* (2006), demonstrando que os CRDs da ConM, uma lectina de sementes da *Canavalia maritima*, e da ConA apresentam uma conformação ligeiramente diferente devido a uma mutação que altera o resíduo 202, o qual naquela é uma Serina e nesta é uma Prolina, mostrando que essa modificação deixa o domínio da ConM cerca de 0.70 Å mais aberto que o da ConA, permitindo, assim, que carboidratos mais complexos possam interagir com maior facilidade com aquela proteína.

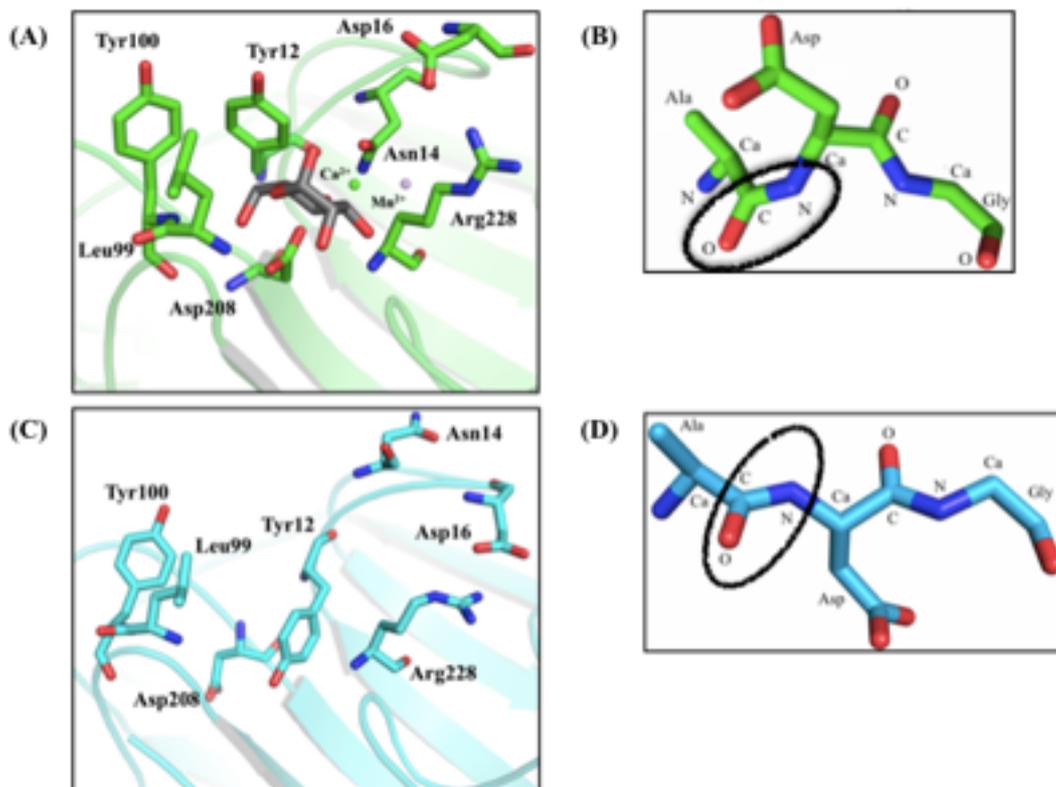
As moléculas de carboidratos nas lectinas da subtribo *Diocleinae* são coordenadas, de maneira geral, pelos resíduos Tyr12, Asn14, Leu99¹, Tyr100, Asp208 e Arg228 (FIGURA 5A) (LORIS *et al.*, 1998) por meio de interações de hidrogênio e de forças de Van der Waals (PEUMANS; VAN DAMME, 1995). Um sétimo resíduo pode participar dessa coordenação: o Asp16 interage com o núcleo do trimanosídeo Man α 1-3[Man α 1-6]Man, como demonstrado pela resolução das estruturas cristalográficas da ConA e da DGL complexadas com esse trissacarídeo (LORIS *et al.*, 1996; NAISMITH; FIELD, 1996; ROZWARSKI *et al.*, 1998).

Para as lectinas *Diocleinae* se ligarem a carboidratos é necessário que a ligação peptídica entre os resíduos Ala207 e Asp208, que encontram-se em uma fita- β ao lado de S2, esteja na configuração *cis* e que os íons metálicos Mn²⁺ e Ca²⁺ estejam em seus respectivos sítios (FIGURA 5B). Nesse estado, denominado de *locked* porque envolve um aumento na afinidade da proteína pelos dois íons (KOENIG; BREWER; BROWN, 1978), o CRD encontra-se estruturado. Embora esse estado possa ainda existir na completa ausência dos íons, mantendo a capacidade do CRD se ligar de maneira específica à glicose e manose

¹ Na CPL, uma lectina de sementes de *Camptosema pedicellatum*, ocorre uma mutação conservativa onde o resíduo de leucina nessa posição é substituído por uma valina (TEIXEIRA *et al.*, 2012).

(BROWN; KOENIG, BREWER, 1982), sua presença é essencial para a permanência dessa arquitetura. Do contrário, esse estado *locked* é modificado de tal maneira que onde antes encontrava-se a cadeia lateral do Asp208, agora está a Tyr12 (FIGURA 5C), o que impede a formação de complexos lectina-carboidrato. Esse novo estado, caracterizado pela configuração *trans* da ligação Ala207-Asp208 (FIGURA 5D) (BOUCKAERT *et al.*, 1995), é denominado *unlocked*.

Figura 5 – Domínio de Reconhecimento a Carboidrato (CRD) de lectinas *Diocleinae*.



Fonte: Elaborada pelo autor.

(A) CRD em sua forma estruturada (*locked*) com uma molécula de carboidrato ligada, sendo destacado os resíduos que o compõem e os íons metálicos (esferas). (B) Ligação peptídica *cis* entre Ala207 e Asp208 (elipse). (C) CRD em sua forma não estruturada (*unlocked*), destacando como os resíduos Tyr12 e Asp208 apresentam uma orientação bastante diferente em relação ao CRD estruturado. Isso impede a ligação de carboidratos. (D) Ligação peptídica *trans* entre Ala207 e Asp208 (elipse).

Interessantemente, o domínio de reconhecimento a carboidrato parece não ser sítio apenas para a ligação de moléculas glicídicas. Inicialmente, Gael *et al.* (2000) mostraram que porfirinas se ligam no CRD da ConA, possivelmente mimetizando interações dessa lectina com monossacarídeos. Mais recentemente, Delatorre *et al.* (2013) reportaram a presença do anel hidrofóbico indol ligado aos resíduos Asp16 e Arg228, constituintes do CRD da ConM, sendo sugerido, neste caso, que a lectina poderia estar relacionada em um processo

de controle da disponibilização desta molécula, a qual é liberada como um composto volátil quando plantas encontram-se sob ataque de insetos herbívoros (FREY *et al.*, 2000; ZHUANG *et al.*, 2012). Por fim, estruturas cristalográficas da ConM (SILVA-FILHO, 2013) e ConBr (NÓBREGA, 2016) complexadas com adenina mostram que essa base nitrogenada, diferentemente dos estudos reportados para outras lectinas vegetais (HAMELRYCK *et al.*, 1999; SHETTY *et al.*, 2013), se liga no mesmo local do anel indol, sugerindo que o CRD possa servir como um sítio para ligação de nucleosídeos (NÓBREGA, 2016).

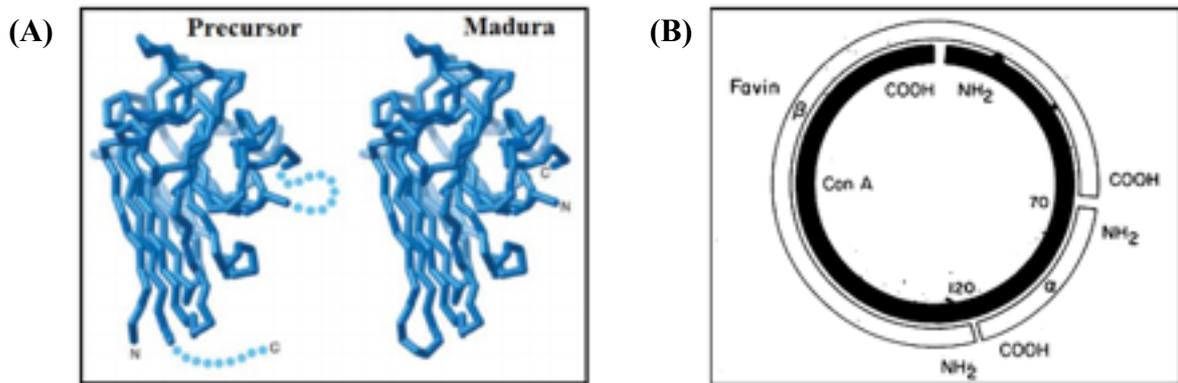
2.2 Maturação de lectinas ConA-like ocorre por permutação circular

Todas as atividades fisiológicas e biológicas atribuídas às lectinas vegetais, tais como defesa vegetal (SHUTTER; VAN DAMME, 2015) e vasorelaxamento (BARROSO-NETO *et al.*, 2014), levam em consideração sua forma madura, a qual é proveniente de processamentos pós-traducionais de uma (glico)proteína precursora (MÜNTZ, 2007). No caso das *Diocleinae*, esse mecanismo é extremamente complexo, sendo único quando comparado com outros eventos de maturação de proteínas de armazenamento. Chamado de permutação circular, esse processo envolve a união das extremidades N- e C-terminais, formando uma proteína temporariamente circular, para em seguida ocorrer uma clivagem proteolítica em outro ponto da cadeia, gerando um polipeptídeo com novas extremidades (FIGURA 6A) e com sequência invertida em relação ao precursor (GOODSELL, 2010).

A primeira observação desse tipo de maturação proteica ocorreu quando Cunningham *et al.* (1979) compararam a sequência de aminoácidos da favina, uma lectina heterodimérica de *Vicia faba*, com a da ConA, observando que a cadeia α da primeira correspondia em homologia ao segmento compreendido entre os resíduos 70 e 119 da segunda e que a cadeia β tinha sua homologia iniciada no resíduo 120, estendia-se ao C-terminal (resíduo 237) e continuava sem interrupção através dos 69 primeiros resíduos do N-terminal da ConA (FIGURA 6B).

Após esse estudo, ficou evidente que a ConA não poderia ser sintetizada nos cotilédones na forma como fora até então caracterizada. Nesse sentido, diversos trabalhos começaram a mostrar que, de fato, essa lectina era sintetizada nos ribossomos do RE na forma de uma pré-pró-proteína (pré-proConA) que ia sendo modificada à medida que seu mRNA era traduzido, dando origem a um polipeptídeo, a proConA, que seria transportado para os corpos proteicos para posterior processamento (BOWLES *et al.*, 1986).

Figura 6 – Permutação circular da ConA e comparação de sua forma madura com a favina.

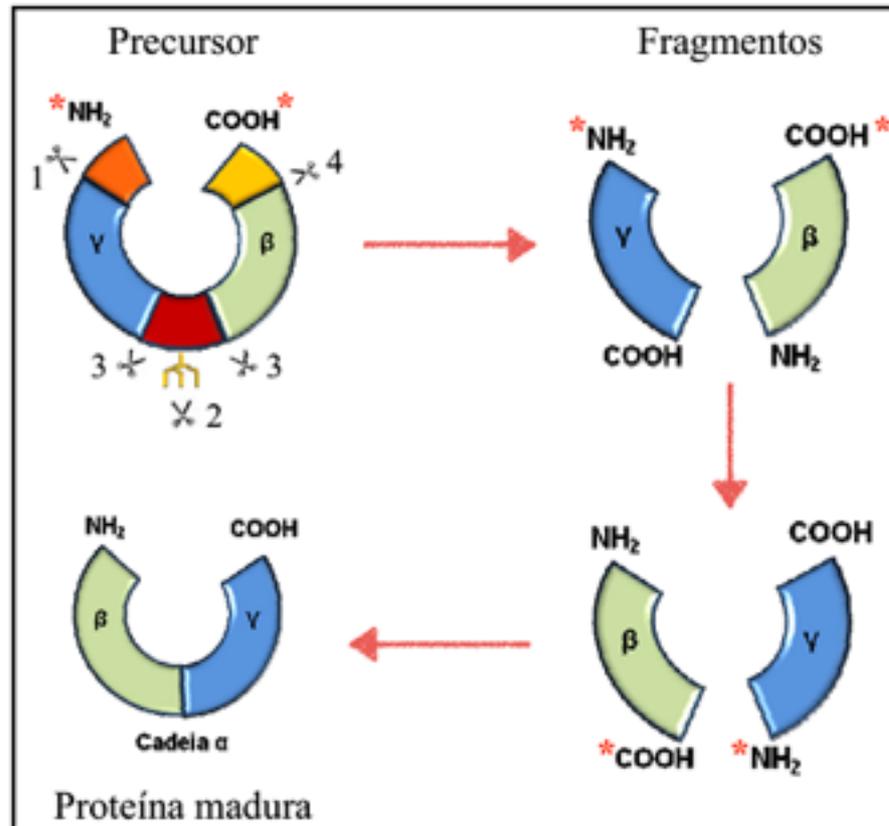


Fonte: Adaptado de Goodsell (2010) e Cunningham *et al.* (1979).

(A) O precursor possui uma extremidade C-terminal e um *loop* extras (pontilhado azul claro). No processamento, o C-terminal é clivado enzimaticamente e a nova extremidade é unida ao N-terminal para formar uma proteína circular. Posteriormente, o *loop* extra também é clivado para dar origem às novas extremidades da, agora, proteína madura. (B) Alinhamento das cadeias α e β da favina (barra clara) com a ConA (barra escura) mostrando que a permutação circular dá a homologia máxima entre as duas sequências.

Inicialmente, a pré-proConA, contendo um peptídeo sinal de direcionamento ao RE com 30 resíduos de aminoácidos em seu N-terminal, sofre ação proteolítica para retirada dessa sequência durante o processo de tradução (HIGGINS, 1984), convertendo-se então na proConA, a qual é glicosilada no próprio retículo pela adição de um oligossacarídeo rico em manose ($\text{Man}_9\text{GlcNAc}_2$) à Asn132 (HERMAN; SHANNON; CHRISPEELS, 1985). Em seguida, a glicoproteína é direcionada para o complexo golgiense, onde ocorre sua deglicosilação e subsequente modificação da cadeia polipeptídica com retirada enzimática de um segmento de 15 resíduos na porção central contendo o sítio de N-glicosilação. Isso gera dois grandes polipeptídeos denominados de fragmentos β e γ . Um nonapeptídeo no C-terminal, envolvido na translocação da proConA aos corpos proteicos (CLAUDE *et al.*, 2005), é então removido e os dois fragmentos são covalentemente ligados um ao outro a partir de suas extremidades originais (CARRINGTON; AUFFRET; HANKE, 1985), formando assim a cadeia α madura de 237 resíduos com uma cadeia em ordem invertida em relação ao seu precursor (FIGURA 7). Esta última etapa de religação dos fragmentos, bem como aquela de remoção do segmento de 15 resíduos, é catalisada por uma asparaginil endopeptidase (MIN; JONES, 1994), a qual não apresenta atividade completamente eficiente, uma vez que nos procedimentos de purificação da ConA (EDMUNDSON *et al.*, 1971; WAXDAL *et al.*, 1971) e de lectinas ConA-like (CECCATTO *et al.* 2002) é muito comum existir uma mistura de 60% de cadeia α e 40% de seus fragmentos β (res. 1-118, 14 kDa) e γ (res. 119-237, 12 kDa).

Figura 7 – Etapas do processamento pós-traducional das lectinas *Diocleinae*.



Fonte: Adaptado de Sousa (2010).

O precursor é mostrado com todos os seus elementos. Os números 1, 2, 3 e 4 indicam, respectivamente, a clivagem do peptídeo sinal da região N-terminal (laranja) relacionado com o transporte da molécula nascente para o RE, deglicosilação, retirada do segmento de 15 resíduos (vermelho) e remoção do nonapeptídeo (amarelo) da região C-terminal. Isso leva à formação dos fragmentos β e γ , cujos respectivos C- e N-terminais se fundem para dar origem à proteína madura. Os asteriscos indicam as extremidades que serão unidas pela asparaginil endopeptidase.

Embora esse complexo mecanismo de maturação seja essencial para que as lectinas *Diocleinae* se tornem funcionais, alguns estudos tem demonstrado que suas formas precursoras também são capazes de se ligar a carboidratos e desenvolver atividades. Inicialmente, Min, Dunn e Jones (1992) mostraram que a forma deglicosilada da proConA recombinante podia não apenas adsorver a uma coluna cromatográfica contendo dextran (Sephadex G-75) como também podia ser eluída por α -metilmanosídeo, um monossacarídeo pelo qual a ConA tem alta afinidade. Posteriormente, Ramis *et al.* (2001) verificaram que a simples deglicosilação não era suficiente para permitir uma atividade lectínica da proConA, sendo necessário também uma acidificação do meio. E mais recentemente, foi visto que a pDGL, precursora deglicosilada da DGL, era capaz de promover a hemaglutinação de eritrócitos de coelho, sendo essa atividade inibida por manose em uma concentração cerca de duas vezes menor em relação àquela necessária para sua contraparte madura (SOUSA *et al.*,

2016b). Na tentativa de elucidar as bases estruturais para essa maior afinidade do precursor pela manose, esse mesmo trabalho realizou predições *in silico* para a ligação desse monossacarídeo às formas precursora e madura da DGL, sendo observado que a presença de uma interação adicional no complexo pDGL-manose seria a responsável pela diferença de afinidade encontrada. Além disso, foi sugerido que a oligomerização da pDGL não favoreceria a formação de tetrâmeros estáveis, diferente de sua forma madura (SOUSA *et al.*, 2016b). Na tentativa de corroborar essas ideias, desenvolveu-se o presente trabalho na perspectiva de que a elucidação da estrutura cristalográfica da pDGL complexada com a manose explique de maneira mais clara porque ocorre essas diferenciações.

3 OBJETIVOS

3.1 Objetivo geral

- Caracterizar os efeitos do processamento pós-traducional sobre a estrutura tridimensional de uma lectina de sementes de *Dioclea grandiflora*.

3.2 Objetivos específicos

- Caracterizar a estrutura cristalográfica do precursor deglicosilado de uma lectina de sementes de *D. grandiflora* (pDGL);
- Clonar o gene da pDGL;
- Expressar o gene da pDGL em um sistema heterólogo;
- Purificar a pDGL;
- Cristalizar a forma recombinante da pDGL na presença de manose;
- Difratar cristais do complexo pDGL-manose;
- Determinar a estrutura tridimensional do complexo pDGL-manose;
- Avaliar a presença de estruturas secundárias não-repetitivas no modelo cristalográfico da pDGL;
- Estudar o padrão de oligomerização da pDGL;
- Investigar as diferenças estruturais de ligação a carboidratos entre a pDGL e sua forma madura.

4 MATERIAL E MÉTODOS

4.1 Obtenção da forma recombinante da pDGL

A produção da pDGL recombinante foi realizada a partir de modificações da metodologia aplicada anteriormente para a forma madura da VML, uma lectina isolada de sementes de *Vatairea macrocarpa* (SOUSA *et al.*, 2016a).

O DNA genômico da pDGL (GenBank AM701773), amplificado pela reação em cadeia da polimerase (PCR) utilizando-se iniciadores dotados de sítios de restrição para as enzimas *NcoI* e *NotI* (NAGANO *et al.*, 2008), foi inserido no vetor de expressão pET-28a (Novagen, Darmstadt, Alemanha) pela enzima T4 DNA ligase (*New England BioLabs*, Reino Unido) para transformação de células *E. coli* TOP10 (Stratagene, La Jolla, Califórnia, Estados Unidos) por choque térmico na presença de cloreto de cálcio (SAMBROOK *et al.*, 1989).

O DNA plasmidial dos clones positivos, identificados por sequenciamento de DNA (LGC Genomics, Berlin, Alemanha), foi utilizado para transformar células *E. coli* BL21 *codon plus* por choque térmico na presença de cloreto de cálcio. A expressão do gene da pDGL nessa cepa bacteriana, que é suplementada com tRNAs para códons eucarióticos, foi induzida por IPTG (isopropil- β -tiogalactopiranosídeo) (Carl Roth GmbH, Karlsruhe, Alemanha) a uma concentração final de 0.5 mM.

A purificação da proteína recombinante, dotada de uma cauda de seis histidinas em sua extremidade C-terminal, ocorreu inicialmente por meio de uma cromatografia em coluna de níquel imobilizado (HisTrap, GE Healthcare), previamente equilibrada com uma solução tampão PBS (NaH_2PO_4 20 mM e NaCl 250 mM) suplementada com 20 mM de imidazol e glicerol 5%. A retirada de proteínas não-retidas foi realizada com a aplicação dessa mesma solução tampão, enquanto a eluição das proteínas retidas se deu pela aplicação dessa solução adicionada de imidazol 500 mM e glicerol 30%.

Na sequência, a fração proteica obtida por essa etapa inicial de purificação foi submetida a uma cromatografia de exclusão molecular em coluna Superdex 200 16/60 (GE Healthcare) previamente equilibrada com ácido L-malico 10 mM, MES 10 mM, Tris 10 mM, NaCl 250 mM e glicerol 5% para aprimorar a pureza da amostra e para realizar a troca do tampão da mesma visando os ensaios estruturais subsequentes.

4.2 Ensaio de cristalização

Para determinar a estrutura tridimensional da pDGL, ensaios de cristalização foram realizados com a proteína na sua forma holoproteica (com os metais Ca^{2+} e Mn^{2+}) na presença de D-manose. Para os ensaios, as concentrações proteicas foram ajustadas para 10 mg/mL com base em ensaios preliminares de solubilidade.

Para se obter essa forma holoproteica, a pDGL descongelada e mantida em uma solução tampão (ácido L-málico 10 mM, MES 10 mM, Tris 10 mM, NaCl 50 mM e glicerol 5%) foi suplementada com soluções de CaCl_2 e MnCl_2 a 1 mM cada. Em seguida, para os ensaios de co-cristalização, a D-manose foi adicionada em uma concentração 30 vezes superior à da proteína, correspondendo, assim, a aproximadamente 30 mM. Estas soluções permaneceram incubadas a 25°C por 30 minutos antes da montagem das placas. Em seguida, gotas de 1 μL , contidas de 0.5 μL de solução de proteína e 0.5 μL de solução de cristalização, foram montadas em placas Swissci (3 gotas por poço, 96 poços) (Hampton Research) utilizando o sistema automático de cristalização Oryx8 (Douglas Instruments), sendo aplicada a técnica de gota sentada mediante o método da matriz esparsa (JANCARIK; KIM, 1991). Diferentes *screens* de cristalização foram utilizados nos ensaios iniciais, sendo eles: JCSG-plus, Morpheus, MIDAS (Molecular Dimensions) e Index screen (Hampton Research). Após a montagem das placas, as mesmas foram seladas e incubadas a 20°C.

As placas montadas foram semanalmente monitoradas com o auxílio de um microscópio, sendo os cristais observados e analisados por meio de um difratômetro laboratorial (Bruker AXS Proteum 6000 X-Ray Diffractometer), acoplado a um detector MAR 345. Antes da análise os cristais foram imersos em solução crioprotetora (glicerol 30%), e congelados instantaneamente com o auxílio de nitrogênio líquido. As condições nas quais foram observados cristais de proteínas foram anotadas para procedimentos de otimização, havendo sido os cristais com resoluções superiores a 3 Å armazenados em nitrogênio líquido para posterior envio à linhas de luz síncrotron.

Diferentes condições foram manualmente preparadas para os ensaios de otimização com base na formulação das condições originais, havendo sido aplicadas variações referentes ao pH, concentração de sal e agente precipitante. Placas de cristalização foram então montadas com o auxílio do sistema automático de cristalização Oryx8, seguindo os mesmos procedimentos iniciais. Dentre os cristais analisados, aqueles de melhor resolução foram armazenados para posterior envio para difração em linhas de luz síncrotron.

4.3 Coleta de dados e processamento inicial

Os conjuntos de dados foram coletados no *European Synchrotron Radiation Facility* (ESRF) (Grenoble, França) nas linhas de luz ID-29 e ID23-2, em uma temperatura de 100 K. Os dados de difração foram indexados e integrados com a utilização do programa XDS (KABSCH, 2010).

4.4 Elucidação das estruturas tridimensionais

Com os dados de difração processados, deu-se então a elucidação da estrutura cristalográfica da pDGL com a utilização de programas computacionais específicos disponíveis no pacote CCP4 (COLLABORATIVE COMPUTATIONAL PROJECT NO. 4, 1994; WINN *et al.*, 2011). Inicialmente, os dados retornados pelo XDS foram analisados quanto à confiança estatística da escolha do grupo espacial. Para tanto, utilizou-se o programa Pointless (EVANS, 2011), o qual sugere uma opção alternativa caso aquele previamente selecionado não esteja correto e já o incorpora em seu arquivo de saída. Em seguida, calculou-se o número de moléculas presentes na unidade assimétrica, e, na sequência, utilizou-se o programa Scala, que a partir da fusão (*merge*) de reflexões (ou pontos) equivalentes presentes em diferentes imagens, faz uma análise estatística do cálculo das intensidades. A confiabilidade desses dados foi dada pela análise de diversos parâmetros, tais como R_{merge} , I/σ , multiplicidade, completeza e coeficiente de correlação ($CC_{1/2}$). O APÊNDICE A traz o significado e o valor esperado para cada um desses índices.

4.4.1 Substituição molecular

Para gerar o modelo tridimensional da pDGL de acordo com os padrões de difração, realizou-se a substituição molecular com o auxílio do programa Phaser-MR (MCCOY *et al.*, 2007), onde utilizou-se como molde inicial as cadeias β e γ da DGL madura (PDB ID 1DGL) para gerar as coordenadas da pDGL. A confiabilidade do modelo gerado foi verificada a partir da análise dos parâmetros TRZ, TFZ e LLG, descritos no APÊNDICE A.

4.4.2 Refinamento do modelo

O modelo tridimensional gerado a partir da substituição molecular foi analisado e modelado através do programa COOT (EMSLEY *et al.*, 2010) para se obter uma maior concordância entre esse modelo e os dados de difração. Nesta etapa, foram realizadas inspeções e correções manuais da cadeia polipeptídica, posicionando e orientando os resíduos de aminoácidos de acordo com o mapa de densidade eletrônica. Em seguida, o modelo já modificado pelos ajustes foi submetido a ciclos de refinamento utilizando máxima verossimilhança com auxílio do programa Refmac5 (MURSHUDOV *et al.*, 2011): inicialmente através de ciclos de corpo rígido seguidos de ciclos por restrição. Novas inspeções e ajustes do modelo foram conduzidas, bem como adição de moléculas de água, íons metálicos e ligantes específicos. Na sequência, novos ciclos de refinamento por restrição foram realizados e ao final de cada um destes a concordância modelo-densidade foi avaliada pelos valores de R_{factor} e R_{free} , índices que representam coeficientes de discordância.

Finalizado o refinamento, a estereoquímica do modelo foi avaliada e validada por ferramentas *on-line* disponíveis na plataforma de validação do *Protein Data Bank* (<http://wwpdb-validation.wwpdb.org/validservice/>), para, então, ser realizado o depósito das coordenadas atômicas e dados de difração nesse banco de dados de estruturas proteicas (BERNSTEIN *et al.*, 1977). As figuras representativas dos modelos foram geradas com o auxílio do programa PyMOL (Schrödinger, LLC).

4.5 Análises *in silico*

Para se realizar uma análise mais completa acerca das características estruturais da pDGL e das bases de sua maior afinidade à manose, realizou-se os ensaios *in silico* descritos a seguir.

4.5.1 Análise do padrão de oligomerização

A predição do estado oligomérico em solução da pDGL foi realizada a partir do servidor *on-line* PISA (*Proteins, Interfaces, Structures and Assemblies*) (KRISSINEL; HENRICK, 2007), cujos resultados foram analisados de acordo com os valores de ΔG^{int} e ΔG^{diss} , parâmetros termodinâmicos que representam, respectivamente, o ganho de energia

livre decorrente do processo de oligomerização e a energia livre necessária para promover a separação dos componentes do oligômero.

4.5.2 Cálculos de ângulos psi (Ψ) e phi (Φ)

Para se investigar a presença de estruturas secundárias não-repetitivas no modelo cristalográfico da pDGL, semelhante ao que fora realizado para a ConM (SILVA-FILHO, 2013), utilizou-se o software Swiss-PdbViewer 4.1 (JOHANSSON *et al.*, 2012) para o cálculo dos valores dos ângulos de torção phi (Φ) e psi (Ψ) de resíduos de aminoácidos específicos (ver seção 5.2).

4.5.3 Simulações de docking molecular e cálculo de dimensões do CRD

Na busca por melhor compreender as bases estruturais que permeiam a maior afinidade da pDGL pela manose, realizaram-se cálculos de *docking* molecular da lectina em suas formas precursora e madura contra essa hexose e o trimanosídeo 3,6-di-O-(α -D-manopiranosil)- α -D-manopiranosose, um ligante cuja DGL apresenta alta afinidade (CHERVENAK; TOONE, 1995; GUPTA *et al.*, 1996).

Para se alcançar esse objetivo, utilizou-se o programa AutoDock Vina 1.1.2 (TROTT; OLSON, 2010) para realizar as simulações dessas interações. Inicialmente, preparou-se os modelos da proteína precursora e madura (PDB ID 1DGL), utilizando-se apenas as cadeias D e A, respectivamente, as quais foram modificadas manualmente para retirada de heteroátomos (moléculas de água, ligantes e íons). Em seguida, utilizou-se a interface gráfica AutoDockTools 1.5.6 para adição de átomos de hidrogênio aos resíduos polares das cadeias polipeptídicas seguida da adição de cargas parciais pelo método de Kollman, o qual atribui valores aos resíduos de aminoácidos de acordo com seus respectivos potenciais elétricos (BEST *et al.*, 2012). Os ligantes, manose e trimanosídeo, foram retirados manualmente de seus complexos cristalográficos com a pDGL e DGL, respectivamente, e tratadas pelo AutoDockTools como moléculas flexíveis. O espaço de busca foi definido por uma caixa 22 Å x 24 Å x 28 Å centralizada no CRD das lectinas. Para cada simulação (pDGL x manose, pDGL x trimanosídeo, DGL x manose e DGL x trimanosídeo), as melhores "poses" foram analisadas de acordo com a energia teórica de ligação, dada em kcal/mol, e a presença de

ligações específicas entre os ligantes e o CRD, quais sejam: (i) interação entre o oxigênio O4 da manose e os átomos Asp99/208² (OD1), Arg119/228 (N) e Asn157/14 (ND2); e (ii) oxigênio O6 da manose com Asp99/208 (OD2) (SHARMA; SUROLIA, 1997).

Para validar essa metodologia, calculou-se o RMSD (*root mean square deviation*) da melhor pose na simulação DGL x trimanosídeo (*redocking*) com o trissacarídeo do modelo cristalográfico (PDB ID 1DGL). Considerou-se que uma solução aceitável seria dada por um RMSD com valor máximo de 2.0 Å (JONES *et al.*, 1997). O cálculo de RMSD foi realizado pelo *software* UCSF Chimera (PETTERSEN *et al.*, 2004).

Além disso, verificou-se se as dimensões do CRD, nas formas precursora e madura, poderiam ter alguma influência sobre suas afinidades para os ligantes analisados nas simulações de *docking*. Nesse caso, utilizou-se o servidor *on-line* DogSiteScorer (VOLKAMER *et al.*, 2012) para calcular o volume (Å³), a área superficial (Å²) e a profundidade (Å) dos sítios.

² Primeiro número indica a pDGL e o seguinte a posição correspondente na lectina madura, DGL.

5 RESULTADOS

5.1 Cristalização e Difração de Raios X

A forma recombinante da pDGL, submetida ao processo de co-cristalização com uma solução 30 mM de D-manose, apresentou cristais a partir da otimização da condição de número 80 do Index Screen da Hampton Research®: acetato de amônia 1 M, PEG (polietilenoglicol) 3350 25%, monometil éter 2000 25%, HEPES 250 mM e pH 7,5. Após a formação dos cristais e seu subsequente processo de difração, obteve-se um conjunto de dados que foram analisados *in silico*, sendo observado que os cristais da pDGL-manose apresentaram-se sob o grupo espacial monoclinico P12₁1, com dimensões de célula unitária de $a = 71,722$, $b = 96,609$ e $c = 72,130$. O coeficiente de Matthews resultou em um valor de $2.21 \text{ \AA}^3 \text{ Da}^{-1}$, indicando a presença de quatro moléculas na unidade assimétrica, com 44,36% de conteúdo de solvente. Dados acerca do escalonamento, substituição molecular e refinamento estão sumarizados na Tabela 3 e referentes à validação estão no Apêndice B.

5.2 Estrutura Geral da pDGL

As quatro moléculas da pDGL presentes na unidade assimétrica estão arranjadas como o característico tetrâmero das lectinas *ConA-like*, uma associação de dois dímeros canônicos (interface tipo II) estabilizada por interações entre resíduos que formam uma interface X2. A Figura 8 mostra os elementos estruturais que caracterizam os *fold*s monoméricos, diméricos e tetraméricos da pDGL.

Todas as quatro cadeias polipeptídicas presentes na unidade assimétrica não apresentaram densidade eletrônica para seus 12 primeiros resíduos N-terminais, sendo, assim, os modelos de suas estruturas tridimensionais iniciadas pelo resíduo Asp13. Em três dos monômeros (cadeias A, B e C) não foi possível modelar o segmento de 15 resíduos (entre as posições 129 e 143) que durante o processamento pós-traducional das lectinas *Diocleinae* é removido para dar origem às novas extremidades N- e C-terminais na forma madura. Na cadeia D, esse *loop* foi modelado de maneira quase completa, excetuando-se as posições 131-134 que não apresentaram densidade eletrônica para suas respectivas cadeias laterais e assim foram modeladas como resíduos de glicina.

Tabela 3 - Dados estatísticos da estrutura cristalográfica da pDGL complexada com manose.

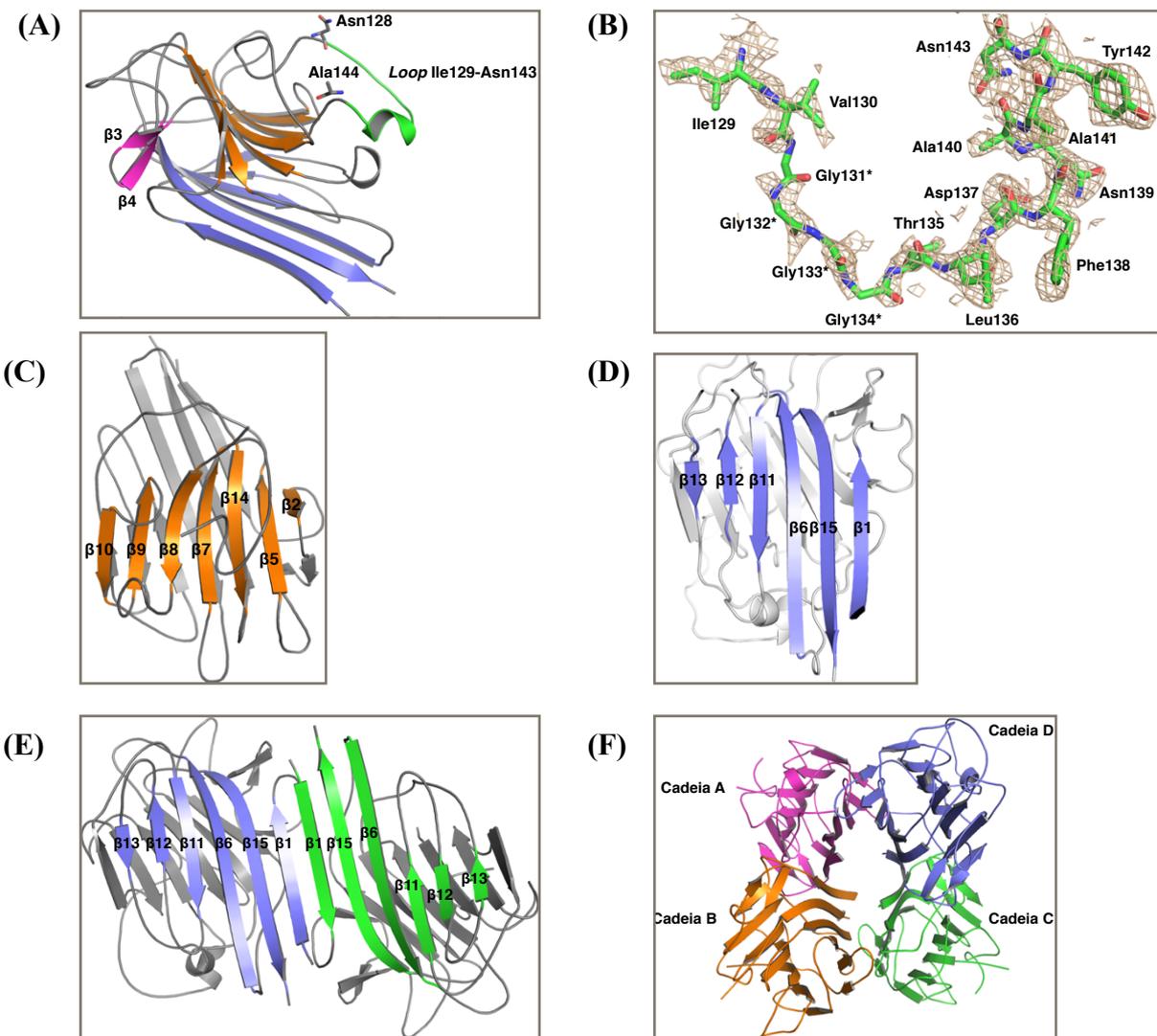
Parâmetros	Valores
Coleta dos dados	
Número total de observações	122.540
Número total de observações únicas	45.938
R_{merge} (%)	4,4 (11,5)
Limite de resolução (Å)	48,30-2,20 (2,32-2,20)
$CC_{1/2}$	99,8 (97,9)
Completeza (%)	93,1 (95,3)
Multiplicidade	2,7 (2,6)
(I)/ σ	16,6 (8,0)
Comprimento de onda (Å)	1,033
Grupo espacial	P12 ₁
Parâmetros de cela (Å)	a = 71,722 b = 96,609 c = 72,130
Substituição Molecular	
RFZ	10,8
TFZ	12,9
LLG	554
Refinamento	
Limite de resolução (Å)	71,42-2,20
R_{factor} (%)	16,39
R_{free} (%)	21,92
Número de aminoácidos na unidade assimétrica	1044
Número de moléculas de água	586
RMSDs dos valores ideais	
Comprimento das ligações (Å)	0,016
Ângulos das ligações (graus)	1,837
Fatores de temperature	
Média do B-fator para toda a proteína (Å)	20,95
Gráfico de Ramachandran	
Resíduos nas regiões mais favorecidas (%)	96,96
Resíduos nas regiões permitidas (%)	2,93
Resíduos nas regiões não-permitidas (%)	0,11

Fonte: Elaborada pelo autor.

Valores entre parênteses correspondem àqueles calculados para a última camada.

Cada monômero apresenta o enovelamento em β -sanduíche, com sete fitas anteriores sobrepondo-se a seis fitas posteriores. Esses dois arranjos em folha- β são conectados entre si por diversos segmentos em *loop* e uma pequena estrutura secundária formada pelas fitas $\beta 3$ e $\beta 4$ compostas pelos resíduos Phe38-Asp40 e Gly43-Gln46, respectivamente. Além disso, em cada cadeia polipeptídica há a presença dos íons divalentes Ca^{2+} e Mn^{2+} , necessários há estruturação do domínio de reconhecimento a carboidrato (ver seção 5.3).

Figura 8 - Características gerais da estrutura cristalográfica da pDGL.



Fonte: Elaborada pelo autor.

(A) Monômero da pDGL, destacando as folhas- β anteriores (laranja) e posteriores (azul), que são conectadas por diversos *loops* e pelas fitas 3 e 4 (magenta). Também é mostrado o *loop* Ile129-Asn143 (verde) que após retirado faz com que os resíduos Asn128 e Ala144 passem a ser, respectivamente, as extremidades C- e N-terminal na lectina madura. (B) Omit map (1.0σ) do *loop* Ile129-Asn143, onde se observa que entre as posições 131 e 134 (*) não há densidade para cadeia lateral, sendo, por isso, modeladas como resíduos de Gly. (C) e (D) Elementos formadores das folhas- β anteriores e posteriores, respectivamente. (E) e (F) Arranjos diméricos e tetraméricos da pDGL. RMSD dos arranjos monoméricos e tetraméricos da pDGL em relação à sua forma madura (PDB ID 2JE9) foram de 0.50 Å (valor calculado para as cadeias D e A, respectivamente) e 0.99 Å, respectivamente.

Uma análise do padrão de ligações de hidrogênio existentes na formação de estruturas secundárias demonstrou a presença de elementos estruturais reportados por Silva-Filho (2013) para a ConM. Estes elementos são chamados de estruturas secundárias não-repetitivas, onde, diferentemente do que ocorre com α -hélices e folhas- β , cujos resíduos formadores possuem aproximadamente os mesmos valores para os ângulos Φ (phi) e Ψ (psi), cada resíduo apresenta um valor característico para estes ângulos de rotação (RICHARDSON, 1981).

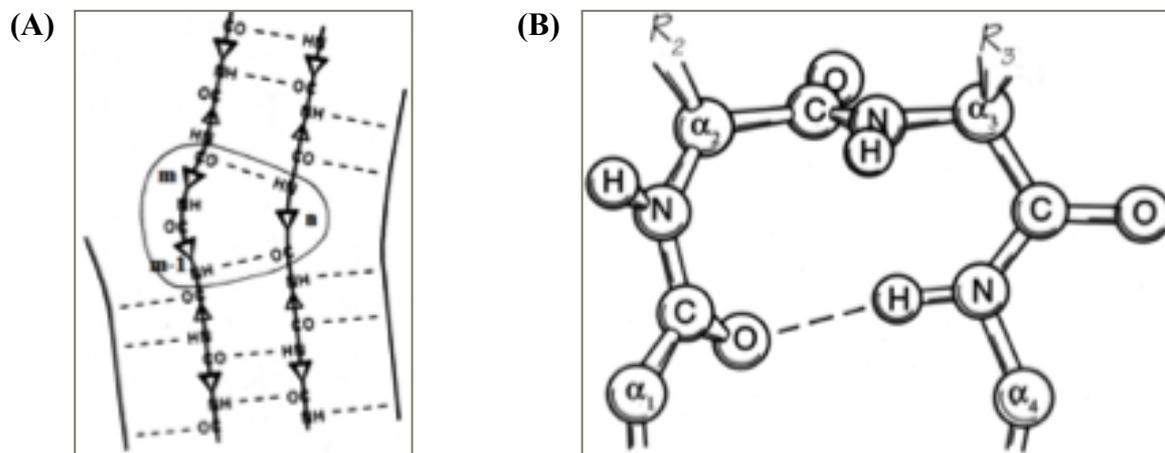
Estruturas secundárias não-repetitivas promovem uma mudança na direção da cadeia polipeptídica, podendo ser agrupadas em dois grupos: *tight-turns* e β -*bulges*. As primeiras são definidas como sendo motivos estruturais onde os átomos de carbono $C\alpha$ de dois resíduos separados por 1-5 unidades peptídicas (ou ligações peptídicas) encontram-se a uma distância de menos 7 Å, com os ângulos Φ e Ψ não sendo necessariamente os mesmos para todos os resíduos. Por sua vez, β -*bulges* rompem a regularidade de folhas- β , sendo mais facilmente identificadas em folhas- β antiparalelas. Esta estrutura é definida como sendo uma região presente entre duas ligações de hidrogênio que estabilizam duas fitas sucessivas e que inclui um resíduo de uma fita e dois da outra fita (RICHARDSON, 1981) (FIGURA 9A).

Segundo Kabasch e Sander (1983), *n-turn* é um segmento de estrutura secundária onde o grupo carbonil de um resíduo, arbitrariamente denominado “i”, perfaz uma ligação de hidrogênio com o grupo -NH do resíduo “i+n”, onde $n = 2, 3, 4$ ou 5 . Neste contexto, *tight-turns*, descritas inicialmente para um sistema de três unidades peptídicas [*3-turn*] (VENKATACHALAN, 1968), onde o grupo CO do primeiro resíduo (i) interage com o grupo -NH do quarto (i+3) (FIGURA 9B), envolvem, além deste tipo, denominado também de β -*turn*, aqueles descritos como γ -*turn* (MILNER-WHITE, 1990) e π -*turn* (RAJASHANKAR; RAMAKUMAR, 1996), os quais correspondem a *2-turns* e *5-turns*, respectivamente. *4-turns*, também chamadas de α -*turns*, apresentam ângulos de torção Φ e Ψ iguais para todos os três resíduos posicionados entre os resíduos “i” e “i+4”, de modo que este tipo de segmento participa da formação de α -hélices, estruturas secundárias ordenadas e repetitivas (BARLOW; THORNTON, 1988; RICHARDSON, 1981), embora também possam ser encontradas isoladamente, sendo estabilizadas por uma única ligação de hidrogênio e apresentando valores diferentes para Φ e Ψ nos três resíduos centrais (PAVONE *et al.*, 1996).

Dentre os quatro tipos de *tight-turns*, as “ β ” merecem destaque especial devido sua maior abundância nas estruturas tridimensionais de proteínas globulares (HUTCHINSON; THORNTON, 1994). Além disso, são caracterizadas por estarem divididas em subclasses de

acordo com os valores preditos como ideais para os ângulos Φ e Ψ dos resíduos “ $i+1$ ” e “ $i+2$ ” (TABELA 4). A análise estrutural mostrou que os monômeros da pDGL contêm diversas estruturas secundárias não repetitivas do tipo β -turn, embora segmentos que formem esses motivos estruturais em uma cadeia não necessariamente as formem em outro monômero. Por exemplo, os resíduos Thr158-Gly161 formam uma β -turn nas cadeias A e B mas não nas cadeias C e D, como mostrado na Tabela 5, que lista as β -turn identificadas na pDGL quanto aos resíduos que as formam e distâncias entre o grupo $-\text{CO}$ do primeiro resíduo (“ i ”) e o $-\text{NH}$ do último (“ $i+3$ ”). Na Tabela 6 são listados os valores dos ângulos Φ e Ψ dos resíduos “ $i+1$ ” e “ $i+2$ ” que categorizam as subclasses dessas estruturas secundárias na pDGL. Interessantemente, algumas β -turns aparecem de forma consecutiva, sendo representadas como hélices- 3_{10}^3 (FIGURA 10A) apesar de seus resíduos formadores “ $i+1$ ” e “ $i+2$ ” não possuírem valores iguais de Φ e Ψ (TABELA 6). Por outro lado, a hélices- 3_{10} presente no *loop* II129-Asn143 da cadeia D (FIGURA 10B) pode, de fato, ser assim chamada, já que seus resíduos constituintes apresentam ângulos de torção dentro do esperado.

Figura 9 - Representações esquemáticas de estruturas secundárias não repetitivas.



Fonte: Adaptado de Richardson (1981).

(A) Ilustração de uma β -bulge do tipo clássica formada entre fitas consecutivas de uma folha- β antiparalela. Observa-se que o padrão de ligações de hidrogênio é rompida pela realização de uma interação entre o grupo carbonil do resíduo “ n ” de uma fita e o grupo $-\text{NH}$ do resíduo “ $m-1$ ” da fita adjacente (em uma condição regular, a interação ocorreria com o $-\text{NH}$ do resíduo m , como visualizado nos padrões que flanqueiam a β -bulge). (B) Ilustração de uma *tight-turn*, evidenciando a mudança de direção tomada pela cadeia polipeptídica. É destacada ainda a ligação de hidrogênio que conecta o grupo carbonil do resíduo “ i ” (α_1) e o grupo $-\text{NH}$ do resíduo “ $i+3$ ” (α_4). Esferas contendo “ α ” representam os átomos de carbono $\text{C}\alpha$. R_2 e R_3 representam as cadeias laterais dos resíduos “ $i+1$ ” e “ $i+2$ ”, respectivamente.

³Hélices- 3_{10} são formadas por β -turns do tipo III (TABELA 4). Tal nomenclatura, “ 3_{10} ”, refere-se ao fato de que entre a ligação de hidrogênio ($i, i+3$) existem três unidades peptídicas, onde 10 átomos estão presentes no anel formado por tal interação.

Tabela 4 - Classificação das β -turns.

Tipo	Φ	Ψ	Φ	Ψ
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
III	-60	-30	-60	-30
IV	-	-	-	-
V	-80	80	80	-80
V'	80	-80	-80	80
Via	-60	120	-90	0
VIb	-135	135	-75	160
VII	-	-	-	-

Fonte: Adaptado de Richardson (1981).

Valores em graus.

^a Tipos I', II' e V' são imagens espaciais de I, II e V, respectivamente. Ou seja, os valores são os mesmos, trocando-se apenas o sinal.

^b Tipo especial de β -turn com caráter helicoidal (hélices-3₁₀), devido conservação dos valores de Φ e Ψ para ambos os resíduos.

^c Qualquer β -turn cujos resíduos apresentem valores de Φ e Ψ que difiram mais de 40° em relação aqueles reportados como ideais para os demais tipos.

^d Reportados em algumas estruturas, mas com contínuo debate sobre sua categorização como grupos separados. No caso do tipo VII, sua definição leva em consideração apenas dois ângulos: $\Phi_{i+2} = 180^\circ$ e $\Psi_{i+1} < 60^\circ$ ou $\Phi_{i+2} < 60^\circ$ e $\Psi_{i+1} = 180^\circ$.

^e Presença de uma cis-prolina na posição "i+2". Diferença entre subtipos "a" e "b" de acordo com conformação deste resíduo.

Tabela 5 - β -turns identificadas nos monômeros da estrutura cristalográfica da pDGL.

Sequência	A	B	C	D
Pro28-Leu31	3,56	3,02	3,26	2,99
Gln34-Ala37	3,45	3,49	3,32	3,41
Ser51-Gly54	2,65	2,77	2,76	2,86
Glu74-Ala77	2,91	3,21	2,76	2,99
Asn107-Thr110	3,27	2,93	3,09	2,91
Gly117-Leu120	3,21	3,08	3,18	3,03
Gly118-Leu121	2,89	2,96	2,89	3,01
Asp137-Ala140				2,79
Phe138-Ala141				3,36
Asn139-Tyr142				3,29
Ala140-Asn143				3,08
Asn157-Ile160	3,01	3,02	2,92	2,90
Thr158-Gly161	3,18	3,17	3,66	3,65
Glu186-Lys189	3,02	3,07	3,02	3,08
Tyr210-Ser213	3,33	3,17	2,97	3,24
Asp223-Asn226	2,91	3,00	2,88	3,04
Leu224-Val227	3,03	3,02	3,10	3,03

Fonte: Elaborada pelo autor.

Valores referem-se às distâncias (Å) entre o grupo –CO do primeiro resíduo (“i”) e o –NH do último (“i+3”) em cada segmento nas quatro cadeias polipeptídicas do modelo cristalográfico da pDGL.

^a Valores que não correspondem à formação de uma interação de hidrogênio e, conseqüentemente, indicam a não formação das respectivas β -turns.

^b Valores que indicam a formação de β -turns consecutivas, as quais são representadas como hélices-3₁₀.

^c Sequências de resíduos presentes no *loop* Ile129-Asn143, o qual foi modelado apenas na cadeia D.

Tabela 6 - Tipos de β -turns identificadas na estrutura cristalográfica da pDGL.

Sequência	A			B			C			D										
	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}								
P28-L31	-77.5	-5.6	-107.1	17.3	I	-67.3	-14.5	-97.9	14.4	I	-72.9	-14.5	-95.1	14.0	I	-61.6	-14.9	-101.1	15.6	I
Q34-A37	63.2	-134.0	-87.4	5.1	II'	67.0	-134.7	-85.7	-0.1	II'	57.7	-125.6	-88.3	3.6	II'	67.1	-133.7	-89.0	5.8	II'
S51-G54	-28.7	-32.4	-90.5	1.0	I	-61.0	-38.9	-62.1	-23.1	III	-48.8	-42.7	-77.7	14.4	III	-56.7	-34.6	-73.0	-12.0	III
E74-A77	-54.0	-41.2	-85.5	10.7	I	-49.4	-57.1	-67.6	-12.0	I	-60.0	-31.6	-98.4	27.1	I	-48.0	-55.5	-88.6	23.6	I
N107-T110	-64.6	-27.2	-99.0	24.1	I	-52.2	-34.9	-90.8	19.1	I	-67.4	-25.6	-95.5	17.2	I	-55.0	-31.0	-94.7	23.2	I
G117-L120 ^b	44.8	-120.4	-53.6	-18.4	II'	56.5	-124.0	-50.9	-20.3	II'	40.5	-113.3	-59.1	-19.1	II'	46.7	-117.3	-63.4	-15.6	II'
G118-L121 ^b	-53.6	-18.4	-100.1	4.8	I	-50.9	-20.3	-93.2	2.1	I	-59.1	-19.1	-94.4	3.8	I	-63.4	-15.6	-91.7	5.2	I
D137-A140 ^a																-56.4	-27.2	-68.9	-23.4	III
F138-A141 ^a																-68.9	-23.4	-58.0	-26.5	III
N139-Y142 ^a																-58.0	-26.5	-52.4	-26.1	III
A140-N143 ^{a,b}																-52.4	-26.1	-85.7	-11.7	I
N157-II60 ^b	-60.5	-27.3	-66.2	-32.3	III	-63.6	-24.0	-69.3	-19.9	III	-61.7	-22.5	-63.0	-21.1	III	-60.6	-14.0	-60.7	-44.3	III
T158-G161 ^b	-66.2	-32.3	-103.0	13.0	I	-69.3	-19.9	-111.7	25.6	I	-63.0	-21.1	-128.0	38.5	I	-60.7	-44.3	-118.6	32.7	I
E186-K189	-48.3	136.4	84.6	-2.7	II	-54.2	143.3	89.3	-15.4	II	-56.3	137.2	88.6	-5.6	II	-53.4	133.7	91.0	-0.8	II
Y210-S213	-61.3	121.3	84.7	3.4	II	-52.1	119.9	74.7	21.3	II	-60.7	132.4	81.2	-8.5	II	-61.1	131.6	69.3	11.1	II
D223-N226 ^b	-63.3	-27.0	-69.4	-14.3	III	-66.2	-18.2	-63.6	-20.1	III	-65.5	-24.9	-69.6	-17.0	III	-64.4	-16.0	-70.4	-20.4	III
L224-V227 ^b	-69.4	-14.3	-91.4	8.8	I	-63.6	-20.1	-92.4	0.1	I	-69.6	-17.0	-89.7	11.1	I	-70.4	-20.4	-97.0	3.8	I

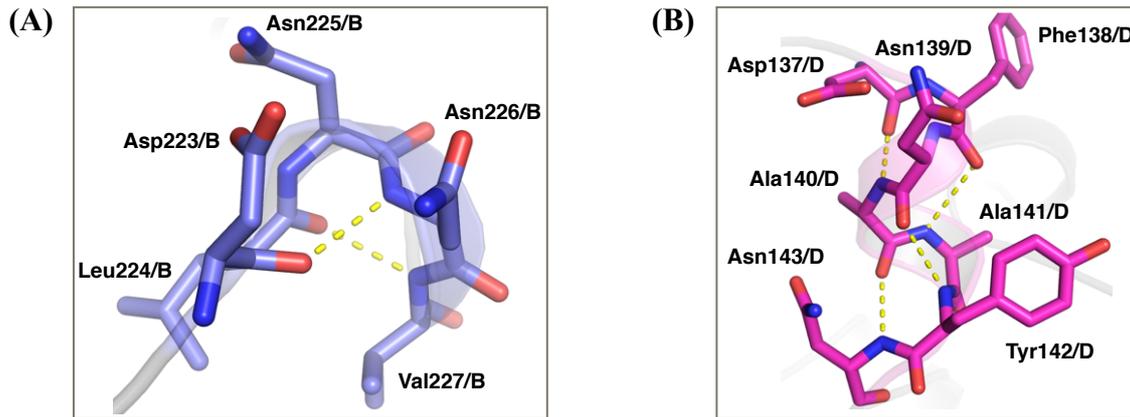
Fonte: Elaborada pelo autor.

As β -turns estão categorizadas (números romanos) de acordo com os valores dos ângulos Φ e Ψ (dados em graus) para os resíduos $i+1$ e $i+2$ (ver texto) em cada uma das cadeias.

^a β -turns presentes no *loop* Ile129-Asn143, modelado apenas na cadeia D.

^b β -turns consecutivas que são representadas como hélices-3₁₀ mas que são de tipos diferentes daquele considerado ideal para formar tais estruturas helicoidais (ver TABELA 4).

Figura 10 - Exemplos de β -turns encontradas no modelo cristalográfico da pDGL.



Fonte: Elaborada pelo autor.

(A) β -turns formadas pelos segmentos Asp223-Asn226 e Leu224-Val227 na cadeia B, destacando as interações de hidrogênio (traçado amarelo) entre os resíduos "i" (Asp223 e Leu224, respectivamente) e "i+3" (Asn226 e Val227, respectivamente). Devido à formação de maneira consecutiva, essa estrutura secundária é representada como uma hélice- 3_{10} , apesar de seus ângulos de torção Φ e Ψ não serem os ideais (TABELAS 4 e 6). (B) Verdadeira hélice- 3_{10} formada pelos segmentos consecutivos Asp137-Ala140, Phe138-Ala141, Asn139-Tyr142 e Ala140-Asn143 encontrados no *loop* Ile129-Asn143.

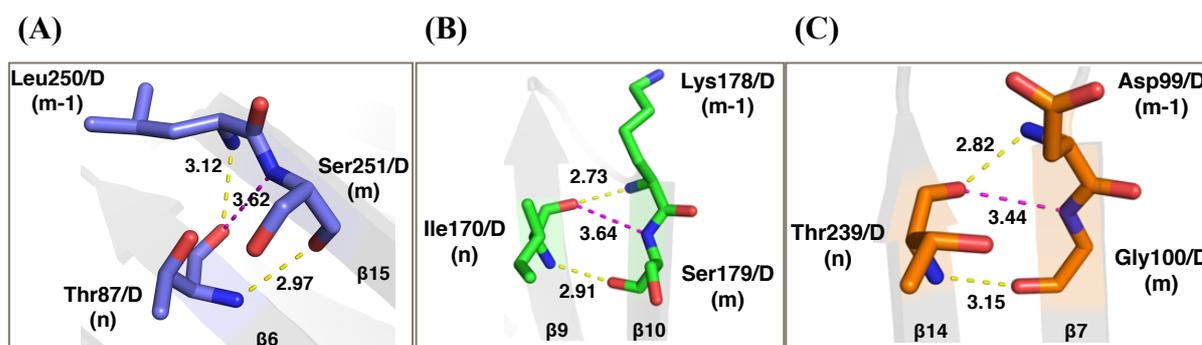
Outras estruturas secundárias não-repetitivas encontradas no modelo cristalográfico da pDGL foram as chamadas β -bulges. Em sua forma clássica, esse tipo de arranjo ocorre quando a carbonila de um resíduo de aminoácido (n) em uma fita atua como aceptora de hidrogênio para o grupo -NH do resíduo "m-1" da fita adjacente, podendo ou não também aceitar um hidrogênio do resíduo "m" (CHAN *et al.*, 1993). No primeiro caso, esta forma clássica é denominada de C+ e a ligação de hidrogênio entre -CO (n) e -NH (m) é mais fraca que aquela performada entre "n" e "m-1" devido à distância entre os átomos doador e receptor (FIGURA 9A, acima). Já no segundo caso, a interação entre "n" e "m" não existe e a β -bulge clássica é denominada C-. Em uma folha- β regular, -CO (n) interage apenas com o -NH do resíduo "m" (KABSCH; SANDER, 1983).

Diferentemente do que ocorre com a β -bulge clássica, onde o resíduo "m-1" encontra-se dentro da fita, a subclasse G1 apresenta este resíduo posicionado ou no início de uma fita ou no final de um *loop*, onde, na maioria das vezes, este resíduo é uma glicina. Quando não o é, a β -bulge é denominada G1A (CHAN *et al.*, 1993).

A Figura 11 ilustra três β -bulges presentes nos monômeros da pDGL. A primeira envolve o resíduo Thr87 (n) da fita β_6 e os resíduos Leu250 (m-1) e Ser251 (m) da fita β_{15} . A segunda β -bulge, por sua vez, ocorre entre o resíduo Ile170 (n) da fita β_9 e os resíduos Lys178

(m-1) e Ser179 (m) da fita β 10. Já a terceira estrutura não repetitiva deste tipo é formada entre o resíduo Thr239 (n) da fita β 14 e os resíduos Asp99 (m-1) e Gly100 (m) da fita β 7. A Tabela 7 lista essas β -bulges quanto às distâncias entre seus resíduos formadores em cada um dos monômeros da pDGL.

Figura 11 - Exemplos de β -bulges encontradas no modelo cristalográfico da pDGL.



Fonte: Elaborada pelo autor.

(A), (B) e (C) Correspondem, respectivamente, às estruturas formadas pela trinca de resíduos Thr87-Leu250-Ser251, Ile170-Lys178-Ser179 e Thr239-Asp99-Gly100 na cadeia D. Tracejados amarelos representam interações de hidrogênio entre $-NH$ (n) e $-CO$ (m) e entre $-CO$ (n) e $-NH$ (m-1). Tracejados magentas representam as distâncias entre $-CO$ (n) e $-NH$ (m), mostrando que, de acordo com seus valores, ligações de hidrogênio podem ser formadas (C) ou não (A e B). Devido à localização e identidade dos resíduos “m-1”, todas as β -bulges da pDGL são do tipo G1A.

Tabela 7 - Distâncias (Å) entre os elementos que formam as β -bulges em cada monômero do modelo cristalográfico da pDGL.

β -bulges	A	B	C	D
Thr87-Leu250-Ser251	3,06 / 2,88 / 3,62	3,06 / 2,93 / 3,68	3,14 / 3,02 / 3,70	2,97 / 3,12 / 3,62
Ile170-Lys178-Ser179	2,95 / 2,73 / 3,72	2,97 / 2,77 / 3,73	2,98 / 2,82 / 3,68	2,91 / 2,73 / 3,64
Thr239-Asp99-Gly100	3,18 / 2,88 / 3,54	3,27 / 2,84 / 3,48	3,22 / 2,79 / 3,56	3,15 / 2,82 / 3,44

Fonte: Elaborada pelo autor.

Para cada relação monômero: β -bulge os valores representam as distâncias entre $-NH$ (n) e $-CO$ (m), $-CO$ (n) e $-NH$ (m-1) e entre $-CO$ (n) e $-NH$ (m), respectivamente.

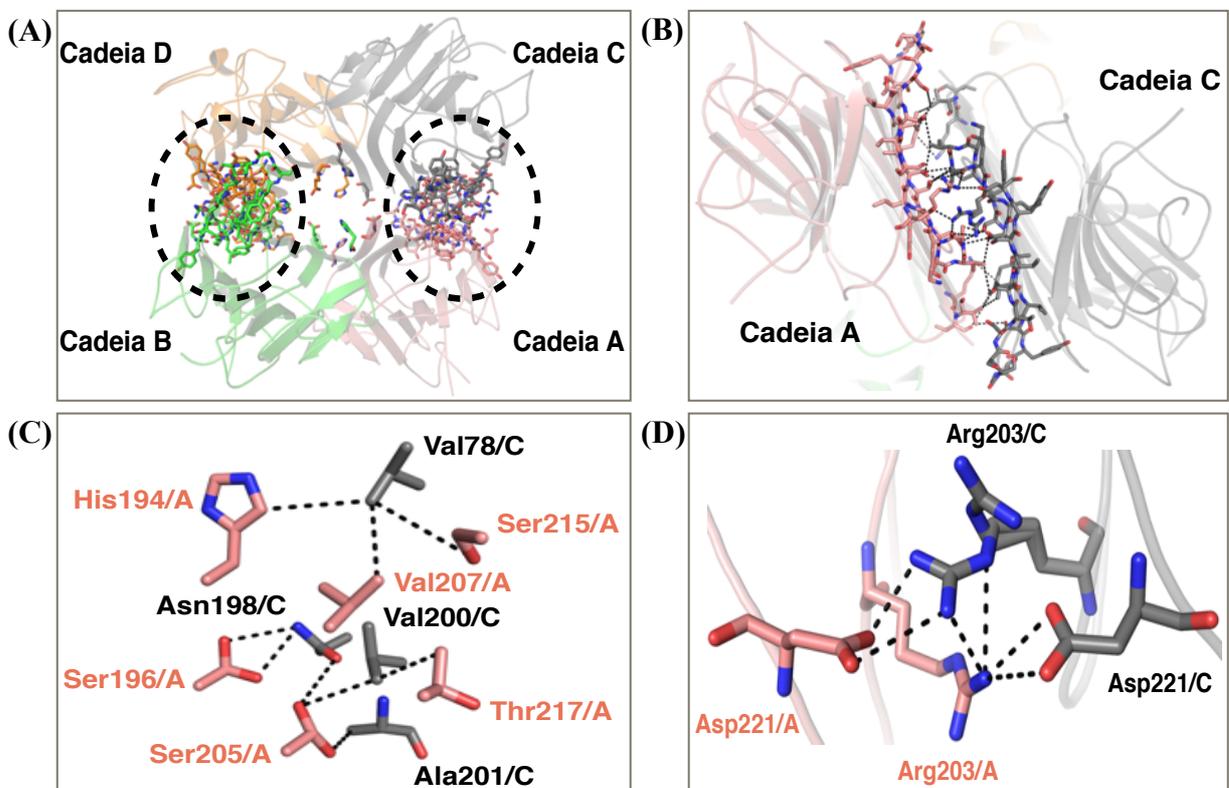
5.2.1 Tetrâmero da pDGL

Como já mencionado, as quatro cadeias polipeptídicas da pDGL presentes na unidade assimétrica estão arranjadas como o clássico tetrâmero das lectinas *Diocleinae*. Esse arranjo foi predito pelo PISA como sendo o estado de oligomerização mais provável em solução, embora seus parâmetros termodinâmicos sejam bem inferiores ao de sua contra-parte madura: para a pDGL, ΔG^{int} e ΔG^{diss} foram de -27.6 e 7.4 kcal/mol, respectivamente, enquanto que

para a DGL esses valores foram de -225.2 e 20.2, respectivamente. Isso indica que a formação dessa estrutura quaternária pela pDGL apresenta pouca estabilidade (ver seção 6.3).

Essa associação oligomérica ocorre apenas por interações periféricas, realizadas entre resíduos que constituem as interfaces X2 (cadeia A interagindo com cadeia C e cadeia B interagindo com cadeia D), como observado na Figura 12 e Tabela 8. Esses contatos interdiméricos ocorrem entre regiões compostas pelos resíduos homólogos 196-217 de monômeros opostos (cadeia A interage com C e B interage com D). Essa região da pDGL corresponde aos resíduos 53-78 da DGL.

Figura 12 - Interações periféricas que estabilizam o tetrâmero da pDGL.



Fonte: Elaborada pelo autor.

(A) Tetrâmero da pDGL, sendo destacados os resíduos (pontilhados) que compõem as interfaces X2. Ao centro são mostrados os resíduos de His22 que não participam da estabilização do tetrâmero. (B) Detalhe de uma das interfaces mostradas em (A). (C) e (D) Interações entre os resíduos. O mesmo padrão de interações são vistos nas posições correspondentes das duas interfaces.

Tabela 8 - Interações entre as cadeias A e C em uma das interfaces X2.

Cadeia A	Cadeia C	Distância (Å)	Cadeia A	Cadeia C	Distância (Å)
Val78 CG1	His194 CD2	4.63	Ser205 OG, A	Asn196 CG	4.38
	Val207 CG1	4.06		Asn196 ND2	4.67
	Val207 CG2	3.71		Asn196 OD1	3.68
	Ser215 CB	4.31		Val200 CB	3.71
Val78 CG2	Ser209 OG	3.95		Val200 CG1	3.79
	Ser215 CB	4.75		Val200 CG2	4.65
His194 CD2	Val78 CG1	4.25	Ser205 OG, B	Val200 CB	4.63
His194 CE1	Lys259 NZ	4.53		Val200 CG1	4.32
	Lys259 CE	4.64		Ala201 CA	3.30
His194 ND1	Lys259 NZ	4.22		Ala201 CB	3.32
Ser196 OG,A	Asn198 ND2	3.38	Ser205 CB	Val200 CB	4.13
Ser196 OG,B	Asn198 ND2	3.21		Val200 CG1	3.75
	Asn198 OD1	3.60	Val207 CG2	Val78 CG1	3.61
Asn198 ND2	Ser196 OG,B	3.32	Ser209 OG	Val78 CG1	3.78
Val200 CG1	Ser205 OG,A	3.84		Val78 CG2	3.65
	Ser205 OG,B	4.27		Val78 CB	4.32
	Thr217 CG2	4.01		Asn261 ND2	4.64
Val200 CG2	Thr217 CG2	4.64		Asn261 OD1	3.52
Ala201 CA	Ser205 OG,B	3.35		Asn261 CB	4.31
Ala201 CB	Ser205 OG,B	3.31	Ser209 CB	Asn261 ND2	4.31
Ala201 O	Arg203 NH1, A	2.65	Thr217 CB	Val200 CG1	3.83
	Arg203 NE, B	2.92	Thr217 CG2	Val200 CG1	3.84
	Arg203 NH2, B	4.10		Val200 CG2	4.63
Arg203 NE	Ala201 O	3.24	Asp221 OD1	Arg203 NH1, A	2.88
Arg203 NH1	Ala201 O	3.64		Arg203 NH2, A	3.24
Arg203 CZ	Ala201 O	3.27		Arg203 CZ, A	3.43
Arg203 NH1	Arg203 NH2, A	2.74	Asp221 OD2	Arg203 NH1, A	3.99
	Arg203 NE, A	3.48		Arg203 NH2, A	3.40
	Arg203 CZ, A	3.26		Arg203 CZ, A	4.17
	Asp221 OD1	3.29	Lys259 NZ	His194 CE1	4.87
	Asp221 OD2	3.49		His194 ND1	4.60

Fonte: Elaborada pelo autor.

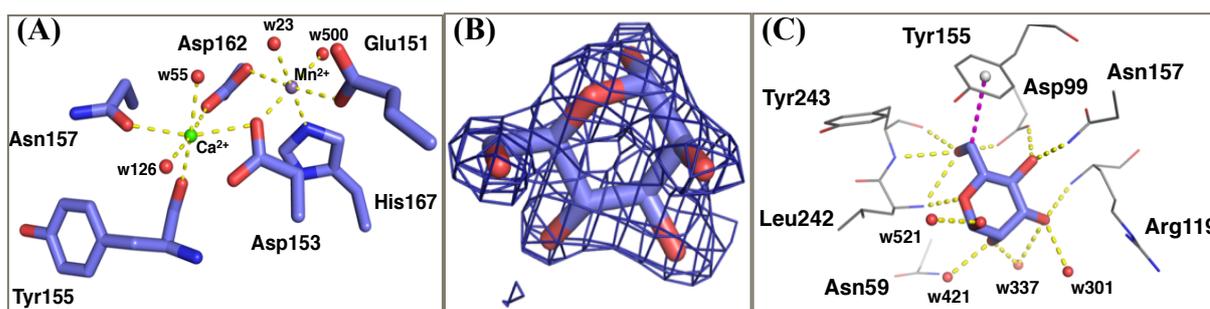
Esse mesmo padrão de interações acontece para a outra interface X2, envolvendo os monômeros B e D.

5.3 Ligação da manose à pDGL

A análise estrutural do complexo pDGL-manose revelou que a presença dos íons bivalentes Ca^{2+} e Mn^{2+} (FIGURA 13A) promoveu a formação de um CRD estruturado nas quatro cadeias do tetrâmero, onde o ligante encontra-se estabilizado (FIGURA 13B). A manose performa interações de hidrogênio com os resíduos Asp99, Arg119, Asn157, Leu242 e Tyr243, além de interagir com três moléculas de água estruturais que medeiam a interação de sua hidroxila OH-2 com os resíduos Asn59 e Gly118 (FIGURA 13C e TABELA 9).

Além dessas ligações, a manose parece também ser estabilizada adicionalmente por uma interação de hidrogênio não-convencional do tipo $-\text{CH} \cdots \pi$, onde seu átomo de carbono C6 interage com o centro do anel benzeno da Tyr155.

Figura 13 - Estabilização da manose.



Fonte: Elaborada pelo autor.

(A) Coordenação dos íons divalentes Ca^{2+} e Mn^{2+} necessários à estruturação do CRD. Note-se que esses íons são estabilizados da mesma maneira que na lectina madura. (B) Omit map (1.4 σ) da manose. (C) Interações que estabilizam a manose no CRD da pDGL. Tracejado magenta representa a interação não-convencional $-\text{CH} \cdots \pi$ entre o C6 do açúcar e o anel benzeno da Tyr155. Esse padrão de interação se repete nas demais cadeias da pDGL, excetuando-se que na cadeia A não há molécula de água correspondente à w521.

5.4 Interação da pDGL com o trimanosídeo e da DGL com manose

Para melhor compreender as bases estruturais que fazem com que a pDGL apresente maior afinidade pela manose, realizou-se quatro ensaios de docking molecular, quais sejam, pDGL x manose, pDGL x trimanosídeo, DGL x manose, DGL x trimanosídeo. Para validar a metodologia utilizada, calculou-se o RMSD entre a melhor pose da simulação DGL x trimanosídeo com o modelo cristalográfico desse trissacarídeo, obtendo-se um valor de 1.22 Å.

A Tabela 10 lista as afinidades de ligação, dadas em Kcal/mol, para as melhores poses de cada simulação. Observa-se que a pDGL apresentou maior afinidade para a manose e afinidade semelhante para o trimanosídeo. No caso da manose, as melhores poses para a lectina precursora e madura possuem pouca variabilidade entre si (FIGURA 14), apresentando um RMSD de apenas 0.2 Å. Quando se compara essas poses com a estrutura cristalográfica, obtém-se valores ligeiramente maiores: 0.62 Å para pDGL x manose e 0.72 Å para DGL x manose. A Tabela 11 lista as interações existentes nas duas simulações com manose.

Tabela 9 - Distâncias (Å) entre os átomos da manose e dos resíduos que formam os CRDs nas quatro cadeias da pDGL.

Manose	Resíduo	A	B	C	D
O1	Água 1	-	2.64	2.87	2.89
O2	Água 2	2.89	2.71	2.65	2.78
	Água 3	2.27	2.64	2.79	2.51
O3	Água 3	2.94	3.23	3.18	3.35
	Água 4	2.87	2.55	2.88	2.82
	Arg119 N	2.90	3.05	3.02	2.87
O4	Asp99 OD1	2.47	2.75	2.48	2.56
	Arg119 N	2.74	3.21	3.41	3.41
	Asn157 ND2	3.34	2.95	2.94	3.00
O5	Leu242 N	3.01	2.85	3.07	3.01
O6	Asp99 OD2	2.89	2.93	2.88	2.82
	Leu242 N	3.23	3.00	3.21	3.18
	Tyr243 N	3.00	3.05	3.05	3.10
	Tyr243 O	3.47	3.25	3.58	3.25
C6	Tyr155 π	3.80	3.90	3.80	3.70

Fonte: Elaborada pelo autor.

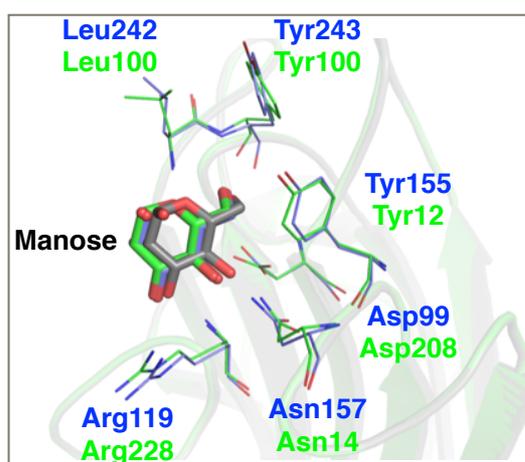
^a Moléculas de água arbitrariamente chamadas de 1, 2, 3 e 4 correspondem, respectivamente, às moléculas w410, w454, w114 e w99 na cadeia B; w302, w477, w156 e w93 na cadeia C e w521, w421, w337 e w301 na cadeia D. Como não há correspondência para a água na cadeia A, as moléculas 2, 3 e 4 correspondem, respectivamente, à w161, w457 e w305.

Tabela 10 - Energias de ligação teórica (kcal/mol) para as interações da pDGL e da DGL à manose e ao trimanosídeo.

Lectina	Manose	Trimosídeo
pDGL	-5.5	-7.6
DGL	-5.1	-7.7

Fonte: Elaborada pelo autor.

Figura 14 - Ligação da manose gerada por docking molecular.



Fonte: Elaborada pelo autor.

Note-se a grande semelhança na acomodação da manose (bastão) tanto em relação às simulações como ao modelo cristalográfico. Resíduos que compõem o CRD estão representados como linhas. pDGL x manose (azul), DGL x manose (verde), pDGL cristal (cinza).

Tabela 11 - Distâncias (Å) relacionadas à interação da manose com a pDGL e a DGL.

Manose	Resíduo	pDGL cristal	pDGL DM	DGL DM
O3	Arg119/228	2.87	3.00	3.00
O4	Asp99/208	2.56	3.09	3.34
	Arg119/228	3.41	3.26	3.16
	Asn157/14	3.00	2.97	2.82
O5	Leu242/99	3.01	2.83	2.80
O6	Asp99/208	2.82	3.35	3.60
	Leu242/99	3.18	2.93	2.80
	Tyr243/100	3.10	2.92	3.07
	Tyr243/100	3.25	3.57	4.10
C6	Tyr155/12	3.70	4.00	3.90

Fonte: Elaborada pelo autor.

As distâncias listadas referem-se ao modelo cristalográfico da pDGL-manose (cristal) e aos modelos gerados por docking molecular (DM) para as formas precursora e madura da lectina. Note-se que nas simulações as interações para o resíduo-chave Asp99/208 são mais favoráveis no complexo da pDGL.

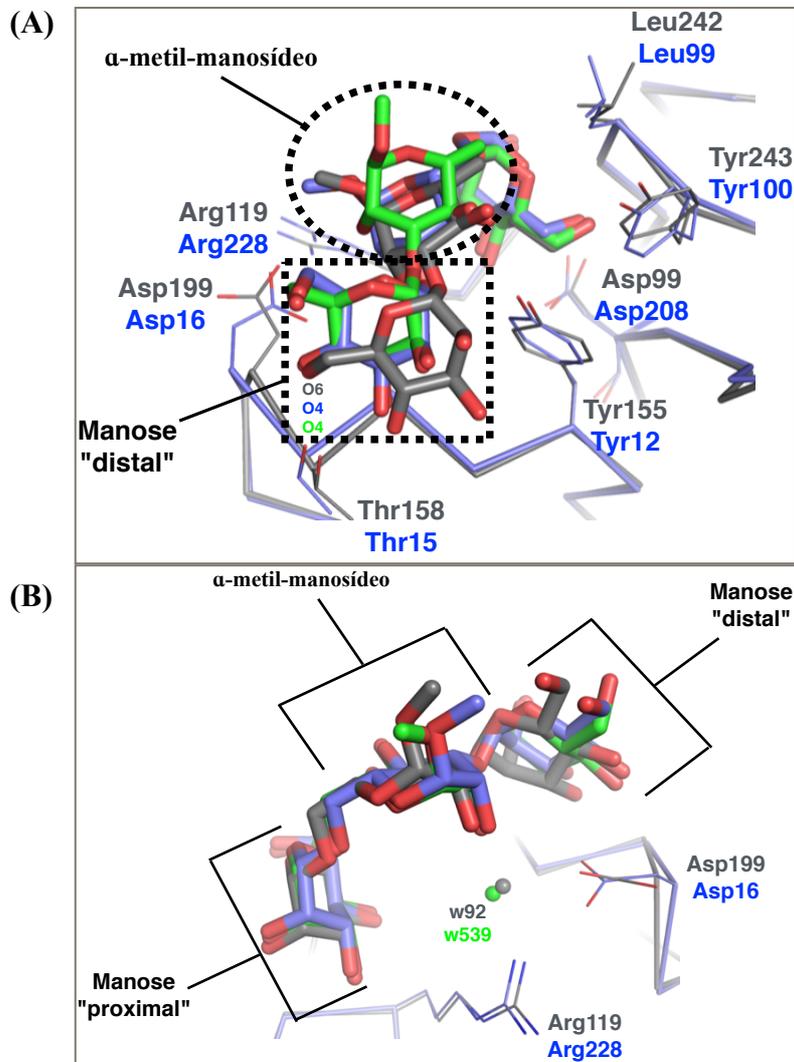
^a Resíduos correspondentes na pDGL e DGL, respectivamente.

A análise das simulações com o trimanosídeo revelou resultados interessantes. Como observado na Tabela 10, a afinidade das duas formas de lectina apresenta valores muito similares, embora o RMSD das melhores poses entre si seja de 1.53 Å, maior do que a comparação feita no *redocking* (de 1.22 Å). O RMSD para pDGL x trimanosídeo e o modelo cristalográfico do trissacarídeo foi de 1.76 Å. Quando analisa-se a sobreposição desses resultados, percebe-se o motivo dessas diferenças: apesar da molécula de manose “proximal” (que interage com o Asp99/208) ter uma acomodação similar nos três complexos, as moléculas de α -metil-manosídeo e manose “distal” estão posicionadas diferentemente (FIGURA 15A). Enquanto no modelo cristalográfico o grupo α -metil encontra-se totalmente voltado para o solvente, sem interagir com a DGL, nas poses teóricas esse grupo está voltado para o resíduo Asp159 (posição 16 na proteína madura), performando interações de van der Waals. Por outro lado, a manose “distal” está acomodada de maneira mais semelhante nos complexos da DGL (modelos cristalográfico e teórico), onde o oxigênio O4 interage com a cadeia lateral da Thr15 (posição 158 na pDGL) e a cadeia principal do Asp16. No modelo teórico da pDGL, essas interações são realizadas pelo oxigênio O6 da manose “distal”, como listado na Tabela 12.

Interessantemente, quando analisamos porque os modelos teóricos se diferenciam do experimental, observa-se no complexo cristalográfico da DGL-trimanosídeo que o oxigênio O2 da molécula de α -metil-manosídeo interage com uma molécula de água, a qual, presente entre os resíduos Asn14, Asp16 e Arg228, é conservada na grande maioria das estruturas cristalográficas de lectinas *Diocleinae* complexadas com açúcar (BEZERRA *et al.*, 2007), inclusive na da pDGL-manose, aqui descrita. Quando se sobrepõe os três complexos (pDGL-trimanosídeo DM, DGL-trimanosídeo DM e DGL-trimanosídeo cristal) na presença dessa molécula de água, percebe-se a ocorrência de um impedimento estérico com o oxigênio O2 do α -metil-manosídeo nos modelos teóricos, o qual interage com as cadeias laterais da Asn157 e do Asp159 (correspondentes às posições 14 e 16, respectivamente, na proteína madura). Na tentativa de corroborar a importância dessa molécula de água para a estabilização do trissacarídeo, realizou-se duas outras simulações de docking molecular da pDGL e DGL contra o trimanosídeo em sua presença. As melhores poses obtidas apresentaram afinidade de -7.3 e -6.9 kcal/mol, respectivamente, e o RMSD em relação ao modelo cristalográfico do ligante passou para 0.94 e 0.62 Å, respectivamente (FIGURA 15B). Com isso, o grupo α -metil adquire a mesma conformação do modelo cristalográfico, voltado para o solvente e sem

interagir com a cadeia lateral do Asp159/16, e o oxigênio O2 perde sua interação com a Asn157/14, possivelmente explicando a diminuição na afinidade em relação às simulações na ausência das moléculas de água (TABELA 10).

Figura 15 - Ligação do trimanosídeo gerada por docking molecular.



Fonte: Elaborada pelo autor.

(A) Sobreposição dos modelos teóricos da pDGL-trimannosídeo (cinza) e DGL-trimannosídeo (azul) em relação ao modelo cristalográfico desse trissacarídeo (verde). Note-se como os modelos teóricos se assemelham entre si quanto à conformação do α -metil-manosídeo, mas diferem quanto à manose "distal". Neste último caso, destaca-se a posição do oxigênio O6 da pDGL em relação ao O4 da lectina madura. (B) Sobreposição dos modelos teóricos da pDGL-trimannosídeo (cinza) e DGL-trimannosídeo (azul) em relação ao modelo cristalográfico desse trissacarídeo (verde) após simulação na presença das moléculas de água (esferas - cinza para pDGL e verde para DGL). Note-se o aumento na semelhança de conformação do α -metil-manosídeo e da manose "distal", fato este refletido nos menores valores de RMSD (ver texto).

Tabela 12 - Distâncias (Å) relacionadas à interação do trimanosídeo com a pDGL e a DGL. Continua.

Componente triman	Resíduo lectina	DGL cristal	Ausência de água		Presença de água			
			pDGL DM	DGL DM	pDGL DM	DGL DM		
Manose "proximal"	O2	Gly241/98	3.33	3.27	3.21	3.28	3.45	
		Leu242/99	3.67	3.43	3.67	3.71	3.61	
	O3	Arg119/228	2.94	3.15	3.05	3.04	3.14	
		Asp99/208	2.89	2.84	3.30	2.94	3.01	
		Arg119/228	3.20	3.16	3.08	3.22	3.08	
	O5	Asn157/14	2.82	3.23	3.10	3.01	2.93	
		Leu242/99	2.90	2.99	2.91	2.88	3.05	
		Leu242/99	3.09	3.26	3.12	3.38	3.11	
	O6	Asp99/208	3.34	3.32	3.34	3.37	3.66	
		Leu242/99	2.90	2.98	2.87	2.96	2.80	
		Tyr243/100	3.04	2.83	3.24	2.82	2.92	
		Tyr243/100	3.83	3.41	4.12	3.45	3.96	
	α-metil	O2	w92/w539	2.80	-	-	3.40	3.23
			Asn157/14	-	2.89	3.11	-	-
Asp159/16			3.90/3.72	4.29	4.34/3.62	3.60	3.89/3.82	
O3		Tyr155/12	3.45	2.94	3.07	4.17	3.89	
O4		Tyr155/12	2.68	3.31	3.22	3.24	3.01	
C6		Leu242/99	3.80	4.11	4.10	3.76	4.13	
		Leu242/99	4.12	4.16	4.62	3.94	4.35	
C7	Asp159/16	-	3.64	3.76	-	-		
Manose "distal"	O2	Pro156/13	4.33	-	4.47	3.35	4.27	
	O3	Glu96/205	-	4.00	-	-	-	
		Glu96/205	-	3.61	-	-	-	
		Pro156/13	2.97	3.98	3.38	3.46	3.20	
		Pro156/13	-	4.39	-	-	-	
		Thr158/15	2.85	-	3.17	2.84	2.84	
		Thr158/15	2.88	-	2.69	2.87	2.81	
	Asp159/16	4.50	-	4.71	3.10	4.28		

Tabela 12 - Distâncias (Å) relacionadas à interação do triman com a pDGL e a DGL. Conclusão.

O4	Pro156/13	-	3.11	-	-	-
	Thr158/15	3.47	3.95	3.71	-	3.94
	Thr158/15	2.60	2.71	2.69	4.09	2.77
	Asp159/16	2.91	-	3.20	4.05	3.24
	Asp159/16	2.85	-	3.06	4.00	3.17
	Asn164/21	-	4.32	-	-	-
O6	Thr158/15	-	2.76	-	-	-
	Asp159/16	-	3.14	-	-	-
C6	Asp159/16	4.39	-	4.20	-	-

Fonte: Elaborada pelo autor.

As distâncias listadas referem-se ao modelo cristalográfico da pDGL-manose (cristal) e aos modelos gerados por docking molecular (DM) para as formas precursora e madura da lectina.

^a Resíduos correspondentes na pDGL e DGL, respectivamente.

^b Moléculas de água presentes nas estruturas cristalográficas da pDGL e DGL, respectivamente, utilizadas nas simulações DM.

^c Devido a uma mudança de conformação em relação ao resíduo correspondente na pDGL, o átomo O2 do α -metil interage com os oxigênios OD1 e OD2 do Asp16.

Na busca por entender esses resultados, que corroboram a maior afinidade da pDGL pela manose em comparação à sua contra-parte madura (SOUSA *et al.*, 2016b), analisou-se as dimensões dos CRDs nessas duas formas da lectina, sendo observado que o precursor apresenta maior volume mas menores valores de área superficial e profundidade (TABELA 13).

Tabela 13 - Dimensões dos CRDs na pDGL e DGL.

Lectina	Volume (Área (Profundidade (Å)
pDGL	139.39	135.59	8.06
DGL	110.98	228.48	8.16

Fonte: Elaborada pelo autor.

Dimensões calculadas pelo servidor DogSiteScorer (VOLKAMER *et al.*, 2012) a partir das cadeias D e A da pDGL e DGL, respectivamente, não dissociadas de seus arranjos quaternários.

6 DISCUSSÃO

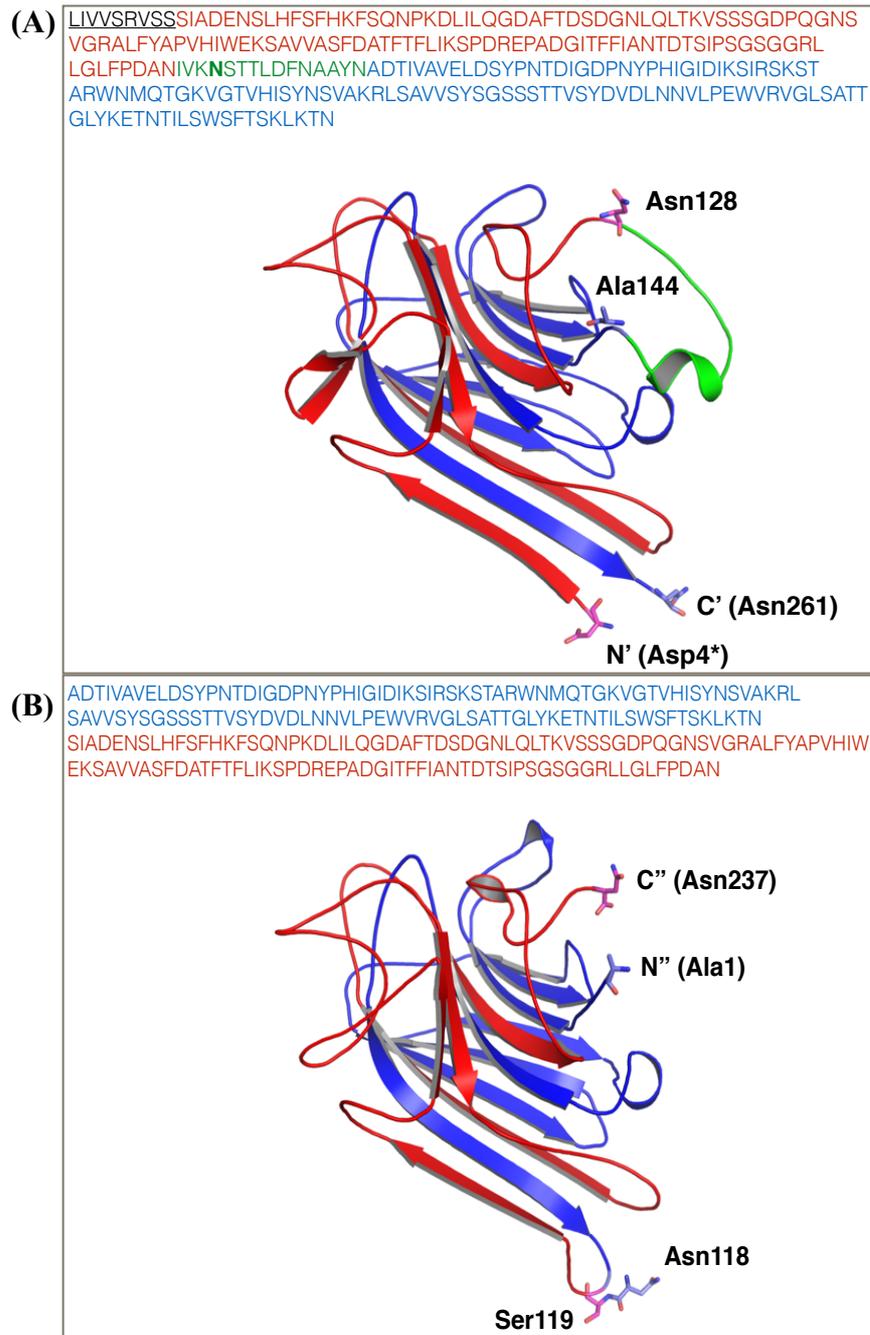
6.1 A pDGL é estruturalmente semelhante à sua contra-parte madura

No presente trabalho, descreveu-se, pela primeira vez, a estrutura cristalográfica de um precursor de lectina *Diocleinae*. Utilizou-se como modelo de estudo a pDGL, precursor deglicosilado de uma lectina isolada de sementes de *Dioclea grandiflora* (DGL), popularmente conhecida como mucunã.

A pDGL, assim como a proConA, molécula precursora da Concanavalina A (ConA), é uma glicoproteína contendo 290 resíduos de aminoácidos (FIGURA 16A) que por um complexo processamento pós-traducional dá origem à lectina madura, com 237 resíduos e desprovida de glicano. Uma primeira etapa desse processo é a retirada de um peptídeo sinal de direcionamento ao retículo endoplasmático com 29 resíduos e localizado na extremidade N-terminal, seguida da clivagem do glicano rico em manose, o qual está ligado à Asn132, por uma N-glicanase. Posteriormente, um segmento de 15 resíduos, contendo o sítio de glicosilação, é clivado proteolicamente por uma asparaginil endopeptidase no grupo carbonila de dois resíduos de Asn que o flanqueiam (posições 128 e 143) e um segmento C-terminal de nove resíduos é retirado, resultando em dois fragmentos polipeptídicos chamados de β e γ , que são unidos a partir das extremidades N- e C-terminais originais (FIGURA 16B), dando origem assim a uma proteína madura de sequência invertida em relação à molécula precursora, mecanismo esse conhecido como permutação circular.

A estrutura cristalográfica da pDGL revela que seus arranjos monoméricos, diméricos e tetraméricos são semelhantes à forma madura (FIGURA 8, acima), indicando que o processamento, apesar de muito complexo, não altera significativamente o *folding* da lectina, como demonstrado recentemente por modelagem molecular (SOUSA *et al.*, 2016b). De fato, as principais diferenças estruturais estão localizadas nas regiões que efetivamente são modificadas durante a permutação circular. A Figura 17, por exemplo, mostra as conformações dos resíduos que flanqueiam o *loop* Ile129-Asn143 na pDGL e que se tornam as extremidades N- e C-terminais na lectina madura após a clivagem proteolítica desse segmento de 15 resíduos.

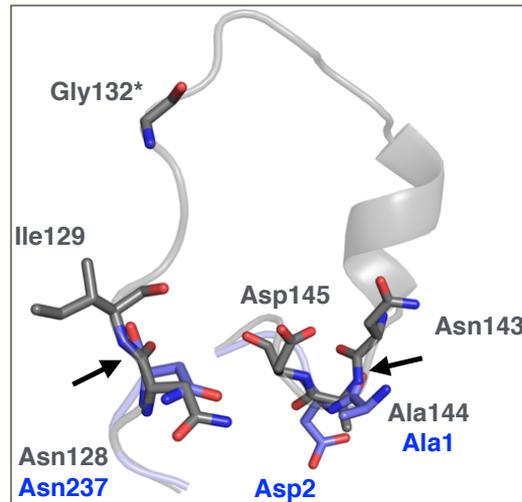
Figura 16 - Estruturas primária e terciária da pDGL e DGL.



Fonte: Elaborada pelo autor

(A) Estrutura primária parcial (acima) da pDGL (UNIPROT A9J251). O segmento sublinhado em preto refere-se aos nove últimos resíduos do peptídeo sinal de direcionamento ao retículo endoplasmático que é retirado na primeira etapa do processamento pós-traducional. Em vermelho, verde e azul destacam-se, respectivamente, o fragmento γ (res. 1-128), o *loop* Ile129-Asn144 (a asparagina destacada é o ponto de glicosilação) e o fragmento β (res. 144-261). Abaixo está representada o monômero da pDGL. Os resíduos em bastão são as extremidades N- e C-terminais e os pontos que flanqueiam o segmento de 15 resíduos que é clivado. * No modelo cristalográfico o resíduo N-terminal é um aspartato porque não havia densidade eletrônica para a serina que inicia o fragmento γ . (B) Estrutura primária (acima) da DGL (UNIPROT P08902). Note-se que os fragmentos β e γ inverteram suas localizações em relação à pDGL: o primeiro passa a ser o N-terminal e o segundo o C-terminal da lectina madura. Desse modo, os resíduos nas extremidades do precursor passam a fazer parte do *loop* Thr117-Glu122, assumindo as posições 118 e 119 (bastões), e os resíduos que na pDGL flanqueavam o segmento de 15 resíduos, tornam-se as novas extremidades N- e C-terminais na DGL após a clivagem proteolítica.

Figura 17 - Exemplo de diferenças conformacionais entre resíduos correspondentes da pDGL e DGL.



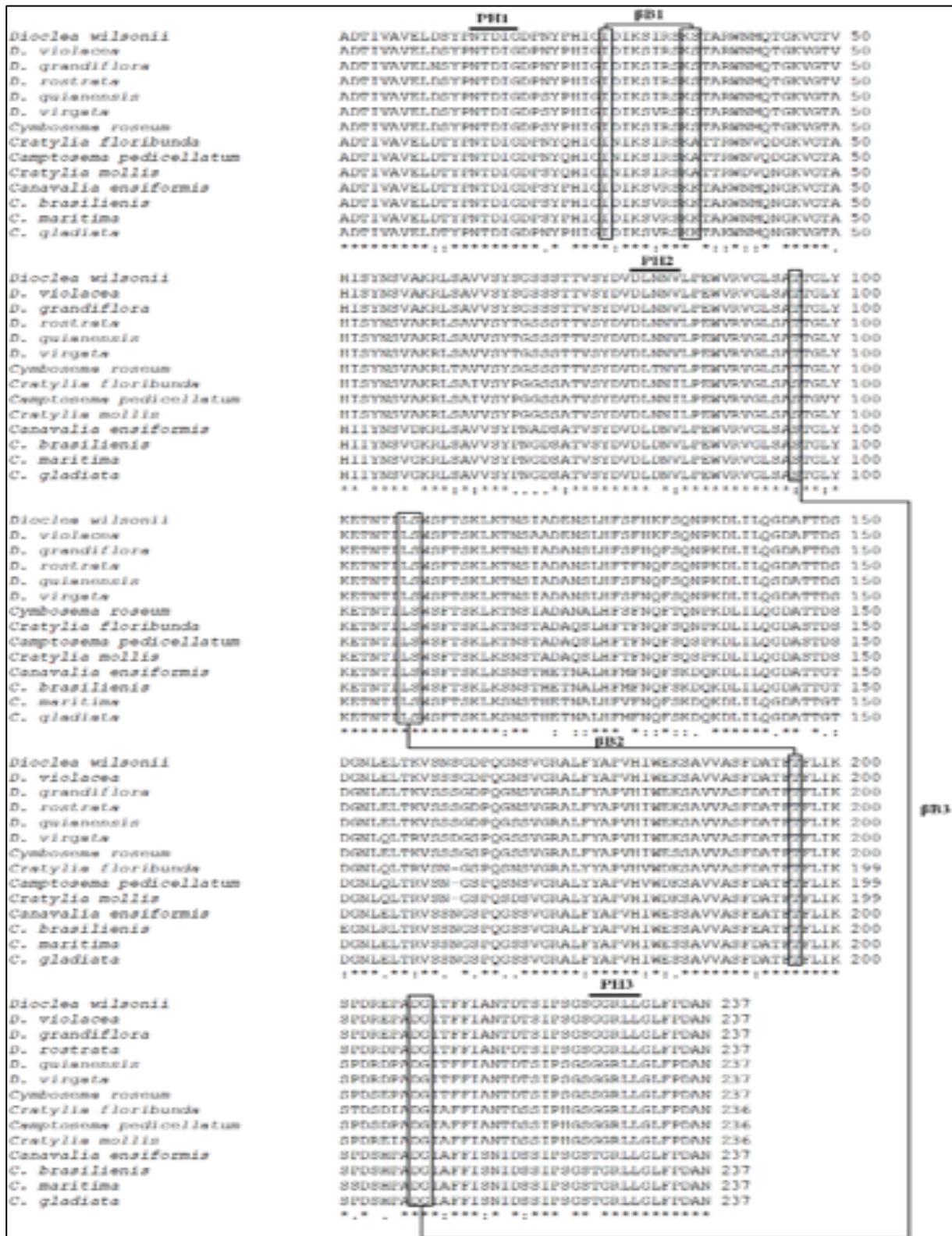
Fonte: Elaborada pelo autor.

Note-se que a retirada do *loop* Ile129-Asn143 durante o processamento da pDGL (cinza), através da atividade da asparaginil endopeptidase, que cliva a cadeia polipeptídica nos pontos indicados pelas setas, altera profundamente as conformações dos resíduos que passam a ser as extremidades N- e C-terminais na lectina madura (azul), Ala1 e Asn237, respectivamente. * Posição 132 é uma Asn na sequência da pDGL, mas foi modelada como Gly por não apresentar densidade eletrônica para a cadeia lateral na estrutura cristalográfica.

Além da grande conservação no *folding* dos arranjos monoméricos, diméricos e tetraméricos, a pDGL apresenta diversas estruturas secundárias não-repetitivas, como descrito anteriormente para a forma madura da ConM (SILVA-FILHO, 2013). Apesar dessas características estruturais ainda não terem sido descritas para a DGL, esse trabalho mostrou que os resíduos que as formam na ConM são conservados não só na DGL mas em diversas outras lectinas *Diocleinae* (FIGURA 18). De fato, quando se analisa o padrão de ligações de hidrogênio nessas regiões da DGL, percebe-se a existência de tais estruturas, podendo, assim, ser concluído que tais elementos são conservados na pDGL e em sua contra-parte madura.

A presença de β -turns e β -bulges é mais comum em sítios ativos/de ligação (RICHARDSON; GETZOFF; RICHARDSON, 1978) e a sua descrição para a ConM revelou que a maioria está localizada em regiões de ligação a compostos, como carboidratos (CRD) e fitohormônios (SILVA-FILHO, 2013). Como não existem estudos acerca da capacidade de lectinas isoladas de espécies do gênero *Dioclea* se ligarem a compostos não-glicídicos, não se sabe se o mesmo padrão visto para a ConM em relação à presença dessas estruturas não-repetitivas em sítios de ligação se mantém para as lectinas desse gênero. Por extensão, o mesmo pode-se dizer para a pDGL, já que o presente trabalho trata de sua capacidade de se ligar a carboidratos e não há nenhum outro estudo acerca de sua ligação a moléculas não-glicídicas.

Figura 18 - Estruturas secundárias não-repetitivas em lectinas *Diocleinae*.



Fonte: Adaptado de Silva-Filho (2013).

Alinhamento de seqüências de algumas lectinas *Diocleinae* com estruturas depositadas no PDB, destacando a conservação dos resíduos que compõem estruturas secundárias não-repetitivas. PH1, PH2 e PH3 referem-se a β -turns consecutivas que são representadas como hélices-3₁₀, mesmo não possuindo ângulos Φ e Ψ iguais (ver seção 5.2). β B1, β B2 e β B3 são β -bulges. * = conservação; : = modificações conservadas, . = modificações semi-conservativas. Espaços em branco referem-se à modificações de um grupo químico para outro (por exemplo de um polar por um apolar).

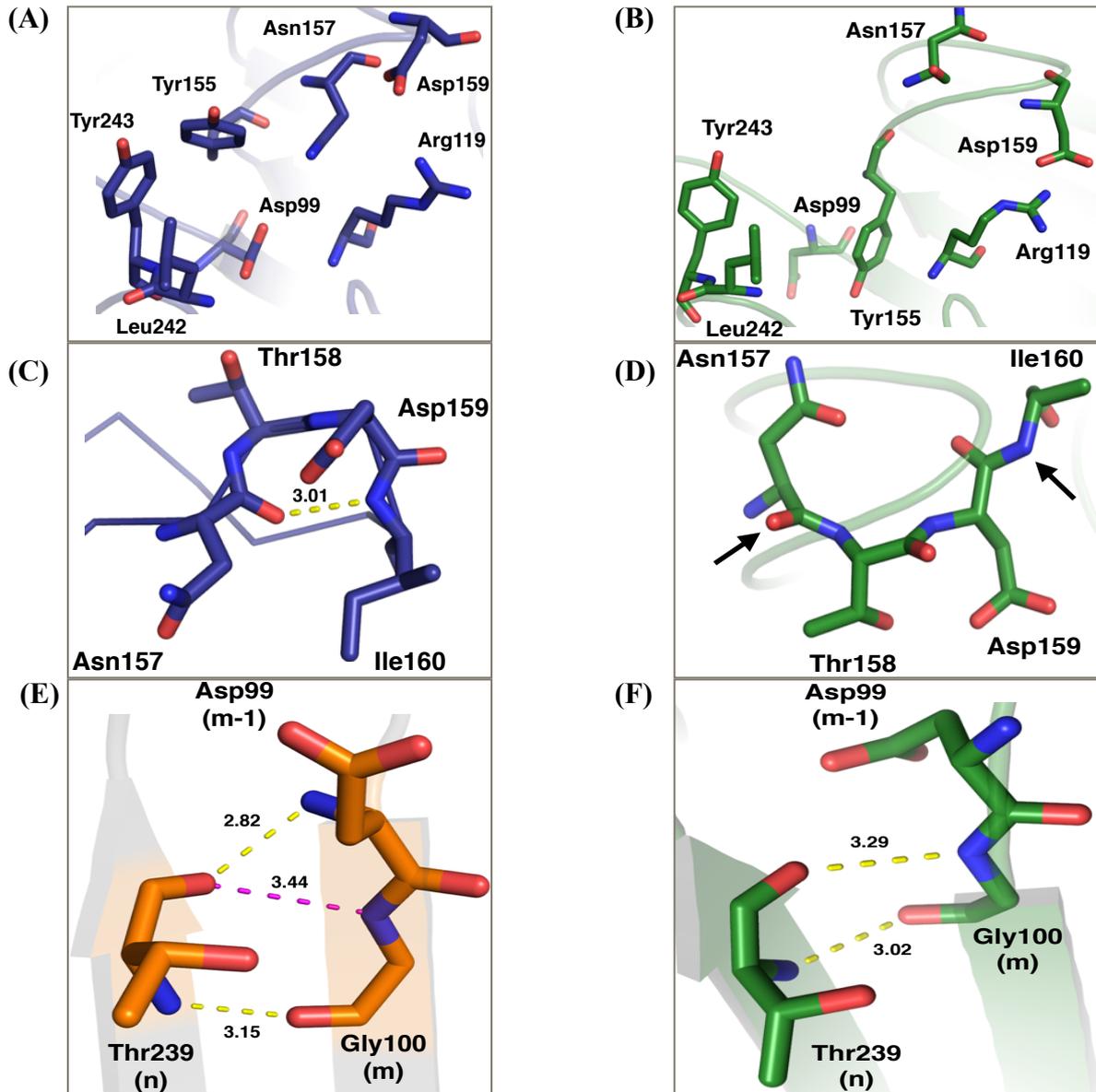
Observa-se que no CRD das quatro cadeias do modelo cristalográfico da pDGL há estruturas secundárias não-repetitivas: duas β -turns formadas pelos resíduos Asn157-Ile160 e Gly117-Leu120 e uma β -bulge formada pela trinca Thr239-Asp99-Gly100 (ver seção 5.2). De fato, essas três estruturas contêm resíduos que são essenciais para a ligação e estabilização da manose: Asn157 e Asp99 interagem com os átomos O4 e O6 da hexose e o nitrogênio da cadeia principal da Arg119 interage com O3 e O4. Assim, a formação de tais elementos estruturais parece ser fundamental para que esses resíduos assumam uma conformação dentro do CRD que seja capaz de ligar a manose. Isso fica evidente quando comparamos o complexo pDGL-manose com a estrutura cristalográfica (ainda em fase de elucidação) da pDGL em sua forma apoproteica, ou seja, sem os íons metálicos Ca^{2+} e Mn^{2+} . Nesse estado, o CRD encontra-se desestruturado (FIGURA 19), com a ligação peptídica entre Ala98 e Asp99 em configuração trans (FIGURA 5, seção 2.1.3), a cadeia lateral da Tyr155 ocupando o centro do sítio e a cadeia lateral da Asn157 voltada para o solvente. Essas mudanças fazem com que a β -turn Asn157-Ile160 e a β -bulge Thr239-Asp99-Gly100 não sejam formadas, impedindo a ligação de carboidratos. Interessantemente, a β -turn Gly117-Leu120 é conservada na forma apoproteica da pDGL, indicando que sua participação na estruturação do CRD é minoritária em relação aos outros dois elementos não-repetitivos.

6.2 A pDGL parece não formar tetrâmeros estáveis

Uma das características mais marcantes das lectinas *Diocleinae* é a capacidade de modificar seu estado de oligomerização de acordo com o pH: para valores iguais ou menores do que 5.5 se comportam como dímeros e para valores acima de 6.5 como tetrâmeros (CALVETE *et al.*, 1999). Esse processo é chamado de equilíbrio dímero-tetrâmero dependente de pH e a grande maioria das *ConA-like* apresentam essa característica (SENEAR; TELLER, 1980). Uma exceção é a DGL, que mantém um estado tetramérico em uma ampla faixa de pH (NAGANO *et al.*, 2008). Essa distinção da DGL em relação às outras lectinas reside na mudança de um resíduo de aminoácido localizado na posição 131: nela (e nas *D. wilsonii* e *D. violacea*) há uma histidina e nas demais uma asparagina. A presença de uma histidina nessa posição possibilita a interação desse resíduo com a Ile120 presente no *loop* Thr117-Glu122, o qual é estabilizado por outro resíduo de histidina (His51), na cavidade central do tetrâmero, estabilizando esse arranjo quaternário. Por outro lado, a presença de uma

asparagina impede esse contato central e o tetrâmero torna-se menos estável, alterando sua oligomerização de acordo com o pH do meio.

Figura 19 - Características estruturais da pDGL em suas formas holo- e apoproteica.



Fonte: Elaborada pelo autor.

(A) e (B) Disposição dos resíduos que compõem o CRD na pDGL holo- e apoproteica, respectivamente. (C) β -turn Asn157-Ile160 na pDGL-holo. (D) Região correspondente na pDGL-apo, destacando a disposição completamente diferente dos resíduos formadores. Note-se que os grupos -CO da Asn157 e -N da Ile160 (setas) estão voltados para o solvente, impedindo que os dois resíduos interajam entre si. Ile160 representada como alatina porque não há densidade eletrônica para modelar toda sua cadeia lateral. (E) β -bulge Thr239-Asp99-Gly100 na pDGL-holo. (F) Mesma região na pDGL-apo, destacando que o grupo -CO da Thr239 (n) perfaz uma interação de hidrogênio com o -N da Gly100 (m), como em uma folha- β convencional (ver FIGURA 9).

Na pDGL não há formação do *loop* Thr117-Glu122, pois os resíduos correspondentes fazem parte da extremidade N-terminal. Assim, não há contato dessa região com a His22 (posição 131 na forma madura), o que indica a incapacidade de formação de tetrâmeros estáveis. De fato, o estado de oligomerização da pDGL predito pelo PISA revela que essa formação quaternária é possível, porém com um caráter termodinâmico muito inferior ao de sua contra-parte madura, sendo já demonstrado por simulações de dinâmica molecular essa distinção nos estados oligoméricos das duas formas da lectina (SOUSA *et al.*, 2016b).

Estudos realizados com sedimentação por equilíbrio, e preditos por cristalografia (OLIVEIRA *et al.*, 2008), revelaram que a oligomerização independente de pH apresentado pela DGL dependia da participação igualmente importante dos resíduos His51 e His131: o primeiro estabilizaria o *loop* para que o segundo pudesse interagir com a Ile120 (ZAMORA-CABALLERO *et al.*, 2015). Mutações em um ou outro fazem com que a lectina apresente padrões similares de sedimentação, transformando a DGL em uma estrutura com limitada transição dímero-tetrâmero, e apenas a mutação simultânea desses resíduos é capaz de romper com a estabilidade tetramérica (ZAMORA-CABALLERO *et al.*, 2015). Esse aspecto é observado no modelo cristalográfico do tetrâmero da pDGL, onde, além da falta de interação entre His22 e Ile2, como visto no parágrafo anterior, a His194 não mantém o mesmo padrão de interações visto para a lectina madura: apesar de performar ligações com os resíduos Val78/189, Val79/190 e Leu259/116, essa histidina não interage com a Thr260/117 (FIGURA 20), que na DGL é o primeiro resíduo do *loop* e assim fundamental para sua estabilização. Dessa maneira, a falta dessas duas interações (His22-Ile2 e His194-Thr260), devido serem as regiões onde a mudança no sentido da cadeia polipeptídica ocorre, parecem ser as responsáveis pela baixa estabilidade do tetrâmero da pDGL.

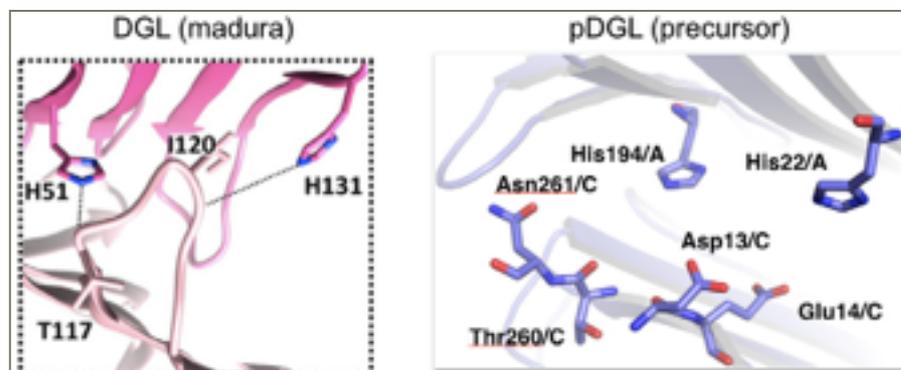
6.3 Interação CH- - π estabiliza a manose no CRD da pDGL

Embora a descrição convencional de interações de hidrogênio presentes em macromoléculas levem em consideração, de maneira geral, apenas os átomos de oxigênio e nitrogênio como doadores, diversos estudos tem demonstrado que ligações “fora do padrão”, onde grupos CH- atuam como doadores, são fundamentais para a estabilização estrutural de proteínas e ácidos nucleicos, bem como são requeridas para o estabelecimento de interações entre proteínas e ligantes (HOROWITZ; TRIEVEL, 2012; SCHEINER, 2011).

No presente trabalho, descreveu-se uma dessas ligações de hidrogênio ditas “não-convencionais”, onde o carbono C6 da manose interage com o centro do anel benzeno da Tyr155 (FIGURA 13). Isso acontece porque os elétrons em uma estrutura aromática, como a cadeia lateral da tirosina, fenilalanina e triptofano, se distribuem uniformemente, fazendo com que essa estrutura adquira um caráter negativo (sistema π), atraindo eletricamente átomos de hidrogênio que estejam ligados covalentemente a outros mais eletronegativos, como o carbono (NISHIO *et al.*, 2014). Apesar da diferença de eletronegatividade entre carbono e hidrogênio ser pequena, fato que dificultaria a formação de uma carga parcial positiva no lado do hidrogênio, a presença de átomos mais eletronegativos ligados ao carbono, como o oxigênio O6 na molécula de manose, faz com que os elétrons se desloquem para o átomo mais eletronegativo (nesse caso o oxigênio), contribuindo, assim, para uma maior polarização da ligação $-CH$, dando-a um caráter mais ácido e tornando a interação de hidrogênio mais forte e favorável (WAHL; SUNDARALINGAM, 1997; STEINER; DESIRAJU, 1998).

Interessantemente, esse tipo de interação parece ser comum entre lectinas *Diocleinae* e carboidratos. Uma análise das estruturas depositadas no PDB dessas proteínas complexadas com diferentes açúcares, mostra que esse tipo de interação deve ser considerada parte essencial na caracterização dos contatos que estabilizam os complexos cristalográficos, bem como na escolha das melhores poses durante ensaios de docking molecular, uma vez que isso pode ser um fator determinante para a afinidade e especificidade de ligação, como demonstrado para as interações diferenciadas da lectina LecA de *Pseudomonas aeruginosa* com galactosídeos aromáticos (KADAM *et al.*, 2013).

Figura 20 - Interações interdiméricas na cavidade central da DGL e da pDGL.



Fonte: Adaptado de Zamora-Caballero *et al.* (2015) e elaborada pelo autor.

Na DGL (à esquerda) existem os contatos interdiméricos orquestrados pelos resíduos His51 e His131 com Thr117 e Ile120. Na pDGL (à direita), os resíduos correspondentes (His194 e His22, respectivamente), não participam da formação de contatos interdiméricos com os resíduos de treonina e isoleucina correspondentes.

6.4 A pDGL tem maior afinidade para manose e trimanosídeo do que a DGL

Os resultados de docking demonstraram que a pDGL de fato se liga com maior afinidade à manose do que sua contra-parte madura (SOUSA *et al.*, 2016b). Quando as simulações foram realizadas com o trimanosídeo, um carboidrato complexo que representa o núcleo trimanosídico de glicanos N-ligados (DAM *et al.*, 1998; DAM *et al.*, 2000), verificou-se que a afinidade entre as duas formas da lectina eram similares (TABELA 11) e, embora a conformação do componente manose “distal” diferisse, o número de interações eram os mesmos (aproximadamente 30 - TABELA 13), explicando essa similaridade de afinidade. Porém, como o componente α -metil-manosídeo apresentou uma disposição diferente em relação ao modelo cristalográfico do ligante, fato esse caracterizado pela ausência de uma molécula de água que conecta-se aos resíduos Arg119/228, Asn157/14 e Asp159/16, simulações adicionais na presença da referida molécula de água resultaram em um padrão energético mais favorável para a interação da pDGL (-7.3 Kcal/mol contra -6.9 Kcal/mol da DGL). Juntos, esses dados indicam que a pDGL além de apresentar alta afinidade pelo monossacarídeo manose, também o tem para carboidratos mais complexos, como o trimanosídeo; para este último caso, a presença de uma molécula de água específica, e que é conservada na maioria das estruturas cristalográficas de lectinas *Diocleinae* (BEZERRA *et al.*, 2007), parece ser essencial para que o mesmo se estabilize em uma conformação energeticamente mais favorável, corroborando a importância de moléculas de solvente para a estabilização de ligantes glicídicos (TSCHAMPEL; WOODS, 2003).

O fato da pDGL apresentar essa maior afinidade tanto para a manose quanto para o trimanosídeo indica que seu CRD, embora com um mesmo *fold* em relação à sua forma madura, contenha elementos que o tornam diferente. Em verdade, quando associamos as suas dimensões (TABELA 11) com distâncias relativas entre resíduos que o compõem, verifica-se que a profundidade desse sítio na pDGL é menor, o que estaria relacionado com uma menor distância entre os os átomos Arg119/228 NE e Asn157/14 ND2 (NÓBREGA *et al.*, 2012). Essa distância é de 5.14 Å na pDGL e de 5.51 Å na DGL. Nesse sentido, um sítio mais raso proporciona interações mais fortes entre seus resíduos formadores e ligantes (BEZERRA *et al.*, 2013). Isso é o que é percebido quando analisa-se os padrões de interação da manose e do trimanosídeo com o resíduo-chave Asp99/208 na pDGL e DGL: na forma precursora essa distância é sempre menor, tanto no modelo cristalográfico quanto nos teóricos, fazendo,

assim, com que sua interação seja energeticamente mais favorável, como verificado pelos valores preditos, e explicando sua maior afinidade.

7 CONCLUSÕES E PERSPECTIVAS

- A estrutura cristalográfica da pDGL aqui apresentada mostra que o processamento pós-traducional não altera o *folding* entre as formas precursora e madura da lectina. Isso pode indicar que o precursor já é sintetizado de uma maneira a exercer atividade biológica, dependendo, para tanto, apenas da deglicosilação e da acidificação do meio;
- A presença de estruturas secundárias não-repetitivas na pDGL holoproteica (contendo os íons divalentes Ca^{2+} e Mn^{2+}) corrobora a ideia de que o precursor é sintetizado em uma forma pronta para exercer alguma função. Considerando que tais elementos estruturais estão presentes de maneira mais frequente em locais de ligação a compostos, como visto para a forma madura da ConM, é de se supor que a pDGL também utilize-os para este fim, mas estudos adicionais de co-cristalização e de atividade biológica deverão ser realizados para corroborar essa hipótese;
- A estabilização do tetrâmero da DGL madura em uma ampla faixa de pH depende da interação dos resíduos His51 e His131 com o *loop* Thr117-Glu122. Como na pDGL essa região corresponde à extremidades N- e C-terminais, tais interações não existem e o precursor apresenta um arranjo tetramérico pouco estável. De fato, análises preliminares de SAXS (*Small Angle X-Ray Scattering*, espalhamento de raios X a baixo ângulo) mostram que a pDGL é mais estável na forma dimérica, corroborando os achados do presente estudo e aquilo que foi predito por dinâmica molecular (SOUSA *et al.*, 2016b). Análises de ultracentrifugação por sedimentação podem ser realizados posteriormente para corroborar em definitivo esses dados;
- Através das análises *in silico* mostrou-se que, de fato, a pDGL não só apresenta maior afinidade pela manose como, inesperadamente, também pelo trimanosídeo 3,6-di-O-(α -D-manopiranosil)- α -D-manopiranosil. Junto com os dados cristalográficos, essas análises indicam que pequenas mudanças na disposição dos resíduos que compõem o CRD fortalecem as interações entre a pDGL e esses ligantes, o que se reflete em sua maior afinidade em relação à lectina madura. Estudos de atividade biológica, como modelos de indução de produção de óxido nítrico, podem ser desenvolvidos para corroborar esses achados.

REFERÊNCIAS

- AGRAWAL, B. B. L.; GOLDSTEIN, I. J.. Physical and chemical characterization of concanavalin A, the hemagglutinin from jack bean (*Canavalia ensiformis*). **Biochim. Biophys. Acta**, [s.l.], v. 133, n. 2, p.376-379, fev. 1967.
- ALENCAR, N. M. N. *et al.* Leguminous Lectins as Tools for Studying the Role of Sugar Residues in Leukocyte Recruitment. **Mediators Inflamm.**, [s.l.], v. 8, n. 2, p.107-113, jan. 1999.
- AMBROSI, M.; CAMERON, N.; DAVIS, B. Lectins: tools for the molecular understanding of glycode. **Org. Biomol. Chem.**, [s.l.], v. 3, n. 9, p. 1593-1608, maio 2005.
- ANDRADE, J. L. *et al.* Lectin-Induced Nitric Oxide Production. **Cell. Immunol.**, [s.l.], v. 194, n. 1, p.98-102, maio 1999.
- ARCALIS, E. *et al.* The dynamic behavior of storage organelles in developing cereal seeds and its impact on the producti. **Front. Plant Sci.**, [s.l.], v. 5, n. 1, p.1-12, 3 set. 2014.
- ASSREUY, A. M. *et al.* Prevention of cyclophosphamide-induced hemorrhagic cystitis by glucose-mannose binding plant lectins. **J. Urol.**, [s.l.], v. 161, n. 6, p.1988-1993, jun. 1999.
- BARAUNA, S. *et al.* Antidepressant-like effect of lectin from *Canavalia brasiliensis* (ConBr) administered centrally in mice. **Pharmacol. Biochem. Behav.**, [s.l.], v. 85, n. 1, p.160-169, set. 2006.
- BARLOW, D. J.; THORNTON, J. M. Helix geometry in proteins. **J. Mol. Biol.**, [s.l.], v. 201, n. 3, p. 601-619, jun. 1988.
- BARRAL-NETO, M. *et al.* Human lymphocyte stimulation by legume lectins from the Diocleae tribe. **Immunol. Invest.**, [s.l.], v. 21, n. 4, p.297-303, jul. 1992.
- BARROSO-NETO, I. L. *et al.* Vasorelaxant activity of *Canavalia grandiflora* seed lectin: A structural analysis. **Arch. Biochem. Biophys.**, [s.l.], v. 543, n. 1, p.31-39, fev. 2014.
- BECKER, J. W. *et al.* The covalent and three-dimensional structure of concanavalin A: III. Structure of the monomer and its interactions with metals and saccharides. **J. Biol. Chem.**, [s.l.], v. 250, n. 4, p.1513-1524, fev. 1975.
- BENTO, C. A. *et al.* Rat paw edema and leukocyte immigration induced by plant lectins. **Agents Actions**, [s.l.], v. 38, n. 1, p.48-54, jan. 1993.
- BERNSTEIN, F. C. *et al.* The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. **Eur. J. Biochem.**, [s.l.], v. 80, n. 2, p.319-324, nov. 1977.
- BEST, R. B. *et al.* Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. **J. Chem.Theory Comput.**, [s.l.], v. 8, n. 9, p.3257-3273, set. 2012.

- BEZERRA, G. A. *et al.* Structural analysis of *Canavalia maritima* and *Canavalia gladiata* lectins complexed with different dimannosides: New insights into the understanding of the structure–biological activity relationship in legume lectins. **J. Struct. Biol.**, [s.l.], v. 160, n. 2, p.168-176, nov. 2007.
- BEZERRA, E. H. S. *et al.* Structural analysis of ConBr reveals molecular correlation between the carbohydrate recognition domain and endothelial NO synthase activation. **Biochem. Biophys. Res. Commun.**, [s.l.], v. 408, n. 4, p.566-570, maio 2011.
- BEZERRA, M. J. B. *et al.* Crystal structure of *Dioclea violacea* lectin and a comparative study of vasorelaxant properties with *Dioclea rostrata* lectin. **Int. J. Biochem. Cell Biol.**, [s.l.], v. 45, n. 4, p.807-815, abr. 2013.
- BOMPARD-GILLES, C. *et al.* Substrate mimicry in the active center of a mammalian alpha-amylase: structural analysis of an enzyme-inhibitor complex. **Structure**, [s.l.], v. 4, n. 12, p. 1441-1452, dez. 1996.
- BOUCKAERT, J. *et al.* Crystallographic structure of metal-free concanavalin A at 2.5 Å resolution. **Proteins: Structure, Function, and Genetics**, [s.l.], v. 23, n. 4, p.510-524, dez. 1995.
- BOUCKAERT, J. *et al.* Novel structures of plant lectins and their complexes with carbohydrates. **Curr. Opin. Struct. Biol.**, [s.l.], v. 9, n. 5, p.572-577, out. 1999.
- BOUCKAERT, J. *et al.* Sequential Structural Changes upon Zinc and Calcium Binding to Metal-free Concanavalin A. **J. Biol. Chem.**, [s.l.], v. 271, n. 27, p.16144-16150, 5 jul. 1996.
- BOWLES, D. J. Posttranslational processing of concanavalin A precursors in jackbean cotyledons. **J. Cell Biol.**, [s.l.], v. 102, n. 4, p.1284-1297, 1 abr. 1986.
- BOYD, W. C.; SHAPLEIGH, E. Specific Precipitating Activity of Plant Agglutinins (Lectins). **Science**, [s.l.], v. 119, n. 3091, p.419-419, 26 mar. 1954.
- BREWER, C. F.; BHATTACHARYYA, L. Specificity of concanavalin A binding to asparagine-linked glycopeptides: A nuclear magnetic relaxation dispersion study. **J. Biol. Chem.**, [s.l.], v. 261, n. 16, p.7306-7310, jun. 1986.
- BRINDA, K. V.; SUROLIA, A.; VISHVESHWARA, S. Insights into the quaternary association of proteins through structure graphs: a case study of lectins. **Biochem. J.**, [s.l.], v. 391, n. 1, p.1-15, 1 out. 2005.
- BROWN, R. D.; KOENIG, S. H.; BREWER, C. F. Conformational equilibrium of demetalized concanavalin A. **Biochemistry**, [s.l.], v. 21, n. 3, p.465-469, fev. 1992.
- CALVETE, J. J. *et al.* Molecular characterization and crystallization of Diocleinae lectins. **Biochim. Biophys. Acta**, [s.l.], v. 1430, n. 2, p.367-375, mar. 1999.

CARRINGTON, D. M.; AUFFRET, A.; HANKE, D. Polypeptide ligation occurs during post-translational modification of concanavalin A. **Nature**, [s.l.], v. 313, n. 5997, p.64-67, 3 jan. 1985.

CAVADA, B. S. *et al.* Revisiting proteus: do minor changes in lectin structure matter in biological activity?: Lessons from and potential biotechnological uses of the Diocleinae subtribe lectins. **Curr. Protein Pept. Sci.**, [s.l.], v. 2, n. 2, p.123-135, jun. 2001.

CECCATTO, V. *et al.* Purification and Partial Characterization of a Lectin from *Canavalia grandiflora* Benth. Seeds. **Protein Pept. Lett.**, [s.l.], v. 9, n. 1, p.67-73, 1 fev. 2002.

CHAN, A. W. E. *et al.* Identification, classification, and analysis of beta-bulges in proteins. **Prot. Sci.**, [s.l.], v. 2, n. 10, p.1574-1590, out. 1993.

CHERVENAK, M. C.; TOONE, E. J. Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities. **Biochemistry**, [s.l.], v. 34, n. 16, p. 5685-5695, abr. 1995.

CHRISPEELS, M. J. *et al.* Characterization of the Endoplasmic Reticulum-associated Precursor of Concanavalin A. **J. Biol. Chem.**, [s.l.], v. 261, n. 22, p.10021-10024, ago. 1986.

CLAUDE, S. *et al.* Targeting of proConA to the Plant Vacuole depends on its Nine Amino-acid C-terminal Propeptide. **Plant Cell Physiol.**, Kyoto, v. 46, n. 10, p.1603-1612, out. 2005.

COLLABORATIVE COMPUTATIONAL PROJECT NUMBER 4. The CCP4 suite: programs for protein crystallography. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 50, n. 5, p.760-763, 1 set. 1994.

CUNNINGHAM, B. A. *et al.* Favin versus concanavalin A: Circularly permuted amino acid sequences. **Proc. Natl. Acad. Sci. Usa**, [s.l.], v. 76, n. 7, p.3218-3222, jul. 1979.

DALKIN, K.; BOWLES, D. J. Analysis of inter-relationship of jackbean seed components by two-dimensional mapping of iodinated tryptic peptides. **Planta**, [s.l.], v. 157, n. 6, p. 536-539, maio 1983.

DAM, T. K. *et al.* Differential Solvation of "Core" Trimannoside Complexes of the *Dioclea grandiflora* Lectin and Concanavalin A Detected by Primary Solvent Isotope Effects in Isothermal Titration Microcalorimetry. **J. Biol. Chem.**, [s.l.], v. 273, n. 49, p.32826-32832, dez. 1998.

DAM, T. K. *et al.* Diocleinae Lectins Are a Group of Proteins with Conserved Binding Sites for the Core Trimannoside of Asparagine-linked Oligosaccharides and Differential Specificities for Complex Carbohydrates. **J. Biol. Chem.**, [s.l.], v. 273, n. 20, p.12082-12088, maio 1998.

DAM, T. K. *et al.* Thermodynamic binding studies of lectins from Diocleinae subtribe to deoxy analogs of the core trimannoside of asparagine-linked oligosaccharides. **J. Biol. Chem.**, [s.l.], v. 275, n. 21, p.16119-16126, maio 2000.

DANG, L.; VAN DAMME, E. J. M. Toxic proteins in plants. **Phytochemistry**, [s.l.], v. 117, p.51-64, set. 2015.

DELATORRE, P. *et al.* Crystal structure of a lectin from *Canavalia maritima* (ConM) in complex with trehalose and maltose reveals relevant mutation in ConA-like lectins. **J. Struct. Biol.**, [s.l.], v. 154, n. 3, p.280-286, jun. 2006.

DELATORRE, P. *et al.* Interactions between indole-3-acetic acid (IAA) with a lectin from *Canavalia maritima* seeds reveal a new function for lectins in plant physiology. **Biochimie**, Paris, v. 95, n. 9, p.1697-1703, set. 2013.

DRICKAMER, K. Two Distinct Classes of Carbohydrate-recognition Domains in Animal Lectin. **J. Biol. Chem.**, [s.l.], v. 263, n. 20, p.9557-9560, jul. 1988.

EDELMAN, G. M. *et al.* The covalent and three-dimensional structure of concanavalin A. **Proc. Natl. Acad. Sci. Usa**, [s.l.], v. 69, n. 9, p.2580-2584, set. 1972.

EDELMAN, G. M.; WANG, J. L. Binding and functional properties of concanavalin A and its derivatives: III. Interactions with indoleacetic acid and other hydrophobic ligands. **J. Biol. Chem.**, [s.l.], v. 253, n. 9, p.3016-3022, maio 1978.

EDMUNDSON, A. B. *et al.* Isolation and characterization of concanavalin A polypeptide chains. **Biochemistry**, [s.l.], v. 10, n. 19, p.3554-3559, set. 1971.

EMSLEY, P. *et al.* Features and development of Coot. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 66, n. 4, p.486-501, 24 mar. 2010.

EVANS, P. R.. An introduction to data reduction: space-group determination, scaling and intensity statistics. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 67, n. 4, p.282-292, 18 mar. 2011.

FERREIRA, R. R. *et al.* Characteristics of the histamine release from hamster cheek pouch mast cells stimulated by lectins from Brazilian beans and concanavalin A. **Inflamm. Res.**, [s.l.], v. 45, n. 9, p.442-447, jul. 1996.

FREY, M. *et al.* An herbivore elicitor activates the gene for indole emission in maize. **Proc. Natl. Acad. Sci. Usa**, [s.l.], v. 97, n. 26, p.14801-14806, 5 dez. 2000.

GADELHA, C. A. A. *et al.* Native crystal structure of a nitric oxide-releasing lectin from the seeds of *Canavalia maritima*. **J. Struct. Biol.**, [s.l.], v. 152, n. 3, p.185-194, dez. 2005.

GASTEIGER, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. In: WALKER, John M. (Ed.). **The Proteomics Protocols Handbook**. [s.i]: Humana Press, 2005. p. 571-607.

GOMES, J. C. *et al.* Histamine release induced by glucose (mannose)-specific lectins isolated from Brazilian beans: Comparison with concanavalin A. **Agents Actions**, [s.l.], v. 41, n. 3, p. 132-135, maio 1994.

- GOODSELL, D. S. Concanavalin A and Circular Permutation. **Rcsb Protein Data Bank**, [s.l.], v. 1, n. 1, p.1-1, 1 abr. 2010. Protein Data Bank, Rutgers University. Acesso em: 05 jun. 2017. Disponível em: <http://dx.doi.org/10.2210/rcsb_pdb/mom_2010_4>.
- GRAHAM, P. H.; VANCE, C. P. Legumes: Importance and Constraints to Greater Use. **Plant Physiol.**, [s.l.], v. 131, n. 3, p.872-877, 1 mar. 2003.
- GUPTA, D. *et al.* A Comparison of the Fine Saccharide-Binding Specificity of *Dioclea grandiflora* Lectin and Concanavalin A. **Eur. J. Biochem.**, [s.l.], v. 242, n. 2, p.320-326, dez. 1996.
- GUPTA, Garima.; SUROLIA, Avadhesh. **Glycomics: an overview of the complex glycode**. In: SUDHAKARAN, Perumana; SUROLIA, Avadhesh (eds). Biochemical Roles of Eukaryotic Cell Surface Macromolecules. Advances in Experimental Medicine and Biology. New York: Springer, 2012, 293 p.
- HARDMAN, K. D.; AINSWORTH, C. F. Structure of concanavalin A at 2.4-Å resolution. **Biochemistry**, [s.l.], v. 11, n. 26, p.4910-4919, dez. 1972.
- HELLIWELL, J. R. **Macromolecular crystallography with synchrotron radiation**. [s.l.]: Cambridge University Press, 1992.
- HERMAN, E. M.; SHANNON, L. M.; CHRISPPEELS, M. J.. Concanavalin A is synthesized as a glycoprotein precursor. **Planta**, [s.l.], v. 165, n. 1, p.23-29, jul. 1985.
- HERMAN, E. M.; SHANNON, L. M.; CHRISPPEELS, M. J.. Concanavalin A is synthesized as a glycoprotein precursor. **Planta**, [s.l.], v. 165, n. 1, p.23-29, jul. 1985.
- HIGGINS, T. J. V. Synthesis and Regulation of Major Proteins in Seeds. **Ann. Rev. Plant Physiol.**, [s.l.], v. 35, n. 1, p.191-221, jun. 1984.
- HOROWITZ, S.; TRIEVEL, R. C.. Carbon-Oxygen Hydrogen Bonding in Biological Structure and Function. **J. Biol. Chem.**, [s.l.], v. 287, n. 50, p.41576-41582, out. 2012.
- HU, D.; TATENO, H.; HIRABAYASHI, J. Lectin engineering, a molecular evolutionary approach to expanding the lectin utilities. **Molecules**, [s.l.], v. 20, n. 5, p. 7637-7656, Abr. 2015.
- HUTCHINSON, E. G.; THORNTON, J. M. A revisited set of potential for beta-turns n proteins. **Prot. Sci.**, [s.l.], v. 3, n. 12, p. 2207-2216, dez. 1994.
- JANCARIK, J.; KIM, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. **J. Appl. Cryst.**, [s.l.], v. 24, n. 4, p.409-411, 1 ago. 1991.
- JOHANSSON, M. U. *et al.* Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. **BMC Bioinformatics**, [s.l.], v. 13, n. 1, p.173-182, jul. 2012.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, [s.l.], v. 22, n. 12, p. 2577-2637, dez. 1983.

KABSCH, W. XDS. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 66, n. 2, p.125-132, 22 jan. 2010.

KARPLUS, P. A.; DIEDERICHS, K. Linking Crystallographic Model and Data Quality. **Science**, [s.l.], v. 336, n. 6084, p.1030-1033, 24 maio 2012.

KAUSHIK, S.; MOHANTY, D.; SUROLIA, A. The Role of Metal Ions in Substrate Recognition and Stability of Concanavalin A: A Molecular Dynamic. **Biophys. J.**, Cambridge, v. 96, n. 1, p.21-34, jan. 2009.

KOMATH, S. S.; KAVITHA, M.; SWAMY, M. J. Beyond carbohydrate binding: new directions in plant lectin research. **Org. Biomol. Chem.**, [s.l.], v. 4, n. 6, p.973-988, jan. 2006.
KRISSEL, E.; HENRICK, K. Inference of Macromolecular Assemblies from Crystalline State. **J. Mol. Biol.**, [s.l.], v. 372, n. 3, p.774-797, set. 2007.

LAM, S. K.; NG, T. B. Lectins: production and practical applications. **Appl Microbiol Biotechnol**, [s.l.], v. 89, n. 1, p.45-55, 3 out. 2011.

LATTMANN, E. E.; LOLL, P. J. **Protein Crystallography: a concise guide**. [s.l.]: The John Hopkins University Press, 2008.

LIS, H.; SHARON, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. **Chem. Rev.**, [s.l.], v. 98, n. 2, p.637-674, abr. 1998.

LORIS, R. *et al.* A structure of the complex between concanavalin A and methyl-3,6-di-O-(alpha-D-mannopyranosyl)-alpha-D-mannopyranoside reveals two binding modes. **J. Biol. Chem.**, [s.l.], v. 271, n. 48, p.30614-30618, nov. 1996.

LORIS, R. *et al.* Legume lectin structure. **Biochim. Biophys. Acta**, [s.l.], v. 1383, n. 1, p. 9-36, mar. 1998.

MACEDO, M.; OLIVEIRA, C.; OLIVEIRA, C. Insecticidal Activity of Plant Lectins and Potential Application in Crop Protection. **Molecules**, [s.l.], v. 20, n. 2, p.2014-2033, jan. 2015.

MALAGUTI, M. *et al.* Bioactive Peptides in Cereals and Legumes: Agronomical, Biochemical and Clinical Aspects. **Int. J. Mol. Sci.**, [s.l.], v. 15, n. 11, p.21120-21135, 14 nov. 2014.

MANOJ, N.; SUGUNA, K. Signature of quaternary structure in the sequences of legume lectins. **Protein Eng.**, [s.l.], v. 14, n. 10, p.735-745, 1 out. 2001.

MCCOY, A. J. *et al.* Phaser crystallographic software. **J. Appl. Cryst.**, [s.l.], v. 40, n. 4, p. 658-674, 13 jul. 2007.

- MILLERD, A. Biochemistry of Legume Seed Proteins. **Ann. Rev. Plant Physiol.**, [s.l.], v. 26, n. 1, p.53-72, jun. 1975.
- MIN, W.; DUNN, A. J.; JONES, D. H. Non-glycosylated recombinant pro-concanavalin A is active without polypeptide cleavage. **Embo J.**, [s.l.], v. 11, n. 4, p.1303-1307, abr. 1992.
- MIN, W.; JONES, D. H. In vitro splicing of concanavalin A is catalized by asparaginyl endopeptidase. **Nat. Struct. Mol. Biol.**, [s.l.], v. 1, n. 8, p.502-504, ago. 1994.
- MIRKOV, T. E. *et al.* Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-alpha-amylase inhibitor family of the common bean and its relatives. **Plant. Mol. Biol.**, [s.l.], v. 26, n. 4, p.1103-1113, nov. 1994.
- MOREIRA, R. A. *et al.* Isolation and characterization of a lectin from the seeds of *Dioclea grandiflora* (Mart.). **Planta**, [s.l.], v. 158, n. 1, p.63-69, jun. 1983.
- MÜNTZ, K. Protein dynamics and proteolysis in plant vacuoles. **J. Exp. Bot.**, [s.l.], v. 58, n. 10, p.2391-2407, 7 jun. 2007.
- MURSHUDOV, G. N. *et al.* REFMAC 5 for the refinement of macromolecular crystal structures. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 67, n. 4, p.355-367, 18 mar. 2011.
- NAGANO, C. S. *et al.* Insights into the structural basis of the pH-dependent dimer-tetramer equilibrium through crystallographic analysis of recombinant *Diocleinae* lectins. **Biochem. J.**, [s.l.], v. 409, n. 2, p.417-428, 15 jan. 2008.
- NAISMITH, J. H.; FIELD, R. A. Structural basis of trimannoside recognition by concanavalin A. **J. Biol. Chem.**, [s.l.], v. 271, n. 2, p.972-976, jan. 1996.
- NASCIMENTO, K.S. *et al.* An overview of lectins purification strategies. **J. Mol. Recognit.**, [s.l.], v. 25, n. 11, p. 527-541, nov. 2012.
- NISHIO, M. *et al.* CH- π hydrogen bonds in biological macromolecules. **Phys. Chem. Chem. Phys.**, [s.l.], v. 16, n. 25, p.12648-12683, jul. 2014.
- NÓBREGA, R. B. *et al.* Structure of *Dioclea virgata* lectin: relations between carbohydrate binding site and nitric oxide induction. **Biochimie**, [s.l.], v. 94, n. 2, p.900-906, nov. 2012.
- NÓBREGA, R. B. **Estruturas cristalográficas da lectina de *Canavalia brasiliensis* - ConBr complexadas com adenina e ribose: análise estrutural e perspectivas.** 2016. 102 f. Tese (Doutorado) - Curso de Bioquímica, Universidade Federal do Ceará, Fortaleza, 2016.
- OLIVEIRA, T. M. *et al.* Crystal structure of *Dioclea rostrata* lectin: Insights into understanding the pH-dependent dimer-tetramer equilibrium and the structural basis for carbohydrate recognition in Diocleinae lectins. **J. Struct. Biol.**, [s.l.], v. 164, n. 2, p.177-182, nov. 2008.
- OSBORNI, T. C. *et al.* Insecticidal Activity and Lectin Homology of Arcelin Seed Protein. **Science**, [s.l.], v. 240, n. 4849, p.207-210, 8 abr. 1988.

- PAVONE, V. *et al.* Discovering protein secondary structures: classification and description of isolated alpha-turns. **Biopolymers**, [s.l.], v. 38, n. 6, p. 705-721, jun. 1996.
- PERNOLLET, J. Protein bodies of seeds: ultrastructure, biochemistry, biosynthesis and degradation. **Phytochemistry**, London, v. 17, n. 9, p.1473-1480, mar. 1978.
- PETTERSEN, E. *et al.* UCSF Chimera--A visualization system for exploratory research and analysis. **J. Comput. Chem.**, [s.l.], v. 25, n. 13, p.1605-1612, out. 2004
- PEUMANS, W. J.; HAO, Q.; VAN DAMME, E. J. Ribosome-inactivating proteins from plants: more than RNA N-glycosidases? **Faseb J.**, [s.l.], v. 15, n. 9, p.1493-1506, 1 jul. 2001.
- PEUMANS, W. J.; VAN DAMME, E. J. M. Lectins as plant defense proteins. **Plant Physiol.**, [s.l.], v. 109, n. 2, p.347-352, out. 1995.
- PILOBELLO, K.T.; MAHAL, L.K. Deciphering the glycode: the complexity and analytical challenge of glycomics. **Curr. Op. Chem. Biol.**, [s.l.], v. 11, n. 3, p. 300-305, jun. 2007.
- POWELL, H. R.; JOHNSON, O.; LESLIE, A. G. W. Autoindexing diffraction images with iMosflm. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 69, n. 7, p.1195-1203, 13 jun. 2013.
- PRABU, M. M. *et al.* Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution. **J. Mol. Biol.**, [s.l.], v. 276, n. 4, p. 787-796, mar. 1998.
- RAJASHANKAR, K. R.; RAMAKUMAR, S. Pi-turns in proteins and peptides: classification, conformation, occurrence, hydration and sequence. **Prot. Sci.**, [s.l.], v. 5, n. 5, p. 932-946, maio 1996.
- RAMIS, C. *et al.* Deglycosylation is necessary but not sufficient for activation of proconcanavalin A. **J. Exp. Bot.**, [s.l.], v. 52, n. 358, p.911-917, maio 2001.
- RICHARDSON, J. S.; GETZOFF, E. D.; RICHARDSON, D. C. The beta bulge: a common small unit of nonrepetitive protein structure. **Proc. Natl. Acad. Sci. USA**, [s.l.], v. 75, n. 6, p. 2574-2578, jun. 1978.
- RICHARDSON, J. S. The anatomy and taxonomy of protein structure. **Adv. Prot. Chem.**, [s.l.], v. 34, p. 167-339, 1981.
- RINI, J. M. Lectin Structure. **Ann. Rev. Biophys. Biomol. Struct.**, [s.l.], v. 24, n. 1, p. 551-577, jun. 1995.
- ROBERTS, D. D.; GOLDSTEIN, I. J. Binding of hydrophobic ligands to plant lectins: titration with arylaminonaphthalenesulfonates. **Arch. Biochem. Biophys.**, [s.l.], v. 224, n. 2, p.479-484, jul. 1983.

ROCHA, B. A. M. *et al.* Structural basis for both pro- and anti-inflammatory response induced by mannose-specific legume lectin from *Cymbosema roseum*. **Biochimie**, [s.l.], v. 93, n. 5, p.806-816, maio 2011.

ROCHA, B. A. M. *et al.* Structural basis of ConM binding with resveratrol, an anti-inflammatory and antioxidant polyphenol. **Int. J. Biol. Macromol.**, [s.l.], v. 72, n. 1, p. 1136-1142, jan. 2015.

ROSSMANN, M. G. **The molecular replacement method**. [s.i]: Gordon & Breach, 1972.

ROZWARSKI, D. A. *et al.* Crystal Structure of the Lectin from *Dioclea grandiflora* Complexed with Core Trimannoside of Asparagine-linked Carbohydrates. **J. Biol. Chem.**, [s.l.], v. 273, n. 49, p.32818-32825, dez. 1998.

ŠÁ, R. C. S.; ALMEIDA, R. N.; BHATTACHARYYA, J. Pharmaceutical properties and toxicology of *Dioclea grandiflora*. **Pharm. Biol.**, [s.l.], v. 51, n. 5, p.659-667, fev. 2013.
SAMBROOK, J. **Molecular Cloning: A Laboratory Manual**. [s.l]: Cold Spring Harbor Laboratory Press, 2002.

SANZ-APARÍCIO, J. *et al.* The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from Concanavalin A. **Febs Lett.**, [s.l.], v. 405, n. 1, p.114-118, mar. 1997.

SCHUTTER, K.; VAN DAMME, E. Protein-Carbohydrate Interactions as Part of Plant Defense and Animal Immunity. **Molecules**, [s.l.], v. 20, n. 5, p.9029-9053, 19 maio 2015.

SENEAR, D. F.; TELLER, D. C.. Thermodynamic studies of concanavalin a dimer-tetramer equilibria. **Biophys. J.**, [s.l.], v. 32, n. 1, p.433-436, out. 1980.

SENEAR, D. F.; TELLER, D. C. Effects of saccharide and salt binding on dimer-tetramer equilibrium of concanavalin A. **Biochemistry**, [s.l.], v. 20, n. 11, p.3083-3091, maio 1981.

ŠEBESTÍK, Jaroslav; REINIŠ, Milan; JEŽEK, Jan. **Sugar code (glycocode)**. In: Biomedical applications of peptide-, glyco- and glycopeptide dendrimers, and analogous dendrimeric structures. Vienna: Springer, 2012, p. 195.

SHARMA, V.; SUROLIA, A. Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. **J. Mol. Biol.**, [s.l.], v. 267, n. 2, p. 433-445, mar. 1997.

SHARON, N.; LIS, H. History of lectins: from hemagglutinins to biological recognition molecules. **Glycobiology**, [s.l.], v. 14, n. 11, p.53-62, 30 jun. 2004.

SCHEINER, S.. Weak H-bonds. Comparisons of CH \cdots O to NH \cdots O in proteins and PH \cdots N to direct P \cdots N interactions. **Phys. Chem. Chem. Phys.**, [s.l.], v. 13, n. 31, p.13860-13872, ago. 2011.

SHELDON, P. S.; BOWLES, D. J. The glycoprotein precursor of concanavalin A is converted to an active lectin by deglycosylation. **Embo J.**, [s.l.], v. 11, n. 4, p.1297-1301, abr. 1992.

- SHOHAM, M.; KALB, A. J.; PECHT, I. Specificity of metal ion interaction with concanavalin A. **Biochemistry**, [s.l.], v. 12, n. 10, p.1914-1917, maio 1973.
- SILVA, L. L. S. *et al.* Avaliação da atividade antimicrobiana de extratos de *Dioclea grandiflora* Mart. ex. Benth., Fabacea. **Rev. Bras. Farmacogn.**, Curitiba, v. 20, n. 2, p. 208-214, abr. 2010.
- SILVA-FILHO, J.C. **Estrutura Cristalográfica da Lectina de Sementes da *Canavalia maritima* Aub. em Complexo com o Ácido Gama-Aminobutírico (GABA) e a Adenina Revela Novas Características Estruturais e Prediz Mecanismo Antiherbivoria.** 2013. 107 f. Dissertação (Mestrado) - Curso de Biologia Molecular e Celular, Universidade Federal da Paraíba, João Pessoa, 2013.
- SINHA, S. *et al.* Subunit assembly of plant lectins. **Curr. Opin. Struct. Biol.**, [s.l.], v. 17, n. 5, p.498-505, out. 2007.
- SOUSA, B. L. **Caracterização parcial de uma Pro-Lectina funcional de sementes de *Dioclea grandiflora* Benth expressa em *Escherichia coli*.** 2010. 93 f. Dissertação (Mestrado) - Curso de Bioquímica, Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2010.
- SOUSA, B. L. *et al.* Structural characterization of a *Vatairea macrocarpa* lectin in complex with a tumor-associated antigen: A new tool for cancer research. **Int. J. Biochem. Cell Biol.**, [s.l.], v. 72, n. 1, p.27-39, mar. 2016a.
- SOUSA, B. L. *et al.* Structure prediction and functional analysis of a non-permuted lectin from *Dioclea grandiflora*. **Biochimie**, [s.l.], v. 131, p.54-67, dez. 2016b.
- SRINIVAS, V. R. *et al.* Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. **Biochim. Biophys. Acta**, [s.l.], v. 1527, n. 3, p.102-111, ago. 2001.
- STEINER, T.; DESIRAJU, G. R.. Distinction between the weak hydrogen bond and the van der Waals interaction. **Chem. Commun.**, [s.l.], n. 8, p.891-892, fev. 1998.
- STINISSEN, H. M.; PEUMANS, W. J.; CARLIER, A. R. Two-step processing of in vivo synthesized rice lectin. **Plant. Mol. Biol.**, [s.l.], v. 2, n. 1, p.33-40, jan. 1983.
- TEIXEIRA, E. H. *et al.* In vitro inhibition of *Streptococci* binding to enamel acquired pellicle by Plant Lectins. **J. Appl. Microbiol.**, [s.l.], v. 101, n. 1, p.111-116, jul. 2006.
- TEIXEIRA, C. S. *et al.* Crystal structure of the lectin of *Camptosema pedicellatum*: implications of a conservative substitution at the hydrophobic subsite. **J. Biochem.**, [s.l.], v. 152, n. 1, p.87-98, 2 maio 2012.
- TROTT, O; OLSON, A. J.. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **J. Comput. Chem.**, [s.l.], p.455-461, jun. 2009.

- TSCHAMPEL, S. M.; WOODS, R. J.. Quantifying the Role of Water in Protein–Carbohydrate Interactions. **J. Phys. Chem. A**, [s.l.], v. 107, n. 43, p.9175-9181, out. 2003.
- VAN DAMME, E. J. M.; LANNOO, N.; PEUMANS, W. J. Plant Lectins. **Adv. Bot. Res.**, [s.l.], v. 48, n. 1, p.107-209, jan. 2008.
- VANDENBORRE, G.; SMAGGHE, G.; VAN DAMME, E. J. M. Plant lectins as defense proteins against phytophagous insects. **Phytochemistry**, [s.l.], v. 72, n. 13, p.1538-1550, set. 2011.
- VASCONCELOS, I. M.; OLIVEIRA, J. T. A. Antinutritional properties of plant lectins. **Toxicon**, [s.l.], v. 44, n. 4, p.385-403, set. 2004.
- VENKATACHALAM, C. M. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. **Biopolymers**, [s.l.], v. 6, n. 10, p. 1425-1436, out. 1968.
- VIJAYAN, M.; CHANDRA, N. Lectins. **Curr. Opin. Struct. Biol.**, [s.l.], v. 9, n. 6, p. 707-714, dez. 1999.
- VOLKAMER, A. *et al.* DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. **Bioinformatics**, [s.l.], v. 28, n. 15, p.2074-2075, maio 2012.
- WAHL, M.; SUNDARALINGAM, M. C-H...O hydrogen bonding in biology. **Trends Biochem. Sci.**, [s.l.], v. 22, n. 3, p.97-102, mar. 1997.
- WAXDAL, M. J. *et al.* Isolation and order of the cyanogen bromide fragments of concanavalin A. **Biochemistry**, [s.l.], v. 10, n. 18, p.3343-3347, ago. 1971.
- WEATHERMAN, R. V. *et al.* Specificity of C -Glycoside Complexation by Mannose/Glucose Specific Lectins. **Biochemistry**, [s.l.], v. 35, n. 11, p.3619-3624, jan. 1996.
- WINN, M. D. *et al.* Overview of the CCP 4 suite and current developments. **Acta Crystallogr. D Biol. Cryst.**, [s.l.], v. 67, n. 4, p.235-242, 18 mar. 2011.
- WLODAWER *et al.* Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. **FEBS J.**, [s.l.], v. 280, n. 22, p. 5705-5736, nov. 2013.
- YAU, T. *et al.* Lectins with Potential for Anti-Cancer Therapy. **Molecules**, [s.l.], v. 20, n. 3, p. 3791-3810, 26 fev. 2015.
- YOUNG, N. M.; OOMEN, R. P. Analysis of sequence variation among legume lectins: A ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. **J. Mol. Biol.**, [s.l.], v. 228, n. 3, p.924-934, dez. 1992.

ZAMORA-CABALLERO, S. *et al.* Quaternary structure of *Dioclea grandiflora* lectin assessed by equilibrium sedimentation and crystallographic analysis of recombinant mutants. **FEBS Letters**, [s.l.], v. 589, n. 18, p.2290-2296, jul. 2015.

ZHUANG, X. *et al.* Biosynthesis and emission of insect herbivory-induced volatile indole in rice. **Phytochemistry**, [s.l.], v. 73, p.15-22, jan. 2012.

**APÊNDICE A – INDÍCES DE AVALIAÇÃO DA QUALIDADE DOS DADOS
OBTIDOS**

Parâmetro	Significado	Valor esperado	Referência
R_{merge}	Diferença estatística no cálculo da intensidade de um spot localizado em diferentes imagens	< 10%	WLODAWER <i>et al.</i> , 2013
I/σ (σ)	Razão média entre a intensidade calculada e o erro associado a esse cálculo	> 2	WLODAWER <i>et al.</i> , 2013
Multiplicidade	Número de vezes que um mesmo spot foi utilizado para o cálculo da intensidade	> 3	WLODAWER <i>et al.</i> , 2013
Completeza	Percentual de spots utilizados para o cálculo da intensidade	~ 100%	WLODAWER <i>et al.</i> , 2013
CC	Coefficiente de correlação das intensidades quando estas são separadas aleatoriamente em dois conjuntos	~ 100%	KARPLUS; DIEDERICHS, 2012
RFZ	Z-score da função de rotação, onde z-score é o número de desvios-padrões abaixo ou acima da média	> 4	MCCOY <i>et al.</i> , 2007
TFZ	Z-score da função de translação	> 8	MCCOY <i>et al.</i> , 2007
LLG (<i>Log Likelihood Gain</i>)	Probabilidade de que a solução encontrada seja melhor do que um arranjo aleatório de átomos	> 120	MCCOY <i>et al.</i> , 2007
R_{factor}	Discordância entre F	< 0.2	WLODAWER <i>et al.</i> , 2013
R_{free}	Discordância entre F aleatório de 1000-5000 reflexões excluídas do refinamento	< 0.2	WLODAWER <i>et al.</i> , 2013

APÊNDICE B – VALIDAÇÃO DO MODELO CRISTALOGRAFICO DA pDGL

Preliminary Full wwPDB X-ray Structure Validation
Report ⓘ

Jul 24, 2016 – 08:52 AM EDT

This is a Preliminary Full wwPDB X-ray Structure Validation Report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027790
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027790

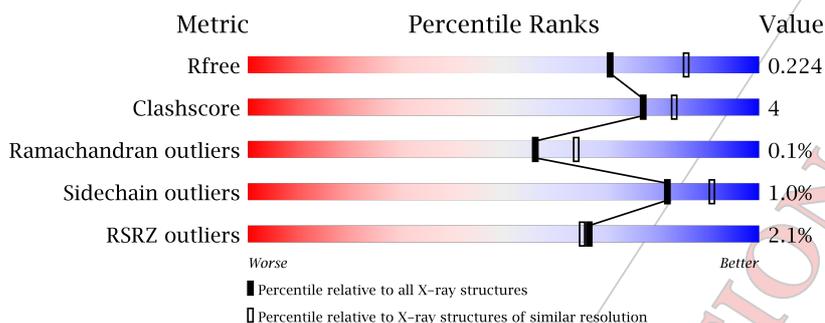
1 Overall quality at a glance [i](#)

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	249	 2% 86% 7% 6%
1	B	249	 3% 87% 6% 7%
1	C	249	 0% 86% 8% 6%
1	D	249	 2% 90% 10%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	EDO	A	265	-	-	-	X

PRELIMINARY VALIDATION REPORT

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 7922 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	232	1778	1125	295	357	1	0	3	0
1	A	233	1783	1128	295	359	1	0	2	0
1	C	234	1799	1137	301	360	1	0	4	0
1	D	249	1904	1202	320	381	1	0	3	0

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Ca	0	0
			1	1		
2	A	1	Total	Ca	0	0
			1	1		
2	C	1	Total	Ca	0	0
			1	1		
2	D	1	Total	Ca	0	0
			1	1		

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Mn	0	0
			1	1		
3	A	1	Total	Mn	0	0
			1	1		
3	C	1	Total	Mn	0	0
			1	1		
3	D	1	Total	Mn	0	0
			1	1		

- Molecule 4 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: unknown).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			12	6	6		
4	A	1	Total	C	O	0	0
			12	6	6		
4	C	1	Total	C	O	0	0
			12	6	6		
4	D	1	Total	C	O	0	0
			12	6	6		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: unknown).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	0
			4	2	2		
5	A	1	Total	C	O	0	0
			4	2	2		
5	D	1	Total	C	O	0	0
			4	2	2		
5	D	1	Total	C	O	0	0
			4	2	2		

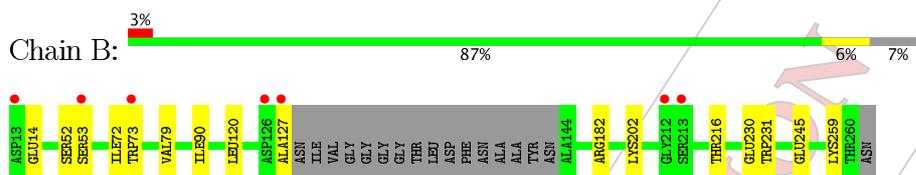
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	H	586	Total	O	0	0
			586	586		

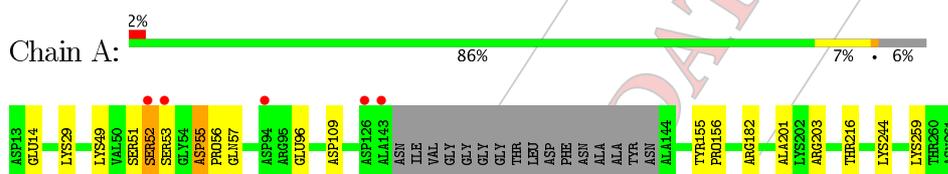
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

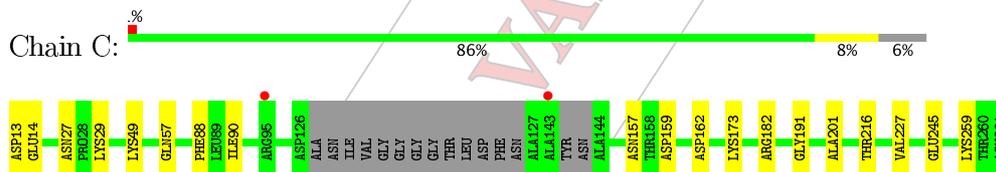
- Molecule 1:



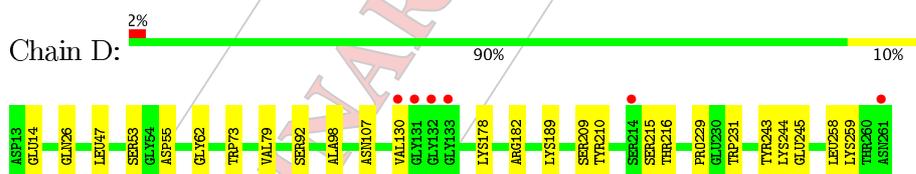
- Molecule 1:



- Molecule 1:



- Molecule 1:



4 Data and refinement statistics 

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	71.72Å 96.61Å 72.13Å 90.00° 98.02° 90.00°	Depositor
Resolution (Å)	71.42 - 2.20 47.18 - 2.20	Depositor EDS
% Data completeness (in resolution range)	92.8 (71.42-2.20) 92.8 (47.18-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.04	Depositor
$\langle I/\sigma(I) \rangle$ ¹	7.42 (at 2.20Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.164 , 0.219 0.178 , 0.224	Depositor DCC
R_{free} test set	2330 reflections (5.35%)	DCC
Wilson B-factor (Å ²)	20.8	Xtrriage
Anisotropy	0.091	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 36.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.088 for l,-k,h	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7922	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.81% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: CA, MN, EDO, MAN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.80	0/1830	0.88	0/2493
1	B	0.82	0/1828	0.85	0/2490
1	C	0.78	0/1852	0.85	0/2523
1	D	0.80	0/1957	0.90	0/2667
All	All	0.80	0/7467	0.87	0/10173

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts i

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1783	0	1731	12	0
1	B	1778	0	1727	13	0
1	C	1799	0	1751	15	0
1	D	1904	0	1844	16	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	12	0	12	0	0
4	B	12	0	12	0	0
4	C	12	0	12	0	0
4	D	12	0	12	0	0
5	A	4	0	6	0	0
5	B	4	0	6	0	0
5	D	8	0	12	2	0
6	H	586	0	0	10	0
All	All	7922	0	7125	53	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

All (53) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:14:GLU:OE2	1:D:259:LYS:HE2	1.73	0.88
1:C:182:ARG:O	1:C:216:THR:HG21	1.74	0.86
1:A:56:PRO:O	1:A:244:LYS:HD2	1.88	0.73
1:D:79:VAL:HG13	1:D:259:LYS:HG3	1.71	0.72
1:B:120:LEU:HD21	1:B:127:ALA:HB2	1.72	0.70
1:B:73:TRP:O	1:B:230:GLU:HG2	1.90	0.69
1:B:14:GLU:OE2	1:B:259:LYS:HE2	1.93	0.68
1:C:27:ASN:HB2	6:H:378:HOH:O	1.94	0.67
1:D:14:GLU:OE2	1:D:259:LYS:CE	2.43	0.65
1:C:157:ASN:O	1:C:162:ASP:HB2	1.98	0.64
1:D:73:TRP:HA	1:D:258:LEU:HD11	1.79	0.64
1:A:96:GLU:HB2	6:H:297:HOH:O	1.98	0.64
1:A:14:GLU:OE2	1:A:259:LYS:HE3	1.98	0.63
1:B:182:ARG:O	1:B:216:THR:HG21	2.02	0.59
1:C:182:ARG:O	1:C:216:THR:CG2	2.50	0.58
1:C:49:LYS:HE2	6:H:244:HOH:O	2.04	0.58
1:B:127:ALA:CB	6:H:196:HOH:O	2.54	0.55
1:B:202:LYS:HE3	6:H:395:HOH:O	2.04	0.55
1:D:189:LYS:HD3	1:D:210:TYR:HB3	1.88	0.54
1:C:182:ARG:NH2	6:H:414:HOH:O	2.34	0.54
1:B:127:ALA:HB3	6:H:196:HOH:O	2.07	0.53
1:C:29:LYS:HD3	1:D:231:TRP:CG	2.44	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:90:ILE:HG23	1:B:245:GLU:HB2	1.91	0.53
1:A:49:LYS:HB2	1:A:57:GLN:HB3	1.92	0.52
1:C:49:LYS:HB3	1:C:57:GLN:HB2	1.91	0.51
1:C:90:ILE:HG23	1:C:245:GLU:HB2	1.94	0.50
1:D:47:LEU:O	1:D:62:GLY:HA3	2.12	0.50
1:D:209:SER:HB3	1:D:215:SER:HB3	1.93	0.50
1:D:178:LYS:NZ	5:D:266:EDO:H12	2.27	0.49
1:A:203:ARG:HD2	1:C:201:ALA:HB1	1.95	0.49
1:D:107:ASN:OD1	1:D:229:PRO:HG3	2.13	0.49
1:D:130:VAL:HG11	5:D:266:EDO:H11	1.96	0.48
1:B:79:VAL:HG13	1:B:259:LYS:HD2	1.95	0.48
1:D:79:VAL:CG1	1:D:259:LYS:HG3	2.41	0.47
1:D:98:ALA:HB3	1:D:243:TYR:O	2.15	0.47
1:B:120:LEU:HD21	1:B:127:ALA:CB	2.42	0.46
1:A:55:ASP:HA	1:A:56:PRO:HD2	1.69	0.46
1:A:109:ASP:OD1	1:A:109:ASP:N	2.43	0.46
1:C:14:GLU:OE2	1:C:259:LYS:HE2	2.17	0.45
1:B:79:VAL:CG1	1:B:259:LYS:HD2	2.48	0.44
1:C:13:ASP:CG	6:H:585:HOH:O	2.55	0.44
1:C:88:PHE:CZ	1:C:191:GLY:HA3	2.53	0.44
1:A:201:ALA:HB2	6:H:15:HOH:O	2.18	0.43
1:C:173:LYS:HD3	1:C:227:VAL:HG13	2.00	0.43
1:B:72:ILE:O	1:B:73:TRP:HB3	2.19	0.42
1:A:182:ARG:O	1:A:216:THR:HG21	2.19	0.42
1:A:51:SER:O	1:A:52:SER:HB3	2.19	0.42
1:D:26:GLN:HG2	6:H:402:HOH:O	2.20	0.42
1:C:88:PHE:CE2	1:C:191:GLY:HA3	2.55	0.41
1:D:92:SER:HB2	1:D:245:GLU:HA	2.02	0.41
1:A:155:TYR:HA	1:A:156:PRO:HD3	1.83	0.41
1:B:231:TRP:HB2	1:A:29:LYS:HE3	2.03	0.41
1:D:182:ARG:O	1:D:216:THR:HG21	2.21	0.41

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	231/249 (93%)	222 (96%)	8 (4%)	1 (0%)	39	42
1	B	231/249 (93%)	223 (96%)	8 (4%)	0	100	100
1	C	234/249 (94%)	223 (95%)	11 (5%)	0	100	100
1	D	250/249 (100%)	243 (97%)	7 (3%)	0	100	100
All	All	946/996 (95%)	911 (96%)	34 (4%)	1 (0%)	56	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	52	SER

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	202/209 (97%)	200 (99%)	2 (1%)	82	91
1	B	201/209 (96%)	199 (99%)	2 (1%)	82	91
1	C	203/209 (97%)	202 (100%)	1 (0%)	92	96
1	D	212/209 (101%)	209 (99%)	3 (1%)	74	85
All	All	818/836 (98%)	810 (99%)	8 (1%)	82	91

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	52	SER
1	B	53	SER
1	A	53	SER
1	A	55	ASP
1	C	159	ASP
1	D	53	SER
1	D	55	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	244	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	22	HIS

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 16 ligands modelled in this entry, 8 are modelled with single atom - leaving 8 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
4	MAN	A	264	-	12,?,?	0.42	0	17,?,?	0.51	0
5	EDO	A	265	-	3,?,?	0.44	0	2,?,?	0.40	0
4	MAN	B	264	-	12,?,?	0.43	0	17,?,?	0.50	0
5	EDO	B	265	-	3,?,?	0.76	0	2,?,?	0.17	0
4	MAN	C	264	-	12,?,?	0.42	0	17,?,?	0.50	0
4	MAN	D	264	-	12,?,?	0.43	0	17,?,?	0.51	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
5	EDO	D	265	-	3,?,?	0.33	0	2,?,?	0.60	0
5	EDO	D	266	-	3,?,?	0.49	0	2,?,?	0.18	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	MAN	A	264	-	-	0/2/?/?	0/1/?/?
5	EDO	A	265	-	-	0/1/?/?	0/0/?/?
4	MAN	B	264	-	-	0/2/?/?	0/1/?/?
5	EDO	B	265	-	-	0/1/?/?	0/0/?/?
4	MAN	C	264	-	-	0/2/?/?	0/1/?/?
4	MAN	D	264	-	-	0/2/?/?	0/1/?/?
5	EDO	D	265	-	-	0/1/?/?	0/0/?/?
5	EDO	D	266	-	-	0/1/?/?	0/0/?/?

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	D	266	EDO	2	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	A	233/249 (93%)	-0.31	5 (2%) 67 65	10, 18, 36, 40	0
1	B	232/249 (93%)	-0.38	7 (3%) 54 53	10, 18, 38, 40	0
1	C	234/249 (93%)	-0.39	2 (0%) 85 85	11, 19, 36, 40	0
1	D	249/249 (100%)	-0.33	6 (2%) 62 61	11, 19, 38, 40	0
All	All	948/996 (95%)	-0.35	20 (2%) 67 65	10, 19, 38, 40	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	143	ALA	6.0
1	D	131	GLY	5.5
1	D	130	VAL	5.1
1	B	127	ALA	4.4
1	C	143	ALA	4.3
1	D	261	ASN	3.4
1	B	126	ASP	3.2
1	A	52	SER	2.8
1	D	132	GLY	2.8
1	B	213	SER	2.5
1	B	13	ASP	2.4
1	B	53	SER	2.4
1	B	73	TRP	2.3
1	D	133	GLY	2.3
1	B	212	GLY	2.2
1	A	94	ASP	2.2
1	A	53	SER	2.2
1	D	214	SER	2.0
1	A	126	ASP	2.0
1	C	95	ARG	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
5	EDO	A	265	4/?	0.82	0.23	5.24	20,20,20,20	0
4	MAN	A	264	12/?	0.95	0.11	0.21	16,19,22,24	0
4	MAN	D	264	12/?	0.95	0.10	-0.17	17,18,19,19	0
4	MAN	B	264	12/?	0.96	0.10	-0.35	15,18,21,22	0
4	MAN	C	264	12/?	0.95	0.09	-0.98	18,21,22,24	0
3	MN	D	263	1/?	1.00	0.07	-1.68	14,14,14,14	0
2	CA	B	262	1/?	0.98	0.06	-1.94	15,15,15,15	0
3	MN	A	263	1/?	1.00	0.07	-2.03	16,16,16,16	0
3	MN	B	263	1/?	1.00	0.07	-2.97	16,16,16,16	0
2	CA	A	262	1/?	0.99	0.05	-3.35	15,15,15,15	0
2	CA	C	262	1/?	0.98	0.06	-3.49	18,18,18,18	0
2	CA	D	262	1/?	1.00	0.05	-3.88	13,13,13,13	0
3	MN	C	263	1/?	0.99	0.04	-4.11	17,17,17,17	0
5	EDO	D	265	4/?	0.84	0.19	-	33,36,39,39	0
5	EDO	B	265	4/?	0.93	0.18	-	23,25,25,27	0
5	EDO	D	266	4/?	0.91	0.46	-	33,34,35,36	0

6.5 Other polymers [i](#)

There are no such residues in this entry.