**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO**

**“Clonagem, expressão, purificação e estudos estruturais dos
domínios de reconhecimento de carboidratos (CRDs) da
galectina-4 humana”**

Ana Lucia Ribeiro Latorre Zimbardi

Ribeirão Preto

2009

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

“Clonagem, expressão, purificação e estudos estruturais dos domínios de reconhecimento de carboidratos (CRDs) da galectina-4 humana”

Dissertação de Mestrado apresentada à Faculdade de Ciências Farmacêuticas de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Mestre em Ciências Farmacêuticas.

Área de Concentração: Física Biológica

Orientada: Ana Lucia Ribeiro Latorre Zimbardi

Orientadora: Profa. Dra. Maria Cristina Nonato Costa

Ribeirão Preto

2009

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.

Zimbardi, Ana Lucia Ribeiro Latorre

Clonagem, expressão, purificação e estudos estruturais dos domínios de reconhecimento de carboidratos (CRDs) da galectina-4 humana / Ana Lucia Ribeiro Latorre Zimbardi; orientadora Maria Cristina Nonato Costa, Ribeirão Preto, 2009.

115p.: il.; 30cm

Dissertação de Mestrado – Programa de Pós-Graduação em Ciências Farmacêuticas - Área de Concentração: Física Biológica – Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP.

1. Lectinas. 2. Galectina-4 humana. 3. CRD. 4. Clonagem. 5. Expressão. 6. Purificação. 7. Cristalização. 8. Estrutura cristalográfica.

FOLHA DE APROVAÇÃO

Ana Lucia Ribeiro Latorre Zimbardi

Clonagem, expressão, purificação e estudos estruturais dos domínios de reconhecimento de carboidratos (CRDs) da galectina-4 humana.

Dissertação de mestrado apresentada à Faculdade de Ciências Farmacêuticas de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Mestre em Ciências Farmacêuticas.

Área de Concentração: Física Biológica

Aprovado em: ____/____/____

Banca examinadora

Prof(a). Dr(a). _____

Instituição: _____

Assinatura: _____

Prof(a). Dr(a). _____

Instituição: _____

Assinatura: _____

Prof(a). Dr(a). _____

Instituição: _____

Assinatura: _____

DEDICATÓRIA

À minha família e alguns grandes
amigos que sempre me ensinaram, incentivaram, confiaram e acreditaram em mim,
muitas vezes mais do que eu mesmo.....

AGRADECIMENTOS

À minha orientadora, Profa. Maria Cristina Nonato, pela oportunidade de ingressar no Laboratório de Cristalografia de Proteínas de Ribeirão Preto (LCP-RP) como aluna, e também pela paciência e atenção que sempre demonstrou.

À Profa Eliane C.A. Braga, por ceder alguns equipamentos e espaço físico do seu laboratório, e todas as confraternizações e momentos descontraídos que tive o prazer de compartilhar.

Aos Profs. Marcelo Dias Baruffi e Sérgio Akira Uyemura, por disponibilizar prontamente equipamentos e/ou reagentes que foram imprescindíveis para a realização deste trabalho.

Ao Laboratório Nacional de Luz Síncrotron (LNLS), por colocar a disposição a área experimental e o laboratório,

Aos alunos, funcionários e professores com quem convivi em algum momento neste meu ingresso no LCP-RP e que de alguma forma me apoiaram, deram idéias, ajuda ou simplesmente gastaram seu tempo conversando: Karla, Flávia, Matheus Pinheiro, Paty, Ricardo, Joane, Aline, Carol, Fernando, Cristiane, Matheus Baldo, Ana Paula, Camila Vinicius Vitale, Vinicius Adriano, Gabriel, Cíntia e Cidinha.

À Faculdade de Ciências Farmacêuticas de Ribeirão Preto, pela oportunidade de realização do curso de mestrado e aos funcionários desta instituição, pela ajuda que de alguma maneira me foi prestada.

“Muitas pessoas passam por nossa vida,
algumas levam muito de nós consigo e outras não levam nada,
outras deixam muito de si conosco, mas não há pessoa que não nos deixe algo.”

RESUMO

ZIMBARDI, A.L.R.L. . Clonagem, expressão, purificação e estudos estruturais dos domínios de reconhecimento de carboidratos (CRDs) da galectina-4 humana. 2009. 115 f. Dissertação Mestrado. Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

A família das galectinas compreende um grupo de lectinas cujos domínios de reconhecimento de carboidratos (CRDs) possuem afinidade específica para β -galactosídeos. Estas se encontram amplamente distribuídas em células normais e neoplásicas de diferentes organismos e estão envolvidas em uma grande diversidade de mecanismos celulares. As galectinas têm sido foco de estudos recentes, principalmente pelo seu envolvimento em processos inflamatórios e neoplásicos, entretanto, muitas perguntas sobre as interações com diferentes carboidratos, a especificidade destas interações e o papel específico das galectinas em inflamação, adesão celular, progressão tumoral e metástase permanecem ainda sem resposta. O presente projeto focou os estudos estruturais dos domínios de reconhecimento de carboidratos (CRDs) da galectina-4 humana (HGal-4). Nosso trabalho envolveu a clonagem, expressão, purificação dos domínios de reconhecimento de carboidratos (CRD-I e CRD-II) de forma independente. O domínio CRD-I da HGal-4 foi cristalizado e sua estrutura determinada por técnicas de cristalografia de raios-X a 2 Å de resolução. A estrutura cristalográfica do domínio CRD-I da galectina-4 humana possui duas folhas- β compostas de seis fitas (S1-S6) e cinco fitas (F1-F6) enoveladas na forma de um β -sanduíche. Uma comparação estrutural entre membros da classe das galectinas mostra que este enovelamento global dos CRDs é conservado e que a diferença em especificidade pelos carboidratos observado pelas diferentes galectinas é consequência de mutações pontuais de aminoácidos. Os resultados obtidos no desenvolvimento do presente projeto serão utilizados como uma ferramenta importante para o entendimento de processos celulares que envolvem a galectina-4 humana, como inflamação, progressão celular e metástase, e conseqüentemente, contribuir para o planejamento de novas estratégias de diagnóstico e tratamento de neoplasias.

ABSTRACT

ZIMBARDI, A.L.R.L . Cloning, expression, purification and structural studies of the carbohydrate-recognition domains (CRDs) of human galectin-4. 2009. 115 f. Dissertation (Master) . Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

The galectin family comprises a group of lectins where the carbohydrate-recognition domains (CRDs) display specific affinity for β -galactosides. They are widely distributed in normal and neoplastic cells of different organisms and are involved in a great diversity of cellular mechanisms. The galectins have been focus of recent studies, mainly for their involvement in inflammatory and neoplastic processes, however, many questions about the interactions with different carbohydrates, the specificity of these interactions and the specific role of the galectins in inflammation, cell adhesion, tumor progression and metastasis remain unanswered. The present project focused the structural studies of human galectin-4 (HGal-4) carbohydrate-recognition domains (CRDs). Our work involved the independent cloning, heterologous expression and purification of both carbohydrate-recognition domains (CRD-I and CRD-II). The HGal-4 CRD-I domain has been successfully crystallized and its structure solved by X-ray crystallography techniques at 2 Å resolution. The crystallographic structure of HGal-4 CRD-I domain comprises two β -sheets containing six (S1- S6) and five strands (F1-F6) each packed as a β -sandwich domain. A structural comparison among members of galectin class of proteins shows that this folding is highly conserved and that the difference in specificity for carbohydrate molecules is consequence of punctual aminoacid mutations. Our results will be used as an important tool towards a better understanding of cellular processes such as inflammation, cell progression and metastasis, and consequently, contribute for the development of new strategies for diagnosis and treatment of neoplasies.

LISTA DE FIGURAS

Figura 01 -	Esquema ilustrativo dos três tipos de arquitetura das galectinas: protótipo, químera e <i>Tanden-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 Tandem-repeat é composto de dois CRD's homólogos.	5
Figura 02 -	Representação esquemática das estruturas proto, químera e <i>tandem-repeat</i> onde os resíduos conservados presentes no CRD estão indicados. Os resíduos indicados em maiúsculo indicam os principais resíduos envolvidos na coordenação com o carboidrato. Os resíduos indicados com asterisco são os resíduos mais conservados em todas as galectinas indicadas até o presente momento, a letra c indica a posição dos resíduos de cisteína, cuja oxidação resulta na perda de afinidade de ligação com o açúcar. A numeração utilizada corresponde à utilizada para descrever a galectina-1 humana. Adaptada de (KASAI; HIRABAYASHI, 1996.	6
Figura 03 -	Modelo de coordenação da galectina-1 humana com o carboidrato lactose. A figura mostra os principais resíduos envolvidos na coordenação com o carboidrato (LOPES-LUCENDO <i>et al.</i> , 2004), PDB código 1GZW. Representação construída no programa <i>Pymol</i> (DELANO, 2002).....	7
Figura 04 -	Estrutura tridimensional do domínio de reconhecimento de carboidrato (CRD) encontrado nas galectinas. Modelo construído com base nas coordenadas da estrutura da galectina-1 humana (LOPES-LUCENDO <i>et al.</i> , 2004) complexada com lactose, PDB código 1GZW. Representação estereográfica construída no programa <i>Pymol</i> (DELANO, 2002).....	8
Figura 05 -	Alinhamento sequencial entre as galectinas humanas 1, 2 e 3. As flechas em azul indicam os resíduos que participam da interação com o carboidrato. O alinhamento foi realizado no programa <i>MultiAlin</i> (CORPET, 1988), e a imagem foi criada no programa <i>ESPrin</i> (GOUET <i>et al.</i> , 1999).....	8
Figura 06	Estrutura primária da galectina-4 humana (HUFLEJT <i>et al.</i> , 1997). Os resíduos em amarelo representam os aminoácidos que interagem com o carboidrato. O resíduo em pink representa a única substituição conservativa observada para os resíduos que coordenam moléculas de carboidrato e os resíduos em vermelho pertencem ao peptídeo de ligação....	10
Figura 07 –	Alinhamento sequencial entre os CRD-I e CRD-II encontrado na galectina-4 humana. O alinhamento foi realizado no programa <i>MultiAlin</i> (CORPET, 1988), e a imagem foi criada no programa <i>ESPrin</i> (GOUET <i>et al.</i> 1990).....	10
Figura 08 –	Alinhamento sequencial entre as galectinas-4 humana, de porco e de rato. O alinhamento foi realizado no programa <i>MultiAlin</i> (CORPET, 1988), e a imagem foi criada no programa <i>ESPrin</i> (GOUET <i>et al.</i> , 1990).....	11
Figura 09-	Mapa do vetor pET29a (Novagen). Este plasmídeo contém um marcador que confere resistência a kanamicina, o gene <i>lacI</i> , o promotor de transcrição T7, a região do operador <i>lac</i> e um sítio múltiplo de clonagem. Também existem duas origens de replicação - a origem <i>f1</i> que permite a produção de vetor de fita simples sob condições apropriadas e uma origem de replicação convencional.....	15
Figura 10 -	Sequência de nucleotídeos e aminoácidos da Hgal-4 (HUFLEJT <i>et al.</i> , 1997). A figura mostra os sítos específicos de início e término utilizados no processo de clonagem dos CRD-I e CRD-II. A sequência do peptídeo de ligação está apresentada em laranja, GenBank NM_006149.3.....	17

Figura 11 -	Estrutura tridimensional do domínio CRD-II de Hgal-4. À esquerda os resíduos Thr-181 (localizado no peptídeo de ligação) e Arg-196 (localizados no, CRD-II) se encontram a distância de aproximadamente 3.7Å. À direita, o resíduo Pro-193 indica a região onde se inicia a região globular, PDB 1X50.....	18
Figura 12 -	Sequência dos oligonucleotídeos sintetizados para utilização na amplificação dos domínios de reconhecimento de carboidratos CRD-I e CRD-II da galectina-4 humana.....	19
Figura 13 -	Mapa do vetor pTZ5R/T (Fermentas). Este plasmídeo contém um marcador que confere resistência a ampicilina e um sítio múltiplo de clonagem. Também existem duas origens de replicação - a origem f1 que permite a produção de vetor de fita simples sob condições apropriadas e uma origem de replicação convencional.....	30
Figura 14-	Mapa do vetor pET28a (Novagen). Este plasmídeo contém um marcador que confere resistência a kanamicina, o gene lacI, o promotor de transcrição T7, a região do operador lac e um sítio múltiplo de clonagem.....	31
Figura 15 -	Interação entre a cauda de histidina da proteína (em azul) e o Ni ²⁺ da resina Ni-NTA Agarose (em vermelho).....	44
Figura 16 -	Alinhamento sequencial entre os domínios N-terminais da galectina-4 de rato e humana. Em vermelho estão selecionados os aminoácidos conservados em ambas sequencias.....	49
Figura 17 –	Eletroforese em gel de agarose 0,8% (w/v) mostrando o perfil da amplificação do gene de interesse através da técnica de PCR para diferentes temperaturas de anelamento. Os genes que codificam os domínios CRD-I e CRD-II possuem 450 bp e 435 bp respectivamente. As pistas 1, 2 e 3 pertencem ao CRD-I e as pistas 4, 5 e 6 ao CRD-II	52
Figura 18 -	Eletroforese em gel de agarose 0,8% (w/v). A figura mostra o perfil da digestão do CRD-I, CRD-II e do vetor pET-29a (Novagen) com as enzimas de restrição <i>Nde</i> -I e <i>Not</i> I (Invitrogen). As pistas 1 e 2 referem-se ao CRD-I e CRD-II respectivamente e apresentam uma banda de aproximadamente 450 e 435 pb. A pista 3 refere-se ao vetor pET-29a, e apresenta uma banda de aproximadamente 5,3 kb. As massas obtidas no processo de quantificação foram, 8,0 ng/μL, 8,0 ng/μL e 10,0 ng/ μL para os CRD-I, CRD-II e vetor pET29a respectivamente.	53
Figura 19 –	Eletroforese em gel de agarose 0,8% (w/v). Dentre as nove colônias selecionadas e testadas, o gel de agarose nos mostra a presença do gene referente ao domínio CRD-I nas colônias designadas como 1, 8 e 9. O tamanho da banda observada equivale aproximadamente 450 pb quando comparada ao padrão . Os clones obtidos foram designados de I-1, I-2 e I-3. O clone I-1 foi utilizado nos ensaios de expressão.....	54
Figura 20 -	Eletroforese em gel de agarose 0,8% (w/v). Dentre as nove colônias selecionadas e testadas, o gel de agarose nos mostra a presença do gene referente ao domínio CRD-II em todas as colônias. O tamanho da banda observada equivale aproximadamente 435 pb quando comparada ao padrão. Os clones obtidos foram designados como II-1, II-2, II-3, II-4, II-5, II-6, II-7,II-8 e II-9. O clone II-4 foi utilizado nos ensaios de expressão.....	55
Figura 21 –	Eletroforese em gel de poliacrilamida (12% SDS-PAGE) para o ensaio de expressão do domínio CRD-II utilizando-se 1,5 mM de IPTG a 37°C por 2, 3, 4, 5, 18 e 24 horas de indução. As legendas PC, PI, SC e SI correspondem a amostras do pellet controle e induzido, sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-II é de aproximadamente 17 kDa.....	56
Figura 22 -	Eletroforese em gel de poliacrilamida (12% SDS-PAGE) para o ensaio de expressão do domínio CRD-II utilizando-se 1,5 mM de IPTG a 37°C por 2, 3, 4, 5, 18 e 24 horas de indução.As legendas PC, PI, SC e SI correspondem a amostras do pellet controle e induzido, sobrenadante controle e induzido respectivamente.....	56

Figura 23	Eletoforese em gel de poliacrilamida (12% SDS-PAGE) para o ensaio de expressão do domínio CRD-II utilizando-se 1,5 mM de IPTG a 18°C, por 24 horas. As legendas PC, PI, SC e SI correspondem às amostras: pellet controle e induzido; sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-II é de aproximadamente 17 kDa.....	57
Figura 24 -	Eletoforese em gel de poliacrilamida (12% SDS-PAGE) para o ensaio de expressão do domínio CRD-I induzido com 1,5 mM de IPTG a 37°C, por 5 e 24 horas. As legendas PC, PI, SC e SI correspondem a amostra do pellet controle e induzido; sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-I é de aproximadamente 17 kDa.....	58
Figura 25 -	A figura mostra o perfil cromatográfico de purificação da Hgal-4 humana inteira, CRD-I e CRD-II em coluna com resina α -Lactose-Agarose (Sigma-Aldrich).....	59
Figura 26 -	Eletoforese em gel de poliacrilamida (SDS-PAGE 12%). Perfil das frações eluídas no processo de purificação da Hgal-4 completa e dos domínios CRD-I e CRD-II. A proteína purificada elui na presença de 100mM de lactose.....	59
Figura 27 -	Eletoforese em gel de agarose 0,8% (w/v) mostrando o perfil da amplificação dos genes de interesse através da técnica de PCR, para diferentes temperaturas de anelamento. As pistas 1, 2 e 3 pertencem ao CRD-I, as pistas 4, 5 e 6 ao CRD-II e as pistas 7, 8 e 9 pertencem a galectina-4 humana inteira. Os genes que codificam os domínios CRD-I e CRD-II e a galectina-4 humana possuem 450 bp, 435 bp e 969 pb respectivamente.....	60
Figura 28 -	Eletoforese em gel de agarose 0,8% (w/v) para o processo quantificação dos fragmentos referentes aos domínios CRDI e CRDII. As pistas 1 e 2 referem-se ao CRD-I (450pb) e apresentam aproximadamente 25 ng/ μ L (1ª eluição) e 8,4ng (2ª eluição) respectivamente; as pistas 3 e 4 referem-se ao CRD-II e apresentam aproximadamente 9,6 ng/ μ L (1ª eluição) e 8,4ng (2ª eluição) respectivamente, quando comparadas ao padrão.....	61
Figura 29 -	Eletoforese em gel de agarose 0,8% (w/v) para a reação de PCR de colônia. As pistas 1, 2, 3 e 4 referem-se ao domínio CRD-I e as pistas 5, 6, 7, 8 e 9 referem-se ao domínio CRD-II. Os clones obtidos foram designados de I-1, I-2, I-3 e I-4 para o CRD-I e II-1, II-2, II-3, II-4 e II-5 para o CRD-II.....	62
Figura 30 -	Eletoforese em gel de agarose 0,8% (w/v) para a reação de digestão das construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II (pTZ57R/T)(II-2). As pistas 1, 2 e 3 referem-se as construções Hgal-4CRD-I(pTZ57R/T)(I-1), Hgal-4CRD-II (pTZ57R/T)(II-2) e ao vetor pET28a respectivamente.	63
Figura 31	Eletoforese em gel de agarose 0,8% (w/v) para a reação de digestão do domínio CRD-I. Observa-se através do gel a presença do gene de interesse na pista 2 (450 pb). O clone obtido foi designado de CRD-I(I-1).....	64
Figura 32 -	Eletoforese em gel de agarose 0,8% (w/v) para a reação de PCR de colônia para o domínio CRD-II. Observa-se através do gel a presença do gene de interesse nas pistas 1 e 7 (435 pb). Os clones obtidos foram designados de CRD-II(II-7) e CRD-II(II-23).....	64
Figura 33 -	Eletoforese em gel de poliacrilamida (12% SDS-PAGE). A figura a mostra o padrão e perfil das lavagens na presença de 0, 25, 50, 100, 150, 200 e 250 mM Imidazol. Observa-se através do gel a eluição em 100 mM de imidazol do CRD-I. A figura b mostra o controle (pista 1) e as oito eluições com 100 mM imidazol. O domínio CRD-I migra com massa molecular de aproximadamente 17 kDa.....	65
Figura 34 -	Eletoforese em Gel de poliacrilamida (12% SDS-PAGE). A figura a mostra o padrão e perfil das lavagens na presença de 0, 25, 50, 100, 150, 200 e 250 mM de Imidazol. Observa-se através do gel a eluição em 150 mM de imidazol e 14 mM β -Mercaptoetanol do domínio CRD-II.....	66

Figura 35 -	Cristal do domínio CRD-I da galectina-4 humana obtido na condição 46 do Crystal Screen TM (PEG 8K a 18 % v/v, 0,1M Cacodilato de Sódio pH 6,5 e 0,2M Acetato de Calcio). Tamanho aproximado de 0,1mm.	67
Figura 36 -	Cristal do domínio CRD-I da galectina-4 humana obtido na condição 22 do Crystal Screen TM (PEG 4K a 30%, 0,1M Tris HCl pH8,5 e 0,2M Acetato de Sódio). Tamanho aproximado de 0.15mm.....	68
Figura 37-	Cristal do domínio CRD-I da galectina-4 humana obtido nas condições: PEG 8K, 0,1M Tris HCl pH 8,5, 0,2M de Acetato de Cálcio. a: cristais com 4 dias b: cristais com 10 dias e com tamanho aproximado de 0,3mm.....	68
Figura 38 -	Cristal do domínio CRD-I da galectina-4 obtido nas condições: PEG 8K, 0,1M Tris HCl pH 8,5, 0,2M de Acetato de Cálcio na presença de lactose. Cristais com 20 dias.....	69
Figura 39 -	Exemplo de uma imagem do padrão de difração do cristal CRD-I _{t2g23} coletado no LNLS.....	71
Figura 40 -	A figura acima demonstra a qualidade do mapa de densidade eletrônica $2F_o - F_c$, com σ igual 1.0, representado em azul.	72
Figura 41 -	Gráfico Ramachandram da estrutura final do domínio CRD-I da galectina-4 humana.....	73
Figura 42 -	Estrutura cristalográfica do CRD-I da galectina-4 humana. A folha β em verde é composta de 6 fitas (S1- S6) e a folha β em azul é composta de cinco fitas (F1-F6), conectadas por uma curta hélice, que se empacotam em um domínio do tipo β -sanduiche.....	74
Figura 43 -	Alinhamento sequencial entre alguns CRD's das galectinas do grupo <i>tandem-repeat</i> : HGAL-9_CRDI: CRD-I da galectina-9 humana; MGAL-9_CRDI: CRD-I da galectina-9 de camundongo; MGAL-4_CRDI: CRD-I da galectina-4 de camundongo, HGAL-4_CRDI: CRD-I da galectina-4 humana e HGAL-4_CRDII: CRD-II da galectina-4 humana.....	75
Figura 44 -	Superposição estrutural entre a estrutura tridimensional do domínio CRD-I da galectina-4 humana com as estruturas dos domínios CRD-I da galectina-4 de camundongo e da galectina-9 humana, e do domínio CRD-II da galectina-4 humana. As maiores diferenças observadas são encontradas nas regiões de voltas que conectam as fitas.....	75
Figura 45 -	Estrutura cristalográfica do CRD-I da galectina-4 humana, indicando os resíduos envolvidos na coordenação com a molécula de carboidrato.....	76
Figura 46 -	Superposição estrutural entre os dímeros descritos para as estruturas da galectina-1 humana (LOPES-LUCENDO <i>et al.</i> , 2004) e da galectina-9 de camundongo (NAGAE, 2006).....	77
Figura 47 -	Perfil de eluição cromatográfico do domínio CRD-I da galectina-4 em coluna de exclusão molecular SEPHACRYL S100 (Ge Healthcare).....	78
Figura 48 -	Perfil de eluição cromatográfico para padrões de massa molecular e a correspondente curva de calibração da coluna de exclusão molecular SEPHACRYL S100 (Ge Healthcare). Dados gentilmente cedidos por Ricardo Augusto Pereira de Pádua.....	78

LISTA DE TABELAS

Tabela 01-	Protocolo para realização de reação de PCR do domínio CRD-I.....	20
Tabela 02 -	Protocolo para realização de reação de PCR do domínio CRD-II.....	20
Tabela 03 -	Protocolo para realização da reação de digestão do domínio CRD-I.....	21
Tabela 04 -	Protocolo para realização da reação de digestão do domínio CRD-II.....	22
Tabela 05 -	Protocolo para realização da reação de digestão do vetor pET29a.....	22
Tabela 06 -	Protocolo para realização da reação de digestão do domínio CRD-I.....	23
Tabela 07 -	Protocolo para realização da reação de digestão do domínio CRD-II.....	23
Tabela 08	Protocolo para realização da reação de digestão do vetor pET29a.....	23
Tabela 09 -	Protocolo para realização da reação de ligação do CRD-I no vetor pET29a.....	24
Tabela 10 -	Protocolo para realização da reação de ligação do CRD-II no vetor pET29a.....	24
Tabela 11 -	Protocolo para realização da reação de PCR de colônia.....	26
Tabela 12 -	Protocolo para realização de reação de PCR do domínio CRD-I.....	32
Tabela 13 -	Protocolo para realização de reação de PCR do domínio CRD-II.....	32
Tabela 14 -	Protocolo para realização de reação de adenilação dos domínios CRD-I.....	33
Tabela 15 -	Protocolo para realização de reação de adenilação dos domínios CRD-I.....	34
Tabela 16 -	Protocolo para realização da reação de ligação do CRD-I ao vetor pTZ57R/T.....	34
Tabela 17 -	Protocolo para realização da reação de ligação do CRD-II ao vetor pTZ57R/T.....	35
Tabela 18 -	Protocolo para realização da reação de PCR de colônia (CRD-I).....	36
Tabela 19 -	Protocolo para realização da reação de PCR de colônia (CRD-II).....	36
Tabela 20 -	Protocolo para realização da reação de digestão da Construção Hgal-4CRD-I(pTZ57R/T)(I-1).	37
Tabela 21 -	Protocolo para realização da reação de digestão da construção Hgal-4CRD-II(pTZ57R/T)(II-2).	37
Tabela 22 -	Protocolo para realização da reação de digestão do vetor pET28a (Novagen).....	38
Tabela 23 -	Protocolo para realização da reação de ligação do domínio CRD-I no vetor pET28a	39
Tabela 24	Protocolo para realização da reação de ligação do domínio CRD-II ao vetor pET28a.	39
Tabela 25 -	Protocolo para realização da reação de digestão (CRD-I).....	40
Tabela 26 -	Protocolo para realização da reação de PCR (CRD-II).....	41
Tabela 27 -	Dados cristalográficos e estatísticos para os diferentes conjuntos de dados coletados no LNLS ($\lambda=1,437 \text{ \AA}$), a 100 K.	71

LISTA DE ABREVIATURAS

C	Carbono
CRDs	Domínio de reconhecimento de carboidratos
dNTPs	Desoxinucleosídeos -fosfato
DNA	<i>Deoxiribonucleic acid</i>
DO	Densidade ótica
Fuc	Fucose
Gal-4	Galectina-4
Gal	Galactose
GalNac	<i>N</i> -acetilgalactosamina
Glc	Glicose
GlcNac	<i>N</i> -acetilglicosamina
IPTG	Isopropiltio- β -D-galactosídeo
K _d	Constante de dissociação
Lac	Lactose
LB	Luria-Bertani
Mol	Moles
SMC	Sítio múltiplo de clonagem
NMR	<i>Nuclear Magnetic Resonance</i>
PDB	<i>Protein Data Bank</i>
PI	Ponto isoeletrico
PMSF	Fluoreto fenilmetilsulfonil
RNA	<i>Ribonucleic acid</i>
SDS	Dodecil sulfato de sódio
SDS-PAGE	Eletroforese em gel de poliacrilamida contendo SDS
TEMED	N,N,N',N' - tetrametiletilenodiamina

LISTA DE SÍMBOLOS

α	Letra grega alfa, encontrada em “ α -Hélice”
β	Letra grega beta, encontrada em “Folha β ”
Å	Angstrom λ
λ	Letra grega lambda, comprimento de ondas

SUMÁRIO

RESUMO	i
ABSTRACT	ii
LISTA DE FIGURAS	iii
LISTA DE TABELAS	vii
LISTA DE ABREVIATURAS	viii
LISTA DE SIMBOLOS	ix
1. INTRODUÇÃO	1
1.1 Galectinas.....	1
1.2 Galectina-4.....	9
2. OBJETIVOS	14
2.1 Objetivos Específicos	14
3. MATERIAIS E MÉTODOS	15
3.1 Clonagem dos CRD-I e CRD-II da HGal-4 em vetor de expressão pET29(Novagen).....	15
3.1.1 Planejamento dos oligonucleotídeos.....	16
3.1.2 Amplificação dos CRDs da Galectina-4 humana.....	20
3.1.3 Digestão dos CRDs da Galectina-4 humana e do vetor pET29a (Novagen).....	21
3.1.4 Ligação dos CRDs da Galectina-4 humana no vetor pET29a (Novagen).....	23
3.1.5 Transformação das construções Hgal-4CRD-I(pET29a) e Hgal-4CRD- II (pET29a) em bactérias Escherichia coli DH5α	25
3.1.6 Expressão dos domínios CRD-I e CRD-II da Galectina-4 humana.....	26
3.1.6.1 Expressão do domínio CRD-II.....	26
3.1.6.2 Análise do padrão de expressão do domínio CRD-II.....	27
3.1.6.3 Expressão do domínio CRD-I.....	27
3.1.6.4 Análise do padrão de expressão do domínio CRD-I.....	28
3.1.7 Purificação dos domínios CRD-I e CRD-II da galectina-4 humana.....	28
3.1.7.1 Cromatografia por Afinidade.....	29

3.2 Clonagem dos domínios CRD-I e CRD-II da galectina-4 humana em vetor de expressão pET28a(Novagen).....	29
3.2.1 Estratégia.....	30
3.2.2 Amplificação dos domínios CRD-I e CRD-II da galectina-4 humana.....	31
3.2.3 Processo de Adenilação dos CRDs da galectina-4 humana.....	33
3.2.4 Ligação dos CRDs da galectina-4 no vetor de subclonagem pTZ57R/T.....	34
3.2.5 Transformação das construções Hgal-4CRD-I(pTZ57R/T) e Hgal-4CRD-II (pTZ57R/T) em bactérias Escherichia coli DH5 α	35
3.2.6 Processo de digestão das construções Hgal-4CRD-I(pTZ57R/T) (I-1) e Hgal-4CRD-II (pTZ57R/T) (II-2) e do vetor pET-28a (Novagen).....	37
3.2.7 Ligação dos CRDs da galectina-4 humana no vetor pET28a (Novagen).....	38
3.2.8 Transformação das construções Hgal-4CRD-I(pET28a) e Hgal-4CRD-II(pET28a) em bactérias Escherichia coli DH5 α	39
3.2.9 Expressão e purificações domínios CRD-I e CRD-II da galectina-4 humana.....	41
3.2.9.1 Expressão e obtenção do domínio CRD-I da galectina-4 humana.....	41
3.2.9.2 Expressão e obtenção do domínio CRD-II da galectina-4 humana.....	42
3.2.10 Purificação dos CRD-I e CRD-II da galectina-4 humana.....	43
3.2.10.1 Cromatografia por Afinidade (Ni-NTA).....	43
3.2.11 Dosagem dos domínios CRD-I e CRD-II da galectina-4 humana.....	45
3.2.11.1 Método do coeficiente de extinção molar.....	45
3.3 Cristalização dos domínios CRD-I e CRD-II da galectina-4 humana.....	45
3.3.1 Cristalização do domínio CRD-I da galectina-4 na forma “apo”.....	46
3.3.2 Cristalização do domínio CRD-I da galectina-4 com lactose.....	47
3.3.3 Experimentos de “soaking” do domínio CRD-I com lactose.....	47
3.3.4 Cristalização do domínio CRD-II da galectina-4 na forma “apo”.....	48
3.4 Coleta e Processamento de Dados para cristais do domínio CRD-I.....	48
3.5 Resolução da Estrutura do domínio CRD-I.....	48

3.6 Refinamento.....	49
3.7 Avaliação dos dados estereoquímicos.....	51
4. RESULTADOS	52
4.1 Clonagem dos CRD-I e CRD-II em vetor de expressão pET-29a (Novagen).....	52
4.1.1 Amplificação dos CRD-I e CRD-II da galectina-4 humana.....	52
4.2 Expressão dos domínios CRD-I e CRD-II da galectina-4 humana em <i>E.coli</i> BL-21(DE3)pLys	55
4.2.1 Expressão do domínio CRD-II da galectina-4 humana.....	55
4.2.2 Expressão do domínio CRD-I da galectina-4 humana.....	57
4.3 Purificação dos domínios CRD-I, CRD-II e da galectina-4 inteira.....	58
4.4 Subclonagem dos CRD-I e CRD-II em vetor pTZ57R/T (Fermentas).....	59
4.4.1 Amplificação dos CRD-I e CRD-II da galectina-4 humana.....	59
4.4.2 Processo de Adenilação dos CRDs da Galectina-4 humana.....	60
4.4.3 Processo de digestão das construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II (pTZ57R/T)(II-2) e do vetor pET-28a (Novagen).....	62
4.4.4 Processo de ligação dos domínios CRD-I e CRD-II no vetor pET-28a (Novagen)	63
4.4.5 Expressão e purificação dos domínios CRD-I e CRD-II em <i>E.coli</i> BL-21(DE3)..	65
4.4.5.1 Expressão e purificação do domínio CRD-I da galectina-4 humana em <i>E.coli</i> BL-21(DE3).....	65
4.4.5.2 Expressão e purificação do domínio CRD-II da galectina-4 humana em <i>E.coli</i> BL-21(DE3).....	66
4.5 Cristalização do domínios CRD-I e CRD-II da galectina-4 humana.....	67
4.5.1 Cristalização do domínio CRD-I na presença de Lactose.....	67
4.6 Coleta, Processamento de Dados e Resolução da Estrutura.....	69
4.7 Refinamento.....	72
4.8 Avaliação dos dados estereoquímicos.....	73
4.9 Estrutura cristalográfica do domínio CRD-I da galectina-4 humana.....	74

5. DISCUSSÕES E CONCLUSÕES	80
6. REFERÊNCIAS BIBLIOGRÁFICAS	84
APÊNDICES	94

1. INTRODUÇÃO

1.1 *Galectinas*

Dentro do vasto e complexo mundo das macromoléculas biológicas, as lectinas correspondem a um importante grupo de proteínas cuja função é reconhecer estruturas de carboidratos e glicoconjugados. Através da habilidade em discriminar entre uma miríade de estruturas de carboidratos encontrados na superfície das células, em matrizes extracelulares e anexadas a glicoproteínas solúveis, as lectinas desempenham um papel vital em vários processos celulares como: reconhecimento de células-alvo, adesão celular, interação entre células, aglutinação, desenvolvimento animal, fisiologia e em vários estágios patológicos (DRICKAMER; TAYLOR, 1993).

As lectinas são proteínas cuja distribuição é ubíqua, sendo encontradas em vírus, bactérias, plantas e animais. Ao longo de mais de um século, as lectinas de origem vegetal foram extensivamente estudadas e caracterizadas (LIS & SHARON ANNU VER, 1986). Mais recentemente, as pesquisas têm também focado o estudo de lectinas de origem animal e sua capacidade de reconhecer moléculas de carboidrato de origem endógena ou exógenas as células. Até o presente momento, mais de 100 lectinas animais tiveram sua estrutura primária identificada e alguns poucos exemplos de estrutura terciária determinada (BIANCHET *et al.*, 2000; LOPEZ-LUCENDO *et al.*, 2004; WALSER *et al.*, 2004; LOBSANOV *et al.*, 1993; SEETHARAMAN *et al.*, 1998; LEONIDAS *et al.*, 1998; ACKERMAN *et al.*, 2002; MASAMICHI *et al.*, 2006.). Uma análise global revela que embora a arquitetura global de uma lectina possa variar enormemente, a capacidade destas proteínas em se ligar à moléculas de carboidratos são confinadas a um domínio específico da proteína designado por domínio de reconhecimento de carboidratos (CRD's) (DRICKAMER; TAYLOR, 1993). Os CRDs compartilham alta similaridade em termos de estrutura primária, no entanto, exibem distintas características com relação à natureza de seus ligantes, sua dependência por cátions bivalentes, sua localização celular e nos processos biológicos nos quais esses participam (DRICKAMER; TAYLOR, 1993). Por exemplo, os CRD's de lectinas do tipo-C, encontradas preferencialmente em ambientes extracelulares como membrana plasmática e matriz extracelular, dependem do íon cálcio para sua atividade e se ligam a diferentes tipos de açúcares. Esta classe de lectinas animais está associada às diversas funções

celulares, atuando como receptores de membrana que mediam endocitose de glicoproteínas, receptores de manose para macrófagos e células endoteliais hepáticas que mediam a ligação e internalização de glicoconjugados.

Já as lectinas do tipo-P se ligam especificamente a molécula de manose 6-fosfato (M6P) e seus CRDs são os mais distintos entre as classes de lectinas existentes. (DAHMS; HANCOCK, 2002). Como exemplo desta classe de proteínas, estão as pentraxinas, que fazem parte de uma família de proteínas plasmáticas altamente conservadas incluindo as pentraxinas “curtas”, o componente P do amiloide sérico (SAP), a proteína reativa C (CRP) e a Pentraxina 3. Estas participam do processo inflamatório durante a resposta de fase aguda e no processo de defesa do hospedeiro (GEWURZ *et al.*, 1995). As lectinas do tipo-P são caracterizadas pela dependência por cálcio e pela capacidade de formar grandes oligômeros (EMSLEY *et al.*, 1994).

As galectinas, originalmente denominadas de lectinas do tipo-S, formam uma classe importante de lectinas animais onde os domínios de reconhecimento por carboidratos apresentam afinidade de ligação específica por carboidratos do tipo β -galactosídeos (BARONDES *et al.*, 1994; KASAI; HIRABAYASHI, 1996; COOPER, 2002; COOPER; BARONDES, 1999). Membros desta classe de proteínas são encontrados em nematódeos, insetos e mamíferos com diversas funções celulares a elas atribuídas. As galectinas são proteínas solúveis, estão presentes em uma grande variedade de tecidos, em ambientes extracelular e intracelular, incluindo o núcleo. Uma vez que a maioria dos glicoconjugados conhecidos estão localizados extracelularmente, não existem ainda dados concretos que expliquem porque proteínas que supostamente devem interagir com glicoconjugados são encontradas intracelularmente. No entanto, com base na estrutura protéica, características como N-terminal acetilado, resíduos de cisteína no estado livre e ausência de sequência sinal (OHYAMA *et al.*, 1986; HYRABAYASHI *et al.*, 1987; HYRABAYASHI *et al.*, 1988) sugerem que as galectinas possuem algum papel biológico dentro das células. Por não possuírem um peptídeo sinal, as galectinas são liberadas para o ambiente extracelular por um mecanismo ainda não bem compreendido. Ensaio imunológicos e histoquímicos relatam que as galectinas encontram-se concentradas em evaginações formadas pela membrana plasmática, sugerindo desse modo um possível mecanismo para externalização dessas proteínas. Supõe-se que a localização das galectinas, assim como sua especificidade por diferentes carboidratos está relacionada à sua função biológica (LEFFLER *et al.*, 2004; PATTERSON *et al.*, 2004).

A função biológica das galectinas varia dependendo do local, tempo e circunstância. O grande número de situações no qual se supõe o envolvimento da galectina faz no momento impossível definir um papel geral para as galectinas. Na verdade, as galectinas participam em uma variedade de fenômenos fisiológicos e patológicos que podem ser divididos em dependentes e independentes da ligação com carboidratos (LIU *et al.*, 2002). Das muitas atividades biológicas atribuídas às galectinas estão; progressão do câncer (GRASSADONIA *et al.*, 2004; BIDON-WAGNER *et al.*, 2004; TAKENAKA *et al.*, 2004; VAN DEN BRULE *et al.*, 2004), em processos específicos de desenvolvimento normal de tecidos, processos de regeneração e cicatrização (HUGHES *et al.*, 2004; WATT *et al.*, 2004) e participação na regulação imunológica e inflamatória (HIRASHIMA *et al.*, 2004; SATO *et al.*, 2004; RABINOVITCH *et al.*, 2004; PACE *et al.*, 2004; ALMLVIST *et al.*, 2004). Além disso, estudos têm demonstrado que as galectinas participam no desenvolvimento de neurite, inibição do crescimento de células não-neurais (WELLS; MALLUCCI, 1991; ADAMS *et al.*, 1996; KOPITZ *et al.*, 2001), estimulam o crescimento celular (PITTS; YANG, 1981; SANFORD; HARRISHOOKER, 1990), induzem apoptose de timócitos imaturos (PERILLO, *et al.*, 1997; GALVAN, *et al.*, 2000), células T ativadas (PERILLO *et al.*, 1995; PACE *et al.*, 1999), gênese tumoral, processos auto-imunes, diferenciação, morfogênese, metástase de tumor, apoptose e função imunorregulatória (ELOLA *et al.*, 2005; RABINOVICH *et al.*, 2002; RUBISTEIN *et al.* 2004).

Uma característica particular das galectinas é a presença de resíduos de cisteína, onde não se é observado a presença de pontes dissulfeto e os grupos SH se encontram livres. Uma vez que a capacidade de ligação aos açúcares é perdida na ausência do agente redutor β -mercaptoetanol, inicialmente se supunha que a presença dos grupos SH livres era essencial para sua atividade (LEVI; TEICHBERG, 1981). Contudo, estudos de mutação sítio-dirigida mostraram que a presença dos grupos SH não é essencial para a ligação com o açúcar (HIRABAYASHI; KASAI, 1991). Na verdade, a relação entre as cisteínas e a atividade se baseia no fato que a oxidação dos grupos SH causa uma mudança conformacional que resulta na perda da habilidade da galectina em se ligar ao açúcar (TRACEY *et al.*, 1992).

Com base no número e na organização dos CRD's, os membros da família das galectinas são classificados em três subgrupos: o grupo proto, o grupo quimera e o grupo "tandem repeat" (**Figura 1**) (KASAI; HIRABAYASHI, 1996). Das 15 galectinas descritas até o momento (LIU *et al.*, 2005) os membros do tipo proto, compreendendo as galectinas-1,-2,-5,-7,-10,-11,-13, -14 e 15, são caracterizadas por um único domínio CRD de

aproximadamente 15 kDa (130 aa), adicionado a uma sequência N-terminal curta. Membros do grupo proto existem tanto como homodímeros em condições fisiológicas, como observado para galectinas-1, -2, -11, -14, e 15 (LOPES-LUCENDO *et al.*, 2004, LOBSANOV *et al.*, 2003; LEONIDAS *et al.*, 1998) ou como monômeros, observado para galectina-5, -7 e -10 (GITT *et al.*, 1995). As galectinas do tipo proto normalmente apresentam atividade de hemaglutinação.

A galectina-3, o único membro do subgrupo químera, e somente descrita em mamíferos, possui massa molecular de aproximadamente duas vezes a encontrada para membros do grupo proto (30 a 35 kDa) e é composta por dois diferentes domínios (**Figura 1**). O grupo químera tem como característica a presença de um domínio CRD na região C-terminal, enquanto que a metade N-terminal consiste em domínio pequeno conectado a uma extensão rica em aminoácidos de prolina, tirosina e glicina pela qual a proteína é capaz de formar oligômeros (DUMIC *et al.*, 2006; MORRIS *et al.*, 2004). Este domínio, rico em resíduos de prolina, tirosina e glicina, consiste de repetições de 7 a 10 aminoácidos com uma sequência consenso do tipo PGAYPG(X)₁₋₄ (onde X é qualquer aminoácido) e é similar a domínios encontrados em proteínas que possuem a característica de auto agregação (MASSA *et al.*, 1993). O número de monômeros do oligômero varia entre as espécies resultando em proteínas de tamanhos diferentes. Estudos apontam esta característica estar diretamente relacionada a localização celular desta proteína (MAHONEY *et al.*, 2000; MEHUL *et al.*, 1995; GAUDIN *et al.*, 2000; MENON *et al.*, 1995).

O grupo *tandem repeat*, é formado por proteínas que apresentam dois CRDs distintos, conectados por um peptídeo de ligação (galectinas 4, 6, 8, 9 e 12) (MORRIS, *et al.*, 2004). A análise dos domínios de reconhecimento de carboidratos para membros pertencentes a este grupo de proteínas permitiu observar que os CRDs se encontram posicionados de tal forma a permitir a ligação simultânea com distintos ligantes (ARATA *et al.*, 1997). Além disso, estudos demonstram que este peptídeo de ligação encontrado no grupo *tandem-repeat* pode apresentar um número variável de resíduos, mas que alguns resíduos são altamente conservados entre as subfamílias. Na verdade, a estrutura do peptídeo de ligação, através das características observadas quanto à proximidade ou orientação frente aos CRDs parece influenciar nas propriedades de ligação frente a diferentes ligantes (COOPER, 2002).

Figura 1: Esquema ilustrativo dos três tipos de arquitetura das galectinas: protótipo, quimera e *tandem-repeat*. O tipo proto é composto somente de um domínio lectínico, o CRD. O tipo quimera 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 CRD's homólogos. Figura adaptada do site: <http://www.glycoforum.gr.jp/science/word/lectin/LEA01E.html>.

Em termos estruturais, a cristalografia de proteínas tem sido extensivamente utilizada para determinar o enovelamento deste núcleo comum para diferentes galectinas e de uma forma geral, a análise estrutural de membros desta família como galectinas de mamíferos (LOBSONOV *et al.*, 1993; BOURNE *et al.*, 1994; LEÔNIDAS *et al.*, 1995; SEETHARAMAN *et al.*, 1998; SWAMINATHAN *et al.*, 1999; MASAMICHI *et al.*, 2006.), galectina de frango (VARELA *et al.*, 1999), de sapo (BIANCHET *et al.*, 2000), e de enguia (SHIRAI *et al.*, 1999) mostram que esta família de proteínas se enovela em um arranjo globular similar consistindo de duas folhas beta antiparalelas formadas por cinco e seis fitas betas cada (denominadas de F1 a F5 e S1 a S6) (**Figura 2**).

Figura 2: Estrutura tridimensional do domínio de reconhecimento de carboidrato (CRD) encontrado nas galectinas. Modelo construído com base nas coordenadas da estrutura da galectina-1 humana (LOPES-LUCENDO *et al.*, 2004) complexada com uma molécula de lactose, **PDB código 1ZW**. Representação construída no programa *Pymol* (DELANO, 2002).

O núcleo comum compartilhado pelas galectinas, CRD, consiste de aproximadamente 130 aminoácidos, muitos dos quais são altamente conservados (**Figura 3**) (COOPER; BARONDES, 1999). A identidade sequencial entre os CRDs das várias galectinas de mamíferos varia de 20 a 40% (ODA *et al.*, 1993) enquanto que a identidade da mesma galectina de diferentes espécies é de 80 a 90%. Os resíduos dos CRDs das galectinas diretamente envolvidos na ligação com o carboidrato são parte de um bolsão formado por quatro fitas β adjacentes (S3, S4, S5 e S6a e S6b) (**Figura 4**). Além disso, a afinidade e a especificidade observada para as galectinas frente aos diversos carboidratos é consequência da distribuição e características dos aminoácidos que compõem cada CRD. Um exemplo é a galectina-3, onde os sítios de ligação a carboidratos, lactose (Lac), N-acetil-lactosamina (LacNAc) e N-acetil-glicosamina (GlcNAc) são, como esperado, formados pelas cadeias β S4-S6a/S6b (SEETHARAMAN *et al.*, 1998), no entanto os aminoácidos envolvidos na interação da galectina-3 com Lac ou LacNAc são: Asn-160, Arg-162, His-158, Glu-184, Asn-174 e Trp-181, enquanto que a interação com N-acetil-glicosamina (GlcNAc) explora outro grupo de resíduos do CRD, Trp-181, Arg-186, Arg-162 e Glu-184 (SEETHARAMAN *et al.*,

1998) (**Figura 5**). No caso da galectina-1, os aminoácidos envolvidos na interação com Lac ou LacNac são Arg-48 e Glu-71, enquanto que a interação com N-acetil-glicosamina (GlcNac) envolvem os resíduos Arg-73, Asp-54 e His-52 (LOPES-LUCENDO *et al.*, 2004). Na galectina-2, os resíduos envolvidos na interação com Lac são His- 45, Asn47, Arg 49, Glu 68, Arg 70 e Trp 65. (LOBSANOV *et al.*, 2003) (**Figura 5**).

Figura 3. Representação esquemática das estruturas proto, quimera e *tandem-repeat* onde os resíduos conservados presentes no CRD estão indicados. Os resíduos indicados em maiúsculo indicam os principais resíduos envolvidos na coordenação com o carboidrato. Os resíduos indicados com asterisco são os resíduos mais conservados em todas as galectinas até o presente momento, a letra c indica a posição dos resíduos de cisteína, cuja oxidação resulta na perda de afinidade de ligação com o açúcar. A numeração corresponde à utilizada para descrever a galectina-1 humana. Adaptada de KASAI; HIRABAYASHI, 1996.

Embora os dados relatados na literatura, sejam de caracterização funcional em nível molecular ou com relação ao seu envolvimento das galectinas em mecanismos celulares e

patológicos apoiam a importância do estudo das galectinas, fica também evidente que este campo de pesquisa poderá ainda responder a muitas perguntas que continuam sem resposta. Por exemplo: Qual a principal função das galectinas? Por que elas são encontradas extracelularmente apesar de elas terem sido desenhadas como proteínas intracelulares? Como elas são externalizadas das células? Por que as galectinas são específicas para β -galactosídeos? Quais são seus ligantes endógenos intra e extracelulares? Por que elas estão localizadas nos sítios onde o risco de inativação oxidativa é alta? Por que galectinas e lectinas de plantas possuem a mesma topologia? As respostas a estas questões certamente contribuirão não somente para compreender o papel das galectinas, como também, compreender o significado do complexo conjunto de glicoconjugados que circundam as células.

Neste trabalho, a biologia estrutural, aliando técnicas de biologia molecular, bioquímica e cristalografia de raios-X, tem um importante papel na caracterização funcional desta classe de biomoléculas. Neste projeto, visamos a caracterização estrutural preliminar dos domínios da galectina-4 humana. Nosso objetivo é que através dos estudos de correlação entre estrutura e função, o resultado deste trabalho possa ser utilizado para o mapeamento das interações com as moléculas de carboidrato, principalmente aquelas responsáveis pela especificidade, e a caracterização de potenciais resíduos envolvidos na localização celular.

1.2 *Galectina-4*

Dentre as galectinas de mamíferos conhecidas, a galectina-4 (Gal-4) pertence ao subgrupo de proteínas *tandem-repeat* (**Figura 1**). Correspondendo a um monômero de aproximadamente 36 kDa, contém dois domínios de reconhecimento de carboidratos (CRD-I no N-terminal e CRD-II no C-terminal) conectados por um peptídeo ligante, em uma cadeia peptídica única (ODA, Y. *et al.*, 1993; HUFLEJT *et al.*, 1997) (**Figura 6**). A sequência de aminoácidos correspondente a cada domínio foi inicialmente sugerida por Huflejt e colaboradores (HUFLEJT *et al.*, 1997). As regiões denominadas de CRD-I e CRD-II correspondem aproximadamente aos resíduos de 1 a 150 (peso molecular estimado de 17 kDa, com pI teórico estimado em aproximadamente 7), e aos resíduos 179 a 323 (peso molecular estimado de 16 kDa, com pI teórico estimado em aproximadamente 9,7), respectivamente. 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) (**Figura 6**).

1	MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDVAFHF	64
65	NPRFDGWDKVVFNLTQGGKWGSEERKRSMFPKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRL	128
129	PLQMVTHLQVDGDLQLQSINFI	150
151	<u>GGQPLRPQGPPMPPYPGPGHCHQQ LNS</u>	178
179	LPTMEGPPTFNPPVPYFGRLQGGLTARRTIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGNGTV	246
247	VRNSLLNGSWGSEEEKITHNPFPGQGFFDLSIRCGLDRFKQVANGQHLFDFAHRLSA FQRV DTLEI	312
313	QGDVTLSTYVQI	323

Figura 6: Estrutura primária da galectina-4 humana (HUFLEJT *et al.*, 1997). Os resíduos em amarelo representam os aminoácidos que interagem com o carboidrato. O resíduo em pink representa a única substituição conservativa observada para os resíduos que coordenam moléculas de carboidrato e os resíduos em vermelho pertencem ao peptídeo de ligação.

Na Gal-4 humana, a sequência dos dois CRDs compartilha aproximadamente 34,0% de identidade sequencial (RECHRECHE *et al.*, 1997). Na verdade, dos 7 resíduos envolvidos na interação com o carboidrato, característicos dos domínios CRDs, somente o resíduo de arginina 261 sofre uma mutação conservativa por uma lisina no CRD-II (**Figura 7**).

Figura 7. Alinhamento sequencial entre os CRD-I e CRD-II encontrado na galectina-4 humana. O alinhamento foi realizado no programa *MultAlin* (CORPET, 1988) e a imagem foi criada no programa *ESPrIt* (GOUET *et al.*, 1999). Os aminoácidos supostamente envolvidos na coordenação com moléculas de carboidrato estão identificados por ▼. A mutação conservativa identificada entre CRD-I e CRD-II está representado por ▲

Comparando os CRDs da Gal-4 com outras galectinas, observamos que estes compartilham uma identidade sequencial de aproximadamente 35% com o CRD da galectina-3, 27% com a galectina-1 e 35% com a galectina-7 humana (RECHRECHE *et al.*, 1997). Os domínios de reconhecimento de carboidratos nas galectinas-4 são altamente conservados entre

os membros da família (RECHERCHE *et al.*, 1997), revelando alta similaridade entre a galectina-4 humana e de porco (79% de identidade) e com 76% de rato (RECHERCHE *et al.*, 1997) (**Figura 8**).

Figura 8. Alinhamento sequencial entre as galectinas-4 humana, de rato e de porco. As setas em azul indicam os resíduos que participam da interação com o carboidrato. As setas em verde indicam as substituições conservativas. O alinhamento foi realizado no programa *MultAlin* (CORPET, 1988), a figura foi criada no programa *ESPrIt* (GOUET *et al.*, 1999).

Ambos os domínios CDR-I e CDR-II da Gal-4 ligam lactose (Gal β 1 $\rightarrow$ 4Glc) com afinidade similar a descrita para outras galectinas (k_d de aproximadamente 0.5-1 mM) mas suas respectivas preferências por outros dissacarídeos são diferentes. Estudos comparativos entre os CRD-I e CRD-II da galectina-4 de rato descrevem as diferentes especificidades frente aos carboidratos (ODA *et al.*, 1993). O resultados demonstram que o domínio CRD-I (N-terminal) 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 $\rightarrow$ 3GalNac). Em contraste, o domínio CRD-II (C-terminal) possui maior afinidade por A-tetra (GalNac α 1 $\rightarrow$ 3-(fuc α 1 $\rightarrow$ 2)Gal β 1 $\rightarrow$ 4) (ODA *et al.*, 1993). Os resultados demonstram que os CRDs possuem similar afinidade por

oligossacarídeos dos tipos $\text{SO}_3^- \rightarrow 3\text{Lac-O-pNP}$, A-tetra ($\text{GalNAc}\alpha 1 \rightarrow \alpha 1 \rightarrow 2\text{Gal}\beta 1 \rightarrow 4\text{Glc}$) e LNF-1 ($\text{Fuc}\alpha 1 \rightarrow 2\text{Gal} \beta 1 \rightarrow 3\text{GlcNAc} \beta 1 \rightarrow 3\text{Gal} \beta 1 \rightarrow 4\text{Glc}$), por outro lado, do CRD-II apresentou maior afinidade por $\text{SO}_3^- \rightarrow \text{core } 1$ ($\text{Gal}\beta 1 \rightarrow 3\text{GalNAc}$)-*O-Bn* (IDEO *et al.*, 2005).

A galectina-4 é expressa no epitélio da mucosa oral, do esôfago, do intestino e colônica (ODA *et al.*, 1993; RECHRECHE *et al.*, 1997). 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; HANSEN *et al.*, 2001). A Gal-4 tem sido detectada como proteína de adesão celular, no epitélio da mucosa oral, no esôfago, e tem sido encontrada mais especificamente na membrana citoplasmática (WASANO; HIRAKAWA, 1999). Estudos recentes demonstram que a Gal-4 se liga a distintas estruturas nas bordas intercelulares do epitélio coloretal. Os domínios de reconhecimento de carboidratos na molécula reconhecem distintos ligantes, sendo que o domínio N-terminal possui afinidade por ligantes localizados em espaços intercelular, enquanto o domínio C-terminal se encontra associado a ligantes presentes na membrana celular. Este fato sugere que a Gal-4 possui como possível função mediar interações entre glicoconjugados presentes na membrana citoplasmática em células do epitélio intestinal e glicoconjugados encontrados em espaço intercelular (WASANO; HIRAKAWA, 1999).

Com relação ao seu envolvimento em fenômenos patológicos, a maioria dos estudos que relacionam tumor à galectina-4, reportam apenas modificações nos níveis de RNAm. Em câncer de cólon intestinal, a expressão de RNAm da Gal-4 foi encontrada estar abaixo dos níveis normais, enquanto estava presente em altas concentrações em carcinoma hepatocelular (KONDOH *et al.*, 1999) e em câncer gástrico, com aumento do poder metastático (HIPPO *et al.*, 2001) quando comparado com o tecido normal. O diferencial na expressão do RNAm da Gal-4 em câncer de cólon indica a participação desta lectina na malignidade do câncer coloretal (RECHRECHER *et al.*, 1997; LAHML *et al.*, 2001). Embora o papel de algumas galectinas em metástase tumoral já ter sido constatado, a participação da Gal-4, assim como a importância das diferentes especificidades de cada um dos seus CRDs neste processo ainda é desconhecida (LIU *et al.*, 2005). Recentemente, trabalhos demonstram que a Gal-4 está alterada nos episódios de malignidade de células humanas (HUFLEJT *et al.*, 1997; RECHRECHE *et al.*, 1997; KONDOH *et al.*, 1999; HUFLEJT *et al.*, 2004; LAHM *et al.*, 2001; HIPPO *et al.*, 2001; WOLLINA *et al.*, 2002; NAGY *et al.*, 2003; BIDON-WAGNER *et al.*, 2004; HIRASHIMA *et al.*, 2004; RUMILLA *et al.*, 2006), sendo identificada como alvo de anticorpos autólogos em pacientes com câncer coloretal (SCANLAN *et al.*, 1998).

A expressão mais controlada da Gal-4 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). Além disso, a Gal-4 não é apenas um marcador de malignidade, mas também promove a carcinogenicidade, e a inibição de sua atividade pode ser um componente importante na terapia anti-câncer (HUFLEJT *et al.*, 1997; SAUSSEZ *et al.*, 2005).

Caracterizar os CRDs da Gal-4, assim como, determinar a especificidade frente diferentes ligantes naturais e específicos de cada domínio de reconhecimento de carboidratos é um passo importante para compreender a ação fisiológica da galectina-4. Sendo assim, nosso projeto tem como objetivo o estudo funcional da galectina-4 humana através da análise estrutural dos domínios de reconhecimento de carboidratos CRDs. O projeto, aqui apresentado, propõe a clonagem, expressão, purificação e cristalização de cada domínio de forma independente. A estratégia de trabalhar com os CRDs isoladamente também favorece os estudos estruturais por difração de raios-X, uma vez que a presença do peptídeo que conecta os dois CRDs introduz uma grande flexibilidade na proteína dificultando os experimentos de cristalização. Além disso, a expressão heteróloga dos CRDs abrirá a oportunidade de realização de novos ensaios biológicos e bioquímicos, como a produção de anticorpos mono e policlonais, imunodeteção da Gal-4 em tecidos normais e em tumores, estudos da atividade dos CRDs de modo independente e a comparação destes com os estudos de atividade realizados para a Gal-4 inteira.

Os estudos propostos neste projeto constituem um passo importante para um melhor entendimento dos processos de desenvolvimento de câncer e metástase, e conseqüentemente, contribuir para o planejamento de novas estratégias de tratamento em processos envolvendo neoplasias.

2. OBJETIVOS

2.1 *Objetivo Geral*

O objetivo deste trabalho é dar início aos estudos de caracterização estrutural dos domínios de reconhecimento de carboidratos, CRD-I (no N-terminal) e CRD-II (no C-terminal), da galectina-4 humana.

2.2 *Objetivos Específicos*

- Clonagem dos domínios CRD-I e CRD-II da galectina-4 humana;
- Expressão heteróloga dos domínios CRD-I e CRD-II na bactéria *E. coli*;
- Purificação dos domínios CRD-I e CRD-I;
- Cristalização dos domínios CRD-I e CRD-II;
- Coleta e processamento de dados através de técnicas de difração de raios-X em monocristais;
- Resolução da estrutura cristalográfica dos CRDs por técnicas de substituição molecular;
- Estudos de correlação entre estrutura e função.

3. MATERIAIS E MÉTODOS

3.1 Clonagem dos CRD-I e CRD-II da HGal-4 em vetor de expressão pET29(Novagen)

As seqüências de nucleotídeos que codificam os domínios de reconhecimento de carboidratos (CRD-I e CRD-II) da galectina-4 humana foram inicialmente clonadas no vetor de expressão pET29a (Novagen), (**Figura 9**). Os vetores de expressão do sistema pET utilizam-se de vários elementos importantes para o controle da transcrição gênica como o gene *lac I* que codifica a proteína repressora que se liga ao operador *lac*, também presente para bloquear a transcrição e o promotor de transcrição T7 o qual é específico para T7 RNA polimerase (e não para a RNA polimerase de bactéria). Além disso, o vetor possui uma origem de replicação (**O**), uma seqüência de DNA que permite que o vetor seja replicado na célula hospedeira; um sítio denominado de Múltiplo Sítio de Clonagem (**MSC**), local onde o gene de interesse é incorporado ao vetor de clonagem, assim como, um gene que confere resistência ao antibiótico kanamicina. O controle do sistema de expressão pET é feito através do promotor e operador *lac* que controla a expressão tanto da T7 RNA polimerase, presente na célula hospedeira, assim como o nosso gene de interesse clonado no vetor de expressão. Na presença de IPTG, a proteína repressora muda sua conformação e se desliga do operador *lac*, permitindo a transcrição de ambos os genes (CAMPBELL, 2003).

Figura 9. Mapa do vetor de clonagem pET29a (Novagen). A figura mostra o gene *lacI*, o promotor T7, o operador *lac* e o sítio múltiplo de clonagem. O vetor apresenta tamanho aproximado de 5300pb.

3.1.1 Planejamento dos oligonucleotídeos

O objetivo deste trabalho é a obtenção dos CRDs de forma independente, para posteriores estudos de caracterização estrutural, assim como, futuros ensaios cinéticos e imunológicos. Para estudos da determinação estrutural, a presença de um peptídeo de ligação entre os CRDs revelado na análise da sequência dos aminoácidos da proteína Hgal-4 inteira (BARONDES et al., 1994), introduz uma alta flexibilidade à proteína dificultando assim os experimentos de cristalização. Sendo assim, a remoção do peptídeo de ligação teve que ser planejada cuidadosamente, pois esta região poderia de alguma forma interagir com os domínios CRD-I e CRD-II de forma a auxiliar na estabilidade estrutural e funcional de cada unidade.

A primeira etapa do projeto consistiu no planejamento dos oligonucleotídeos para o processo de amplificação dos CRD-I e CRD-II da Gal-4 humana (Hgal-4). A sequência de aminoácidos correspondente a cada domínio foi inicialmente sugerida por Huflejt e colaboradores (HUFLEJT *et al.*, 1997). As regiões denominadas de CRD-I e CRD-II correspondem aos resíduos de 1 a 150 e 179 a 323, respectivamente (**Figura 10**). No caso específico do domínio CRD-I, os oligos foram planejados para amplificar a região limitada pelos resíduos 1 a 150.

Para a amplificação do domínio CRD-II, foi possível utilizar a estrutura deste domínio recentemente determinada por RMN (dados não publicados) e cujas coordenadas se encontram disponíveis no banco de dados de estrutura de proteínas (www.rcsb.org) (**Figura 11**). A estrutura, cujas coordenadas estão disponibilizadas no banco de dados sob o código PDBID 1X50, cobre o intervalo de resíduos de His-173, até Ile-323, contendo, além do CRD-II, 6 resíduos extras pertencentes ao peptídeo de ligação (HQQLNS) (**Figura 10**).

↓ Início da construção do CRD-I		
1	ATGGCCTATGTCCCGCACCGGGCTACCAGCCACCTACAACCCGACGCTGCCTTACTACCAGCCCA	65
	M A Y V P A P G Y Q P T Y N P T L P Y Y Q P I	23
66	TCCCGGCGGGCTCAACGTGGGAATGTCTGTTTACATCCAAGGAGTGGCCAGCGAGCACATGAAGCGGTTCTTCGTGAACCTTTGTGGTGG	155
24	P G G L N V G M S V Y I Q G V A S E H M K R F F V N F V V G	53
156	GGCACGATCCGGGCTCAGACGTCGCCTTCCACTTCAATCCGCGGTTTGACGGCTGGGACAAGGTGGTCTTCAACACGTTGCAGGGCGG	244
54	Q D P G S D V A F H F N P R F D G W D K V V F N T L Q G G	82
245	GAAGTGGGGCAGCGAGGAGAGGAAGAGGAGCATGCCCTTAAAAAGGGTGCCGCTTTGAGCTGGTCTTCATAGTCATGGCTGAGCAC	333
83	K W G S E E R K R S M P F K K G A A F E L V F I V M A E H	111
334	TACAAGTGGTGGTAAATGGAAATCCCTTCTATGAGTACGGGCACCGGCTTCCCTACAGATGGTCACCCACCTGCAAGTGGATGGGGA	422
112	Y K V V V N G N P F Y E Y G H R L P L Q M V T H L Q V D G D	141
Término da construção do CRD-I ↓		
Início da construção do CRD-II ↓		
423	TCTGCAACTTCAATCAACTTCATCGGAGGCCAGCCCCCGGGCCCGAGGACCCCGATGATGCCACCTTACCCTGGTCCCGGACA	512
142	L Q L Q S I N F I G G Q P L R P Q G P P M M P P Y P G P G H	171
513	TTGCCATCAACAGCTGAACAGCCTGCCAACCATGGAAGGACCCCAACCTTCAACCCGCCGGTGCCATATTTGGGAGGCTGCAAGGAGG	602
172	C H Q Q L N S L P T M E G P P T F N P P V P Y F G R L Q G G	201
603	GCTCACAGCTCGAAGAACCATCATCATCAAGGGCTATGTGCCTCCACAGGCAAGAGCTTTGCTATCAACTTCAAGTGGGCTCCTCAGG	692
202	L T A R R T I I I K G Y V P P T G K S F A I N F K V G S S G	231
693	GGACATAGCTCTGCACATTAACCCGAGGATGGCAACGGTACCCTGGTCCGGAACAGCCTTCTGAATGGCTCGTGGGGATCCGAGGAGAA	781
232	D I A L H I N P R M G N G T V V R N S L L N G S W G S E E K	261
782	AGAAGACTACCCACAACCCATTTGGTCCCGGACAGTTCTTTGATCTGTCCATTGCTGTGGCTTGGATCGCTTCAAGGTTTACGCCAATGG	872
262	K I T H N P F G P G Q F F D L S I R C G L D R F K V Y A N G	291
873	CCAGCACCTCTTTGACTTTGCACATCCGCTGTCTGGCCTTCCAGAGGGTCGACACATTGGAATCCAGGGTGATGTCACCTTGTCTATGTC	964
292	Q H L F D F A H R L S A F Q R V D T L E I Q G D V T L S Y V	321
↓ Término da construção do CRD-II		
965	CAGATCTAATCTAT TCCTGGCATA	988
322	Q I *	323

Figura 10: Sequência de nucleotídeos e aminoácidos da Hgal-4 (HUFLEJT *et al.*, 1997; RECHRECHE *et al.*, 1997). A figura mostra os sítos específicos de início e término utilizados no processo de clonagem dos CRD-I e CRD-II. A sequência do peptídeo de ligação está apresentada em vermelho, **GenBank código NM-006149.3**.

Através da observação entre interações existentes na estrutura, não foi possível identificar nenhuma interação direta entre os resíduos pertencentes ao peptídeo de ligação e o CRD-II. A única interação observada entre a extensão N-terminal, correspondente ao peptídeo de ligação e o domínio globular se faz através de uma possível ligação de hidrogênio fraca (3.7 Å) entre as cadeias laterais da Thr-181 e Arg-197 (**Figura 11**). Embora, a estrutura sugere que seria possível expressar um domínio globular estável a partir da região por volta do resíduo Pro-193 (**Figura 11**), as várias conformações adotadas pela proteína nas diferentes estruturas obtidas por NMR mostram que a região N-terminal é bastante flexível, o que então não descarta a possibilidade de que haja alguma importante interação, mesmo que dinâmica, entre o peptídeo de ligação e o CRD-II. Sendo assim, com base em nossas análises, optamos por clonar o domínio relativo ao CRD-II fusionado a um peptídeo cuja sequência corresponde a um pequeno fragmento do peptídeo de ligação. Optamos assim por clonar o domínio CRD-II correspondente aos resíduos Gly-170 e Ile-323 (**Figura 10**).

Figura 11: Estrutura tridimensional do domínio CRD-II de Hgal-4. À esquerda os resíduos Thr-181 (localizado no peptídeo de ligação) e Arg-197 (localizado no CRD-II) se encontram a distância de aproximadamente 3.7Å. À direita, o resíduo Pro-193 indica a região onde se inicia a região globular, **PDB código1X50**.

Os oligonucleotídeos foram desenhados manualmente para serem utilizados no processo de clonagem dos genes no vetor de clonagem pET29a (Novagen) (**Figura 12**). A escolha do vetor pET29a (Novagen) para o processo de clonagem dos CRD-I e CRD-II da Hgal-4 inicialmente baseou-se em nosso interesse de expressar os domínios de forma íntegra, sem a necessidade de fusioná-los a qualquer peptídeo. A vantagem em se produzir uma proteína intacta, sem um peptídeo de fusão é que ela seria utilizada sem a preocupação de que o peptídeo de fusão pudesse interferir em sua estrutura e atividade. Baseados na evidência de

que os CRD-I e CRD-II possuem afinidade por lactose ($K_d \sim 0.5-1,0\text{mM}$), estes seriam isolados primeiramente utilizando-se o processo de cromatografia por afinidade em resina de α -Lactose-Agarose (Sigma-Aldrich). Os oligonucleotídeos foram planejados de forma a incorporar os sítios de restrição para as enzimas *Nde* I e *Not* I (**Figura 12**). A escolha dos sítios de restrição para as enzimas *Nde* I e *Not* I foi feita de forma a atender os seguintes requisitos: Ausência dos sítios de restrição no gene de interesse e escolha de enzimas de mais baixo custo e de fácil aquisição. Os oligonucleotídeos foram adquiridos da Integrated DNA Technologies, Inc. (IDT). Os oligonucleotídeos utilizados para amplificar os genes de interesse consistem de 3 segmentos: uma âncora (GACGAC) para facilitar a digestão com as enzimas de restrição, os sítios de restrição e um fragmento de sequência do gene a ser amplificado (**Figura 12**).

Figura 12: Sequência dos oligonucleotídeos sintetizados para amplificação dos domínios de reconhecimento de carboidratos CRD-I e CRD-II da galectina-4 humana.

3.1.2 Amplificação dos CRDs da galectina-4 humana

Para os experimentos de amplificação dos peptídeos codificantes dos domínios CRD-I e CRD-II, utilizou-se uma reação de PCR (Reação em Cadeia da Polimerase). Como molde para as reações de PCR, foi utilizada a construção onde a sequência de nucleotídeos que codifica a galectina-4 humana completa se encontra clonada no vetor de expressão pQE (Qiagen). Esta construção foi gentilmente cedida pelo Prof. Dr. Marcelo D. Baruffi do Departamento de Análises Clínicas, Toxicológicas e Bromatológicas da Faculdade de Ciências Farmacêuticas de Ribeirão Preto.

Para o processo de amplificação, testes preliminares foram realizados para obtenção da temperatura ótima de anelamento dos oligonucleotídeos sintetizados. As temperaturas iniciais testadas foram de 50°C, 55°C e 60°C.

As reações seguiram os seguintes protocolos:

Tabela 1: Protocolo para realização de reação de PCR do domínio CRD-I.

	Volume (μL)	[] Final
EnzimaTaq DNA Polymerase High Fidelity (Invitrogen)	0,20	1 unidade
Tampão da Taq DNA Polymerase High Fidelity (Invitrogen)	5,00	1,0 X
MgSO ₄ (50mM)	2,00	2,0 mM
Oligonucleotídeos sintetizados (primer I) (IDT) (4 μM)	5,00	0,4 μM
Oligonucleotídeos sintetizados (primer II) (IDT) (4μM)	5,00	0,4 μM
dNTP mix (Eppendorf)	2,50	0,2 μM
DNA template	1,00	---
Água autoclavada	29,30	---
Volume final da reação	50,00	---

Tabela 2: Protocolo para realização de reação de PCR do domínio CRD-II.

	Volume (μL)	[] Final
EnzimaTaq DNA Polymerase High Fidelity (Invitrogen)	0,20	1 unidade
Tampão da Taq DNA Polymerase High Fidelity (Invitrogen)	5,00	1,0 X
MgSO ₄ (50mM)	2,00	2,0 mM
Oligonucleotídeos sintetizados (primer III) (IDT) (4μM)	5,00	0,4 μM
Oligonucleotídeos sintetizados (primer IV) (IDT) (4μM)	5,00	0,4 μM
dNTP mix (Eppendorf)	2,50	0,2 μM
DNA template	1,00	---
Água autoclavada	29,30	---
Volume final da reação	50,00	---

As reações de PCR seguiram conforme protocolo específico para a enzima EnzimaTaq DNA Polymerase High Fidelity (Invitrogen), sendo de 25 ciclos, com uma desnaturação inicial à temperatura de 94°C por um tempo de 2 minutos. Cada ciclo consistiu de: 30 segundos a 94 °C para desnaturação, 30 segundos a 60 °C (CRD-I) e 50 °C (CRD-II) para o anelamento dos oligonucleotídeos e 30 segundos a 68 °C para o processo de extensão. Após o término dos 25 ciclos, a reação foi mantida a 68 °C para o processo de extensão final. Para a reação de PCR, foi utilizado um termociclador Bioqclin Termociclador Mastercycler gradient (Eppendorf) gentilmente cedido pelo Prof. Dr. Sérgio Akira Uyemura.

Os fragmentos obtidos no processo de amplificação foram submetidos a uma eletroforese em gel de agarose a 0,8 % (w/v) contendo brometo de etídio. O brometo de etídeo é um composto intercalador, o qual se insere na dupla hélice da molécula de DNA e fluoresce na presença de luz ultra-violeta, permitindo portanto uma visualização dos fragmentos amplificados. Os fragmentos obtidos foram então purificados do gel utilizando-se o Kit QIAquick® Gel Extraction Kit Protocol (Qiagen) seguindo-se para tal procedimento as especificações do fabricante.

3.1.3 Digestão dos CRDs da galectina-4 humana e do vetor pET29a (Novagen)

Amostras do vetor de clonagem pET-29a (Novagen), assim como, dos fragmentos de DNA amplificados pela reação de PCR, foram submetidos à reação de digestão com as enzimas de restrição *Not-I* (Invitrogen) e *Nde-I* (Invitrogen). As reações procederam-se separadamente, pois as enzimas possuem máxima eficiência em diferentes tampões. Primeiramente realizou-se as reações de digestão dos CRDs assim como do vetor de clonagem pET29a (Novagen) utilizando-se a enzima de restrição *Nde-I* (Invitrogen). As reações foram conduzidas a 37°C por um período de 4 horas.

As reações seguiram os seguintes protocolos

Tabela 3: Protocolo para realização da reação de digestão do domínio CRD-I.

	Volume (μL)	[] Final
Fragmentos amplificados (CRD-I)	10,0	0,1μg/μL
Enzima de restrição <i>Nde I</i> (Invitrogen)	1,0	5 unidades
Tampão React® 6 (10x) (Invitrogen)	3,0	1,0X
H ₂ O (autoclavada)	16,0	---
Volume final da reação	30,0	---

Tabela 4: Protocolo para realização da reação de digestão do domínio CRD-II.

	Volume (μL)	[] Final
Fragmentos amplificados (CRDII)	10,0	0,1μg/μL
Enzima de restrição <i>Nde I</i> (Invitrogen)	1,0	5 unidades
Tampão React® 6 (10x) (Invitrogen)	3,0	1,0X
H ₂ O (autoclavada)	16,0	---
Volume final da reação	30,0	---

Tabela 5: Protocolo para realização da reação de digestão do vetor pET29a.

	Volume (μL)	[] Final
Vetor pET29a (Novagen)	7,5	0,1μg/μL
Enzima de restrição <i>Nde I</i> (Invitrogen)	1,0	5 unidades
Tampão React® 6 (10x) (Invitrogen)	3,0	1,0X
H ₂ O (autoclavada)	18,5	---
Volume final da reação	30,0	---

Após o processo de digestão dos fragmentos (CRD-I e CRD-II), assim como, das amostras do vetor pET-29a (Novagen), utilizando-se a enzima de restrição *NdeI* (Invitrogen), as amostras foram submetidas a um processo de purificação para a remoção da enzima, assim como do tampão utilizado na reação. Para tal, utilizou-se o kit QIAquick® PCR Purification Kit (Qiagen) segundo as especificações de uso descritas pelo fabricante. Subsequentemente, as amostras purificadas foram então utilizadas no processo de digestão com a enzima de restrição *Not I* (Invitrogen), e novamente submetidas a limpeza com o kit QIAquick® PCR Purification Kit (Qiagen) para serem utilizadas nas reações de ligação. As reações foram conduzidas a 37°C por um período de 4 horas.

As reações seguiram os seguintes protocolos:

Tabela 6: Protocolo para realização da reação de digestão do domínio CRD-I.

	Volume (µL)	[] Final
Fragmentos amplificados (CRD-I)	10,0	0,1µg/µL
Enzima de restrição <i>Not I</i> (Invitrogen)	1,0	5 unidades
Tampão React® 3 (10x) (Invitrogen)	3,0	1,0X
H ₂ O (autoclavada)	16,0	---
Volume final da reação	30,0	---

Tabela 7: Protocolo para realização da reação de digestão do domínio CRD-II.

	Volume (µL)	[] Final
Fragmentos amplificados (CRD-II)	10,0	0,1µg/µL
Enzima de restrição <i>Not I</i> (Invitrogen)	1,0	5 unidades
Tampão React® 3 (10x) (Invitrogen)	3,0	1,0X
H ₂ O (autoclavada)	16,0	---
Volume final da reação	30,0	---

Tabela 8: Protocolo para realização da reação de digestão do vetor pET29a.

	Volume (µL)	[] Final
Vetor pET29a (Novagen)	7,5	0,1µg/µL
Enzima de restrição <i>Not I</i> (Invitrogen)	1,0	5 unidades
Tampão React® 3 (10x) (Invitrogen)	3,0	1,0X
H ₂ O (autoclavada)	18,5	---
Volume final da reação	30,0	---

3.1.4 Ligação dos CRDs da galectina-4 humana no vetor pET29a (Novagen)

As reações de ligação dos fragmentos referentes aos domínios CRD-I e CRD-II da Gal-4 humana no vetor pET29a (Novagen) previamente digeridos foram realizadas utilizando-se a enzima T4 DNA Ligase (Invitrogen). A DNA ligase catalisa a formação de duas ligações

fosfodiéster entre a hidroxila da extremidade 3' de uma fita de DNA com o fosfato presente na extremidade 5' de outra fita.

No planejamento dos experimentos de ligação, inicialmente fez-se uma estimativa da quantidade de inserto e do vetor após os ensaios de digestão. Para isto, as amostras foram submetidas à eletroforese em um gel de agarose 0,8% (v/w) e as intensidades das bandas foram comparadas às intensidades obtidas em uma amostra padrão de DNA (Biolabs). Para maximizar as chances de sucesso nos experimentos de ligação, preparou-se diferentes reações, onde diferentes proporções entre o produto de PCR (CRD-I e CRD-II) e vetor pET29a (Novagen) digeridos foram testadas. Para controle do experimento, o sucesso da ligação foi monitorado através de uma reação de ligação na ausência do inserto (Tabela 4, reação 4). As reações de ligação foram realizadas a temperatura de 16°C por um período de 16 horas.

As reações seguiram os seguintes protocolos:

Tabela 9: Protocolo para realização da reação de ligação do CRD-I no vetor pET29a.

	1	2	3	4
Vetor pET29a (Novagen)	2 µL	3µL	4µL	4 µL
Inserto CRD-I	6 µL	9 µL	12 µL	--
Enzima T4 DNA ligase (Invitrogen)	1 µL	1 µL	1 µL	1 µL
Tampão T4 DNA Ligase (Invitrogen) 5X	4 µL	4 µL	4 µL	4 µL
H ₂ O (autoclavada)	7 µL	3 µL	---	11 µL
Volume final da reação	20 µL	20 µL	21 µL	20 µL

Tabela 10: Protocolo para realização da reação de ligação do CRD-II no vetor pET29a.

	1	2	3	4
Vetor pET29a (Novagen)	2 µL	3µL	4µL	4 µL
Inserto CRD-II	6 µL	9 µL	12 µL	--
Enzima T4 DNA ligase (Invitrogen)	1 µL	1 µL	1 µL	1 µL
Tampão T4 DNA Ligase (Invitrogen) 5X	4 µL	4 µL	4 µL	4 µL
H ₂ O (autoclavada)	7 µL	3 µL	---	11 µL
Volume final da reação	20 µL	20 µL	21 µL	20 µL

Como resultado do processo de ligação dos fragmentos amplificados por PCR codificando os domínios de reconhecimento (CRDs) da proteína galectina-4 humana no vetor de expressão pET29a (Novagen), obteve-se as construções designadas por **Hgal-4CRD-I(pET29a)** e **Hgal-4CRD-II(pET29a)**.

3.1.5 Transformação das construções *Hgal-4CRD-I(pET29a)* e *Hgal-4CRD-II(pET29a)* em bactérias *Escherichia coli DH5α*

O processo de transformação constitui um evento de alta importância na técnica de manipulação gênica. O processo envolve a introdução de DNA exógeno nas células bacterianas e este DNA passa então a fazer parte do material genético da bactéria, podendo ser herdado pelas novas células todas as vezes que a bactéria se multiplicar. No entanto, para permitir a entrada de DNA plasmidial nas células bacterianas, as bactérias devem apresentar condições fisiológicas especiais. Este estado fisiológico especial é chamado de competência. Uma bactéria competente é uma bactéria que está apta a receber DNA exógeno.

A competência pode ser produzida através de diferentes técnicas. Uma destas técnicas envolve tratamento com uma solução de cloreto de cálcio e mudanças bruscas de temperatura (MANDEL: HIGA, 1970). Os íons de metal e as mudanças bruscas de temperatura alteram a permeabilidade da parede celular fazendo com que estas células permitam a entrada do DNA exógeno através da membrana celular. As bactérias competentes são bastante frágeis e devem ser manuseadas com cuidado. No presente trabalho utilizou-se a técnica que emprega o tratamento das células bacterianas com uma solução de cloreto de cálcio (*ver apêndice II*).

Após a obtenção das bactérias competentes, as construções **Hgal-4CRD-I(pET29a)** e **Hgal-4CRD-II(pET29a)** foram transformadas na bactéria *Escherichia coli*, utilizando a cepa DH5α (Novagen), (*ver apêndice III*). Para a verificação da presença do gene de interesse, algumas colônias obtidas foram selecionadas e utilizadas para inocular 5 mL de meio LB (Luria-Bertani) + (50 µg/mL) de antibiótico kanamicina. Após o período de incubação (*overnight* a 37°C/250 rpm), a análise da presença dos insertos em cada colônia foi realizada através de uma reação de um PCR de colônia. O produto de PCR foi submetido a uma eletroforese em gel de agarose 0,8% (w/v) para visualização da presença do inserto dentre as colônias selecionadas.

As reações de PCR de colônia seguiram os seguintes protocolos:

Tabela11 : Protocolo para realização da reação de PCR de colônia.

	Volume (μL)	[] Final
EnzimaTaq DNA Polymerase Recombinante (Promega)	0,25	1 unidade
Tampão da Taq DNA Polymerase Recombinante (Promega)	5,00	1,0 X
MgCl ₂ (25mM)	2,00	2,0mM
Oligonucleotídeos sintetizados (IDT) (4,0 μM)	5,00	0,4 μM
dNTP mix (Eppendorf)	2,50	0,2 μM
Amostras de cada colônia (pré-inóculo)	2,00	---
Água autoclavada	23,25	---
Volume final da reação	50,00	---

As colônias testadas e que apresentaram resultado positivo foram utilizadas para preparar estoque em glicerol e armazenadas em nitrogênio líquido. Os clones positivos foram purificados com o Kit QIAprep® Miprep segundo o protocolo sugerido no manual do mesmo. Para os ensaios de expressão, amostras das construções Hgal-4CRD-I(pET29a) e Hgal-4CRD-II(pET29a) foram transformados em bactérias *Escherichia coli* utilizando-se cepas BL-21(DE3) pLysS, resistentes aos antibióticos kanamicina (50 μg/mL) e cloranfenicol (34 μg/mL) e foram mantidas a -80°C sob a forma de glicerol estoque.

3.1.6 Expressão dos domínios CRD-I e CRD-II da galectina-4 humana

3.1.6.1 Expressão do domínio CRD-II

Os primeiros ensaios de expressão foram para o domínio CRD-II. Para a realização dos experimentos utilizou-se o clone designado como *Hgal-4CRD-II(pET29a)(II-4)*. Todos os experimentos de expressão foram realizados em ambiente estéril. Para a preparação do pré-inóculo, amostras do glicerol estoque contendo a construção Hgal-4CRD-II(pET29a)(II-4) foram utilizadas para inocular 10,0 mL de meio LB (Luria-Bertani), na presença dos antibióticos apropriados, kanamicina (50 μg/mL) e cloranfenicol (34 μg/mL) e mantido em uma incubadora sob agitação por aproximadamente 16 horas a 250 rpm e 37°C. O material foi então utilizado para inocular 1L de meio de cultura LB (Luria-Bertani), na proporção (1:100) e mantido sob agitação a 250 rpm e 37°C. O crescimento das células foi monitorado medindo-se a DO₆₀₀ (densidade ótica do meio de cultura em um comprimento de onda de 600nm) até atingir um valor de absorbância entre 0,45 e 0,50. A indução da expressão da proteína de interesse foi feita através da adição de IPTG (isopropyl β-D-1-tiogalactopiranosídeo).

Como controle da expressão, amostras de 2,0 mL da cultura de células foram retiradas em diferentes tempos de indução: 0h (controle), 2, 3, 4 e 5, 18 e 24 horas. Foram realizados ensaios de expressão na presença de diferentes concentrações de IPTG (0,5 mM, 1,0 mM e 1,5 mM) e nas temperaturas de 37 e 18°C. As amostras das células foram isoladas do meio de cultura por centrifugação e armazenadas no freezer a -20°C.

3.1.6.2 Análise do padrão de expressão do domínio CRD-II.

As células congeladas foram ressuspensas em tampão de lise consistindo de tampão PBS 1X (NaCl 137mM; KCl 2,6mM; Na₂HPO₄ 8mM; KH₂PO₄ 15mM, pH 7,4), 14mM de β-mercaptoetanol e 1,0 mM de inibidor de protease PMSF (Roche). Para ressuspender as células foram utilizados aproximadamente 3,0 mL do tampão de lise para cada 1,0 grama de células. O material ressuspendido foi sonificado por 8 vezes em intervalos de tempo de 20 segundos. Após a sonicação, a amostra foi centrifugada por 45 minutos a 10.000g e 4°C para separação do material solúvel (*sobrenadante*) e insolúvel (*pellet*).

A análise do padrão de expressão foi feita através da eletroforese de gel de poliacrilamida 12% na presença do detergente dodecil sulfato de sódio (SDS-PAGE). Na eletroforese em presença de SDS, as proteínas são separadas quase exclusivamente com base em sua massa molecular. O detergente SDS ligado à proteína contribui com uma grande carga líquida negativa e quando um campo elétrico é aplicado, as proteínas movem-se para o interior do gel; proteínas menores movem-se mais rapidamente no gel do que proteínas maiores, portanto, serão encontradas na extremidade inferior, onde se localiza o polo positivo (NELSON; COX, 2002).

3.1.6.3 Expressão do domínio CRD-I

Para a realização dos experimentos de expressão utilizou-se o clone designado como *Hgal-4CRD-I(pET29a)(I-1)*. Todos os experimentos foram realizados em ambiente estéril. Para a preparação do pré-inóculo, amostras do glicerol estoque contendo a construção *Hgal-4CRD-I(pET29a)(I-1)* foram utilizadas para inocular 10,0 mL de meio LB (Luria-Bertani), na presença dos antibióticos apropriados, kanamicina (50 µg/mL) e

cloranfenicol (34 $\mu\text{g/mL}$) e mantido em uma incubadora sob agitação por aproximadamente 16 horas a 250 rpm e 37°C. O material foi então utilizado para inocular 1L de meio de cultura LB (Luria-Bertani), na proporção (1:100) e mantido sob agitação a 250 rpm e 37°C. O crescimento das células foi monitorado medindo-se a DO_{600} (densidade ótica do meio de cultura em um comprimento de onda de 600nm) até atingir um valor de absorbância entre 0,45 e 0,50. A indução da expressão da proteína de interesse é feita através da adição de IPTG (isopropyl β -D-1-tiogalactopiranosídeo).

Como controle da expressão, amostras de 2,0 mL da cultura de células foram retiradas em tempos de indução: 0h (controle), 5 e 24 horas. Foram realizados ensaios de expressão na presença de IPTG (1,5 mM) e nas temperaturas de 37°C. As amostras das células foram isoladas do meio de cultura por centrifugação e armazenadas no freezer a -20°C.

3.1.6.4 *Análise do padrão de expressão do domínio CRD-I*

As células congeladas foram ressuspendidas em tampão de lise consistindo de tampão PBS 1X (NaCl 137mM; KCl 2,6mM; Na_2HPO_4 8mM; KH_2PO_4 15mM, pH 7,4), 14mM de β -mercaptoetanol e 1,0 mM de inibidor de protease PMSF (Roche). Para ressuspender as células, foram utilizados aproximadamente 3,0 mL do tampão de lise para cada 1,0 grama de células. O material ressuspendido foi sonicado por oito vezes em intervalos de tempo de 20 segundos. Após a sonicação, a amostra foi centrifugada por 45 minutos a 10.000g e 4°C para separação do material solúvel (*sobrenadante*) e insolúvel (*pellet*).

A análise do padrão de expressão foi feita através da eletroforese de gel de poliacrilamida 12% na presença do detergente dodecil sulfato de sódio (SDS-PAGE).

3.1.7 *Purificação dos domínios CRD-I e CRD-II da galectina-4 humana*

Todos os passos da purificação dos domínios CRD-I e CRD-II foram realizados a temperatura de 4°C, para minimizar efeitos de desnaturação da proteína. Para os ensaios de purificação adotou-se o protocolo utilizado na obtenção da galectina-4 humana completa, onde foi empregada a técnica de **purificação por afinidade**, utilizando-se uma coluna de α -Lactose-Agarose (Sigma-Aldrich).

Para obtenção das proteínas de interesse, após o processo de expressão dos domínios CRD-I e CRD-II, as células congeladas foram ressuspensas em tampão de lise consistindo de tampão PBS 1X (NaCl 137mM; KCl 2,6mM; Na₂HPO₄ 8mM; KH₂PO₄ 15mM, pH 7,4), 14mM de β-mercaptoetanol e 1,0 mM de inibidor de protease PMSF (Roche). Para ressuspender as células foram utilizados aproximadamente 3,0 mL do tampão de lise para cada 1,0 grama de células. O material ressuspensionado foi sonificado por oito vezes em intervalos de tempo de 20 segundos. Após a sonicação, a amostra foi centrifugada por 45 minutos a 10.000g e 4°C para separação do material solúvel (*sobrenadante*) e insolúvel (*pellet*). Após este processo, o material solúvel foi aplicado em uma coluna de α-Lactose-Agarose (Sigma-Aldrich).

3.1.7.1 Cromatografia por Afinidade

A cromatografia por afinidade é um método empregado na purificação de proteínas o qual separa as proteínas de interesse por suas especificidades de ligação. No presente trabalho, utilizou-se a cromatografia por afinidade em uma coluna de α-Lactose-Agarose (Sigma-Aldrich), uma vez que os domínios de reconhecimento de carboidratos (CRD-I e CRD-II) da Hgal-4 apresentam, ainda que baixa, afinidade por lactose.

A cromatografia por afinidade foi realizada em um sistema de cromatografia líquida ÄKTA *prime*TM, utilizando-se uma coluna contendo um volume de 8,0 mL da resina α-Lactose-Agarose (Sigma-Aldrich). Em um primeiro passo, a coluna de α-Lactose-Agarose foi equilibrada com uma solução de PBS 1X + 14 mM de β-Mercaptoetanol designado como tampão de lavagem. O extrato bruto foi então aplicado e a coluna lavada com 4 vezes o volume de coluna com tampão de lavagem. O material ligado à coluna foi então eluído utilizando-se uma solução de PBS 1X + 14 mM de β-Mercaptoetanol contendo 100mM de lactose, designada por tampão de eluição. A cromatografia foi realizada com um fluxo na coluna de 1,0 mL/min. As frações eluídas da purificação por cromatografia por afinidade dos domínios CRD-I e CRD-II e da galectina-4 inteira foram analisadas por eletroforese em gel de poliacrilamida (12% SDS-PAGE).

3.2.1 Estratégia

Primeiramente as sequências de nucleotídeos que codificam os domínios de reconhecimento de carboidratos (CRD-I e CRD-II) da galectina-4 humana foram inicialmente subclonadas no vetor de subclonagem pTZ57R/T (Fermentas) (**Figura 13**). A utilização do vetor pTZ57R/T(Fermentas) favorece a eficiência no processo de digestão dos fragmentos amplificados com as enzimas de restrição, facilitando os futuros experimentos de clonagem dos domínios CRDs em um sistema de expressão.

Após o processo de subclonagem em sistema pTZ57R/T (Fermentas), os domínios CRD-I e CRD-II foram então clonados em vetor de expressão pET28a (Novagen) (**Figura 14**). Este vetor possui todas as características dos vetores de expressão do sistema pET, como também, possui uma região codificadora de uma cauda de 6 histidinas na região 5' anterior ao sítio de restrição múltiplo. O processo de clonagem dos CRD-I e CRD-II da Hgal-4 em vetor pET28a (Novagen) baseou-se em nosso interesse de expressar os domínios em fusão com a

cauda de histidina. Uma vez que os domínios CRD-I e CRD-II se encontram fusionados a uma cauda de histidina, estes serão isolados utilizando-se o processo de cromatografia por afinidade em coluna de Ni-NTA (Qiagen).

Figura 14. Mapa do vetor pET28a (Novagen). Este plasmídeo contém um marcador que confere resistência a kanamicina, o gene *lacI*, o promotor de transcrição T7, a região do operador *lac* e um sítio múltiplo de clonagem.

3.2.2 Amplificação dos domínios CRD-I e CRD-II da galectina-4 humana

Para a subclonagem dos domínios CRD-I e CRD-II em vetor pTZ5R/T(Fermentas), novos ensaios para amplificação dos peptídeos codificantes dos CRDs foram realizados. Para tal utilizou-se uma reação de PCR (Reação em Cadeia da Polimerase) e como molde para as reações foi utilizada a construção onde a sequência de nucleotídeos que codifica a galectina-4 humana completa se encontra clonada no vetor de expressão pQE (Qiagen).

Para o processo de amplificação, testes preliminares foram realizados para obtenção da temperatura ótima de anelamento dos oligonucleotídeos sintetizados. As temperaturas iniciais testadas foram de 50°C, 55°C e 60°C.

As reações de PCR seguiram os seguintes protocolos:

Tabela 12: Protocolo para realização de reação de PCR do domínio CRD-I.

	Volume (μL)	[] Final
EnzimaTaq DNA Polymerase True Start (Fermentas)	0,2	1 unidade
Tampão da Taq DNA Polymeras True Start	5,0	1,0 X
MgCl ₂ (25 mM)	5,0	2,5 mM
Oligonucleotídeos sintetizados (primer I) (IDT) (4μM)	1,0	0,8 μM
Oligonucleotídeos sintetizados (primer II) (IDT) (4μM)	1,0	0,8 μM
dNTP mix (Eppendorf) (10mM)	1,0	0,2 mM
DNA template	2,0	4,0 ng/μL
Água autoclavada	34,8	---
Volume final da reação	50,0	---

Tabela 13: Protocolo para realização de reação de PCR do domínio CRD-II.

	Volume (μL)	[] Final
EnzimaTaq DNA Polymerase True Start (Fermentas)	0,2	1 unidade
Tampão da Taq DNA Polymeras True Start (Fermentas)	5,0	1,0 X
MgCl ₂ (25 mM)	5,0	2,5 mM
Oligonucleotídeos sintetizados (primer III) (IDT) (4μM)	1,0	0,8 μM
Oligonucleotídeos sintetizados (primer IV) (IDT) (4μM)	1,0	0,8 μM
dNTP mix (Eppendorf) (10mM)	1,0	0,2 mM
DNA template	2,0	4,0 ng/ μL
Água autoclavada	34,8	---
Volume final da reação	50,0	---

As reações foram mantidas segundo especificações descritas pelo fabricante da enzima Taq DNA polymerase True Start (Fermentas). A reação de PCR seguiu-se por 30 ciclos, com uma desnaturação inicial à temperatura de 95°C por um tempo de 2 minutos. Cada ciclo consistiu de: 60 segundos a 95 °C para desnaturação, para o anelamento dos oligonucleotídeos foram testadas as temperaturas de 50°C, 55°C e 60°C por 60 segundos, e para o processo de extensão utilizou-se a temperatura de 72 °C por 35 segundos. Após o término dos 30 ciclos, a reação foi mantida por 15 minutos a 72 °C para o processo de extensão final. Para a realização da reação de PCR utilizou-se um termociclador Bioqclin Termociclador Mastercycler gradient (Eppendorf) gentilmente cedido pelo Prof. Dr. Sérgio Akira Uyemura.

Os fragmentos obtidos no processo de amplificação foram submetidos a uma eletroforese em gel de agarose a 0,8 % (w/v) contendo brometo de etídio para possível visualização dos segmentos amplificados. Os fragmentos obtidos e visualizados foram então purificados do gel de agarose utilizando-se o Kit QIAquick® Gel Extraction Kit Protocol (Qiagen) segundo as especificações do fabricante.

3.2.3 Processo de Adenilação dos CRDs da galectina-4 humana.

Para a utilização do sistema pTZ57R/T (Fermentas), os fragmentos amplificados por uma reação de PCR e purificados a partir do gel de agarose foram primeiramente submetidos a um processo de adenilação. Na reação de adenilação, os fragmentos amplificados por PCR sofrem a adição de nucleotídeos de adenina, através de uma reação de PCR (Reação em Cadeia da Polimerase) mediada pela enzima Taq DNA polymerase True Start (Fermentas). Os fragmentos adenilados foram ligados o vetor pTZ57R/T (Fermentas) através do pareamento de bases já que o vetor apresenta um sítio de restrição com extremidades contendo resíduos de Timina.

A reação foi mantida segundo especificações descritas pelo fabricante da enzima Taq DNA polymerase True Start (Fermentas). A reação de PCR seguiu-se inicialmente por 1 minuto a 94°C, seguidamente por 30 minutos a 72°C para a reação de adição dos nucleotídeos de adenina. Para a reação de PCR, foi utilizado um termociclador Bioqclin Termociclador Mastercycler gradient (Eppendorf) gentilmente cedido pelo Prof. Dr. Sérgio Akira Uyemura.

As reações de adenilação seguiram os seguintes protocolos

Tabela 14: Protocolo para realização de reação de adenilação do domínio CRD-I.

	Volume (µL)	[] Final
EnzimaTaq DNA Polymerase Recombinante (Fermentas)	1,0	5 unidades
Tampão da Taq DNA Polymeras True Start 10 X (Fermentas)	1,0	1,0 X
MgCl ₂ (50 mM)	0,4	2,0 mM
Fragmentos amplificados CRD-I	6,0	60,0ng/ µL
dATP mix (Eppendorf) (3mM)	0,7	0,2 mM
Água autoclavada	0,9	---
Volume final da reação		

Tabela 15: Protocolo para realização de reação de adenilação do domínio CRD-II.

	Volume (µL)	[] Final
EnzimaTaq DNA Polymerase Recombinante (Fermentas)	1,0	5 unidades
Tampão da Taq DNA Polymeras True Start 10 X (Fermentas)	1,0	1,0 X
MgCl ₂ (50 mM)	0,4	2,0 mM
Fragmentos amplificados CRD-I	6,0	60,0ng/ µL
dATP mix (Eppendorf) (3mM)	0,7	0,2 mM
Água autoclavada	0,9	---
Volume final da reação	10,0	---

3.2.4 Ligação dos CRDs da galectina-4 no vetor de subclonagem pTZ57R/T(Fermentas)

As reações de ligação dos fragmentos previamente amplificados e adenilados no vetor pTZ57R/T foram realizadas utilizando-se a enzima T4 DNA Ligase (Invitrogen). No planejamento dos experimentos de ligação, inicialmente fez-se uma estimativa da quantidade de inserto amplificado e adenilado. Para isto, as amostras foram submetidas a uma eletroforese em um gel de agarose 0,8% (w/v) e as intensidades das bandas foram comparadas às intensidades obtidas em uma amostra padrão de DNA (Biolabs). As reações de ligação foram realizadas a temperatura de 16°C por um período de 16 horas conforme especificações do fabricante.

As reações de ligação seguiram os seguintes protocolos:

Tabela 16: Protocolo para realização da reação de ligação do CRD-I ao vetor pTZ57R/T.

	Volume	[] Final
Vetor pTZ57R/T (Fermentas)	3 µL	165 ng/µL
Inserto adenilado CRD-I	10 µL	250 ng/µL
Enzima T4 DNA Ligase (Invitrogen)	1 µL	5 unidades
Tampão T4 DNA Ligase (Invitrogen) (5X)	6 µL	1X
H ₂ O (autoclavada)	10 µL	-----
Volume final da reação	30 µL	-----

Tabela 17: Protocolo para realização da reação de ligação do CRD-II ao vetor pTZ57R/T.

	Volume	[] Final
Vetor pTZ57R/T (Fermentas)	3 µL	165ng/µL
Inserto adenilado CRD-II	21 µL	176,4 ng/µL
Enzima T4 DNA Ligase (Invitrogen)	1 µL	5 unidades
Tampão T4 DNA Ligase (Invitrogen) (5X)	6 µL	1X
H ₂ O (autoclavada)	--	-----
Volume final da reação	31 µL	-----

Como resultado do processo de ligação dos fragmentos codificando os domínios de reconhecimento (CRDs) da proteína galectina-4 humana ao vetor de subclonagem pTZ57R/T (Fermentas), obteve-se as construções designadas por **Hgal-4CRD-I(pTZ57R/T)** e **Hgal-4CRD-II(pTZ57R/T)**.

3.2.5 Transformação das construções *Hgal-4CRD-I(pTZ57R/T)* e *Hgal-4CRD-II(pTZ57R/T)* em bactérias *Escherichia coli DH5α*

Para o processo de transformação das construções *Hgal-4CRD-I(pTZ57R/T)* e *Hgal-4CRD-II(pTZ57R/T)* em bactérias *Escherichia coli*, utilizou-se primeiramente a técnica de tratamento com uma solução de cloreto de cálcio e mudanças bruscas de temperatura para

indução da competência nas cepas da bactéria *Escherichia coli* (Ver apêndice II). Após a obtenção das bactérias competentes, as construções **Hgal-4CRD-I(pTZ57R/T)** e **Hgal-4CRD-I(pTZ57R/T)** foram transformadas na bactéria *Escherichia coli*, utilizando-se a cepa DH5 α (Novagen) (Ver apêndice III).

Para a verificação da presença do gene de interesse, as colônias obtidas e selecionadas foram utilizadas para inocular 5 mL de meio LB (Luria-Bertani) + (50 μ g/mL) de Ampicilina. Após o período de incubação (*overnight* a 37°C/250 rpm), a análise da presença dos insertos em cada colônia foi realizada através de uma reação de um PCR de colônia. O produto de PCR foi submetido a uma eletroforese em gel de agarose 0,8% (w/v) para visualização da presença do inserto dentre as colônias selecionadas.

As reações de PCR de colônia seguiram os seguintes protocolos:

Tabela 18: Protocolo para realização da reação de PCR de colônia (CRD-I)

	Volume (μ L)	[] Final
EnzimaTaq DNA Polymerase Recombinante (Promega)	0,10	0,5 unidades
Tampão da Taq DNA Polymerase Recombinante 10X (Promega)	1,00	1,0 X
MgCl ₂ (25mM)	0,80	2,0mM
Oligonucleotídeos sintetizados (IDT) (primer I) (4,0 μ M)	0,25	1,0 μ M
Oligonucleotídeos sintetizados (IDT) (primer II) (4,0 μ M)	0,25	1,0 μ M
dNTP mix (10mM) (Eppendorf)	0,20	0,2 mM
Amostras de cada colônia	---	---
Água autoclavada	7,40	---
Volume final da reação	10,00	---

Tabela 19: Protocolo para realização da reação de PCR de colônia (CRD-II)

	Volume (μ L)	[] Final
EnzimaTaq DNA Polymerase Recombinante (Promega)	0,10	0,5 unidades
Tampão da Taq DNA Polymerase Recombinante 10X (Promega)	1,00	1,0 X
MgCl ₂ (25mM)	0,80	2,0mM
Oligonucleotídeos sintetizados (IDT) (primer III) (4,0 μ M)	0,25	1,0 μ M
Oligonucleotídeos sintetizados (IDT) (primer IV) (4,0 μ M)	0,25	1,0 μ M
dNTP mix (10mM) (Eppendorf)	0,20	0,2 mM
Amostras de cada colônia	---	---
Água autoclavada	7,40	---
Volume final da reação	10,00	---

As colônias selecionadas e que apresentaram resultado positivo foram utilizadas para preparar estoque em glicerol e armazenadas à temperatura de -80°C. Para a realização dos experimentos de clonagem dos domínios CRD-I e CRD-II em vetor de expressão pET28a (Novagen), utilizou-se os clones designados como *Hgal-4CRD-I(pTZ57R/T)(I-1)* e *Hgal-4CRD-II (pTZ57R/T)(II-2)*.

3.2.6 Processo de digestão das construções *Hgal-4CRD-I(pTZ57R/T)(I-1)* e *Hgal-4CRD-II (pTZ57R/T)(II-2)* e do vetor *pET-28a* (Novagen)

As construções *Hgal-4CRD-I(pTZ57R/T)(I-1)* e *Hgal-4CRD-II(pTZ57R/T)(II-2)* transformadas em bactérias *E.coli DH5α*, assim como, amostras do vetor de clonagem pET28a (Novagen) foram submetidas a um processo de purificação utilizando-se com o Kit QIAprep® Miprep segundo o protocolo sugerido no manual do mesmo. Amostras das construções *Hgal-4CRD-I(pTZ57R/T)(I-1)* e *Hgal-4CRD-II (pTZ57R/T)(II-2)* purificadas e do vetor de clonagem pET-28a (Novagen), foram submetidos à reação de dupla digestão com as enzimas de restrição *Not-I* (Fermentas) e *Nde-I* (Fermentas). As reações foram conduzidas a 37°C por um período de 4 horas.

As reações de digestão seguiram os seguintes protocolos:

Tabela 20: Protocolo para realização da reação de digestão da Construção *Hgal-4CRD-I(pTZ57R/T)(I-1)*

	Volume (μL)	[] Final
Construção <i>Hgal-4CRD-I(pTZ57R/T)(I-1)</i>	10,0	1,4μg
Enzima de restrição <i>Nde I Fast Digest</i> (Fermentas)	1,0	10 unidades
Enzima de restrição <i>Not I Fast Digest</i> (Fermentas)	1,0	10 unidades
Tampão 0 (Fermentas)	2,0	1,0X
H ₂ O (autoclavada)	6,0	---
Volume final da reação	20,0	---

Tabela 21: Protocolo para realização da reação de digestão da construção *Hgal-4CRD-II(pTZ57R/T)*

	Volume (μL)	[] Final
Construção <i>Hgal-4CRD-II(pTZ57R/T)(II-2)</i>	10,0	1,4μg
Enzima de restrição <i>Nde I Fast Digest</i> (Fermentas)	1,0	10 unidades
Enzima de restrição <i>Not I Fast Digest</i> (Fermentas)	1,0	10 unidades
Tampão 0 (Fermentas)	2,0	1,0X
H ₂ O (autoclavada)	6,0	---
Volume final da reação	20,0	---

Tabela 22: Protocolo para realização da reação de digestão do vetor pET28a (Novagen)

	Volume (μL)	[] Final
Vetor pET28a (Novagen)	10,0	1,0μg
Enzima de restrição <i>Nde I Fast Digest</i> (Fermentas)	1,0	10 unidades
Enzima de restrição <i>Not I Fast Digest</i> (Fermentas)	1,0	10 unidades
Tampão 0 (Fermentas)	2,0	1,0X
H ₂ O (autoclavada)	6,0	---
Volume final da reação	20,0	---

Após o processo de digestão das construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II (pTZ57R/T)(II-2), as amostras foram purificadas do gel utilizando-se o Kit QIAquick® Gel Extraction Kit Protocol (Qiagen) seguindo-se para tal procedimento as especificações do fabricante.

Amostras do vetor pET-28a (Novagen), foram submetidas a um processo de purificação para a remoção das enzimas, assim como do tampão utilizado na reação. Para tal, utilizou-se o kit QIAquick® PCR Purification Kit (Qiagen) segundo as especificações de uso descritas pelo fabricante.

3.2.7 Ligação dos CRDs da galectina-4 humana no vetor pET28a (Novagen)

As reações de ligação dos fragmentos previamente purificados a partir do gel de agarose e digeridos (CRD-I e CRD-II) no vetor pET28a (Novagen) foram realizadas utilizando-se a enzima T4 DNA Ligase (Invitrogen). No planejamento dos experimentos de ligação, inicialmente fez-se uma estimativa da quantidade de inserto e do vetor após os ensaios de digestão. Para isto, as amostras foram submetidas a uma eletroforese em um gel de agarose 0,8% (w/v) e as intensidades das bandas foram comparadas às intensidades obtidas em uma amostra padrão de DNA (Biolabs). Para maximizar as chances de sucesso nos experimentos de ligação, preparou-se diferentes reações, onde diferentes proporções entre os fragmentos (CRD-I e CRD-II) e vetor pET28a (Novagen) foram testadas. Para controle do experimento, o sucesso da digestão foi monitorado através de uma reação de ligação na ausência do inserto. As reações de ligação foram realizadas a temperatura de 16°C por um período de 16 horas.

As reações de ligação seguiram os seguintes protocolos:

Tabela 23: Protocolo para realização da reação de ligação do domínio CRD-I no vetor pET28a.

	1	2	3	4	5	6
Vetor pET28a (Novagen)	2 µL	2 µL	2 µL	4 µL	2 µL	2 µL
Inserto CRD-I	2 µL	4 µL	8 µL	12 µL	---	---
Enzima T4 DNA Ligase (Invitrogen)	1 µL	1 µL	1 µL	1 µL	1 µL	---
Tampão T4 DNA Ligase (Invitrogen)	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL
H ₂ O (autoclavada)	11 µL	9 µL	5 µL		13 µL	14 µL
Volume final da reação	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

Tabela 24: Protocolo para realização da reação de ligação do domínio CRD-II ao vetor pET28a.

	1	2	3	4	5	6
Vetor pET28a (Novagen)	2 µL	2 µL	2 µL	4 µL	2 µL	2 µL
Inserto CRD-II	2 µL	4 µL	8 µL	12 µL	---	---
Enzima T4 DNA Ligase (Invitrogen)	1 µL	1 µL	1 µL	1 µL	1 µL	---
Tampão T4 DNA Ligase (Invitrogen)	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL
H ₂ O (autoclavada)	11 µL	9 µL	5 µL		13 µL	14 µL
Volume final da reação	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

Como resultado do processo de ligação dos fragmentos codificando os domínios de reconhecimento (CRDs) da proteína galectina-4 humana no vetor de expressão pET28a (Novagen), obteve-se as construções designadas por **Hgal-4CRD-I(pET28a)** e **Hgal-4CRD-II(pET28a)**.

3.2.8 Transformação das construções *Hgal-4CRD-I(pET28a)* e *Hgal-4CRD-II(pET28a)* em bactérias *Escherichia coli* DH5α

No processo de transformação das construções *Hgal-4CRD-I(pET28a)* e *Hgal-4CRD-II(pET28a)* em bactérias *Escherichia coli* utilizou-se a técnica de tratamento com uma solução de cloreto de cálcio e mudanças bruscas de temperatura para indução da competência nas cepas da bactéria *Escherichia coli* DH5α. (ver apêndice II). Após a obtenção das bactérias competentes, as construções **Hgal-4CRD-I(pET28a)** e **Hgal-4CRD-II(pET28a)** foram transformadas na bactéria *Escherichia coli*, utilizando a cepa DH5α (Novagen) (ver apêndice III).

Para a verificação da presença do gene de interesse referente ao domínio **CRD-I**, as colônias obtidas e selecionadas foram utilizadas para inocular 5 mL de meio LB (Luria-Bertani) + (50 µg/mL) de kanamicina. Após o período de incubação (overnight a 37°C/250 rpm), as amostras foram purificadas utilizando-se com o Kit QIAprep® Miprep segundo o protocolo sugerido no manual do mesmo e a análise da presença dos insertos em cada colônia foi

realizada através de uma reação de digestão utilizando-se as enzimas de restrição *NdeI* Fast digest (Fermentas) e *NotI* Fast digest (Fermentas) seguindo especificações do fabricante. O produto da digestão foi submetido a uma eletroforese em gel de agarose 0,8% (w/v) para visualização da presença do inserto dentre as colônias selecionadas. A reação permaneceu por 16 horas a temperatura de 37°C.

A reação de digestão seguiu o seguinte protocolo:

Tabela 25: Protocolo para realização da reação de digestão (CRD-I)

	Volume (μL)	[] Final
Amostra da construção Hgal-4CRDI(pET 28a)	10,0	1,4μg
Enzima de restrição <i>Nde I Fast Digest</i> (Fermentas)	1,0	10 unidades
Enzima de restrição <i>Not I Fast Digest</i> (Fermentas)	1,0	10 unidades
Tampão (10) (Fermentas)	2,0	1,0X
H ₂ O (autoclavada)	6,0	---
Volume final da reação	20,0	---

As colônias testadas e que apresentaram resultado positivo foram mantidas a - 80°C sob a forma de glicerol estoque. Para a realização dos experimentos de expressão do domínio CRD-I utilizou-se o clone designado como ***Hgal-4CRD-I(pET28a)(I-2)***.

Para os ensaios de expressão, amostras da construção ***Hgal-4CRD-II(pET28a)(I-2)*** foram transformadas em *Escherichia coli*, utilizando a cepa BL-21(DE3) (Novagen), resistentes ao antibiótico kanamicina e mantidas a -80°C, sob a forma de glicerol estoque.

Para a verificação da presença do gene de interesse referente ao domínio **CRD-II**, as várias colônias obtidas e selecionadas foram utilizadas para inocular 5,0 mL de meio LB (Luria-Bertani) + (50 μg/mL) de antibiótico kanamicina. Após o período de incubação (*overnight* a 37°C/250 rpm), a análise da presença dos insertos em cada colônia foi realizada através de uma reação de PCR de colônia. O produto de PCR foi submetido a uma eletroforese em gel de agarose 0,8% (w/v) para visualização da presença do inserto dentre as colônias obtidas.

As reações de PCR seguiram o seguinte protocolo:

Tabela 26: Protocolo para realização da reação de PCR (CRD-II).

Volume (μL)	[] Final
-------------	-----------

EnzimaTaq DNA Polymerase Recombinante (Promega)	0,25	1,25 unidades
Tampão da Taq DNA Polymerase Recombinante 10X (Promega)	2,00	1,0 X
MgCl ₂ (25mM)	0,80	2,0mM
Oligonucleotídeos sintetizados (IDT) (primer III) (4,0μM)	0,25	0,5μM
Oligonucleotídeos sintetizados (IDT) (primer IV) (4,0μM)	0,25	0,5μM
dNTP mix (10mM) (Eppendorf)	0,20	0,2 mM
Amostras de cada colônia	2,00	---
Água autoclavada	14,25	---
Volume final da reação	20,00	---

As colônias testadas e que apresentaram resultado positivo foram mantidas a - 80°C sob a forma de glicerol estoque. Para a realização dos experimentos de expressão do domínio CRD-II utilizou-se o clone designado como *Hgal-4CRD-II(pET28a)(II-23)*.

O clone designado como *Hgal-4CRD-II(pET28a)(II-23)* foi purificado com o Kit QIAprep® Miprep segundo o protocolo sugerido no manual do fabricante. Para os ensaios de expressão, amostras da construção *Hgal-4CRD-II(pET28a)(II-23)* foram transformadas em *Escherichia coli*, utilizando a cepa BL-21(DE3) (Novagen), resistentes ao antibiótico kanamicina e mantidas a -80°C, sob a forma de glicerol estoque.

3.2.9 Expressão e purificações domínios CRD-I e CRD-II da galectina-4 humana.

3.2.9.1 Expressão e obtenção do domínio CRD-I da galectina-4 humana.

Os primeiros ensaios de expressão em *Escherichia coli* BL-21(DE3) foram realizados para o domínio CRD-I. Todos os experimentos de expressão foram realizados em ambiente estéril. Para a preparação do pré-inóculo, uma amostra do glicerol estoque contendo a construção *Hgal-4CRD-I(pET28a)(I-2)/(BL-21(DE3))* foi utilizada para inocular 10,0 mL de meio LB (Luria-Bertani), na presença do antibiótico kanamicina (50 μg/mL) e mantido em uma incubadora sob agitação (*overnight* a 250 rpm/37°C). O material foi então utilizado para inocular 1L de meio de cultura LB na presença do antibiótico kanamicina (50 μg/mL) na proporção (1:100) e mantido sob agitação a 250 rpm e 37°C. O crescimento das células foi monitorado medindo-se a DO₆₀₀ (densidade ótica em um comprimento de onda de 600nm) até atingir o valor de absorbância entre 0,45 e 0,50. Ao final desse processo, a indução da expressão da proteína de interesse foi feita através da adição de IPTG (isopropyl β-D-1-tiogalactopiranosídeo) a uma concentração de 1mM e permaneceu sob agitação (*overnight* / 250 rpm /30°C).

As amostras das células foram isoladas do meio de cultura por centrifugação e armazenadas no freezer a -20°C.

Para obtenção da proteína de interesse, as células congeladas foram ressuspensas em tampão de lise (50mM NaH₂PO₄, 300mM NaCl, pH 8,0) e 1,0 mM de Inibidor de protease PMSF) (Roche) . Para ressuspender as células foram utilizados aproximadamente 3,0 mL do tampão de lise para cada 1 grama de células. O material ressuspensionado foi sonificado por 8 vezes em intervalos de tempo de 20 segundos. Após a sonificação, a amostra foi centrifugada por 45 minutos a 10.000g e 4°C, para separação do material solúvel (*sobrenadante*) e insolúvel (*pellet*) que foi mantido a temperatura de 4°C.

A análise do padrão de expressão foi feita através de uma eletroforese em gel de poliacrilamida (SDS-PAGE).

3.2.9.2 Expressão e obtenção do domínio CRD-II da galectina-4 humana.

Para os ensaios de expressão do domínio CRD-II da galectina-4 humana, amostras do glicerol estoque contendo a construção Hgal-4CRD-II(pET28a)(II-23)/(BL-21(DE3)) foram utilizadas para inocular 10,0 mL de meio LB (Luria-Bertani), na presença do antibiótico kanamicina (50 µg/mL) e mantido sob agitação (*overnight* a 250 rpm/37°C). O material foi então utilizado para inocular 1L de meio de cultura LB na presença do antibiótico kanamicina (50 µg/mL) na proporção de 1:100 e mantido sob agitação a 250 rpm e 37°C. O crescimento das células foi monitorado medindo-se a DO₆₀₀ (densidade ótica do meio de cultura em um comprimento de onda de 600nm) até atingir o valor de absorbância entre 0,45 e 0,50. Ao final desse processo, a indução da expressão da proteína de interesse foi feita através da adição de IPTG (isopropyl β-D-1-tiogalactopiranosídeo) a uma concentração de 1mM e permaneceu sob agitação (*overnight* /250 rpm/30°C).

As amostras das células foram isoladas do meio de cultura por centrifugação e armazenadas no freezer -20°C.

Para obtenção e purificação do domínio CRD-II, todos os ensaios foram executados na presença do agente redutor, β-mercaptoetanol a uma concentração de 14 mM. Este procedimento foi empregado devido ao fato do domínio CRD-II apresentar em sua sequência primária 2 resíduos de cisteína, os quais, segundo estudos realizados, sob a forma oxidada, puderam induzir a uma mudança conformacional na molécula, dificultando os experimentos de cristalização e subsequentemente o estudo cristalográfico desta proteína.

As células congeladas foram ressuspensas em tampão de lise consistindo de tampão (50mM NaH₂PO₄, 300mM NaCl, pH 8,0), 1,0 mM de Inibidor de protease PMSF) (Roche) e β-mercaptoetanol 14 mM. Para ressuspender as células foram utilizados aproximadamente 3,0 mL do tampão de lise para cada 1 grama de células. O material ressuspendido foi sonificado por 8 vezes em intervalos de tempo de 20 segundos. Após a sonicação, a amostra foi centrifugada por 45 minutos a 10.000g e 4°C, para separação do material solúvel (*sobrenadante*) e insolúvel (*pellet*) e mantidas a temperatura de 4°C.

A análise do padrão de expressão foi feita através de uma eletroforese em gel de poliacrilamida (SDS-PAGE).

3.2.10 Purificação dos CRD-I e CRD-II da galectina-4 humana

3.2.10.1 Cromatografia por Afinidade (Ni-NTA)

O segundo método cromatográfico utilizado para a purificação dos domínios de reconhecimento de carboidratos (CRD-I e CRD-II) da galectina-4 humana foi a cromatografia de afinidade. Neste processo, as proteínas são separadas de acordo com a afinidade por um determinado ligante. Nestes ensaios, os domínios CRD-I e CRD-II foram expressos em fusão com uma cauda contendo seis aminoácidos de histidina, que confere a proteína em pH apropriado, afinidade por íons bivalentes carregados positivamente. A resina utilizada nesta etapa possui como matriz (conjunto ligante-polímero) íons de níquel ligados a agarose, Ni-NTA (Qiagen) (**Figura 15**).

Figura 15: Interação entre a cauda de histidina da proteína (em azul) e o Ni^{2+} da resina Ni-NTA Agarose (em vermelho).

Para os experimentos de purificação do domínio CRD-I, a resina foi inicialmente equilibrada com o tampão de ligação (pH 8, 50 mM NaH_2PO_4 e 300 mM NaCl), o extrato bruto foi adicionado e o eluente coletado.

Para a remoção de contaminantes que eventualmente possam interagir com a resina Ni-NTA (Qiagen), a mesma é lavada com um gradiente crescente de imidazol (0 mM à 50 mM imidazol em 100 mL). A proteína (**Hgal-4-CRD-I**) foi eluída na presença de 100 mM de imidazol (pH 8, 50 mM NaH_2PO_4 , 300 mM NaCl e 100 mM imidazol) em 8 alíquotas de 2 mL cada.

O grau de pureza da proteína foi monitorado por eletroforese em gel de poliacrilamida (12 % SDS-PAGE). Na revelação do gel, empregou-se o corante “*Coomassie brilliant blue*”.

Para os experimentos de purificação do domínio CRD-II, utilizou-se em todas as etapas o agente redutor β -mercaptoetanol. A resina Ni-NTA (Qiagen) foi inicialmente equilibrada com o tampão de ligação (pH 8, 50 mM NaH_2PO_4 e 300 mM NaCl e β -mercaptoetanol 14mM), o extrato bruto foi adicionado e o eluente coletado. Para a remoção de contaminantes que eventualmente possam interagir com a resina Ni-NTA (Qiagen), a mesma foi lavada com um gradiente crescente de imidazol (0 mM à 100 mM imidazol e β -mercaptoetanol 14mM). A proteína (**Hgal-4-CRD-II**) foi eluída na presença de 150 mM de imidazol (pH 8, 50 mM NaH_2PO_4 , 300 mM NaCl e 150 mM imidazol e β -mercaptoetanol 14mM) em 9 alíquotas de 2 mL cada. O grau de pureza da proteína foi monitorado por eletroforese em gel de poliacrilamida (12% SDS PAGE).

3.2.11. Dosagem dos domínios CRD-I e CRD-II da galectina-4 humana.

3.2.11.1 Método do coeficiente de extinção molar

A estimativa da concentração das proteínas Hgal-4-CRD-I e Hgal-4-CRD-II foi determinada através da Lei de Lambert-Beer. Esta lei relaciona a intensidade da luz incidente, I_0 , e a intensidade da luz transmitida, I , por um meio transparente de espessura b e concentração c através da relação dada por:

$$A = \log I_0/I = \epsilon.b.c$$

ϵ , o coeficiente de proporcionalidade, é denominado coeficiente de extinção, o qual é característico da espécie absorvente e varia com o comprimento de onda e o solvente empregado; A é a absorbância da solução e b é denominado caminho óptico.

No caso específico de macromoléculas, os aminoácidos triptofano, tirosina e fenilalanina, devido à presença de grupos aromáticos em suas cadeias laterais, são responsáveis pelas propriedades de absorção de luz das proteínas, em particular para comprimento de onda de 280 nm. Sendo assim, com base na sequência de aminoácidos que compõem a enzima foi possível estimar o coeficiente de extinção molar esperado para dos domínios CRD-I e CRD-II, (NELSON; COX, 2002). O coeficiente de extinção molar teórico a 280 nm, calculado pelo programa *ProtParam* (www.expasy.ch/tools/protparam.html), para os CRD-I e CRD-II equivale a 24,41 mM⁻¹cm⁻¹ e 11,58 mM⁻¹cm⁻¹, respectivamente.

3.3 *Cristalização dos domínios CRD-I e CRD-II da galectina-4 humana*

A cristalização dos CRDs da galectina-4 humana visa o estudo estrutural desta classe de proteínas através da utilização de métodos de difração de raios-X em monocristais.

Para a formação de um cristal, a solução de proteína deve ser levada para um estado de supersaturação; estado termodinamicamente instável, onde as moléculas de proteína são agrupadas de forma organizada. Diversos fatores podem ser combinados de forma a levar a solução de proteína para o estado de supersaturação e conseqüentemente formação do cristal. Os mais comumente utilizados são: temperatura, concentração do agente precipitante, pH, concentração de proteína, adição de ligantes que modificam a solubilidade da proteína, aumento e diminuição da concentração salina (“salting out” e salting in”), dentre inúmeros outros.

Existem várias técnicas utilizadas para a cristalização de proteínas como: diálise, difusão líquido-líquido e difusão de vapor. Dois dos métodos mais comuns para cristalização utilizam a técnica por difusão de vapor. Estes são conhecidos como o método da gota suspensa (“hanging drop”) e gota sentada (“sitting drop”) (MCPHERSON, 1982). Em ambos, uma gota contendo a solução de cristalização (tampão, precipitantes e aditivos, também chamada de líquido-mãe) é misturada a solução de proteína e equilibrada contra um reservatório (poço) contendo a mesma solução de cristalização. Desta forma, a solução de cristalização encontrada na gota se encontra inicialmente mais diluída do que a solução encontrada no poço, e a diferença de concentração inicial entre a gota e poço depende da razão de diluição inicial da gota. Este sistema é selado e o equilíbrio é alcançado por meio da difusão das espécies voláteis (água e solventes orgânicos) até que a pressão de vapor na gota se iguale

àquela do reservatório (poço). A concentração dos precipitantes na gota aumenta, diminuindo assim a solubilidade da solução de proteína e permitindo, em condições específicas, a formação de cristais.

Os métodos de gota suspensa e sentada diferem somente na posição em que a gota é colocada no experimento em relação à solução do poço. A técnica de difusão de vapor, em gota sentada, foi a escolhida para os ensaios de cristalização realizados neste projeto.

3.3.1 *Cristalização do domínio CRD-I da galectina-4 na forma “apo”*

A cristalização da proteína na forma “**apo**” (livre de ligantes) constitui o primeiro objetivo do nosso trabalho. Além de fornecer dados estruturais relevantes da proteína, esses cristais podem ser utilizados em experimentos de “*soaking*” para obtenção de complexos ligantes.

Foram testadas inicialmente 98 condições disponíveis nos kits de cristalização Crystal Screen™ e Crystal Screen 2™ (Hampton Research) e 48 condições disponíveis no kit Peg Ion (Hampton Research). Neste experimento são testados um grande intervalo de pH, sais, precipitantes e aditivos (JANCARIK; KIM, 1991). Os ensaios de cristalização foram realizados a temperatura de 22°C.

Nos primeiros experimentos de cristalização utilizando o método de difusão de vapor utilizou-se 500 µL das soluções de cristalização no reservatório. As gotas consistiam de 2 µL da solução do reservatório e 2 µL de proteína (7 mg/mL) (volume total da gota igual a 4 µL).

3.3.2 *Cristalização do domínio CRD-I da galectina-4 com lactose*

Além da cristalização do domínio CRD-I na forma apo, outro objetivo do nosso trabalho era o de determinar a estrutura da proteína em complexo com o carboidrato lactose, um ligante natural. Ensaios de cristalização do domínio CRD-I na presença de lactose foram realizados com o objetivo de obter cristais do complexo-proteína, e dessa maneira, a partir dos dados cristalográficos e estudos de caracterização estrutural, analisar as principais interações ocorridas entre a proteína e o carboidrato.

Os ensaios de cristalização do domínio CRD-I na presença de lactose foram realizados utilizando-se as melhores condições obtidas na obtenção dos cristais na forma apo.

Foram realizados ensaios de cristalização primeiramente incubando a proteína com uma solução de lactose a 50mM (*overnight*).

3.3.3 Experimentos de “soaking” do domínio CRD-I com lactose

Nesta técnica, os cristais na forma apo, previamente obtidos, são transferidos para uma solução contendo os ligantes, no caso a lactose e incubados por um determinado período de tempo antes da realização da coleta de dados. A concentração do inibidor e o tempo de incubação utilizados nestes experimentos são parâmetros que foram variados na tentativa de uma melhora na qualidade dos dados coletados e na incorporação do ligante. Neste ensaio adicionamos a lactose na solução crioprotetora testando diferentes concentrações (0,5 a 2,0 mM) e o cristal foi transferido para a solução.

3.3.4 Cristalização do domínio CRD-II da galectina-4 na forma “apo”

Para os ensaios de cristalização do CRD-II, foram testadas inicialmente 98 condições disponíveis nos kits de cristalização Crystal ScreenTM e Crystal Screen 2TM (Hampton Research). Neste experimento são testados um grande intervalo de pH, sais, precipitantes e aditivos (JANCARIK; KIM, 1991). Os ensaios de cristalização foram realizados a temperatura de 22°C.

Nos primeiros experimentos de cristalização utilizando o método de difusão de vapor utilizou-se 500 µL das soluções de cristalização no reservatório. As gotas consistiam de 2 µL da solução do reservatório e 2 µL de proteína (8,9 mg/mL), volume total da gota igual a 4 µL.

3.4 Coleta e Processamento de Dados para cristais do domínio CRD-I

Os experimentos de difração de raios-X foram realizados em fonte de radiação Síncrotron da linha de cristalografia de proteínas do Laboratório Nacional de Luz Síncrotron (DO3BMX1 - LNLS). As coletas foram realizadas com comprimento de onda, λ , igual a

1,437 Å, em temperatura criogênica (100 K) e os dados registrados em um detector do tipo marCCD165 (MarResearch).

Cada conjunto de imagens coletado foi processado usando o programa *MOSFLM* (LESLIE, 1992) e escalonado com o programa *SCALA (CCP4)* (EVANS, 1997), separando-se 5 % das reflexões totais para o conjunto teste (R_{free}) a ser utilizado no teste de validação cruzada (BRUNGER et al., 1992).

3.5 Resolução da Estrutura do domínio CRD-I

O conteúdo de solvente, assim como o número de subunidades da proteína na unidade assimétrica foi estimado calculando o coeficiente de Matthews (MATTHEWS, 1968) utilizando o programa *MATTHEWS_COEF* dentro do pacote de programas *CCP4* (CCP4, 1994).

A resolução da estrutura do domínio N-terminal, CRD-I, da galectina-4 humana foi efetuada através da técnica de substituição molecular, implementada no programa *MOLREP (CCP4)* (VAGIN; TEPLYAKOV, 1997) utilizando, no primeiro momento, as coordenadas da estrutura do domínio CRD N-terminal da galectina-4 de rato (PDB código **2DYC**) como modelo. A solução encontrada pelo programa de substituição molecular, rotação e translação, foi aplicada ao modelo do domínio N-terminal da galectina-4 de rato (KATO-MURAYAMA et al., manuscrito em preparação), que possui 81% de identidade sequencial com o domínio N-terminal da galectina-4 humana (**Figura 16**). Após a obtenção do primeiro modelo do domínio CRD-I da galectina-4 humana, este foi utilizado na resolução da estrutura com um novo conjunto de dados de melhor qualidade.

	1	10	20	30	40	50	60
CRD1_HUMANA	MAYVPAP	GYQPTYNPTLPYYQ	IPGGLN	VGMSVYIQGV	ASEHMKRFFVNFV	VGQDP	GSDV
CRD1_RATO	.GSSGSS	GYQPTYNPTLPYKR	IPGGLS	VGMSVYIQGM	AKENMRRFHVNF	AVGQDD	GADV
	70	80	90	100	110	120	
CRD1_HUMANA	AFHFNPRFDGWDKVVFN	TLQSGK	WGSSEERKR	SMPFKKGA	AFELVFIVLA	EHYKV	VVNGNP
CRD1_RATO	AFHFNPRFDGWDKVVFN	TLQSGQW	GKEEKKS	SMPFQK	GFELVFMV	MP	EHYKV
	130	140	150				
CRD1_HUMANA	FYEYGHRLPLQ	MVTHLQVDGDL	QLQSINF	I			
CRD1_RATO	FYEYGHRLPLQ	MVTHLQVDGDL	QLQSINF	LGGQP	AAAP	Y	

Figura 16: Alinhamento sequencial entre os domínios N-terminais da galectina-4 de rato e humana. Em vermelho estão selecionados os aminoácidos conservados em ambas sequencias.

3.6 Refinamento

O refinamento da estrutura cristalográfica foi realizado utilizando o programa *REFMAC (CCP4)* (MURSHUDOV et al., 1999) onde está implementado o princípio de máxima entropia (“Maximum Likelihood”). A primeira etapa de refinamento consistiu em um refinamento de corpo rígido, cujo objetivo é aperfeiçoar os parâmetros rotacionais e translacionais obtidos na etapa de substituição molecular. A construção visual da estrutura foi realizada no programa gráfico *Coot* (EMSLEY; COWTAN, 2004). O ajuste manual das posições atômicas foi realizado levando-se em consideração as restrições estereoquímicas, assim como a concordância com os mapas pesados de densidade eletrônica $2F_o - F_c$ ($2mF_o - DF_c$), com σ igual a 1, e o mapa diferença $F_o - F_c$ ($mF_o - DF_c$), com σ igual a 3. F_o (hkl) corresponde ao fator de estrutura observado para a reflexão hkl e F_c corresponde ao fator de estrutura para a reflexão hkl calculado baseado no modelo.

Após o ajuste de todos os aminoácidos na densidade eletrônica ($2F_o - F_c$), moléculas de água foram automaticamente adicionadas pelo programa *Coot* (EMSLEY; COWTAN, 2004) através das posições dos picos no mapa diferença $F_o - F_c$, com σ igual a 3, e validadas através da checagem visual levando-se em consideração o ambiente químico de cada molécula de solvente. Algumas águas, que não foram encontradas automaticamente, foram adicionadas manualmente. A busca de ligantes na estrutura, compostos utilizados nas condições de cristalização e agente crioprotetor, foi realizada utilizando o programa *Coot* (EMSLEY; COWTAN, 2004). As coordenadas iniciais para os ligantes foram obtidas do site HIC-Up (<http://xray.bmc.uu.se/hicup/>).

O sucesso de cada ciclo de refinamento foi monitorado através da diminuição dos valores de R_{factor}^1 e R_{free}^2 , até a obtenção de valores aceitáveis para a resolução na qual o modelo estava sendo refinado.

$$^1R_{factor} = \frac{\sum_{(100-T)\%hkl} \|F_{obs} - k|F_{cal}|\|}{\sum_{(T-100)\%hkl} |F_{obs}|} \times 100\%$$

$$^2R_{free} = \frac{\sum_{(T)\%hkl} \|F_{obs} - k|F_{cal}|\|}{\sum_{(T)\%hkl} |F_{obs}|} \times 100\%$$

T é a porcentagem de reflexões escolhidas para o conjunto teste

3.7 Avaliação dos dados estereoquímicos

A validação estereoquímica da estrutura foi realizada utilizando o programa *PROCHECK* (LASKOWSKI *et al.*, 1998), onde são verificados diversos parâmetros, como: comprimentos e ângulos de ligação, quiralidade, planaridade das ligações peptídicas, entre outros.

4. RESULTADOS

4.1 Clonagem dos CRD-I e CRD-II em vetor de expressão pET-29a (Novagen).

4.1.1 Amplificação dos CRD-I e CRD-II da galectina-4 humana.

No presente estudo, os genes que codificam os domínios de reconhecimento de carboidratos CRD-I e CRD-II da galectina-4 humana foram amplificados por PCR (reação em cadeia da polimerase), utilizando como molde para a reação a construção onde o gene completo que codifica a galectina-4 humana se encontra clonado no vetor pQE (Qiagen). Foram inicialmente testadas três diferentes temperaturas de anelamento: 50°C, 55°C e 60°C (Figura 17). Baseado no tamanho dos fragmentos amplificados, a análise dos resultados sugere inicialmente que os primers foram eficientes para amplificar os domínios de interesse: CRD-I (450pb) e CRD-II (432pb). Das temperaturas testadas para o anelamento, concluímos que as reações de PCR são mais eficazes às temperaturas de 60°C para CRD-I e 50°C para CRD-II.

Figura 17: Eletroforese em gel de agarose 0,8% (w/v) mostrando o perfil da amplificação do gene de interesse através da técnica de PCR para diferentes temperaturas de anelamento. Os genes que codificam os domínios CRD-I e CRD-II possuem 450 bp e 435 bp respectivamente. As pistas 1, 2 e 3 pertencem ao CRD-I e as pistas 4, 5 e 6 ao CRD-II.

Os fragmentos amplificados por PCR e purificados através do gel de agarose, assim como amostras do vetor pET29a (Novagen), foram submetidos ao processo de digestão utilizando-se as enzimas de restrição *Nde I* e *Not I* (Invitrogen). Os produtos da digestão foram submetidos à eletroforese em gel de agarose 8% (w/v) para quantificação previamente aos ensaios de ligação. Para o processo de quantificação dos fragmentos resultantes do processo de digestão, amostras do vetor pET29a, assim como, dos fragmentos CRD-I e CRD-II digeridos foram aplicadas em gel de agarose 8% (w/v). As bandas visualizadas em gel foram então comparadas às bandas apresentadas por uma amostra padrão de DNA (Biolabs). Os fragmentos referentes aos domínios CRD-I, CRD-II e o vetor pET29a apresentaram uma banda de aproximadamente 40 ng, 40 ng e 30 ng respectivamente, quando comparadas a amostra padrão, (**Figura 18**).

Figura 18: Eletroforese em gel de agarose 0,8% (w/v). A figura mostra o perfil da digestão do CRD-I, CRD-II e do vetor pET-29a (Novagen) com as enzimas de restrição *Nde-I* e *Not I* (Invitrogen). As pistas 1 e 2 referem-se ao CRD-I e CRD-II respectivamente e apresentam uma banda de aproximadamente 450 e 435 pb. A pista 3 refere-se ao vetor pET-29a, e apresenta uma banda de aproximadamente 5300 pb. As massas obtidas no processo de quantificação foram, 8,0 ng/ μ L, 8,0 ng/ μ L e 10,0 ng/ μ L para os CRD-I, CRD-II e vetor pET29a respectivamente.

Os fragmentos digeridos e quantificados foram então submetidos a uma reação de ligação no vetor pET29a. Como resultado do processo de ligação obteve-se as construções designadas por Hgal-4CRD-I(pET29a) e Hgal-4CRD-II(pET29a). Amostras das construções obtidas foram então transformadas em bactérias *Escherichia coli DH5a*. Após o término do

período de incubação em estufa a uma temperatura de 37°C, por um período de 16 horas, as colônias obtidas e selecionadas, foram submetidas a uma reação de PCR de colônia para verificação da presença do gene de interesse. A análise da presença dos insertos em cada colônia foi realizada através de eletroforese em gel de agarose 0,8% (w/v) (**Figuras 19 e 20**).

Para os experimentos relacionados ao CRD-I, foram testadas nove colônias, dentre essas, o gel de agarose nos revelou a presença do gene de interesse nas colônias designadas por 1, 8 e 9, (**Figura 19**), já para os experimentos relacionados ao domínio CRD-II, o gel de agarose nos revelou a presença do gene de interesse em todas as nove colônias selecionadas e testadas (**Figura 20**). As colônias que apresentaram resultados positivos referentes ao CRD-I e CRD-II foram armazenadas como glicerol estoque e mantidas a – 80°C.

Figura 19: Eletroforese em gel de agarose 0,8% (w/v). Dentre as nove colônias selecionadas e testadas, o gel de agarose nos mostra a presença do gene referente ao domínio CRD-I nas colônias designadas como 1, 8 e 9. O tamanho da banda observada equivale aproximadamente 450 pb quando comparada ao padrão. Os clones obtidos foram designados de I-1, I-2 e I-3. O clone I-1 foi utilizado nos ensaios de expressão.

Padrão 1 2 3 4 5 6 7 8 9

Figura 20: Eletroforese em gel de agarose 0,8% (w/v). Dentre as nove colônias selecionadas e testadas, o gel de agarose nos mostra a presença do gene referente ao domínio CRD-II em todas as colônias. O tamanho da banda observada equivale aproximadamente 435 pb quando comparada ao padrão. Os clones obtidos foram designados como II-1, II-2, II-3, II-4, II-5, II-6, II-7, II-8 e II-9. O clone II-4 foi utilizado nos ensaios de expressão.

4.2 Expressão dos domínios CRD-I e CRD-II da galectina-4 humana em *E.coli* BL21(DE3)pLys

4.2.1 Expressão do domínio CRD-II da galectina-4 humana

Para os experimentos de expressão do domínio CRD-II, utilizou-se o clone designado como Hgal-4CRD-II(pET29a)(II-4). O domínio CRD-II da Hgal-4 foi expresso em cepas de *E.coli* BL-21(DE3)pLys. Procederam-se vários experimentos no sentido de otimizar as condições de expressão como também de maximizar o rendimento da proteína para os experimentos de cristalização. A expressão do CRD-II da Hgal-4 foi primeiramente induzida com 1,5 mM de IPTG a 37 °C, condições estas otimizadas para a expressão da Hgal-4 completa. O perfil de expressão proteica foi monitorado em períodos de 2, 3, 4, 5, 18 e 24 horas de indução. Após a lise celular, amostras do sobrenadante e pellets obtidos foram submetidas à eletroforese em gel de poliacrilamida (SDS-PAGE) para visualização do padrão de expressão (**Figura 21**).

Figura 21: Eletroforese em gel de poliacrilamida (12% SDS-PAGE) para o ensaio de expressão do domínio CRD-II utilizando-se 1,5 mM de IPTG a 37°C por 2, 3, 4, 5, 18 e 24 horas de indução. As legendas PC, PI, SC e SI correspondem a amostras do pellet controle e induzido, sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-II é de aproximadamente 17 kDa.

Como não foi possível visualizar nenhuma banda de expressão para o domínio CRD-II decidimos testar diferentes parâmetros utilizados como concentração de IPTG e redução da temperatura de indução. Foram realizados ensaios de expressão, onde diferentes concentrações de IPTG foram testadas: 0,5 mM e 1,0 mM de IPTG a 37 °C por 5 horas de indução, assim como, a redução da temperatura de indução para 18 °C (1,5 mM de IPTG a 18°C por 24 horas). Amostras do sobrenadante e pellet foram submetidos à eletroforese em gel de poliacrilamida (SDS-PAGE) para visualização do padrão de expressão (**Figuras 22 e 23**).

Figura 22: Eletroforese em gel de poliacrilamida (SDS-PAGE-12%). Ensaio de expressão do domínio CRD-II, utilizando-se 0,5 mM e 1,0 mM de IPTG a 37°C, por 5 horas. As legendas PC, PI, SC e SI correspondem às amostras: pellet controle e induzido; sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-II é de aproximadamente 17 kDa.

Figura 23: Eletroforese em gel de poliacrilamida (SDS-PAGE 12%). Ensaio de expressão do domínio CRD-II induzido com 1,5 mM de IPTG a 18°C, por 24 horas. As legendas PC, PI, SC e SI correspondem às amostras: pellet controle e induzido; sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-II é de aproximadamente 17 kDa.

Mesmo com uma análise detalhada dos experimentos de eletroforese para as diferentes condições testadas, não foi possível visualizar nenhuma banda de expressão para o domínio CRD-II.

4.2.2 Expressão do domínio CRD-I da galectina-4 humana

Para os experimentos de expressão do domínio e CRD-I, utilizou-se o clone designado como Hgal-4CRD-I(pET29a)(I-1). O domínio CRD-I da Hgal-4 foi expresso em cepas de *E.coli* BL-21(DE3)pLys. A expressão do CRD-I da Hgal-4 foi induzida com 1,5 mM de IPTG a 37 °C, condições estas utilizadas com sucesso na expressão da Hgal-4 completa. O perfil de expressão proteica foi monitorado em períodos de 5 e 24 horas de indução. Após a lise celular, amostras do sobrenadante e pellets obtidos foram submetidas à eletroforese em gel de poliacrilamida 12% (SDS-PAGE) para visualização do padrão de expressão (**Figura 24**).

Figura 24: Eletroforese em gel de poliacrilamida (SDS-PAGE 12%). Ensaio de expressão do domínio CRD-I induzido com 1,5 mM de IPTG a 37°C, por 5 e 24 horas. As legendas PC, PI, SC e SI correspondem a amostra do pellet controle e induzido; sobrenadante controle e induzido respectivamente. A massa molecular esperada para o CRD-I é de aproximadamente 17 kDa.

Não foi possível visualizar nenhuma banda de expressão associada ao domínio de interesse.

4.3 Purificação dos domínios CRD-I, CRD-II e da galectina-4 inteira

Uma vez que os ensaios de expressão em pequena escala para os domínios CRD-I e CRD-II não evidenciava a presença das proteínas de interesse, decidimos utilizar a purificação por afinidade em coluna de lactose, como estratégia para concentrar as nossas proteínas de interesse. Como controle dos experimentos de purificação, os resultados obtidos no processo de purificação dos domínios CRD-I e CRD-II foram comparados aos resultados observados para a galectina-4 humana inteira. Para os ensaios de purificação adotou-se o protocolo utilizado para isolar a proteína HGal-4 inteira. Para os testes, foram utilizadas células provenientes de 1,0 litro de meio de cultura LB em um ensaio de expressão utilizando-se IPTG a 1,5 mM, por um período de 24 horas à temperatura de 37 °C. O perfil de eluição da purificação dos CRD-I e CRD-II da Hgal-4, assim como o perfil de purificação da Hgal-4 inteira, estão representados no cromatograma da **Figura 25**. Os domínios CRD-I, CRD-II e a proteína inteira foram eluídos em 5, 5 e 6 frações de 2,0 mL cada respectivamente. Os

materiais eluídos foram submetidos a uma eletroforese em gel de poliacrilamida 12% (SDS-PAGE) para visualização do grau de pureza das proteínas purificadas (**Figura 26**).

Figura 25: A figura mostra o perfil cromatográfico de purificação da Hgal-4 humana inteira, CRD-I e CRD-II em coluna com resina α -Lactose-Agarose (Sigma-Aldrich).

Figura 26: Eletroforese em gel de poliacrilamida (SDS-PAGE 12%). Perfil das frações eluídas no processo de purificação da Hgal-4 completa e dos domínios CRD-I e CRD-II. Em a): perfil das frações eluídas da Hgal-4 completa em 47, 48, 49, 50 e 51 mL; b) perfil das frações eluídas do domínio CRD-I em 48, 49, 50, 51 e 52 mL; c) perfil das frações eluídas do CRD-II em 51, 52, 53, 54, 55 e 56 mL. As proteínas purificadas eluem na presença de 100mM de lactose.

4.4 Subclonagem dos CRD-I e CRD-II em vetor pTZ57R/T (Fermentas)

4.4.1 Amplificação dos CRD-I e CRD-II da galectina-4 humana

No presente estudo, os genes que codificam os domínios de reconhecimento de carboidratos CRD-I e CRD-II da galectina-4 humana foram amplificados por PCR (reação em cadeia da polimerase), utilizando como molde para a reação a construção onde o gene completo que codifica a galectina-4 humana se encontra clonado no vetor pQE (Qiagen). Foram inicialmente testadas três diferentes temperaturas de anelamento: 50°C, 55°C e 60°C (Figura 27). Das temperaturas testadas para o anelamento, concluímos que as reações de PCR são mais eficazes às temperaturas de 55°C para o CRD-I e CRD-II. Na verdade, para todas as temperaturas testadas o gene é amplificado com sucesso. No entanto, visualizamos também fragmentos de menor tamanho sendo amplificados, os quais acreditamos ser consequência do anelamento entre os “primers” utilizados.

Figura 27: Eletroforese em gel de agarose 0,8% (w/v) mostrando o perfil da amplificação dos genes de interesse através da técnica de PCR, para diferentes temperaturas de anelamento. As pistas 1, 2 e 3 pertencem ao CRD-I, as pistas 4, 5 e 6 ao CRD-II e as pistas 7, 8 e 9 pertencem a galectina-4 humana inteira. Os genes que codificam os domínios CRD-I e CRD-II e a galectina-4 humana possuem 450 bp, 435 bp e 969 pb respectivamente.

4.4.2 Processo de Adenilação dos CRDs da Galectina-4 humana

Os fragmentos amplificados por PCR e purificados a partir do gel de agarose de 8% (w/v) foram submetidos a uma reação de adenilação para subsequentemente serem utilizados na reação de ligação ao vetor de subclonagem pTZ57R/T (Fermentas). Após a reação de adenilação e posterior limpeza dos componentes da reação, os fragmentos foram visualizados e quantificados a partir do gel de agarose de agarose 8% (w/v) (Figura 28).

Figura 28: Eletroforese em gel de agarose 0,8% (w/v) para o processo quantificação dos fragmentos referentes aos domínios CRDI e CRDII. As pistas 1 e 2 referem-se ao CRD-I (450pb) e apresentam aproximadamente 25 ng/μL (1ª eluição) e 8,4ng (2ª eluição) respectivamente; as pistas 3 e 4 referem-se ao CRD-II e apresentam aproximadamente 9,6 ng/μL (1ª eluição) e 8,4ng (2ª eluição) respectivamente, quando comparadas ao padrão.

Os fragmentos adenilados e quantificados foram então submetidos a uma reação de ligação ao vetor de subclonagem pTZ57R/T(Fermentas). Como resultado do processo de ligação obteve-se as construções designadas por Hgal-4CRD-I(pTZ57R/T) e Hgal-4CRD-II(pTZ57R/T).

Amostras das construções Hgal-4CRD-I(pTZ57R/T) e Hgal-4CRD-II(pTZ57R/T) foram então transformadas em bactérias *Escherichia coli DH5α*. As colônias obtidas e selecionadas foram submetidas a uma reação de PCR de colônia para verificação da presença dos genes de interesse. A análise da presença genes de interesse em cada colônia selecionada

foi realizada através de uma eletroforese em gel de agarose 0,8% (w/v), (**Figura 29**). Para todas as colônias testadas, os fragmentos de interesse foram amplificados.

Figura 29: Eletroforese em gel de agarose 0,8% (w/v) para a reação de PCR de colônia. As pistas 1, 2, 3 e 4 referem-se ao domínio CRD-I e as pistas 5, 6, 7, 8 e 9 referem-se ao domínio CRD-II. Os clones obtidos foram designados de I-1, I-2, I-3 e I-4 para o CRD-I e II-1, II-2, II-3, II-4 e II-5 para o CRD-II.

Para a realização dos ensaios de clonagem dos domínios CRD-I e CRD-II em vetor de clonagem pET28a (Novagen), foram utilizados os clones Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II(pTZ57R/T)(II-2).

4.4.3 Processo de digestão das construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II (pTZ57R/T)(II-2) e do vetor pET-28a (Novagen)

As construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II (pTZ57R/T)(II-2) transformadas em em bactérias *Escherichia coli DH5α*, assim como amostras do vetor de clonagem pET28a (Novagen) foram submetidas a um processo de purificação utilizando-se com o Kit QIAprep® Miprep segundo o protocolo sugerido no manual do mesmo. Posterior ao processo de purificação as construções referentes aos domínios CRD-I e CRD-II e o vetor

pET28a (Novagen) foram submetidos à reação de digestão utilizando-se as enzimas de restrição *Not-I* (Fermentas) e *Nde-I* (Fermentas) (Figura 30).

Figura 30: Eletroforese em gel de agarose 0,8% (w/v) para a reação de digestão das construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II (pTZ57R/T)(II-2). As pistas 1, 2 e 3 referem-se as construções Hgal-4CRD-I(pTZ57R/T)(I-1), Hgal-4CRD-II (pTZ57R/T)(II-2) e ao vetor pET28a respectivamente.

Após o processo de digestão dos fragmentos referentes às construções Hgal-4CRD-I(pTZ57R/T)(I-1) e Hgal-4CRD-II(pTZ57R/T)(II-2), as amostras foram submetidas a um processo de purificação a partir do gel de agarose. As amostras referentes aos domínios CRD-I e CRD-II purificadas e quantificadas foram então utilizadas no processo de ligação no vetor de clonagem pET28-a (Novagen).

4.4.4 Processo de ligação dos domínios CRD-I e CRD-II no vetor pET-28a (Novagen).

Como resultado do processo de ligação dos domínios CRD-I e CRD-II no vetor pET28a, obteve-se as construções designadas por Hgal-4CRD-I(pET28a) e Hgal-4CRD-II(pET28a). Amostras das construções foram então transformadas em bactérias *Escherichia coli DH5α*. A análise da presença do gene de interesse nas colônias obtidas e selecionadas para o domínio CRD-I foi realizada através de uma reação de digestão utilizando-se as enzimas de restrição *NdeI* Fast digest (Fermentas) e *NotI* Fast digest (Fermentas), (Figura

31). Para a análise da presença do gene de interesse para o CRD-II as colônias obtidas e selecionadas, foram submetidas a uma reação de PCR de colônia (Figura 30). Os resultados obtidos foram visualizados através de uma eletroforese em gel de agarose 0,8%.

Figura 31: Eletroforese em gel de agarose 0,8% (w/v) para a reação de digestão para o domínio CRD-I. Observa-se através do gel a presença do gene de interesse na pista 2 (450 pb) . O clone obtido foi designado de CRD-I(I-1).

Figura 32: Eletroforese em gel de agarose 0,8% (w/v) para a reação de PCR de colônia para o domínio CRD-II. Observa-se através do gel de agarose a presença do gene de interesse nas pistas 1 e 7 (435 pb) . Os clones obtidos foram designados de CRD-II(II-17) e CRD-II(II-23)

4.4.5 Expressão e purificação dos domínios CRD-I e CRD-II em *E.coli* BL-21(DE3)

4.4.5.1 Expressão e purificação do domínio CRD-I da galectina-4 humana em *E.coli* BL-21(DE3)

Para os ensaios de expressão do CRD-I foi utilizada a construção designada como Hgal-4(CRD-I)(pET28a)(I-1). O domínio CRD-I da Hgal-4 foi expresso em bactérias *Escherichia coli*, utilizando a cepa BL21(DE3) (Novagen). A expressão do CRD-I da Hgal-4 foi induzida com 1,0 mM de IPTG a 30 °C a 250 rpm, overnight. A purificação do domínio CRD-I da galectina-4 foi realizada em uma única etapa que envolve cromatografia por afinidade em resina de níquel.

O processo de purificação utilizando cromatografia por afinidade com a resina Ni-NTA agarose é analisado por gel de poliacrilamida. A proteína elui na presença de 100 mM de Imidazol. As lavagens na presença de 0, 25, e 50, mM imidazol, foram eficientes para retirar os contaminantes (**Figuras 33a e 33b**).

Figura 33: Eletroforese em gel de poliacrilamida (12% SDS-PAGE) . A figura **a** mostra o padrão e perfil das lavagens na presença de 0, 25, 50, 100, 150, 200 e 250 mM Imidazol. Observa-se através do gel a eluição em 100 mM de imidazol do CRD-I. A figura **b** mostra o controle (pista 1) e as oito eluições com 100 mM imidazol. O domínio CRD-I migra com massa molecular de aproximadamente 17 kDa.

4.4.5.2 Expressão e purificação do domínio CRD-II da galectina-4 humana em *E.coli* BL-21(DE3)

Para os ensaios de expressão do domínio CRD-II foi utilizada a construção designada como Hgal-4(CRD-II)(pET28a)(II-23). O domínio CRD-II da Hgal-4 foi expresso em bactérias *Escherichia coli*, utilizando a cepa BL21(DE3) (Novagen). A expressão do CRD-II da Hgal-4 foi induzida com 1,0 mM de IPTG a 30 °C a 250 rpm, overnight. A purificação do domínio CRD-II da galectina-4 foi realizada em uma única etapa que envolve cromatografia por afinidade em resina de níquel.

Devido à presença de duas cisteínas supostamente livres, todos os passos da purificação do domínio CRD-II foram realizados na presença do agente redutor β -Mercaptoetanol a 14mM. A proteína é eluída na presença de 150mM de imidazol + β -Mercaptoetanol a 14mM. As lavagens na presença de 0, 25, e 100, mM imidazol + β -Mercaptoetanol a 14mM, foram eficientes para retirar os contaminantes. O processo de purificação utilizando cromatografia por afinidade com a resina Ni-NTA agarose é analisado por gel de poliacrilamida SDS-PAGE (**Figuras 34a e 34b**).

Figura 34: Eletroforese em Gel de poliacrilamida (12% SDS-PAGE). A figura **a** mostra o padrão e perfil das lavagens na presença de 0, 25, 50, 100, 150, 200 e 250 mM de Imidazol. Observa-se através do gel a eluição em 150 mM de imidazol e 14 mM β -Mercaptoetanol do domínio CRD-II. A figura **b** mostra o padrão e as nove

primeiras eluições com 150 mM imidazol e 14 mM β -Mercaptoetanol. O domínio CRD-II migra com massa molecular de aproximadamente 17 kDa.

4.5 *Cristalização dos domínios CRD-I e CRD-II da galectina-4 humana*

Os ensaios iniciais de cristalização dos domínios CRD-I e CRD-II da galectina-5 humana foram realizados com as proteínas purificadas através de cromatografia por afinidade (Ni-NTA). Para a realização dos ensaios de cristalização determinou-se previamente a concentração da proteína. Para a estimativa da concentração utilizou-se o método do coeficiente de extinção molar teórico, os valores teóricos estimados para os CRD-I e CRD-II foram de $24,41 \text{ mM}^{-1}\text{cm}^{-1}$ e $11,58 \text{ mM}^{-1}\text{cm}^{-1}$ respectivamente.

Inicialmente, os ensaios de cristalização do **CRD-I** foram realizados utilizando-se os kits de cristalização Crystal ScreenTM, Crystal Screen 2TM (Hampton Research) e PEG Ion (Hampton Research). Das condições presentes nos kits, testadas a 22°C, as condições **46** do Crystal ScreenTM (PEG 8K a 18 % v/v, 0,1M Cacodilato de Sódio pH 6,5 e 0,2M Acetado de Calcio) e **22** do Crystal ScreenTM (PEG 4K a 30%, 0,1M Tris HCl pH8,5 e 0,2M Acetado de Sódio) apresentaram a formação de vários cristais de pequena dimensão (**Figuras 35 e 36**).

Figura 35: Cristal do domínio CRD-I da galectina-4 humana obtido na condição 46 do Crystal ScreenTM (PEG 8K a 18 % v/v, 0,1M Cacodilato de Sódio pH 6,5 e 0,2M Acetado de Calcio). Tamanho aproximado de 0,1mm.

Figura 36: Cristal do domínio CRD-I da galectina-4 humana obtido na condição 22 do Crystal Screen™ (PEG 4K a 30%, 0,1M Tris HCl pH8,5 e 0,2M Acetado de Sódio). Tamanho aproximado de 0.15mm.

Após a realização de uma varredura completa em diferentes concentrações e peso molecular de PEG (Polietileno Glicol), concentração de Acetato de Cálcio, assim como, diferentes pHs, obtivemos uma melhora significativa na qualidade dos cristais. Experimentos de difração de raios-X mostraram uma boa qualidade do padrão de difração nos cristais obtidos na presença de PEG 8K, 0,1M Tris HCl pH 8,5 e 0,2M de acetato de cálcio que aparecem cerca de 4 dias após a realização dos experimentos, para a concentração de proteína em torno de 7,0 mg/mL (**Figura 37**).

Figura 37: Cristal do domínio CRD-I da galectina-4 humana obtido nas condições: PEG 8K, 0,1M Tris HCl pH 8,5, 0,2M de Acetato de Cálcio. **a:** cristais com 4 dias **b:** cristais com 10 dias e com tamanho aproximado de 0,3mm.

4.5.1 *Cristalização do domínio CRD-I na presença de Lactose.*

Ensaio de cristalização do domínio CRD-I na presença de lactose foram realizados com o objetivo de obter cristais do complexo proteína-ligante. Os ensaios de cristalização do domínio CRD-I na presença de 50 mM de lactose foram realizados utilizando-se as melhores condições obtidas na obtenção dos cristais “apo”. Os primeiros ensaios de cristalização do CRD-I na presença de lactose, apresentaram a formação de cristais na condição de PEG 8K a 8%, 0,1M Tris HCl pH 8,5, 0,2 M acetato de cálcio e volume de gota 4 μ L: 2 μ L (solução de proteína), 2 μ L (solução de cristalização) para a proteína a uma concentração em torno de 7.3 mg/mL (**Figura 38**).

Figura 38: Cristal do domínio CRD-I da galectina-4 obtido nas condições: PEG 8K, 0,1M Tris HCl pH 8,5, 0,2M de Acetato de Cálcio na presença de lactose. Cristais com 20 dias.

4.6 *Coleta, Processamento de Dados e Resolução da Estrutura*

Todas as coletas de dados foram realizadas em temperatura criogênica (100 K), deste modo, é necessário promover a proteção do cristal com uma solução crioprotetora. O agente crioprotetor utilizado foi o glicerol, em concentração final de 20 %. Os melhores resultados obtidos foram dois conjuntos de dados denominados: **CRD-I_t2g20** e **CRD-I_t2g23**, cuja resolução máxima foi de 2,20 e 2,00 Å, respectivamente.

O primeiro conjunto de dados processado foi o conjunto **CRD-I_t2g20**, onde o cristal foi obtido a 22°C na presença de 10 % PEG 8 K, 0,1 M Tris pH 8,5 e 0,2 M acetato de cálcio. Foram coletadas 71 imagens. A distância detetor-cristal foi de 100 mm e o cristal difratou com resolução máxima na borda de 2,2 Å para uma exposição de 45 segundos e oscilação de 1° por imagem. No primeiro momento, o conjunto de dados foi indexado em um grupo espacial de baixa simetria, trigonal P3. A análise das ausências sistemáticas mostrou

claramente que existem condições de difração $00l = 6n$, o que caracteriza a presença de um eixo helicoidal de ordem seis (6_1 ou 6_5) ao longo do eixo “c”. No segundo momento, o conjunto de dados foi processado no grupo espacial hexagonal ($P6_1$) e em um grupo espacial hexagonal de maior simetria ($P6_122$). A análise dos valores de R_{merge} obtidos para os dois processamentos indicou que o grupo espacial correto se tratava do de maior simetria, pois os valores de R_{merge} obtidos entre os dois processamentos (15 % em $P6_1$ e 16 % em $P6_122$) eram praticamente iguais. O número de moléculas na unidade assimétrica foi determinado após o cálculo do coeficiente de Mathews. Análise do conteúdo de solvente, para os grupos espaciais $P6_122$ e $P6_522$, de 47,06 % com coeficiente de Mathews de $2,32 \text{ \AA}^3\text{Da}^{-1}$ prevê um monômero na unidade assimétrica.

A resolução da estrutura foi realizada por substituição molecular utilizando como modelo a estrutura do domínio N-terminal da galectina-4 de rato, disponível no *Protein Data Bank* sob o código de acesso **2DYC**, que apresenta 81% de identidade sequencial com o domínio N-terminal da proteína galectina-4 humana. O problema de ambigüidade no grupo espacial ($P6_122$ e seu enantiomórfo $P6_522$) foi resolvido a partir da determinação da estrutura por substituição molecular utilizando os dados processados no grupo espacial $P6_122$. Os dados relacionados a coleta e processamento dos dados estão sumarizados na tabela 27.

O segundo e melhor conjunto de dados processado foi o conjunto **CRD-I_t2g23**, onde o cristal foi obtido a 22 °C na presença de 16 % PEG 8 K, 0,1 M Tris pH 8,5 e 0,2 M acetato de cálcio. Foram coletadas 38 imagens (**Figura 39**). A distância detetor-cristal foi de 90 mm e o cristal difratou com resolução máxima na borda de $2,0 \text{ \AA}$ para uma exposição de 60 segundos e oscilação de 1° por imagem. Análise do conteúdo de solvente de 47,51 % com coeficiente de Mathews de $2,34 \text{ \AA}^3\text{Da}^{-1}$ preve um monômero na unidade assimétrica. A resolução da estrutura foi realizada por substituição molecular utilizando, como modelo, a estrutura obtida com o conjunto de dados CRD-I_t2g20. Os dados relacionados à coleta e processamento dos dados estão sumarizados na tabela 27.

Figura 39: Exemplo de uma imagem do padrão de difração do cristal CRD-I_t2g23 coletada no LNLS.

Tabela 27: Dados cristalográficos e estatísticos para os diferentes conjuntos de dados coletados no LNLS ($\lambda=1,437 \text{ \AA}$), a 100 K. Os valores relativos à última camada de resolução estão mostrados entre parênteses, quando apropriado.

Dados e Parâmetros	CRD-I_t2g20	CRD-I_t2g23
Sistema Cristalino	Hexagonal	Hexagonal
Grupo especial	P6 ₁ 22	P6 ₁ 22
Parâmetros de rede (Å)	a=b=71,01, c=108,47	a=b=71,25, c=108,66
Mosaicidade (°)	0,25	0,21
Intervalo de Resolução (Å)	40,69 – 2,50	40,79 – 2,20
Moléculas por unidade assimétrica	1	1
V_m (Coeficiente de Matthews)	2,32 Å ³ Da ⁻¹	2,34 Å ³ Da ⁻¹
Solvente (%)	47,06	47,51
Completeza (%)	100 (100)	100 (100)
I / sigma(I)	14,1 (3,1)	10,1 (2,5)
^a R _{sym} (%)	12,9 (67,4)	11,8 (59,0)
Nº reflexões únicas	6 072 (854)	8 809 (1 250)
Multiplicidade	8,0 (8,1)	4,3 (4,4)
Substituição Molecular		
R _{factor}	54,6 %	38,1 %
Correlação	45,1 %	73,8 %

^a $R_{sym} = \sum |I_{obs} - \langle I \rangle| / \sum I_{obs}$ onde $\langle I \rangle$ é a média das intensidades calculada a partir das observações para as reflexões relacionadas por simetria.

4.7 Refinamento

Após a etapa de substituição molecular realizada com o conjunto de dados CRD-I_t2g20, o modelo inicialmente obtido foi submetido a dez ciclos de refinamento de corpo rígido realizado no programa *REFMAC* (MURSHUDOV *et al.*, 1999). Os valores de R_{factor} e R_{free} de 50,5 e 51,2 %, respectivamente, foram obtidos para o modelo inicial. As mutações correspondentes a sequência da proteína de interesse foram introduzidas no programa *Coot* (EMSLEY; COWTAN, 2004). Este novo modelo foi submetido a trinta ciclos de refinamento posicional no programa *REFMAC* (MURSHUDOV *et al.*, 1999), resultando em um modelo com valores de R_{factor} e R_{free} de 28,3 e 38,0 %, respectivamente. Esta estrutura foi utilizada como modelo inicial na substituição molecular realizada com os dados a mais alta resolução, conjunto de dados CRD-I_t2g23. Os cinco primeiros ciclos de refinamento de corpo rígido, realizado no programa *REFMAC*, resultaram em um modelo com valores de R_{factor} e R_{free} de 31,1 e 32,8 %, respectivamente. Várias moléculas de água foram adicionadas manualmente e através da utilização do programa *Coot*. Após inúmeros ciclos de refinamento posicional realizados no programa *REFMAC*, intercalados com adição de solvente e manipulação manual dos resíduos no programa *Coot*, o modelo final apresenta os valores de R_{factor} e R_{free} de 19,97 e 25,53 %, respectivamente. A figura 40 abaixo mostra a boa qualidade do mapa de densidade eletrônica obtida para o modelo final.

Figura 40: A figura acima demonstra a qualidade do mapa de densidade eletrônica $2F_o - F_c$, com σ igual 1.0, representado em azul.

4.8 Avaliação dos dados estereoquímicos

As ligações N-C α e C α -C podem sofrer rotações em ângulos torsionais denominados φ (*phi*) e ψ (*psi*), respectivamente. O gráfico de Ramachandram mostra a distribuição dos ângulos torsionais associados a cada aminoácido (RAMACHANDRAM *et al.*, 1963). O diagrama de Ramachandram (**Figura 41**), da estrutura final do domínio CRD-I da galectina-4 humana, mostra que 87,6 % dos resíduos encontram-se nas regiões mais favoráveis, 12,4 % na região adicional permitida, e nenhum resíduo encontra-se na região generosamente permitida ou não permitida. Os valores dos desvios médio quadrático (rmsd) em relação aos valores ideais obtidos para o modelo final foi de 0,012 Å para distância de ligações e 1,4° para ângulos de ligação.

Figura 41: Gráfico Ramachandram da estrutura final do domínio CRD-I da galectina-4 humana.

4.9 Estrutura cristalográfica do domínio CRD-I da galectina-4 humana

A estrutura cristalográfica do domínio CRD-I da galectina-4 humana foi determinada a 2.0 Å de resolução. A estrutura cristalina mostra que o CRD-I compreende duas folhas compostas de seis fitas (S1- S6) e cinco fitas (F1-F6) cada (**Figura 42**).

Figura 42: Estrutura cristalográfica do CRD-I da galectina-4 humana. A folha β em verde é composta de 6 fitas (S1- S6) e a folha β em azul é composta de cinco fitas (F1-F6), conectadas por uma curta hélice, que se empacotam em um domínio do tipo β -sanduiche.

A estrutura global do domínio CRD-I da galectina-4 humana é bastante similar às estruturas de galectinas já descritas, embora a identidade sequencial não exceda 40% (**Figura 43**). Por exemplo, uma superposição estrutural utilizando as coordenadas do C α entre diferentes membros da classe das galectinas mostra que o CRD-I da Hgal-4 compartilha um r.m.s.d de 0.81 Å com CRD-I da galectina-4 de camundongo (código PDB: 2DYC) (KREJCIRIKOVA, 2008), um r.m.s.d de 1.41 Å com CRD-I da galectina-9 humana (código PDB: 2EAK) (NAGAE *et al.*; 2009), um r.m.s.d de 0.96 Å com o CRD-I da galectina-9 de camundongo (código PDB: 2D6K) (NAGAE, *et al.*; 2006) e um r.m.s.d de 1.41 com o CRD-II da galectina-4 humana (código PDB: 1X50) (TOMIZAWA *et al.*, 2005. As maiores

diferenças são observadas nas regiões de loops e turns que conectam as fitas (**Figura 44**) que predomina quando comparamos os diferentes domínios da Hgal-4 (CRD-I x CRD-II) e quando comparamos CRD's de diferentes espécies (humana x camundongo).

Figura 43: Alinhamento sequencial entre alguns CRD's das galectinas do grupo *tandem-repeat*: HGAL-9_CRDI: CRD-I da galectina-9 humana; MGAL-9_CRDI: CRD-I da galectina-9 de camundongo; MGAL-4_CRDI: CRD-I da galectina-4 de camundongo, HGAL-4_CRDI: CRD-I da galectina-4 humana e HGAL-4_CRDII: CRD-II da galectina-4 humana.

Figura 44: Superposição estrutural entre a estrutura tridimensional do domínio CRD-I da galectina-4 humana com as estruturas dos domínios CRD-I da galectina-4 de camundongo e da galectina-9 humana, e do domínio CRD-II da galectina-4 humana. As maiores diferenças observadas são encontradas nas regiões de voltas que conectam as fitas.

O sítio de ligação do carboidrato, assim como observado para outras galectinas, é formado pelas fitas S4, S5 e S6, sugerindo que o mecanismo de reconhecimento das moléculas de carboidrato utilizado para o domínio CRD-I é similar às observadas para outras galectinas (**Figura 45**). Importantes estudos estruturais combinados com estudos de mutação sítio-dirigida para a galectina-9 humana demonstraram que resíduos de aminoácidos não conservados nesta superfície côncava são importantes para a determinação da especificidade da proteína por alvos específicos (NAGAE *et al.*, 2008).

Figura 45: Estrutura cristalográfica do CRD-I da galectina-4 humana, indicando os resíduos envolvidos na coordenação com a molécula de carboidrato.

Em nossos estudos, a estrutura do CRD-I foi encontrada na forma monomérica na unidade assimétrica. Estudos realizados para outros CRD's de galectinas do grupo *tandem-repeat* sugerem diferentes níveis de oligomerização. Por exemplo, a estrutura do CRD-I de galectina-9 de camundongo foi determinada na forma dimérica para os diferentes complexos estudados (NAGAE *et al.*, 2006). Já a estrutura do CRD-I de galectina-9 humana foi determinada em diferentes formas cristalinas e se apresentam tanto na forma monomérica,

dimérica, trimérica e tetramérica (NAGAE, 2006; NAGAE, 2008). As diferentes formas oligoméricas para os CRD's podem ser consequência do empacotamento cristalino, ou mesmo possam aparecer em solução, em sistemas onde os CRD's se encontram isolados. No entanto, existem sugestões que as galectinas da classe *tandem-repeat* podem ser encontradas na forma de oligômeros.

Para determinar o grau de oligomerização do domínio CRD-I da galectina-4 humana, analisamos o empacotamento cristalino e o perfil de migração em coluna de exclusão molecular. O grupo espacial identificado por esta forma cristalina P6₁22, permite apenas a formação de um dímero através de um eixo de ordem 2 cristalográfico, que foi comparado com o dímero encontrada na estrutura do domínio CRD-I da galectina-9 de camundongo (NAGAE, 2006) e da galectina-1 humana (LOPES-LUCENDO *et al.*, 2004) (**Figura 46**).

Figura 46: Superposição estrutural entre os dímeros descritos para as estruturas da galectina-1 humana (LOPES-LUCENDO *et al.*, 2004) e da galectina-9 de camundongo (NAGAE, 2006).

A diferença encontrada para as formas diméricas descritas para as galectinas sugere que se a forma dimérica é funcional, esta não segue padrão característico nas galectinas. Para então analisarmos o comportamento do nosso domínio de interesse, a proteína foi submetida a uma coluna de filtração em gel (Sephacryl S100) previamente padronizada (**Figuras 47 e 48**)

Figura 47: Perfil de eluição cromatográfico do domínio CRD-I da galectina-4 em coluna de exclusão molecular SEPHACRYL S100 (Ge Healthcare).

Figura 48: Perfil de eluição cromatográfico para padrões de massa molecular e a correspondente curva de calibração da coluna de exclusão molecular SEPHACRYL S100 (Ge Healthcare). Dados gentilmente cedidos por Ricardo Augusto Pereira de Pádua.

A estimativa da massa molecular para o domínio CRD-I através do uso da curva de calibração é de aproximadamente 2,5 kDa, incompatível com qualquer forma oligomérica esperada uma vez que o monômero possui massa molecular estimada em aproximadamente 17 kDa. Acreditamos que, embora o perfil de migração do CRD-I sugere que a proteína

produzida em nossos experimentos esteja essencialmente pura e homogênea na sua forma oligomérica, a proteína deve ter interagido com a coluna de exclusão molecular prejudicando a sua caracterização em termos da estrutura quaternária. Novos experimentos de calibração e corrida serão realizados na presença de lactose para testar nossa hipótese.

5.DISSCUSSÕES E CONCLUSÕES

As galectinas correspondem a uma classe importante de proteínas, envolvidas na regulação de uma grande variedade de funções celulares. Este trabalho focou o início da caracterização estrutural dos domínios de reconhecimento de carboidrato da galectina-4 humana, denominados CRD-I e CRD-II.

O desenvolvimento do presente projeto deu início com a clonagem dos fragmentos que codificam os domínios CRD-I e CRD-II da galectina-4 humana no vetor pET29a (Novagen). Como uma estratégia para avaliar o comportamento da expressão protéica foram realizados inúmeros ensaios de expressão onde foram testadas diferentes concentrações de IPTG (0.5, 1.0 e 1.5 mM) e diferentes temperaturas (18°C e 37°C). As análises do perfil de expressão protéica indicaram que o nível de expressão, se existisse, era muito baixo. Sendo assim, considerar a hipótese de que os domínios estavam sendo expressos em baixa concentração, tornando a visualização em SDS-PAGE inviável, decidimos utilizar a técnica de cromatografia por afinidade por lactose como estratégia para isolar e concentrar os domínios de interesse. Como mostrado na **figura 25**, o perfil cromatográfico sugere o sucesso do processo de purificação. No entanto, os resultados da eletroforese SDS-PAGE, surpreendentemente indicam que as frações isoladas apresentam massa molecular entre 30 kDa e 45 kDa, num perfil de migração eletroforética muito similar para os CRD-I e CRD-II e para a proteína Hgal-4 inteira, utilizada como controle (**Figura 26**).

Várias hipóteses para explicar nossos resultados foram levantadas. Inicialmente, consideramos que as galectinas possuem como uma característica importante a presença de cisteínas livres e a oxidação das mesmas leva a uma mudança conformacional que altera a capacidade destas proteínas em se ligar a moléculas de carboidratos (TRACEY *et al.*, 1992).

Além disso, a presença destas cisteínas livres poderia induzir agregação através da formação de pontes dissulfeto intermoleculares. Portanto, nossa hipótese inicial para explicar o perfil de migração eletroforética com massa molecular aparentando o dobro do esperado seria a possível formação de pontes dissulfeto entre duas moléculas de proteína. Uma vez que a massa molecular calculada para os domínios CRD-I e CRD-II é de 17 kDa, um dímero de massa molecular de aproximadamente 34 kDa, seria compatível com o perfil de migração eletroforética.

No entanto, embora esta hipótese explique o comportamento do CRD-II, que possui dois resíduos de cisteínas, não explica o mesmo comportamento observado para o CRD-I, já que

este domínio de reconhecimento de carboidrato não possui nenhuma cisteína na sua sequência e portanto não poderia formar uma estrutura dimérica. Em relação aos resultados apresentados nos ensaios elaborados com a Hgal-4 inteira, a análise do perfil de purificação (**Figura 25**), assim como, da migração eletroforética (**Figura 26**), fornece um dado relevante a ser considerado. Quando comparamos os resultados obtidos no processo de purificação da proteína inteira com os resultados obtidos para CRD-I e CRD-II, observamos o mesmo comportamento. A proteína Hgal-4, elui na mesma posição que os domínios independentes (**Figura 26**), sugerindo assim possuir uma mesma afinidade por lactose. Uma possível explicação para estes resultados seria hipotetizar que na verdade, a Hgal-4 apresenta uma baixa afinidade por lactose, pelo menos nas condições utilizadas em nossos estudos bioquímicos, e que na verdade, a proteína purificada através da cromatografia por afinidade poderia ser uma proteína expressa pela bactéria *E. coli* utilizada para os experimentos de expressão heteróloga. Experimentos de purificação utilizando o extrato da bactéria serão realizados para testar esta hipótese.

Um estudo minucioso dos trabalhos de expressão heteróloga de galectinas sugere que o nível de expressão da Hgal-4 em *Escherichia coli* é baixo (IDEO *et al.*, 2002) e alguns artifícios, como a utilização da cepa M15 (Qiagen), têm sido extensivamente utilizados para melhorar o nível de expressão protéica (MARKOVA *et al.*, 2006; KREJCIRIKOVA *et al.*, 2008). Além disso, dando apoio à nossa suposta hipótese, a qual sugere que a afinidade da Hgal-4 por lactose não é suficiente para isolar a proteína inteira, assim como, os CRD-I e CRD-II, vários trabalhos envolvendo a galectina-4, humana ou de outras espécies, utilizam-se do artifício da presença da cauda de histidina para facilitar os experimentos de purificação (IDEO *et al.*, 2002; LEONIDAS *et al.*, 1998) foram observados. Desta forma, decidimos modificar a construção utilizada neste projeto para tentarmos melhorar o nível de expressão protéica, assim como, utilizar outra estratégia cromatográfica para o isolamento dos domínios CRD-I e CRD-II em quantidades suficientes para os nossos estudos de caracterização estrutural.

Para a clonagem dos fragmentos que codificam os domínios CRD-I e CRD-II da galectina-4 humana no vetor pET28a (Novagen) utilizamos uma diferente estratégia.

A fim de maximizar o processo de digestão, decidimos inicialmente utilizar o vetor de clonagem pTZ57R/T (Fermentas). O fragmento foi clonado com sucesso conforme identificado por técnicas de PCR de colônia (**Figuras 29, 31 e 32**), o processo de digestão eficiente (**Figura 30**), e o produto de ligação entre o fragmento digerido e o vetor, transformado com sucesso em BL21(DE3) para os ensaios de expressão.

Ao contrário da estratégia utilizada inicialmente para as construções em pET29a, decidimos por, inicialmente realizar um único ensaio de expressão em larga escala (1L) onde a expressão da proteína foi induzida a 1 mM de IPTG a 30°C. Estas condições foram eficientes para a expressão dos domínios na forma solúvel, que posteriormente foram isolados por técnicas de cromatografia por afinidade em coluna de Níquel. A metodologia empregada para a purificação de ambos domínios foi essencialmente a mesma, onde a utilização de um gradiente de concentração de imidazol foi eficiente para isolar as proteínas de interesse com excelente grau de pureza (**Figuras 33 e 34**). A única diferença entre as metodologias foi a utilização do agente redutor β -mercaptoetanol para a purificação do domínio CRD-II, devido a presença de cisteínas livres neste fragmento da Hgal-4. O rendimento obtido foi de aproximadamente 2 mg de proteína para cada litro de cultura para ambos os domínios.

A purificação dos domínios CRDs da galectina-4 humana de forma independente permitiu a realização dos experimentos de cristalização. O domínio CRD-I na forma apo foi cristalizado com sucesso e os dados de difração de raios-X permitiram a resolução da estrutura por técnicas de substituição molecular a resolução máxima de 2,0 Å. Recentemente, obtivemos cristais do domínio CRD-II em complexo com lactose, no entanto, ainda não houve tempo de feixe disponível no Laboratório Nacional de Luz Síncrotron para a determinação da estrutura do complexo.

Para o domínio CRD-II, embora vários experimentos de cristalização foram realizados, até o presente momento não foi possível observar a presença de nenhuma condição promissora de cristalização.

O modelo cristalográfico obtido, embora de média resolução, permitiu a realização de estudos comparativos preliminares entre o domínio CRD-I da galectina-4 humana e membros das galectinas, onde foi observada uma grande similaridade estrutural que reforça a hipótese de que as diferenças de especificidade observadas para as moléculas de carboidrato são consequência de substituições pontuais que podem alterar a distribuição estérica e de carga, ainda que sutilmente.

Os resultados, aqui apresentados, abrem oportunidades para a realização de novos estudos com diferentes abordagens. A produção heteróloga dos domínios CRD-I e CRD-II da galectina-4 humana poderão ser utilizados para a produção de anticorpos mono e policlonais, como uma ferramenta na detecção de níveis diferenciados de expressão da galectina-4 humana. Este trabalho será realizado em colaboração com o Prof. Dr. Marcelo D. Baruffi, do Departamento de Análises Clínicas e Bromatológicas da Faculdade de Ciências Farmacêuticas de Ribeirão Preto. Além disso, experimentos para a completa caracterização biofísica, como dicroísmo circular, calorimetria já estão em andamento através da colaboração com o Prof. Dr. Antonio José da Costa Filho do Instituto de Física de São Carlos.

Novos estudos estruturais que visam a determinação da estrutura do CRD-I com diferentes ligantes serão realizados. Os cristais na presença de lactose foram recentemente obtidos e serão submetidos a coleta de dados no próximo turno disponibilizado pelo Laboratório Nacional de Luz Síncrotron. Pretendemos também realizar novos ensaios de cristalização para o domínio CRD-II, na tentativa de determinarmos a estrutura cristalográfica deste domínio. Isto porque, embora a estrutura do domínio CRD-II já tenha sido determinada por técnicas de ressonância magnética nuclear, isto nos permitirá a realização de análises comparativas detalhadas entre os dois domínios, a cristalização deste domínio constituirá uma ferramenta importante para a determinação de estruturas de diferentes complexos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, L.; SCOTT, G.K.; WEINBERG, C.S. Biphasic modulation of cell growth by recombinant human galectin-1. **Biochimica and Biophysical Acta**, v.1312, p.137-144, 1996.

ALMKVIST, J.; KARLSSON, A. Galectins as inflammatory mediators. **Glycoconjugate Journal**, v.19, p.575-581, 2004.

ALMKVIST, J.; FALDT, J.; DAHLGREN, C.; LEFFLER, H.; KARLSSON, A. Lipopolysaccharide-induced gelatinase granule mobilization primes neutrophils for activation by galectin-3 and formymethionyl-Leu-Phe. **Infection and Immunity**, v.69, p.832-837, 2001.

ALMKVIST, J.; DAHLGREN, C.; LEFFLER, H.; KARLSSON, A. Activation of the neutrophil nicotinamide dinucleotide phosphate oxidase by galectin1-1. **Journal of Immunology**, v.168(8), p.4034-41, 2002 .

ARATA, Y.; HIRABAYASHI, J.; KASAI, K. Structure of the 32-kDa galectin gene of the nematode *Caenorhabditis elegans*. **The Journal of Biological Chemistry**, v.272, n.42, 1997.

BARONDES, S.H.; COOPER, D.W.; GITT, M.A.; LEFFLER, H. Structure and function of a large family of animal lectins. **The Journal of Biological Chemistry**, v.269, p.20807-20810, 1994.

BIANCHET, M.A.; AHMED, H.; VASTA, G.R.; AMZEL, L.M. Soluble beta-galactosyl-binding lectin (galectin) from toad ovary: crystallographic studies of two protein-sugar complexes. **Proteins**, v. 40, p. 378-80, 2000.

BIDON-WAGNER, N.; LE PENNEC, J. Intracellular sorting of galectin-8 based on carbohydrate fine specificity. **Glycobiology**, v.19, p.557-563, 2004.

BOURNE, Y.; BOLGIANO, B.; LIAO, D.I.; STRECKER, G., CANTAU, P.; HERZBERG, O.; FEIZI, T.; CAMBILLAU, C.; Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. **Nature Structure Biology**, v.1, p.863-870, 1994.

BRUNGER, A.T.; ADAMS, P.D.; CLORE, G.M.; DELANO, W.L.; GROS, P.; GROSSE-KUNSTLEVE, R.W.; JIANG, J.S.; KUSZEWSKI, J.; NILGES, M.; PANNU, N.S.; READ, R.J.; RICE, L.M.; SIMONSON, T.; WARREN, G.L. Crystallography and NMR System: A new software suite for macromolecular structure determination. **Acta Crystallographica Section D**, v.54, p.905-921, 1998.

COOPER D.N.W.; BARONDES, S.H. Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. **Journal of Cell Science**, v.110, p.1681-1691, 1990.

COOPER, D.N.W.; BARONDES, S.H. God must love galectins; He made so many of them. **Glycobiology**, v.9, p.979-984, 1999.

COOPER, D.N.W.; Galectinomics: finding themes in complexity. **Biochimica et Biophysica Acta**, v.1572, p.209-231, 2002.

DAHMS, N.M.; HANCOCK, M.K. P-type lectins. P-type. **Biochimica et Biophysica Acta**, v. 1572, p.317-340, 2002.

DUMIC, J.; DABELIC, S. FLOGEL, M. Galectin-3: an open-ended story. **Biochimica et Biophysica Acta**, v.1760, n.4, p.616-635, 2006.

DRICKAMER, K.; TAYLOR, M.E. Biology of animal lectins. **Annual Reviews Cell Biology**, v.9, p.237-264, 1993.

ELOLA, M.T.; CHIESA, M.E., ;ALBERTI, A.F.; MORDOH, J.; FINK, N.E. Galectin-1 receptors in different cell types. **Journal of Biomedical Science**, v.12, p.13-29, 2005.

EMSLEY, J.; WHITE, H.E.; O'HARA, B.P.; OLIVIA, G.; SRINIVASAN, N.; TICKLE, I.J.; BLUNDELL, T.L.; PEPYS, M.B.; WOOD, S.P. Structure of pentameric human serum amyloid P component. **Nature**, v.367, p.338-345, 1994.

EMSLEY, P; COWTAN, K. Coot: Model-Building Tools for Molecular Graphics. **Acta Crystallographica Section D**, v.60, p.2126-2132, 2004.

EVANS, P.R. **Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography**, v.33, p.22-24, 1997.

GALVAN, M.; TSUBOI, S.; FUKUDA, M. ;BAUM, L.G.; Expression of a specific glycosyltransferase enzyme regulates T cell death mediated by galectin-1. **The Journal of Biological Chemistry**, v.275, p.16730-16737, 2000.

GAUDIN, J.C.; MONSIGNV, M.; LEGRAND, A. Cloning of the cDNA encoding rabbit galectin-3. [Gene](#), v.163, p.249-252, 1995.

GEWURZ, H.; ZHANG, X.; LINT, T.F. Structure and function of the pentraxins. **Current Opinion Immunology**, Chicago, v.7, p.54-64, 1995.

GITT, M.A.; MASSA, S.M.; BARONDES, S.H. Isolation and expression of a gene encoding L-14-II, a new human soluble lactose binding lectin. **The Journal Biological Chemistry**, v.267, p.10601-06, 1992.

GITT, M.A.; WISER, M.F.; LEFFLER, H.; HERMANN, J.; XIA, Y.R.; MASSA, S.M.; COOPER, D.N.W.; LUSIS, A.J.; BARONDES, S.H. Sequence and mapping of galectin-5, a beta-galactoside-binding lectin, found in raat erythrocytes. **The Journal of Biological Chemistry**, v.270, p.5032-38, 1995.

GRASSADONIA, A.; TINARI, N.; IURISCI, I.; PICCOLO, E.; CUMASCHI, A.; INNOMINATO, P.; D'EGIDIO, M.; NATOLI, C.; ICCOBELLI, S. Galectins: matricellular glycan-binding proteins linking cell adhesion, migration and survival. **Glycoconjugate Journal**, v.19, p.551-556, 2004.

GOUET, P.; COURCELLE, E.; STUART, D.I.; METOZ, F. ESPript:multiple sequence alignments in PostScript. **Bioinformatics.**, v.15, p.305-308, 1999.

HANSEN, G.H.; IMMERDAL, L.; THORSEN, E.; NIELS-CHRISTIANSEN, L.L.; NYSTROM, B.T.; DEMANT, E.J.F.; DANIELSEN, E.M. **Lipid Rafts Exist as Stable Cholesterol-independent Microdomains in the Brush Border Membrane of Enterocytes.** **Journal Biology Chemistry**, v.276, p.32338-32344, 2001.

HERRMANN, J.; TURCK, C.W.; ATCHISON, R.E.; POULTER, L.; GITT, M.A.; BURLINGAME, A.L.; BARONDES, S.H.; LEFFLER, H. Primary structure of the soluble lactose binding lectin L-29 from rat and dog and interaction of its non-collagenous proline-, glycine-, tyrosine-rich sequence with bacterial and tissue collagenase. **The Journal of Biological Chemistry**, v.268, p.6704-11, 1993.

HIPPO, Y.; YASHIRO, M.; ISHII, M.; TANIGUCHI, H.; TSUTSUMI, S.; HIRAKAWA, K.; KODAMA, T.; ABURATANI, H. Differential gene expression profiles of scirrhous gastric cancer cells with high metastatic potential to peritoneum or lymph nodes. **Cancer Research**, v.61, p.889-895, 2001.

HIRABAYASHI, J.; KASAI, K. Effect of amino acid substitution by sited-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa beta-galactoside-binding lectin. **Journal Biology Chemistry**, v.3, p.23648-53, 1991.

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

HIRABAYASHI, H.; KAWASAKI, K.; SUZUKI, K.; KASAI, K. Complete amino acid sequence of 14 kDa **β -Galactoside-Binding Lectin of Chick Embryo**. **Journal Biochemistry**, v. 101, p.775-778, 1987.

HIRASHIMA, M.; KASHIO, Y.; NISHI, N.; YAMAUCHI, A.; IMAIZUMI, T.A.; KAGESHITA, T.;SAITA, N.; NAKAMURA, T. Galectin-9 in physiological and pathological conditions. **Glycoconjugate Journal**, v.19, p.593-600, 2004.

HIRASHIMA, J. Lectin-based structural glycomics: Glycoproteomics and glycan profiling. **Glycoconjugate Journal**, v.21, p.34-40, 2003.

HIRABAYASHI, Y.; KIMURA, M.; MATSUMOTO, M. YAMAMOTO, K.; KADOWAKI, S.; TOCIKURA, T. **A Novel Glycosphingolipid Hydrolyzing Enzyme, Glycosphingolipid Ceramide Deacylase, Which Cleaves the Linkage between the Fatty Acid and Sphingosine Base in Glycosphingolipids** . **Journal Biochemistry**, v.103, p.1-4, 1988.

HUFLEJT, M.E.; JORDAN, E.T.; GITT, M.A.; BARONDES, S.H.; LEFFLER, H. Strikingly different localization of Galectin-3 and Galectin-4 in human colon adenocarcinoma T84 cells. **The Journal of Biological Chemistry**, v.272, p.14294-03, 1997.

HUFLEJT, M.J.; MOSSINE, V.V.; NAIDENKI, O.; JAZAYERI, M.; ROGERS, P.; TINARI, N.; IACOBELLI, S.M.E.; LUSTAGARTEN, J.M.C. 24th Annual San Antonio Breast Câncer Symposium abstract. **Glycoconjugate**, v.19, p.621-629, 2004.

HUGHES, R.C. Galectins in kidney development. **Glycoconjugate Journal**, v.19, p.621-629, 2004.

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.. **Journal of Biological Chemistry**, v.280, p.4730-37, 2005.

KASAI, K.; HIRABAYASHI, J. Galectins: A family of animal lectins that decipher glycocodes. **The Journal of Biochemistry**, v.119, p.1-18, 1996.

KREJCIRIKOVA, V.; FABRY, M.; MARKOVA, V.; MALY, P.; REZACOVA, P.; BRYNDA, J. Crystallization and preliminary X-ray diffraction analysis of mouse galectin-4 N-terminal carbohydrate recognition domain in complex with lactose. **Acta Crystallographica**, v.64, p.665-667, 2008.

KONDOH, N.; WAKATSUKI, T.; RYO, A.; HADA, A.; AIHARA, T.; HORIUCHI, S.; GOSEKI, N.; MATSUBARA, O.; TAKENAKA, K.; SHICHITA, M.; TANAKA, K.; SHUDA, M.; YAMAMOTO, M. **Identification and Characterization of Genes Associated with Human Hepatocellular Carcinogenesis. Cancer Research**, v.59, p. 4990-96, 1999.

KOPITZ, J.; ANDRE, S.; REITZENSTEIN, V.; VERSLUIS, K.; KALTNER, H.; PIETERS, R.J.; WASANO, K.; KUWABARA, I.; LIU, F.T.; CANTZ, M.; HECK, A.J.; GABIUS, H.J. A guide to signaling pathways connecting protein-glycan interaction. **Oncogene**, v.22, p.6277-88, 2003.

KOPTIZ, J.; VON REITZENSTEIN, C.; ANDRÉ, S.; KALTNER, H.; UHL, J.; EHEMANN, V.; CANTZ, M.; GABIUS, H.J. Negative regulation of neuroblastoma cell growth by carbohydrate-dependent surface binding of galectin-1 and functional divergence from galectin-3. **The Journal of Biological Chemistry**, v.276, p.35917-23, 2001.

LAHM, H.; ANDRE, S.; HOEFELICH, A.; FISCHER, J.R.; SORDAT, B.; KALTNER, H.; WOLF, E.; GABIUS, H.J. Changes in galectin-7 and cytokeratin-19 expression during the progression of malignancy in thyroid tumors: Diagnostic and biological implications. **Cancer Research Clinical Oncology**, v.127, p.375-386, 2001.

LASKOWSKI, R.A.; MACARTHUR, M.W.; THORNTON, J.M. Validation of protein models derived from experiment. **Current Opinion in Structural Biology**, v.8, p.631-639, 1998.

LEFFLER, H.; CARLSSON, S.; HEDLUND, M.; QIAN, Y.; POIRIER, F. Introduction to galectins. **Glycobiology**, v.19, p.433-440, 2004.

LEONIDAS, D.D.; VATZAKI, E.H.; VORUM, H.; CELIS, J.E.; MADSEN, P.; ACHARYA, K.R. **Structural Basis for the Recognition of Carbohydrates by Human Galectin-7. Biochemistry**, v.37, p.13930-40, 1998.

LESLIE, A.G.W. Recent changes to MOSFLM package for processing film and image plate data. **Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography**, v.26, 1992.

LEVI, G.; TEICHBERG, V.I. Isolation and physicochemical characterization of Eletrolectin, a β -D-galactoside binding lectin from the electric organ of *Electrophorus Electricus*. **The Journal of Biological Chemistry**, v.256, p.5735-40, 1981.

LIU F.T.; PATTERSON, R.J.; WANG, J.L. Intracellular functions of galectins. **Biochimistry Biophysics Acta**, v.1572, p.263-273, 2002.

LIU, F.T.; RABINOVICH, G.A. Galectins as novel modulators of tumor progression. **Nature Reviews Cancer**, v.5, p.29-41, 2005.

LOBSANOV, Y.D.; GITT, M.A.; LEFFLER, H.; BARONDES, S.H.; RINI, J.M. **X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9Å resolution. Journal of Biological Chemistry**, v.268, p.27034-38, 2003.

LOPES-LUCENDO, M.I.F.; SOLIS, D.; KALTNER, H.; GABIUS, H.J. ROMERO, A. Crystal Structure Analysis of Human Galectin-I. **Journal Molecular Biology**, v.343(4), p.957-970, 2004.

MAHONEY, S.A.; WILKINSON, M.; SMITH, S.; HAYNES, S.W. Stabilization of neurites in cerebellar granule cells by transglutaminase activity: identification of midkine and galectin-3 as substrates. **Neuroscience**, v.101, p.141-155, 2000.

MARKOVA, V.; SMETANA, K.J.; JENÍKOVA, G.; LÁCHOVA, J.; KREJCIRÍKOVA, V.; POPLSTEINS, M.; FÁBRY, M.; BRYNDA, J.; ALVAREZ, R.A.; CUMMINGS, R.D.; MALY, P. Role of the carbohydrate recognition domains of mouse galectin-4 in oligosaccharide binding and epitope recognition and expression of galectin-4 and galectin-6 in mouse cells and tissues. **International Journal of Molecular Medicine**, v.18, p.65-76, 2006.

MASSA, S.M.; COOPER, D.N.W.; LEFFLER, H.; BARONDES, S.H.. L-29 an endogenous lectin binds to glycoconjugate ligands with positive cooperativity. **Biochemistry**, v.32, p.260-267, 1993.

MATTHEWS, B.W. Solvent content of protein crystals. **Journal of Molecular Biology**, v.33(2), p.491-497, 1968.

MEHUL, B.; BAWUMIA, S.; HUGHES, R.C. Cross-linking of galectin 3, a galactose-binding protein of mammalian cells, by tissue-type transglutaminase. **European Journal of Biochemistry** v.360, p.160-164, 1995.

MENON, R.P.; HUGHES, R.C. Determinants in the N-terminal domains of galectin-3 for secretion by a novel pathway circumventing the endoplasmic reticulum-Golgi complex. **European journal of Biochemistry**, v.264, p.569-572, 1995.

MORRIS, S.; AHMAD, N.; ANDRÉ, SABINE; KALTNER, H.; GABIUS, HAN-J; BRENOWITZ, M.; BREWER, F. Quaternary solution structures of galectins-1, -3, and -7. **Glycobiology**, v.14, p.293-300, 2004.

MURSHUDOV, G.N.; VAGIN, A.A.; LEBEDEV, A.; WILSON, K.S.; DODSON, E.J. Efficient anisotropic refinement of macromolecular structures using FFT. **Acta Crystallographica Section D**, v.55, p.247-255, 1999.

NAGY, N.; LEGENDRE, H.; ENGLES, O.; ANDRE, S.; KALTNER, H. WASANO, K.; ZICK, Y.; PECTOR, J.C. DECAESTECKER, C. Cell surface overexpression of galectin-3 and the presence of its ligand in the blood plasma as determinantes in colon neoplastic lesions. **Cancer**, v.97, p.1849-58, 2003.

ODA, Y.; HERRMANN, J.; GITT, M.A. TURCK, C.W.; BURLINGAME, A.L.; BARONDES, S.H.; LEFFLER, H. Soluble lactose-binding lectin from rat intestine with two

different carbohydrate-binding domains in the same peptide chain. **Journal of Biological Chemistry**, v.268, p.5929-5939, 1993.

OHYAMA, Y.; HIRABAYASHI, J.; ODA, Y.; OHNO, S. KAWASAKI, H.; SUZUKI, K.; KASAI, K. Nucleotide sequence of chick 14K beta-galactoside-binding lectin . **Biochemical and Biophysical Research Communication**, v. 134, p.51-56, jan. 1986.

PACE, K.E.; BAUM, L.G. Galectin-1 receptors in different cell types. **Glycoconjugate Journal**, v.19, p.607-614, 2004.

PATTERSON, R.J.; LIU, F.T.; WANG, J.L.; Intracellular functions of galectins. **Biochimica et Biophysica Acta** v.1572, p.263-273, 2004.

PERILLO, N.L.; PACE, K.E.; SEILHAMER, J.J.; BAUM, L.G. Apoptosis of T cells mediated by galectin-1. **Nature**, v.378, p.736-739, 1995.

PITTS, M.J; YANG, D.C.H. Mitogenicity and binding properties of β -galactoside-binding lectin chick-embryo kidney. **Biochemical Journal**, v.195, 435-439, 1981.

RABINOVICH, G.A.; RUBINSTEINS, N.; TOSCANO, M.A. Role of galectins in inflammatory and immunomodulatory processes. **Biochimica et Biophysica Acta**, v.19, p.274-84, 2002.

RAMACHANDRAM, G.N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. **Journal of Molecular Biology**, v.7, p.95-99, 1963.

RECHRECHE, H.; MALLO, G.V.; MONTALTO, G.; DAGOM, J.; IOVANNA, J.L. Cloning and expression of the mRNA of human galectin-4, an S-type lectin down-regulated in colorectal cancer. **European Journal of Biochemistry**, v.248, p.225-230, 1997.

RUBINSTEIN, N.; IARREGUI, J.M.; TOSCANO, M.A.; RABINOVICH, G.A. The role of galectins in the initiation, amplification and resolution of the inflammatory response. **Tissue Antigens**, v. 64, p.1-12, 2004.

RUMILLA, K.M.; ERICKSON, L.A.; ERICKSON, A.K.; LLOYD, V. Galectin-4 expression in carcinoid tumors. **Endocrine Pathology**, v.17, p.243-250, 2006.

SANFORD, G.L.; HARRIS-HOOKER, S. Stimulation of vascular cell proliferation by beta-galactoside specific lectins. **The FASEB Journal**, v.4, p.2912-18, 1990.

SATO, S.; BURDETT, I.; HUGHES, R.C. Secretion of the baby hamster kidney 30-kDa galactose-binding lectin from polarized and nonpolarized cells: a pathway independent of the endoplasmic reticulum-Golgi complex. **Experimental Cell Research**, v.207, p.8-18, 1993.

SATO, S.; NIEMINEN, J. Galectin-3 interacts with naive and primed neutrophils, inducing innate immune response. **Glycoconjugate Journal**, v.19, p.583-591, 2004.

SAUSSEZ, S.; NONCLERCQ, D.; LAURENT, G.; WATTIEZ, R.; ANDRE, S.; GABIUS, H.J.; KISS, R.; TOUBEAU, G. Galectin-1: A small proteins with major function. **Histochemistry and Cell Biology**, v.123. p.29-41, 2005.

SCANLAN, M.J.; CHEN, Y.T.; WILLIAMSON, B.; GURE, A.; STOCKERT, E.; GORDON, J.D.; TURECI, O.; SAHIN, U.; PFREUNDSCHUH, M.; OLD, L. Characterization of human colon cancer antigens recognized by autologous antibodies. **Internation Journal of Cancer**, v.76. p.652-658, 1998.

SEETHARAMAN, J.; KANOGBERG, A.; SLAABY, R.; LEFFLER, H. BARONDES, S.H.; RINI, J.M. **X-ray Crystal Structure of the Human Galectin-3 Carbohydrate Recognition Domain at 2.1-Å Resolution**. **Journal of Biological Chemistry**, v.273, p.13047-13052, 1999.

SHIRAI, T.; MITSUYAMA, C.; NIWA, Y.; MATSUI, Y.; HOTTA, H.; YAMANE, T.; KAMIYA, H.; ISHII, C.; OGAWA, T.; MURAMOTO, K. Structure of galectin in lactose-ligand and ligand free forms: emergence of a new structure class by eccelerated evolution. **Structure**, v.7, p.1223-33, 1999.

SWAMINATHAN, G.J.; LEONIDAS, D.D.; SAVAGE, M.P.; ACHERMAN, S.J.; ACHARYA, K.R. Selective recognition of mannose by the human eosinophil charcot-leyden crystal protein (Galectin-10): a crystallographic study at 1.8 Å resolution. **Biochemistry**, v.38, p.13837-43, 1999.

TAKENAKA, Y.; FUKUMORI, T.; RAZ, A. Galectin-3 and metastasis. **Journal of Biological Chemistry**, v. 19, p.543-549, 2004.

TRACEY, B.M.; FEIZI, T.; ABBOTT, M.W.; CARRUTHERS, R.A.; GREENLL, B.N.; LAWSON, A.M. Subunit molecular mass assignment of 14,654 Da to the soluble β -Galactoside-binding Lectin from B ovine heart muscle and demonstration of intramolecular disulfide bonding associated with oxidative inactivation. **The Journal of Biological Chemistry**, v.267, p.10342-47, 1992.

VAGIN, A.A.; TEPLYAKOV, A. MOLREP: an automated program for molecular replacement. **Journal of Applied Crystallography**, v.30, p.1022-1025, 1997.

VAN DEN BRÛLE, F.; CALIFICE, S.; CASTRONOVO, V. Expression of galectins in cancer: a critical review. **Glycoconjugate Journal**, v.19, p.532-542, 2004.

VARELA, P.F.; SOLIS, D.; DIAZ-MAURINO, T.; KALTNER, H.J.; GABIUS ROMERO, A.; The 2.15Å crystal structure of Cg-16, the developmentally regulated homodimeric chicken galectin. **Journal Molecular Biology**, v.294, p.537-549, 1999.

WASSANO, K.; HIRAKAWA, Y. Two domains of rat galectin-4 bind to distinct structures of the intercellular borders of colorectal epithelia. **Journal of Histochemistry & Cytochemistry**, v.47, p.75-82, 1999.

WATT, D.; JONES, G.E.; GOLDRING, K. The involvement of gal-1 in skeletal muscle determination differentiation and regeneration. **Glycoconjugate Journal**, v.19, p.615-619, 2004.

WELLS, V.; MALLUCCI, L. Identification of an autocrine negative growth factor: Mouse β -galactoside-binding protein is a cytostatic factor and cell growth regulator. **Cell**, v.64, p.91-97, 1991.

WOLLINA, U.; GRAEFE, T.; FELDRAPPE, S.; ANDRE, S.; WASANO, K. KALTNER, H.; ZICK, Y.; GABIUS, H.J. Homodimeric galectin-7 (p53-induced gene 1) is a negative growth regulator for human neuroblastoma cells. **Oncogene**, v.22, p.6277-88, 2003.

Apêndice I: O *lac* operon

A lactose é um dissacarídeo composto por uma molécula de glicose e uma molécula de galactose. Este açúcar pode servir como fonte de carbono para a bactéria *Escherichia coli*. Para tanto, a lactose deve penetrar na célula da bactéria e ser quebrada em seus monossacarídeos constituintes, a galactose e a lactose. Os genes que controlam o metabolismo da lactose no genoma de *E.coli* encontram-se organizados em um *operon*, denominado *operon lac*, constituído por três genes estruturais contínuos, precedidos por uma região reguladora que controla coordenadamente a transcrição destes genes. O gene *lacZ* codifica a proteína β -galactosidase que quebra a lactose em glicose e galactose. O gene *lacY* codifica a proteína β -galactosidase-permease que se associa a membrana plasmática da bactéria e funciona como uma proteína de transporte que carrega a lactose para dentro do citoplasma. O gene *lacA* codifica a síntese da enzima transacetilase responsável pela eliminação dos galactosídeos não hidrolisados. A região controladora é constituída por uma sequência de nucleotídeos denominada promotor, que funciona como sítio de ligação da enzima RNA polimerase, sucedida por uma curta sequência de nucleotídeos denominada operador que funciona como sítio de ligação da proteína repressora da expressão do operon. O gene de interesse a ser expresso é inserido em um sítio múltiplo de clonagem (SMC), próximo a região controladora, a indução da transcrição é obtida pela adição de um análogo de lactose sintético e não degradável (isopropiltio- β -D-galactosídeo, IPTG).

Na ausência de lactose, o operador está sempre associado a uma proteína repressora. Essa proteína é codificada pelo gene *i*, denominado gene repressor e que não faz parte do *operon lac*. A proteína impede a enzima RNA polimerase de transcrever os genes estruturais. Se houver a entrada de lactose na célula, uma molécula deste açúcar se liga a proteína repressora associada ao operador alterando sua conformação espacial, provocando sua dissociação do operador. Com o operador desimpedido, a RNA polimerase pode efetuar a transcrição dos genes estruturais do *operon lac*, como também do gene de interesse.

Apêndice II: Preparação de células competentes utilizando CaCl_2

1. Preparar um pré-inóculo com 5,0 mL de meio LB (Luria-Bertani);
2. Incubar a cultura overnight a 37°C / 250 rpm;
3. Transferir 1 mL do pré-inóculo para 100 mL de meio LB (Luria-Bertani) pré-aquecido a 37°C ;
4. Incubar a cultura a 37°C , 250 rpm, e monitorar o crescimento espectofotometricamente até alcançar uma densidade óptica (D.O._{600}) ~ 0.37 ;
5. Transferir o volume para dois tubos falcons de 50,0 mL estéreis, previamente mantidos em gelo. Manter os tubos em repouso no gelo por 10 minutos;
6. Em seguida, centrifugar a suspensão a 1600g, 7 minutos a 4°C ;
7. Descartar os sobrenadantes e ressuspender gentilmente o precipitado em 10,0 mL de uma solução A gelada;
8. Centrifugar a suspensão a 1100g, 5 minutos a 4°C ;
9. Descartar o sobrenadante e ressuspender gentilmente o precipitado em 10,0 mL de solução de cloreto de cálcio gelada;
10. Incubar a suspensão em gelo por 30 minutos;
11. Centrifugar a 1100g, 5 minutos a 4°C ;
12. Descartar o sobrenadante e ressuspender o precipitado muito gentilmente em 2,0mL da solução A gelada;
13. Aliquotar a suspensão em volumes de 200 μL e manter a uma temperatura de -80°C .

Solução A (estéril)

- Tampão Hepes 10 mM pH 7.0
- CaCl_2 60 mM
- Glicerol 15% (v/v)

Apêndice III: Protocolo de transformação utilizando células competentes

1. Descongelar as bactérias *E.coli* DH5 α competentes em gelo (~ 15 minutos);
2. Misturar 10,0 μ L do produto da ligação a 100,0 μ L da suspensão contendo as bactérias competentes
3. Manter em repouso por 30 minutos;
4. Dar um choque térmico a 42°C por 2 minutos;
5. Manter em repouso no gelo por 2 minutos;
6. Adicionar 270 μ L de meio LB e incubar a 37°C por 1 hora a 200rpm;
7. Plaquear até 200 μ L em meio de cultura LB contendo antibiótico kanamicina (50 μ g/mL);
8. Manter as culturas em estufas a 37°C por um período de 16 horas.