UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

Caracterização estrutural e avaliação de aspectos funcionais de
galectinas humanas do grupo *tandem-repeat*

Joane Kathelen Rustiguel Bonalumi

Ribeirão Preto

2014

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

**Caracterização estrutural e avaliação de aspectos funcionais
de galectinas humanas do grupo *tandem-repeat***

Tese de Doutorado apresentada ao Programa
de Pós-Graduação em Ciências Farmacêuticas
para obtenção do Título de Doutor em Ciências

Área de Concentração: Química e Física
Biológica

Orientada: Joane Kathelen Rustiguel Bonalumi

Orientadora: Profa. Dra. Maria Cristina Nonato
Costa

Co-orientador: Prof. Dr. Marcelo Dias Baruffi

Versão corrigida da Tese de Doutorado apresentada ao Programa de Pós-Graduação
em Ciências Farmacêuticas em 12/11/2014. A versão original encontra-se disponível
na Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP.

Ribeirão Preto

2014

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Rustiguel Bonalumi, Joane Kathelen

Caracterização estrutural e avaliação de aspectos funcionais de galectinas humanas do grupo *tandem-repeat*. Ribeirão Preto, 2014.
178 p.

Tese de Doutorado, apresentada à Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP – Área de concentração: Química e Física biológica.

Orientadora: Nonato, Maria Cristina

Co-orientador: Dias-Baruffi, Marcelo

1. galectinas. 2. Domínio de reconhecimento de carboidratos.
3. Cristalografia de proteínas.

Joane Kathelen Rustiguel Bonalumi

Caracterização estrutural e avaliação de aspectos funcionais de galectinas humanas do grupo *tandem-repeat*

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas para obtenção do Título de Doutor em Ciências

Área de Concentração: Química e Física Biológica.

Orientadora: Profa. Dra. Maria Cristina Nonato Costa

Co-orientador: Prof. Dr. Marcelo Dias Baruffi

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Ao meu querido e amado esposo

Carlos

“Você é luz ...”

AGRADECIMENTOS

À Faculdade de Ciências Farmacêuticas de Ribeirão Preto e ao Programa de Pós-graduação em Ciências Farmacêuticas pela oportunidade de realização do doutorado

À Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP pela concessão da bolsa de doutorado e pelo apoio financeiro para a realização dessa pesquisa

À minha orientadora Profa. Dra. Maria Cristina Nonato Costa pela oportunidade de realização desse projeto em seu laboratório

Ao meu co-orientador Prof. Dr. Marcelo Dias-Baruffi pelo apoio durante o desenvolvimento do projeto

A todos do laboratório de Glicoimunologia em especial ao Camillo e ao Daniel Cerri, pelo apoio no desenvolvimento de algumas etapas

Às técnicas Francine e Lilian

A todos do laboratório de Biofísica Molecular

Ao Instituto de Física de São Carlos – Grupo de Biofísica Molecular “Sérgio Mascarenhas”

Ao Laboratório Nacional de Luz Síncrotron

Em especial, aos meus amigos e companheiros de trabalho Paty, Renata, Valéria, Juliana, Iara, Ricardo Pádua, Ricardo Soares, Éder, Giovani, Guilherme, Vinícius, Matheus e João Paulo

À minha família e aos meus queridos, Carlos, Isis e Farid, por tornarem essa etapa mais suave e amena

RESUMO

RUSTIGUEL BONALUMI, J. K. **Caracterização estrutural e avaliação de aspectos funcionais de galectinas do grupo *tandem-repeat***. 2014. 178f. Tese de Doutorado. Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2014.

As galectinas, proteínas que compartilham características como afinidade por β -galactosídeos e a presença de um domínio de reconhecimento ao carboidrato (CRD), tem como importante papel a decodificação de mensagens moleculares. As galectinas são encontradas em uma grande variedade de células e tecidos animais e estão envolvidas em diversos processos celulares relacionados a resposta imune e inflamatória. O grupo *tandem-repeat* das galectinas é formado por proteínas que apresentam dois CRDs distintos (CRD1 e CRD2) conectados por um peptídeo de ligação, ao qual pertencem as galectinas-4 e -12 humanas, alvos de nosso estudo. Durante o desenvolvimento do presente projeto foram utilizadas abordagens multidisciplinares através de técnicas biofísicas, cristalográficas, computacionais e imunoquímicas. Foi iniciado o estudo de caracterização estrutural da galectina-4 humana (hGal4) e seus domínios hGal4-CRD1 e hGal4-CRD2, de forma independente, através da produção heteróloga das proteínas e ensaios de estabilidade térmica e cristalização. As estruturas cristalográficas dos domínios foram determinadas a 1,48 e 2,46 Å de resolução, respectivamente e utilizadas para a construção de um modelo para a hGal4. Com base nos estudos de dinâmica molecular e estabilidade térmica foi possível propor um modelo de interação entre os CRDs através do peptídeo de ligação. Ensaio de dinâmica da hGal4 com LacNAc sugeriram uma diferença de plasticidade dos CRDs no reconhecimento do ligante. Foram também realizadas incessantes tentativas de desenvolver um protocolo de expressão heteróloga para as isoformas e domínios da galectina-12 humana (hGal12). Como resultado, observamos uma alta tendência à formação de corpos de inclusão e agregados resistentes a desnaturação, além da susceptibilidade ao ataque proteolítico. Os modelos estruturais dos domínios da hGal12 não permitiram identificar nenhuma característica que justificasse o comportamento das proteínas. Foram realizados estudos de localização celular em células HL-60 e foi possível identificar a presença da hGal12 na superfície das gotas de lipídio, organelas responsáveis pelo armazenamento de energia e o processo de catabolismo celular. A alta insolubilidade observada para as isoformas e domínios da hGal12, a sua localização em gotas de lipídeos, a divergência evolutiva da hGal12 quando comparada com outras galectinas do mesmo grupo e incluindo o fato de um dos CRDs não apresentar os requerimentos estruturais básicos para ligar β -galactosídeos, nos leva a especular se essa proteína estaria, na verdade, sendo expressa como parte de um heterocomplexo proteico, que estabilizaria a estrutura da hGal12 e a endereçaria para as gotas de lipídeo. Em nossa hipótese, o domínio hGal12-CRD2 poderia ter evoluído para favorecer a interação com outros parceiros macromoleculares. Devido ao importante papel desempenhado pelas galectinas em inúmeros processos celulares, os resultados aqui obtidos representam uma importante contribuição no entendimento do papel que as galectinas hGal4 e hGal12 exercem nos diferentes eventos celulares, além de fornecerem ferramentas experimentais para desenvolvimento de futuros estudos funcionais.

Palavras-chave: galectinas, domínio de reconhecimento de carboidratos, cristalografia de proteínas

ABSTRACT

RUSTIGUEL BONALUMI, J. K. **Structural characterization and evaluation of functional aspects of human galectins from tandem-repeat group.** 2014. 178f. Thesis of Doctoral. Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2014.

Galectins are proteins that share characteristics such as affinity for β -galactosides and the presence of a carbohydrate recognition domain (CRD). They belong to a family of proteins that display the important role of decoding molecular messages. Galectins are found in a variety of cell types and are involved in several biological phenomena related to immune response and inflammation. The tandem-repeat group of galectins consists of proteins with two distinct CRDs (CRD1 and CRD2) connected by a peptide linker, in which belong galectins-4 and -12, targets of our study. During the development of the present project a multidisciplinary approach was used including the use of biophysical, crystallographic, computational and immunochemical techniques. The structural characterization of human galectin-4 (hGal4) and its hGal4-CRD1 and hGal4-CRD2 independent domains was initiated by their heterologous protein production, thermal stability and crystallization assays. The crystallographic structure of domains were determined at 1.48 e 2.46 Å resolution, respectively, and used for constructing a model for hGal4. Based on molecular dynamics and thermal stability studies we propose a model of interaction between CRDs mediated by linker peptide. Dynamics simulation of hGal4 with LacNAc suggested a difference in plasticity between CRDs and ligand recognition. In addition, several attempts have been made towards the development of a protocol for expression of isoforms and independent domains from human galectin-12 (hGal12). As result, we observed a high tendency to body inclusion formation and denaturing resistant aggregates, besides high susceptibility to proteolytic attack. Structural models for the hGal12 domains did not allow the identification of any feature to justify proteins behavior. Cell location studies in HL-60 cells were performed and hGal12 was found to be located on the surface of lipid droplets, organelles responsible for energy storage and catabolism. The high insolubility displayed by isoforms and domains from hGal12, together with its location on lipid droplets, the evolutionary divergence of hGal12 compared to tandem-repeat proteins, and the fact that hGal12-CRD2 does not present the structural requirements for β -galactosides binding, suggest that hGal12 is, in fact, expressed as part of a protein complex that could both stabilize hGal12 structure and guide it to the lipid droplets. In our hypothesis, the hGal12-CRD2 could have evolved in order to favor the interaction with other macromolecular partners. Due to crucial roles displayed by galectins in innumerable biological process, we believe that our results represent an important contribution for the understanding of hGal4 e hGal12 proteins roles within the cell, besides providing experimental tools for the development of further functional studies.

Keywords: galectins, carbohydrate recognition domains, protein crystallography

LISTA DE FIGURAS

Figura 1 - Representação estrutural dos diferentes tipos de enovelamentos característicos em lectinas. A) α/β misto em lectinas do tipo-C em invertebrados (PDB: 2VUZ); B) β -prisma I em lectinas que ligam a manose (PDB: 1C3K); C) β -prisma II em aglutinina de plantas (PDB: 1KJ1); D) β -propeller de lectina de fungo (PDB: 2C4D); E) β -trefoil em mamífero (PDB: 1FWV); F) tipo-C em bactérias (PDB: 1PWB); G) fibrinogênio- <i>like</i> em caranguejo-ferradura (PDB: 1JC9); H) imunoglobulina- <i>like</i> em lectinas de mamífero da família siglec (PDB: 1QFO); I) <i>jelly-roll</i> em lectinas de legumes (PDB: 3ZYR); J) tipo-P em receptores de manose-6-fosfato de mamífero (PDB: 1M6P).....	4
Figura 2 - Representação da superfície de lectinas evidenciando os bolsos de interação com os ligantes. A) lectina de leguminosas. Em amarelo, um complexo de glicanos interage na superfície da proteína (PDB: 3ZYR). B) galectina-1 de mamífero. Em amarelo, um complexo de glicanos interage entre dímeros promovendo o <i>crosslinking</i> entre eles (PDB: 1SLA).....	5
Figura 3 - Esquema ilustrativo dos três subgrupos das galectinas: proto, químera e <i>tandem-repeat</i> . O tipo proto é composto somente de um domínio lectínico, o CRD. O tipo químera é composto de duas partes: o domínio C-terminal correspondendo ao CRD e um domínio N-terminal não lectínico. O tipo <i>tandem-repeat</i> é composto de dois CRDs distintos em uma cadeia polipeptídica única.	7
Figura 4 - Topologia das galectinas. A) Representação em <i>cartoon</i> do monômero da estrutura cristalográfica da galectina-1 humana em complexo com lactose. PDB: 1GZW. B) Representação da topologia da galectina-1 humana pelo gráfico gerado pelo banco de dados PDBsum (http://www.ebi.ac.uk/pdbsum/).....	8
Figura 5 - Representação do homodímero da galectina-1 humana. A lactose está localizada no bolso de ligação do CRD compreendido pelas fitas S4, S5 e S6. PDB: 1GZW.	9
Figura 6 - Representação dos principais resíduos envolvidos na coordenação com o carboidrato. PDB 1GZW.	9
Figura 7 – Alinhamento sequencial entre o CRD1 e CRD2 da hGal4. Os aminoácidos envolvidos na coordenação das moléculas de carboidratos estão marcados em ciano. A mutação conservativa em um dos resíduos envolvidos na coordenação está identificada em rosa. O alinhamento foi feito com o programa MULTALIN (Corpet, 1988) e a imagem foi gerada com o programa Esript (Robert e Gouet, 2014).....	18
Figura 8 – Região de clonagem múltipla do vetor pET-28a-c(+). Fonte: http://richsingiser.com/4402/pET28.pdf . 23	
Figura 9 – Representação esquemática da construção resultante da clonagem no vetor pET-28a-sumo.....	24
Figura 10 – Representação esquemática de um poço da placa <i>Crychem</i> utilizada nos experimentos de cristalização... 30	
Figura 11 – Representação esquemática do procedimento de <i>streak seeding</i>	32
Figura 12 – Perfil de purificação da hGal4 em coluna de níquel. MM – PageRuler marcador de peso molecular (kDa) (<i>Thermo Scientific</i>) A) Purificação por cromatografia de afinidade em coluna de níquel: 1) pellet 2) sobrenadante 3) fração 1 do efluente 4) fração 1 da lavagem com tampão sem imidazol 5) fração 1 da lavagem com tampão 25 mM imidazol 6) fração 1 da lavagem com 40 mM imidazol 7-10) frações 1, 2, 3 e 5 da eluição com 500 mM de imidazol. B) Purificação por cromatografia de afinidade em coluna de níquel após a clivagem com a protease ULP1: 1) amostra proteica pós clivagem com ULP1 2-4) frações correspondentes ao efluente 1, 2 e 3, respectivamente, do material aplicado na coluna 5-6) frações correspondentes a eluição com tampão sem imidazol 1 e 2; 7-11) frações de 1 a 5 da eluição com 25 mM de imidazol da proteína clivada; 12-14) frações de 1 a 3 da eluição das proteínas sumo e ULP1, com 500 mM de imidazol.....	41

Figura 13 – Perfil cromatográfico de purificação da hGal4 por cromatografia de exclusão por tamanho em coluna Superdex™ 200 10/300. SDS-PAGE 14% das frações coletadas da cromatografia de exclusão por tamanho. Poços 1) 8-9 ml; 2) 11-12 ml; 3) 13-14 ml; 4)16-17 ml; 5-11) 17,5-21 ml; 12) 22-23 ml; 13) 24-25 ml.....	41
Figura 14 – Perfil de purificação do domínio hGal4-CRD1 por cromatografia de afinidade em coluna de níquel; MM – PageRuler marcador de peso molecular (kDa) (<i>Thermo Scientific</i>): 1) pellet 2) sobrenadante 3) fração 1 do efluente 4) fração da lavagem com tampão sem imidazol 5) fração 1 da lavagem com tampão 25 mM imidazol 6) fração 1 da lavagem com 40 mM imidazol 7-13) frações 1, 2, 3, 5 e 10 da eluição com 500 mM de imidazol.....	42
Figura 15 – Perfil cromatográfico de purificação da hGal4-CRD1 por cromatografia de exclusão por tamanho em coluna Superdex™ 200 10/300. SDS-PAGE 14% das frações coletadas: Poços 1) 8-9 ml; 2) 11-12 ml; 3) 15-16 ml; 4)16-17 ml; 5-11) 17,5-21 ml; 12) 22-23 ml; 13) 24-25 ml.....	43
Figura 16 – Perfil de purificação do domínio hGal4-CRD2 por cromatografia de afinidade em coluna de níquel; MM – PageRuler marcador de peso molecular (kDa) (<i>Thermo Scientific</i>): 1) pellet 2) sobrenadante 3) fração 1 do efluente 4) fração da lavagem com tampão sem imidazol 5) fração 1 da lavagem com tampão 25 mM imidazol 6) fração 1 da lavagem com 40 mM imidazol 7-13) frações 1, 2, 3, 4, 5 e 10 da eluição com 500 mM de imidazol.	44
Figura 17 – Perfil cromatográfico de purificação da hGal4-CRD2 por cromatografia de exclusão por tamanho em coluna Superdex™ 200 10/300. SDS-PAGE 14% das frações coletadas: Poços 1) 8-9 ml; 2) 11-12 ml; 3) 14-15 ml; 4) 15-16 ml; 5) 16-17 ml; 6) 17-18 ml; 7) 18-18,5 ml; 8-12) 18,5-21 ml; 13) 24-25 ml	44
Figura 18 – Cristais da proteína nas condições contendo 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado.	45
Figura 19 – Cristal de hGal4-CRD1 obtidos na presença de 10 mM de lactose, nas condições contendo 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado.....	47
Figura 20 – A) Representação em <i>cartoon</i> da unidade assimétrica da estrutura cristalográfica do domínio hGal4-CRD1. B) Representação em superfície da molécula na unidade assimétrica e sua simétrica gerado pela operação (x-y,x,z+1/6). Em azul, representado como sticks, estão os principais resíduos que interagem com o carboidrato	48
Figura 21 – Cristais de hGal4-CRD1 obtidos na presença do ligante C3S, nas condições contendo 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado.	49
Figura 22 – Foto de duas gotas de cristalização da proteína hGal4-CRD2 na condição de cristalização contendo 1,8 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 2 % (v/v) PEG MME 550.	51
Figura 23 – Foto dos cristais obtidos no experimento de variação do volume da gota. Cada cristal foi obtido com diferentes volumes entre a solução de proteína e a solução do reservatório no poço.....	52
Figura 24 – Foto de um experimento de microsemeadura nas condições descritas acima.	52
Figura 25 – Foto dos cristais obtidos no experimento de <i>streak seeding</i>	53
Figura 26 – Foto de cristais típicos da hGal4-CRD2.....	53
Figura 27 – Mapa de densidade eletrônica 2mFobs-DFcalc, com σ (nível de contorno) igual a 2, representado em azul.	55
Figura 28 – Representação da topologia da hGal4-CRD1.....	56
Figura 29 – Representação em <i>cartoon</i> das fitas S4, S5, S6a e S6b compreendendo os principais resíduos de interação com o carboidrato.....	57
Figura 30 – Representação da superfície de potencial eletrostático da hGal4-CRD1 em diferentes posições, mostrando as posições frontal das folhas S1-S6a/b e F1-F6 e lateral das fitas F1 e S1 e F6 e S6a/b.	57

Figura 31 – Representação em <i>cartoon</i> do modelo de saída da etapa de substituição molecular nos dados da hGal4-CRD2. A figura mostra a unidade assimétrica com 4 moléculas, nas posições A) lateral, B) topo e C) frontal.	58
Figura 32 - Mapa de densidade eletrônica 2mFobs-DFcalc, com σ (nível de contorno) igual a 1, representado em azul.	59
Figura 33 - Representação da topologia da hGal4-CRD2.	60
Figura 34 - Representação em <i>cartoon</i> das fitas S4, S5, S6 compreendendo os principais resíduos de interação com o carboidrato.	60
Figura 35 - Representação da superfície de potencial eletrostático da hGal4-CRD1 em diferentes posições, mostrando as posições frontal das folhas S1-S6a e F1-F6 e lateral das fitas F1 e S1 e F6 e S6a.	61
Figura 36 – Representação em <i>cartoon</i> dos modelos gerados para o peptídeo de ligação entre os CRDs da galectina-4.	62
Figura 37 – Representação em <i>cartoon</i> dos dois melhores modelos gerados para a galectina-4.	62
Figura 38 – Representação dos modelos inicial e final da galectina-4. A) Vetores do tipo porco-espinho (“porcupine”) feitos com ajuda do script “Porcupine” do Laboratório de Modelagem e Dinâmica Molecular da UFRJ (Prof. Pascutti http://lmdm.biof.ufrj.br/software/porcupine/index.html). B) Representação em <i>cartoon</i> . As figuras foram feitas no programa VMD.	63
Figura 39 – Gráfico de Ramachandran do modelo final da galectina-4 humana.	63
Figura 40 – A) Gráfico de RMSD dos domínios hGal4-CRD1 e hGal4-CRD2 e do peptídeo de ligação em cada momento da simulação comparado com a posição inicial dos mesmos e B) matriz de RMSD da hGal4 em cada momento da simulação comparado com a sua posição inicial.	64
Figura 41 – Clusterização das estruturas visitadas durante a trajetória. Os dados indicam que a estrutura da proteína apresenta predominância de dois estados (clusters). B) A flexibilidade da cadeia, de acordo com o RMSF. Maiores flexibilidades se encontram no N-terminal e no peptídeo de ligação.	65
Figura 42 – Representação em <i>cartoon putty</i> com coloração do fator de temperatura do modelo da hGal4.	65
Figura 43 – Perfil da ocorrência de ligações de hidrogênio ao longo da trajetória. Cada curva descreve a evolução de uma ligação de H em particular (de acordo com o par indicado na legenda). Para fins de clareza, cada gráfico apresenta um subconjunto de dados, indicando como cada CRD interage com o LacNAc, com o peptídeo de ligações, ou com outro CRD. Análise feita com o programa <i>g_hbond</i> e o <i>script python readHBmap</i> , disponível em http://www.gromacs.org/Downloads/User_contributions/Other_software .	66
Figura 44 – Representação em sticks das interações do ligante LacNAc com os resíduos no CRD1 e CRD2 após a etapa de dinâmica molecular.	68
Figura 45 – Superfície de potencial eletrostático para os CRDs no início e no final da dinâmica molecular.	68
Figura 46 – Perfil de espalhamento das amostras e gráfico de Kratky do espalhamento, com subtração da curva de espalhamento para o tampão. Em vermelho, está a curva de espalhamento para a hGal4, seguida do perfil de espalhamento para a hGal4 com lactose, em verde, e com a sacarose em azul.	69
Figura 47 – Função de distribuição de distâncias.	70
Figura 48 – Gráfico do perfil de fluorescência relativa das proteínas em função do aumento da temperatura.	71
Figura 49 – Gráfico da variação da temperatura de desnovelamento ΔT_m em função dos compostos que estabilizaram todas as proteínas testadas, hGal4, hGal4-CRD1 e hGal4-CRD2.	73

Figura 50 – Gráfico da variação da temperatura de desnvolvamento (ΔT_m) em função da concentração de ligante. S corresponde aos valores de ΔT_m obtidos para a sacarose e L corresponde aos valores de ΔT_m obtidos para lactose.....	74
Figura 51 – Gráfico da variação da temperatura de desnvolvamento em função da concentração do ligante.....	75
Figura 52 – Gráfico da variação da temperatura de desnvolvamento da hGal4 em função dos compostos VA1-VA10.	76
Figura 53 – Representação em <i>ribbon</i> da sobreposição da cadeia principal das estruturas da hGal4-CRD1 e hGal4-CRD2, RMS=1,21 Å.	77
Figura 54 – Representação da superfície de potencial eletrostático para os domínios hGal4-CRD1 e hGal4-CRD2 em diferentes posições.....	77
Figura 55 – Superfície de potencial eletrostático calculado para o modelo da galectina-4.	79
Figura 56 – A) Representação da topologia da galectina-4. Topologia gerada no servidor PDBsum. Representação em <i>cartoon</i> do enovelamento global e posicionamento dos CRDs no modelo da galectina-4: B) visão frontal e C) visão do topo.	80
Figura 57 – Representação em <i>cartoon</i> da estrutura do peptídeo de ligação da galectina-4 evidenciando em <i>sticks</i> as prolínas presentes nessa região.	80
Figura 58 – Alinhamento das isoformas 1, 2, 3, 4 e 5 da galectina-12 humana. A sequência em azul claro, dos resíduos 1 ao 183 representam o domínio N-terminal; a sequência marcada em azul escuro, dos resíduos 211 ao 337 representam o domínio C-terminal e dos resíduos 184 ao 210 formam o peptídeo de ligação dos domínios, em preto.	85
Figura 59 - Representação esquemática do alinhamento das isoformas 1, 2, 3, 4 e 5 da galectina 12 humana. A região marcada em azul claro representa os resíduos que representam o CRD1 N-terminal; a região marcada em azul escuro representa o CRD-2 C-terminal e na região marcada em rosa está o peptídeo de ligação dos domínios.	86
Figura 60 - Alinhamento das sequências dos membros da família <i>tandem-repeat</i> (galectinas 4, 6, 8 e 9 humanas) com uma isoforma longa e uma isoforma curta da galectina 12humana. Os resíduos marcados em ★ mostram os resíduos envolvidos na coordenação com o carboidrato ligante.	86
Figura 61 – Anatomia de uma gota de lipídio. À esquerda micrografia eletrônica de uma gota de lipídio; à direita, representação esquemática das características estruturais da gota de lipídio. Figura retirada do artigo <i>Lipid Droplets Finally get a little R-E-S-P-E-C-T</i> (Farese e Walther, 2009). Micrografia eletrônica cortesia de S. Stone e J. Wong. Imagem reimpressa de Walther and Farese(Walther e Farese, 2009).	88
Figura 62 – Representação esquemática da construção resultante da clonagem no vetor pET-28a-sumo.	93
Figura 63 – Representação esquemática do alinhamento entre as 5 isoformas da galectina-12. Em azul claro a região que representa o CRD1; em rosa, a região que representa o peptídeo de ligação e em azul escuro, a região que representa o CRD2. A região hachurada representa o lugar na sequência em que cada primer foi baseado.	94
Figura 64 – Representação esquemática do bacmídeo recombinante na região de transposição do gene de interesse. Adaptado do manual Bac-to-Bac® <i>Baculovirus Expression System</i> (Invitrogen™).	99
Figura 65 – Representação esquemática das isoformas 2 e 3. As regiões referentes ao CRD1 (azul claro), peptídeo de ligação (rosa) e CRD2 (azul escuro) podem ser identificadas na figura. Em amarelo está representado o	

oligonucleotídeos <i>Forward</i> -peptídeo e em laranja, o oligo <i>Reverse</i> -peptídeo, desenhados para a clonagem da isoforma 3 a partir da isoforma 2 por <i>overlap extension</i>	101
Figura 66 – Representação esquemática da disposição das células durante o teste de expressão.	106
Figura 67 – Análise da expressão da proteína por SDS-PAGE 14% com o aumento da concentração de IPTG. MM - Marcador de proteínas de baixo peso molecular (GE <i>Healthcare</i>). 1) controle negativo, 2) 10 μ M, 3) 25 μ M, 4) 50 μ M, 5) 75 μ M, 6) 100 μ M, 7) 150 μ M e 8) 200 μ M nas frações insolúveis (gel A) e solúveis (gel B).	117
Figura 68 – A) Perfil cromatográfico da eluição da proteína CRD1c-inteína, após o processo de clivagem com o agente redutor DTT e da cauda de afinidade inteína após processo de limpeza com 1% de SDS. B) SDS-PAGE 15%. MM) Marcador de peso molecular. Amostras referentes a diferentes etapas de purificação: 1) volume de 2 ml, 2) volume de 5 ml, 3) volume de 10 ml, 4) volume de 80 ml, 5) volume de 82 ml e 6) volume de 84 ml.	117
Figura 69 – Perfil cromatográfico de eluição do domínio CRD1c após clivagem com agente redutor. O gráfico mostra a sobreposição das curvas de eluição nas diferentes condições testadas.	118
Figura 70 – Alinhamento local entre as sequências primárias da proteína inteína Sce VMA1 e os resíduos identificados no processo de sequenciamento N-terminal da amostra purificada.	120
Figura 71 – Perfil eletroforético de purificação em coluna de níquel do domínio CRD1c em fusão com a proteína sumo. MM) Marcador molecular. A) Eluição com 50 mM de imidazol 1-9) frações 1, 2, 3, 4, 5, 6, 8 e 10; B) Eluição com 100 mM de imidazol 1-9) frações 1, 2, 3, 4, 5, 6, 8 e 10; C 1-4) frações 1, 2, 3 e 4 da eluição com 250 mM de imidazol; 5-8) frações 1, 2, 3 e 4 da eluição com 500 mM de imidazol, 9) amostra da resina de níquel após a limpeza da coluna; D) Purificação pós-clivagem 1-5) frações da eluição da proteína CRD1c; 6-8) frações referentes a eluição com 500 mM de imidazol 9) amostra concentrada das frações de 1 a 5.	122
Figura 72 – Perfil eletroforético de migração das proteínas antes e depois da clivagem. 1) fração correspondente ao pico da eluição em 100 mM de imidazol de uma purificação típica do domínio CRD1c-sumo. 2) Fração 1 após clivagem com a ULP1, protease específica da SUMO.	123
Figura 73 – Perfil eletroforético do extrato proteico solúvel após lise celular e centrifugação. MM- Marcador molecular. As amostras aplicadas nos poços estão descritas na figura.	125
Figura 74 – Perfil eletroforético da purificação em coluna de níquel para as construções Iso2-3-pETsumo e CRD2-5-pETsumo. MM) Marcador molecular. Gel A – 1) precipitado, 2) sobrenadante, 3) efluente, 4-8) tampão 0 de imidazol, 10-14) tampão 25 mM de imidazol; Gel B 1-7) tampão 50 mM de imidazol, 8-14) tampão 100 mM de imidazol; Gel C – 1-7) tampão 250 mM imidazol e 8-14) tampão 500 mM imidazol.	126
Figura 75 – Perfil de purificação em coluna de níquel sob condições desnaturantes. A) CRD1c-3-pET29 B) CRD1c-1-pETsumo. 1) precipitado pós solubilização 2) corpo de inclusão solubilizado 3) fração 1 tampão 0 imidazol 4-6) frações 1, 3 e 5 tampão 25 mM imidazol 7-12) frações 1, 2, 3, 5, 7 e 10 tampão 200 mM imidazol 13) fração 1 500 mM imidazol 14) resina de níquel.	126
Figura 76 – Purificação das construções CRD1c-3-pET29 e CRD1c-1-pETsumo em coluna de afinidade. MM – marcador molecular. CRD1c-3-pET29 1-6) frações 1, 2, 3, 4, 5 e 10 da eluição; 7) resina de níquel; CRD1c-1-pETsumo 8-13) frações 1, 2, 3, 4, 5 e 10 da eluição; 14) resina de níquel.	127
Figura 77 – Imunodeteção por <i>western blot</i> das proteínas recombinantes CRD1c e CRD2 provenientes das construções CRD1c-3-pET29 e CRD2-pET29, respectivamente. De 1 a 8 diluições de 1:10000, 1:20000, 1:40000, 1:80000, 1:160000, 1:320000, 1:640000 e 1:1280000.	129
Figura 78 – Análise eletroforética (SDS-PAGE) e a reação de <i>western blot</i> dos extratos proteicos de bactérias submetidos a imunodeteção com o anticorpo policlonal anti-peptídeo sintético da hGal12. MM) marcador molecular; 1)	

induzido e 2) não induzido CRD2-pET29; 3) induzido e 4) não induzido Iso2-2-pET29; 5-6) BL21(DE3) sem plasmídeo; 7) não induzido e 8) induzido Iso2-3-pETsumo; 9) não induzido e 10) induzido Iso4-1-pETsumo; 11) não induzido e 12) induzido CRD2-5-pETsumo.....	130
Figura 79 – Análise da imunorreatividade do anticorpo comercial anti-peptídeo da hGal12 para outras galectinas humanas. As diferentes preparações proteicas foram submetidas a eletroforese em gel de poliacrilamida (SDS-PAGE) e eletrotransferidas para membranas de nitrocelulose. As amostras utilizadas foram: MM) mistura de marcadores de massa molecular composta por proteínas biotiniladas; 1) hGal1, 2) hGal3, 3) hGal4, 4) hGal12, 5) hGal1, 6) hGal3, 7) hGal4, 8) hGal12, 9) hGal1, 10) hGal3, 11) hGal4 e 12) hGal12. Nos poços de 1 a 4 foram utilizados 100 ng, nos poços de 5 a 12 foram utilizadas 500 ng e 250 ng de cada proteína, respectivamente. .	130
Figura 80 – Determinação das condições experimentais do uso do laser para a análise da microscopia confocal A) Sobreposição dos canais na fluorescência no azul para o marcador DAPI e no vermelho para o anticorpo secundário AlexaFluor® 546 anti-IgG de cabra. B) Sobreposição dos canais na fluorescência no azul para o marcador DAPI e no verde para o marcador BODIPY.....	131
Figura 81 – Composição das imagens de microscopia confocal em células HL-60. A) Sobreposição das imagens com DAPI e BODIPY. B) Sobreposição das imagens com DAPI e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 no vermelho.....	132
Figura 82 – Imagens tridimensionais de microscopia confocal em células HL-60. A) Sobreposição das imagens com DAPI (marcação em azul) e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 (marcação em vermelho). B) Sobreposição das imagens com BODIPY (marcação verde) e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 (marcação em vermelho).....	132
Figura 83 – Amplificação da imagem B da figura 82 da fusão das imagens com BODIPY e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 no vermelho.	133
Figura 84 – Representação esquemática do enovelamento dos CRDs das isoformas 1, 2, 3, 4 e 5 descritas para a galectina-12 humana. CRD1a – CRD1 da isoforma 1; CRD1b – CRD1 das isoformas 2 e 3; CRD1c – CRD1 da isoforma 4; CRD2 – CRD2 para as isoformas 1, 2, 3, 4 e 5.....	134
Figura 85 – Representação da superfície de potencial eletrostático para os diferentes CRDs: CRD1a, CRD1b, CRD1c e CRD2 nas posições $\gamma=0^\circ$, 90° , 180° e 270°	135

LISTA DE TABELAS

Tabela 1 – Classificação geral da família de lectinas †.....	4
Tabela 2 – Sequência dos oligonucleotídeos utilizados para amplificar o gene que codifica a proteína humana galectina-4.....	25
Tabela 3 – Sequência dos oligonucleotídeos previamente utilizados para amplificação dos domínios hGal4-CRD1 e hGal4-CRD2, que foram clonados no vetor pET-28a(+) (Novagen) para expressão heteróloga em <i>Escherichia coli</i> . Em parênteses estão as sequências referentes aos sítios de restrição.	26
Tabela 4 – Estatística da coleta e processamento de dados. Em parênteses estão os valores para a camada mais externa de resolução.....	46
Tabela 5 – Estatística da coleta e processamento de dados. Em parênteses estão os valores para a camada mais externa de resolução.....	47
Tabela 6 – Estatística da coleta e processamento de dados. Em parênteses estão os valores para a camada mais externa de resolução.....	49
Tabela 7 – Condições iniciais de cristalização do domínio hGal4-CRD2.	50
Tabela 8 – Varredura das condições iniciais de cristalização obtidas para o domínio hGal4-CRD2.....	51
Tabela 9 – Estatística de coleta e processamento dos dados. Em parênteses estão os valores para a camada mais externa de resolução.....	54
Tabela 10 – Estatística do refinamento da estrutura da hGal4-CRD1.....	56
Tabela 11 – Estatística do refinamento da estrutura da hGal4-CRD2.....	59
Tabela 12 – Relação dos pares de interação durante a simulação. As ligações de hidrogênio com tempo de permanência maior que 30% foram isoladas. As interações foram classificadas para as interações entre CRD1-CRD2, CRD1-peptídeo, CRD2-peptídeo e peptídeo-peptídeo.	66
Tabela 14 – Tabela dos valores de ΔT_m obtidos para a hGal4, hGal4-CRD1 e hGal4-CRD2 no experimento de termofluor realizado para o kit de solubilidade e estabilidade da Hampton Research. ■ valores < 1, ■ valores 1 > 2, ■ valores 2 > 5 e ■ valores > 5.....	72
Tabela 15 – Biblioteca de glicoconjugados testados e a nomenclatura utilizada	75
Tabela 16 – Oligonucleotídeos utilizados no processo de clonagem dos fragmentos de DNA codificantes das isoformas e CRDs da galectina-12 humana. Em parênteses estão as sequências referentes aos sítios de restrição.	95
Tabela 17 – Relação de clones obtidos durante o processo de clonagem das isoformas e domínios da hGal12 e seus respectivos vetores e sistema de expressão.	97
Tabela 19 – Quantidade percentual de códons raros para <i>E. coli</i> presentes nas isoformas da hGal12.....	116
Tabela 20 – Valores obtidos para os picos majoritários da distribuição de tamanho das partículas em solução.	119
Tabela 21 – Valores obtidos para os picos majoritários da distribuição de tamanho das partículas em solução	127

SUMÁRIO

Capítulo 1 Introdução	i
Introdução.....	1
1.1 Glicanos e glicoconjugados.....	1
1.2 Lectinas	2
1.3 Galectinas.....	5
1.3.1 Galectinas do grupo <i>tandem-repeat</i>	10
Capítulo 2 Galectina-4.....	17
Introdução.....	18
Objetivos	22
Materiais e métodos.....	23
2.1 Produção proteica.....	23
2.1.1 Clonagem	23
2.1.2 Expressão e purificação	26
2.2 Estudo estrutural	29
2.2.1 Cristalização	29
2.2.2 Coleta de dados e processamento.....	33
2.2.3 Determinação da estrutura e refinamento.....	34
2.2.4 Modelagem molecular	34
2.2.5 Dinâmica molecular.....	35
2.3 Estudos de estabilidade térmica – Termofluor™	38
2.3.1 <i>Screening</i> de solubilidade e estabilidade	38
2.3.2 Glicoconjugados.....	38
Resultados e Discussões.....	40

2.4 Produção proteica.....	40
2.4.1 Clonagem, expressão e purificação da hGal4	40
2.4.2 Clonagem, expressão e purificação da hGal-CRD1.....	42
2.4.3 Clonagem, expressão e purificação da hGal4-CRD2.....	43
2.5 Estudo estrutural	45
2.5.1 Cristalização, coleta de dados e processamento	45
2.5.2 Determinação da estrutura	55
2.5.3 Modelagem Molecular.....	61
2.5.4 Dinâmica Molecular.....	62
2.5.5 Espalhamento a baixo ângulo	69
2.6 Estudos de estabilidade térmica	70
2.6.1 <i>Screening</i> de solubilidade e estabilidade	70
2.6.2 Glicanos.....	73
2.7 Comparação estrutural.....	76
Conclusões	81
Capítulo 3.....	84
Galectina-12.....	84
Introdução.....	85
Objetivos.....	89
Materiais e métodos.....	90
3.1 Obtenção do DNA molde	90
3.1.1 Construção do cDNA	90
3.1.2 Otimização de códons.....	91
3.2 Clonagem	91
3.2.1 Vetores de clonagem e expressão	92
3.2.2 Clonagem das isoformas e domínios da hGal12.....	94
3.3 Expressão heteróloga.....	101

3.3.1 Expressão das construções em bactérias	102
3.3.2 Expressão das construções em células de inseto.....	105
3.4 Purificação	106
3.4.1 Purificação por cromatografia de afinidade em coluna de quitina	106
3.4.2 Purificação por cromatografia de afinidade em coluna de níquel.....	107
3.5 Ensaio imunológico	110
3.5.1 Produção de anticorpos policlonais	110
3.5.2 <i>Western blot</i>	111
3.5.3 Microscopia confocal	112
3.6 Modelagem molecular.....	113
Resultados e discussões.....	115
3.7 Produção das macromoléculas	115
3.7.1 Produção em bactérias das macromoléculas obtidas a partir do cDNA	115
3.7.2 Produção das macromoléculas em bactérias a partir do DNA otimizado	116
3.7.3 Produção das macromoléculas em células de inseto a partir do cDNA	128
3.8 Ensaio imunológico	128
3.9 Modelagem molecular.....	133
Conclusões	136
Capítulo 4.....	141
Referências Bibliográficas	141

CAPÍTULO 1

INTRODUÇÃO

INTRODUÇÃO

1.1 Glicanos e glicoconjugados

Os carboidratos, também denominados sacarídeos, constituem a classe de biomoléculas mais abundante na Terra. O termo, derivado de hidratos de carbono, descreve substâncias que ocorrem naturalmente com a fórmula geral $C_x(H_2O)_n$. A unidade básica, os monossacarídeos, podem ser divididos em dois grupos: poliidroxialdeídos e poliidroxicetonas, caracterizados pela presença de um grupamento aldeído e um grupamento cetona, respectivamente.

A adição de um grupo hidroxil de um monossacarídeo ao carbono anomérico de outro, chamada de ligação glicosídica, resulta em um dissacarídeo. Esses podem se associar a novos monossacarídeos pela ligação glicosídica formando complexos com até 20 monossacarídeos (oligossacarídeos) ou ainda maiores (polissacarídeos). Em solução, geralmente aldoses com 4 carbonos e todos os monossacarídeos com 5 ou mais átomos de carbono no esqueleto ocorrem predominantemente como estruturas cíclicas, que protege o carboidrato da oxidação (Lehninger, Nelson e Cox, 2011).

Além de serem armazenadores de energia (amido, glicogênio, dextranas) e elementos estruturais (celulose, quitina, peptidoglicanos, glicosaminoglicanos), os polissacarídeos e oligossacarídeos são transportadores de informação, possuindo um papel essencial em muitos processos fisiológicos, incluindo a biologia do desenvolvimento, defesa imune, diferenciação, proliferação e apoptose celular, interação célula-matriz e adesão entre células (Varki *et al.*, 2009), bem como na patogênese de agentes infecciosos como no processo de adesão de vírus, bactérias e parasitas às células hospedeiras e em uma diversidade de doenças que vão de desordens congênitas até diabetes e câncer (Campbell e Yarema, 2005; Arnaud, Audfray e Imberty, 2013).

Nas células, de um modo geral, a maioria dos oligossacarídeos constituídos por três ou mais unidades não ocorre como moléculas livres, mas ligada covalentemente a lipídeos ou proteínas formando os glicoconjugados, através de um processo pós-traducional denominado de glicosilação (Varki *et al.*, 2009; Lehninger, Nelson e Cox, 2011). Na molécula bioativa, o glicoconjugado, o componente carboidrato é denominado como glicona (*glycone*) e o componente não carboidrato como aglicona (*aglycone*) (Varki *et al.*, 2009). Os diferentes tipos de glicoconjugados compreendem proteoglicanos, glicoproteínas e glicolipídeos. Os proteoglicanos são glicoconjugados nos quais um ou mais glicosaminoglicanos sulfatados estão covalentemente ligados a uma proteína central, sendo os principais componentes de todas as matrizes extracelulares. Agem como organizadores de tecidos e influenciam várias atividades celulares. As glicoproteínas são glicoconjugados nos quais

os glicanos, covalentemente ligados a resíduos de Asp ou Ser/Thr, são ramificados, menores e estruturalmente diversos. Os glicolipídeos são lipídeos covalentemente ligados a glicanos, em sua maioria, glicosfingolipídeos (Varki *et al.*, 2009; Lehninger, Nelson e Cox, 2011).

De fato, a glicosilação é a modificação pós-traducional mais comum em proteínas e lipídeos (Opdenakker *et al.*, 1993) e sua natureza versátil gera complexidade estrutural e diversidade biológica, e consequente aumento na escala de funcionalidade das moléculas biológicas (Gabiús *et al.*, 2011). A importância da glicosilação é evidente não somente por trazer uma variabilidade química e estrutural, mas também pela descoberta de que alterações e/ou defeitos nesse processo pós-traducional estão intimamente relacionados à manifestação de vários distúrbios genéticos (Freeze *et al.*, 2014).

De um modo geral, os carboidratos são considerados sinais bioativos únicos, com habilidade insuperável de codificar a informação (aproximadamente de $2 \cdot 10^7$ e $3 \cdot 10^{11}$ vezes maior do que peptídeos e oligonucleotídeos, respectivamente) (Laine, 1996; Gabiús *et al.*, 2011). Essa capacidade ocorre em virtude de suas propriedades químicas intrínsecas que permitem uma versatilidade de configurações espaciais, como a isomeria ótica, bem como a introdução de ramificações e de substituições sítio específicas tais como acetilação, fosforilação ou sulfatação a que os oligossacarídeos podem ser submetidos, o que constitui a base dessa diversidade estrutural (Gabiús *et al.*, 2011). Adicionalmente, o conjunto de glicoconjugados que compõe o glicoma de uma célula ou organismo é dinâmico, respondendo de forma diferente a sinais intrínsecos ou extrínsecos (Varki *et al.*, 2009).

Os glicanos possuem uma capacidade de codificação extremamente abrangente, tornando-os biomoléculas ideais para a geração de mensagens moleculares – o glicocódigo. Decodificar essa informação é uma função atribuída às proteínas que se ligam a glicanos – GBP (*Glycan binding proteins*). A decodificação do glicocódigo é de fundamental importância para a compreensão dos mecanismos fundamentais da vida, com grande aplicabilidade dentro das ciências da saúde de um modo geral (Rabinovich e Toscano, 2009; Rabinovich e Croci, 2012; Solís *et al.*, 2014).

1.2 Lectinas

A natureza se aproveita da variedade de glicanos expressos por diferentes organismos através da produção de proteínas que são capazes de reconhecer seletivamente os diferentes carboidratos, mediando processos fisiológicos e patológicos importantes. De fato, não existem organismos onde as GBPs não tenham sido encontradas. Com exceção dos anticorpos específicos

de glicanos, pode-se classificar as GBPs em dois grandes grupos de proteínas: as proteínas que ligam a glicosaminoglicanos e as lectinas.

As lectinas (Boyd e Shapleigh, 1954) são definidas como proteínas de origem não imune que ligam a carboidratos sem alterá-los enzimaticamente. O termo lectina, que do latim *lego* significa escolha, foi o nome proposto por Boyd e Shapley em 1954 às aglutininas de plantas capazes de distinguir eritrócitos de diferentes tipos sanguíneos. As evidências da presença na natureza de proteínas capazes de aglutinar eritrócitos datam do final do século 19. Estudos pioneiros de Ehrlich, que empregava as primeiras lectinas descritas, a ricina por Stillmark e a abrina por Hellin em 1888, como modelos de antígenos em estudos imunológicos, demonstraram a especificidade da resposta imune, o fenômeno da memória imunológica e a transferência da imunidade humoral (Sharon e Lis, 2004).

Na sequência, duas novas descobertas foram feitas no início dos anos 60, a atividade de estimular linfócitos a sofrer mitose por uma lectina encontrada no feijão vermelho (PHA - *Phytohemagglutinin*) (Nowell, 1960) e a habilidade de aglutinar preferencialmente células malignas por uma lectina do gérmen do trigo (WGA - *Wheat germen agglutinin*) (Aub, Sanford e Cote, 1965).

Desde então, o ritmo do isolamento de diversos tipos de lectinas aumentou drasticamente, o que possibilitou a identificação de diferentes grupos da proteína presentes em plantas e animais. O interesse nessa classe de proteínas foi fortemente estimulado por se mostrarem ferramentas importantíssimas na detecção, isolamento e caracterização de glicoconjugados, como as glicoproteínas, para a histoquímica de células e tecidos e para o exame de mudanças que ocorrem na superfície de células durante processos fisiológicos e patológicos, em particular, na diferenciação celular para o estudo de processos cancerígenos (Sharon e Lis, 2004).

A partir dos anos 70 houve uma intensificação dos estudos das propriedades moleculares individuais das lectinas, bem como a sua identificação na maioria dos organismos desde vírus, bactérias e fungos até plantas e animais (Wu, A. M. *et al.*, 2009; Vasta *et al.*, 2011; Arnaud, Audfray e Imberty, 2013).

A análise da sequência de amino ácidos das lectinas conhecidas mostrou a presença de um segmento polipeptídico responsável pela interação com o ligante, designado domínio de reconhecimento ao carboidrato – CRD (Drickamer, 1988). Baseado na sequência conservada de aminoácidos e no arranjo tridimensional dos diferentes tipos de CRDs, foi possível classificar a maioria das lectinas em diferentes famílias (Varki *et al.*, 2009; Vasta *et al.*, 2011) (tabela 1). Contudo, mesmo com a crescente descoberta de diferentes famílias de lectinas, um número limitado de enovelamentos proteicos tem sido descrito (Gabius *et al.*, 2011; Arnaud, Audfray e Imberty, 2013) (figura 1).

Tabela 1 – Classificação geral da família de lectinas †

Tipo-C – lectinas dependentes de cálcio
Galectinas – antigas lectinas Tipo-S
Tipo-P – receptores de manose-6-fosfato
Tipo-I – membros da superfamília das imunoglobulinas
Tipo-F – receptores de oligossacarídeos que terminam em fucose
Calnexinas
Tipo-M – lectinas relacionadas a α -manosidase
Tipo-L – lectinas de sementes de leguminosas
Tipo-R – ricina, outras lectinas de plantas, receptores de GalNAc-SO ₄
F- <i>box</i>
Ficolinas – lectinas contendo domínio de fibrinogênio/colágeno
Quitinase- <i>like</i>
Pentraxinas
Intelectinas

† Tabela adaptada de (Vasta *et al.*, 2011)

Figura 1 - Representação estrutural dos diferentes tipos de enovelamentos característicos em lectinas. A) α/β misto em lectinas do tipo-C em invertebrados (PDB: 2VUZ); B) β -prisma I em lectinas que ligam a manose (PDB: 1C3K); C) β -prisma II em aglutinina de plantas (PDB: 1KJ1); D) β -propeller de lectina de fungo (PDB: 2C4D); E) β -trefoil em mamífero (PDB: 1FWV); F) tipo-C em bactérias (PDB: 1PWB); G) fibrinogênio-*like* em caranguejo-ferradura (PDB: 1JC9); H) imunoglobulina-*like* em lectinas de mamífero da família siglec (PDB: 1QFO); I) *jelly-roll* em lectinas de legumes (PDB: 3ZYR); J) tipo-P em receptores de manose-6-fosfato de mamífero (PDB: 1M6P).

Independente das características estruturais das lectinas, os sítios de interação ao carboidrato estão geralmente na superfície da proteína, em sulcos rasos ou na interface entre oligômeros, favorecendo o *crosslinking* entre unidades funcionais (Bourne *et al.*, 1994; Arnaud, Audfray e Imberty, 2013) (figura 2).

Figura 2 - Representação da superfície de lectinas evidenciando os bolsos de interação com os ligantes. A) lectina de leguminosas. Em amarelo, um complexo de glicanos interage na superfície da proteína (PDB: 3ZYR). B) galectina-1 de mamífero. Em amarelo, um complexo de glicanos interage entre dímeros promovendo o *crosslinking* entre eles (PDB: 1SLA).

Por sua capacidade de reconhecer uma miríade de glicoconjugados, as lectinas possuem diversas funções biológicas. Atuam no reconhecimento de glicanos da superfície celular do hospedeiro, como receptores por patógenos, nos processos de infecção (Imberty e Varrot, 2008; Varki *et al.*, 2009; Arnaud, Audfray e Imberty, 2013); na defesa e proteção contra organismos fitopatogênicos, nematoides ou insetos, em plantas (Peumans e Van Damme, 1995; Vandenborre, Smaghe e Van Damme, 2011); como parte do sistema de imunidade inata, em muitos invertebrados (Vasta, Ahmed e Odom, 2004); no transporte e liberação de glicoproteínas, na defesa imune, no desenvolvimento entre outros processos biológicos, em animais (Varki *et al.*, 2009; Arnaud, Audfray e Imberty, 2013), consequentemente, desempenhando um papel essencial na decodificação do glicocódigo (Arnaud, Audfray e Imberty, 2013).

1.3 Galectinas

As galectinas são proteínas membros de uma família evolutivamente conservada de lectinas que compartilham características como a afinidade por β -galactosídeos e a presença de um domínio de reconhecimento ao carboidrato (CRDs) (Barondes *et al.*, 1994), além da habilidade de regular diferencialmente várias respostas biológicas (Rabinovich *et al.*, 2007). Membros das galectinas são amplamente distribuídos na natureza, presente desde organismos primitivos como esponjas e nematoides até fungos e animais superiores (Kasai e Hirabayashi, 1996; Houzelstein *et al.*, 2004;

Yang, Rabinovich e Liu, 2008; Vasta *et al.*, 2012). Sua presença em espécies anciãs sugere o desempenho de um papel fundamental na biologia celular, enquanto que a presença de múltiplas galectinas em uma única espécie indica que elas divergiram para atuar especificamente em uma variedade de funções possivelmente não correlatas (Cooper, 2002).

Possuem características de proteínas citosólicas e embora não contenham um peptídeo sinal clássico, alguns membros são direcionados para o núcleo ou outros sítios subcelulares, podendo até mesmo serem secretados por vias não clássicas (Hughes, 1999; Leffler *et al.*, 2002). Estudos recentes mostram a habilidade das galectinas em se ligar seletivamente à superfície de glicoreceptores (glicoproteínas ou glicolípídeos) (Rabinovich e Toscano, 2009).

Uma vez no ambiente extracelular, as galectinas podem se ligar a múltiplos parceiros glicosilados e traduzir a informação codificada pelos glicanos em processos como a ativação celular imune, diferenciação e homeostase. Além disso, as galectinas podem formar redes ordenadas de interação com glicanos da superfície das células (Rabinovich *et al.*, 2007; Yang, Rabinovich e Liu, 2008; Rabinovich e Toscano, 2009; Varki *et al.*, 2009).

Esse quadro mostra que as galectinas secretadas não possuem um receptor específico, mas possuem a capacidade de mediar a comunicação entre as células imunes reconhecendo um conjunto preferencial de glicoconjugados da superfície celular (Yang, Rabinovich e Liu, 2008). Além disso, resultados apontam que, interessantemente, as galectinas podem ter função dentro das células de modo independente de sua atividade de ligação ao carboidrato (Liu, Patterson e Wang, 2002; Di Lella *et al.*, 2011). Por exemplo, todas as galectinas reconhecem o dissacarídeo N-acetilactosamina (LacNAc) que é encontrado em N- e O-glicanos da superfície celular (Hirabayashi *et al.*, 2002). Entretanto, há entre as galectinas preferências na interação com os glicanos (Hirabayashi *et al.*, 2002; Stowell, Arthur, Mehta, *et al.*, 2008) que podem ser a base de diferenças fundamentais nas suas atividades biológicas. Tais variações são geralmente associadas a modificações do sacarídeo terminal como sialilação ou fucosilação ou mesmo devido a vínculos conformacionais da superfície da célula (Stowell, Arthur, Mehta, *et al.*, 2008; Rabinovich e Toscano, 2009). Além disso, a interação seletiva com diferentes parceiros glicoproteicos pode resultar da orientação espacial dos CRDs e da topologia que o número de N-glicanos delineia em uma dada glicoproteína (Brewer, Miceli e Baum, 2002; Rabinovich e Toscano, 2009).

As galectinas estão envolvidas em diversas atividades biológicas e eventos relacionados a imunidade inata e adaptativa. Por mecanismos extracelulares ou intracelulares, as galectinas podem influenciar a capacidade das células da imunidade inata (neutrófilos, células dendríticas, monócitos/macrófagos, eosinófilos e mastócitos) de responder a gradientes quimiotáticos, migrar através da superfície celular do endotélio, sintetizar e liberar mediadores pro e anti-inflamatórios e

reconhecer, fagocitar e matar microrganismos e células danificadas (Rabinovich e Ilarregui, 2009; Vasta, 2009; Garner *et al.*, 2010; Bi *et al.*, 2011; Cerliani *et al.*, 2011; Davicino *et al.*, 2011; Di Lella *et al.*, 2011).

Além disso, as galectinas podem moldar a imunidade adaptativa por influenciar a sinalização e ativação de células T, alterar o balanço de citocinas, regular o compartimento de células B e agir como mediadoras da função supressora das células T regulatórias (Rabinovich e Toscano, 2009). Devido a influência seletiva de galectinas individuais em processos biológicos diferentes, o resultado dos seus efeitos sincronizados contribuem para orquestrar a ativação, polarização e resolução das respostas da imunidade adaptativa (Di Lella *et al.*, 2011).

Até o momento, 16 membros foram identificados em vertebrados. Com base no número e na organização dos CRDs, as proteínas da família das galectinas são classificadas em três subgrupos: proto, químera e *tandem-repeat* (figura 3) (Kasai e Hirabayashi, 1996).

Figura 3 - Esquema ilustrativo dos três subgrupos das galectinas: proto, químera e *tandem-repeat*. O tipo proto é composto somente de um domínio lectínico, o CRD. O tipo químera é composto de duas partes: o domínio C-terminal correspondendo ao CRD e um domínio N-terminal não lectínico. O tipo *tandem-repeat* é composto de dois CRDs distintos em uma cadeia polipeptídica única.

As proteínas do subgrupo proto compreendem as galectinas 1, 2, 5, 7, 10, 11, 13, 14, 15 e 16, que são caracterizadas pela presença de um único domínio CRD de aproximadamente 130 aminoácidos. Algumas dessas galectinas podem formar homodímeros pela associação de seus monômeros, como é o caso da galectina-1 (Gitt *et al.*, 1992) ou permanecem monoméricas, tal como a galectina-5 (Gitt *et al.*, 1995). O subgrupo químera, com um único membro identificado em vertebrados, a galectina-3, é composto por um domínio C-terminal lectínico unido a um domínio não lectínico N-terminal. Esse domínio apresenta várias repetições de uma sequência polipeptídica rica em prolina, glicina e tirosina, através da qual a proteína é capaz de formar oligômeros (Massa *et al.*, 1993).

O subgrupo *tandem-repeat*, composto pelas galectinas 4, 6, 8, 9 e 12, apresenta dois domínios de reconhecimento ao carboidrato não idênticos unidos por um peptídeo de ligação de comprimento variável (Hirabayashi e Kasai, 1993; Cooper, 2002). Alguns estudos específicos para

galectinas do grupo *tandem-repeat* permitiram concluir que esses dois domínios estão posicionados de forma a possibilitar a ligação simultânea de carboidratos distintos em seus sítios de ligação (Hirabayashi *et al.*, 2002).

A determinação da estrutura tridimensional das galectinas tem contribuído para o entendimento das bases da interação entre proteína e glicanos. Estudos de comparação entre as estruturas dos CRDs tem sido realizadas (Guardia *et al.*, 2011) concomitante com o depósito no banco de dados de novas estruturas das galectinas resolvidas em diferentes condições, na ausência e presença de ligantes. Novas estruturas de galectinas tornam possível a correlação entre as mais diversas funções biológicas em que tais proteínas participam e as bases estruturais que governam as características físicas e químicas que regem o ambiente de interação com o carboidrato ligante.

A análise estrutural de membros desta família mostra que o CRD das galectinas se enovela em um arranjo β -sanduíche disposto em uma topologia *jelly-roll* típica de galectinas (López-Lucendo *et al.*, 2004) (figura 4).

Figura 4 - Topologia das galectinas. A) Representação em *cartoon* do monômero da estrutura cristalográfica da galectina-1 humana em complexo com lactose. PDB: 1GZW. B) Representação da topologia da galectina-1 humana pelo gráfico gerado pelo banco de dados PDBsum (<http://www.ebi.ac.uk/pdbsum/>).

Representativamente, a estrutura da galectina-1 humana consiste de 135 amino ácidos arranjados em duas folhas β antiparalelas formadas por cinco (F1-F5) e seis (S1 a S6a/b) fitas β . Os resíduos diretamente envolvidos na ligação com o carboidrato são parte de um bolsão formado por quatro fitas β adjacentes (S4, S5, S6a e S6b) (figura 4). A afinidade e a especificidade observada para as galectinas frente aos diversos carboidratos é consequência da distribuição e das características dos aminoácidos que compõe o bolsão de ligação no CRD. Substituições sutis na sequência de cada CRD levam a diferenças na afinidade por diferentes ligantes.

Em cada subunidade, o sítio de ligação ao carboidrato está localizado em um sulco compostos das fitas S4 - S6, localizado no lado oposto da superfície de dimerização (figura 5). A dimerização envolve auto associação dos monômeros pelos lados opostos de seus CRDs, podendo

formar *crosslinking* de ligantes agrupados (Bourne *et al.*, 1994; Cooper, 2002; López-Lucendo *et al.*, 2004).

Figura 5 - Representação do homodímero da galectina-1 humana. A lactose está localizada no bolso de ligação do CRD compreendido pelas fitas S4, S5 e S6. PDB: 1GZW.

Os resíduos His44, Asn46, Arg48, Asn61, Trp68, Glu71 e Arg73, geralmente conservados entre as galectinas, se mostram cruciais no estabelecimento de interações chaves que determinam a especificidade de ligação ao carboidrato (López-Lucendo *et al.*, 2004; Di Lella *et al.*, 2011) (figura 6).

Figura 6 - Representação dos principais resíduos envolvidos na coordenação com o carboidrato. PDB 1GZW.

Na busca por potenciais determinantes da especificidade, vários estudos têm sido conduzidos para descrever em detalhes essas interações usando abordagens computacionais, bioquímicas, biofísicas (Ford *et al.*, 2003; López-Lucendo *et al.*, 2004; Di Lella *et al.*, 2007; Di Lella *et al.*, 2009; Meynier *et al.*, 2009; Miller *et al.*, 2009; Echeverria e Amzel, 2011). Embora a estrutura terciária das galectinas seja bastante conservada entre os membros da família, a estrutura quaternária pela associação de seus monômeros pode diferir. Nesse ponto de vista, a maioria das

galectinas são bivalentes ou multivalentes. Enquanto as galectinas proto podem dimerizar, as galectinas *tandem-repeat* são ao menos bivalentes e a galectina-3 pode formar oligômeros (Ahmad *et al.*, 2004).

A oligomerização multivalente é, portanto, essencial para alcançar estabilidade e funcionalidade biológica (Garner e Baum, 2008). Essa característica dota as galectinas com a capacidade de formar redes ordenadas de galectina-glicano na superfície celular (Rabinovich e Toscano, 2009). A combinação de galectinas e glicanos multivalentes permite a formação de distintos tipos de redes, que pode vir a ser o motivo da divergência em funções biológicas (Brewer, Miceli e Baum, 2002; Di Lella *et al.*, 2011).

1.3.1 Galectinas do grupo *tandem-repeat*

A ação de galectinas em múltiplas funções biológicas tem sido amplamente relatada (Elola *et al.*, 2007; Yang, Rabinovich e Liu, 2008; Liu e Rabinovich, 2010). Estudos mostram a expressão alterada de galectinas em células associadas ao câncer (Danguy, Camby e Kiss, 2002; Takenaka, Fukumori e Raz, 2002; Van Den Brûle, Califice e Castronovo, 2002; Lahm *et al.*, 2003; Ahmed *et al.*, 2009), em processos de metástase tumoral incluindo adesão (Ellerhorst *et al.*, 1999; Van Den Brûle *et al.*, 2003), agregação (Tinari *et al.*, 2001; Glinisky *et al.*, 2003), invasão (Le Marer e Hughes, 1996; Hittetlet *et al.*, 2003), angiogênese (Clausse *et al.*, 1999; Nangia-Makker *et al.*, 2000; Thijssen *et al.*, 2006; Thijssen, Hulsmans e Griffioen, 2008; Delgado *et al.*, 2011) e inflamação (Liu e Rabinovich, 2010). O conjunto de processos aos quais essas proteínas estão envolvidas indica que as galectinas podem modular a progressão do tumor e influenciar a evolução da doença (Liu e Rabinovich, 2005).

De fato, diferentes galectinas estão emergindo como mediadores cruciais em doenças inflamatórias agudas e crônicas, na autoimunidade e no câncer e estão sendo reconhecidas como alvos moleculares para a descoberta de medicamentos inovadores (Honjo *et al.*, 2001; Inufusa *et al.*, 2001; Camby *et al.*, 2002; Song, Billiar e Lee, 2002; John *et al.*, 2003; Califice *et al.*, 2004; Rubinstein *et al.*, 2004; Ueda, Kuwabara e Liu, 2004; Le *et al.*, 2005; Rabinovich, 2005; Norling, Perretti e Cooper, 2009; Wu, M. H. *et al.*, 2009; Espelt *et al.*, 2011; Salatino e Rabinovich, 2011). Entretanto, a maioria dessas funções foram descritas apenas para as galectinas-1 e -3, e os papéis das galectinas *tandem-repeat* permanecem pouco conhecidos (Troncoso *et al.*, 2012).

Galectinas do grupo *tandem-repeat* possuem dois CRDs unidos por um peptídeo de ligação. Funções como interações proteína-proteína, inserção na membrana e regulação do posicionamento do CRD têm sido propostas para o peptídeo de ligação desse grupo de proteínas (Troncoso *et al.*,

2012). Utilizando formas mutadas e/ou truncadas da galectina-8, tem sido demonstrado que os CRDs isolados têm atividade biológica prejudicada. Interessantemente, um efeito semelhante foi demonstrado, em algumas circunstâncias, quando houve o encurtamento do peptídeo de ligação. Esses resultados indicam que galectinas *tandem-repeat* requerem interações cooperativas entre os dois CRDs e uma região de ligação corretamente orientada para que o desempenho da função seja efetivo (Levy *et al.*, 2006).

Numerosos estudos têm demonstrado que galectinas *tandem-repeat* são consideravelmente mais potentes do que outros membros da família das galectinas na indução da mesma resposta celular: por exemplo, as galectinas-4, -8 e -9 são muito mais potentes do que as galectinas-1 e -3 em provocar a sinalização de neutrófilos e células T (Kashio *et al.*, 2003; Stillman *et al.*, 2006; Stowell, 2006; Bi *et al.*, 2008; Stowell, Arthur, Slanina, *et al.*, 2008; Stowell, Qian, *et al.*, 2008). Estudos demonstraram que a inserção de uma região ligante entre os dois CRDs de uma galectina *tandem-repeat* confere um aumento na potência de sinalização devido interações intermoleculares desses CRDs, o que promove uma multimerização de ordem mais alta e aumento na formação da rede na superfície celular (Earl, Bi e Baum, 2011).

1.3.1.1 Galectina-4

Inicialmente foram identificadas em extrato de intestino e na mucosa de rato como as proteínas RI-H, de 17 kDa (Leffler, Masiarz e Barondes, 1989) e L-36, de 36 kDa (Oda *et al.*, 1993), sendo mais tarde demonstrado que a proteína RI-H era um fragmento proteolítico da L-36. A proteína L-36 seria posteriormente classificada como a primeira proteína da família das galectinas com dois CRDs na mesma cadeia polipeptídica, a galectina-4.

Na mesma época, a galectina-4 foi reportada como uma proteína aderente no epitélio da língua de porco (Chiu, Jones e O'keefe, 1992; Chiu *et al.*, 1994) e depois como principal componente de *lipid rafts* em microvilosidades do intestino delgado (Danielsen e Hansen, 2008). O RNA mensageiro da galectina-4 foi identificado no trato alimentar de humanos, porcos, ratos e camundongos e em bexiga de coelhos (Huflejt e Leffler, 2003; Wooters, Ropp e Erickson, 2005).

A caracterização imuno-histoquímica da pGal4 (galectina-4 porcina) mostrou a presença da proteína no citoplasma das células epiteliais que revestem as vilosidades intestinais absorventes, com expressão maior na superfície apical (Wooters *et al.*, 2005). Essa distribuição é consistente com o papel da galectina-4 na formação e estabilização de *lipid rafts*. Estudos de RNA de interferência em adenocarcinoma do cólon humano sugerem que a galectina-4 pode desempenhar um papel importante na liberação apical de proteínas (Delacour *et al.*, 2005).

O estudo comparativo da especificidade a glicanos mostra que a galectina-4 se liga preferencialmente ao antígeno T (Gal β 1 $\rightarrow$ 3GalNAc), dissacarídeos 3'-sulfatado e 3'-sialilado, colesterol 3-sulfato e antígenos ABH do grupo sanguíneo humano (Ideo *et al.*, 2002; Huflejt e Leffler, 2003; Ideo, Seko e Yamashita, 2005; Ideo, Seko e Yamashita, 2007; Stowell *et al.*, 2010). Experimentos realizados com os domínios CRDs separadamente mostraram preferências distintas de ligação aos glicanos (Vokhmyanina *et al.*, 2012).

1.3.1.2 Galectina-6

A galectina-6, uma proteína de 33 kDa, é um membro do grupo *tandem-repeat* que foi isolado no curso da clonagem do cDNA da galectina-4 a partir do cólon de camundongo (Gitt, Colnot, *et al.*, 1998; Gitt, Xia, *et al.*, 1998). A principal diferença entre a galectina-4 e a galectina-6 de camundongo está no peptídeo de ligação dos CRDs, que é mais curto para a galectina-6.

O gene *LGALS6*, que codifica a galectina-6, surgiu por uma duplicação tandem do gene *LGALS4* no genoma de camundongo (Gitt, Colnot, *et al.*, 1998; Gitt, Xia, *et al.*, 1998; Houzelstein *et al.*, 2004) após a divergência entre ratos e camundongos. A sequência do gene *LGALS6* evoluiu sob influência de uma forte pressão seletiva, provavelmente levando ao ganho de uma nova função (Troncoso *et al.*, 2012). Interessantemente, enquanto a maioria das espécies tem apenas o gene *LGALS4*, alguns camundongos possuem os dois parálogos *LGALS4* e *LGALS6* (Houzelstein *et al.*, 2013).

No trato digestivo normal, seu padrão de expressão, da língua até o cólon, é bastante similar, o que sugere uma redundância funcional. Entretanto, duas novas propriedades foram notadas para a galectina-6: sua expressão em células enteroendócrinas e sua tendência para formar agregados maiores no núcleo e em processos de inflamação induzida na mucosa colônica (Houzelstein *et al.*, 2013). Além disso, estudos de localização celular mostraram vários aspectos diferentes na localização dessas proteínas, evidenciando uma nova identidade para a galectina-6 ao longo da evolução (Houzelstein *et al.*, 2013).

1.3.1.3 Galectina-8

A galectina-8, membro do grupo das galectinas *tandem-repeat*, foi inicialmente clonada da biblioteca de cDNA de fígado de rato e é expressa em vários tecidos incluindo pulmão, fígado, rim, cérebro, membros posteriores e músculo cardíaco (Hadari *et al.*, 1995). Em humanos, o antígeno 1 tumoral de carcinoma de próstata (PCTA-1) que é seletivamente expresso em células de carcinoma de próstata, mas não em próstata normal ou hipertrofia benigna de próstata, foi identificado como

a galectina-8 (Su *et al.*, 1996). Similarmente, Po66, um anticorpo IgG1 monoclonal de camundongo produzido por imunização com células cancerígenas escamosas, reconhece uma proteína que liga a carboidrato, Po66-CBP, que foi identificada como a galectina-8 (Dazord *et al.*, 1987; Henno *et al.*, 2002).

Uma característica interessante do gene da galectina-8 (*LGALS8*) é a capacidade de codificar vários RNAs mensageiros por splicing alternativo dando origem a três variantes *tandem-repeat* da galectina-8, mas também a quatro variantes proto, presentes tanto em linhagens de células normais quanto tumorais (Bidon *et al.*, 2001; Cooper, 2002; Ahmed, Banerjee e Vasta, 2007). Uma vez que o domínio N-terminal da galectina-8 dimeriza, o estudo dos domínios N- e C-terminais possibilita desvendar as potenciais vias de sinalização iniciadas pelos domínios separadamente (Stowell, Arthur, Slanina, *et al.*, 2008). Além do mais, o peptídeo de ligação da galectina-8 mostra sensibilidade à clivagem por trombina (Nishi *et al.*, 2006), provendo ao menos uma possível via regulatória capaz de separar as potenciais consequências funcionais desses dois domínios.

Os estudos de Stowell e colaboradores contribuem para o entendimento das estruturas quaternárias da galectina-8, bem como suas interações na superfície celular e sinalização através de cada CRD (Stowell, Arthur, Slanina, *et al.*, 2008).

Em relação à especificidade de ligação, a galectina-8 possui alta afinidade por galactosídeos 3'-O-sialilados e 3'-O-sulfatados, característica principalmente atribuída ao CRD N-terminal (Ideo *et al.*, 2003). A resolução da estrutura cristalográfica do N-terminal da galectina-8 complexado com lactose, 3'-sialil-lactose e 3'-sulfo-lactose, aliado a estudos de mutagênese sítio-dirigida e SPR (*Surface Plasmon Resonance*) mostraram que os resíduos Arg45, Gln47 e Arg59 são indispensáveis e contribuem de forma coordenada para a afinidade de ligação a oligossacarídeos sialilados e sulfatados (Ideo *et al.*, 2011).

Além disso, a galectina-8 reconhece e mata cepas de *E. coli* que expressam antígenos associados ao grupo sanguíneo de mamíferos, demonstrando que lectinas da defesa inata podem prover imunidade contra patógenos que possuem antígenos do grupo sanguíneo em sua superfície (Stowell *et al.*, 2010). Dependendo do contexto celular e o modo de apresentação (solúvel ou imobilizada) a galectina-8 pode regular a adesão, o espalhamento, o crescimento e a apoptose celular (Hadari *et al.*, 2000; Levy *et al.*, 2001; Levy *et al.*, 2003; Zick *et al.*, 2004; Diskin *et al.*, 2009).

1.3.1.4 Galectina-9

A galectina-9, originalmente isolada de células embrionárias extraídas de rim de camundongo (Wada e Kanwar, 1997), é amplamente distribuída em vários tecidos com diferentes

graus de expressão, exibindo uma expressão temporal durante estágios diferentes da gestação, sendo mais pronunciada no desenvolvimento do timo, onde possui um papel essencial na deleção clonal de células T (Wada *et al.*, 1997). Possui três variantes derivadas de *splicing* detectadas em células Jurkat, cuja diferença está na inserção de 32 amino ácidos e em uma deleção de 12 amino ácidos, respectivamente, no peptídeo de ligação da isoforma de tamanho intermediário de 35,9 kDa (Chabot *et al.*, 2002).

A galectina-9 humana foi detectada como um antígeno a partir da biblioteca de cDNA derivada de pacientes com a doença de Hodgkin, sendo expressa em leucócitos do sangue periférico e tecido linfático; foi sugerido ter um papel importante na regulação do sistema imune (Türeci *et al.*, 1997). Possui uma potente atividade quimioatraente de eosinófilos (Matsumoto *et al.*, 1998) e é produzida pelas células endoteliais estimuladas com IFN- γ tendo sua expressão aumentada em lesões inflamatórias do sistema vascular humano (Imaizumi *et al.*, 2002). Além do mais, uma expressão alterada dessa lectina tem sido reportada em fibrose cística derivada de células brônquio-epiteliais (Vega-Carrascal *et al.*, 2011).

Ambos os domínios, N e C-terminais, exibem alta afinidade por cadeias de açúcar do tipo complexo ramificadas, especialmente por glicanos tri e tetra ramificados N-ligados que carregam múltiplas unidades de N-acetilactosamina (Sato *et al.*, 2002). Mais recentemente, a estrutura cristalográfica do N-terminal da galectina-9 de camundongo em complexo com N-acetilactosamina e também com o antígeno T (Gal β 1 $\rightarrow$ 3GalNAc) revelou o mecanismo básico de reconhecimento do carboidrato, mediado pela Arg46 (Nagae *et al.*, 2006). Por outro lado, a análise estrutural do N-terminal da galectina-9 humana revelou propriedades que diferem completamente daquelas observadas na proteína ortóloga de camundongo (Nagae *et al.*, 2008; Nagae *et al.*, 2009). De fato, a galectina-9 humana exibe alta afinidade pelo pentassacarídeo Forssman e dois possíveis modos de reconhecimento para a ligação a poli-N-acetilactosamina; além disso, estudos de mutação demonstram que resíduos de amino ácidos não conservados foram importantes na determinação das especificidades do domínio (Nagae *et al.*, 2008; Nagae *et al.*, 2009). A galectina-9 possuiu uma variedade de funções celulares, inclusive diferenciação, adesão, agregação, apoptose, sobrevivência e ativação celular (Hirashima *et al.*, 2002).

1.3.1.5 Galectina-12

O gene que codifica a galectina-12 foi identificado em linhagens de células T humanas Jurkat (Yang *et al.*, 2001). Análises de *Northen blot* mostraram que o RNA mensageiro da galectina-12 foi quase indetectável em muitos tecidos, mas sua expressão foi confirmada por PCR em tempo

real (RT-PCR), no coração, pâncreas, baço, timo e leucócitos do sangue periférico (Yang *et al.*, 2001). Também está presente em níveis mais baixos no pulmão, músculo esquelético, rim, próstata, testículo, ovário e cólon, mas indetectável no cérebro, placenta e fígado (Yang *et al.*, 2001).

Interessantemente, o RNA mensageiro da galectina-12 foi encontrado ser preferencialmente expresso no tecido adiposo de camundongos e humanos e em adipócitos 3T3-L1 diferenciados (Hotta *et al.*, 2001). Enquanto a expressão do RNA mensageiro da galectina-12 não diferiu entre animais obesos e diabéticos, a restrição calórica e o tratamento com troglitazona (uma droga anti diabética que induz apoptose de grandes adipócitos e melhora a resistência à insulina) induziu o aumento do RNA mensageiro da galectina-12 e diminuiu o tamanho da célula do tecido adiposo (Hotta *et al.*, 2001).

Esses estudos sugeriram que a galectina-12 pode regular a fisiologia do adipócito através da regulação da progressão do ciclo celular e/ou apoptose (Hotta *et al.*, 2001; Yang *et al.*, 2001). Além disso, foi demonstrado que isoproterenol, insulina, o fator α de necrose tumoral (TNF- α) e glicocorticóides inibem a expressão do RNA mensageiro da galectina-12 em adipócitos 3T3-L1 (Fasshauer *et al.*, 2002). A inibição da galectina-12 por esses hormônios sugere sua contribuição para a resistência à insulina. Dentro desse contexto, os resultados sugerem que a diminuição da expressão da galectina-12 promove uma redução de células de gordura resistentes à insulina *via* apoptose o que, por sua vez, contribui para a resistência à insulina (Troncoso *et al.*, 2012).

Os resultados reforçam um potencial papel para a galectina-12 na patogênese do diabetes tipo-2, caracterizado por resistência à insulina de tecidos periféricos. Além disso, estudos mais recentes mostraram que a galectina-12 tem um papel na diferenciação de adipócitos, sugerindo que essa galectina pode servir como um regulador no desenvolvimento do tecido adiposo e um potencial alvo para o tratamento de doenças metabólicas e inflamatórias associadas a obesidade (Yang *et al.*, 2004).

Embora menos estudadas do que membros do grupo proto e quimera, as galectinas *tandem-repeat* emergiram como reguladores centrais de processos imunes, inflamatórios e neoplásicos (Troncoso *et al.*, 2012). Essas lectinas exercem uma variedade de funções provendo sinais estimulatórios ou inibitórios que agem em acordo com outras citocinas, quimiocinas ou mediadores para regular processos fisiológicos e patológicos. A expressão espaço-temporal de galectinas, tanto na iniciação quanto na resolução de processos inflamatórios pode controlar a magnitude e o desfecho dessas respostas (Troncoso *et al.*, 2012). Além disso, similar a citocinas e fatores de crescimento, as mesmas galectinas podem também mostrar efeitos divergentes dependendo da compartimentalização celular, concentração local e status de diferenciação das células alvo (Toscano *et al.*, 2007).

Além disso, a produção ou o mascaramento de diferentes glicanos, assim como mudanças nos perfis de glicosilação da superfície celular, podem ter efeitos dramáticos na ligação e função das galectinas em papéis como imunomodulação, angiogênese, diferenciação e apoptose, dentre outros (Troncoso *et al.*, 2012). Estudos adicionais são imprescindíveis na análise de características como dimerização, comprimento do peptídeo de ligação dos CRDs e sensibilidade à clivagem proteolítica, bem como a natureza oligomérica do reconhecimento de ligantes. O entendimento da relevância biológica e a distribuição nos tecidos das diferentes isoformas das galectinas *tandem-repeat*, assim como os diferentes ligantes e as vias de sinalização desencadeadas por essas lectinas proverá *insights* para intervenções terapêuticas em doenças neoplásicas e inflamatórias (Troncoso *et al.*, 2012).

No câncer, galectinas *tandem-repeat* podem regular negativamente ou positivamente a sobrevivência de células tumorais e a adesão ao endotélio ou matriz extracelular, bem como a vascularização do tumor, reações imuno-inflamatórias e outros processos que são cruciais para o espalhamento metastático. Entretanto, há muito o que investigar a fim de entender completamente o papel de galectinas *tandem-repeat* nos processos biológicos, no contexto das respostas inflamatórias e na progressão tumoral (Toscano *et al.*, 2006; Croci *et al.*, 2007; Ciliberti *et al.*, 2009).

Diante da grande relevância biológica das galectinas do grupo *tandem-repeat* associada à limitação de dados estruturais e funcionais disponíveis, o presente projeto tem interesse na caracterização desse grupo de proteínas através de uma abordagem multidisciplinar combinando ferramentas bioquímicas, biofísicas, computacionais e cristalográficas. Com o estudo estrutural, funcional e de localização celular das galectinas-4 e -12 humanas, nossa meta é mapear os mecanismos que governam as interações e a seletividade, relevantes ao reconhecimento de carboidratos, de forma a estabelecer uma correlação entre estrutura e função em distintos eventos celulares, como por exemplo, apoptose, proliferação celular e modulação da resposta imunológica, onde estão envolvidas as galectinas do grupo *tandem-repeat*.

CAPÍTULO 2

GALECTINA-4

INTRODUÇÃO

A galectina-4 humana (hGal4), pertencente ao grupo *tandem-repeat* da família das galectinas, é um monômero de aproximadamente 36 kDa, composto de dois domínios de reconhecimento de carboidratos distintos, hGal4-CRD1 no N-terminal e hGal4-CRD2 no C-terminal, conectados por um peptídeo ligante, em uma cadeia peptídica única (Leffler, Masiarz e Barondes, 1989; Oda *et al.*, 1993; Huflejt *et al.*, 1997). A sequência de aminoácidos correspondente a cada domínio foi definida por Huflejt e colaboradores (Huflejt *et al.*, 1997), sendo o domínio CRD1 correspondente aos resíduos de 1 a 150 (massa molecular de 17 kDa, com pI teórico de 6,97) e o domínio CRD2 aos resíduos 179 a 323 (massa molecular estimada de 16 kDa, com pI teórico de 9,66). Já o peptídeo ligante, correspondendo aos resíduos 151 a 178 (aproximadamente 3 kDa), é homólogo ao domínio repetitivo rico em prolinas e glicinas encontrado na sequência da galectina-3, altamente sensível ao ataque de proteases (Oda *et al.*, 1993). Na hGal4 os CRDs compartilham entre si aproximadamente 35,5 % de identidade sequencial (Rice, Longden e Bleasby, 2000) (figura 7).

Figura 7 – Alinhamento sequencial entre o CRD1 e CRD2 da hGal4. Os aminoácidos envolvidos na coordenação das moléculas de carboidratos estão marcados em ciano. A mutação conservativa em um dos resíduos envolvidos na coordenação está identificada em rosa. O alinhamento foi feito com o programa MULTALIN (Corpet, 1988) e a imagem foi gerada com o programa Esript (Robert e Gouet, 2014).

Ambos os domínios CDR1 e CDR2 da gal-4 ligam lactose (um dissacarídeo composto pelos monossacarídeos galactose e glicose, Gal β 1 $\rightarrow$ 4Glc) com afinidade similar à descrita para outras galectinas (k_d de aproximadamente 0,5-1,0 mM). Contudo, estudos comparativos entre o CRD1 e CRD2 da galectina-4 de rato descrevem diferentes especificidades frente aos carboidratos (Oda *et al.*, 1993). Os resultados demonstram que o domínio hGal4-CRD1 possui similar afinidade por lactose (Gal β 1 $\rightarrow$ 4Glc), A-tetra (GalNac α 1 $\rightarrow$ 3-(fuc α 1 $\rightarrow$ 2)Gal β 1 $\rightarrow$ 4) e core 1 (Gal β 1

→3GalNac). Em contraste, o domínio hGal4-CRD2 (C-terminal) possui maior afinidade por A-tetra (GalNac α 1→3-(fuc α 1→2)Gal β 1→4) (Oda *et al.*, 1993; Hansen *et al.*, 2001). Os resultados demonstram que os CRDs possuem afinidade similar por oligossacarídeos dos tipos SO₃-→3Lac-O-pNP, A-tetra (GalNac α 1→ α 1→2)Gal β 1→4Glc) e LNF-1 (Fuc α 1→2Gal β 1→3GlcNac β 1→3Gal β 1→4Glc); por outro lado, o CRD2 apresenta maior afinidade por SO₃-→core 1 (Gal β 1→3GalNac)-O-Bn (Ideo, Seko e Yamashita, 2005).

Estudos realizados em células Raji sugerem que a galectina-4 se liga à oligolactosaminas e que o motivo básico de ligação Gal β 1→3GlcNAc ou Gal β 1→3GalNAc foi melhor do que o canônico Gal β 1→4GlcNAc. Além disso, foi demonstrado que a galectina-4 se liga a dissacarídeos 3'-O-sulfatados e 3'-sialilados e antígenos do sistema sanguíneo com alta afinidade (Vokhmyanina *et al.*, 2012). Dados na literatura indicam que o domínio N-terminal, hGal4-CRD1, é responsável por ancorar a hGal4 na superfície celular via ligação a glicanos do tipo lactosaminas, altamente abundantes na superfície celular, enquanto que o domínio C-terminal, hGal4-CRD2, é exposto para interagir com ligantes externos (Vokhmyanina *et al.*, 2012).

Uma característica única a hGal4 é sua capacidade de se ligar a colesterol 3-sulfato de uma forma dependente da Arg45 presente no domínio N-terminal (Ideo, Seko e Yamashita, 2007). O colesterol 3-sulfato é amplamente distribuído em diferentes tecidos e é o esterol sulfato conhecido como o mais importante no plasma humano (Strott e Higashi, 2003). É um importante componente de membranas celulares, protegendo eritrócitos da lise osmótica e regulando a capacitação do esperma. Também pode controlar as atividades de serino proteases, que estão envolvidas na coagulação sanguínea, fibrinólise e adesão de células epidérmicas (Ideo, Seko e Yamashita, 2007). Além disso, o colesterol sulfato inibe a esterificação de colesterol (Nakagawa e Kojima, 1976) e pode agir na síntese do mesmo por modular a hidroximetilglutaril-CoA (Williams, Hughes-Fulford e Elias, 1985). A ligação da galectina-4 ao colesterol 3-sulfato levanta questões interessantes sobre o seu possível envolvimento nesses processos.

A galectina-4 é expressa no epitélio da mucosa oral, do esôfago, do intestino e cólon (Huflejt e Leffler, 2003). No trato gastrointestinal, mais especificamente, localiza-se na borda das células em escova dos enterócitos formando um complexo insolúvel enzima-proteína (Oda *et al.*, 1993). A hGal4 tem sido detectada como proteína de adesão celular, e está presente no epitélio da mucosa oral, no esôfago e mais especificamente na membrana citoplasmática (Wasano e Hirakawa, 1999).

Diferentes estudos mostraram o envolvimento da galectina-4 em processos inflamatórios associados à doença inflamatória do intestino (doença de Crohn e colite ulcerativa), um grupo de

condições inflamatórias crônicas e reincidentes (Podolsky, 2002) cuja patogênese envolve fatores como desequilíbrio imune, interação desregulada entre hospedeiro e micróbios e susceptibilidade genética (Strober, Fuss e Blumberg, 2002; Mizoguchi, Mizoguchi e Bhan, 2003; Mizoguchi e Mizoguchi, 2008).

Foi mostrado que a galectina-4 contribui para acentuar a colite crônica mediada por células T CD4⁺ e retardar a recuperação da lesão intestinal aguda, por estimular a produção da citocina IL-6 envolvida na patogênese da inflamação intestinal e do câncer de cólon (Hokama *et al.*, 2004; Hokama, Mizoguchi e Mizoguchi, 2008). O envolvimento da galectina-4 na inflamação pode estar associado a mudanças do glicoma das células T CD4⁺ ocasionadas por uma regulação negativa de enzimas associadas à glicosilação sob condições inflamatórias (Mizoguchi e Mizoguchi, 2007).

Por outro lado, estudos mostraram que a galectina-4 possui um papel único no controle da inflamação intestinal por induzir seletivamente a apoptose de células T e a restrição do ciclo celular (Paclik, Danese, *et al.*, 2008), além de se ligar nas células epiteliais e promover a regeneração do tecido (Paclik, Lohse, *et al.*, 2008). Esses resultados sugerem que a galectina-4 pode ser um agente anti-inflamatório na doença inflamatória do intestino (Paclik, Danese, *et al.*, 2008; Paclik, Lohse, *et al.*, 2008).

Em conjunto, esses resultados suportam um papel central, embora controverso, da hGal4 na inflamação intestinal. É possível que a ativação e apoptose de células T representem dois eventos modulados pela hGal4 que ocorrem em diferentes momentos durante o curso da resposta imune (Troncoso *et al.*, 2012).

Já em processos cancerígenos, a expressão do RNA mensageiro da hGal4 está abaixo dos níveis normais em câncer colorretal (Rechreche *et al.*, 1997) e em adenocarcinoma (Lee *et al.*, 2006); de fato, a hGal4 foi descrita como um supressor de tumor em câncer colorretal humano (Satelli *et al.*, 2011). Estudos de imunohistoquímica mostraram que durante a transformação maligna é evidente a perda progressiva dos agregados densos supra-nucleares que a hGal4 forma em condições normais (Huflejt e Leffler, 2003). Adicionalmente, a galectina-4 pode estar envolvida nos estágios iniciais do desenvolvimento de carcinoma de cólon humano (Nagy *et al.*, 2003). Estudos recentes demonstraram que a supressão da expressão da galectina-4 promove a proliferação de células cancerosas *in vitro* e *in vivo* (Kim, S. W. *et al.*, 2013).

Diferente do observado em câncer colorretal, os níveis de expressão da hGal4 foram maiores em carcinomas hepatocelulares (Kondoh *et al.*, 1999), em células de câncer gástrico com potencial metastático aumentado em relação a tecidos normais (Hippo *et al.*, 2001) e em tumores carcinoides gástricos e do íleo (Rumilla *et al.*, 2006). O gene da galectina-4 é um dos mais altamente representados em tumores gastroenteropancreático neuroendócrino (Duerr *et al.*, 2008) e em

esôfago de Barret (Van Baal *et al.*, 2005). Significativa alteração na expressão do gene da galectina-4 foi mostrado em adenocarcinomas sinonasais (Tripodi *et al.*, 2009) e em câncer mucinoso das células epiteliais de ovário (Marquez *et al.*, 2005; Heinzelmann-Schwarz *et al.*, 2006).

Em câncer de linhagens de células pancreáticas foi mostrado que a galectina-4 inibe a migração e previne a metástase por agir como uma molécula de adesão entre as células (Belo *et al.*, 2013). A expressão diferenciada da hGal4 nos diferentes processos de malignidade faz com que o uso desta galectina seja particularmente promissor como marcador diferencial no diagnóstico e prognóstico de certos tipos de câncer (Huflejt *et al.*, 1997; Kopitz *et al.*, 2003).

Poucos estudos de correlação entre estrutura e função estão disponíveis para a galectina-4. Características estruturais podem ser a chave para explicar as controversas funções da galectina-4 dentro do mesmo contexto patológico, além de poder elucidar as diferenças na preferência por ligantes entre os CRDs. Entendemos que a caracterização estrutural da galectina-4 é um passo importante no processo de desvendar os mecanismos estruturais pelos quais a proteína exerce sua função biológica tanto em processos normais quanto patológicos.

OBJETIVOS

Considerando a grande relevância da galectina-4 em processos relacionados à imunidade inata, inflamação e câncer, o desenvolvimento do presente projeto teve por objetivo contribuir para a elucidação das bases estruturais que correlacionam a ação da galectina-4 com os mais diversos contextos celulares. Nosso trabalho tem como foco a caracterização estrutural da galectina-4 humana (hGal4), além dos seus domínios N- (hGal4-CRD1) e C-terminal (hGal4-CRD2) de forma independente, através do uso de uma combinação de técnicas biofísicas que envolveram cristalografia de proteínas, modelagem molecular, dinâmica molecular e termofluor.

MATERIAIS E MÉTODOS

2.1 Produção proteica

2.1.1 Clonagem

2.1.1.1 Vetores de clonagem

Para a clonagem da hGal4 e seus domínios N- e C-terminal foram utilizados os vetores do sistema pET. Tais vetores são plasmídeos desenhados para a clonagem e expressão de proteínas recombinantes em *E. coli*. Os principais elementos que compõe o plasmídeo são o gene *lacI* que codifica o repressor lac, o promotor T7, altamente específico para a T7 RNA polimerase, o operador *lac*, onde se liga o repressor *lac* e que controla o processo de transcrição, um gene que confere resistência a um antibiótico e um sítio múltiplo de clonagem, onde o gene de interesse é inserido.

pET-28a(+)

O vetor pET-28a(+) (Novagen) foi utilizado com o objetivo de permitir a produção recombinante da proteína de interesse em fusão com uma cauda de poli histidina (figura 8) que permite a purificação por afinidade em resina de níquel.

Figura 8 – Região de clonagem múltipla do vetor pET-28a-c(+). Fonte: <http://richsingiser.com/4402/pET28.pdf>.

pET-28a(+)-sumo

O vetor pET-28a(+)-sumo, gentilmente cedido pelo Dr. André Ambrósio do LNBio (Campinas), consiste em uma versão modificada do vetor pET-28a(+) (figura 8) com o fragmento

de DNA que codifica a proteína Smt3 de 11 kDa, ubiquitina de *Saccharomyces cerevisiae* (UniProtKB/Swiss-prot: Q12306.1) clonado entre os sítios de restrição *NheI* e *BamHI*. A proteína Smt3, referida aqui como SUMO, é um membro da família *ubiquitin-like*. Essa construção permite a clonagem da proteína alvo em fusão com a proteína SUMO na extremidade N-terminal da proteína de interesse (figura 9). O processo de clivagem da fusão é realizado com a SUMO-protease ULP1 (*Ubiquitin-like-specific Protease 1* – EC 3.4.22.68) de *Saccharomyces cerevisiae* (UniProtKB/Swiss-Prot: Q02724.1). A clivagem ocorre pelo reconhecimento do motivo Gly-Gly na região C-terminal da SUMO (Mossessova e Lima, 2000), resultando na liberação da proteína apenas com os resíduos referentes ao sítio de restrição inseridos entre a SUMO e a proteína de interesse. O fragmento de DNA que codifica a SUMO-protease ULP1 também foi clonado no vetor pET-28a(+) e gentilmente cedido pelo Dr. André Ambrósio, de forma a ser produzida heterologicamente para a realização dos nossos experimentos.

Figura 9 – Representação esquemática da construção resultante da clonagem no vetor pET-28a-sumo.

A sequência de nucleotídeos que codifica a galectina-4 humana, hGal4, clonado originalmente no vetor pQE (QIAGEN), foi gentilmente cedido pelo Prof. Dr. Marcelo Dias Baruffi e utilizado como molde para a clonagem da hGal4 e seus respectivos domínios em vetores de expressão da série pET.

2.1.1.2 Clonagem da galectina-4 humana

A clonagem da hGal4 foi iniciada pelo planejamento e desenho manual dos oligonucleotídeos iniciadores (*primers*) complementares às extremidades 3' e 5' do fragmento de DNA de interesse. Os oligonucleotídeos *forward* e *reverse* foram desenhados para amplificar a sequência de interesse e introduzir os sítios de restrição para as enzimas *EcoRI* e *XhoI* (tabela 2). A temperatura de anelamento para cada oligonucleotídeos foi calculada com base na expressão: $T_m = 60 + (41N_{CG} - 500)/L$, onde N_{CG} representa o número de nucleotídeos guanina e citosina; L é o número total de nucleotídeos utilizados no desenho dos oligonucleotídeos.

Tabela 2 – Sequência dos oligonucleotídeos utilizados para amplificar o gene que codifica a proteína humana galectina-4.

Oligonucleotídeo Forward	GAC(GAATTC)ATGGCCTATGTCCCCGCA
Oligonucleotídeo Reverse	GAC(CTCGAG)TTAGATCTGGACATAGGACAAGG

† Os oligonucleotídeos foram sintetizados pela empresa Exxtend.

A amplificação foi realizada pela aplicação da técnica de reação em cadeia da polimerase – PCR, em um termociclador (*Mastercycler gradiente* – Eppendorf), utilizando 0,4 µM de cada oligonucleotídeo, 5 ng/µl de DNA molde (plasmídeo hGal4-pQE), 0,2 mM dNTPmix (Invitrogen™), 1 X *High Fidelity Buffer* (NEB), 1 U da enzima *Phusion High Fidelity* (NEB) em 50 µl de reação. A reação foi iniciada por uma desnaturação inicial a 98 °C por 30s, seguida de 25 ciclos de desnaturação (98 °C por 10s), anelamento (60 °C por 20 s) e extensão (72 °C por 15 s), com extensão final a 72 °C por 5 minutos. O produto da reação de amplificação foi analisado por visualização em luz UV após migração em gel de agarose 1% (m/v) e coloração com brometo de etídeo (AMRESCO). Após a confirmação de que a sequência foi amplificada no tamanho esperado, as reações de PCR foram purificadas utilizando o kit *QLAquick PCR Purification* (QIAGEN), segundo as instruções do fabricante.

Em seguida, tanto o material amplificado quanto o vetor de expressão pET-28a-sumo foram digeridos com as enzimas de restrição *EcoRI*-HF (NEB) e *XhoI* (NEB) em uma reação de 50 µl contendo 20 U de cada enzima, 1 X tampão *CutSmart* (NEB) e 1 µg total de amostra, a 37 °C por 30 minutos. Os produtos das reações foram submetidos a eletroforese em gel de agarose 1% (m/v), excisadas e purificadas do gel utilizando o kit *QLAquick Gel Purification* (QIAGEN), em acordo com o manual. Os fragmentos digeridos foram ligados através da reação de ligação contendo 1 µl da enzima T4 DNA ligase (NEB), 1 X o tampão da enzima (NEB) em 20 µl de reação, na qual as proporções entre inserto e vetor variaram de 3:1, 6:1 a 9:1. As reações foram conduzidas a 22 °C, por 20 minutos.

Uma alíquota do meio reacional de ligação foi utilizada para transformar bactérias competentes (Davis, 1993) por meio de choque térmico (Sambrook, Fritsch e Maniatis, 1989). As células foram semeadas em meio LB sólido previamente homogeneizado com 30 µg/ml de canamicina, em uma placa de Petri, e incubadas em uma estufa bacteriológica a 37 °C por 16 horas. As colônias selecionadas foram utilizadas como molde em uma reação de PCR e também inoculadas em meio LB. Os clones positivos selecionados foram propagados, os DNAs plasmidiais extraídos, quantificados a 260 nm e submetidos à digestão com as enzimas de restrição específicas a fim de confirmar a presença do inserto. Para confirmar a fidelidade das sequências, o DNA de

interesse foi então submetido ao sequenciamento automático no equipamento ABI 3130xL (*Applied Biosystems*) pelo Grupo de Biofísica Molecular – “Sérgio Mascarenhas”.

2.1.1.3 Clonagem dos domínios CRD1 e CRD2 da galectina-4 humana

Os domínios hGal4-CRD1 e hGal4-CRD2 foram anteriormente clonados pelo nosso grupo no vetor pET-28a(+) (Novagen) (Zimbardi *et al.*, 2010). As informações referentes à clonagem estão resumidas na tabela (tabela 3).

Tabela 3 – Sequência dos oligonucleotídeos previamente utilizados para amplificação dos domínios hGal4-CRD1 e hGal4-CRD2, que foram clonados no vetor pET-28a(+) (Novagen) para expressão heteróloga em *Escherichia coli*. Em parênteses estão as sequências referentes aos sítios de restrição.

Organismo de origem	Homo sapiens	
Oligonucleotídeo Forward	hGal4-CRD1	GACGAC(CATATG)GCCTATGTCCCCGCAC
	hGal4-CRD2	AT(CATATG)CTGCCCACCATGGAAGGACC
Oligonucleotídeo Reverse	hGal4-CRD1	GACGAC(GCGGCCGC)ATTAGCCTCCGATGAAGTTGATTGATTGAAG
	hGal4-CRD2	AT(GCGGCCGC)ATTAGATCTGGACATAGGAC

2.1.2 Expressão e purificação

2.1.2.1 Expressão e purificação da hGal4

Teste de expressão

Uma vez confirmada a sequência correta do fragmento de DNA de interesse, o plasmídeo recombinante foi utilizado para transformar cepas de *E. coli* Rosetta (DE3) (Novagen) e dar início aos ensaios preliminares de expressão da proteína de interesse. O teste foi conduzido pela variação da concentração do agente indutor, temperatura e tempo de expressão, de forma a maximizar o rendimento e a solubilidade da proteína de interesse.

O pré-inóculo foi preparado em meio LB (Acumedia) suplementado com 30 µg/ml de canamicina (Sigma-Aldrich) e 34 µg/ml de cloranfenicol (Sigma-Aldrich) cultivado a 37 °C durante 16 horas sob agitação constante de 180 rpm. A cultura foi então diluída na proporção 1:100 em meio de cultura LB, acrescido dos antibióticos canamicina e cloranfenicol, e incubada a 37 °C sob agitação até atingir uma densidade ótica entre 0,5 e 0,6 u.a. a 600 nm. Nesse momento, a cultura foi dividida em quatro volumes de 10 ml aos quais foram adicionadas as diferentes concentrações de isopropil-β-D-1-tiogalactopiranosídeo, IPTG (Sigma-Aldrich): 0 (controle negativo), 5, 50 e 500

μM . A cultura foi incubada sob agitação a 180 rpm. Foram testadas as temperaturas de 18 °C e 25 °C, durante 6 e 24 horas. Depois da incubação, as células foram coletadas por centrifugação a 4 °C, 10000g, ressuspensas em 0,5 ml de tampão A contendo 1 mM de fluoreto de fenilmetilsulfonila-PMSF (USB) e sonicadas com 5 W de potência por 5 vezes de 10 s com intervalos de 30 s. Os precipitados e sobrenadantes obtidos foram separados e analisados por eletroforese em gel de poliacrilamida em condições desnaturante (SDS-PAGE) 14 % para verificar o rendimento e a solubilidade da proteína.

Expressão

Uma única colônia de células *E. coli* Rosetta (DE3) transformadas com o plasmídeo expressando a proteína recombinante hGal4 em fusão com SUMO foi utilizada para inocular 10 ml de meio LB suplementado com 30 $\mu\text{g/ml}$ de canamicina e 34 $\mu\text{g/ml}$ de cloranfenicol cultivados a 37 °C durante 16 horas sob agitação constante 180 rpm.

A cultura foi então diluída na proporção 1:100 em meio de cultura LB, acrescido dos antibióticos canamicina e cloranfenicol, e incubada a 37 °C sob agitação até atingir uma densidade ótica entre 0,5 e 0,6 u.a. a 600 nm. A superexpressão da proteína de interesse foi induzida com 50 μM de IPTG por 24 h a 25 °C. Depois da incubação, as células foram coletadas por centrifugação a 4 °C e 10000g e armazenadas a -20 °C.

Purificação por cromatografia de afinidade

A primeira etapa de purificação foi realizada por cromatografia de afinidade em uma coluna *Poly-Prep*® empacotada com 2 ml de resina de níquel Ni-NTA (QIAGEN). O tampão base (denominado tampão A) utilizado nessa etapa conteve 50 mM NaH_2PO_4 (Sigma-Aldrich), 300 mM NaCl (JT Baker) pH 8,0 e 14 mM de 2-mercaptoetanol (2ME) (Sigma-Aldrich).

As células foram ressuspensas em tampão A contendo 1 comprimido do coquetel de inibidor de proteases SIGMAFAST™ sem EDTA (Sigma-Aldrich) e rompidas por ultrassonicação 10 W de potência, usando 10 pulsos de 30 s com intervalos de 30 s, em gelo. O sobrenadante foi separado por centrifugação durante 30 minutos a 16000g, 4 °C, e aplicado na coluna de níquel pré-empacotada e equilibrada com o tampão A. As impurezas foram separadas pela aplicação de um gradiente em degrau de 0, 25 e 40 mM de imidazol (Sigma-Aldrich) adicionado no tampão A, com 10 volumes de coluna cada. A proteína de interesse eluiu com 5 volumes de coluna de tampão A acrescido de 500 mM de imidazol.

A hGal4 foi concentrada usando um concentrador Amicon® Ultra-15 de 10 kDa (Millipore), dialisada com tampão A e submetida a clivagem proteolítica da cauda de afinidade com a protease ULP1, a 4 °C por 12 horas. Após a clivagem, o material foi repassado na coluna de níquel previamente equilibrada com tampão A. A proteína recombinante eluiu com 25 mM de imidazol enquanto que a proteína de fusão, Sumo, e a protease ULP1 eluíram da coluna com 500 mM de imidazol.

Purificação por cromatografia de exclusão por tamanho

A hGal4 foi concentrada e submetida a etapa de purificação por cromatografia de exclusão por tamanho usando a coluna Superdex™ 200 10/300 (GE Healthcare) no sistema de cromatografia Äkta purifier (GE Healthcare). A coluna foi equilibrada com tampão contendo 50 mM Hepes, 150 mM NaCl e 14 mM 2ME pH 7,2.

Em todas as etapas de purificação, o grau de pureza da amostra foi monitorado por SDS-PAGE 14%. A revelação do gel foi feita com o corante azul de Coomassie R-250 (Vetec). As frações correspondentes ao pico da eluição da proteína de interesse foram combinadas, concentradas e quantificadas com base no coeficiente de extinção teórico (Pace *et al.*, 1995) por espectroscopia no UV a 280 nm utilizando o NanoDrop 2000 (Thermo Scientific).

2.1.2.2 Expressão e purificação dos domínios hGal4-CRD1 e hGal4-CRD2

Expressão

O processo de expressão e purificação dos domínios CRD1 e CRD2 da galectina-4 foram otimizados nas mesmas condições. Uma única colônia de células *E. coli* Rosetta (DE3) transformadas com os plasmídeos das proteínas recombinantes hGal4-CRD1 e hGal4-CRD2, em pET-28a, foi utilizada para inocular 10 ml de meio LB suplementado com 30 µg/ml de canamicina e 34 µg/ml de cloranfenicol cultivados a 37 °C durante 16 horas sob agitação constante de 180 rpm.

A cultura foi então diluída na proporção 1:100 em meio de cultura LB, acrescido dos antibióticos canamicina e cloranfenicol, e incubada a 37 °C sob agitação até atingir uma densidade ótica entre 0,5 e 0,6 u.a. a 600 nm. A superexpressão das proteínas de interesse foi induzida com 1mM de IPTG por 24 h a 30 °C. Depois da incubação, as células foram coletadas por centrifugação a 4 °C e 10000g e armazenadas a -20 °C.

Purificação por cromatografia de afinidade

A primeira etapa de purificação foi realizada por cromatografia de afinidade em uma coluna *Poly-Prep*® empacotada com 2ml de resina de níquel Ni-NTA. O tampão base (denominado tampão A) utilizado nessa etapa contém 50 mM NaH₂PO₄, 300 mM NaCl pH 8,0 e 14 mM de 2ME.

As células foram ressuspensas em tampão A contendo 1 mM de PMSF e rompidas por ultrassonicação usando 10 pulsos de 30 s com intervalos de 30 s, em gelo. O sobrenadante foi separado por centrifugação a 4 °C por 30 minutos a 16000g e aplicado na coluna de níquel pré-empacotada e equilibrada com o tampão A. As impurezas foram separadas pela aplicação de um gradiente em degrau de 0, 25 e 40 mM de imidazol adicionado no tampão A, com 10 volumes de coluna cada. A proteína de interesse eluiu com 10 volumes de coluna de tampão A acrescido de 500 mM de imidazol.

Purificação por cromatografia de exclusão por tamanho

As proteínas foram concentradas usando um concentrador Amicon® Ultra-15 de 3 kDa (Millipore) e submetidas a segunda etapa de purificação por cromatografia de exclusão por tamanho usando a coluna Superdex 200 10/300 no sistema Äktapurifier. A coluna foi equilibrada com tampão contendo 50 mM Hepes, 150 mM NaCl e 14 mM 2ME.

Em todas as etapas de purificação, o grau de pureza das amostras foi monitorado por SDS-PAGE 14%. A revelação do gel foi feita com o corante azul de Coomassie R-250 (Vetec). As frações correspondentes ao pico da eluição de cada proteína, hGal4-CRD1 e hGal4-CRD2, foram combinadas, concentradas e quantificadas com base no coeficiente de extinção teórico (Pace *et al.*, 1995) por espectroscopia no UV a 280 nm utilizando o NanoDrop 2000 (*Thermo Scientific*).

2.2 Estudo estrutural

2.2.1 Cristalização

Os estudos de cristalização foram realizados com as proteínas recombinantes hGal4-CRD1 e hGal4-CRD2 fusionados a cauda de poli-histidina na região N-terminal. As concentrações finais foram calculadas utilizando o coeficiente de extinção molar teórico calculado no programa ProtParam (Gasteiger *et al.*, 2005) baseado na sequência de aminoácidos das proteínas, incluindo a cauda de poli-histidina. Os experimentos iniciais de cristalização foram realizados em placas *Cryschem* (Hampton Research) de 24 poços pelo método de difusão de vapor utilizando kits

comercialmente disponíveis: *Crystal Screen*, *Crystal Screen 2*, *PEG/ION*, *PEG/ION 2*, *PEGRx 1 e 2*, *Index* e *SaltRx1 e 2* (Hampton Research).

O método de difusão de vapor se baseia no aumento da concentração da solução de proteína até a um estado de supersaturação por evaporação dos componentes voláteis da solução de cristalização. Esses componentes são gradualmente transferidos da solução do poço para a solução do reservatório por difusão de vapor, até que a pressão de vapor entre ambas esteja em equilíbrio. O método foi implementado utilizando a técnica de gota sentada (figura 10).

Figura 10 – Representação esquemática de um poço da placa *Crychem* utilizada nos experimentos de cristalização.

Como padrão experimental inicial foram utilizados no reservatório 500 µl de solução de cristalização e no poço foram misturados 2 µl de solução de proteína e 2 µl de solução do reservatório. O sistema foi selado com a fita adesiva *Crystal clear* (Henkel *Consumer Adhesives*) e armazenado à temperatura constante de 21 °C.

2.2.1.1 Cristalização do domínio hGal4-CRD1

Os ensaios de cristalização da proteína nativa procederam conforme descrito (Zimbardi *et al.*, 2010). Os cristais otimizados foram obtidos na presença de 0,1 M Tris-HCl (Sigma-Aldrich) pH 8,5, 16% (m/v) polietilenoglicol 8000 (Sigma-Aldrich) e 0,2 M acetato de cálcio hidratado (Sigma-Aldrich) a 21 °C. O domínio hGal4-CRD1 foi submetido a experimentos de co-cristalização com ligantes utilizando a técnica de gota sentada.

Experimentos de co-cristalização com lactose

A primeira estratégia de co-cristalização foi realizada utilizando 500 µl de solução de cristalização (0,1 M Tris-HCl pH 8,5, 16% (m/v) polietilenoglicol 8000 e 0,2 M acetato de cálcio hidratado) no reservatório. As gotas consistiram de 2 µl de solução do reservatório, 2,5 µl de solução de proteína e 0,5 µl de solução de lactose. As concentrações finais de lactose foram 5 e 10 mM.

A segunda estratégia de co-cristalização foi realizada pela incubação da proteína com 100 mM de lactose e posterior varredura utilizando os kits de cristalização disponíveis. Foram utilizados

500 µl de solução de cristalização no reservatório, 2 µl de solução de proteína adicionadas de 2 µl de solução do reservatório para compor as gotas. As placas de cristalização foram mantidas a 21 °C e monitoradas regularmente.

Experimento de co-cristalização com colesterol-3-sulfato (C3S)

No primeiro experimento de co-cristalização, o composto C3S foi solubilizado em 100% de DMSO – dimetilsulfóxido (Sigma-Aldrich) para uma concentração final de 5 mM. As placas foram preparadas com 500 µl da solução de cristalização no reservatório, e as gotas com diferentes volumes entre solução do reservatório : proteína [(4,5 µl : 5 µl) e (2,0 µl : 2,5 µl)] e acrescido da quantidade fixa de 0,5 µl do composto C3S.

No segundo experimento, foi preparada uma solução estoque de 100 mM de C3S, adicionado à proteína para uma concentração final de 5 mM e incubado por 12 horas. A mistura foi centrifugada por 10 minutos a 16000g, 4 °C e o precipitado foi separado do sobrenadante. O experimento foi feito com 500 µl da solução de cristalização no reservatório, 2 µl da solução proteína-C3S e 2 µl da solução do reservatório na gota.

Em um terceiro experimento, a solução estoque de 100 mM de C3S foi adicionada à solução de cristalização para uma concentração final de 5 mM e incubado por 12 horas. A mistura foi centrifugada por 10 minutos a 16000g, 4 °C e o precipitado separado do sobrenadante. O experimento foi feito com 500 µl da solução de cristalização no reservatório, 2 µl da solução de proteína e 2 µl da solução de cristalização-C3S na gota.

2.2.1.2 Cristalização do domínio hGal4-CRD2

O domínio hGal4-CRD2, a uma concentração de aproximadamente 10 mg/ml, foi submetido a cristalização utilizando os kits comerciais disponíveis. Os primeiros cristais foram identificados em 5 diferentes condições. A fim de favorecer a formação de monocristais adequados para os experimentos de difração de raios X, foi realizada uma varredura na concentração dos componentes das condições iniciais através da variação da concentração do agente precipitante, aditivo e tipo de tampão. A partir dos novos experimentos foi escolhida a condição contendo 1,8 M de sulfato de amônio, 0,1 M cacodilato de sódio e 2% (v/v) PEG MME 550 para dar continuidade aos experimentos de cristalização. Os novos testes envolveram a variação do volume do reservatório, do volume da gota, da concentração do agente precipitante, da concentração do aditivo, experimentos de semeadura e a cristalização na presença de aditivos e óleo.

Experimentos de semeadura

O método de semeadura é uma ferramenta versátil e poderosa para a otimização das condições de crescimento de cristais (Bergfors, 2007). Foram utilizadas duas técnicas para a semeadura dos cristais, *streak seeding* e *microseeding*.

a) *Streak seeding*

A técnica de *streak seeding* consiste na transferência dos cristais para uma nova gota utilizando uma ferramenta de semeadura (Hampton Research) (Bergfors, 2003). As sementes foram depositadas em uma nova gota, contendo a proteína e a solução de cristalização, em uma linha reta na direção horizontal (figura 11). Os experimentos foram conduzidos a 21 °C.

Figura 11 – Representação esquemática do procedimento de *streak seeding*.

b) *Microseeding*

As sementes foram preparadas utilizando a ferramenta *Seed Bead™* (Hampton Research) e um agitador tipo Vortex, conforme descrito (Luft e Detitta, 1999). A partir da solução estoque foram preparadas as diluições 1:2, 1:10, 1:20, 1:40, 1:80, 1:160 e 1:320 no tampão de cristalização contendo 1,8 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 2% (v/v) PEG MME 550. Os experimentos de cristalização com micro-semeadura foram feitos utilizando uma alíquota da solução de proteína e no lugar da solução do reservatório foram adicionadas alíquotas das soluções das sementes nas diferentes diluições. Os parâmetros variados no experimento foram concentração da proteína e a proporção entre o volume da proteína e o volume da solução das sementes.

Cristalização na presença de óleo

A técnica de cristalização na presença de óleo (Chayen, 1997) foi utilizada para controlar a taxa de difusão de vapor entre a solução de proteína e a solução do reservatório. Nesse experimento uma camada de óleo foi adicionada sobre a solução do reservatório, agindo como uma barreira

para a difusão de vapor. Foram testadas diferentes proporções entre parafina e silicone (10:0, 9:1, 8:2, 7:3, 6:4, 5:5, 4:6, 3:7, 2:8, 1:9 e 0:10) em um volume total de 200 μl e 400 μl de óleo. O experimento foi realizado para a condição de cristalização contendo 1,8 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 2% (v/v) PEG MME 550, com gotas de tamanho final de 2 μl e 4 μl .

Cristalização na presença de aditivos

Foram realizados testes de cristalização com aditivos do *kit Additive Screen*TM contendo 96 aditivos distribuídos nas categorias de cátions multivalentes, sais, aminoácidos, agentes dissociativos, *linkers*, poliaminas, caotrópicos, cofatores, agentes redutores, polímeros, agentes quelantes, carboidratos, polióis, não detergentes, moléculas anfífilas, detergentes, osmólitos, reagentes orgânicos voláteis e não voláteis. Os experimentos de cristalização foram feitos pela adição de 2,5 μl de proteína, 2 μl de solução do reservatório e 0,5 μl da solução de aditivos, em uma placa de cristalização de 24 poços com 500 μl de solução de cristalização no reservatório, contendo 1,8 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 2% (v/v) de PEG MME 550.

2.2.2 Coleta de dados e processamento

O experimento de coleta de dados consiste na incidência de um feixe de raios X sobre um cristal e na medida das intensidades resultantes de cada família de planos desse cristal colocados em condição de difração (Mcpherson, 2011). Desse modo, os cristais de hGal4-CRD1 e hGal4-CRD2 foram submetidos a difração por raios X. A coleta de dados de difração de raios X foi realizada em fonte de radiação síncrotron em duas linhas de luz diferentes: I04-1 no Diamond Light Source, Oxfordshire – Inglaterra e W01B-MX2 no CNPEM/MCT – LNLS, Campinas – Brasil (Guimaraes *et al.*, 2009).

Os cristais foram envolvidos em solução crioprotetora contendo a solução de cristalização com agente precipitante em torno de 10% mais concentrado acrescido de 20% de glicerol. Os cristais coletados em Diamond foram resfriados em nitrogênio líquido e armazenados em botijão criogênico antes da coleta de dados; os cristais coletados no CNPEM/MCT-LNLS foram transportados nas placas de cristalização para Campinas, transferidos para a solução crioprotetora e resfriados diretamente em fluxo de nitrogênio gasoso, na cabeça goniométrica do difratômetro.

A coleta de dados foi realizada em temperatura criogênica (100 K), com comprimento de onda igual a 0,92 Å (linha I04-1) e 1,45 Å (linha W01B-MX2). Os dados foram registrados em detectores PILATUS 2M e MarCCD, para cada linha de luz, respectivamente. As imagens foram

processadas e escalonadas utilizando os programas MOSFLM (Leslie e Powell, 2007), SCALA (Evans, 2006) e Aimless (Evans e Murshudov, 2013).

2.2.3 Determinação da estrutura e refinamento

As fases iniciais para os domínios da galectina-4 foram obtidas por aplicação do método de substituição molecular (Rossmann, 1990) implementado no programa phaser.phenix (McCoey *et al.*, 2007; Adams *et al.*, 2010).

2.2.3.1 Determinação da estrutura e refinamento do domínio hGal4-CRD1

Para a determinação das fases iniciais, foi utilizado como modelo de busca as coordenadas do modelo preliminar obtido a partir do uso do primeiro conjunto de dados obtido (Zimbardi *et al.*, 2010). O refinamento da estrutura foi realizado utilizando o programa phenix.refine (Adams *et al.*, 2010) seguido de vários ciclos de ajustes manuais do modelo feitos utilizando o programa COOT (Emsley *et al.*, 2010), em que as posições atômicas do modelo inicial foram modificadas de forma a satisfazer a estereoquímica e o mapa de densidade eletrônica. A validação da estrutura foi realizada no MolProbity Server (Chen *et al.*, 2010).

2.2.3.2 Determinação da estrutura e refinamento do domínio hGal4-CRD2

Para a determinação das fases iniciais foram utilizadas as coordenadas do domínio C-terminal da galectina-8 humana, hGal8-CRD2 (PDB 4GXL, 3OJB, 44 % de identidade sequencial). O refinamento da estrutura foi realizado utilizando o programa phenix.refine (Adams *et al.*, 2010) seguido de vários ciclos de ajustes manuais do modelo feitos utilizando o programa COOT (Emsley *et al.*, 2010), em que as posições atômicas do modelo inicial foram modificadas de forma a satisfazer a estereoquímica e o mapa de densidade eletrônica. A validação da estrutura foi realizada no MolProbity Server (Chen *et al.*, 2010).

2.2.4 Modelagem molecular

Como uma primeira estratégia para a obtenção de um modelo que descreva a estrutura da galectina-4 inteira decidimos por utilizar as ferramentas de modelagem molecular comparativa implementadas no programa MODELLER (Sali e Blundell, 1993) e o servidor ROBETTA (Kim, Chivian e Baker, 2004) que provê ferramentas automáticas para predição de estruturas proteicas.

2.2.4.1 Modelagem do peptídeo de ligação da hGal4

Como primeira etapa, a sequência de 33 aminoácidos contendo o peptídeo de ligação entre os CRDs da galectina-4 foi submetida à modelagem *ab initio* no servidor ROBETTA. Os modelos gerados foram submetidos a 5 macrociclos de minimização de energia (considerando-se os termos geométrico de comprimento de ligação, distâncias não-ligadas, ângulos de ligação, ângulo diedral, quiralidade e planaridade) implementados no programa phenix.geometry_minimization (Adams *et al.*, 2010) e posteriormente analisado no servidor MolProbity (Chen *et al.*, 2010).

2.2.4.2 Modelagem da hGal4

As estruturas cristalográficas dos domínios hGal4-CRD1 e hGal4-CRD2 e a estrutura do modelo selecionado para o peptídeo de ligação foram utilizadas para a construção da estrutura da galectina-4 inteira utilizando o programa MODELLER v9.14. Os modelos foram refinados utilizando o método da função alvo variável (VTFM – *Variable target function method*) com gradiente conjugado e ciclos de dinâmica molecular com *simulated annealing*, implementados na rotina do programa de construção dos modelos. Os modelos gerados foram submetidos a minimização de energia e analisados no servidor MolProbity como descrito no item anterior.

2.2.5 Dinâmica molecular

Os ensaios de dinâmica molecular foram inicialmente empregados no modelo da galectina-4 com o intuito de encontrar o mínimo energético mais favorável e as posições ótimas para os CRDs, uma vez que não há dados na literatura sobre o posicionamento correto desses domínios. As estruturas cristalográficas que foram resolvidas para a galectina-8 do grupo *tandem-repeat* com os dois domínios na mesma cadeia polipeptídica são resultantes de uma forma resistente a proteólise da proteína, na qual apenas dois resíduos de aminoácidos constituem o elo entre um domínio e outro (Yoshida *et al.*, 2012; Kim, B. W. *et al.*, 2013). As simulações de dinâmica molecular foram realizadas pelo Dr. Ricardo Oliveira dos Santos Soares, atualmente pós-doutorando no Laboratório de Cristalografia de Proteínas de Ribeirão Preto.

Todas as simulações de dinâmica molecular foram feitas através do pacote GROMACS (Pronk *et al.*, 2013), e, a não ser que explicitamente indicado, contaram essencialmente com os mesmos parâmetros a seguir. Para o conjunto de potenciais que envolvem as interações entre átomos ligados e não ligados, utilizamos o campo de força AMBER99SB-ILDN (Lindorff-Larsen *et al.*, 2010). O solvente foi descrito explicitamente pelo modelo de água TIP3P. As interações

eletrostáticas foram tratadas pelo esquema PME (*Particle Mesh Ewald*) (Darden, York e Pedersen, 1993), com raios de corte de Coulomb e van der Waals ambos iguais a 1 nm, enquanto as interações covalente tiveram restrições pelo algoritmo LINCS (*Linear Constraint Solver*) (Hess *et al.*, 1997).

A temperatura do sistema foi descrita inicialmente pelas velocidades geradas por uma distribuição de Maxwell-Boltzmann e então controlada pelo algoritmo de Berendsen (durante o equilíbrio) e V-Rescale (durante a produção) (Berendsen *et al.*, 1984; Bussi, Donadio e Parrinello, 2007). A pressão do sistema foi determinada indiretamente pelo volume da unidade de simulação (sistemas dodecaédricos) e foi controlada pelos algoritmos de Berendsen e Parrinello-Rahman, nas fases de equilíbrio e de produção, respectivamente (Parrinello e Rahman, 1981; Berendsen *et al.*, 1984). Os efeitos de borda foram minimizados pela utilização de condições de contorno periódicas. O tempo de integração de cada passo do algoritmo *leap-frog* foi ajustado para 2 fs.

As etapas típicas das simulações descritas seguem basicamente o mesmo esquema, com os programas GROMACS em cada passo indicados entre parênteses. Primeiramente o modelo da proteína no formato pdb foi transformado para o formato GROMACS ao mesmo tempo em que geramos a topologia equivalente a um pH 7,0, de acordo com os parâmetros do campo de força (pdb2gmx). Em seguida, criamos uma caixa de simulação em forma de dodecaedro, de maneira que a proteína foi centralizada e teve seu eixo de simetria mais longo alinhado a 1,2 nm de distância das paredes (editconf). Em seguida, preenchemos todo o volume dodecaédrico com moléculas de água, representadas pelo modelo TIP3P, para então removermos aquelas que sobrepõem ao espaço ocupado pela proteína (genbox). De acordo com o estado de ionização da proteína, correspondente ao pH neutro, a carga total do sistema foi neutralizada, ao mesmo tempo em que íons Na⁺ e Cl⁻ suficientes foram adicionados para se alcançar o valor padrão de força iônica fisiológica (150 mM) (genion).

A introdução de várias moléculas no sistema pode causar o surgimento de forças espúrias, por isso procedemos com uma minimização de energia com o algoritmo *steepest-descent* até que o sistema atingisse um patamar de energia potencial menor que 2.10³ KJ/mol (grompp e mdrun). Para o próximo passo, procedemos com uma simulação de equilíbrio, onde a temperatura e a pressão foram estabilizadas. Para não danificar a estrutura inicial da proteína, aplicamos uma restrição de movimento nos átomos da cadeia principal enquanto a camada de solvatação foi ajustada (grompp e mdrun). Após 2 ns de dinâmica de restrição, procedemos com a dinâmica de produção, onde o tempo de simulação varia de acordo com cada sistema descrito nos próximos itens.

2.2.5.1 Determinação da conformação do peptídeo de ligação da hGal4

Em simulações de equilíbrio, em condições normais, os domínios da hGal4 mostraram uma forte tendência de aglomeração. No entanto, a conformação final do peptídeo de ligação se mostrou variável, e aconteceu de acordo com o ângulo em que ocorre a aproximação inicial entre os CRDs. Por isso, utilizamos a técnica de *flooding* (Lange, Schäfer e Grubmüller, 2006), em que foi aplicada uma força vetorial em estados de baixa energia, o que estimula o sistema a explorar o espaço conformacional em uma velocidade muito superior que uma simulação convencional. Para isso, procedemos com uma simulação de produção da hGal4 (40 ns) e extraímos uma trajetória baseada na matriz de covariância (programa *g_covar*) e PCA (*Principal Component Analysis*, programa *g_anaeig*), os quais foram utilizados para guiar a aplicação das energias na simulação de flooding (programas *make_ed*, *grompp* e *mdrun*).

2.2.5.2 Parâmetros e simulações dos ligantes LacNAc

Uma das perguntas que levantamos era se a o peptídeo de ligação entre os domínios poderia modular a ligação ao carboidrato via posicionamento dos CRDs. Para dar início a esses estudos, primeiramente foram gerados modelos da hGal4 em complexo com duas moléculas de LacNAc (N-acetil-D-lactosamina), uma em cada CRD. Para chegar a esses modelos, a estrutura da galectina-8 complexada com lactose (PDB 3VKL) foi submetida ao alinhamento estrutural com o modelo da hGal4, um CRD por vez. Uma vez superpostos, a lactose de cada domínio da hGal8 foi transferida para o modelo da hGal4. Para posicionar o ligante, as unidades de galactose do LacNAc foram superpostas com as de lactose. As moléculas de LacNAc foram obtidas da estrutura do N-terminal da galectina-9 (PDB 2D6N) (Kumar, Frank e Schwartz-Albiez, 2013).

As moléculas do LacNAc em formato *pdb* foram parametrizadas pelo uso do módulo *Carbohydrate Builder* da interface do website www.glycam.org (Kirschner *et al.*, 2008), segundo suas sequências descritas na literatura (Kumar, Frank e Schwartz-Albiez, 2013). Após a geração dos modelos, os parâmetros foram convertidos do formato nativo do programa AMBER para o formato GROMACS com o uso do *script acpype* (Wang *et al.*, 2004; Sousa Da Silva e Vranken, 2012). Os modelos foram então incluídos na topologia da cadeia proteica da hGal4 e tiveram seus resíduos renomeados de forma a serem mais diretamente identificados nas análises subsequentes.

2.3 Estudos de estabilidade térmica – Termofluor™

Os estudos de estabilidade térmica foram iniciados pelo uso da técnica de *Differential Scanning Fluorimetry* ou Termofluor™. Termofluor monitora o desenovelamento térmico de proteínas na presença de um corante fluorescente (Poklar *et al.*, 1997; Pantoliano *et al.*, 2001) que é altamente fluorescente em ambientes apolares, tais como sítios hidrofóbicos em proteínas desenoveladas (Niesen, Berglund e Vedadi, 2007).

Termofluor é um método com amplas aplicações, tanto para varredura de condições que estabilizam proteínas quanto para busca de ligantes (Cummings, Farnum e Nelen, 2006; Boivin, Kozak e Meijers, 2013). A intensidade de fluorescência é plotada como uma função da temperatura, o que gera usualmente uma curva sigmoideal que pode ser descrita por transição de dois estados. O ponto de inflexão da curva permite a determinação da temperatura de desenovelamento, T_m , da proteína.

Neste projeto, esta técnica foi utilizada para a caracterização da estabilidade térmica e identificação de ligantes da galectina-4 e seus domínios separadamente. Os experimentos foram realizados em um termociclador em tempo real - Mx3005P (*Agilent Technologies*) equipado com filtro específico para medir sinais de fluorescência na presença do corante fluorescente SYPRO® orange (492/610 nm) (Invitrogen™). Os experimentos de estabilidade térmica foram realizados para a hGal4, hGal4-CRD1 e hGal4-CRD2. As proteínas foram filtradas em uma membrana de 0,2 µm (Milipore) e quantificadas a 280 nm com base no coeficiente de extinção molar teórico.

2.3.1 *Screening* de solubilidade e estabilidade

Como primeira etapa no estudo de estabilidade térmica, os experimentos de termofluor foram realizados utilizando o kit *Solubility and stability screen* (Hampton Research) em placa de PCR contendo 96 poços. Em cada poço foi adicionado 10 µl de proteína (20 µM em tampão 50 mM Hepes 150 mM NaCl pH 7,2 14 mM β-ME), 10 µl de cada solução do kit e 1 µl de SYPRO® orange (100 X) (Invitrogen™). A temperatura foi varrida de 25 a 95 °C (1 °C/min) e as medidas de fluorescência realizadas. Os dados foram processados conforme protocolo descrito por Niesen e colaboradores (Niesen, Berglund e Vedadi, 2007).

2.3.2 Glicoconjugados

Na segunda etapa, o experimento de termofluor foi realizado na presença de diferentes carboidratos, que incluiu sacarose e lactose (em diferentes concentrações) e uma biblioteca de

compostos em colaboração com a Dra. Vanessa Leiria Campo. O experimento foi realizado nas mesmas condições descritas anteriormente. A afinidade de interação do ligante foi estimada através da variação na temperatura de desnaturação térmica da proteína na presença dos compostos.

RESULTADOS E DISCUSSÕES

2.4 Produção proteica

2.4.1 Clonagem, expressão e purificação da hGal4

A galectina-4 foi clonada com sucesso no vetor pET-28a(+)-sumo e expressa em *E. coli* Rosetta (DE3). O teste de expressão revelou as condições ótimas para a produção da proteína de forma solúvel com 50 μ M de IPTG a 25 °C por aproximadamente 24 h. A análise do gel de poliacrilamida mostrou que a construção His-SUMO-hGal4 (49 kDa) foi expressa com sucesso na fração solúvel (figura 12a poço 2) e ligou majoritariamente na coluna de níquel (figura 12a poço 3-efluente). Na etapa de lavagem dos contaminantes foi visto que o processo de eluição da proteína iniciou-se na presença de 25 mM de imidazol. Desse modo, foi adicionada apenas mais uma etapa de lavagem com 40 mM de imidazol, uma vez que observamos que a adição de outros passos de lavagem com concentrações mais altas de imidazol não eram suficientes para separar a proteína dos contaminantes e resultavam em um baixo rendimento.

Assim, após a lavagem com 40 mM de imidazol, toda a proteína foi eluída da coluna na presença de 500 mM de imidazol (figura 12a poço 7-10). A amostra foi concentrada e dialisada para a retirada do imidazol, de forma a maximizar a atividade proteolítica da ULP1. Foram adicionados 2,9 mg da protease ULP1 na amostra proteica contendo a construção His-SUMO-hGal4, por 12 horas a 4 °C.

Após o processo de clivagem (figura 12b poço 1), a solução de proteínas foi repassada na coluna de níquel. Foram coletadas amostras referentes ao efluente (figura 12b poços 2 a 4), a lavagem com tampão A (figura 12b poços 5 e 6), a eluição da hGal4 (36 kDa) com 25 mM de imidazol (figura 10b poços 7 a 11) e a eluição da SUMO e ULP1 com 500 mM de imidazol (figura 12b poços 12 a 14). Tanto a protease ULP1 como a SUMO possuem cauda de histidinas em seu N-terminal e essa característica permite que tais proteínas permanecessem ligadas a resina de níquel enquanto a hGal4 elui. Todo o processo de purificação em resina de níquel foi monitorado a 280 nm.

Figura 12 – Perfil de purificação da hGal4 em coluna de níquel. MM – PageRuler marcador de peso molecular (kDa) (*Thermo Scientific*) A) Purificação por cromatografia de afinidade em coluna de níquel: 1) pellet 2) sobrenadante 3) fração 1 do efluente 4) fração 1 da lavagem com tampão sem imidazol 5) fração 1 da lavagem com tampão 25 mM imidazol 6) fração 1 da lavagem com 40 mM imidazol 7-10) frações 1, 2, 3 e 5 da eluição com 500 mM de imidazol. B) Purificação por cromatografia de afinidade em coluna de níquel após a clivagem com a protease ULP1: 1) amostra proteica pós clivagem com ULP1 2-4) frações correspondentes ao efluente 1, 2 e 3, respectivamente, do material aplicado na coluna 5-6) frações correspondentes a eluição com tampão sem imidazol 1 e 2; 7-11) frações de 1 a 5 da eluição com 25 mM de imidazol da proteína clivada; 12-14) frações de 1 a 3 da eluição das proteínas sumo e ULP1, com 500 mM de imidazol.

Após a clivagem e o segundo processo de eluição, as frações referentes a hGal4 foram combinadas, concentradas e submetidas a cromatografia de exclusão por tamanho (figura 13).

Figura 13 – Perfil cromatográfico de purificação da hGal4 por cromatografia de exclusão por tamanho em coluna Superdex™ 200 10/300. SDS-PAGE 14% das frações coletadas da cromatografia de exclusão por tamanho. Poços 1) 8-9 ml; 2) 11-12 ml; 3) 13-14 ml; 4) 16-17 ml; 5-11) 17,5-21 ml; 12) 22-23 ml; 13) 24-25 ml.

O pico da eluição foi coletado manualmente e analisado por eletroforese em gel de poli(acrilamida) (figura 13). Em todas as etapas de purificação foi verificada a presença de um contaminante de tamanho entre 15 kDa e 20 kDa. O aparecimento dessa banda pode ser resultado de degradação uma vez que o peptídeo de ligação entre os CRDs é bastante suscetível à proteólise (Oda *et al.*, 1993) ou pode ser referente a contaminação com a SUMO. A proporção entre o contaminante majoritário e a proteína de interesse variam entre os diferentes experimentos. As frações foram concentradas e quantificadas a 280 nm, utilizando o coeficiente de extinção molar teórico $37360 \text{ M}^{-1}\text{cm}^{-1}$ baseado na sequência de aminoácidos da proteína (inclusive três resíduos adicionais, Ser-Glu-Phe, no N-terminal, devido ao processo de clonagem) e considerando todas as cisteínas reduzidas.

2.4.2 Clonagem, expressão e purificação da hGal-CRD1

O domínio N-terminal da galectina-4, hGal4-CRD1, foi clonado e expresso com sucesso em nosso laboratório. O plasmídeo recombinante pET28a-hGal4-CRD1 foi utilizado para transformar cepas *E. coli* Rosetta (DE3) e a expressão seguiu como reportado previamente (Zimbardi *et al.*, 2010). A análise do gel de poli(acrilamida) mostrou que a construção His-hGal4-CRD1 (19 kDa) foi expressa com sucesso na fração solúvel (figura 14 poço 2) e ligou majoritariamente na coluna de níquel (figura 14 poço 3-efluente). Como observado para a construção His-SUMO-hGal4, a eluição do domínio His-hGal4-CRD1 pode ser observada com 25 mM de imidazol. Assim, a etapa de lavagem e eluição foi feita exatamente como no item anterior, sendo que após a lavagem com 40 mM de imidazol, toda a proteína foi eluída da coluna na presença de 500 mM de imidazol (figura 14 poço 7-12). O processo de purificação em resina de níquel foi monitorado a 280 nm.

Figura 14 – Perfil de purificação do domínio hGal4-CRD1 por cromatografia de afinidade em coluna de níquel; MM – PageRuler marcador de peso molecular (kDa) (*Thermo Scientific*): 1) pellet 2) sobrenadante 3) fração 1 do efluente 4) fração da lavagem com tampão sem imidazol 5) fração 1 da lavagem com tampão 25 mM imidazol 6) fração 1 da lavagem com 40 mM imidazol 7-13) frações 1, 2, 3, 5 e 10 da eluição com 500 mM de imidazol.

As frações referentes a eluição da hGal4-CRD1 foram combinadas, concentradas e submetidas a cromatografia de exclusão por tamanho (figura 15).

Figura 15 – Perfil cromatográfico de purificação da hGal4-CRD1 por cromatografia de exclusão por tamanho em coluna Superdex™ 200 10/300. SDS-PAGE 14% das frações coletadas: Poços 1) 8-9 ml; 2) 11-12 ml; 3) 15-16 ml; 4) 16-17 ml; 5-11) 17,5-21 ml; 12) 22-23 ml; 13) 24-25 ml

O pico da eluição foi coletado manualmente e analisado por eletroforese em gel de poliacrilamida (figura 15). As frações referentes aos poços 5-11 foram concentradas e quantificadas a 280 nm, utilizando o coeficiente de extinção molar teórico $24410 \text{ M}^{-1}\text{cm}^{-1}$ baseado na sequência de aminoácidos da proteína (inclusive 20 resíduos adicionais, correspondente a cauda de histidina no N-terminal) e considerando todas as cisteínas reduzidas.

2.4.3 Clonagem, expressão e purificação da hGal4-CRD2

O domínio C-terminal da galectina-4, hGal4-CRD2, foi clonado (em colaboração com o laboratório de Biofísica Molecular chefiado pelo Prof. Dr. Antonio José da Costa Filho) e expresso com sucesso em nosso laboratório. O plasmídeo recombinante pET-28a-hGal4-CRD2 foi utilizado para transformar cepas *E. coli* Rosetta (DE3) e a expressão da proteína solúvel ocorreu com 1 mM de IPTG a 30 °C por aproximadamente 24 h. A análise do gel de poliacrilamida mostrou que a construção His-hGal4-CRD2 (18 kDa) foi expressa majoritariamente na fração solúvel (figura 16 poço 2) e ligou na coluna de níquel (figura 16 poço 3-efluente). Como observado para as outras construções, a eluição do domínio His-hGal4-CRD2 se inicia com 25 mM de imidazol. Assim, a etapa de lavagem e eluição foi feita como descrito anteriormente, sendo que, após a lavagem com

40 mM de imidazol, toda a proteína foi eluída da coluna na presença de 500 mM de imidazol (figura 16 poço 7-11). O processo de purificação em resina de níquel foi monitorado a 280 nm.

Figura 16 – Perfil de purificação do domínio hGal4-CRD2 por cromatografia de afinidade em coluna de níquel; MM – PageRuler marcador de peso molecular (kDa) (*Thermo Scientific*): 1) pellet 2) sobrenadante 3) fração 1 do efluente 4) fração da lavagem com tampão sem imidazol 5) fração 1 da lavagem com tampão 25 mM imidazol 6) fração 1 da lavagem com 40 mM imidazol 7-13) frações 1, 2, 3, 4, 5 e 10 da eluição com 500 mM de imidazol.

As frações referentes a eluição da hGal4-CRD2 foram combinadas, concentradas e submetidas a cromatografia de exclusão por tamanho (figura 17). O pico da eluição foi coletado manualmente e analisado por eletroforese em gel de poliácridamida (figura 17). As frações referentes aos poços 8-12 foram concentradas e quantificadas a 280 nm, utilizando o coeficiente de extinção molar teórico $11460 \text{ M}^{-1}\text{cm}^{-1}$ baseado na sequência de aminoácidos da proteína (inclusive 20 resíduos adicionais, correspondentes a cauda de histidina no N-terminal) e considerando todas as cisteínas reduzidas.

Figura 17 – Perfil cromatográfico de purificação da hGal4-CRD2 por cromatografia de exclusão por tamanho em coluna Superdex™ 200 10/300. SDS-PAGE 14% das frações coletadas: Poços 1) 8-9 ml; 2) 11-12 ml; 3) 14-15 ml; 4) 15-16 ml; 5) 16-17 ml; 6) 17-18 ml; 7) 18-18,5 ml; 8-12) 18,5-21 ml; 13) 24-25 ml

2.5 Estudo estrutural

2.5.1 Cristalização, coleta de dados e processamento

2.5.1.1 Cristalização da hGal4-CRD1, coleta de dados e processamento

A proteína foi concentrada a aproximadamente 8 mg/ml e utilizada nos ensaios de cristalização conforme descrito (Zimbardi *et al.*, 2010). Os cristais otimizados foram obtidos na presença de 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado a 21 °C (figura 18).

Figura 18 – Cristais da proteína nas condições contendo 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado.

Os melhores dados para os cristais nativos foram obtidos de uma coleta realizada na linha de luz I04-1 no Diamond Light Source, Oxfordshire – Inglaterra, utilizando um comprimento de onda de 0,92 Å. Um total de 360 imagens com ângulo de oscilação de 0,25° foram coletadas, com 0,3 s de exposição por imagem, e 205,94 mm de distância entre cristal e detector. As imagens foram gravadas em um detector Pilatus 2M e processadas utilizando os programas MOSFLM e SCALA. As estatísticas estão descritas na tabela abaixo (tabela 4).

Tabela 4 – Estatística da coleta e processamento de dados. Em parênteses estão os valores para a camada mais externa de resolução.

Fonte de difração	I04-1, Diamond
Comprimento de onda (Å)	0,92
Temperatura (K)	100
Detector	Pilatus 2M
Distância detector-cristal (mm)	205,94
Ângulo de rotação por imagem (°)	0,25
Rotação total (°)	90
Tempo de exposição por imagem (s)	0,3
Grupo espacial	P6 ₁ 22
<i>a</i> , <i>c</i> (Å)	72,50, 110,20
α , β , γ (°)	90,90, 120
Mosaicidade (°)	0,29
Resolução (Å)	36,73-1,68(1,77-1,68)
Total de reflexões	181229 (26707)
Total de reflexões únicas	20341(2898)
Completeza (%)	99,9(100)
Multiplicidade	8,9(9,2)
$\langle I/\sigma(I) \rangle$	13,5(3,6)
$R_{\text{pim.}}$ (%)	3,4(20,1)
CC(1/2)	0,998(0,831)
<i>B</i> factor Wilson plot (Å ²)	15,3

Ensaio de co-cristalização com lactose

Das várias estratégias de co-cristalização realizadas na presença de lactose, os melhores cristais foram obtidos após 5 dias para os experimentos em que a lactose foi adicionada diretamente na gota de cristalização (figura 19). Os dados de difração foram obtidos através de uma coleta realizada na linha de luz I04-1 no Diamond Light Source, Oxfordshire – Inglaterra. A estatística dos resultados está descrita abaixo (tabela 5).

Figura 19 – Cristal de hGal4-CRD1 obtidos na presença de 10 mM de lactose, nas condições contendo 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado.

Tabela 5 – Estatística da coleta e processamento de dados. Em parênteses estão os valores para a camada mais externa de resolução.

Fonte de difração	I04-1, Diamond
Comprimento de onda (Å)	0,92
Temperatura (K)	100
Detector	Pilatus 2M
Distância detector-cristal (mm)	174,8
Ângulo de rotação por imagem (°)	0,25
Rotação total (°)	120
Tempo de exposição por imagem (s)	0,3
Grupo espacial	P6 ₁ 22
a, c (Å)	72,55, 110,30
α, β, γ (°)	90,90, 120
Mosaicidade (°)	0,29
Resolução (Å)	31,73-1,48(1,56-1,48)
Total de reflexões	350098(51170)
Total de reflexões únicas	29321(4172)
Completeza (%)	100(100)
Multiplicidade	11,9(12,3)
$\langle I/\sigma(I) \rangle$	23,5(4,8)
R_{pim} (%)	1,7(15,7)
CC(1/2)	0,999(0,905)
B factor Wilson plot (Å ²)	15,4

A análise preliminar do mapa de densidade eletrônica revelou que embora os experimentos de cristalização tenham sido realizados na presença de lactose, não foi possível a obtenção do complexo proteína-ligante. Uma análise mais cuidadosa do empacotamento cristalino revelou que

nessas condições de cristalização o bolsão de ligação ao carboidrato encontra-se parcialmente obstruído, impedindo que cristais do complexo hGal4-CRD1-lactose sejam obtidos (figura 20).

Figura 20 – A) Representação em *cartoon* da unidade assimétrica da estrutura cristalográfica do domínio hGal4-CRD1. B) Representação em superfície da molécula na unidade assimétrica e sua simétrica gerado pela operação $(x-y, x, z+1/6)$. Em azul, representado como sticks, estão os principais resíduos que interagem com o carboidrato

Sendo assim, novos experimentos de cristalização foram realizados onde a proteína foi previamente incubada com 100 mM de lactose previamente aos ensaios de cristalização. Além disso, o complexo foi submetido aos novos experimentos de cristalização utilizando os kits comerciais disponíveis na tentativa de encontrar uma nova condição em que o sítio não estivesse obstruído devido o empacotamento. Até o presente momento não obtivemos sucesso na obtenção de cristais do complexo hGal4-CRD1-lactose. Novas estratégias estão sendo planejadas, assim como a busca de ligantes que possuam maior afinidade ao domínio hGal4-CRD1.

Experimento de co-cristalização com colesterol-3-sulfato (C3S)

Para o experimento de co-cristalização com colesterol-3-sulfato, a primeira dificuldade encontrada foi a alta insolubilidade do composto. Em todos os casos, no momento em que o ligante era diluído, seja na gota ou na solução de proteína e até mesmo na solução de cristalização foi observado precipitação imediata do composto. Essas observações sugeriam que a quantidade final de C3S em solução era bastante inferior e impossível de ser estimada, prejudicando a ocupação total dos sítios disponíveis para a interação com o ligante. Outra desvantagem foi a necessidade de utilizar DMSO na solubilização do C3S, resultando na presença de 10% de DMSO (para a menor diluição) na solução de proteína. Ficou evidente que a insolubilidade do C3S em solução aquosa afeta significativamente as condições físico-químicas necessárias para a obtenção dos cristais. Ainda assim, alguns poucos cristais foram obtidos, embora não reproduzíveis, quando o C3S foi adicionado como aditivo na gota (figura 21).

hGal4-CRD1-C3S

Figura 21 – Cristais de hGal4-CRD1 obtidos na presença do ligante C3S, nas condições contendo 0,1 M Tris-HCl pH 8,5, 16% (m/v) PEG 8000 e 0,2 M de acetato de cálcio hidratado.

Os dados de difração foram obtidos através de uma coleta realizada na linha de luz I04-1 no Diamond Light Source, Oxfordshire – Inglaterra. A estatística dos resultados está descrita abaixo (tabela 6).

Tabela 6 – Estatística da coleta e processamento de dados. Em parênteses estão os valores para a camada mais externa de resolução.

Fonte de difração	I04-1, Diamond
Comprimento de onda (Å)	0,92
Temperatura (K)	100
Detector	Pilatus 2M
Distância detector-cristal (mm)	236,19
Ângulo de rotação por imagem (°)	0,25
Rotação total (°)	50
Tempo de exposição por imagem (s)	0,3
Grupo espacial	P6 ₁ 22
a, c (Å)	72,98, 110,94
α, β, γ (°)	90,90, 120
Mosaicidade (°)	0,36
Resolução (Å)	41,69-1,85(1,95-1,85)
Total de reflexões	76304(11256)
Total de reflexões únicas	15418(2195)
Completeza (%)	99,2(99,6)
Multiplicidade	4,9(5,1)
$\langle I/\sigma(I) \rangle$	11,0(3,2)
R_{pim} (%)	4,4(22,5)
CC(1/2)	0,998(0,747)
B factor Wilson plot (Å ²)	16,6

A análise dos dados processados mostra que apesar da aparente mudança na morfologia externa dos cristais, o empacotamento cristalino se assemelha ao obtido para proteína na forma nativa. Pela análise preliminar do mapa de densidade eletrônica, não foi possível identificar qualquer região associada ao C3S. Experimentos adicionais para a cristalização da hGal4-CRD1 na presença de ligantes estão sendo conduzidos. Além de nos concentrarmos na identificação de novas condições de cristalização, também estamos avaliando melhores estratégias para a incubação da proteína com os ligantes desejados, ou análogos sintéticos.

É importante ressaltar que, ao contrário da lactose, os estudos de mutação sítio dirigida mostram que o resíduo Arg45, na fita adjacente ao sítio de interação com o carboidrato, é o responsável pela interação com o C3S (Ideo, Seko e Yamashita, 2007). Isso implica que diferentes estratégias deverão ser adotadas para a obtenção do complexo.

2.5.1.2 Cristalização da hGal4-CRD2, coleta de dados e processamento

Os ensaios de cristalização para o domínio hGal4-CRD2 foram iniciados pela varredura de diferentes condições nos kits de cristalização comerciais disponíveis em nosso laboratório. Foram obtidos cristais em 5 diferentes condições (tabela 7).

Tabela 7 – Condições iniciais de cristalização do domínio hGal4-CRD2.

Condição nº 15 Crystal Screen	Condição nº 20 Crystal Screen 2	Condição nº 26 Crystal Screen 2

0,1 M de cacodilato de sódio pH 6,5 30% PEG 8000 e 0,2 M sulfato de amônio	0,1 M MES pH 6,5 e 1,6 M sulfato de magnésio	0,1 MES pH 6,5 30% PEG MME 5000 e 0,2 M sulfato de amônio
Condição nº 11 PEGRx 2	Condição nº 49 Index

0,1 M Bis-Tris pH 6,5 2% v/v PEG MME 550 1,8 M sulfato de amônio	0,1 M Bis-Tris pH 6,5 45% v/v MPD e 0,2 M cloreto de cálcio diidratado	

Os experimentos de otimização foram realizados pela variação das concentrações de agente precipitante, tampão e aditivo (tabela 8).

Tabela 8 – Varredura das condições iniciais de cristalização obtidas para o domínio hGal4-CRD2.

	Agente precipitante	Tampão	Aditivo
Crystal Screen II nº 26	PEG MME 5000 (15 a 40 %)	MES pH 6,5	Sulfato de amônio (0,1 a 0,4 M)
Crystal Screen II nº 20	Sulfato de magnésio (0,1 a 1,6 M)	MES pH 6,5	-
Crystal Screen I nº15	PEG MME 8000 (15 a 40 %)	Cacodilato de sódio pH 6,5	Sulfato de amônio (0,1 a 0,4 M)
PEGRx 2 nº3	MPD (35-60%)	Acetato de sódio pH 5,5 Cacodilato de sódio pH 6,5 MES pH 6,5 Hepes pH 7,5 Tris-HCl pH 8,5	Cloreto de cálcio (0,1 a 0,25 M)
PEGRx 2 nº 11	Sulfato de amônio (1 a 2 M)	Cacodilato de sódio pH 6,5 Acetato de sódio pH 5,5 Cacodilato de sódio pH 6,5 MES pH 6,5 Hepes pH 7,5 Tris-HCl pH 8,5	2% PEG MME 550

A condição 11 do PEGRx 2 foi escolhida para dar prosseguimento aos experimentos de cristalização observando dois critérios, tamanho e reprodutibilidade dos cristais. A variação das condições iniciais de cristalização confirmou a cristalização da proteína com 1,8 M de sulfato de amônio em 0,1 M de cacodilato de sódio pH 6,5 e 2% (v/v) PEG MME 550. Em uma gota típica na cristalização da hGal4-CRD2, há cristais de morfologias diferentes e a formação de policristais é dominante (figura 22).

Figura 22 – Foto de duas gotas de cristalização da proteína hGal4-CRD2 na condição de cristalização contendo 1,8 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 2 % (v/v) PEG MME 550.

Na tentativa de melhorar a morfologia externa dos cristais através do controle da taxa de difusão de vapor, novos ensaios de cristalização foram realizados, em que foram variados os volumes do reservatório. No entanto, não foi observada nenhuma diferença no tempo de

surgimento, tamanho e morfologia externa dos cristais e, desse modo, o volume de 500 μ l de solução no reservatório foi utilizado como padrão para os ensaios subsequentes.

Foram também realizados experimentos onde foram variados o volume da gota mudando a proporção entre os volumes de solução de proteína : solução do reservatório no poço. Os volumes testados foram 1:1, 1:2, 1:3, 2:1, 2:2 e 2:3 μ l. Os melhores cristais foram obtidos na condição contendo 1:1, 2:1 e 2:2 μ l de solução de proteína : solução do reservatório (figura 23). Para as demais condições, em que a solução de cristalização estava em maior quantidade houve muita nucleação e a formação de agregados cristalinos.

Figura 23 – Foto dos cristais obtidos no experimento de variação do volume da gota. Cada cristal foi obtido com diferentes volumes entre a solução de proteína e a solução do reservatório no poço.

Em paralelo foram realizados os experimentos de semeadura. A microsemeadura foi realizada inicialmente com duas concentrações de proteína, 5,8 mg/ml e 11,6 mg/ml. As diluições das sementes utilizadas foram 1:1, 1:2, 1:10, 1:20, 1:40 e 1:80. Em todas as gotas houve a formação de muitos cristais pequenos. Assim, dois novos experimentos foram realizados, o primeiro com as diluições de 1:160 e 1:320 e a proteína a 5,8 mg/ml e o segundo com a diluição de 1:320 e a proteína nas concentrações de 3, 4, 5 e 6 mg/ml. Em ambos os casos foram observados cristais de tamanho inadequados para a realização de experimentos de difração de raios X (figura 24)

Figura 24 – Foto de um experimento de microsemeadura nas condições descritas acima.

Apesar da semeadura resultar em cristais muito pequenos para o experimento de coleta de dados, a formação dos cristais foi uniforme na gota toda, sem a presença de precipitados ou policristais como observado anteriormente. Na tentativa de otimizar o processo de semeadura, o próximo experimento foi a realização de *streak seeding* por esgotamento. Nesse caso, a haste de semeadura foi inserida em uma gota com cristais e mergulhada diretamente em gotas novas contendo a mistura das soluções de proteína e de cristalização. A haste foi passada da gota 1 para a 2 e sucessivamente para que houvesse diluição das sementes. Mesmo na maior diluição foram obtidos inúmeros cristais pequenos (figura 25).

Figura 25 – Foto dos cristais obtidos no experimento de *streak seeding*

Em todos os experimentos, os cristais surgiam em 1 dia e atingiam seu crescimento máximo após 1 mês. Em muitos deles foi observado crescimento irregular ao longo desse tempo resultando na formação de placas cristalinas sobrepostas (figura 26). A formação de placas é indesejável, podendo dificultar ou até impossibilitar a indexação do cristal majoritário.

Figura 26 – Foto de cristais típicos da hGal4-CRD2

Considerando que esse padrão de crescimento poderia ser consequência da alta taxa de crescimento dos cristais, novos experimentos foram realizados com o objetivo de diminuir a taxa

de difusão de vapor. O primeiro foi a cristalização em baixa temperatura, 4 °C, e o segundo utilizando o método de cristalização na presença de óleo. Em ambos os casos não houve formação de cristais.

Os melhores cristais, possuindo tamanho aproximado de 0,1 mm x 0,1 mm x 0,1 mm, foram obtidos usando 1,8 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 2% PEG MME 550 como condição de cristalização. Esses cristais foram submetidos a coleta de dados na linha W01B-MX2 no CNPEM/MCT – LNLS, em Campinas. A estatística dos dados está na tabela abaixo (tabela 9).

Tabela 9 – Estatística de coleta e processamento dos dados. Em parênteses estão os valores para a camada mais externa de resolução.

Fonte de difração	W01B-MX2, LNLS
Comprimento de onda (Å)	1,459
Temperatura (K)	100
Detector	MarCCD
Distância detector-cristal (mm)	110
Ângulo de rotação por imagem (°)	1
Rotação total (°)	180
Tempo de exposição por imagem (s)	20
Grupo espacial	C2
a, b, c (Å)	58,69 68,02 129,71
α, β, γ (°)	90 97,45 90
Mosaicidade (°)	0,5
Resolução (Å)	30,06 – 2,46(2,56-2,46)
Total de reflexões	66617 (7036)
Total de reflexões únicas	18528(2082)
Completeza (%)	99,8(99,5)
Multiplicidade	3,6(3,4)
$\langle I/\sigma(I) \rangle$	9,6 (2,3)
R_{pim}	7,8(40,9)
CC(1/2)	0,991(0,773)
B factor Wilson plot (Å ²)	39,56

2.5.2 Determinação da estrutura

2.5.2.1 Determinação da estrutura da hGal4-CRD1

O conjunto de dados a 1,48 Å de resolução foi utilizado para refinar a estrutura da hGal4-CRD1. O cálculo do coeficiente de Matthews (Matthews, 1968) sugeriu a presença de 1 molécula na unidade assimétrica ($2,32 \text{ Å}^3 \text{Da}^{-1}$, com conteúdo de solvente estimado de 47%). As fases iniciais foram obtidas por substituição molecular com o programa phaser.phenix. Como modelo de busca foram utilizadas as coordenadas da estrutura preliminar obtida na primeira coleta de dados (Zimbardi *et al.*, 2010). O refinamento foi feito usando o programa phenix.refine e ainda está em fase final. O modelo parcial atingiu R_{factor} 16,97% e R_{free} 17,88% e a qualidade do mapa de densidade eletrônica pode ser visualizado na figura abaixo (figura 27).

Figura 27 – Mapa de densidade eletrônica 2mFobs-DFcalc, com σ (nível de contorno) igual a 2, representado em azul.

O modelo final compreende os resíduos de 5 a 152. Duzentos sítios de solventes foram tratados como 1 íon acetato, 1 íon cálcio e 198 moléculas de água. Os fatores de temperatura médio para a macromolécula foi $22,20 \text{ Å}^2$. A estatística do refinamento e os parâmetros do modelo são dados na tabela 10.

Tabela 10 – Estatística do refinamento da estrutura da hGal4-CRD1

Refinamento	
Resolução (Å)	31,73-1,48
Número de átomos	
Proteína	1226
Ligante	2
Água	198
^a R _{factor} (%)	16,97
^b R _{free} (%)	17,88
Rmsd da geometria ideal	
Comprimento das ligações (Å)	0.007
Ângulo das ligações (°)	1,168
Fator de temperatura médio (Å ²)	22,2
Ramachandran	
Mais favoráveis (%)	99,3
Permitidos (%)	0,7
Proibidos (%)	0,0

^aR-factor = $\sum_{hkl} |F_o| - k |F_c| / \sum_{hkl} |F_o|$, onde F_o e F_c são os fatores de estrutura medido e calculado, respectivamente.

^bR-free = $\sum_{hkl} |F_o| - k |F_c| / \sum_{hkl} |F_o|$, R-factor calculado a partir de 5% das reflexões.

Estrutura global

A análise do enovelamento do domínio N-terminal no banco de dados PDBsum mostrou que a hGal4-CRD1 é composta de 13 fitas β (52%), 1 hélice 3_{10} (2%) e 46% do conteúdo de estrutura secundária distribuídos em elementos do tipo β -hairpin, β -bulge, β -turn e γ -turn. O enovelamento global consiste em um arranjo β -sanduiche disposto em uma topologia *jelly-roll*. O monômero do domínio hGal4-CRD1 consiste de 152 resíduos de aminoácidos arranjados em duas folhas β antiparalelas formadas por seis fitas beta cada (denominadas de F1-F6 e S1-S6a/b) como pode ser demonstrado no gráfico de topologia e na representação do enovelamento global (figura 28).

Figura 28 – Representação da topologia da hGal4-CRD1.

O sítio de ligação ao carboidrato está localizado em um sulco composto pelas fitas S4, S5 e S6, compreendendo os resíduos His63, Asn65, Arg67, Asn77, Trp84, Glu87 e Arg89, geralmente conservados (figura 29).

Figura 29 – Representação em *cartoon* das fitas S4, S5, S6a e S6b compreendendo os principais resíduos de interação com o carboidrato.

Superfície de potencial eletrostático

A superfície de potencial eletrostático da hGal4-CRD1 foi calculada usando o campo de força AMBER através do servidor PDB2PQR (Dolinsky *et al.*, 2004) que utiliza o programa APBS (Baker *et al.*, 2001). O estado de protonação dos resíduos foi calculado utilizando o programa PROPKA (Li, Robertson e Jensen, 2005) no pH 7,0 (figura 30).

Figura 30 – Representação da superfície de potencial eletrostático da hGal4-CRD1 em diferentes posições, mostrando as posições frontal das folhas S1-S6a/b e F1-F6 e lateral das fitas F1 e S1 e F6 e S6a/b.

A posição em $y=0^\circ$ mostra a folha 1 do sanduiche β frontalmente, contendo as fitas S1-S6a/b, já a posição $y=180^\circ$ mostra a folha oposta (fitas F1-F6). Nas posições $y=90^\circ$ e $y=270^\circ$ as

fitas F1-S1 e F6-S6, respectivamente, ficam evidenciadas (figura 30). É importante notar que na porção referente as fitas S4-S6a/b há uma distribuição de cargas majoritariamente positiva. Essa região é exatamente a de interação com glicanos, sugerindo que a preferência ao carboidrato ligante possa estar relacionada a fatores eletrônicos.

2.5.2.2 Determinação da estrutura da hGal4-CRD2

O conjunto de dados a 2,46 Å de resolução foi utilizado para refinar a estrutura nativa da hGal4-CRD2. O cálculo do coeficiente de Matthews (Matthews, 1968) sugere a presença de 3 moléculas na unidade assimétrica ($2,39 \text{ Å}^3 \text{Da}^{-1}$, com conteúdo de solvente estimado de 48,5%). Entretanto, o melhor resultado obtido na etapa de substituição molecular com o programa phaser.phenix sugeriu a presença de 4 moléculas na unidade assimétrica (coeficiente de Matthews de $1,79 \text{ Å}^3 \text{Da}^{-1}$ e conteúdo de solvente estimado em 31,33 %).

Foram utilizadas as coordenadas da estrutura do domínio C-terminal da galectina-8 humana (PDB 4GXL) que compartilha 44 % de identidade sequencial, como modelo de busca. O melhor resultado foi obtido com funções score LLG 360,48 e TFZ 8,9 (Mccoy *et al.*, 2007; Adams *et al.*, 2010), e 4 moléculas na unidade assimétrica (figura 31).

Figura 31 – Representação em *cartoon* do modelo de saída da etapa de substituição molecular nos dados da hGal4-CRD2. A figura mostra a unidade assimétrica com 4 moléculas, nas posições A) lateral, B) topo e C) frontal.

O modelo inicial obtido para a hGal4-CRD2 foi submetido a etapa de refinamento utilizando o programa phenix.refine e os ajustes manuais no mapa de densidade eletrônica foram realizados no programa COOT. Como estratégia inicial foi feito o refinamento dos parâmetros de deslocamento atômico (ADP ou B-factor) e no espaço real aliado a utilização de vínculos de simetria não cristalográfica (NCS). O refinamento parcial alcançou um R_{factor} 21,21% e R_{free} 29,85 e a qualidade do mapa de densidade eletrônica pode ser visualizado na figura 32.

Figura 32 - Mapa de densidade eletrônica 2mFobs-DFcalc, com σ (nível de contorno) igual a 1, representado em azul.

Os monômeros compreendem os resíduos de 9-145 para a cadeia A, 6-145 para as cadeias B e C e 8-145 para a cadeia D. Cinquenta e oito sítios de solventes foram tratados como moléculas de água. O fator de temperatura médio (B_{factor}) para a macromolécula foi $48,3 \text{ \AA}^2$. A estatística do refinamento e os parâmetros do modelo são dados na tabela 11.

Refinamento	
Resolução ($\text{\AA}$)	30,06-2,46
Número de átomos	
Proteína	4333
Água	58
$^a R_{\text{factor}} (\%)$	21,21
$^b R_{\text{free}} (\%)$	29,85
Rmsd da geometria ideal	
Comprimento das ligações ($\text{\AA}$)	0,011
Ângulo das ligações ($^\circ$)	1,50
Fator de temperatura médio ($\text{\AA}^2$)	48,3
Ramachandran	
Mais favoráveis (%)	98,0
Permitidos (%)	1,6
Proibidos (%)	0,4

$^a R\text{-factor} = \sum_{hkl} | |F_o| - k |F_c| | / \sum_{hkl} |F_o|$, onde F_o e F_c são os fatores de estrutura medido e calculado, respectivamente.
 $^b R\text{-free} = \sum_{hkl} | |F_o| - k |F_c| | / \sum_{hkl} |F_o|$, R-factor calculado a partir de 5% das reflexões.

Estrutura global

A análise do enovelamento do domínio C-terminal no banco de dados PDBsum mostrou que a hGal4-CRD2 é composta de 12 fitas β (52,1%), 1 hélice 3_{10} (2,1%) e 45,7% dos elementos de estrutura secundária distribuídos em β -hairpin, β -bulge, β -turn e γ -turn. O enovelamento global consiste em um arranjo β -sanduiche disposto em uma topologia jelly-roll, como esperado. O

monômero do domínio hGal4-CRD2 (cadeia B) consiste de 145 resíduos de aminoácidos arranjados em duas folhas β antiparalelas formadas por seis fitas beta cada (denominadas de F1-F6 e S1-S6a) como pode ser demonstrado no gráfico de topologia e na representação do enovelamento global (figura 33).

Figura 33 - Representação da topologia da hGal4-CRD2.

A análise comparativa com outras estruturas cristalográficas em complexo com lactose sugere que o sítio de ligação ao carboidrato está localizado em um sulco composto das fitas S4, S5 e S6, compreendendo os resíduos His58, Asn60, Arg62, Asn71, Trp78, Glu81 e Lys83, sendo a Lys83 a única substituição conservada na sequência (figura 34).

Figura 34 - Representação em *cartoon* das fitas S4, S5, S6 compreendendo os principais resíduos de interação com o carboidrato.

Superfície de potencial eletrostático

A superfície de potencial eletrostático da hGal4-CRD2 foi calculada usando o campo de força AMBER através do servidor PDB2PQR (Dolinsky *et al.*, 2004) que utiliza o programa APBS (Baker *et al.*, 2001). O estado de protonação dos resíduos foi calculado utilizando o programa PROPKA (Li, Robertson e Jensen, 2005) no pH 7,0 (figura 35).

Figura 35 - Representação da superfície de potencial eletrostático da hGal4-CRD1 em diferentes posições, mostrando as posições frontal das folhas S1-S6a e F1-F6 e lateral das fitas F1 e S1 e F6 e S6a.

A posição em $y=0^\circ$ mostra a folha 1 do sanduiche β frontalmente, contendo as fitas S1-S6a/b, já a posição $y=180^\circ$ mostra a folha oposta (fitas F1-F6). Nas posições $y=90^\circ$ e $y=270^\circ$ as fitas F1-S1 e F6-S6, respectivamente, ficam evidenciadas (figura 35). A distribuição de cargas para o hGal4-CRD2 foi majoritariamente positiva ao longo de toda a superfície (devido a contribuição da grande quantidade de resíduos de argininas, lisinas e histidinas na superfície), evidenciando uma característica interessante da superfície de potencial do domínio C-terminal, que difere completamente do N-terminal.

2.5.3 Modelagem Molecular

Devido as dificuldades encontradas na obtenção de cristais da hGal4, como uma primeira estratégia para a obtenção de um modelo que descreva a estrutura da galectina-4 inteira decidimos por utilizar as ferramentas de modelagem molecular comparativa implementadas no programa MODELLER (Sali e Blundell, 1993) e o servidor ROSETTA (Kim, Chivian e Baker, 2004) que provê ferramentas automáticas para predição de estruturas proteicas. Na primeira etapa, foram gerados 5 modelos para o peptídeo de ligação no servidor ROSETTA, que, após a etapa de minimização de energia, foram reduzidos a dois modelos finais (figura 36). Esses modelos foram posteriormente utilizados para a modelagem da galectina-4 inteira.

Figura 36 – Representação em *cartoon* dos modelos gerados para o peptídeo de ligação entre os CRDs da galectina-4.

Assim, de posse das estruturas cristalográficas dos domínios hGal4-CRD1 e hGal4-CRD2 e a estrutura dos modelos selecionados para o peptídeo de ligação foi possível construir 6 diferentes modelos para a galectina-4. Os melhores foram selecionados com base em critérios geométricos avaliados com o programa MolProbity. As estruturas finais foram minimizadas, comparadas e uma delas foi utilizada para dar início aos estudos de dinâmica molecular (figura 37).

Figura 37 – Representação em *cartoon* dos dois melhores modelos gerados para a galectina-4.

2.5.4 Dinâmica Molecular

O modelo da galectina-4 com menor energia potencial, foi selecionado para refinamento conformacional por métodos *in silico*. O modelo foi submetido a uma simulação de dinâmica molecular de equilíbrio por 40 ns, para a obtenção de uma trajetória a qual aplicamos análise dos componentes principais (PCA – *Principal Component Analysis*). Em simulação subsequente, metadinâmica molecular, utilizamos o primeiro autovetor para guiar a busca do espaço conformacional da hGal4, em que obtivemos uma estimativa dos mínimos locais de energia potencial em várias conformações durante a trajetória. Aliando tais informações com os autovetores da cadeia principal obtidos por dinâmica essencial, foi possível obter um modelo para a galectina-4, com baixo nível de energia potencial (figura 38).

Figura 38 – Representação dos modelos inicial e final da galectina-4. A) Vetores do tipo porco-espinho (“porcupine”) feitos com ajuda do script “Porcupine” do Laboratório de Modelagem e Dinâmica Molecular da UFRJ (Prof. Pascutti <http://lmdm.biof.ufrj.br/software/porcupine/index.html>). B) Representação em *cartoon*. As figuras foram feitas no programa VMD.

O modelo final foi obtido após 5 ciclos de minimização da geometria no programa phenix.geometry_minimization (Adams *et al.*, 2010) e analisado no servidor MolProbity (Chen *et al.*, 2010). O gráfico de Ramachandran mostrou que 96,6% dos resíduos estão nas regiões mais favoráveis e 3,1% nas regiões permitidas (figura 39) (Lovell *et al.*, 2003).

Figura 39 – Gráfico de Ramachandran do modelo final da galectina-4 humana.

No modelo final obtido por técnicas de dinâmica molecular, observamos a formação de uma estrutura compacta para a hGal4, em que os domínios hGal4-CRD1 e hGal4-CRD2 interagem

indiretamente através de ligações de hidrogênio com resíduos do peptídeo de ligação. Nesse modelo, os respectivos sítios de ligação ao carboidrato se encontram diametralmente opostos e expostos para o solvente.

2.5.4.1 hGal4-LacNAc

Ao modelo final da galectina-4 foram adicionadas duas moléculas de LacNAc, uma em cada CRD, a fim de avaliar o comportamento proteico nessas condições. O procedimento utilizado para esse fim foi descrito na seção 2.2.5.2. As análises de dinâmica mostraram que o modelo apresentou níveis distintos de deformação ao longo da cadeia. Em geral, o desvio em relação à estrutura inicial foi apenas residual, sendo que os valores mais significativos foram decorrentes principalmente das flutuações do peptídeo de ligação (figura 40a).

Figura 40 – A) Gráfico de RMSD dos domínios hGal4-CRD1 e hGal4-CRD2 e do peptídeo de ligação em cada momento da simulação comparado com a posição inicial dos mesmos e B) matriz de RMSD da hGal4 em cada momento da simulação comparado com a sua posição inicial.

Comparando a conformação em cada momento da simulação com o seu estado inicial obtemos uma matriz de RMSD. A matriz obtida na simulação hGal4-LacNAc confirma a baixa deformação que a proteína sofre ao longo da trajetória, já que a maior parte dos valores de RMSD estiveram significativamente abaixo de 1 nm (figura 40b). Apesar da baixa variação estrutural, em geral, existem dois estados predominantemente visitados pela proteína, que correspondem àquelas estruturas que permeiam valores de RMSD de 0,15 nm e 0,4 nm (figura 41a). Tais estruturas têm como principal diferença a conformação de um *loop* relativamente flexível (entre os resíduos Gly159 e Gly168, aproximadamente), localizado adjacente à α -hélice do peptídeo de ligação (figura 41b).

Figura 41 – Clusterização das estruturas visitadas durante a trajetória. Os dados indicam que a estrutura da proteína apresenta predominância de dois estados (clusters). B) A flexibilidade da cadeia, de acordo com o RMSF. Maiores flexibilidades se encontram no N-terminal e no peptídeo de ligação

Com exceção do *loop* supracitado e do N-terminal (figura 42), de um modo geral o modelo da hGal4 apresentou uma estrutura bastante rígida. Aproximadamente 48% dos resíduos permanecem em conformação de folhas β , 4 % α -hélice e 48% distribuídos em outros elementos de estrutura secundária.

Figura 42 – Representação em *cartoon putty* com coloração do fator de temperatura do modelo da hGal4.

Com o objetivo de identificar os resíduos que interagiram mais durante a simulação, as ligações de hidrogênio mais relevantes, com uma permanência total maior que 30% do tempo, foram isoladas. As glutaminas se mostraram como resíduos bastante importantes na interface direta entre os CRDs, especialmente a Gln145, que forma o único par perene com a Gln322, e mais tardiamente com Glu183. Os pares Gln143-Tyr320 e Gln36-Thr181 se mostraram aparentemente coordenados, apresentando aumento de interação gradual na segunda metade da simulação (figura 43). No entanto, a interface entre os CRDs foi mais fortemente representada pela ligação indireta CRD1-peptídeo-CRD2, onde seis entre sete pares perenes foram formados. Por outro lado, os resíduos que formam ligações de hidrogênio exclusivamente intrapeptídeo foram aqueles que constituem a α -hélice (tabela 12).

Embora o sítio de ligação dos CRDs tenha uma constituição bastante conservada, as duas moléculas de LacNAc se comportaram de maneiras levemente distintas. Na primeira metade da simulação, o LacNAc ligado ao CRD1 abandona gradualmente a interação consigo próprio, bem como com a His63, para o estabelecimento de uma forte coordenação com os resíduos Asp69 e Lys44, fora do sítio canônico de interação com o carboidrato, o que resultou em uma ligação bastante estável. Já no CRD2 há uma movimentação de todo sítio e um ajuste na posição do LacNAc em uma posição que permitiu a manutenção da interação com cinco dos sete resíduos do sítio de ligação ao carboidrato. Os pares das ligações de hidrogênio com permanência maior que 25% estão descritas na tabela 13.

Tabela 13 – Relação dos pares de interação durante a simulação. As ligações de hidrogênio com tempo de permanência maior que 25% foram isoladas. As interações foram classificadas entre CRD1-LacNAc e CRD2-LacNAc.

	Resíduo 1	Resíduo 2	Ocupação (%)
CRD1-LacNAc	2 0LB (H20)	Asp69 (OD2)	26,9
	2 0LB (H40)	Lys44 (O)	26,8
CRD2-LacNAc	Asn249 (D21)	2 0LB (O6)	94,7
	Arg240 (H21)	2 0LB (O4)	49,6
	Arg240 (H21)	1 4YB (O3)	64,8
	Arg240 (H11)	1 4YB (O3)	84,4
	2 0LB (H60)	Glu259 (OE2)	42,3
	2 0LB (H60)	Glu259 (OE1)	53,1
	1 4YB (H30)	Glu259 (OE2)	49,7
	1 4YB (H30)	Glu259 (OE1)	59,5

Observando o modelo final obtido pela dinâmica molecular podemos visualizar as interações entre as moléculas de LacNAc e os respectivos domínios CRD1 e CRD2 (figura 44), assim como uma comparação entre a superfície de potencial eletrostático da molécula no início e no final da trajetória (figura 45). Pelas análises de interação de hidrogênio ao longo da trajetória, vimos que houve uma diferença marcante na acomodação das moléculas de LacNAc no bolso de ligação da proteína. Para o CRD1, apesar da molécula ter estabelecido uma interação forte com o resíduo Asp69 no final da dinâmica, poucas interações entre proteína e ligante existem. Por outro lado, para o CRD2, foi observada a presença de várias interações com os resíduos do sítio de reconhecimento ao carboidrato, porém a análise do gráfico de interações de hidrogênio (figura 43) mostrou que a interação proteína-ligante, para esse sítio, ainda não estabilizou. Portanto, um eventual alongamento da trajetória provavelmente beneficiará a consolidação dessa interação e a consequente manifestação de novos contatos e movimentos coordenados. Podemos observar, de um modo geral, que as transições nas ligações de hidrogênio entre CRD1 e LacNAc acontecem concomitantemente à medida que as interações significativas entre os domínios CRD1 e CRD2 se intensificam, bem como durante a ocorrência de perturbações entre as interações CRD1-peptídeo,

CRD2-peptídeo e peptídeo-peptídeo. Essas perturbações indicam fortemente a possibilidade de uma coordenação entre essas regiões.

Figura 44 – Representação em sticks das interações do ligante LacNAc com os resíduos no CRD1 e CRD2 após a etapa de dinâmica molecular.

Figura 45 – Superfície de potencial eletrostático para os CRDs no início e no final da dinâmica molecular.

2.5.5 Espalhamento a baixo ângulo

Como uma segunda abordagem no estudo estrutural da hGal4, foram iniciados ensaios de espalhamento a baixo ângulo (SAXS) a fim de construir o envelope dessa proteína e comparar com o modelo obtido da simulação de dinâmica molecular. Além disso, temos interesse em verificar se há diferenças significativas no envelope na presença e ausência de ligantes. Foram preparadas três amostras a 5 mg/ml em tampão contendo 50 mM Hepes 150 mM NaCl pH 7,2 e 14 mM β ME, para a proteína nativa, na presença de 20 mM de sacarose (controle negativo) e 20 mM de lactose (ligante natural). De um modo geral, a análise dos dados revelou mudanças conformacionais sutis (figura 46-1) que são dominadas pelas diferenças no perfil das curvas de intensidade para altos valores de q (figura 46-2) como indicado.

Figura 46 – Perfil de espalhamento das amostras e gráfico de Kratky do espalhamento, com subtração da curva de espalhamento para o tampão. Em vermelho, está a curva de espalhamento para a hGal4, seguida do perfil de espalhamento para a hGal4 com lactose, em verde, e com a sacarose em azul.

A análise do gráfico de Kratky mostrou que as intensidades de espalhamento para as proteínas com sacarose e lactose aumentam drasticamente em altos valores de q em comparação com o espalhamento da proteína nativa (figura 46-4). Esse comportamento é característico de proteínas desenoveladas, mas pode ocorrer também devido a uma incompatibilidade entre o tampão da amostra e o utilizado para a subtração do espalhamento pelos componentes do tampão. Entretanto, a diferença sutil entre as curvas para a sacarose e lactose (figura 46-3) pode revelar diferenças de conformação na presença do ligante.

Por outro lado, o cálculo do raio de giração (R_g) para as três amostras resultou em valores muito similares. Se um desenovelamento estivesse ocorrendo somente para as amostras com

lactose e sacarose, o raio de giração deveria ser maior do que para a proteína sem carboidratos. Para cada uma das 3 amostras há um leve aumento no R_g com a concentração, o que indica que as proteínas tendem a associar nessas condições. Entretanto, esperaríamos um grau maior da associação dependente da concentração para proteínas desenhovadas. A análise da função de distribuição de distâncias mostrou três características na curva compatíveis com a presença de dois domínios (figura 47-5) e a presença de alguma proteína agregada (figura 47-6).

Figura 47 – Função de distribuição de distâncias.

Entretanto, o cálculo da qualidade da curva de distribuição de distâncias (figura 47) resultou em aproximadamente 47% de confiança, o que sugere que novas medidas deverão ser realizadas.

2.6 Estudos de estabilidade térmica

Os estudos de estabilidade térmica tiveram como objetivo mapear as diferenças de comportamento e afinidade por ligantes entre a galectina-4 e seus domínios separadamente. Assim, não só a busca por possíveis ligantes é uma meta nesse estudo, como também entender as diferenças de comportamento das proteínas na presença de compostos variados.

2.6.1 *Screening* de solubilidade e estabilidade

Foi utilizado nessa etapa o kit *Solubility and Stability Screen* (Hampton Research) como ponto de partida nos ensaios de estabilidade térmica. O primeiro resultado obtido mostrou que as três proteínas possuem diferentes temperaturas de desenhovamento (T_m), sendo a hGal4 a menos

estável com 55,9 °C, seguida do domínio hGal4-CRD1 com 57,3 °C e o domínio hGal4-CRD2 com 71,9 °C (figura 48).

Figura 48 – Gráfico do perfil de fluorescência relativa das proteínas em função do aumento da temperatura.

As pequenas variações observadas entre as temperaturas de desenovelamento para a hGal4 e o domínio hGal4-CRD1 podem sugerir inicialmente que o processo de desenovelamento da hGal4 é governado pelo domínio hGal4-CRD1. No entanto, não podemos descartar a hipótese de que a presença do peptídeo de ligação também tenha influência na estabilidade térmica da galectina-4. Outro fato bastante importante se baseia no perfil sigmoidal observado tanto para a proteína inteira quanto para seus domínios independentes. Era de se esperar, considerando que a hGal4-CRD1 e hGal4-CRD2 possuem T_m bastante distintos e os domínios como estruturas independentes na proteína hGal4 inteira, que o perfil de desnaturação térmica seguisse um modelo de dois passos, caracterizado por dois pontos de inflexão, em que cada ponto correspondesse a transição de um domínio.

Entretanto, com base nos resultados obtidos e nos dados de dinâmica molecular é coerente propor que os domínios hGal4-CRD1 e hGal4-CRD2 interajam entre si, seja diretamente ou por meio do peptídeo de ligação, de tal forma que a desestruturação de um domínio leva a consequente desestruturação de toda a proteína.

Dando continuidade aos nossos estudos, avaliando a estabilidade da proteína e seus domínios na presença de diferentes aditivos e adotando o critério de estabilização para um ΔT_m

acima de 0,5 °C, observamos que a hGal4 teve sua temperatura de desenovelamento aumentada por 26 dos 96 compostos do kit, seguida de 15 para a hGal4-CRD1 e 7 para a hGal4-CRD2 (tabela 14). É importante mencionar que o domínio hGal4-CRD2 se mostrou bastante susceptível às variações físico-químicas testadas, exibindo em várias condições valores negativos para o ΔT_m .

Tabela 14 – Tabela dos valores de ΔT_m obtidos para a hGal4, hGal4-CRD1 e hGal4-CRD2 no experimento de termofluor realizado para o kit de solubilidade e estabilidade da Hampton Research. valores < 1, valores 1 > 2, valores 2 > 5 e valores > 5.

Categoria	Composto	Concentração	hGal4	hGal4-CRD1	hGal4-CRD2
aminoácidos	L-Arginina	125 mM	0,9	0,1	-8,0
	Glicina	250 mM	0,8	0,3	-1,7
	L-Histidina	60 mM	0,6	-0,4	-2,3
	B-Alanina	250 mM	0,7	0,1	-1,7
	L-Serina	250 mM	0,7	-0,2	-1,4
Osmólito	N-óxido de trimetilamina	1250 mM	3,1	3,1	1,3
	Metil α -D-glicopiranosídeo	1000 mM	1,8	0,6	-0,1
	Betáina monoidratada	1250 mM	1,7	0,6	2,5
	D-(+)-Trealose diidratada	375 mM	3,7	2,7	1,1
	Xilitol	1000 mM	4,7	3,4	2,2
	D-sorbitol	1000 mM	4,4	3,5	2,4
	Sacarose	1000 mM	7,4	6,2	5,5
	Hidroxiectoína	250 mM	0,9	-0,3	-2,2
Não-detergente	Non-Detergent Sulfobetina 211	500 mM	0,5	-0,7	-0,6
Ácido orgânico	Malonato de sódio pH 7,0	250 mM	1,6	1,8	-3,2
	Ácido succínico pH 7,0	250 mM	1,3	1,5	-3,9
	Tacsinato pH 7,0	2,5 % v/v	0,8	-0,2	-4,0
Sal	Sulfato de magnésio hidratado	250 mM	0,5	0,4	-4,2
	Ácido DL-málico pH 7,0	250 mM	1,5	1,5	-4,1
	Citrato tribásico de lítio tetraidratado	250 mM	2,4	2,6	-3,1
	Cloreto de potássio	700 mM	0,6	-0,3	-5,6
	Fosfato monobásico de sódio monoidratado, fosfato dibásico de potássio	350 mM/650 mM	1,6	7,2	-4,7
	Sulfato de sódio decaidratado	500 mM	2,6	5,2	-4,8
Poliol/sal	Glicerol/cloreto de lítio	20% v/v / 200 mM	3,5	1,7	0,3
Poliol	Glicerol	25% v/v	4,4	2,5	2,4
	Polipropileno glicol P 400	25% v/v	8,5		

É interessante notar que os compostos que estabilizaram tanto a galectina-4 como seus domínios são principalmente osmólitos (aminoácidos, carboidratos e polióis) moléculas usadas por organismos vivos em ambiente de estresse celular e que são ferramentas de estabilização de

macromoléculas de um modo geral (Arakawa e Timasheff, 1985; Harries e Rösger, 2008). No entanto, notamos que embora algumas classes de moléculas induzam o aumento do T_m nos nossos 3 modelos experimentais, hGal4, hGal4-CRD1 e hGal4-CRD2, simultaneamente (figura 49), algumas são seletivas para a proteína inteira, como os aminoácidos (tabela 16). Especulativamente, podemos considerar que a proteína inteira é muito mais estável, em diferentes ambientes químicos, que os domínios independentes e, portanto, mais fácil de ser estabilizada.

Outro dado importante é observarmos que existe, em geral, uma maior similaridade entre o comportamento da hGal4 com o domínio hGal4-CRD1, do que quando comparamos hGal4 e hGal4-CRD2 e mesmo hGal4-CRD1 com hGal4-CRD2. Esses resultados corroboram nossas hipóteses preliminares de que os domínios interagem entre si, na proteína inteira e que o domínio hGal4-CRD1 governa o comportamento da proteína como um todo.

Figura 49 – Gráfico da variação da temperatura de desenovelamento ΔT_m em função dos compostos que estabilizaram todas as proteínas testadas, hGal4, hGal4-CRD1 e hGal4-CRD2.

2.6.2 Glicanos

Como segunda etapa do experimento de termoestabilidade fizemos a titulação da lactose e da sacarose de 0,1 a 12,8 mM para a hGal4, hGal4-CRD1 e hGal4-CRD2. A lactose é um ligante de baixa afinidade ($K_D \sim 0,8$ mM (Ideo *et al.*, 2002) para a galectina-4 e foi titulado a fim de estabelecermos um controle positivo no *screening* de ligantes. Já a sacarose, não é ligante para a galectina-4 e, apesar de ter aumentado a temperatura de desenovelamento da hGal4 e seus CRDs, a concentração utilizada para a sacarose no primeiro experimento foi de 1M. É esperado que a baixas concentrações o efeito estabilizador da sacarose seja reduzido (figura 50).

Figura 50 – Gráfico da variação da temperatura de desenovelamento (ΔT_m) em função da concentração de ligante. S corresponde aos valores de ΔT_m obtidos para a sacarose e L corresponde aos valores de ΔT_m obtidos para a lactose.

A análise do gráfico acima mostra que com o aumento da concentração, a variação na temperatura de desenovelamento é maior para a lactose do que para a sacarose. Entretanto, para a hGal4-CRD2 nas concentrações de 0,8 e 1,6 mM, a sacarose teve um efeito estabilizador maior do que a lactose. Esse resultado mostra que é necessário traçar um intervalo de confiança para a interpretação dos experimentos de busca por ligantes.

Uma informação importante é que os formatos da curva de resposta ao aumento da concentração do ligante são reveladoras para avaliar os detalhes do mecanismo de interação (Matulis *et al.*, 2005). Por exemplo, a figura 49 revela que ainda que a sacarose induza aumento na estabilidade térmica, seu comportamento aleatório frente à variação da concentração, sugere fortemente que o processo de estabilidade, nesse caso, se faz pelo efeito osmótico, através da interação com a cadeia polipeptídica. A habilidade dos osmólitos em proteger o enovelamento é devido ao seu efeito solvofóbico, desfavorecendo contato com o *backbone* e protegendo do solvente.

Para confirmar que um composto estabiliza a proteína de uma maneira dependente da concentração, é importante fazer a titulação, que é crítica para investigar se pequenos efeitos estabilizadores seriam válidos e reproduzíveis. Além disso, esse experimento ajuda a determinar a concentração na qual o máximo efeito estabilizador pode ser obtido (Niesen, Berglund e Vedadi, 2007). Para avaliar o comportamento das curvas com o aumento da concentração do ligante, procedemos a titulação da lactose para a hGal4, hGal4-CRD1 e hGal4-CRD2 (figura 51).

Figura 51 – Gráfico da variação da temperatura de desnvolvimento em função da concentração do ligante.

O gráfico mostra que o efeito estabilizante da lactose na proteína é dependente da concentração. Matulis e colaboradores demonstram o uso dessa curva, obtida dos experimentos de termofluor, para extrair as constantes de ligação (Matulis *et al.*, 2005). Entretanto, nosso intuito inicial é utilizar o método para fazer uma varredura de ligantes. Os cálculos das constantes de ligação utilizando o termofluor poderão ser implementados no futuro, para investigações quantitativas.

Demos início ao *screening* de ligantes utilizando uma biblioteca de moléculas (tabela 15) sintetizadas pela Dr. Vanessa Leiria Campo e validadas como ligantes para outras galectinas humanas.

Tabela 15 – Biblioteca de glicoconjugados testados e a nomenclatura utilizada

VA1	VA2	VA3	VA4	VA5
α GlcNAc-ThrOH	α LacNAc-ThrOH	α GalNAc-SerOH	α GalNAc-ThrOH	β Gal(1,3)- α GalNAc-ThrOH
VA6	VA7	VA8	VA9	VA10
Glicopeptídeo (α GalNAc) ₃ Thr ₃ OH	Glicopeptídeo (β Gal(1,3)- α GalNAc) ₃ Thr ₃ OH	β Gal-SerOH	3-Thr- α GalOMe	Derivado triazólico de (β Lac) ₂

Como primeira etapa, iniciamos a varredura dos compostos com a hGal4. As moléculas foram testadas na concentração de 0,5 mM com a proteína a 10 μ M. Os ensaios foram feitos como descrito anteriormente (figura 52).

Figura 52 – Gráfico da variação da temperatura de desenovelamento da hGal4 em função dos compostos VA1-VA10.

Os resultados dessa varredura inicial com a biblioteca de potenciais ligantes resultaram que a maioria deles promoveu alguma variação no ΔT_m da hGal4, exceto os compostos VA6 e VA10. As variações mais significativas foram observadas para os compostos VA3 com $\Delta T_m = 0,47$ °C e VA7 com $\Delta T_m = 0,61$ °C. Com base nos resultados obtidos, ainda não é possível traçar uma correlação entre os valores de ΔT_m obtidos e a estrutura dos compostos. O próximo passo para verificar o real efeito de estabilização térmica será proceder com a titulação dos compostos. Uma vez que os compostos estejam validados, testaremos a estabilidade com os domínios separadamente, a fim de identificar o grau de afinidade de cada um deles pelas moléculas em questão.

2.7 Comparação estrutural

A comparação entre as estruturas dos domínios hGal4-CRD1 e hGal4-CRD2 revela que ambos modelos compartilham alta similaridade estrutural. A sobreposição entre os elementos de estruturas secundárias (*Secondary Structure Matching Superpose* – Coot) dos domínios hGal4-CRD1 e hGal4-CRD2 resultou em um RMS médio de 1,21 Å (Emsley *et al.*, 2010) sendo as maiores diferenças encontradas principalmente nas regiões de *loop* (figura 53).

Figura 53 – Representação em *ribbon* da sobreposição da cadeia principal das estruturas da hGal4-CRD1 e hGal4-CRD2, RMS=1,21 Å.

No entanto, quando comparamos a superfície de potencial eletrostático observamos grande divergência em termos de distribuição de carga na superfície dos domínios. Como pode ser visualizado na figura abaixo, a distribuição de carga do domínio C-terminal é majoritariamente positiva, principalmente na folha F1-F6, enquanto que para o domínio N-terminal, apenas a região referente ao sítio de ligação ao carboidrato se mostrou mais positivamente carregada (figura 54).

Figura 54 – Representação da superfície de potencial eletrostático para os domínios hGal4-CRD1 e hGal4-CRD2 em diferentes posições.

Embora tenhamos obtido com sucesso as estruturas dos domínios hGal4-CRD1 e hGal4-CRD2, estruturalmente similares entre si e também quando comparados com outros CRDs, pouco

se sabe a respeito da estrutura global das proteínas do grupo *tandem-repeat*. A presença do peptídeo de ligação nas galectinas *tandem-repeat* é também uma questão não totalmente elucidada. Em vertebrados, tais peptídeos são os principais alvos de variações em duplicação gênica e em variantes de *splicing*. Isso sugere que essas sequências podem ter a função de regular a orientação dos domínios um em relação ao outro ou mesmo providenciar um sítio para a interação com outras proteínas (Cooper, 2002). De fato, a alta variabilidade sequencial observada para o peptídeo revela que essa deve ser uma região importante para a seletividade e consequente distintas funções biológicas das proteínas do grupo *tandem-repeat*.

Nossos estudos de modelagem e dinâmica molecular levaram à identificação de uma hélice no peptídeo de ligação e do empacotamento dos domínios hGal4-CRD1 e hGal4-CRD2. De fato, em nossos experimentos preliminares de estabilidade térmica, observamos diferença de comportamento entre a galectina-4 inteira e seus domínios, quando comparados nas mesmas condições, sugerindo que a resposta da proteína ao ambiente foi influenciada pela presença ou ausência do peptídeo de ligação.

Um estudo realizado com a galectina-8 mostrou que os dois CRDs não são funcionalmente independentes, sugerindo que essa proteína deve ser vista como uma entidade funcional única, na qual é essencial que os CRDs estejam propriamente orientados e agindo em conjunto para o desempenho das funções de adesão e sinalização da galectina-8 (Levy *et al.*, 2006).

Recentemente, um estudo combinando duas variantes da galectina-4, que sofreram alterações no comprimento do peptídeo de ligação, com a síntese de glicoclusters, mostrou que o comprimento do peptídeo de ligação determinou as propriedades de ligação nos ensaios realizados. Quando comparado com a proteína nativa, os dados revelaram um nível de reatividade diminuída das variantes para com os compostos testados, nos casos em que o peptídeo foi diminuído e truncado. Assim, os autores definiram que a apresentação espacial é um dos fatores que estabelece a reatividade de um glicano para a lectina (André *et al.*, 2014).

Dependendo do ambiente químico, os CRDs podem ser orientados de forma a minimizar a energia entre eles, principalmente observando as características de distribuição de cargas para o CRD2. Uma análise do posicionamento final dos CRDs mostra que há uma tendência em que eles fiquem com os CRDs nas posições opostas um ao outro. Nessa posição, a região altamente positiva da hGal4-CRD2 permanece voltada para o peptídeo de ligação, modificando a distribuição de cargas na superfície da proteína e provavelmente a habilidade de interação com os ligantes (figura 55).

Figura 55 – Superfície de potencial eletrostático calculado para o modelo da galectina-4.

Assim, entendemos que o posicionamento dos CRDs mediado pelo peptídeo de ligação não só regula a orientação pelo qual o CRD deve se ligar ao carboidrato, mas também modula o tipo de carboidrato a que o CRD pode se ligar, pelas restrições de carga que impõe à proteína por expor ou não regiões com uma carga marcadamente positiva ou negativa.

Nesse modelo (figura 56), a quantidade total de hélices foi estimada em 2,79%. Contudo, estudos de SRCD (Synchrotron Radiation Circular Dichroism) realizados pelo grupo de Biofísica Molecular chefiado pelo Prof. Dr. Antonio José da Costa Filho, sugerem que a quantidade de hélices para a hGal4 de aproximadamente 8%. Essa discrepância em termos de porcentagem de conteúdo de estrutura secundária pode ser apenas consequência da imprecisão da técnica e da limitação no uso de banco de dados adequadas, como também poderia ser explicada como resultado de reajustes, no caso ganho e perda de hélices, que podem ocorrer sob diferentes condições físico-químicas ou mesmo na presença de diferentes glicanos.

Figura 56 – A) Representação da topologia da galectina-4. Topologia gerada no servidor PDBsum. Representação em *cartoon* do enovelamento global e posicionamento dos CRDs no modelo da galectina-4: B) visão frontal e C) visão do topo.

Interessante também ressaltar que 28,5% da sequência do peptídeo é composta por prolinas, o que do ponto de vista estrutural impõe rigidez ao movimento dessa região (figura 57).

Figura 57 – Representação em *cartoon* da estrutura do peptídeo de ligação da galectina-4 evidenciando em *sticks* as prolinas presentes nessa região.

O papel dessas prolinas para a rigidez da conformação do peptídeo e consequente relação com a afinidade exibida pela hGal4 pelos carboidratos poderá ser investigada através de estudos de mutagênese sítio-dirigida.

CONCLUSÕES

A galectina-4 e seus domínios, hGal4-CRD1 e hGal4-CRD2, foram clonados, expressos e purificados com sucesso através do uso de cromatografia por afinidade em coluna de níquel e exclusão por tamanho. O rendimento médio da hGal4 geralmente está em torno 2 mg por litro de cultura, seguido de aproximadamente 4 mg para a hGal4-CRD1. Por outro lado, a hGal4-CRD2 se mostrou bastante solúvel nas mesmas condições de purificação, resultando em um rendimento de aproximadamente 45 mg por litro de cultura.

Ambos domínios hGal4-CRD1 e hGal4-CRD2 foram cristalizados com sucesso. Enquanto os cristais da hGal4-CRD1 difratam a alta resolução, os cristais da hGal4-CRD2 não difratam regularmente e, apesar da aparente morfologia externa adequada, o padrão de difração sugeria a presença de mais um cristal adjunto na amostra.

Um outro problema durante a coleta de dados para os cristais de hGal4-CRD2 é a alta concentração do agente precipitante, 1,8 M de sulfato de amônio. Uma vez que as gotas de cristalização são abertas para a coleta do cristal, o sulfato de amônio presente na solução começa a cristalizar rapidamente. Além disso, muitos cristais estavam ou presos a película que forma na superfície da gota ou aderidos no poço, tornando o manuseio dos mesmos mais demorado. A demora no manuseio dos cristais nessas condições pode ser um fator limitante na obtenção de cristais com qualidade.

Para minimizar o efeito da formação de cristais de sulfato de amônio, a solução crioprotetora foi adicionada diretamente a gota de cristalização o que ocasionou a formação de correntes de convecção acentuadas. Todos esses fatores podem ter contribuído para a obtenção de dados com qualidade inferior aos dados da hGal4-CRD1. A estrutura final apresenta várias regiões flexíveis, dificultando a construção do modelo. A baixa qualidade dos dados, associada a mobilidade de diversas regiões da proteína, podem ser responsáveis pela diferença de aproximadamente 8% entre os valores de R_{free} e R_{factor} na etapa atual de refinamento. O uso de pesos adequados para os dados experimentais e geométricos pode ser essencial no refinamento final da estrutura.

No caso do domínio hGal4-CRD1, a principal dificuldade encontrada foi a obtenção das estruturas dos complexos com ligantes. Como já demonstrado, devido ao empacotamento cristalino, o sítio de ligação ao carboidrato está parcialmente obstruído por moléculas simétricas. A busca de uma nova condição de cristalização na presença de lactose está em andamento. Entretanto, dados na literatura mostram, para as galectinas no geral, que a multivalência tanto da proteína quanto do ligante, é essencial para a avidéz na ligação. Assim, a co-cristalização com ligantes de baixa afinidade como a lactose pode ser um desafio.

Em relação à estrutura cristalográfica dos domínios, vimos que o enovelamento global é bastante similar, como é o esperado para galectinas do grupo *tandem-repeat*. A distribuição de cargas na superfície de potencial dos domínios mostrou-se a maior divergência entre eles até o momento e pode ser a chave para o entendimento das diferenças de afinidades entre eles.

Ensaio de cristalização para a galectina-4 foram realizados recentemente, mas não temos nenhum indício da formação de cristais no momento. Como uma primeira abordagem para analisar a estrutura terciária da hGal4 procedemos com a construção de um modelo a partir das estruturas dos domínios separados. O maior desafio foi construir um modelo para o peptídeo de ligação, já que não existe no banco de dados nenhuma estrutura para uma galectina do grupo *tandem-repeat* com o peptídeo de ligação íntegro.

Na verdade, devido à alta suscetibilidade à proteólise e a flexibilidade imposta pelo peptídeo, a estrutura da galectina-8 (do grupo *tandem-repeat*) foi resolvida em uma forma truncada, apenas com dois resíduos de aminoácidos compondo o peptídeo de ligação. Desse modo, utilizamos um servidor *online*, Robetta, para a construção de modelos para o peptídeo de ligação da hGal4. Como resultado obtivemos vários modelos, sendo os melhores escolhidos para dar início a modelagem da galectina-4. O critério de escolha foi baseado na análise da geometria dos ângulos e das ligações entre os átomos, a quantidade de *clashes* e a análise do gráfico de Ramachandran. Assim, apenas dois dos modelos foram aceitáveis para o passo seguinte.

Com as estruturas cristalográficas dos domínios hGal4-CRD1 e hGal4-CRD2 e o modelo do peptídeo de ligação, foi possível construir o modelo para a galectina-4, que passou pelo mesmo critério de análise geométrica. O modelo da hGal4 foi submetido a dinâmica molecular, a fim de encontrarmos um mínimo de energia para o posicionamento correto dos CRDs. Foi visto que a interação entre os CRDs foi mais fortemente representada pela ligação indireta CRD1-peptídeo-CRD2. Pelas análises de interação de hidrogênio ao longo da dinâmica foi observado que as moléculas de LacNAc estabeleceram interações com diferentes resíduos no domínio hGal4-CRD1 e hGal4-CRD2. Essas diferenças podem ser devido a modificações sutis na sequência e, consequentemente, na distribuição de cargas na superfície da proteína que conferem especificidade e afinidade diferenciada para os domínios hGal4-CRD1 e hGal4-CRD2.

Os ensaios de estabilidade térmica, utilizando a técnica de termofluor, foram realizados para a hGal4, hGal4-CRD1 e hGal4-CRD2. Em primeira instância, utilizamos um kit contendo diferentes aditivos para testar o comportamento das três proteínas. A análise da curva de fluorescência das proteínas em tampão mostrou que a hGal4 possui a menor temperatura de desenovelamento, seguido da hGal4-CRD1 e depois da hGal4-CRD2. A temperatura da proteína inteira e do domínio N-terminal são bastante próximas, diferindo apenas 1,4 °C. Mas a comparação

com o domínio C-terminal mostrou diferenças de 16,04 °C e 14,64 °C, respectivamente. A galectina-4 mostrou aumento na estabilidade térmica na presença de um maior número de aditivos do que o domínio N- e C-terminal.

Duas moléculas foram encontradas como potenciais ligantes para a galectina-4, entretanto, experimentos de titulação das mesmas deverão ser realizados a fim de validá-las como ligantes. Uma vez validadas, a titulação será feita para os CRDs separadamente para definir as especificidades de ligação e correlacionar com a estrutura, por meio de abordagens computacionais.

CAPÍTULO 3

GALECTINA-12

INTRODUÇÃO

A galectina-12 humana é codificada pelo gene *LGALS12* e pode sofrer *splicing* alternativo resultando em diferentes isoformas. É pertence ao subgrupo das galectinas *tandem-repeat*, portanto, possui dois domínios de reconhecimento de carboidratos (CRD1 no N-terminal e CRD2 no C-terminal) conectados por um peptídeo de ligação de tamanho variável (Yang *et al.*, 2001) (figura 58).

	1	10	20	30	40	50	60	70	80				
isoforma1	MSQPSGGRAPGT	RIYWSWCPTVM	SPGEXLDP	IPDSFILQ	PPVFH	PVVPYVT	TFGGLHAGK	MVMLQG	VVPVLD	DAH	RFQVD		
isoforma2	MSQPSGGRAPGT	RIYWSWCPTVM	SPGEXLDP	IPDSFILQ	PPVFH	PVVPYVT	TFGGLHAGK	MVMLQG	VVPVLD	DAH	RFQVD		
isoforma3	MSQPSGGRAPGT	RIYWSWCPTVM	SPGEXLDP	IPDSFILQ	PPVFH	PVVPYVT	TFGGLHAGK	MVMLQG	VVPVLD	DAH	RFQVD		
isoforma4								MVMLQG	VVPVLD	DAH	RFQVD		
isoforma5								MVMLQG	VVPVLD	DAH	RFQVD		
	90	100	110	120	130	140	150	160					
isoforma1	FQCGCSLCP	RPDIAFHFN	PRFHTTK	PHVICNTL	HGGRWQ	REARWPH	LALRRGSS	FLILFL	FGNEEV	KVSVNG	QHFLHFRY		
isoforma2	FQCGCSLCP	RPDIAFHFN	PRFHTTK	PHVICNTL	HGGRWQ	REARWPH	LALRRGSS	FLILFL	FGNEEV	KVSVNG	QHFLHFRY		
isoforma3	FQCGCSLCP	RPDIAFHFN	PRFHTTK	PHVICNTL	HGGRWQ	REARWPH	LALRRGSS	FLILFL	FGNEEV	KVSVNG	QHFLHFRY		
isoforma4	FQCGCSLCP	RPDIAFHFN	PRFHTTK	PHVICNTL	HGGRWQ	REARWPH	LALRRGSS	FLILFL	FGNEEV	KVSVNG	QHFLHFRY		
isoforma5	FQCGCSLCP	RPDIAFHFN	PRFHTTK	PHVICNTL	HGGRWQ	REARWPH	LALRRGSS	FLILFL	FGNEEV	KVSVNG	QHFLHFRY		
	170	180	190	200	210	220	230	240					
isoforma1	RLPLSHVD	TLGIFG	DILVEAV	GF	LNINPF	VEGSREY	PAGHPFL	LMSPRIE	VP	CSHALP	QGLSPG	QVIVRGL	VLPQEPKHF
isoforma2	RLPLSHVD	TLGIFG	DILVEAV	GF	LNINPF	VEGSREY	PAGHPFL	LMSPRIE	VP	CSHALP	QGLSPG	QVIVRGL	VLPQEPKHF
isoforma3	RLPLSHVD	TLGIFG	DILVEAV	GF	LNINPF	VEGSREY	PAGH.....		VP	CSHALP	QGLSPG	QVIVRGL	VLPQEPKHF
isoforma4	RLPLSHVD	TLGIFG	DILVEAV	GF	LNINPF	VEGSREY	PAGHPFL	LMSPRIE	VP	CSHALP	QGLSPG	QVIVRGL	VLPQEPKHF
isoforma5	RLPLSHVD	TLGIFG	DILVEAV	GF	LNINPF	VEGSREY	PAGH.....		VP	CSHALP	QGLSPG	QVIVRGL	VLPQEPKHF
	250	260	270	280	290	300	310	320					
isoforma1	TVSLRDQAA	HAPVTL	RASFAD	RTLAWIS	RWGQKK	LISAPFL	FYPQRFF	EVLLLF	QEGGLK	LALNGQ	GLGAT	SMNQ	QALEQ
isoforma2	TVSLRDQAA	HAPVTL	RASFAD	RTLAWIS	RWGQKK	LISAPFL	FYPQRFF	EVLLLF	QEGGLK	LALNGQ	GLGAT	SMNQ	QALEQ
isoforma3	TVSLRDQAA	HAPVTL	RASFAD	RTLAWIS	RWGQKK	LISAPFL	FYPQRFF	EVLLLF	QEGGLK	LALNGQ	GLGAT	SMNQ	QALEQ
isoforma4	TVSLRDQAA	HAPVTL	RASFAD	RTLAWIS	RWGQKK	LISAPFL	FYPQRFF	EVLLLF	QEGGLK	LALNGQ	GLGAT	SMNQ	QALEQ
isoforma5	TVSLRDQAA	HAPVTL	RASFAD	RTLAWIS	RWGQKK	LISAPFL	FYPQRFF	EVLLLF	QEGGLK	LALNGQ	GLGAT	SMNQ	QALEQ
	330												
isoforma1	LRELRI	SGSVQ	LYCVHS										
isoforma2	LRELRI	SGSVQ	LYCVHS										
isoforma3	LRELRI	SGSVQ	LYCVHS										
isoforma4	LRELRI	SGSVQ	LYCVHS										
isoforma5	LRELRI	SGSVQ	LYCVHS										

Figura 58 – Alinhamento das isoformas 1, 2, 3, 4 e 5 da galectina-12 humana. A sequência em azul claro, dos resíduos 1 ao 183 representam o domínio N-terminal; a sequência marcada em azul escuro, dos resíduos 211 ao 337 representam o domínio C-terminal e dos resíduos 184 ao 210 formam o peptídeo de ligação dos domínios, em preto.

Com base no alinhamento sequencial das isoformas foi possível denominar os diferentes CRDs como sendo o CRD1a (domínio N-terminal para a isoforma 1), CRD1b (domínio N-terminal comum as isoformas 2 e 3), CRD1c (domínio N-terminal comum as isoformas 4 e 5) e CRD2 (domínio C-terminal), comum a todas as isoformas (figura 59).

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

1 10 20 30

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

40 50 60 70 80 90 100

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

110 120 130 140 150

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

160 170 180 190 200

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

210 220 230 240 250 260 270

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

280 290 300 310 320

galectin4
galectin6
galectin8
galectin9
galectin12a
galectin12d

Figura 60 - Alinhamento das sequências dos membros da família *tandem-repeat* (galectinas 4, 6, 8 e 9 humanas) com uma isoforma longa e uma isoforma curta da galectina 12humana. Os resíduos marcados em ★ mostram os resíduos envolvidos na coordenação com o carboidrato ligante.

Uma comparação entre a galectina-12 e outros membros da família *tandem-repeat* mostra que seu domínio N-terminal apresenta uma identidade significativa em relação as outras galectinas, enquanto que a sequência do domínio C-terminal desvia significativamente da sequência consenso e muitos dos resíduos conservados não estão presentes (Liu, Patterson e Wang, 2002) (figura 60). No entanto, a proteína tem atividade de ligação à lactose, provavelmente devido ao domínio N-terminal.

Análises realizadas por *Northern blot* mostraram que o RNA mensageiro da galectina-12 foi quase indetectável em muitos dos tecidos testados, em contraste com os resultados obtidos da análise feita para o mRNA da galectina 3, detectada em quase todos os tecidos (Liu, Patterson e Wang, 2002). A aplicação da técnica de PCR em tempo real detectou RNA mensageiro da galectina-12 no coração, pâncreas, baço, timo e leucócitos periféricos. Em menor nível foi detectado no pulmão, músculo esquelético, rim, próstata, testículo, ovário e cólon (Liu, Patterson e Wang, 2002). A Gal12 não foi detectada em muitas linhagens de células testadas, mas a sua expressão foi confirmada em células periféricas mono e polinucleares, em células mielóides (U937, HL-60 e KU-812), em células B Wil-2 e em linhagens de células de câncer de mama HBL-100 (Yang *et al.*, 2001).

O mRNA da galectina-12 contém funcionalidades que codificam proteínas com funções reguladoras do crescimento. Estes incluem um códon de iniciação em um contexto de sequência que é sub-ótima para iniciação da tradução e motivos AU-ricos da região 3'-não traduzida, que conferem instabilidade ao mRNA; esses motivos são geralmente encontrados em mRNA de proto-oncogenes e em genes que são codificados para fatores de transcrição nucleares e citocinas (Liu, Patterson e Wang, 2002).

Estudos realizados por Hotta e colaboradores (2001) mostraram que a galectina-12 possui propriedades indutoras de apoptose, tendo sua expressão aumentada concomitante ao aumento da apoptose em células de gordura (Hotta *et al.*, 2001). Além disso, a galectina-12 tem um importante papel na regulação e parada do ciclo celular, levando a considerar a hipótese de que a galectina-12 pode atuar na tumorigênese. O gene que codifica a galectina-12 é encontrado em várias regiões do cromossomo 11q13 humano, que frequentemente é alterado em muitos tipos de cânceres (Yang *et al.*, 2001).

A habilidade da galectina-12 paralisar o ciclo celular e inibir a proliferação de várias linhagens de células cancerígenas a torna uma candidata a supressora de tumor no cromossomo em questão, o que sugere que a terapia gênica usando a galectina-12 pode vir a ser uma ferramenta valiosa para o tratamento de câncer (Yang *et al.*, 2001).

Estudos de localização celular realizados para a galectina-12 de camundongo mostraram que essa proteína, preferencialmente expressa em adipócitos, encontra-se principalmente associada

a gotas de lipídios (Yang *et al.*, 2011). Gotas de lipídios são organelas especializadas no armazenamento de energia, consistindo de uma monocamada fosfolipídica que envolve um núcleo de lipídios neutros, com a superfície envolta por inúmeras proteínas (figura 61) (Farese e Walther, 2009; Guo *et al.*, 2009).

Figura 61 – Anatomia de uma gota de lipídio. À esquerda micrografia eletrônica de uma gota de lipídio; à direita, representação esquemática das características estruturais da gota de lipídio. Figura retirada do artigo *Lipid Droplets Finally get a little R-E-S-P-E-C-T* (Farese e Walther, 2009). Micrografia eletrônica cortesia de S. Stone e J. Wong. Imagem reimpressa de Walther and Farese (Walther e Farese, 2009).

Yang e colaboradores mostraram que a galectina-12 regula a lipólise modulando a sinalização da proteína quinase A (PKA). Seus resultados sugerem que a galectina-12 pode ser alvo para o tratamento de condições metabólicas associadas a obesidade como diabetes tipo 2 (Yang, Havel e Liu, 2012).

Diante desse quadro, há muitas perguntas no que diz respeito às funções das diferentes isoformas da galectina-12 em outros tipos celulares, principalmente em humanos onde a expressão não está restrita a adipócitos. Investigações adicionais são necessárias para elucidar a função da galectina-12 *in vivo*. Portanto, a caracterização estrutural e funcional dos CRDs da galectina-12 é um passo fundamental no entendimento dos mecanismos de ação dessa proteína nos diversos tecidos, na regulação do ciclo celular e na apoptose de células adiposas.

OBJETIVOS

Diante da relevância biológica da galectina-12 na regulação da lipólise e no metabolismo energético, e de poder ser um alvo no gerenciamento de condições metabólicas relacionadas a obesidade, como resistência à insulina e diabetes, o desenvolvimento do presente projeto teve por objetivo contribuir para a elucidação das bases estruturais e funcionais que se correlacionam à ação da galectina-12 humana nos diferentes contextos celulares. Nosso trabalho tem como foco a caracterização funcional e estrutural das isoformas e domínios N- e C-terminais da galectina-12 humana (hGal12), utilizando abordagens computacionais como modelagem molecular e de imunodeteção nos estudos de localização celular.

MATERIAIS E MÉTODOS

3.1 Obtenção do DNA molde

3.1.1 Construção do cDNA

A primeira etapa na produção das isoformas recombinantes da hGal12 consistiu na construção do cDNA a partir de leucócitos totais do sangue periférico humano. O projeto em questão foi aprovado pelo Comitê de Ética em Pesquisa da FCFRP/USP (Protocolo CEP/FCFRP nº 133). O procedimento de separação dos leucócitos foi realizado por aplicação de um gradiente de densidade descontínuo utilizando as soluções HISTOPAQUE® 1077 e 1119 (Sigma-Aldrich). Em um tubo falcon foram adicionados, nessa ordem, 3 ml de Histopaque®-1077, 3 ml de Histopaque®-1119 e 6 ml de amostra de sangue periférico. O material foi centrifugado por 35 minutos, 22 °C a 700g, segundo as instruções do fabricante.

Os eritrócitos formam um precipitado no fundo do tubo, enquanto as células polimorfonucleares formam uma banda na interface entre as soluções Histopaque®-1119 e Histopaque®-1077; já as células mononucleares formam uma banda na interface entre o plasma e a solução Histopaque®-1077. As camadas referentes aos leucócitos totais (células mononucleares e polimorfonucleares) foram separadas por aspiração, transferidas para um tubo de centrífuga de 50 ml no qual foram lavadas em solução tampão PBS 1X e centrifugadas a 200g por 10 minutos. O precipitado de células foi ressuspenso em 5 ml de solução tampão PBS 1X.

3.1.1.1 Purificação do RNA total e transcrição

As células foram contadas adicionando 10 µl da suspensão em 190 µl de solução de Türk. A contagem foi realizada em uma câmara de Neubauer e as células foram utilizadas para a purificação do RNA total pela utilização do kit *Rneasy Plus* (QIAGEN), segundo as instruções do fabricante. Nesse procedimento, todas as moléculas de RNA maiores do que 200 nucleotídeos foram purificadas (18S rRNA, 28S rRNA e mRNA), excluindo seletivamente de 15-20% do RNA total que é composto por 5,8S rRNA, 5S rRNA e tRNAs implicando em uma amostra rica em RNA mensageiro.

As amostras eluídas foram visualizadas por eletroforese em gel de agarose 1% (m/v) em condições não desnaturantes. A integridade da amostra de RNA foi verificada pela visualização sob luz UV das bandas 18S e 28S do rRNA. Em seguida, a amostra foi quantificada em um

espectrofotômetro a 260 nm (Nanodrop 2000) e utilizada como molde para a síntese da primeira fita de cDNA, utilizando a enzima *RevertAidTM H Minus Reverse Transcriptase* (Fermentas).

A reação foi preparada como recomendado no manual da enzima; uma quantidade entre 0,1 ng e 5 µg de RNA molde foi incubada a 65 °C por 5 minutos com 100 pmol de oligo(dT). Em seguida foram adicionados 1 mM de dNTP, 200 U da enzima transcriptase reversa e tampão de reação (1X) para um volume final de 20 µl. A reação foi incubada a 42 °C por 60 minutos e terminada por aquecimento a 70 °C durante 10 minutos. O cDNA foi purificado utilizando o sistema *QLAquick® PCR Purification kit* (QIAGEN). O material resultante foi quantificado a 260 nm (Nanodrop 2000) e armazenado a -20 °C.

3.1.2 Otimização de códons

Tendo em vista a possibilidade da existência de um grande número de códons raros no gene de interesse, as sequências que codificam as isoformas da hGal12 foram submetidas a uma análise da quantidade e frequência de códons pela utilização dos *softwares Rare Codon Caltor* (<http://people.mbi.ucla.edu/sumchan/caltor.html>) e *Rare Codon Calculator – RaCC* (<http://nihserver.mbi.ucla.edu/RACC/>).

Desse modo, a sequência que codifica a Isoforma 2 e o CRD1 das isoformas 4 e 5 (CRD1c) foram submetidas ao um processo de otimização de códons, no qual os raros foram substituídos pelos mais frequentes em *E. coli*. As sequências foram sintetizadas e clonadas pela IDT no vetor pIDTSmart. Com a obtenção da sequência otimizada da isoforma 2, foi possível clonar as demais isoformas e CRDs.

3.2 Clonagem

Para a clonagem das isoformas da hGal12 e seus domínios N- e C-terminal foram utilizados como molde o cDNA obtido de leucócitos totais humanos e as sequências otimizadas para *E.coli* da Isoforma 2 e do domínio CRD1c. Além disso, diferentes estratégias de expressão foram testadas nas quais os vetores pTZ57R/T, pJET1.2/blunt, pTYB2, pET-28a(+), pET-29a(+) e pET-28a(+)-sumo foram utilizados. Uma breve descrição de cada vetor é dada a seguir.

3.2.1 Vetores de clonagem e expressão

3.2.1.1 Vetor de clonagem pTZ57R/T (*Thermo Scientific*)

O vetor pTZ57R/T é um plasmídeo linear que possui nucleotídeos de ddT (deoxitimina) em suas extremidades 3' que permite a inserção direta de produtos de PCR, previamente adenilados, por uma de uma reação de ligação. Em sua sequência, o vetor pTZ57R/T possui o gene *lacZ* que na presença de agente indutor da expressão, IPTG (Isopropil β-D-1-tiogalactopiranosídeo), codifica a enzima β-galactosidase, que por sua vez, degrada o substrato X-GAL formando um produto colorido de cor azul. A inserção correta do fragmento de interesse no vetor de clonagem interrompe a sequência do gene *lacZ* e, como resultado, não há expressão da enzima β-galactosidase. Assim, na presença de IPTG e X-GAL, as colônias de bactérias que possuírem o plasmídeo com a sequência *lacZ* intacta produzirá colônias azuis, enquanto que as colônias que possuírem o inserto interrompendo a sequência *lacZ* produzirá colônias brancas, que indicam o sucesso da clonagem.

3.2.1.2 Vetor de clonagem pJET1.2/blunt (*Thermo Scientific*)

Produtos de PCR com extremidades abruptas geradas pela *Pfu* DNA polimerase podem ser ligados diretamente ao vetor de clonagem pJET1.2/blunt. O vetor pJET1.2/blunt contém um gene letal que é interrompido pela ligação do inserto no sítio de clonagem. Como resultado, somente células com o plasmídeo recombinante são propagadas. O vetor contém um replicon, responsável pela replicação do pJET1.2, uma origem de replicação, uma sequência que confere resistência a ampicilina, o gene letal *eco47IR*, promotor P_{lac} modificado para promover a expressão do gene *eco47IR*, um promotor T7 para transcrição *in vitro*, um sítio múltiplo de clonagem e um sítio de inserção do DNA abrupto.

3.2.1.3 Vetor de expressão pTYB2 - NEB

Plasmídeos do sistema IMPACT (*Intein Mediated Purification with an Affinity Chitin-binding Tag* - NEB) controlam a expressão do gene de interesse utilizando elementos chave como o promotor T7, a proteína repressora *lac*, o operador *lac*, uma origem de replicação de DNA, um sítio de policlonagem e a sequência do gene *Ampr* que confere à bactéria hospedeira resistência ao antibiótico ampicilina. Nesse sistema, a proteína alvo é fusionada a um elemento de *splicing* modificado, inteína, e a um domínio de ligação à quitina que permite a purificação por afinidade

do precursor da fusão na coluna de quitina. O sistema IMPACT-CN contém vetores de expressão PTYB que permitem a fusão da cauda tanto no C-terminal (pTYB1 e pTYB2) quanto no N-terminal (pTYB11 e pTYB12) da proteína alvo.

3.2.1.4 Vetores de expressão pET-28a(+) e pET-29a(+) (Novagen)

Os principais elementos que compõem o plasmídeo são o gene *lacI* que codifica a proteína repressora *lac*, o promotor T7 específico para T7 RNA polimerase, um operador *lac*, um sítio de policlonagem, uma origem f1 de replicação e uma sequência que confere resistência à canamicina. Os vetores pET-28a(+) e pET-29a(+) permitem a clonagem da proteína alvo em fusão com uma cauda de histidina, no N- ou C-terminal da proteína.

3.2.1.5 Vetor de expressão pET-28a(+)-sumo

O vetor pET-28a(+)-sumo, gentilmente cedido pelo Dr. André Ambrósio, consiste em uma modificação no vetor pET-28a(+) pela inserção da proteína Smt3 de 11 kDa, ubiquitina de *Saccharomyces cerevisiae* (UniProtKB/Swiss-prot: Q12306.1) clonada entre os sítios de restrição *NheI* e *BamHI*. A proteína Smt3, referida aqui como SUMO, é um membro da família ubiquitin-*like*. Essa construção permite a clonagem da proteína alvo em fusão com a SUMO no N-terminal (figura 62). O processo de clivagem da cauda é realizado com a SUMO-protease ULP1 (*Ubiquitin-like-specific Protease 1* – EC 3.4.22.68) de *Saccharomyces cerevisiae* (UniProtKB/Swiss-Prot: Q02724.1), também clonada no vetor pET-28a(+). A clivagem ocorre pelo reconhecimento do motivo Gly-Gly no C-terminal da SUMO (Mossessova e Lima, 2000), resultando na liberação da proteína apenas com os resíduos referentes ao sítio de restrição inseridos entre a SUMO e a proteína de interesse.

Figura 62 – Representação esquemática da construção resultante da clonagem no vetor pET-28a-sumo.

3.2.1.6 Vetor de transferência pFast-His (*Life Technologies*)

O vetor de transferência pFast-His permite a expressão controlada pelo promotor da poliedrina (P_H) do vírus da poliedrose múltipla de *Autographa californica* – AcMNPV. O vetor possui

a sequência para a expressão de uma cauda contendo seis histidinas, um sítio de trombina seguido pelo sítio de clonagem múltiplo. Outros componentes presentes na sequência incluem um sinal de poliadenilação, transposons Tn7 flanqueando o cassete de expressão, uma origem de replicação e as sequências que codificam genes de resistência à gentamicina e ampicilina.

3.2.2 Clonagem das isoformas e domínios da hGal12

A próxima etapa consistiu em clonar os fragmentos de DNA que codificam as isoformas e os CRDs da hGal12 através da amplificação das sequências de interesse. Como primeiro passo dessa etapa, procedemos com o planejamento dos oligonucleotídeos iniciadores complementares às extremidades 3' e 5' do fragmento de DNA de interesse. A estratégia de planejamento dos oligonucleotídeos foi baseada nas diferenças observadas entre as isoformas (figura 63). Essas diferenças se localizam principalmente na região N-terminal das proteínas e na região onde se localiza o peptídeo de ligação.

Figura 63 – Representação esquemática do alinhamento entre as 5 isoformas da galectina-12. Em azul claro a região que representa o CRD1; em rosa, a região que representa o peptídeo de ligação e em azul escuro, a região que representa o CRD2. A região hachurada representa o lugar na sequência em que cada primer foi baseado.

Durante o desenvolvimento desse projeto, diferentes estratégias de expressão foram abordadas, assim, foi necessário a realização de várias etapas de clonagem. Todos os oligonucleotídeos utilizados nesse projeto foram desenhados manualmente com base nas respectivas sequências de interesse e sintetizados pelas empresas IDT (Integrated DNA Technologies) e Exxtend. Uma sequência referente ao sítio de restrição utilizado foi adicionada em cada oligo seguido de uma sequência de ancoramento. A temperatura de anelamento para cada oligonucleotídeo foi calculada com base na expressão, $T_m = 60 + (41N_{CG} - 500)/L$, onde N_{CG} representa o número de nucleotídeos guanina e citosina; L é o número total de nucleotídeos utilizados no desenho dos oligonucleotídeos (tabela 16).

Tabela 16 – Oligonucleotídeos utilizados no processo de clonagem dos fragmentos de DNA codificantes das isoformas e CRDs da galectina-12 humana. Em parênteses estão as sequências referentes aos sítios de restrição.

Oligonucleotídeos para clonagem em pTYB2 e pET-28a(+)

<i>Forward</i> CL	GACGACGAC(CATATG)AGTCAGCCCAGTGGG
<i>Forward</i> CC	GACGACGAC(CATATG)GTCATGCTGCAAGGAGTG
<i>Forward</i> CRD2	GACGACGAC(CATATG)AGCCCCAGGCTG
<i>Reverse</i>	GACGAC(CTCGAG)GGAGTGGACACAGTAGAGCT
<i>Reverse</i> CRD1	GACGAC(CTCGAG)CAGGAATCCAAGAG

Oligonucleotídeos otimizados clonagem em pET-28a(+) e pET-29a(+)

<i>oForward</i> CL	GACGACGAC(CATATG)TCTCAGCCGTCTGG
<i>oForward</i> CC	GACGACGAC(CATATG)GTAATGCTGCAGGGC
<i>oForward</i> CRD2	GACGACGAC(CATATG)TCCCCTCGCCTC
<i>oReverse</i> CRD1	GACGAC(CTCGAG)AAGGAAGCCAACTG
<i>oReverse</i>	GACGAC(CTCGAG)AGAATGGACACAGTAC

Oligonucleotídeos otimizados clonagem em pET-28a(+)-sumo

<i>opForward</i> CL	GACGAC(GAATTC)ATGTCTCAGCCGTCTGGC
<i>opForward</i> CC	GACGAC(GAATTC)ATGGTAATGCTGCAGGGC
<i>opForward</i> CRD2	GACGAC(GAATTC)ATGTCCCCTCGCCTCGAA
<i>opReverse</i> CRD1	GACGAC(CTCGAG)TCAAAGGAAGCCAACTG
<i>opReverse</i>	GACGAC(CTCGAG)TCAAGAATGGACACAGTAC
<i>opForward</i> Iso1	CTGGACGCTCACTCTCGCTTCCAAGTT
<i>opReverse</i> Iso1	AAC TTGGAAGCGAGAGTGAGCGTCCAG
<i>opForward</i> -peptídeo	GAATACCCAGCCGGCCATGAAGTCCCC
<i>opReverse</i> -peptídeo	GTGAGAACAGGGGACTTCATGGCCGGC

Oligonucleotídeos para clonagem em pFast

<i>Forward</i> CL-Bac	GACGAC(GAATTC)ATGAGTCAGCCCAGTGG
<i>Forward</i> CC-Bac	GACGAC(GAATTC)ATGGTCATGCTGCAAGGAG
<i>Forward</i> CRD2-bac	GACGACGAATTCATGAGCCCCAGGCTG
<i>Reverse</i> CRD1-bac	GACGACGGGCCCTCACAGGAATCCAACAGCC
<i>Reverse</i> -bac	GACGACGGGCCCTCAGGAGTGGACACAGTA

As etapas de clonagem procederam pela amplificação do gene de interesse, ligação no vetor de clonagem pTZ57R/T ou pJET1.2/blunt, digestão e ligação no vetor de expressão ou apenas pela amplificação, digestão e ligação no vetor de expressão, como descrito a seguir (salvo algumas

exceções que serão descritas particularmente). As reações e as condições dos ciclos de amplificação, temperatura de desnaturação e extensão, assim como os respectivos tempos, foram utilizados segundo as instruções nos manuais das enzimas Taq DNA polimerase utilizadas no decorrer do projeto: Taq DNA recombinante (ThermoScientific), Platinum® Taq DNA polimerase High Fidelity (Invitrogen™), *Pfu* Turbo® DNA polimerase (Stratagene) e Phusion® High Fidelity DNA polimerase (NEB). As enzimas de restrição utilizadas foram *Nde*I, *Eco*RI, *Eco*RI-HF®, *Apa*I e *Xho*I, todas adquiridas da New England Biolabs – NEB.

A amplificação foi realizada pela aplicação da técnica de reação em cadeia da polimerase, PCR, em um termociclador (Mastercycler gradiente – Eppendorf), utilizando de 0,2 a 0,4 µM de cada oligonucleotídeo, 5 ng/µl de DNA molde, 0,2 mM dNTPmix (Invitrogen™), 1 X tampão da enzima, 1 U da enzima Taq DNA polimerase em 50 µl de reação. A reação foi iniciada por uma desnaturação inicial, seguida de 25 a 30 ciclos de desnaturação, anelamento (60 °C) e extensão, com extensão final por 5 a 10 minutos. Os produtos das reações de amplificação foram analisados por visualização em luz UV após migração em gel de agarose 1% (m/v). Após a confirmação de que as sequências foram amplificadas no tamanho esperado, as reações de PCR foram purificadas utilizando o kit QIAquick PCR Purification (QIAGEN), segundo as instruções do fabricante.

Em seguida, tanto o material amplificado quanto o vetor de expressão foram digeridos com as enzimas de restrição em um volume de reação de 50 µl contendo 20 U de cada enzima, 1 X tampão da enzima e 1 µg total de amostra, a 37 °C por 30 minutos. As reações foram submetidas a eletroforese em gel de agarose 1% (m/v), excisadas e purificadas do gel utilizando o kit QIAquick Gel Purification (QIAGEN), em acordo com o manual. Os fragmentos digeridos foram ligados através da reação de ligação contendo 1 µl da enzima T4 DNA ligase (NEB), 1 X o tampão da enzima (NEB) em 20 µl de volume de reação, na qual as proporções entre inserto e vetor variaram de 3:1, 6:1 a 9:1. As reações foram conduzidas a 22 °C, por 20 minutos.

Uma alíquota da reação de ligação foi utilizada para transformar bactérias competentes (Davis, 1993) por meio de choque térmico (Sambrook, Fritsch e Maniatis, 1989). As células foram semeadas em meio LB sólido (Acumedia) previamente homogeneizado com os devidos antibióticos, em uma placa de Petri, e incubadas em uma estufa bacteriológica a 37 °C por 16 horas. As colônias selecionadas foram utilizadas como molde em uma reação de PCR e inoculadas em meio LB. Os clones positivos selecionados foram propagados, os DNAs plasmidiais extraídos, quantificados a 260 nm e submetidos a digestão com as enzimas de restrição específicas a fim de confirmar a presença do inserto. O DNA de interesse foi então submetido ao sequenciamento automático no equipamento ABI 3130xL (*Applied Biosystems*) pelo Grupo de Biofísica Molecular –

“Sérgio Mascarenhas”. Os clones obtidos durante essa etapa estão descritos na tabela a seguir (tabela 17).

Tabela 17 – Relação de clones obtidos durante o processo de clonagem das isoformas e domínios da hGal12 e seus respectivos vetores e sistema de expressão.

Construção	Isoforma/domínio	Vetor	Sistema de expressão
Clonagem a partir do cDNA			
CL6-10-pTYB2	Isoforma 1	pTYB2	Bactéria
CL10-9-pTYB2	Isoforma 2	pTYB2	Bactéria
CC3-3-pTYB2	Isoforma 4	pTYB2	Bactéria
CRD1a-9-pTYB2	CRD1a	pTYB2	Bactéria
CRD1b-1-pTYB2	CRD1b	pTYB2	Bactéria
CRD1c-1-pTYB2	CRD1c	pTYB2	Bactéria
CRD2-10-pTYB2	CRD2	pTYB2	Bactéria
CRD1a9-2.6-pET28	CRD1a	pET-28a(+)	Bactéria
Iso1-7-pFast	Isoforma 1	pFast-His	Inseto
Iso2-5-pFast	Isoforma 2	pFast-His	Inseto
Iso4-5-pFast	Isoforma 4	pFast-His	Inseto
CRD1a-2-pFast	CRD1a	pFast-His	Inseto
CRD1b-5-pFast	CRD1b	pFast-His	Inseto
CRD1c-1-pFast	CRD1c	pFast-His	Inseto
CRD2-4-pFast	CRD2	pFast-His	Inseto
Clonagem a partir do DNA otimizado			
CRD1c-2.2.5-6-pTYB2	CRD1c	pTYB2	Bactéria
CRD1c-9-pET28	CRD1c	pET-28a(+)	Bactéria
CRD1c-3-pET29	CRD1c	pET-29a(+)	Bactéria
Iso2-2-pET29	Isoforma 2	pET-29a(+)	Bactéria
CRD1b-1-pET29	CRD1b	pET-29a(+)	Bactéria
CRD2-pET29	CRD2	pET-29a(+)	Bactéria
CRD1c-1-pETsumo	CRD1c	pET-28a-sumo	Bactéria
Iso1-2-pETsumo	Isoforma 1	pET-28a-sumo	Bactéria
Iso2-3-pETsumo	Isoforma 2	pET-28a-sumo	Bactéria
Iso3-1-pETsumo	Isoforma 3	pET-28a-sumo	Bactéria
Iso4-1-pETsumo	Isoforma 4	pET-28a-sumo	Bactéria
CRD1b-1-pETsumo	CRD1b	pET-28a-sumo	Bactéria
CRD2-5-pETsumo	CRD2	pET-28a-sumo	Bactéria

3.2.2.1 Produção do bacmídeo recombinante

O método utilizado para a obtenção do baculovírus recombinante foi o de transposição, que se baseia na transposição específica de um cassete de expressão dentro do bacmídeo propagado em *E. coli* (Luckow *et al.*, 1993; Ciccarone, Polayes e Luckow, 1997). Para proceder com a produção do vírus recombinante, após a clonagem no vetor de transferência pFast-His, foi necessário utilizar cepas de *E. coli* DH10bacTM, que possuem um bacmídeo e um plasmídeo auxiliar.

Bacmídeo

O vetor de transporte do baculovírus, o bacmídeo bMON14272 (136 kb), contém um baixo número de cópias do replicon mini-F, um marcador de resistência a canamicina e um segmento de DNA que codifica o peptídeo *lacZ* a partir de um vetor de clonagem baseado no pUC, dentro do qual o sítio de ligação para o transposon bacteriano, Tn7 (mini-*att*Tn7), foi inserido (manual Bac-to-Bac Expression System – Invitrogen™). O gene *lacZ* do bacmídeo funciona como gene repórter que na presença do agente indutor da expressão, IPTG, codifica a enzima β -galactosidase que por sua vez degrada o substrato X-Gal formando um produto colorido de cor azul. A transposição do fragmento de interesse para o bacmídeo interrompe a sequência desse gene e, como resultado, na presença de IPTG e X-GAL as colônias de bactérias que possuem o inserto serão brancas, indicando o sucesso da clonagem.

Plasmídeo auxiliar

O plasmídeo auxiliar, pMON7124 (13.2 kb), contém o gene que codifica a transposase, enzima responsável pela transposição do cassete de expressão do vetor pFast-His para o bacmídeo, nas regiões Tn7 de transposição. Além disso, o plasmídeo auxiliar contém o gene que confere resistência a tetraciclina.

Método de transposição

Os plasmídeos recombinantes Iso1-7-pFast, Iso2-5-pFast, Iso4-5-pFast, CRD1a-2-pFast, CRD1b-5-pFast, CRD1c-1-pFast e CRD2-4-pFast foram utilizados para transformar cepas *E. coli* DH10bac™ competentes conforme protocolo descrito para transformação na seção 3.2.2. Após o choque térmico, as células foram incubadas a 37 °C sob agitação a 180 rpm durante 4 horas, em meio de cultura contendo 50 µg/ml de canamicina, 10 µg/ml de tetraciclina e 7 µg/ml de gentamicina. As células foram diluídas nas proporções 1:10, 1:100, 1:1000 e 1:10000 e semeadas em meio LB sólido, suplementado com os antibióticos acima descritos e contendo 4 µmol de IPTG e X-GAL. As placas foram incubadas em estufa bacteriológica durante 36 horas a 37 °C. Os clones positivos foram selecionados visualmente pela presença das colônias azuis e brancas.

Para verificarmos a transposição correta do bacmídeo as colônias selecionadas foram submetidas a uma reação de PCR, conforme descrito na seção 3.2.2, com os oligonucleotídeos *forward* dos insertos e M13 *reverse*, os quais hibridizam conforme esquema a seguir (figura 64).

Figura 64 – Representação esquemática do bacmídeo recombinante na região de transposição do gene de interesse. Adaptado do manual Bac-to-Bac® *Baculovirus Expression System* (Invitrogen™).

Extração do DNA total e transfecção

Os clones em DH10bac™ contendo o bacmídeo recombinante para as construções obtidas foram propagados e o DNA total extraído. As células foram ressuspensas, lisadas, neutralizadas e centrifugadas por 10 minutos em temperatura ambiente, seguindo as instruções do manual do kit de extração de plasmídeo *Wizard® Plus SV minipreps DNA Purification System* (Promega). Ao sobrenadante foi adicionado isopropanol a uma concentração final de 70% (v/v) e as amostras foram centrifugadas a 4 °C, 16000g durante 30 minutos. O sobrenadante foi descartado e o material precipitado, lavado duas vezes com 1 ml de etanol 70% (v/v) e seco em temperatura ambiente. O precipitado, composto de DNA genômico e bacmídeo, foi ressuspendido em água mili-Q autoclavada e armazenado a 4 °C.

O processo de transfecção foi realizado para as isoformas 1, 2 e 4, adicionando 1 µl do DNA total a 100 µl de meio TC-100 (HiMedia) sem soro, em um tubo de centrifuga estéril, segundo as instruções do fabricante. Em outro tubo de centrifuga estéril 8 µl de Cellfectin® II foram diluídos em 100 µl de meio TC-100 sem soro. As soluções diluídas de DNA total e Cellfectin® foram misturadas em uma placa de cultura de seis poços e incubadas a temperatura ambiente por 30 minutos. A essa mistura foram adicionados 2 ml da suspensão de células Tn5b (originárias de células do ovário de *Trichoplusia ni*) e incubadas a 27 °C por aproximadamente 16 horas. As células foram ressuspensas e centrifugadas a 1300 rpm. O sobrenadante contendo os baculovírus recombinantes foi coletado e armazenado a 4 °C.

3.2.2.2 Clonagem das isoformas 1 e 3

Isoforma 1

A isoforma 1 possui apenas um resíduo de aminoácido, Ser75, a mais que a isoforma 2. Assim, após a otimização de códons que foi realizada para a isoforma 2 nós procedemos com a etapa de clonagem da isoforma 1 por inserção do códon para serina baseado no protocolo do kit de mutagênese *QuikChange II Site-Directed Mutagenesis* (Agilent Technologies) utilizando a enzima *Phusion® High Fidelity DNA polimerase* (NEB). O protocolo consistiu inicialmente de uma reação de amplificação do plasmídeo de interesse por PCR utilizando 0,5 µM de cada oligonucleotídeos, 96 ng DNA molde (Iso2-3-pETsumo), 0,2 mM dNTP mix, 1X tampão CG da enzima, 5% DMSO e 2U da enzima *Phusion® High Fidelity DNA polimerase*, em 50 µl de reação. A reação de PCR iniciou-se pela desnaturação inicial a 98 °C por 1 minuto, seguida de 18 ciclos de desnaturação a 98 °C por 10 s, anelamento a 70°C ± 5°C por 1 minuto e extensão a 72 °C por 2 minutos. Ao final foi realizada uma etapa de extensão a 72 °C por 5 minutos. As reações de PCR foram purificadas e submetidas a uma reação de digestão contendo 20 U da endonuclease Dpn I e 1 X tampão da enzima em 50 µl de reação, conforme instruções do fabricante. A reação de digestão foi incubada a 37 °C por 2 horas, o material foi purificado novamente e utilizado para transformar cepas de *E. coli* DH5α. As colônias selecionadas foram submetidas a reação de PCR para verificar a presença do inserto. Os clones positivos foram propagadas e tiveram seus DNAs plasmidiais extraídos para sequenciamento. O clone resultante foi denominado Iso1-2-pETsumo.

Isoforma 3

A isoforma 3 possui 9 resíduos de aminoácidos a menos que a isoforma 2, na região do peptídeo de ligação. Dessa forma, para a clonagem da isoforma 3 foi realizada mutagênese sítio-dirigida por *overlap extension* usando uma reação de PCR (Ho *et al.*, 1989; Horton *et al.*, 1989). Assim, foram preparadas duas reações de PCR contendo a isoforma 2 como molde (Iso2-3-pETsumo). Em cada reação foi utilizado um par de oligonucleotídeos, correspondendo aos oligonucleotídeos *Forward*CL e *Reverse*-peptídeo, para a amplificação do fragmento 1, e os oligonucleotídeos *Forward*-peptídeo e *Reverse*, para a amplificação do fragmento 2 (figura 65).

Figura 65 – Representação esquemática das isoformas 2 e 3. As regiões referentes ao CRD1 (azul claro), peptídeo de ligação (rosa) e CRD2 (azul escuro) podem ser identificadas na figura. Em amarelo está representado o oligonucleotídeo *Forward-peptídeo* e em laranja, o oligo *Reverse-peptídeo*, desenhados para a clonagem da isoforma 3 a partir da isoforma 2 por *overlap extension*.

As reações de PCR foram purificadas e os fragmentos 1 e 2 foram utilizados como molde em uma nova reação de PCR contendo 0,2 mM de dNTP mix, 1 X tampão da enzima, 1 U *Phusion® High Fidelity* e 500 ng de cada fragmento, em 50 µl de reação. A reação de PCR iniciou-se pela desnaturação inicial a 98 °C por 30 s, seguida de 10 ciclos de desnaturação a 98 °C por 10s, anelamento a 60 °C por 30 s e extensão a 72 °C por 15 s, para anelamento dos fragmentos complementares gerados na etapa anterior. Em seguida, os oligonucleotídeos *ForwardCL* e *Reverse* foram adicionados a uma concentração de 0,2 µM e a reação foi prolongada por mais 15 ciclos, seguidos de extensão final a 72 °C por 10 minutos.

O produto da reação foi aplicado em gel de agarose 1 % (m/v) e o inserto no tamanho correto foi excisado e purificado do gel. Assim, com a obtenção da sequência que codifica a isoforma 3, a clonagem foi prosseguida por digestão com as enzimas de restrição e ligação ao vetor pET-28a-sumo. A ligação foi utilizada para transformar cepas de *E. coli* DH5α e as colônias selecionadas foram submetidas a reação de PCR para verificar a presença do inserto. Os clones positivos foram propagados e tiveram seus DNAs plasmidiais extraídos para sequenciamento. O clone resultante foi denominado Iso3-1-pETsumo.

3.3 Expressão heteróloga

Na etapa de expressão, diversos testes foram realizados com as construções obtidas na clonagem, a fim de encontrar uma condição em que as proteínas fossem expressas na forma solúvel. Nesse sentido, procedemos com a variação da concentração do agente indutor, do tempo e da temperatura de expressão. Alternativamente, diferentes cepas de bactérias foram testadas, dentre elas BL21(DE3) (Novagen), ER2566 (NEB), Rosetta (DE3) (Novagen), C41(DE3) (Lucigen) e Arctic Express (*Agilent Technologies*). Ensaio de expressão na presença de osmólitos também foram realizados. De um modo geral, diferentes abordagens foram testadas na tentativa de obter as condições mais favoráveis para a produção da hGal12.

3.3.1 Expressão das construções em bactérias

A etapa de expressão heteróloga das isoformas e CRDs da hGal12 foi iniciada pela varredura das condições de expressão como concentração do agente indutor da expressão IPTG, temperatura e tempo total da expressão, de forma a maximizar o rendimento e a solubilidade das proteínas de interesse. Todas as etapas de expressão foram realizadas em ambiente estéril. De uma forma geral, os testes de expressão foram conduzidos como descrito a seguir. O pré-inóculo foi preparado em meio LB suplementado com o antibiótico adequado e cultivado a 37 °C durante 16 horas sob agitação constante e 180 rpm. A cultura foi então diluída na proporção 1:50 ou 1:100 em meio LB, acrescido do antibiótico, e incubada a 37 °C sob agitação até atingir uma densidade ótica entre 0,5 e 0,6 u.a. a 600 nm. Nesse momento, a cultura foi dividida em diferentes volumes aos quais foram adicionadas as concentrações finais de IPTG a serem testadas e incubada sob agitação a 180 rpm.

Depois da incubação, as células foram coletadas por centrifugação a 4 °C, 16000g e ressuspendidas em 0,5 ml de tampão acrescido de 1 mM PMSF e sonicadas com 5 W de potência por 5 vezes de 10 s com intervalos de 30 s. Os precipitados e sobrenadantes obtidos foram separados e analisados por SDS-PAGE, 12% ou 14%, dependendo do tamanho da construção, para verificar o rendimento e a solubilidade da proteína.

3.3.1.1 Construções clonadas a partir do cDNA

Após o sequenciamento, os clones foram utilizados para transformar cepas de bactérias *E. coli* e dar início aos ensaios preliminares de expressão das proteínas de interesse. O ensaio para as construções CL6-10-pTYB2, CL10-9-pTYB2, CC3-3-pTYB2 foram realizados em *E. coli* BL21 (DE3) variando as concentrações de IPTG (0,25; 0,50 e 1,00 mM) e o tempo de indução (1, 3, 6 e 24 horas) a 25 °C, nas culturas suplementadas com 100 µg/ml de ampicilina. Posteriormente, um ensaio de expressão foi realizado em cepas ER2566 e Rosetta (DE3) transformadas com a construção CC3-3-pTYB2, variando IPTG e o tempo de indução nas mesmas condições anteriores, a 25 °C e 37 °C. As construções CRD1a-9-pTYB2, CRD1b-1-pTYB2, CRD1c-1-pTYB2 e CRD2-10-pTYB2 relativas aos domínios N- e C-terminal das isoformas da hGal12 foram utilizadas para proceder o teste de expressão em *E. coli* BL21(DE3) variando as concentrações de IPTG (0,25; 0,50 e 1,00 mM) e o tempo de indução (3 e 6 horas) a 22 °C. O último teste de expressão em bactérias, para as construções clonadas a partir do cDNA, foi realizado com o CRD1a9-2.6-pET28a em *E. coli* BL21 (DE3) variando o agente indutor IPTG nas mesmas concentrações

anteriores, durante 3 horas a 37 °C. Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 20 mM Tris-HCl pH 8,0, 50 mM NaCl e 1 mM PMSF. Os precipitados e sobrenadantes foram analisados por SDS-PAGE 12 %.

3.3.1.2 Construções clonadas a partir do DNA otimizado

pTYB2

Após a etapa de otimização de códons raros para *E. coli*, na sequência que codifica o CRD1c da hGal12, foi dado início a novos ensaios de expressão. A construção CRD1c-2.2.5-6-pTYB2 em *E. coli* BL21(DE3) foi expressa variando as concentrações de IPTG nas condições acima citadas, durante 3 horas a 37 °C. O procedimento de extração e análise das proteínas totais foi realizado como descrito na seção anterior. O passo seguinte foi realizado no intuito de refinar as condições de expressão, variando a temperatura (18 °C e 25 °C) e o tempo de indução (3, 9 e 24 horas). O último ensaio de expressão com essa construção foi realizado pela variação da concentração de IPTG a 10, 25, 50, 75, 100, 150 e 200 µM, durante 24 horas, a 15 °C. Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 20 mM Tris-HCl pH 8,0, 50 mM NaCl e 1mM PMSF. A condição de expressão na qual foi atingida a melhor solubilidade e rendimento proteico foi obtida com a concentração de 75 µM de IPTG, induzidos por 24 horas a 15 °C. Nessas condições foram realizados ensaios de purificação em coluna de quitina.

pET-29a(+)

A construção CRD1c-3-pET29 em *E. coli* BL21 (DE3) foi expressa variando as condições as concentrações de IPTG (0,25, 0,50 e 1,0 mM), por 3 horas a 37 °C. As condições de expressão foram refinadas pela variação da concentração de IPTG (10, 25, 50, 75, 100, 150 e 200 µM) e do tempo de indução, 6 e 24 horas, a 18 °C, de forma a maximizar o rendimento e a solubilidade da proteína de interesse. Em seguida, a expressão das construções Iso2-2-pET29, CRD1b-1-pET29 e CRD2-pET29 foi testada em cepas de *E. coli* BL21(DE3) com 1 mM de IPTG durante 3 horas, a 37 °C. Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 50 mM Na₂PO₄, 300 mM NaCl pH 8,0, 14 mM β-ME e 1 mM PMSF. As amostras foram separadas em precipitado e sobrenadante e analisadas por SDS-PAGE 14%.

pET-28a(+)-sumo

O teste de expressão para a construção CRD1c-1-pETsumo foi conduzido em *E. coli* BL21 (DE3) variando as concentrações de IPTG (5, 50 e 500 μ M) e a temperatura de indução a 37 °C por 1 hora, 25 °C por 5 horas e 18 °C por 24 horas. A cultura de células foi preparada como descrito anteriormente, suplementado de 30 μ g/ml de canamicina como antibiótico. Um refinamento das condições de expressão foi realizado pela variação de IPTG (5, 10, 25, 50, 75, 100, 150 e 200 μ M) a 15 °C. A cultura foi dividida em duas partes e as células coletadas por centrifugação. Foram utilizados dois diferentes tampões para ressuspender as células: 1) 50 mM Na₂PO₄, 300 mM NaCl pH 8,0, 14 mM β -ME e 1mM PMSF e 2) 50 mM Na₂PO₄, 300 mM NaCl, 10% glicerol, 250 mM lactose, 1 mM thesit® (Sigma-Aldrich), 10 mM imidazol pH 8,0, 14 mM β -ME e 1mM PMSF. Os precipitados e sobrenadantes foram separados após lise das células e centrifugação, a fim de avaliar o rendimento e a solubilidade da proteína, por SDS-PAGE 14%.

As construções Iso2-3-pETsumo, Iso4-1-pETsumo, CRD1b-1-pETsumo e CRD2-5-pETsumo foram utilizadas para transformar cepas *E. coli* BL21 (DE3) e os ensaios de expressão foram conduzidos pela variação da concentração de IPTG (50, 500 e 1000 μ M), a 25°C e 18 °C, por 6 e 24 horas. Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 50 mM Na₂PO₄, 300 mM NaCl pH 8,0, 14 mM β -ME e 1 mM PMSF. As amostras foram separadas em precipitado e sobrenadante e analisadas por SDS-PAGE 14%.

C41(DE3)

O teste de expressão na cepa C41(DE3), que possui uma mutação no gene *lacUV5* promovendo uma diminuição da expressão e favorecendo a produção de proteínas de membranas (Wagner *et al.*, 2008), foi conduzido com o CRD1c-1-pETsumo, pela variação da concentração do agente indutor IPTG (25, 50, 100 e 500 μ M) nas temperaturas de 25 °C e 30 °C (por 3 e 6 horas) e 18 °C (por 6 e 24 horas). Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 50 mM Na₂PO₄, 300 mM NaCl pH 8,0, 14 mM β -ME e 1 mM PMSF. As amostras foram separadas em precipitado e sobrenadante e analisadas por SDS-PAGE 14%.

Arctic Express

A cepa Arctic Express possui as chaperonas Cpn10 e Cpn60 de *O. antarctica*, que compartilham 74% e 54% de identidade sequencial com as chaperonas de *E. coli* GroEL e GroES, respectivamente, e auxiliam no enovelamento proteico a baixas temperaturas, 4 °C - 12 °C. Assim, essa cepa foi transformada com o plasmídeo contendo o domínio CRD1c-sumo (CRD1c-1-pETsumo) e o ensaio de expressão foi feito pela variação da concentração do agente indutor IPTG (50, 500 e 1000 µM) por 24 horas a 10 °C. Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 50 mM Na₂PO₄, 300 mM NaCl pH 8,0, 14 mM β-ME e 1 mM PMSF. As amostras foram separadas em precipitado e sobrenadante e analisadas por SDS-PAGE 14%.

Ensaio de expressão na presença de osmólitos

Osmólitos são uma classe de compostos utilizados para estabilizar e até mesmo alterar a cinética de enovelamento de proteínas (Arakawa e Timasheff, 1985; Ignatova e Gierasch, 2006; Harries e Rösger, 2008). Assim, para as isoformas 2 (Iso2-3-pETsumo), isoforma 4 (Iso4-1-pETsumo), CRD1b (CRD1b-1-pETsumo), CRD1c (CRD1c-1-pETsumo) e CRD2 (CRD2-5-pETsumo) foi realizado um teste de expressão variando a concentração de IPTG (5, 50 e 500 µM), a 25 °C, por 6 horas, adicionando na cultura no momento antes da indução os seguintes osmólitos: 0,1 M glicina, 0,5 M sacarose, 0,5 M L-prolina e 2 M glicerol. Alternativamente, foi realizado um teste de expressão com as mesmas proteínas, porém com 500 µM de IPTG a 25 °C durante 6 horas, em que no momento anterior a indução foi adicionada na cultura 300 mM NaCl e 20 mM L-prolina. Após o término da expressão e coleta das células por centrifugação, as mesmas foram ressuspensas em tampão 50 mM Na₂PO₄, 300 mM NaCl pH 8,0, 14 mM β-ME e 1 mM PMSF. As amostras foram separadas em precipitado e sobrenadante e analisadas por SDS-PAGE 14%.

3.3.2 Expressão das construções em células de inseto

O teste de expressão em pequena escala foi realizado por infecção das células de inseto Tn5b com os baculovírus recombinantes para as isoformas 1, 2 e 4 da hGal12. Em uma placa de cultura de 24 poços foram adicionados 1 ml da suspensão de células em cada poço e incubadas a 27 °C, por 12 horas. O meio de cultura foi aspirado e substituído por 1 ml de meio TC-100 suplementado com 10% de soro bovino fetal, estreptomicina e penicilina. Às células foram adicionados 100 µl do estoque de baculovírus recombinante na seguinte disposição (figura 66).

Figura 66 – Representação esquemática da disposição das células durante o teste de expressão.

As células foram incubadas durante 24, 48, 72 e 96 horas a fim de verificar a produção das proteínas recombinantes nos diferentes estágios de infecção. Como controle negativo foram utilizadas células Tn5b não infectadas. Ao término de cada período, as células foram ressuspensas e transferidas para um tubo de centrífuga estéril. A cultura foi centrifugada e o sobrenadante contendo as proteínas secretadas separado. O precipitado de células foi ressuspensado em tampão PDB 1X, lisado por congelamento a -80°C e centrifugado durante 15 minutos, $16000g$ a 4°C . Foram separadas as amostras correspondentes ao sobrenadante 2 (contendo proteínas intracelulares solúveis) e ao precipitado (contendo proteínas insolúveis). As frações foram analisadas após migração por SDS-PAGE 14%.

3.4 Purificação

O processo de purificação foi realizado com os diferentes testes de expressão, em que uma condição mais promissora era escolhida para dar início a um ensaio de purificação. Desse modo, nem todas as construções chegaram a etapa de purificação, devido a problemas de formação de corpos de inclusão e instabilidade. De um modo geral, após a expressão na condição escolhida, as células foram coletadas por centrifugação a $10000g$, 4°C e armazenadas a -20°C .

3.4.1 Purificação por cromatografia de afinidade em coluna de quitina

O método cromatográfico de afinidade por quitina (Chong *et al.*, 1997), permite a purificação de proteínas clonadas em fusão com uma cauda de afinidade composta pelo domínio

de inteína, de *S. cerevisiae*, seguida por um domínio de ligação à quitina, CBD – *Chitin binding domain* de *B. circulans*. A inteína pode ser induzida a sofrer uma reação de auto clivagem na ligação peptídica na sua extremidade N-terminal na presença de agente indutor (Chong *et al.*, 1997).

Os experimentos de purificação foram realizados no sistema de cromatografia líquida Äkta prime™ (GE Healthcare). A etapa de purificação da proteína de interesse foi iniciada pelo descongelamento das células em gelo, seguida por ressuspensão em solução tampão contendo 20 mM Tris-HCl pH 8,0, 500 mM NaCl e 1 mM de 2ME (tampão B) acrescido de 1 mM de agente inibidor de proteases PMSF. A lise celular foi induzida por sonicação (P=10 W) sendo o procedimento repetido 15 vezes em pulsos de 30 s com intervalos de descanso de 30 s. Após esse processo, a amostra foi centrifugada por 30 minutos a 16000g e 4 °C para separar o material solúvel de debris celulares. O sobrenadante, contendo a proteína de interesse, foi aplicado em uma coluna cromatográfica (C10/10) previamente empacotada com a resina de quitina.

Para a remoção do material não ligado, a coluna foi lavada com 25 volumes de coluna do tampão B sob um fluxo de 0,5 ml/min. O processo de eluição se iniciou com a adição de tampão C contendo 20 mM Tris-HCl pH 8,0, 50 mM NaCl e 30 mM DTT, aplicado na coluna a um fluxo de 1 ml/min. O perfil da condutividade do material eluído da coluna durante a adição do agente redutor foi cuidadosamente monitorado para evitar a eluição antecipada da proteína de interesse.

A coluna foi então desconectada do cromatógrafo, fechada e armazenada a 4 °C por aproximadamente 16 h. Após o período de incubação na presença do agente redutor, a coluna foi reconectada e submetida a um fluxo de 1 ml/min (tampão C) e o efluente foi coletado em frações de 2 ml que posteriormente foram analisadas por SDS-PAGE 12%.

A cauda composta pela proteína inteína fusionada ao domínio de ligação a quitina permanece ligada a resina de quitina após o processo de eluição do domínio CRD1c. Para eluir a cauda, a coluna foi lavada com 6 volumes de coluna de água mili-Q seguida de 6 volumes de coluna de solução contendo 1% de dodecil sulfato de sódio (SDS). As frações contendo a cauda foram coletadas e analisadas por SDS-PAGE 12%. Adicionalmente, a purificação foi feita testando 100 mM de lactose, 5 mM ditiotreitol (DTT) (Sigma-Aldrich), 14 mM β ME e 0,5 % triton X-100 como aditivos nos tampões de purificação B e C.

3.4.2 Purificação por cromatografia de afinidade em coluna de níquel

O método cromatográfico de afinidade por níquel se baseia na clonagem do gene codificante da proteína alvo em fusão com uma cauda de poliistidina. A resina NTA (ácido nitrilotriacético) é um adsorvente quelante tetradentado que ocupa quatro dos seis sítios de ligação

da esfera de coordenação do níquel, permanecendo dois sítios livres para a interação com cauda de histidina. Aminoácidos como histidina apresentam alta afinidade por esses íons, assim, proteínas ricas em histidinas podem ser isoladas por eluição seletiva.

De um modo geral, as células previamente congeladas a -20°C após a etapa de expressão, são ressuspensas em tampão e submetidas a lise celular por sonicação com potência de saída de aproximadamente 10 W, durante 30 s com intervalos de descanso de 30 s, sendo o procedimento repetido 15 vezes. Em seguida a amostra foi centrifugada durante 30 minutos a $16000g$, a 4°C para separar o material solúvel de debris celulares. O sobrenadante, contendo a proteína de interesse, foi aplicado em coluna *Poly-Prep*® previamente empacotada com 2 ml da resina Ni-NTA. O processo de purificação foi realizado para as proteínas clonadas em pET-29a(+) e pET-28a(+)-sumo em condições nativas, pela aplicação de um gradiente em passos de imidazol, e desnaturantes através da variação do pH da solução, conforme descrito.

3.4.2.1 Purificação em condições nativas

As células foram ressuspensas em solução tampão A contendo 50 mM NaH_2PO_4 pH 8,0, 300 mM NaCl e 14 mM de 2ME acrescido de 1 mM de agente inibidor de proteases PMSF. Após o processo de lise e centrifugação, o sobrenadante, contendo a proteína de interesse, foi aplicado em uma coluna de níquel. O processo de eluição dos contaminantes e da proteína de interesse foi realizado pela passagem sucessiva do tampão A acrescido de imidazol nas seguintes concentrações: 0, 25, 50, 100, 250 e 500 mM. Frações de 2 ml foram coletadas e submetidas a SDS-PAGE 14%. Alternativamente, foram adicionados aditivos na solução tampão em diferentes testes de purificação, para aumentar a estabilidade e a solubilidade da proteína. Os aditivos utilizados foram triton X-100, lactose, thesit®, glicerol e fosfato.

3.4.2.2 Purificação em condições desnaturantes

As células foram ressuspensas em solução tampão contendo 100 mM NaH_2PO_4 , 10 mM Tris-HCl e 8 M de ureia pH 8,0 (tampão D). Após a lise e centrifugação, o sobrenadante foi aplicado em coluna de níquel e a eluição dos contaminantes foi realizada pela passagem sucessiva de tampão D nos pHs 6,3 para lavagem, 5,9 e 4,5 para a eluição da proteína. Frações de 2 ml foram coletadas e analisadas por SDS-PAGE 14%.

3.4.2.3 Purificação de corpos de inclusão, solubilização e reenovelamento

Os corpos de inclusão, partículas densas de agregados proteicos, foram purificados por sucessivas etapas de lavagem e centrifugação. Em seguida os corpos de inclusão foram solubilizados e submetidos ao processo de reenovelamento. Nesse sentido, dois protocolos foram utilizados e são descritos a seguir.

Protocolo 1

Esse protocolo, adaptado do trabalho descrito por Singh e colaboradores (Singh e Panda, 2005) foi aplicado para as construções Iso2-2-pET29a, CRD1b-1-pET29a, CRD1c-3-pET29a e CRD2-pET29a. As células previamente coletadas por centrifugação e armazenadas a -20°C foram ressuspensas em solução 20 % (m/v) de sacarose, sonificadas 8 vezes (30 s com potência de saída de 10 W e intervalos de descanso de 30 s) e centrifugadas a 8000g a 4 °C. O sobrenadante foi descartado e o material insolúvel precipitado foi ressuspendido em tampão 50 mM Tris-HCl pH 8,5 e 0,5% (m/v) CHAPS, sonificadas e centrifugadas; esse procedimento foi repetido duas vezes e a amostra foi incubada a 37 °C por 12 horas. O próximo passo foi seguido por centrifugação e ressuspensão em tampão 50 mM Tris-HCl pH 8,5 e incubação por 30 minutos a temperatura ambiente. As amostras foram centrifugadas por 20 minutos, 8000g a 4 °C. Uma vez obtido o corpo de inclusão purificado, procedeu-se a ressuspensão dos mesmos em tampão contendo 100 mM Tris pH 12 e 2 M ureia. Em seguida, a proteína solubilizada foi diluída no tampão de reenovelamento (50 mM Tris-HCl, 2 M ureia, 10% (v/v) glicerol, 5% (m/v) sacarose, 14 mM β ME e 1 mM PMSF pH 8,0) sob agitação constante e de forma pulsada. A amostra final foi filtrada em membrana 0,2 μ m.

Protocolo 2

Esse protocolo foi otimizado para a construção CRD1c-1-pETsumo e aplicado para a construção CRD1c-3-pET29. As células congeladas, foram ressuspensas em tampão de lise contendo PBS 1X, 0,05% (v/v) thesit® e 1 mM PMSF. A suspensão foi sonificada 15 vezes por 30 s com intervalos de descanso de 30 s e potência 10 W. O extrato bruto foi então centrifugado por 30 minutos, a 16000g, 4 °C e o sobrenadante descartado. O precipitado, correspondendo principalmente aos corpos de inclusão, foi ressuspendido em tampão de lavagem (PBS 1X, 0,05% (v/v) thesit®, 25% (m/v) sacarose e 0,5 M ureia), sonificado e centrifugado nas mesmas condições anteriores. Essa etapa foi repetida duas vezes, sendo o sobrenadante descartado. A etapa de

solubilização foi iniciada pela ressuspensão dos corpos de inclusão em tampão contendo 50 mM NaH_2PO_4 pH 8,0, 300 mM NaCl, 8 M ureia, 0,2 % (m/v) SDS e 14 mM 2ME. Em seguida, o material foi sonificado e incubado a temperatura ambiente por 12 horas. As proteínas solubilizadas foram centrifugadas em temperatura ambiente para separar os debris celulares e aplicadas em coluna de níquel, previamente equilibrada com o tampão de solubilização. Os contaminantes foram removidos pela aplicação do tampão de solubilização acrescido de 25 mM de imidazol e a proteína eluiu pelo acréscimo de 200 mM de imidazol no tampão.

As amostras purificadas foram submetidas a etapa de reenovelamento por diluição pulsada em tampão de reenovelamento contendo 20 mM Tris-HCl pH 7,5, 150 mM NaCl, 10% glicerol, 14 mM β ME e 0,005 % (v/v) thesit®, sob agitação lenta a 4 °C. Em seguida, a solução foi repassada na coluna de níquel, eluída com o acréscimo de imidazol, concentrada e dialisada em tampão de reenovelamento. A proteína purificada foi submetida a leitura de tamanho de partícula e índice de polidispersividade por espalhamento dinâmico de luz – DLS, em um Zetasizer Nano (Malvern).

3.5 Ensaios imunológicos

3.5.1 Produção de anticorpos policlonais

3.5.1.1 Imunização de animais de experimentação

Para a produção de anticorpos policlonais, dois coelhos albinos, machos e adultos (New Zeland) foram imunizados cada um com os corpos de inclusão dos domínios CRD1c e CRD2 da hGal12. Os experimentos foram realizados no biotério da Faculdade de Ciências Farmacêuticas de Ribeirão Preto. O projeto em questão foi aprovado pela Comissão de Ética no Uso de Animais da PUSP-RP (Processo nº 11.1.1355.53.0).

As amostras de corpos de inclusão foram dissolvidas em 1,5 ml de tampão PBS 1X e emulsificadas na proporção 1:1 com adjuvante completo de Freund (Sigma-Aldrich). A emulsão foi aplicada em cinco pontos no dorso dos animais. A partir da segunda imunização, utilizou-se adjuvante incompleto de Freund (Sigma-Aldrich) e a amostra foi aplicada com injeção intramuscular na pata. Foram realizadas cinco imunizações, com intervalos de 15 dias. Previamente a cada imunização foram colhidas alíquotas de sangue total dos coelhos, a fim de determinar a eficiência do procedimento de hiperimunização.

O sangue total foi mantido por 1 hora a temperatura ambiente para formação do coágulo e posterior coleta do soro. Em seguida, as amostras foram centrifugadas a 1060g por 10 minutos.

O soro foi separado por aspiração, aquecido a 56 °C por 30 minutos, aliqotado e armazenado a -20 °C. Após a confirmação da eficiência do procedimento de hiperimunização, por *western blot*, foi feito o procedimento de sangria “branca” por punção cardíaca, apenas no animal imunizado com o CRD2, uma vez que o animal que foi desafiado com o CRD1c não sobreviveu até o final desse procedimento. O soro hiperimune foi obtido como descrito anteriormente.

3.5.1.2 Purificação do anticorpo policlonal

A etapa de purificação do anticorpo policlonal de coelho anti-CRD2 foi iniciada pela precipitação dos anticorpos com 40% (m/v) de sulfato de amônio, adicionado gota a gota, sob agitação leve por 30 minutos. O material foi mantido a 4°C por 12 horas e em seguida foi centrifugado a 3000g durante 30 minutos. O precipitado foi lavado com solução 40% (m/v) de sulfato de amônio, centrifugado a 3000g por 10 minutos, ressuspendido em água e dialisado contra PBS 1X a 4 °C, durante 2 dias, com 3 trocas de tampão por dia. O soro dialisado foi esterilizado por filtração com o uso de membrana de 0,2 µm e quantificado a 280 nm. Em seguida, visando a obtenção de uma preparação de imunoglobulinas da classe G (IgG) a partir do soro-imune anti-CRD2, o material biológico foi aplicado em uma coluna de proteína G, previamente equilibrada com PBS 1X. A coluna foi lavada com PBS 1X e os anticorpos ligados a coluna eluíram com solução 0,1 M glicina pH 2,8. As alíquotas foram coletadas e neutralizadas com 1 M K₂HPO₄ pH 10. O eluato foi concentrado em concentrador Vivaspín™ (GE Healthcare) MWCO 100000 e dialisado contra PBS 1X. A solução de anticorpos foi filtrada em uma membrana de 0,22 µm em ambiente estéril, aliqotada e armazenada a 4 °C. A amostra de anticorpos foi quantificada a 280 nm e a concentração calculada utilizando o coeficiente de extinção $\epsilon_{280\text{ nm}}^{\text{mb/ml}}=1,35$ (Fasman, 1989).

3.5.2 Western blot

A técnica de *western blot* foi utilizada para verificar a eficiência da produção de anticorpos policlonais a partir de soros de animais imunizados com CRD1c e CRD2. Além disso, este procedimento foi usado para avaliar a imunorreatividade do soro-imune anti-CRD2 para as isoformas e o domínio CRD2 da hGal12, nos extratos celulares ou nas amostras de corpo de inclusão purificadas. Os anticorpos primários utilizados nessa etapa foram a preparação de IgG de coelho anti-CRD2 e anti-gal12 feito em cabra (Santa Cruz Biotechnology). Os anticorpos secundários, conjugados com a enzima HRP (*horseradish peroxidase*) foram o anti-IgG de coelho feito em cabra (Santa Cruz Biotechnology) e anti-IgG de cabra feito em burro (*Life Technologies*), nas diluições sugeridas pelo fabricante.

As amostras proteicas foram submetidas a SDS-PAGE 14 % e eletrotransferidas, por 1 hora, do gel para a membrana de nitrocelulose (Bio-Rad), utilizando o sistema *Mini Trans-Blot® Cell* (Bio-Rad), a 300 mA e 100 V. Após a transferência a membrana foi submetida a etapa de bloqueio em tampão TBS-tween (Tampão TBS 1X contendo 0,1% (v/v) de tween 20) acrescido de 5% (m/v) de leite em pó desnatado, durante 12 horas a 4 °C, sob agitação lenta. Em seguida a membrana foi lavada 3 vezes com tampão TBS-tween durante 15, 5 e 5 minutos. Na próxima etapa, a membrana foi incubada com o tampão de bloqueio acrescido do anticorpo primário e incubado sob agitação, em temperatura ambiente por 1 hora. A membrana foi lavada 3 vezes com tampão TBS-tween e incubada com o tampão de bloqueio, agora acrescido do anticorpo secundário, por 1 hora a temperatura ambiente sob agitação. Ao término da incubação, as membranas foram lavadas e as proteínas foram detectadas por quimioluminescência pela adição das soluções de detecção, segundo as instruções do fabricante (*ECL western blotting detection reagents – Amersham*). A detecção foi realizada por exposição da membrana a um filme de raios X ou em um fotodocumentador ChemiDoc (Bio-Rad).

3.5.3 Microscopia confocal

A imunolocalização celular da hGal12 foi feita com o uso de células HL-60, anticorpo primário anti-gal12 feito em cabra (*Santa Cruz Technology*), o anticorpo IgG de burro (*Jackson ImmunoResearch*) para o bloqueio, o anticorpo secundário Alexa Fluor® 546 anti IgG de cabra feito em burro (*Life Technologies*) com fluorescência no vermelho e os marcadores DAPI (4',6-diamidino-2-phenylindole) (*Life Technologies*) para DNA com emissão de fluorescência máxima em torno de 460 nm (cor azul) e BODIPY para lipídeos neutros com fluorescência máxima em torno de 503 nm (verde). O marcador BODIPY foi sintetizado pelo laboratório de química heterocíclica e medicinal e gentilmente cedido pelo Prof. Flávio da Silva Emery.

Inicialmente as células HL-60, promieloblastos do sangue periférico humano, foram cultivadas em meio RPMI 1640 (2 mM de glutamina, 1% de penicilina e estreptomicina e 10% (v/v) soro bovino fetal, inativado por calor) em uma concentração entre 1.10^5 e 1.10^6 células/ml em 5% CO₂ a 37 °C. As células foram coletadas, centrifugadas a 240g, por 10 minutos a temperatura ambiente. O material foi ressuspensionado em PBS 1X e uma alíquota de 20 µl foi adicionada à 20 µl de Tripán Blue para a contagem em câmara de Neubauer.

Lamínulas circulares, previamente preparadas em uma placa de cultura de 24 poços, foram utilizadas para fixar as células. As lamínulas foram cobertas com 200 µl de uma solução de poli-L-lisina a 75 µg/ml durante 10 minutos a temperatura ambiente. Em seguida, a solução de poli-L-

lisina foi aspirada e a placa foi colocada em uma estufa a 37 °C para secar. Em seguida, foram adicionados aos poços 200 µl da suspensão de células, em um total de $5 \cdot 10^4$ células por poço. A placa foi incubada a 37 °C por 1 hora, o excesso de células na suspensão foi aspirado e os poços foram lavados com PBS 1X e fixados com 500 µl de formaldeído 2% (v/v) em PBS 1X, durante 20 minutos a 37 °C. As células foram permeabilizadas por 10 minutos com 0,3 % de triton X-100 em PBS. Os sítios livres do formaldeído foram bloqueados com 500 µl de uma solução de 0,25 M de glicina em PBS 1X.

Em seguida, foram realizadas as etapas de bloqueio com uma solução contendo PBS 1X contendo 1% BSA e 5 µg/ml de IgG de burro, por 45 minutos. Os poços foram lavados 3 vezes de 5 minutos com PBS 1X e incubados por 1 hora com anticorpo primário anti-gal12 (10 µg/ml) diluído em tampão de bloqueio (PBS 1X e 1% BSA). Como controle negativo, a um poço não foi adicionado o anticorpo primário. Os poços foram lavados 5 vezes por 5 minutos cada vez com PBS 1X e incubados por 1 hora em uma solução contendo o anticorpo secundário (1:350), o marcador DAP (1:125) e o marcador BODIPY (1:1200). Por fim, as placas foram lavadas 5 vezes de 5 minutos em PBS 1X, enxaguadas rapidamente em água mili-Q e montadas em lâminas com solução de montagem contendo PBS 1X pH 8,2, 90 % (v/v) glicerol e 10 mM de paraifenilendiamina. As lâminas foram seladas com esmalte e armazenadas a 4 °C. O experimento foi realizado como esquematizado (tabela 18).

Tabela 18 – Esquema da distribuição dos marcadores fluorescentes do procedimento de imunolocalização da hGal12 endógena em células HL-60, por microscopia confocal.

	Poço 1	Poço 2	Poço 3
Anticorpo primário	-	Anti-gal12	Anti-gal12
Anticorpo secundário	Alexa Fluor® 546	Alexa Fluor® 546	Alexa Fluor® 546
Marcadores	DAPI	DAPI	DAPI + BODIPY

O ensaio de localização celular da hGal12 endógena foi feito por imunofluorescência com o uso de microscopia confocal e do equipamento LEICA TCS SP8 (Leica *Microsystems*).

3.6 Modelagem molecular

Durante o desenvolvimento do projeto nos deparamos com uma grande dificuldade em obter as isoformas e os domínios CRDs da galectina-12 na sua forma solúvel. Diante dos resultados experimentais obtidos levantamos várias hipóteses que pudessem explicar a formação de corpos de inclusão e a insolubilidade da proteína. Em paralelo aos experimentos traçados para corroborar

nossas hipóteses, fizemos o uso da ferramenta de modelagem comparativa por homologia para obter informações estruturais que pudessem responder tais questões.

Os modelos para as isoformas foram calculados na base de dados ModBase (<http://salilab.org/modbase>) através do programa ModPipe (Pieper *et al.*, 2011). Os modelos para o CRD1a, CRD1b e CRD2 foram obtidos utilizando-se como molde os domínios CRD1 e CRD2 da estrutura da galectina 8 (código PDB 4HAN) com a qual compartilham aproximadamente 42% de identidade sequencial para o CRD1a e CRD1b e 26% para o CRD2. O modelo para o CRD1c foi obtido tendo como estrutura de partida o N-terminal da galectina-4 de camundongo (código PDB 3I8T) com a qual compartilha cerca de 35% de identidade sequencial.

Os modelos de saída foram submetidos a 5 macrociclos de minimização da geometria considerando-se os termos de comprimento de ligação, distâncias não-ligadas, ângulo de ligação, ângulo diedral, quiralidade e planaridade, implementados no programa phenix.geometry_minimization (Adams *et al.*, 2010). O último ciclo de otimização do modelo foi realizado manualmente no programa Coot (Emsley *et al.*, 2010), tendo como referência a validação da geometria e a satisfação do Gráfico de Ramachandran.

A superfície de potencial eletrostático foi calculada com o programa APBS (*Adaptive Poisson-Boltzman Solver*) implementado no servidor PDB2PQR (Dolinsky *et al.*, 2004) empregando-se o campo de força AMBER e utilizando o estado de protonação em pH 7,4 calculado pelo programa PROPKA (Bas, Rogers e Jensen, 2008). Os resultados foram visualizados no programa pymol (Delano, 2004).

RESULTADOS E DISCUSSÕES

3.7 Produção das macromoléculas

3.7.1 Produção em bactérias das macromoléculas obtidas a partir do cDNA

A primeira etapa de obtenção do DNA molde, para a clonagem dos fragmentos que codificam as isoformas da hGal12, foi realizada pela construção do cDNA fita simples a partir do RNA total extraído de leucócitos totais humanos. Aproximadamente 18 ml de sangue periférico humano foram utilizados para a obtenção de 3.10^7 células de leucócitos, que foram separados por um gradiente de densidade descontínuo. O RNA total foi extraído e a integridade da amostra foi visualizada, após migração em gel de agarose, pela presença das bandas 28S e 18S do RNA ribossômico. O RNA total foi quantificado e utilizado como molde para o processo de síntese da fita simples do cDNA. O material resultante foi combinado e purificado, resultando em uma amostra de cDNA a 24,6 ng/ μ l.

O cDNA foi utilizado como molde nas reações de PCR para a amplificação do gene de interesse. Como resultado dessa etapa foram obtidas as isoformas 1, 2 e 4 clonadas no vetor pTZ57R/T e posteriormente subclonadas em pTYB2 descritas como CL6-10-pTYB2 (isoforma 1), CL10-9-pTYB2 (isoforma 2) e CC3-3-pTYB2 (isoforma 4). Foi iniciado um teste preliminar de expressão em cepas *E. coli* BL21 (DE3). Pela análise do gel de poliacrilamida das amostras, observamos que não houve expressão das proteínas. Assim, testamos a expressão da isoforma 4 em cepas ER2566 (indicada para os vetores pTYB2 do sistema IMPACT) e Rosetta (DE3) (indicada para proteínas que possuem códons raros para *E. coli*). Como resultado, nenhuma banda referente à proteína de interesse foi encontrada nas alíquotas referentes ao precipitado ou sobrenadante.

Por sugestão do Prof. Dr. Marcelo D. Baruffi, utilizamos a estratégia de clonar os CRDs das isoformas separadamente, assim como foi feito para a hGal4, por evidência experimental do sucesso na expressão da galectina-9 após a diminuição do peptídeo de ligação que une os CRDs. Deste modo, utilizamos as construções CL6-10-pTYB2, CL10-9-pTYB2 e CC3-3-pTYB2 como molde para a obtenção dos respectivos CRDs clonados no vetor pTYB2: CRD1a (CRD1a-9-pTYB2), CRD1b (CRD1b-1-pTYB2), CRD1c (CRD1c-1-pTYB2) e CRD2 (CRD2-10-pTYB2). Os testes de expressão para os quatro CRDs foram realizados em *E. coli* BL21 (DE3) e a análise

do gel de poliacrilamida não mostrou nenhuma evidência de bandas referentes às proteínas de interesse, no precipitado ou sobrenadante. Nesse caso, a fim de verificar se a ausência de expressão era devido a problemas no vetor pTYB2, foi realizada a subclonagem do CRD1a no vetor pET-28a(+) (CRD1a9-2.6-pET28). A clonagem foi efetuada com sucesso, entretanto, não obtivemos expressão da proteína. Em todos os testes de expressão foram variadas as concentrações de agente indutor IPTG, temperatura e tempo de expressão. A análise das amostras no gel de poliacrilamida foi comparada com um controle negativo, ao qual o agente indutor não foi adicionado.

3.7.2 Produção das macromoléculas em bactérias a partir do DNA otimizado

Em vista do insucesso na expressão das proteínas nas condições testadas, realizamos a análise de códons nas sequências das isoformas da hGal12 e comparamos com os códons preferencialmente utilizados pela bactéria *E. coli*. Foi possível constatar que o gene que codifica as isoformas da hGal12 se utiliza preferencialmente de vários códons, como para expressar os resíduos de arginina, glicina, isoleucina e prolina, que são considerados raros para a *E. coli* (tabela 19).

Tabela 19 – Quantidade percentual de códons raros para *E. coli* presentes nas isoformas da hGal12

aminoácido	Isoforma 1	Isoforma 2	Isoforma 3	Isoforma 4	Isoforma 5
ARG	82%	82%	82%	80%	79%
GLY	65%	65%	65%	68%	68%
ILE	13%	13%	13%	18%	18%
LEU	2%	2%	2%	3%	3%
PRO	36%	36%	36%	50%	44%

Avaliando a contribuição dos códons raros em toda a sequência, temos que para cada isoforma cerca de 14 % dos aminoácidos são codificados por códons raros para *E. coli*. Além disso, a inspeção da localização desses códons na sequência mostra que há formação de *clusters* de códons raros, o que significa que dois ou mais códons raros estão em sequência. Para as isoformas 1 e 2 ocorrem 7 clusters e para as isoformas 3, 4 e 5 ocorrem uma quantidade de 6, 5 e 4 *clusters*, respectivamente. É possível que a formação de *clusters* dificulte ainda mais o reconhecimento de um dado códon raro pela bactéria. Para comprovar essa hipótese, a sequência que codifica o CRD1 da isoforma 4, CRD1c, foi submetida a otimização de códons para *E. coli* e síntese (IDT – Integrated DNA Technologies). O fragmento sintético, clonado originalmente no vetor pIDTsmart foi amplificado com novos oligonucleotídeos e clonado em pTYB2 sendo denominado CRD1c-2.2.5-6-pTYB2.

3.7.2.1 Produção da construção CRD1c-2.2.5-6-pTYB2

O plasmídeo recombinante CRD1c-2.2.5-6-pTYB2 foi utilizado para transformar bactérias *E. coli* BL21 (DE3) e o teste de expressão foi realizado a 37 °C. Foi observada a presença da banda referente à construção CRD1c-inteína, com aproximadamente 70 kDa e dado início ao processo de refinamento das condições de expressão. A melhor condição para a expressão do domínio CRD1c foi obtida com indução a 75 μ M de IPTG, a 15 °C durante 24 horas (figura 67b poço 5).

Figura 67 – Análise da expressão da proteína por SDS-PAGE 14% com o aumento da concentração de IPTG. MM - Marcador de proteínas de baixo peso molecular (GE Healthcare). 1) controle negativo, 2) 10 μ M, 3) 25 μ M, 4) 50 μ M, 5) 75 μ M, 6) 100 μ M, 7) 150 μ M e 8) 200 μ M nas frações insolúveis (gel A) e solúveis (gel B).

Com o sucesso na etapa de expressão do domínio CRD1c-2.2.5-6-pTYB2, demos início a purificação em coluna de quitina. A purificação foi realizada como descrito na seção 3.4.1.

Figura 68 – A) Perfil cromatográfico da eluição da proteína CRD1c-inteína, após o processo de clivagem com o agente redutor DTT e da cauda de afinidade inteína após processo de limpeza com 1% de SDS. B) SDS-PAGE 15%. MM) Marcador de peso molecular. Amostras referentes a diferentes etapas de purificação: 1) volume de 2 ml, 2) volume de 5 ml, 3) volume de 10 ml, 4) volume de 80 ml, 5) volume de 82 ml e 6) volume de 84 ml.

Na análise do perfil cromatográfico não foi possível observar a eluição da proteína após adição de DTT (figura 68a, de 0 a 15 ml), sugerindo que não havia a presença da proteína de interesse na fração solúvel. Entretanto, após a passagem de SDS pela resina (procedimento para

retirada da cauda) pudemos ver um perfil de eluição (figura 68a, de 80 a 100 ml) e constatar a presença de proteínas por SDS-PAGE (figura 68b). A análise do perfil cromatográfico mostrou um pico de absorbância em torno de 600 mAu que indica a presença de uma quantidade considerável de proteína sendo eluída. A análise das bandas no gel de poliácridamida mostrou a presença de um fragmento de aproximadamente 56 kDa, compatível com a cauda de inteína e uma banda de aproximadamente 14 kDa, compatível ao domínio CRD1c. Três hipóteses foram levantadas para explicar esse comportamento. A primeira seria a clivagem antecipada da inteína devido à presença de 1 mM de 2ME no tampão de ressuspensão, lise e lavagem da coluna. A segunda hipótese seria uma possível agregação da proteína após o processo de clivagem, fazendo com que a proteína precipitasse na resina e por último, não descartamos a possibilidade da fusão ter sofrido ataque proteolítico em solução (manual IMPACTTM).

Para testar a hipótese de agregação por falta de ambiente reduzido ou insolubilidade, realizamos 4 purificações testes na presença dos aditivos 100 mM lactose, 5 mM DTT, 14 mM 2ME e 0,1% (v/v) triton X-100, nos tampões de lise, lavagem e eluição (figura 69). Em todos os casos, o perfil de eluição da proteína foi semelhante para os diferentes aditivos. As frações foram coletadas e a coluna foi lavada com solução adstringente 1% SDS.

Figura 69 – Perfil cromatográfico de eluição do domínio CRD1c após clivagem com agente redutor. O gráfico mostra a sobreposição das curvas de eluição nas diferentes condições testadas.

Apesar do perfil de eluição com absorção significativa a 280 nm, não foi possível visualizar qualquer banda no gel de poliácridamida que pudesse estar associada a presença desse pico.

Entretanto, da mesma forma observada na purificação anterior, foi possível eluir a inteína da coluna e visualizar as bandas referentes à cauda no gel de poliacrilamida. A presença de um pico de absorbância para o material eluído, associado à ausência de bandas no gel e à presença da cauda depois da limpeza da coluna embasam nossa hipótese da fusão CRD1c-2.2.5-6-pTYB2 ser expressa na fração solúvel altamente agregada ou agregado após a clivagem. Para confirmar nossa hipótese, amostras do material eluído foram submetidas para análise por DLS (espalhamento dinâmico de luz) que revelou a presença de picos de intensidade em diferentes tamanhos (tabela 20). Os resultados por DLS sugeriram que a amostra encontra-se majoritariamente agregada, uma vez que o tamanho médio de um CRD é de 40 Å (estimativa feita a partir da estrutura cristalográfica de diferentes CRDs).

Tabela 20 – Valores obtidos para os picos majoritários da distribuição de tamanho das partículas em solução.

	diâmetro (Å)	% intensidade
Pico 1	650	60,4
Pico 2	7,5	14,1%
Pico 3	3,9	11,4%

Outra hipótese avaliada foi a de estar ocorrendo clivagem antecipada devido à presença do agente redutor no início da purificação. A presença do β -mercaptoetanol é necessária para estabilizar as cisteínas livres dos CRDs, mas poderia estar ativando o sistema de remoção da cauda. Para responder essa questão foi realizado um experimento de purificação na ausência de agentes redutores em qualquer etapa do processo. A coluna foi carregada com o extrato proteico, lavada e em seguida foi aplicada a solução adstringente de SDS. As frações foram coletadas, submetidas a migração por eletroforese em gel de poliacrilamida e transferidas para uma membrana de PVDF para sequenciamento N-terminal no laboratório de Toxinologia - FCFRP. Os 15 resíduos sequenciados foram submetidos ao alinhamento local utilizando o programa BLAST. O alinhamento mostrou 100% de identidade sequencial para os 14 dos 15 primeiros resíduos da endonuclease com atividade *splicing* de *S. cerevisiae* EC: 2.6.1.34 (figura 70). Levando em consideração que resíduos de cisteínas não são detectados a menos que estejam reduzidos ou alquilados, podemos considerar o primeiro resíduo como cisteína. Portanto, o sequenciamento N-terminal mostrou claramente que o conteúdo majoritário da fração que elui da coluna se trata exclusivamente da cauda de inteína, fortalecendo a hipótese de clivagem antecipada da cauda.

	1	10	20	30	40	50	60	70	80	90	100
1VDE Pi-Scei	QFAKGTNVLMADGSI										
sequenciamento	QFAKGTNVLMADGSI										
	110	120	130	140	150	160	170	180	190	200	
1VDE Pi-Scei	YFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPAPERANELVESYRKASNKAYFENTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHL										
sequenciamento	YFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPAPERANELVESYRKASNKAYFENTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHL										
	210	220	230	240	250	260	270	280	290	300	
1VDE Pi-Scei	TIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTYAEKLNLCASYKDRKEPQVAKTVNLYSKVVRGNGIRNNLTENFLWDAIVGLGFLKDG										
sequenciamento	TIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTYAEKLNLCASYKDRKEPQVAKTVNLYSKVVRGNGIRNNLTENFLWDAIVGLGFLKDG										
	310	320	330	340	350	360	370	380	390	400	
1VDE Pi-Scei	KNIPSFSLTDNIGTRETFLAGLIDSDGYVTDEHGKATIKTIHTSVRDGLVSLARSLGLVSVNAEPARVDMNGTKHKISYAIYMSGGDVLLNVLSKCA										
sequenciamento	KNIPSFSLTDNIGTRETFLAGLIDSDGYVTDEHGKATIKTIHTSVRDGLVSLARSLGLVSVNAEPARVDMNGTKHKISYAIYMSGGDVLLNVLSKCA										
	410	420	430	440	450						
1VDE Pi-Scei	SKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN										
sequenciamento	SKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN										

Figura 70 – Alinhamento local entre as sequências primárias da proteína íntea Sce VMA1 e os resíduos identificados no processo de sequenciamento N-terminal da amostra purificada.

Como resultado desses experimentos, podemos concluir que os insucessos na obtenção da proteína de interesse, nesse caso utilizando a construção CRD1c-2.2.5-6-pTYB2 como molde, são consequência de dois processos. Acreditamos que a maioria da proteína produzida em *E. coli* sofre proteólise liberando o domínio CRD1c no extrato proteico. Por outro lado, a construção íntegra e talvez esteja íntegra sem ataque proteolítico, porque se trata de uma amostra de proteína majoritariamente agregada. Desse modo, optamos por trocar o vetor de expressão por clonagem das sequências de interesse em vetores da série pET.

3.7.2.2 Produção da construção CRD1c-3-pET29

O domínio CRD1c foi então clonado no vetor pET-29a(+) (CRD1c-3-pET29) e submetido aos ensaios de expressão em bactérias *E. coli* BL21 (DE3). Os ensaios de expressão nessa construção não indicaram uma condição em que fosse possível obter a proteína solúvel. Assim, testamos uma purificação em condições desnaturantes. Ainda assim, a proteína majoritariamente permaneceu no precipitado e apenas uma pequena quantidade de proteína foi isolada na presença de contaminantes. A adição dos detergentes triton X-100, n-octil- β -D-glicopiranosídeo e CHAPS foram utilizados em diferentes purificações sob condições desnaturantes, para testar a recuperação da proteína na fração insolúvel e ajudar no processo de eliminação dos contaminantes. Entretanto, não houve recuperação significativa da proteína, sugerindo que a mesma estivesse formando corpos de inclusão.

Na tentativa de isolar a proteína de interesse, foi primeiramente testado o protocolo 1 de purificação de corpos de inclusão. Tal protocolo propõe a manutenção de elementos de estruturas secundária intactos, por meio de condições de desnaturação amenas. Assim, o processo de reenovelamento da proteína seria favorecido e garantiria maior rendimento da proteína ativa. O protocolo foi seguido como descrito na seção 3.4.2.3. Na etapa final, quando a amostra foi filtrada em membrana de 0,2 μ m, toda a amostra proteica ficou retida, indicando a agregação da proteína durante o processo de reenovelamento.

Como alternativa, o experimento foi repetido em condições desnaturantes mais fortes, na presença de 8 M de ureia e 2 M de guanidina. A amostra foi aplicada na resina de níquel para purificação sob condições desnaturantes. Novamente, durante a tentativa de promover o reenovelamento da proteína na coluna, através da diluição das concentrações dos agentes desnaturantes, houve precipitação da proteína na coluna no momento da remoção da guanidina.

3.7.2.3 Produção da construção CRD1c-1-pETsumo

Condições Nativas

Observando o perfil de expressão do CRD1c nas construções em pTYB2 e pET-29a(+) vimos que nas condições testadas só obtivemos uma fração solúvel de proteína quando esta estava fusionada a uma proteína, no caso a inteína. Assim, procedemos com a clonagem do CRD1c em pET-28a(+)-sumo resultando na construção CRD1c-1-pETsumo. A estratégia de utilizar a proteína SUMO é conhecida por melhorar a expressão e a solubilidade de proteínas (Malakhov *et al.*, 2004; Butt *et al.*, 2005; Zuo *et al.*, 2005)

Diferentes ensaios de expressão em *E. coli* BL21 (DE3) foram utilizados e apesar da tendência da proteína à formar corpos de inclusão, obtivemos condições em que uma pequena quantidade de proteína estava na fração solúvel. Nos ensaios preliminares de purificação ficou evidente a instabilidade da proteína devido à alta taxa de precipitação ao longo da purificação. Foram testados o uso de aditivos como lactose, detergentes (triton X-100 e thesit®), fosfato e glicerol, além de variações na quantidade de sal e pH. Em um perfil eletroforético típico de purificação por afinidade em resina de níquel (figura 71), observamos a eluição da proteína durante todo o processo de lavagem com imidazol, o que sugere agregação da amostra, além da presença de uma grande parte da proteína que permanece precipitada na coluna (figura 71c poço 9).

Figura 71 – Perfil eletroforético de purificação em coluna de níquel do domínio CRD1c em fusão com a proteína sumo. MM) Marcador molecular. A) Eluição com 50 mM de imidazol 1-9) frações 1, 2, 3, 4, 5, 6, 8 e 10; B) Eluição com 100 mM de imidazol 1-9) frações 1, 2, 3, 4, 5, 6, 8 e 10; C 1-4) frações 1, 2, 3 e 4 da eluição com 250 mM de imidazol; 5-8) frações 1, 2, 3 e 4 da eluição com 500 mM de imidazol, 9) amostra da resina de níquel após a limpeza da coluna; D) Purificação pós-clivagem 1-5) frações da eluição da proteína CRD1c; 6-8) frações referentes a eluição com 500 mM de imidazol 9) amostra concentrada das frações de 1 a 5.

Apesar do perfil de precipitação, as amostras referentes à eluição em 500 mM de imidazol foram concentradas, dialisadas para a retirada do imidazol e submetidas a clivagem da cauda com a proteína ULP1. Após a etapa de clivagem, as amostras da eluição da proteína (figura 71d poços 1 a 5) foram concentradas e visualizadas no gel (figura 71d poço 9). Observamos a presença de uma banda próxima ao padrão de 14 kDa, compatível com a proteína clivada, e uma banda de aproximadamente 30 kDa, compatível com a proteína não clivada.

Durante o processo de concentração da proteína antes e depois da clivagem, foi observado um processo de precipitação massivo no concentrador. Assim, acreditamos que a proteína não é estável em solução, provavelmente devido a problemas no enovelamento. Além disso, analisando a eluição em 100 mM de imidazol (figura 71b), foi possível notar a presença de uma banda majoritária abaixo do padrão de 20 kDa que pode ser resultado de proteólise. Em outros experimentos de purificação foi visto que esse contaminante estava sempre presente. Para testar se

essa banda correspondia a um possível produto de degradação, amostras do pico referente à eluição em 100 mM foram submetidas a clivagem com a ULP1 (figura 72).

Figura 72 – Perfil eletroforético de migração das proteínas antes e depois da clivagem. 1) fração correspondente ao pico da eluição em 100 mM de imidazol de uma purificação típica do domínio CRD1c-sumo. 2) Fração 1 após clivagem com a ULP1, protease específica da SUMO.

Com base na análise comparativa das amostras, antes e depois da digestão com a ULP1, foi possível observar uma redução no tamanho da banda majoritária, sugerindo que parte do domínio CRD1c na construção CRD1c-1-pETsumo estava sendo alvo de proteólise. Em um novo ensaio de purificação utilizamos, além de PMSF, o coquetel de inibidores de proteases SigmaFAST™ sem EDTA (Sigma-Aldrich). Mesmo assim, o perfil de purificação resultou na presença da mesma banda com perfil de eluição em 100 mM de imidazol. Ainda cogitando a hipótese de que o ataque proteolítico pudesse ser resultado de ataque por metaloproteases e que a presença de EDTA pudesse ser essencial, um novo protocolo de purificação, sem a utilização da coluna de afinidade (por ser incompatível com a presença de EDTA), foi desenvolvido envolvendo a cromatografia de troca iônica (nas condições aniônicas e catiônicas) e de interação hidrofóbica. Entretanto, nas condições de carga e pH adequadas aos experimentos, a proteína permaneceu no precipitado.

Uma vez que o CRD2 da hGal12 possui apenas 21% de identidade sequencial com o CRD1c, supomos que o comportamento desse domínio poderia ser completamente diferente. Assim, procedemos com a otimização de códons para a isoforma 2, já que, por técnicas de clonagem poderíamos obter as demais isoformas e domínios. Foram obtidos os clones em BL21(DE3) para as isoformas 1 (Iso1-2-pETsumo), 2 (Iso2-3-pETsumo), 3 (Iso3-1-pETsumo) e 4 (Iso4-1-pETsumo) e para os domínios CRD1b (CRD1b-1-pETsumo) e CRD2 (CRD2-5-pETsumo). Foram realizados diversos ensaios de expressão, para as construções Iso2-3-pETsumo, Iso4-1-pETsumo, CRD1b-1-pETsumo e CRD2-5-pETsumo, nos quais observamos, de um modo geral, o mesmo perfil de insolubilidade.

Para testar se o CRD2 teria comportamento diferente devido às diferenças na sequência de aminoácidos, procedemos com um ensaio de purificação nas mesmas condições do CRD1c e observamos o mesmo padrão de precipitação e a presença da banda próxima de 20 kDa, em 100 mM de imidazol. Com base nesses resultados, concluímos que, mesmo com a otimização da expressão, as proteínas foram expressas majoritariamente como corpos de inclusão e a fração solúvel correspondia a um agregado solúvel, mas instável. Durante os ensaios de expressão das proteínas também foi observado uma diminuição significativa na quantidade final de células, provavelmente devido à grande quantidade de proteína formando corpos de inclusão.

A galectina-12 é uma proteína associada a gotas de lipídeos, organelas existentes em eucariotos e em alguns casos em procariotos em ambientes com baixa disponibilidade de nutrientes (Wältermann e Steinbüchel, 2005). Até o momento, a galectina-12 não foi encontrada no citosol das células e isso levanta a questão de para onde elas são direcionadas durante o processo de expressão em bactérias. Tomando como modelo de comparação o processo de superexpressão de proteínas de membranas em *E. coli*, traçamos um paralelo para o nosso problema de expressão. A superexpressão de proteínas de membrana em *E. coli* é altamente tóxica porque leva a produção de mais proteína do que o translocon Sec (conjunto de proteínas que translocam as proteínas para membrana) pode processar, tornando impossível a inserção na membrana das proteínas de membrana endógenas, levando a um acúmulo de proteínas no citosol, à agregação e indiretamente afetando a síntese de ATP (Wagner *et al.*, 2008).

Assim, é provável que a localização da galectina-12 em associação com gotas de lipídeos aliado a alta taxa de expressão proteica em bactérias promova a formação de proteínas enoveladas incorretamente, com consequente formação de corpos de inclusão. A fim de testar se a diminuição da expressão poderia ajudar na produção de proteínas solúveis, utilizamos duas diferentes cepas: Arctic express, para expressão em temperaturas entre 4 °C e 12 °C, e C41 (DE3), que possui uma mutação no promotor *lacUV5* que leva a uma diminuição nos níveis de expressão pela RNA T7 polimerase (Miroux e Walker, 1996; Wagner *et al.*, 2008).

Os testes de expressão foram realizados como descritos na seção 3.3.1.2. Para a expressão em Arctic express, nenhuma banda referente a proteína foi observada no precipitado ou sobrenadante. Já para a cepa C41 (DE3), houve expressão majoritária da proteína na fração insolúvel, com algumas quantidades nas frações solúveis a 18 °C por 24 horas. Para verificar melhora no padrão de estabilidade da proteína, por promoção de um enovelamento adequado, a proteína foi purificada após a expressão nessa cepa. O mesmo perfil de precipitação e proteólise foi observado.

Uma outra abordagem para melhorar a estabilidade e a solubilidade da proteína foi tentada através do uso de osmólitos durante o processo de expressão (Ignatova e Gierasch, 2006; Hailu, Foit e Bardwell, 2013). Assim, realizamos os testes de expressão na presença de glicina, sacarose, L-prolina e glicerol adicionados ao meio de cultura, para as construções Iso2-3-pETsumo, Iso4-1-pETsumo, CRD1b-1-pETsumo, CRD1c-1-pETsumo e CRD2-5-pETsumo. O resultado desse experimento não foi conclusivo devido à presença, no controle negativo, de uma banda forte no tamanho das proteínas de interesse. Um novo ensaio de expressão, baseado no trabalho de Ignatova e colaboradores, foi realizado para as proteínas na presença de 20 mM de L-prolina e 300 mM NaCl, com 500 μ M de IPTG, 25 °C durante 6 horas. Para evitar o vazamento da expressão, foi adicionado em todas as etapas da cultura 1 % (m/v) de glicose (figura 73). No gel foram aplicadas as amostras referentes ao sobrenadante após a etapa de lise celular.

Figura 73 – Perfil eletroforético do extrato proteico solúvel após lise celular e centrifugação. MM- Marcador molecular. As amostras aplicadas nos poços estão descritas na figura.

Foram observados para as construções Iso2-3-pETsumo e CRD2-5-pETsumo uma banda clara de expressão. Assim, com base nessas condições de expressão, foi realizada a purificação das duas proteínas em coluna de níquel. Nos dois casos, foi possível ver a banda de proteína solúvel pela análise do gel (figura 74a poço 2), entretanto, na análise do perfil de purificação, a eluição da proteína correspondente ao produto da proteólise é sempre visualizada nos géis de purificação. Além disso, as amostras que eluíram em 500 mM de imidazol, para o CRD2 (figura 74c poços 8-14), foram concentradas e o mesmo perfil de precipitação obtido para a construção CRD1c-1-pETsumo foi observado. Nesse caso, apesar da presença de proteínas solúveis em algumas condições, concluímos que não somente para a construção CRD1c-1-pETsumo mas também para a Iso2-3-pETsumo e CRD2-5-pETsumo foram observados comportamentos de instabilidade e agregação, reafirmando a hipótese de que as proteínas não enovelam apropriadamente.

Figura 74 – Perfil eletroforético da purificação em coluna de níquel para as construções Iso2-3-pETsumo e CRD2-5-pETsumo. MM) Marcador molecular. Gel A – 1) precipitado, 2) sobrenadante, 3) efluente, 4-8) tampão 0 de imidazol, 10-14) tampão 25 mM de imidazol; Gel B 1-7) tampão 50 mM de imidazol, 8-14) tampão 100 mM de imidazol; Gel C – 1-7) tampão 250 mM imidazol e 8-14) tampão 500 mM imidazol.

Condições desnaturantes

O processo de purificação de corpo de inclusão em condições amenas foi testado para a construção clonada em pET-29a(+), sem sucesso. Nessa etapa, procedemos com a purificação de corpos de inclusão, seguido de purificação desnaturante em acordo com o protocolo 2 descrito na seção 3.4.2.3. Foram utilizadas as construções CRD1c-3-pET29 e CRD1c-1-pETsumo para fins de avaliação de estabilidade.

Figura 75 – Perfil de purificação em coluna de níquel sob condições desnaturantes. A) CRD1c-3-pET29 B) CRD1c-1-pETsumo. 1) precipitado pós solubilização 2) corpo de inclusão solubilizado 3) fração 1 tampão 0 imidazol 4-6) frações 1, 3 e 5 tampão 25 mM imidazol 7-12) frações 1, 2, 3, 5, 7 e 10 tampão 200 mM imidazol 13) fração 1 500 mM imidazol 14) resina de níquel

Pela análise do gel de poliacrilamida em ambas as construções foi observado que a proteína solubilizou na presença de 8 M de ureia e 0,2 % SDS (figura 75 poço 2). A proteína desnaturada eluiu com sucesso da coluna de níquel, na presença de 200 mM de imidazol adicionado no tampão de purificação. As amostras referentes à eluição da proteína foram combinadas e submetidas ao processo de reenovelamento por diluição pulsada (Clark, 1998; Tsumoto *et al.*, 2003). Após a diluição em tampão de reenovelamento, um volume total de aproximadamente 200 ml de solução foi obtido. Essa quantidade foi utilizada para o processo de reenovelamento da proteína em baixas concentrações. A solução final de cada proteína foi lentamente aplicada em coluna de níquel, lavada com tampão de reenovelamento e eluída na presença de 25 mM de imidazol (figura 76). As frações foram concentradas a aproximadamente 1 mg/ml, devido a precipitação, e em seguida submetidas ao experimento de espalhamento dinâmica de luz – DLS.

Figura 76 – Purificação das construções CRD1c-3-pET29 e CRD1c-1-pETsumo em coluna de afinidade. MM – marcador molecular. CRD1c-3-pET29 1-6) frações 1, 2, 3, 4, 5 e 10 da eluição; 7) resina de níquel; CRD1c-1-pETsumo 8-13) frações 1, 2, 3, 4, 5 e 10 da eluição; 14) resina de níquel.

A análise do perfil de espalhamento mostrou para as duas construções um índice de polidispersividade de 42,9% e 34,8%, com raio hidrodinâmico de 1059 ± 455 e 245 ± 85 Å, para o CRD1c-3-pET29 e CRD1c-1-pETsumo, respectivamente, refletindo um conteúdo altamente agregado de proteínas (tabela 21).

Tabela 21 – Valores obtidos para os picos majoritários da distribuição de tamanho das partículas em solução

CRD1c-pET29a		CRD1c-pETsumo	
	diâmetro (Å)	% massa	
Pico 1	340	72,9	224,5
Pico 2	1299	27,1	-

3.7.3 Produção das macromoléculas em células de inseto a partir do cDNA

A produção das macromoléculas em células de inseto foi iniciada com a clonagem das isoformas 1, 2 e 4, obtidas do cDNA no vetor de clonagem pJET1.2/blunt e posterior subclonagem no vetor de transferência pFast-His. Em seguida, as cepas de *E. coli* DH10bacTM foram transformadas com os plasmídeos de transferência contendo a sequência das isoformas para promover o processo de transposição do cassete de expressão para o bacmídeo. Os bacmídeos recombinantes foram extraídos das células e utilizados para transfectar células de inseto Tn5b. As células transfectadas foram incubadas para a produção dos baculovírus recombinantes.

Os baculovírus recombinantes foram utilizados para o ensaio preliminar de expressão das proteínas de interesse em células de inseto. Amostras de cultura foram colhidas em diferentes tempos de incubação, 24, 48, 72 e 96 h, centrifugadas e submetidas à lise. A análise por SDS-PAGE das amostras, em diferentes tempos de expressão, indicou uma diferença no perfil de expressão proteica a partir de 72 horas, entretanto, não foram observadas bandas nas frações solúveis referente às proteínas de interesse.

3.8 Ensaio imunológicos

Apesar do insucesso na obtenção de proteína solúvel, a produção de anticorpos policlonais foi realizada por imunização dos coelhos com os domínios CRD1c e CRD2 produzidos a partir das construções CRD1c-3-pET29 e CRD2-pET29 e purificados como corpos de inclusão (protocolo 1). A presença de anticorpos anti-CRD1c e anti-CRD2 nos soros dos animais imunizados com essas proteínas foi avaliada pela imunodeteção dos respectivos domínios com o uso da técnica de *western blot*. O soro contendo os anticorpos policlonais, foi diluído nas proporções 1:10000, 1:20000, 1:40000, 1:80000, 1:160000, 1:320000, 1:640000 e 1:1280000 e testado contra os domínios CRD1c e CRD2 transferidos para uma membrana de nitrocelulose (figura 77).

Figura 77 – Imunodeteccção por *western blot* das proteínas recombinantes CRD1c e CRD2 provenientes das construções CRD1c-3-pET29 e CRD2-pET29, respectivamente. De 1 a 8 diluições de 1:10000, 1:20000, 1:40000, 1:80000, 1:160000, 1:320000, 1:640000 e 1:1280000.

Os dados ilustrados na figura 77 mostram que com o uso de diluições altíssimas (1:1280000 e 1:640000) dos soros-ímenes anti-CRDc1 ou anti-CRD2 foi possível detectar nos corpos de inclusão purificados que continham os domínios da hGal12, uma banda proteica com massa molecular aparente de 15 kDa, correspondente ao esperado para as formas recombinantes do CRD1c e do CRD2. Entretanto, mesmo com o uso de altas diluições de ambos soros-ímenes, foi detectada uma banda proteica de aproximadamente 30 kDa, não correspondente à massa molecular destes domínios da hGal12. Este fato, provavelmente, pode estar associado a presença de contaminantes de *E. coli* contidos nos corpos de inclusão purificados utilizados na etapa de imunização dos coelhos.

Após a etapa de titulação, os soros-ímenes foram submetidos a purificação em resinas de agarose-proteína G. Os resultados mostraram que os anticorpos policlonais anti-CRD2 purificados, diferentemente dos soros-ímenes, não foram eficientes em detectar as isoformas ou o domínio CRD2 da hGal12 em extratos bacterianos (dados não mostrados). Além disso, estes anticorpos purificados apresentaram reação cruzada para as formas purificadas e recombinantes das galectinas 4 e 3 humanas (dados não mostrados). Com base nesses resultados, foi adquirido da Santa Cruz *Biotechnology* o anticorpo anti-galectina12 humana produzido em cabras e dirigido a um peptídeo sintético localizado no domínio CRD2 desta galectina. Este anticorpo comercial foi testado contra o extrato proteico proveniente da expressão de células BL21 (DE3) não transformadas e transformadas (não induzidas e induzidas 500 μ M de IPTG) com os plasmídeos recombinantes das diferentes construções Iso2-3-pETsumo, Iso4-1-pETsumo, CRD2-5-pETsumo, Iso2-2-pET29, CRD2-pET29. Como mostrado na figura 78, o anticorpo comercial anti-peptídeo sintético da hGal12 mostrou reatividade para as seguintes preparações: Iso2-2-pET29 (figura 78 poço 3), Iso2-3-pETsumo (figura 78 poço 8), Iso4-1-pETsumo (figura 78 poço 10) e CRD2-5-pETsumo (figura 78 poço 12).

Figura 78 – Análise eletroforética (SDS-PAGE) e a reação de *western blot* dos extratos proteicos de bactérias submetidos a imunodeteção com o anticorpo policlonal anti-peptídeo sintético da hGal12. MM) marcador molecular; 1) induzido e 2) não induzido CRD2-pET29; 3) induzido e 4) não induzido Iso2-2-pET29; 5-6) BL21(DE3) sem plasmídeo; 7) não induzido e 8) induzido Iso2-3-pETsumo; 9) não induzido e 10) induzido Iso4-1-pETsumo; 11) não induzido e 12) induzido CRD2-5-pETsumo.

Após a constatação da eficiência da imunorreatividade do anticorpo anti-hGal12 comercial para as isoformas (2 e 4) e para o CRD2 da hGal12, foi avaliado o nível de reação deste anticorpo para outros membros da família das galectinas [hGal1 (galectina-1 humana), hGal3 (galectina-3 humana) e hGal4 (galectina-4 humana)] em comparação com a hGal12 (Iso2-3-pETsumo) nas concentrações de 1000, 500 e 250 ng (figura 79).

Figura 79 – Análise da imunorreatividade do anticorpo comercial anti-peptídeo da hGal12 para outras galectinas humanas. As diferentes preparações proteicas foram submetidas a eletroforese em gel de poliacrilamida (SDS-PAGE) e eletrotransferidas para membranas de nitrocelulose. As amostras utilizadas foram: MM) mistura de marcadores de massa molecular composta por proteínas biotinizadas; 1) hGal1, 2) hGal3, 3) hGal4, 4) hGal12, 5) hGal1, 6) hGal3, 7) hGal4, 8) hGal12, 9) hGal1, 10) hGal3, 11) hGal4 e 12) hGal12. Nos poços de 1 a 4 foram utilizados 100 ng, nos poços de 5 a 12 foram utilizadas 500 ng e 250 ng de cada proteína, respectivamente.

Como obtido para o anticorpo policlonal purificado do soro-imune anti-CDR2, o anticorpo policlonal anti-peptídeo sintético da hGal12 também reconheceu as galectinas humanas -3 e -4 recombinantes (figura 79).

Apesar da reação cruzada do anticorpo comercial anti-hGal12 com a hGal3 e a hGal4, foram realizados os experimentos de localização celular da hGal12 com este anticorpo comercial. Esta decisão foi fundamentada nos seguintes aspectos: *i*) a galectina-12 de camundongos foi descrita como o membro da família destas proteínas que está associada à gotas de lipídeos intracelulares (Yang *et al.*, 2011), diferentemente da localização celular para a hGal3 e hGal4 e *ii*) uso simultâneo do anticorpo comercial anti-hGal12 e do BODIPY, um marcador de lipídeos neutros.

A avaliação preliminar da localização celular da hGal12 em células HL-60 foi realizada por microscopia confocal. Nesse experimento, foram utilizados os marcadores DAPI para o núcleo (marcação em azul), BODIPY para gotas de lipídeo (marcação em verde) e o anticorpo comercial anti-hGal12 (marcação em vermelho). O experimento foi realizado como descrito na seção 3.5.3.

Primeiramente foram ajustados os parâmetros de aquisição das imagens, como o ganho e a porcentagem do laser para evitar a detecção do “vazamento” da fluorescência entre os canais no azul, no vermelho e no verde (figura 80). Em seguida, as imagens foram adquiridas (figuras 81 e 82).

Figura 80 – Determinação das condições experimentais do uso do laser para a análise da microscopia confocal A) Sobreposição dos canais na fluorescência no azul para o marcador DAPI e no vermelho para o anticorpo secundário AlexaFluor® 546 anti-IgG de cabra. B) Sobreposição dos canais na fluorescência no azul para o marcador DAPI e no verde para o marcador BODIPY.

A galectina-12 mostrou uma localização bem definida, esférica e compacta nas células HL-60, compatível com a marcação para gotas de lipídeos com o BODIPY (figura 81).

Figura 81 – Composição das imagens de microscopia confocal em células HL-60. A) Sobreposição das imagens com DAPI e BODIPY. B) Sobreposição das imagens com DAPI e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 no vermelho.

Para estabelecer a localização celular da galectina-12 humana em relação as gotas de lipídeos citoplasmáticas foram feitas capturas das imagens de fluorescência em diferentes seções do plano z para compor uma imagem tridimensional (figura 82).

Figura 82 – Imagens tridimensionais de microscopia confocal em células HL-60. A) Sobreposição das imagens com DAPI (marcação em azul) e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 (marcação em vermelho). B) Sobreposição das imagens com BODIPY (marcação verde) e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 (marcação em vermelho).

A análise da imagem contida na figura 82b mostra que a galectina-12 recobre toda a superfície das gotas de lipídeos, em uma distribuição aparentemente homogênea. Além disso, a galectina-12 humana parece estar localizada apenas nas gotas de lipídeos, como descrito para a galectina-12 de camundongo (Yang *et al.*, 2011). A figura 83 mostra uma amplificação da imagem

ilustrada na figura-82b referente as marcações do BODIPY (verde) e da Gal12 humana (vermelho). A análise desta imagem reforça a hipótese da distribuição homogênea, superficial e exclusiva da Gal12 humana na estrutura da gota de lipídeo localizadas no citosol das células HL-60.

Figura 83 – Amplificação da imagem B da figura 82 da fusão das imagens com BODIPY e anticorpo primário anti-gal12 com fluorescência do secundário AlexaFluor® 546 no vermelho.

Até o momento, parece que a marcação celular com o anticorpo comercial contra a hGal12 não apresentou reação cruzada com hGal3 e hGal4, como mostrado nos ensaios de *western blot*. Os achados preliminares ilustrados nas figuras 81 e 82 sugerem que a galectina-12 humana endógena está exclusivamente localizada nas gotas de lipídeos, uma característica biológica descrita apenas para esta galectina (Yang *et al.*, 2011). Entretanto, ensaios de inibição destas marcações com diferentes galectinas (hGal3, hGal4 e hGal12) serão feitos para testar esta proposição, como descrito para outras situações experimentais envolvendo a localização tecidual seletiva de galectinas (Dias-Baruffi *et al.*, 2010). Ensaios de *western blot* com o lisado das células HL-60 estão em andamento e permitirão identificar quais isoformas estão presentes nessas células. Adicionalmente, estudos de diferenciação das células serão realizados com o objetivo de estudar o impacto da hGal12 no processo de diferenciação de leucócitos a partir de células precursoras.

3.9 Modelagem molecular

Como resultado da modelagem comparativa por homologia, obtivemos os modelos das estruturas dos domínios CRD1a, CRD1b, CRD1c e CRD2. Os resultados sugerem que o

enovelamento global é conservado e apresenta um domínio do tipo β -sanduíche constituído de duas folhas betas antiparalelas (figura 84).

Figura 84 – Representação esquemática do enovelamento dos CRDs das isoformas 1, 2, 3, 4 e 5 descritas para a galectina-12 humana. CRD1a – CRD1 da isoforma 1; CRD1b – CRD1 das isoformas 2 e 3; CRD1c – CRD1 da isoforma 4; CRD2 – CRD2 para as isoformas 1, 2, 3, 4 e 5.

A principal característica que as difere quanto a sua topologia é a presença de uma fita a mais, F0, na segunda folha para os domínios CRD1a e CRD1b, e da ausência de uma fita, S1, na primeira folha para o domínio CRD1c, quando comparado com a topologia das galectinas do grupo *tandem-repeat*.

Uma análise da superfície de potencial eletrostático calculada para os diferentes CRDs não nos permitiu evidenciar nenhuma característica que pudesse justificar a instabilidade desta proteína, como por exemplo, um grande domínio hidrofóbico (figura 85). Ao contrário disso, vemos grandes porções carregadas positivamente, característica já vista para os CRDs da galectina-4 que são altamente solúveis.

Além disso, os modelos construídos mostraram que, embora as substituições encontradas (figura 84) não afetem a estrutura tridimensional dos CRDs, os mesmos induzem modificações importantes na superfície da proteína, podendo ser uma das características chaves para a seletividade de reconhecimento de cada CRD por ligantes e/ou localização celular.

Figura 85 – Representação da superfície de potencial eletrostático para os diferentes CRDs: CRD1a, CRD1b, CRD1c e CRD2 nas posições $y=0^\circ$, 90° , 180° e 270° .

CONCLUSÕES

O estudo da galectina-12 humana foi iniciado pela obtenção do cDNA fita simples a partir de leucócitos totais humanos. Diferentes estratégias foram adotadas ao longo do desenvolvimento desse projeto no intuito de obter as isoformas ou os domínios da hGal12 em uma forma solúvel. A primeira dificuldade encontrada foi expressar a proteína em bactérias. A tentativa em obter expressão da proteína foi voltada para a clonagem dos domínios separadamente, já que a presença do peptídeo de ligação poderia influenciar a expressão. Diante do insucesso encontrado, testamos a hipótese de que o vetor pTYB2 pudesse ter sofrido uma mutação e clonamos um dos CRDs, o CRD1a, em pET-28a. A grande quantidade de códons raros para *E. coli*, bem como a formação de clusters de códons raros impediu a expressão das proteínas, mesmo em cepas contendo o RNA transportador para 7 dos 10 códons raros encontrados para *E. coli*. Essa hipótese foi testada primeiramente pela síntese do menor domínio, CRD1c, com códons otimizados para *E. coli*. A clonagem do gene sintético foi feita no vetor pTYB2 permitindo a produção da proteína em fusão com uma cauda de inteína. O teste de expressão a 37 °C mostrou a produção da proteína nas frações insolúveis e confirmou a hipótese da ausência de expressão observada anteriormente, devido à presença e disposição na sequência dos códons raros.

Vários testes de expressão foram realizados, pela variação das concentrações do agente indutor, temperatura e tempo de expressão. Como resultado, obtivemos uma condição em que o domínio foi expresso na forma solúvel. Apesar do resultado aparentemente promissor, a etapa de purificação revelou dois problemas concomitantes. A maior parte da amostra da proteína era sujeita a proteólise na extremidade C-terminal e a pequena quantidade solúvel resultou em proteínas altamente agregadas. Como a proteólise ocorreu especificamente na região da fusão, decidimos trocar o vetor de expressão. Assim, o CRD1c foi clonado em pET-29a e foi dado início aos testes de expressão e purificação. Paralelamente, a isoforma 2 foi submetida a otimização de códons da mesma forma que o CRD1c. A partir da otimização dos códons para a isoforma 2, foi possível obter as construções para o CRD1b e o CRD2 além do CRD1c. As tentativas de obter proteína na forma solúvel foram frustradas para essas construções e foi observado que as amostras eram expressas na forma de corpos de inclusão.

Na tentativa de resolver os problemas de expressão enfrentados para a hGal12, diferentes estratégias foram utilizadas. Primeiramente buscamos entender a razão para tal insolubilidade através do uso de ferramentas computacionais e técnicas de modelagem molecular por homologia. Tentamos prever características estruturais que pudessem justificar a formação de corpos de

inclusão como, por exemplo, a identificação de peptídeo sinal, grandes superfícies hidrofóbicas ou sítios alvo para modificação pós-traducional.

Uma das razões para acreditar que a hGal12 poderia apresentar uma superfície hidrofóbica se baseia no fato de que um trabalho utilizando a galectina-12 de camundongo mostrou sua expressão preferencial em adipócitos e localização celular na superfície de gotas de lipídios.

Os modelos estruturais foram construídos para os diferentes CRDs identificados para as isoformas (CRD1a – CRD1 das isoformas 1; CRD1b – CRD1 das isoformas 2 e 3; CRD1c – CRD1 da isoforma 4; CRD2 – CRD2 para as isoformas 1, 2, 3, 4 e 5) e mostraram que, assim como para outras galectinas, a arquitetura do CRD em um β -sanduíche é mantida intacta. Além disso, a superfície de potencial eletrostático calculada para os diferentes CRDs, não nos permitiu evidenciar nenhuma característica estrutural que pudesse justificar a grande instabilidade desta proteína em permanecer na fração solúvel. Decidimos adotar duas estratégias distintas na tentativa de obter a proteína ou seus domínios na forma solúvel. Considerando a hipótese da proteína sofrer alguma mudança pós-traducional, não realizada pela bactéria, seja a remoção de um peptídeo sinal não predito ou a adição de grupos funcionais como acetato, fosfato, lipídeos e carboidratos, decidimos testar um sistema para expressão heteróloga mais desenvolvido como a expressão em células de inseto utilizando recombinação em baculovírus. Esses experimentos resultaram na presença de proteínas insolúveis e aparentemente com a presença de ataque proteolítico.

Testamos também a hipótese de que a insolubilidade das proteínas de interesse pudesse ser causada por alguma dificuldade no processo de enovelamento, ou mesmo agregação, causada por interação inespecífica do tipo proteína-proteína. Para minimizar estes efeitos, decidimos por testar novas construções onde as proteínas estariam fusionadas à uma proteína conhecida por aumentar a solubilidade de seus parceiros de fusão, a proteína SUMO, clonada no vetor pET28a, dando origem ao vetor pET-28a-sumo. Para essa construção foram clonadas as isoformas 1, 2, 3 e 4 e os domínios CRD1b, CRD1c e CRD2. As isoformas 1 e 3 foram clonadas recentemente, assim, não foram realizados experimentos de expressão e purificação das mesmas.

Vários esforços foram realizados na tentativa de contornar os problemas de instabilidade e formação de corpos de inclusão encontrados ao longo dos testes de expressão e purificação realizados. Estratégias como o uso de aditivos na purificação, além dos ensaios de expressão na presença de osmólitos, não resultaram em uma melhora na estabilidade da proteína. Uma vez que a formação de corpos de inclusão é consequência da alta taxa de expressão proteica, realizamos experimentos que favorecessem a diminuição da taxa de expressão, por temperatura, com uma cepa que permite a expressão em temperaturas entre 4 e 12 °C, e pelo uso de uma cepa com o

promotor *lacUV5* mutado, resultando em uma diminuição da atividade da T7 RNA polimerase. Em nenhuma das estratégias obtivemos um resultado satisfatório.

Como uma alternativa na obtenção de proteína solúvel, testamos a purificação de corpos de inclusão seguida de desnaturação e reenovelamento. Apesar de um alto rendimento no processo de purificação, o reenovelamento não foi alcançado, devido à precipitação e instabilidade da amostra. Além disso, medidas de DLS para as amostras reenoveladas revelaram um perfil de polidispersividade e um diâmetro compatível com o de proteínas agregadas.

Ainda assim, para dar prosseguimento ao projeto, o ensaio de imunização em coelhos foi iniciado com os corpos de inclusão, uma vez que a imunização em camundongos com o gene *LGALS12* depletado havia sido feita a partir de corpos de inclusão da galectina-12 (Yang *et al.*, 2011). Apesar de verificado um alto título na produção dos anticorpos durante as etapas de imunização, o anticorpo purificado em coluna de proteína G se mostrou bastante inespecífico. Esse resultado foi visto no experimento de *western blot* com lisado proteico de *E. coli*. A marcação das bandas para o controle negativo foi o mesmo que na presença das proteínas expressas. Além disso, um ensaio com as diferentes galectinas purificadas revelou que o anticorpo anti-CRD2 foi muito mais específico para a hGal4 e em menor grau para a hGal3 do que para a hGal12. Nesse sentido, adquirimos o anticorpo primário anti-gal12 da empresa Santa Cruz *Biotechnology*. O experimento de *western blot* para o lisado proteico de *E. coli* foi bastante satisfatório, pelo reconhecimento claro das isoformas e CRDs testados. Entretanto, o teste com proteínas purificadas mostrou que também ocorre reação cruzada contra a hGal3 e hGal4. Ainda assim, demos início aos ensaios de imunodeteção da hGal12 em células HL-60 utilizando microscopia confocal. Uma análise preliminar mostrou que a galectina-12 está associada a gotas de lipídeos presentes nas células HL-60 de forma a envolver toda a superfície da gota de lipídeo. Esses resultados confirmam que mesmo em células diferentes de adipócitos, a galectina-12 está associada a gotas de lipídeos. O seu papel biológico ainda não está elucidado e muitas questões permanecem ainda sem resposta.

De um modo geral, as galectinas formam uma família de proteínas que ligam a carboidratos e são estruturalmente conservadas. Os membros das galectinas são amplamente distribuídos por todo o domínio dos eucariotos, presente desde organismos primitivos como esponjas e nematoides até fungos e animais superiores (Kasai e Hirabayashi, 1996; Houzelstein *et al.*, 2004; Yang, Rabinovich e Liu, 2008; Vasta *et al.*, 2012). Sua presença em espécies anciãs sugere o desempenho de um papel fundamental na biologia celular, enquanto que a presença de múltiplas galectinas dentre de uma única espécie indica que elas divergiram para atuar especificamente em uma variedade de funções possivelmente não correlatas (Cooper, 2002).

Em particular, o favorecimento evolutivo da formação de galectinas bivalentes ou por associação dos monômeros ou pela organização como *tandem-repeat* enfatiza a importância da valência nessas proteínas, sugerindo que essa característica seja um aspecto crítico na associação com glicoconjugados (Cooper, 2002). Uma questão interessante e recorrente é a regulação do comprimento dos peptídeos de ligação que une os CRDs nas galectinas do grupo *tandem-repeat*. Essa característica sugere que essas sequências podem ter uma função maior ou na regulação da orientação dos domínios um em relação ao outro ou por prover um sítio de interação com outras proteínas (Cooper, 2002).

Uma visão mais detalhada da organização estrutural das galectinas permite observar que há uma organização bastante conservada. As regiões de *loops* entre as fitas S5 e S6 e entre S6 e F3 são bastante conservadas. Além disso, a maioria das moléculas estão envolvidas diretamente na ligação ao carboidrato, incluindo os resíduos conservados nas fitas- β S4, S5 e S6. Por outro lado, o comprimento dos *loops* entre as fitas S3-S4 e S4-S5 variam significativamente de acordo com os diferentes grupos de galectinas, indicando que esses *loops* podem estar envolvidos na especificidade refinada ao carboidrato (Houzelstein *et al.*, 2004).

Sete resíduos chave estão envolvidos na interação com β -galactosídeos. Esses resíduos estão presentes na maioria dos CRDs. Interessantemente, o C-terminal da galectina-12 assim como para a proteína HSPC159 e a galectina-10, não apresentam vários desses resíduos, sugerindo pouca ou nenhuma atividade β -galactosídica (Dyer e Rosenberg, 1996; Houzelstein *et al.*, 2004). Nesse sentido, embora a galectina-12 seja bivalente (apresente dois CRDs), a sua ação como uma galectina bivalente ainda não foi experimentalmente avaliada.

É possível que o domínio C-terminal da galectina-12 esteja envolvido em outros processos de reconhecimento não necessariamente relacionados ao reconhecimento de glicanos. Uma vez que foi observado que as galectinas proto evoluíram, rapidamente duplicaram e divergiram para formar galectinas *tandem-repeat* e clusters de genes similares (Cooper, 2002) e que alguns desses genes parecem ter perdido a atividade de ligação a galactosídeos evoluindo talvez através do reconhecimento de outros carboidratos ou mesmo outros tipos de ligantes, a nossa hipótese de que a galectina-12 possa participar de outros tipos de interações é plausível.

Com base em nossos experimentos, devido a insolubilidade da hGal12, a sua localização em gotas de lipídeos associado à ausência de peptídeo sinal, a divergência evolutiva e as substituições na sequência de interação com β -galactosídeos no domínio C-terminal, acreditamos que a galectina-12 pode possuir um parceiro biológico como uma proteína ou ligante, que propicie o seu enovelamento e funcionamento corretos dentro da célula, além de explicar o seu direcionamento para gotas de lipídeos. Experimentos adicionais são necessários para entender o

quadro geral do papel biológico dessa proteína. Portanto, demos os primeiros passos na caracterização da galectina-12.

Importante ressaltar que, devido à complexidade da família das galectinas, mesmo em uma única espécie, uma atenção criteriosa deve ser dada durante a caracterização das galectinas e suas variantes, no intuito de compreender melhor o papel dessa classe de proteínas, não só fisiologicamente, mas evolutivamente.

CAPÍTULO 4

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. **Acta Crystallographica Section D-Biological Crystallography**, v. 66, p. 213-221, Feb 2010.

AHMAD, N. et al. Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. **J Biol Chem**, v. 279, n. 12, p. 10841-7, Mar 2004.

AHMED, H.; BANERJEE, P. P.; VASTA, G. R. Differential expression of galectins in normal, benign and malignant prostate epithelial cells: silencing of galectin-3 expression in prostate cancer by its promoter methylation. **Biochem Biophys Res Commun**, v. 358, n. 1, p. 241-6, Jun 2007.

AHMED, H. et al. Evidence of heavy methylation in the galectin 3 promoter in early stages of prostate adenocarcinoma: development and validation of a methylated marker for early diagnosis of prostate cancer. **Transl Oncol**, v. 2, n. 3, p. 146-56, Aug 2009.

ANDRÉ, S. et al. Combining glycocluster synthesis with protein engineering: an approach to probe into the significance of linker length in a tandem-repeat-type lectin (galectin-4). **Carbohydr Res**, v. 389, p. 25-38, May 2014.

ARAKAWA, T.; TIMASHEFF, S. N. The stabilization of proteins by osmolytes. **Biophys J**, v. 47, n. 3, p. 411-4, Mar 1985.

ARNAUD, J.; AUDFRAY, A.; IMBERTY, A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. **Chem Soc Rev**, v. 42, n. 11, p. 4798-813, Jun 2013.

AUB, J. C.; SANFORD, B. H.; COTE, M. N. Studies on reactivity of tumor and normal cells to a wheat germ agglutinin. **Proc Natl Acad Sci U S A**, v. 54, n. 2, p. 396-9, Aug 1965.

BAKER, N. et al. Electrostatics of nanosystems: Application to microtubules and the ribosome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 18, p. 10037-10041, AUG 28 2001 2001.

BARONDES, S. H. et al. GALECTINS - STRUCTURE AND FUNCTION OF A LARGE FAMILY OF ANIMAL LECTINS. **Journal of Biological Chemistry**, v. 269, n. 33, p. 20807-20810, 1994.

BAS, D. C.; ROGERS, D. M.; JENSEN, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. **Proteins**, v. 73, n. 3, p. 765-83, Nov 2008.

BELO, A. I. et al. Galectin-4 Reduces Migration and Metastasis Formation of Pancreatic Cancer Cells. **PLoS One**, v. 8, n. 6, p. e65957, 2013.

BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684-3690, 1984.

BERGFORS, T. Seeds to crystals. **J Struct Biol**, v. 142, n. 1, p. 66-76, Apr 2003.

_____. Succeeding with seeding: some practical advice. In: READ, R. e SUSSMAN, J. (Ed.). **Evolving Methods for Macromolecular Crystallography**: Springer Netherlands, v.245, 2007. cap. 1, p.1-10. (NATO Science Series). ISBN 978-1-4020-6314-5.

BI, S. et al. Structural features of galectin-9 and galectin-1 that determine distinct T cell death pathways. **J Biol Chem**, v. 283, n. 18, p. 12248-58, May 2008.

_____. Galectin-9 binding to cell surface protein disulfide isomerase regulates the redox environment to enhance T-cell migration and HIV entry. **Proceedings of the National Academy of Sciences**, v. 108, n. 26, p. 10650-10655, 2011.

BIDON, N. et al. Two messenger RNAs and five isoforms for Po66-CBP, a galectin-8 homolog in a human lung carcinoma cell line. **Gene**, v. 274, n. 1-2, p. 253-262, 8/22/ 2001.

BOIVIN, S.; KOZAK, S.; MEIJERS, R. Optimization of protein purification and characterization using ThermoFluor screens. **Protein Expr Purif**, v. 91, n. 2, p. 192-206, Oct 2013.

BOURNE, Y. et al. Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. **Nat Struct Biol**, v. 1, n. 12, p. 863-70, Dec 1994.

BOYD, W. C.; SHAPLEIGH, E. Specific Precipitating Activity of Plant Agglutinins (Lectins). **Science**, v. 119, n. 3091, p. 419, Mar 1954.

BREWER, C. F.; MICELI, M. C.; BAUM, L. G. Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. **Current Opinion in Structural Biology**, v. 12, n. 5, p. 616-623, 10/1/ 2002.

BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **J Chem Phys**, v. 126, n. 1, p. 014101, Jan 2007.

BUTT, T. R. et al. SUMO fusion technology for difficult-to-express proteins. **Protein Expr Purif**, v. 43, n. 1, p. 1-9, Sep 2005.

CALIFICE, S. et al. Dual activities of galectin-3 in human prostate cancer: tumor suppression of nuclear galectin-3 vs tumor promotion of cytoplasmic galectin-3. **Oncogene**, v. 23, n. 45, p. 7527-36, Sep 2004.

CAMBY, I. et al. Galectin-1 modulates human glioblastoma cell migration into the brain through modifications to the actin cytoskeleton and levels of expression of small GTPases. **J Neuropathol Exp Neurol**, v. 61, n. 7, p. 585-96, Jul 2002.

CAMPBELL, C. T.; YAREMA, K. J. Large-scale approaches for glycobiology. **Genome Biol**, v. 6, n. 11, p. 236, 2005.

CERLIANI, J. et al. Expanding the Universe of Cytokines and Pattern Recognition Receptors: Galectins and Glycans in Innate Immunity. **Journal of Clinical Immunology**, v. 31, n. 1, p. 10-21, 2011/02/01 2011.

CHABOT, S. et al. Regulation of galectin-9 expression and release in Jurkat T cell line cells. **Glycobiology**, v. 12, n. 2, p. 111-8, Feb 2002.

CHAYEN, N. E. The role of oil in macromolecular crystallization. **Structure**, v. 5, n. 10, p. 1269-74, Oct 1997.

CHEN, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallogr D Biol Crystallogr**, v. 66, n. Pt 1, p. 12-21, Jan 2010.

CHIU, M. L.; JONES, J. C.; O'KEEFE, E. J. Restricted tissue distribution of a 37-kD possible adherens junction protein. **J Cell Biol**, v. 119, n. 6, p. 1689-700, Dec 1992.

CHIU, M. L. et al. An adherens junction protein is a member of the family of lactose-binding lectins. **J Biol Chem**, v. 269, n. 50, p. 31770-6, Dec 1994.

CHONG, S. R. et al. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. **Gene**, v. 192, n. 2, p. 271-281, 1997.

CICCARONE, V. C.; POLAYES, D. A.; LUCKOW, V. A. Generation of Recombinant Baculovirus DNA in E.coli Using a Baculovirus Shuttle Vector. In: (Ed.). **Molecular Diagnosis of Infectious Diseases**, v.13, 1997. p.213-235. (Methods in Molecular Medicine).

CILIBERTI, E. et al. [Tolerance and autoimmunity. Novel therapeutic approaches]. **Medicina (B Aires)**, v. 69, n. 4, p. 460-5, 2009.

CLARK, E. D. B. Refolding of recombinant proteins. **Curr Opin Biotechnol**, v. 9, n. 2, p. 157-63, Apr 1998.

CLAUSSE, N. et al. Galectin-1 expression in prostate tumor-associated capillary endothelial cells is increased by prostate carcinoma cells and modulates heterotypic cell-cell adhesion. **Angiogenesis**, v. 3, n. 4, p. 317-25, 1999.

COOPER, D. Galectinomics: finding themes in complexity. **Biochimica Et Biophysica Acta-General Subjects**, v. 1572, n. 2-3, p. 209-231, SEP 19 2002 2002. ISSN 0304-4165.

CORPET, F. Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Res**, v. 16, n. 22, p. 10881-90, Nov 1988.

CROCI, D. O. et al. Dynamic cross-talk between tumor and immune cells in orchestrating the immunosuppressive network at the tumor microenvironment. **Cancer Immunol Immunother**, v. 56, n. 11, p. 1687-700, Nov 2007.

CUMMINGS, M. D.; FARNUM, M. A.; NELEN, M. I. Universal screening methods and applications of ThermoFluor. **J Biomol Screen**, v. 11, n. 7, p. 854-63, Oct 2006.

DANGUY, A.; CAMBY, I.; KISS, R. Galectins and cancer. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1572, n. 2-3, p. 285-293, 9/19/ 2002.

DANIELSEN, E. M.; HANSEN, G. H. Lipid raft organization and function in the small intestinal brush border. **J Physiol Biochem**, v. 64, n. 4, p. 377-82, Dec 2008.

DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089-10092, 1993.

DAVICINO, R. C. et al. Coupling pathogen recognition to innate immunity through glycan-dependent mechanisms. **International Immunopharmacology**, v. 11, n. 10, p. 1457-1463, 10//2011.

DAVIS, R. E. A new molecular biology bible? Current protocols in molecular biology (1993). Edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J. G. Setdman, John A. Smith and Kevin Struhl. Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. 2200+ pp. ISBN 0-471-50338-X (vols 1 & 2 set). ISBN 0-471-50337-1 (vol. 2 binder). \$415. Update service \$170 p.a. **BioEssays**, v. 15, n. 9, p. 635-635, 1993.

DAZORD, L. et al. A monoclonal antibody (Po66) directed against human lung squamous cell carcinoma immunolocalization of tumour xenografts in nude mice. **Cancer Immunol Immunother**, v. 24, n. 3, p. 263-8, 1987.

DELACOUR, D. et al. Galectin-4 and sulfatides in apical membrane trafficking in enterocyte-like cells. **J Cell Biol**, v. 169, n. 3, p. 491-501, May 2005.

DELANO, W. L. Use of PYMOL as a communications tool for molecular science. **Abstracts of Papers of the American Chemical Society**, v. 228, p. U313-U314, Aug 22 2004.

DELGADO, V. M. et al. Modulation of endothelial cell migration and angiogenesis: a novel function for the "tandem-repeat" lectin galectin-8. **FASEB J**, v. 25, n. 1, p. 242-54, Jan 2011.

DI LELLA, S. et al. Critical role of the solvent environment in galectin-1 binding to the disaccharide lactose. **Biochemistry**, v. 48, n. 4, p. 786-91, Feb 2009.

_____. Characterization of the galectin-1 carbohydrate recognition domain in terms of solvent occupancy. **J Phys Chem B**, v. 111, n. 25, p. 7360-6, Jun 2007.

_____. When galectins recognize glycans: from biochemistry to physiology and back again. **Biochemistry**, v. 50, n. 37, p. 7842-57, Sep 2011.

DIAS-BARUFFI, M. et al. Differential expression of immunomodulatory galectin-1 in peripheral leukocytes and adult tissues and its cytosolic organization in striated muscle. **Glycobiology**, v. 20, n. 5, p. 507-20, May 2010..

DISKIN, S. et al. The role of integrin glycosylation in galectin-8-mediated trabecular meshwork cell adhesion and spreading. **Glycobiology**, v. 19, n. 1, p. 29-37, Jan 2009.

DOLINSKY, T. J. et al. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. **Nucleic Acids Res**, v. 32, n. Web Server issue, p. W665-7, Jul 2004.

DRICKAMER, K. Two distinct classes of carbohydrate-recognition domains in animal lectins. **J Biol Chem**, v. 263, n. 20, p. 9557-60, Jul 1988.

DUERR, E. M. et al. Defining molecular classifications and targets in gastroenteropancreatic neuroendocrine tumors through DNA microarray analysis. **Endocr Relat Cancer**, v. 15, n. 1, p. 243-56, Mar 2008.

DYER, K. D.; ROSENBERG, H. F. Eosinophil Charcot-Leyden crystal protein binds to beta-galactoside sugars. **Life Sci**, v. 58, n. 23, p. 2073-82, 1996.

EARL, L. A.; BI, S.; BAUM, L. G. Galectin multimerization and lattice formation are regulated by linker region structure. **Glycobiology**, v. 21, n. 1, p. 6-12, Jan 2011.

ECHEVERRIA, I.; AMZEL, L. M. Disaccharide binding to galectin-1: free energy calculations and molecular recognition mechanism. **Biophys J**, v. 100, n. 9, p. 2283-92, May 2011.

ELLERHORST, J. et al. Differential expression of endogenous galectin-1 and galectin-3 in human prostate cancer cell lines and effects of overexpressing galectin-1 on cell phenotype. **Int J Oncol**, v. 14, n. 2, p. 217-24, Feb 1999.

ELOLA, M. T. et al. Galectins: matricellular glycan-binding proteins linking cell adhesion, migration, and survival. **Cellular and Molecular Life Sciences**, v. 64, n. 13, p. 1679-1700, 2007/07/01 2007.

EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica Section D-Biological Crystallography**, v. 66, p. 486-501, Apr 2010.

ESPELT, M. V. et al. Novel roles of galectin-1 in hepatocellular carcinoma cell adhesion, polarization, and in vivo tumor growth. **Hepatology**, v. 53, n. 6, p. 2097-106, Jun 2011.

EVANS, P. Scaling and assessment of data quality. **Acta Crystallographica Section D-Biological Crystallography**, v. 62, p. 72-82, Jan 2006.

EVANS, P. R.; MURSHUDOV, G. N. How good are my data and what is the resolution? **Acta Crystallographica Section D**, v. 69, n. 7, p. 1204-1214, 07/01 2013.

FARESE, R. V.; WALTHER, T. C. Lipid droplets finally get a little R-E-S-P-E-C-T. **Cell**, v. 139, n. 5, p. 855-60, Nov 2009.

FASMAN, G. D. **Practical handbook of biochemistry and molecular biology**. Boca Raton, Fla: CRC Press, 1989. 601.

FASSHAUER, M. et al. Negative regulation of adipose-expressed galectin-12 by isoproterenol, tumor necrosis factor alpha, insulin and dexamethasone. **European Journal of Endocrinology**, v. 147, n. 4, p. 553-559, 2002.

FORD, M. G. et al. Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. **Proteins**, v. 53, n. 2, p. 229-40, Nov 2003.

FREEZE, H. H. et al. Solving glycosylation disorders: fundamental approaches reveal complicated pathways. **Am J Hum Genet**, v. 94, n. 2, p. 161-75, Feb 2014.

GABIUS, H. J. et al. From lectin structure to functional glycomics: principles of the sugar code. **Trends Biochem Sci**, v. 36, n. 6, p. 298-313, Jun 2011.

GARNER, O. B. et al. Endothelial galectin-1 binds to specific glycans on nipah virus fusion protein and inhibits maturation, mobility, and function to block syncytia formation. **PLoS Pathog**, v. 6, n. 7, p. e1000993, 2010.

GARNER, O. B.; BAUM, L. G. Galectin-glycan lattices regulate cell-surface glycoprotein organization and signalling. **Biochem Soc Trans**, v. 36, n. Pt 6, p. 1472-7, Dec 2008.

GASTEIGER, E. et al. Protein Identification and Analysis Tools on the ExPASy Server. In: WALKER, J. (Ed.). **The Proteomics Protocols Handbook**: Humana Press, 2005. cap. 52, p.571-607.

GITT, M. A. et al. Galectin-4 and galectin-6 are two closely related lectins expressed in mouse gastrointestinal tract. **J Biol Chem**, v. 273, n. 5, p. 2954-60, Jan 1998.

_____. Isolation and expression of a gene encoding L-14-II, a new human soluble lactose-binding lectin. **Journal of Biological Chemistry**, v. 267, n. 15, p. 10601-10606, 1992.

_____. Sequence and Mapping of Galectin-5, a β -Galactoside-binding Lectin, Found in Rat Erythrocytes. **Journal of Biological Chemistry**, v. 270, n. 10, p. 5032-5038, 1995.

_____. Sequence, structure, and chromosomal mapping of the mouse Lgals6 gene, encoding galectin-6. **J Biol Chem**, v. 273, n. 5, p. 2961-70, Jan 1998.

GLINSKY, V. V. et al. Intravascular Metastatic Cancer Cell Homotypic Aggregation at the Sites of Primary Attachment to the Endothelium. **Cancer Research**, v. 63, n. 13, p. 3805-3811, 2003.

GUARDIA, C. M. A. et al. An Integrated Computational Analysis of the Structure, Dynamics, and Ligand Binding Interactions of the Human Galectin Network. **Journal of Chemical Information and Modeling**, v. 51, n. 8, p. 1918-1930, 2011/08/22 2011.

GUIMARAES, B. G. et al. The MX2 macromolecular crystallography beamline: a wiggler X-ray source at the LNLS. **Journal of Synchrotron Radiation**, v. 16, n. 1, p. 69-75, 01/01 2009.

GUO, Y. et al. Lipid droplets at a glance. **J Cell Sci**, v. 122, n. Pt 6, p. 749-52, Mar 2009.

HADARI, Y. R. et al. Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis. **J Cell Sci**, v. 113 (Pt 13), p. 2385-97, Jul 2000.

_____. Galectin-8. A new rat lectin, related to galectin-4. **J Biol Chem**, v. 270, n. 7, p. 3447-53, Feb 1995.

HAILU, T. T.; FOIT, L.; BARDWELL, J. C. In vivo detection and quantification of chemicals that enhance protein stability. **Anal Biochem**, v. 434, n. 1, p. 181-6, Mar 2013.

HANSEN, G. H. et al. Lipid rafts exist as stable cholesterol-independent microdomains in the brush border membrane of enterocytes. **J Biol Chem**, v. 276, n. 34, p. 32338-44, Aug 2001.

HARRIES, D.; RÖSGEN, J. A practical guide on how osmolytes modulate macromolecular properties. **Methods Cell Biol**, v. 84, p. 679-735, 2008.

HEINZELMANN-SCHWARZ, V. A. et al. A distinct molecular profile associated with mucinous epithelial ovarian cancer. **Br J Cancer**, v. 94, n. 6, p. 904-13, Mar 2006.

HENNO, S. et al. Expression of Po66-CBP, a galectin-8, in different types of primary and secondary broncho-pulmonary tumors. **Oncol Rep**, v. 9, n. 1, p. 177-80, 2002 Jan-Feb 2002.

HESS, B. et al. LINC: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463-1472, 1997.

HIPPO, Y. et al. Differential Gene Expression Profiles of Scirrhus Gastric Cancer Cells with High Metastatic Potential to Peritoneum or Lymph Nodes. **Cancer Research**, v. 61, n. 3, p. 889-895, 2001.

HIRABAYASHI, J. et al. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1572, n. 2-3, p. 232-254, 9/19/ 2002.

HIRABAYASHI, J.; KASAI, K.-I. The family of metazoan metal-independent β -galactoside-binding lectins: structure, function and molecular evolution. **Glycobiology**, v. 3, n. 4, p. 297-304, 1993.

HIRASHIMA, M. et al. Galectin-9 in physiological and pathological conditions. **Glycoconjugate Journal**, v. 19, n. 7-9, p. 593-600, 2002/01/01 2002.

HITTELET, A. et al. Upregulation of galectins-1 and -3 in human colon cancer and their role in regulating cell migration. **Int J Cancer**, v. 103, n. 3, p. 370-9, Jan 2003.

HO, S. N. et al. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. **Gene**, v. 77, n. 1, p. 51-9, Apr 1989.

HOKAMA, A.; MIZOGUCHI, E.; MIZOGUCHI, A. Roles of galectins in inflammatory bowel disease. **World J Gastroenterol**, v. 14, n. 33, p. 5133-7, Sep 2008.

HOKAMA, A. et al. Induced reactivity of intestinal CD4(+) T cells with an epithelial cell lectin, galectin-4, contributes to exacerbation of intestinal inflammation. **Immunity**, v. 20, n. 6, p. 681-93, Jun 2004.

HONJO, Y. et al. Down-regulation of galectin-3 suppresses tumorigenicity of human breast carcinoma cells. **Clin Cancer Res**, v. 7, n. 3, p. 661-8, Mar 2001.

HORTON, R. M. et al. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. **Gene**, v. 77, n. 1, p. 61-8, Apr 1989.

HOTTA, K. et al. Galectin-12, an adipose-expressed galectin-like molecule possessing apoptosis-inducing activity. **Journal of Biological Chemistry**, v. 276, n. 36, p. 34089-34097, 2001.

HOUZELSTEIN, D. et al. Phylogenetic Analysis of the Vertebrate Galectin Family. **Molecular Biology and Evolution**, v. 21, n. 7, p. 1177-1187, 2004.

_____. Expression patterns suggest that despite considerable functional redundancy, galectin-4 and -6 play distinct roles in normal and damaged mouse digestive tract. **J Histochem Cytochem**, v. 61, n. 5, p. 348-61, May 2013.

HUFLEJT, M.; LEFFLER, H. Galectin-4 in normal tissues and cancer. **Glycoconjugate Journal**, v. 20, n. 4, p. 247-255, 2003/05/01 2003.

HUFLEJT, M. E. et al. Strikingly different localization of galectin-3 and galectin-4 in human colon adenocarcinoma T84 cells. Galectin-4 is localized at sites of cell adhesion. **J Biol Chem**, v. 272, n. 22, p. 14294-303, May 1997.

HUGHES, R. C. Secretion of the galectin family of mammalian carbohydrate-binding proteins. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1473, n. 1, p. 172-185, 12/17/1999. ISSN 0304-4165.

IDEO, H. et al. Galectin-8-N-domain recognition mechanism for sialylated and sulfated glycans. **J Biol Chem**, v. 286, n. 13, p. 11346-55, Apr 2011.

_____. The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity. **Glycobiology**, v. 13, n. 10, p. 713-23, Oct 2003.

_____. High-affinity binding of recombinant human galectin-4 to SO₃→3Galβ1→3GalNAc pyranoside. **Glycobiology**, v. 12, n. 3, p. 199-208, 2002.

IDEO, H.; SEKO, A.; YAMASHITA, K. Galectin-4 binds to sulfated glycosphingolipids and carcinoembryonic antigen in patches on the cell surface of human colon adenocarcinoma cells. **J Biol Chem**, v. 280, n. 6, p. 4730-7, Feb 2005.

_____. Recognition Mechanism of Galectin-4 for Cholesterol 3-Sulfate. **Journal of Biological Chemistry**, v. 282, n. 29, p. 21081-21089, 2007.

IGNATOVA, Z.; GIERASCH, L. M. Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. **Proc Natl Acad Sci U S A**, v. 103, n. 36, p. 13357-61, Sep 2006.

IMAIZUMI, T. et al. Interferon-gamma stimulates the expression of galectin-9 in cultured human endothelial cells. **J Leukoc Biol**, v. 72, n. 3, p. 486-91, Sep 2002.

IMBERTY, A.; VARROT, A. Microbial recognition of human cell surface glycoconjugates. **Curr Opin Struct Biol**, v. 18, n. 5, p. 567-76, Oct 2008.

INUFUSA, H. et al. Role of galectin-3 in adenocarcinoma liver metastasis. **Int J Oncol**, v. 19, n. 5, p. 913-9, Nov 2001.

JOHN, C. M. et al. Truncated galectin-3 inhibits tumor growth and metastasis in orthotopic nude mouse model of human breast cancer. **Clin Cancer Res**, v. 9, n. 6, p. 2374-83, Jun 2003.

- KASAI, K.; HIRABAYASHI, J. Galectins: a family of animal lectins that decipher glycocodes. **J Biochem**, v. 119, n. 1, p. 1-8, Jan 1996.
- KASHIO, Y. et al. Galectin-9 induces apoptosis through the calcium-calpain-caspase-1 pathway. **J Immunol**, v. 170, n. 7, p. 3631-6, Apr 2003.
- KIM, B. W. et al. Structural basis for recognition of autophagic receptor NDP52 by the sugar receptor galectin-8. **Nat Commun**, v. 4, p. 1613, 2013.
- KIM, D. E.; CHIVIAN, D.; BAKER, D. Protein structure prediction and analysis using the Robetta server. **Nucleic Acids Res**, v. 32, n. Web Server issue, p. W526-31, Jul 2004.
- KIM, S. W. et al. Abrogation of galectin-4 expression promotes tumorigenesis in colorectal cancer. **Cell Oncol (Dordr)**, v. 36, n. 2, p. 169-78, Apr 2013.
- KIRSCHNER, K. N. et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. **J Comput Chem**, v. 29, n. 4, p. 622-55, Mar 2008.
- KONDOH, N. et al. Identification and characterization of genes associated with human hepatocellular carcinogenesis. **Cancer Res**, v. 59, n. 19, p. 4990-6, Oct 1999.
- KOPITZ, J. et al. Homodimeric galectin-7 (p53-induced gene 1) is a negative growth regulator for human neuroblastoma cells. **Oncogene**, v. 22, n. 40, p. 6277-88, Sep 2003.
- KUMAR, S.; FRANK, M.; SCHWARTZ-ALBIEZ, R. Understanding the specificity of human Galectin-8C domain interactions with its glycan ligands based on molecular dynamics simulations. **PLoS One**, v. 8, n. 3, p. e59761, 2013.
- LAHM, H. et al. Tumor galectinology: Insights into the complex network of a family of endogenous lectins. **Glycoconjugate Journal**, v. 20, n. 4, p. 227-238, 2003/05/01 2003.
- LAINE, R. A. The Information-Storing Potential of the Sugar Code. In: (Ed.). **Glycosciences**: Wiley-VCH Verlag GmbH, 1996. p.1-14.
- LANGE, O. F.; SCHÄFER, L. V.; GRUBMÜLLER, H. Flooding in GROMACS: accelerated barrier crossings in molecular dynamics. **J Comput Chem**, v. 27, n. 14, p. 1693-702, Nov 2006.
- LE MARER, N.; HUGHES, R. C. Effects of the carbohydrate-binding protein galectin-3 on the invasiveness of human breast carcinoma cells. **J Cell Physiol**, v. 168, n. 1, p. 51-8, Jul 1996.
- LE, Q. T. et al. Galectin-1: a link between tumor hypoxia and tumor immune privilege. **J Clin Oncol**, v. 23, n. 35, p. 8932-41, Dec 2005.
- LEE, S. et al. Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon. **Oncol Rep**, v. 16, n. 4, p. 747-54, Oct 2006.
- LEFFLER, H. et al. Introduction to galectins. **Glycoconjugate Journal**, v. 19, n. 7-9, p. 433-440, 2002/01/01 2002.

LEFFLER, H.; MASIARZ, F. R.; BARONDES, S. H. Soluble lactose-binding vertebrate lectins: a growing family. **Biochemistry**, v. 28, n. 23, p. 9222-9229, 1989/11/01 1989.

LEHNINGER, A.; NELSON, D.; COX, M. **Principios de bioquímica de Lehninger**. 5. Porto Alegre: 2011. 1274

LESLIE, A. G. W.; POWELL, H. R. Processing diffraction data with MOSFLM. **Evolving Methods for Macromolecular Crystallography: THE STRUCTURAL PATH TO THE UNDERSTANDING OF THE MECHANISM OF ACTION OF CBRN AGENTS**, v. 245, p. 41-51, 2007 2007.

LEVY, Y. et al. Galectin-8 functions as a matricellular modulator of cell adhesion. **J Biol Chem**, v. 276, n. 33, p. 31285-95, Aug 2001.

_____. It depends on the hinge: a structure-functional analysis of galectin-8, a tandem-repeat type lectin. **Glycobiology**, v. 16, n. 6, p. 463-76, Jun 2006.

_____. Sustained induction of ERK, protein kinase B, and p70 S6 kinase regulates cell spreading and formation of F-actin microspikes upon ligation of integrins by galectin-8, a mammalian lectin. **J Biol Chem**, v. 278, n. 16, p. 14533-42, Apr 2003.

LI, H.; ROBERTSON, A. D.; JENSEN, J. H. Very fast empirical prediction and rationalization of protein pKa values. **Proteins: Structure, Function, and Bioinformatics**, v. 61, n. 4, p. 704-721, 2005.

LINDORFF-LARSEN, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. **Proteins**, v. 78, n. 8, p. 1950-8, Jun 2010.

LIU, F. T.; PATTERSON, R. J.; WANG, J. L. Intracellular functions of galectins. **Biochimica Et Biophysica Acta-General Subjects**, v. 1572, n. 2-3, p. 263-273, 2002.

LIU, F. T.; RABINOVICH, G. A. Galectins as modulators of tumour progression. **Nat Rev Cancer**, v. 5, n. 1, p. 29-41, Jan 2005.

_____. Galectins: regulators of acute and chronic inflammation. **Ann N Y Acad Sci**, v. 1183, p. 158-82, Jan 2010.

LOVELL, S. C. et al. Structure validation by C α geometry: phi,psi and C β deviation. **Proteins**, v. 50, n. 3, p. 437-50, Feb 2003.

LUCKOW, V. A. et al. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. **J Virol**, v. 67, n. 8, p. 4566-79, Aug 1993.

LUFT, J. R.; DETTITA, G. T. A method to produce microseed stock for use in the crystallization of biological macromolecules. **Acta Crystallogr D Biol Crystallogr**, v. 55, n. Pt 5, p. 988-93, May 1999.

- LÓPEZ-LUCENDO, M. F. et al. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. **J Mol Biol**, v. 343, n. 4, p. 957-70, Oct 2004.
- MALAKHOV, M. P. et al. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. **J Struct Funct Genomics**, v. 5, n. 1-2, p. 75-86, 2004.
- MARQUEZ, R. T. et al. Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in normal fallopian tube, endometrium, and colon. **Clin Cancer Res**, v. 11, n. 17, p. 6116-26, Sep 2005.
- MASSA, S. M. et al. L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity. **Biochemistry**, v. 32, n. 1, p. 260-267, 1993/01/01 1993.
- MATSUMOTO, R. et al. Human Ecalectin, a Variant of Human Galectin-9, Is a Novel Eosinophil Chemoattractant Produced by T Lymphocytes. **Journal of Biological Chemistry**, v. 273, n. 27, p. 16976-16984, 1998.
- MATTHEWS, B. W. Solvent content of protein crystals. **Journal of Molecular Biology**, v. 33, n. 2, p. 491-497, 4/28/ 1968.
- MATULIS, D. et al. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. **Biochemistry**, v. 44, n. 13, p. 5258-66, Apr 2005.
- MCCOY, A. J. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, p. 658-674, Aug 2007.
- MCPHERSON, A. **Introduction to Macromolecular Crystallography**. Wiley, 2011.
- MEYNIER, C. et al. NMR and MD investigations of human galectin-1/oligosaccharide complexes. **Biophys J**, v. 97, n. 12, p. 3168-77, Dec 2009.
- MILLER, M. C. et al. The carbohydrate-binding domain on galectin-1 is more extensive for a complex glycan than for simple saccharides: implications for galectin-glycan interactions at the cell surface. **Biochem J**, v. 421, n. 2, p. 211-21, Jul 2009.
- MIROUX, B.; WALKER, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. **J Mol Biol**, v. 260, n. 3, p. 289-98, Jul 1996.
- MIZOGUCHI, A.; MIZOGUCHI, E. Inflammatory bowel disease, past, present and future: lessons from animal models. **J Gastroenterol**, v. 43, n. 1, p. 1-17, 2008.
- MIZOGUCHI, A.; MIZOGUCHI, E.; BHAN, A. K. Immune networks in animal models of inflammatory bowel disease. **Inflamm Bowel Dis**, v. 9, n. 4, p. 246-59, Jul 2003.
- MIZOGUCHI, E.; MIZOGUCHI, A. Is the sugar always sweet in intestinal inflammation? **Immunol Res**, v. 37, n. 1, p. 47-60, 2007.

MOSSESSOVA, E.; LIMA, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. **Mol Cell**, v. 5, n. 5, p. 865-76, May 2000.

NAGAE, M. et al. Crystal structure of the galectin-9 N-terminal carbohydrate recognition domain from *Mus musculus* reveals the basic mechanism of carbohydrate recognition. **J Biol Chem**, v. 281, n. 47, p. 35884-93, Nov 2006.

_____. Structural analysis of the recognition mechanism of poly-N-acetyllactosamine by the human galectin-9 N-terminal carbohydrate recognition domain. **Glycobiology**, v. 19, n. 2, p. 112-7, Feb 2009.

_____. Structural analysis of the human galectin-9 N-terminal carbohydrate recognition domain reveals unexpected properties that differ from the mouse orthologue. **J Mol Biol**, v. 375, n. 1, p. 119-35, Jan 2008.

NAGY, N. et al. Refined prognostic evaluation in colon carcinoma using immunohistochemical galectin fingerprinting. **Cancer**, v. 97, n. 8, p. 1849-58, Apr 2003.

NAKAGAWA, M.; KOJIMA, S. Effect of cholesterol sulfate and sodium dodecyl sulfate on lecithin-cholesterol acyltransferase in human plasma. **J Biochem**, v. 80, n. 4, p. 729-33, Oct 1976.

NANGIA-MAKKER, P. et al. Galectin-3 induces endothelial cell morphogenesis and angiogenesis. **Am J Pathol**, v. 156, n. 3, p. 899-909, Mar 2000.

NIESEN, F. H.; BERGLUND, H.; VEDADI, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. **Nat Protoc**, v. 2, n. 9, p. 2212-21, 2007.

NISHI, N. et al. Galectin-8 and galectin-9 are novel substrates for thrombin. **Glycobiology**, v. 16, n. 11, p. 15C-20C, Nov 2006.

NORLING, L. V.; PERRETTI, M.; COOPER, D. Endogenous galectins and the control of the host inflammatory response. **J Endocrinol**, v. 201, n. 2, p. 169-84, May 2009.

NOWELL, P. C. Phytohemagglutinin: an initiator of mitosis in cultures of normal human leukocytes. **Cancer Res**, v. 20, p. 462-6, May 1960.

ODA, Y. et al. Soluble lactose-binding lectin from rat intestine with two different carbohydrate-binding domains in the same peptide chain. **J Biol Chem**, v. 268, n. 8, p. 5929-39, Mar 1993.

OPDENAKKER, G. et al. Concepts and principles of glycobiology. **FASEB J**, v. 7, n. 14, p. 1330-7, Nov 1993.

PACE, C. N. et al. How to measure and predict the molar absorption coefficient of a protein. **Protein Sci**, v. 4, n. 11, p. 2411-23, Nov 1995.

PACLIK, D. et al. Galectin-4 controls intestinal inflammation by selective regulation of peripheral and mucosal T cell apoptosis and cell cycle. **PLoS One**, v. 3, n. 7, p. e2629, 2008.

_____. Galectin-2 and -4, but not galectin-1, promote intestinal epithelial wound healing in vitro through a TGF-beta-independent mechanism. **Inflamm Bowel Dis**, v. 14, n. 10, p. 1366-72, Oct 2008.

PANTOLIANO, M. W. et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. **J Biomol Screen**, v. 6, n. 6, p. 429-40, Dec 2001.

PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182-7190, 1981.

PEUMANS, W. J.; VAN DAMME, E. J. Lectins as plant defense proteins. **Plant Physiol**, v. 109, n. 2, p. 347-52, Oct 1995.

PIEPER, U. et al. ModBase, a database of annotated comparative protein structure models, and associated resources. **Nucleic Acids Res**, v. 39, n. Database issue, p. D465-74, Jan 2011.

PODOLSKY, D. K. Inflammatory Bowel Disease. **New England Journal of Medicine**, v. 347, n. 6, p. 417-429, 2002/08/08 2002.

POKLAR, N. et al. pH and temperature-induced molten globule-like denatured states of equinatoxin II: a study by UV-melting, DSC, far- and near-UV CD spectroscopy, and ANS fluorescence. **Biochemistry**, v. 36, n. 47, p. 14345-52, Nov 1997.

PRONK, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics**, v. 29, n. 7, p. 845-54, Apr 2013.

RABINOVICH, G. A. Galectin-1 as a potential cancer target. **Br J Cancer**, v. 92, n. 7, p. 1188-92, Apr 2005.

RABINOVICH, G. A.; CROCI, D. O. Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer. **Immunity**, v. 36, n. 3, p. 322-35, Mar 2012.

RABINOVICH, G. A.; ILARREGUI, J. M. Conveying glycan information into T-cell homeostatic programs: a challenging role for galectin-1 in inflammatory and tumor microenvironments. **Immunological Reviews**, v. 230, n. 1, p. 144-159, 2009.

RABINOVICH, G. A.; TOSCANO, M. A. 'Turning 'sweet' on immunity: galectin-glycan interactions in immune tolerance and inflammation. **Nat Rev Immunol**, v. 9, n. 5, p. 338-52, May 2009.

RABINOVICH, G. A. et al. Functions of cell surface galectin-glycoprotein lattices. **Curr Opin Struct Biol**, v. 17, n. 5, p. 513-20, Oct 2007.

RECHRECHE, H. et al. Cloning and expression of the mRNA of human galectin-4, an S-type lectin down-regulated in colorectal cancer. **Eur J Biochem**, v. 248, n. 1, p. 225-30, Aug 1997.

RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOS: The European Molecular Biology Open Software Suite. **Trends in Genetics**, v. 16, n. 6, p. 276-277, 6/1/ 2000.

ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDscript server. **Nucleic Acids Res**, v. 42, n. Web Server issue, p. W320-4, Jul 2014.

ROSSMANN, M. G. The molecular replacement method. **Acta Crystallographica Section A**, v. 46, n. 2, p. 73-82, 1990.

RUBINSTEIN, N. et al. Targeted inhibition of galectin-1 gene expression in tumor cells results in heightened T cell-mediated rejection; A potential mechanism of tumor-immune privilege. **Cancer Cell**, v. 5, n. 3, p. 241-51, Mar 2004.

RUMILLA, K. M. et al. Galectin-4 expression in carcinoid tumors. **Endocr Pathol**, v. 17, n. 3, p. 243-9, 2006.

SALATINO, M.; RABINOVICH, G. A. Fine-tuning antitumor responses through the control of galectin-glycan interactions: an overview. **Methods Mol Biol**, v. 677, p. 355-74, 2011.

SALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **J Mol Biol**, v. 234, n. 3, p. 779-815, Dec 1993.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular cloning: a laboratory manual**. Cold Spring Harbor Laboratory, 1989.

SATELLI, A. et al. Galectin-4 functions as a tumor suppressor of human colorectal cancer. **Int J Cancer**, v. 129, n. 4, p. 799-809, Aug 2011.

SATO, M. et al. Functional analysis of the carbohydrate recognition domains and a linker peptide of galectin-9 as to eosinophil chemoattractant activity. **Glycobiology**, v. 12, n. 3, p. 191-7, Mar 2002.

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

SINGH, S. M.; PANDA, A. K. Solubilization and refolding of bacterial inclusion body proteins. **J Biosci Bioeng**, v. 99, n. 4, p. 303-10, Apr 2005.

SOLÍS, D. et al. A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. **Biochim Biophys Acta**, Mar 2014.

SONG, Y. K.; BILLIAR, T. R.; LEE, Y. J. Role of galectin-3 in breast cancer metastasis: involvement of nitric oxide. **Am J Pathol**, v. 160, n. 3, p. 1069-75, Mar 2002.

SOUSA DA SILVA, A. W.; VRANKEN, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. **BMC Res Notes**, v. 5, p. 367, 2012.

STILLMAN, B. N. et al. Galectin-3 and galectin-1 bind distinct cell surface glycoprotein receptors to induce T cell death. **J Immunol**, v. 176, n. 2, p. 778-89, Jan 2006.

STOWELL, S. R. et al. Innate immune lectins kill bacteria expressing blood group antigen. **Nat Med**, v. 16, n. 3, p. 295-301, Mar 2010.

_____. Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. **J Biol Chem**, v. 283, n. 15, p. 10109-23, Apr 2008.

_____. Dimeric Galectin-8 induces phosphatidylserine exposure in leukocytes through polylactosamine recognition by the C-terminal domain. **J Biol Chem**, v. 283, n. 29, p. 20547-59, Jul 2008.

_____. Differential roles of galectin-1 and galectin-3 in regulating leukocyte viability and cytokine secretion. **J Immunol**, v. 180, n. 5, p. 3091-102, Mar 2008.

STOWELL, S. R. K. S. S. C. J. D.-B. M. M. R. P. C. R. D. Human galectin-1, -2, and -4 induce surface exposure of phosphatidylserine in activated human neutrophils but not in activated T cells. **Blood**, v. 109, n. 1, p. 219-227, 2006.

STROBER, W.; FUSS, I. J.; BLUMBERG, R. S. THE IMMUNOLOGY OF MUCOSAL MODELS OF INFLAMMATION1. **Annual Review of Immunology**, v. 20, n. 1, p. 495-549, 2002/04/01 2002.

STROTT, C. A.; HIGASHI, Y. Cholesterol sulfate in human physiology: what's it all about? **J Lipid Res**, v. 44, n. 7, p. 1268-78, Jul 2003.

SU, Z. Z. et al. Surface-epitope masking and expression cloning identifies the human prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin gene family. **Proc Natl Acad Sci U S A**, v. 93, n. 14, p. 7252-7, Jul 1996.

TAKENAKA, Y.; FUKUMORI, T.; RAZ, A. Galectin-3 and metastasis. **Glycoconjugate Journal**, v. 19, n. 7-9, p. 543-549, 2002/01/01 2002.

THIJSEN, V. L.; HULSMANS, S.; GRIFFIOEN, A. W. The galectin profile of the endothelium: altered expression and localization in activated and tumor endothelial cells. **Am J Pathol**, v. 172, n. 2, p. 545-53, Feb 2008.

THIJSEN, V. L. J. L. et al. Galectin-1 is essential in tumor angiogenesis and is a target for antiangiogenesis therapy. **Proceedings of the National Academy of Sciences**, v. 103, n. 43, p. 15975-15980, 2006.

TINARI, N. et al. Glycoprotein 90K/MAC-2BP interacts with galectin-1 and mediates galectin-1-induced cell aggregation. **Int J Cancer**, v. 91, n. 2, p. 167-72, Jan 2001.

TOSCANO, M. A. et al. Dissecting the pathophysiologic role of endogenous lectins: glycan-binding proteins with cytokine-like activity? **Cytokine Growth Factor Rev**, v. 18, n. 1-2, p. 57-71, 2007 Feb-Apr 2007.

_____. [How do protein-glycan interactions regulate T-cell physiology?]. **Medicina (B Aires)**, v. 66, n. 4, p. 357-62, 2006.

TRIPODI, D. et al. Gene expression profiling in sinonasal adenocarcinoma. **BMC Med Genomics**, v. 2, p. 65, 2009.

TRONCOSO, M. F. et al. Integrating structure and function of 'tandem-repeat' galectins. **Front Biosci (Schol Ed)**, v. 4, p. 864-87, 2012.

TSUMOTO, K. et al. Practical considerations in refolding proteins from inclusion bodies. **Protein Expr Purif**, v. 28, n. 1, p. 1-8, Mar 2003.

TÜRECI, O. et al. Molecular definition of a novel human galectin which is immunogenic in patients with Hodgkin's disease. **J Biol Chem**, v. 272, n. 10, p. 6416-22, Mar 1997.

UEDA, S.; KUWABARA, I.; LIU, F. T. Suppression of tumor growth by galectin-7 gene transfer. **Cancer Res**, v. 64, n. 16, p. 5672-6, Aug 2004.

VAN BAAL, J. W. et al. A comparative analysis by SAGE of gene expression profiles of Barrett's esophagus, normal squamous esophagus, and gastric cardia. **Gastroenterology**, v. 129, n. 4, p. 1274-81, Oct 2005.

VAN DEN BRÛLE, F.; CALIFICE, S.; CASTRONOVO, V. Expression of galectins in cancer: A critical review. **Glycoconjugate Journal**, v. 19, n. 7-9, p. 537-542, 2002/01/01 2002.

VAN DEN BRÛLE, F. et al. Galectin-1 accumulation in the ovary carcinoma peritumoral stroma is induced by ovary carcinoma cells and affects both cancer cell proliferation and adhesion to laminin-1 and fibronectin. **Lab Invest**, v. 83, n. 3, p. 377-86, Mar 2003.

VANDENBORRE, G.; SMAGGHE, G.; VAN DAMME, E. J. Plant lectins as defense proteins against phytophagous insects. **Phytochemistry**, v. 72, n. 13, p. 1538-50, Sep 2011.

VARKI, A. et al. **Essentials of Glycobiology**. 2nd 2009.

VASTA, G. R. Roles of galectins in infection. **Nat Rev Microbiol**, v. 7, n. 6, p. 424-38, Jun 2009.

VASTA, G. R. et al. Galectins as self/non-self recognition receptors in innate and adaptive immunity: An unresolved paradox. **Frontiers in Immunology**, v. 3, 2012.

VASTA, G. R.; AHMED, H.; ODOM, E. W. Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. **Curr Opin Struct Biol**, v. 14, n. 5, p. 617-30, Oct 2004.

VASTA, G. R. et al. Structural and functional diversity of the lectin repertoire in teleost fish: relevance to innate and adaptive immunity. **Dev Comp Immunol**, v. 35, n. 12, p. 1388-99, Dec 2011.

VEGA-CARRASCAL, I. et al. Dysregulation of TIM-3-galectin-9 pathway in the cystic fibrosis airways. **J Immunol**, v. 186, n. 5, p. 2897-909, Mar 2011.

VOKHMYANINA, O. A. et al. Comparative study of the glycan specificities of cell-bound human tandem-repeat-type galectin-4, -8 and -9. **Glycobiology**, v. 22, n. 9, p. 1207-17, Sep 2012.

WADA, J.; KANWAR, Y. S. Identification and characterization of galectin-9, a novel beta-galactoside-binding mammalian lectin. **J Biol Chem**, v. 272, n. 9, p. 6078-86, Feb 1997.

- WADA, J. et al. Developmental regulation, expression, and apoptotic potential of galectin-9, a beta-galactoside binding lectin. **J Clin Invest**, v. 99, n. 10, p. 2452-61, May 1997.
- WAGNER, S. et al. Tuning Escherichia coli for membrane protein overexpression. **Proc Natl Acad Sci U S A**, v. 105, n. 38, p. 14371-6, Sep 2008.
- WALTHER, T. C.; FARESE, R. V. The life of lipid droplets. **Biochim Biophys Acta**, v. 1791, n. 6, p. 459-66, Jun 2009.
- WANG, J. et al. Development and testing of a general amber force field. **J Comput Chem**, v. 25, n. 9, p. 1157-74, Jul 2004.
- WASANO, K.; HIRAKAWA, Y. Two domains of rat galectin-4 bind to distinct structures of the intercellular borders of colorectal epithelia. **J Histochem Cytochem**, v. 47, n. 1, p. 75-82, Jan 1999.
- WILLIAMS, M. L.; HUGHES-FULFORD, M.; ELIAS, P. M. Inhibition of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity and sterol synthesis by cholesterol sulfate in cultured fibroblasts. **Biochim Biophys Acta**, v. 845, n. 3, p. 349-57, Jun 1985.
- WOOTERS, M. A. et al. Immunohistochemical characterization of the distribution of galectin-4 in porcine small intestine. **J Histochem Cytochem**, v. 53, n. 2, p. 197-205, Feb 2005.
- WOOTERS, M. A.; ROPP, S. L.; ERICKSON, A. K. Identification of galectin-4 isoforms in porcine small intestine. **Biochimie**, v. 87, n. 2, p. 143-9, Feb 2005.
- WU, A. M. et al. Lectins as tools in glycoconjugate research. **Glycoconjugate Journal**, v. 26, n. 8, p. 899-913, 2009.
- WU, M. H. et al. Galectin-1-mediated tumor invasion and metastasis, up-regulated matrix metalloproteinase expression, and reorganized actin cytoskeletons. **Mol Cancer Res**, v. 7, n. 3, p. 311-8, Mar 2009.
- WÄLTERMANN, M.; STEINBÜCHEL, A. Neutral lipid bodies in prokaryotes: recent insights into structure, formation, and relationship to eukaryotic lipid depots. **J Bacteriol**, v. 187, n. 11, p. 3607-19, Jun 2005.
- YANG, R.-Y.; RABINOVICH, G. A.; LIU, F.-T. Galectins: structure, function and therapeutic potential. **Expert Reviews in Molecular Medicine**, v. 10, p. null-null, 2008.
- YANG, R.-Y. et al. Ablation of a galectin preferentially expressed in adipocytes increases lipolysis, reduces adiposity, and improves insulin sensitivity in mice. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 46, p. 18696-18701, Nov 15 2011.
- YANG, R. Y.; HAVEL, P. J.; LIU, F. T. Galectin-12: A protein associated with lipid droplets that regulates lipid metabolism and energy balance. **Adipocyte**, v. 1, n. 2, p. 96-100, Apr 2012.
- YANG, R. Y. et al. Galectin-12 is required for adipogenic signaling and adipocyte differentiation. **Journal of Biological Chemistry**, v. 279, n. 28, p. 29761-29766, 2004.

_____. Cell cycle regulation by galectin-12, a new member of the galectin superfamily. **Journal of Biological Chemistry**, v. 276, n. 23, p. 20252-20260, 2001.

YOSHIDA, H. et al. X-ray structure of a protease-resistant mutant form of human galectin-8 with two carbohydrate recognition domains. **FEBS J**, v. 279, n. 20, p. 3937-51, Oct 2012.

ZICK, Y. et al. Role of galectin-8 as a modulator of cell adhesion and cell growth. **Glycoconj J**, v. 19, n. 7-9, p. 517-26, 2004.

ZIMBARDI, A. L. et al. Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of the N-terminal carbohydrate-recognition domain of human galectin-4. **Acta Crystallogr Sect F Struct Biol Cryst Commun**, v. 66, n. Pt 5, p. 542-5, May 2010.

ZUO, X. et al. Enhanced expression and purification of membrane proteins by SUMO fusion in Escherichia coli. **J Struct Funct Genomics**, v. 6, n. 2-3, p. 103-11, 2005.