



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA**

**MARINA DUARTE PINTO LOBO**

**ANÁLISE PROTEÔMICA DE PLASMA DE PACIENTES COM  
CÂNCER DE MAMA UTILIZANDO LECTINAS VEGETAIS E *LABEL-  
FREE* LC-MS<sup>E</sup>**

**FORTALEZA**  
**2016**

**MARINA DUARTE PINTO LOBO**

**ANÁLISE PROTEÔMICA DE PLASMA DE PACIENTES COM  
CÂNCER DE MAMA UTILIZANDO LECTINAS VEGETAIS E *LABEL-  
FREE* LC-MS<sup>E</sup>**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Bioquímica, da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Doutora em Bioquímica.

Orientador: Prof. Dr. Renato de Azevedo Moreira

Co-Orientadora: Profa. Dra. Ana Cristina de Oliveira Monteiro Moreira

**FORTALEZA  
2016**

Esta tese foi apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Ceará como parte dos requisitos necessários à obtenção do grau de Doutor em Bioquímica, outorgado pela Universidade Federal do Ceará, e encontrar-se-á à disposição na Biblioteca Central da referida Universidade.

A transcrição ou utilização de qualquer trecho deste trabalho é permitida desde que seja feito de acordo com as normas da ética científica.



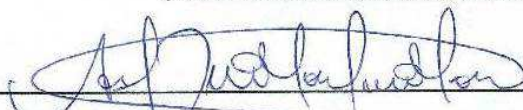
**Marina Duarte Pinto Lobo**

Tese aprovada em: 22 / 01 / 2016

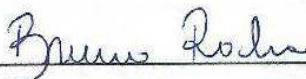
**BANCA EXAMINADORA**



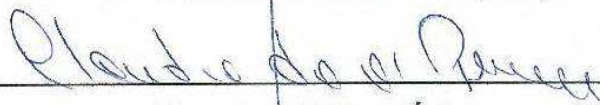
**Prof. Dr. Renato de Azevedo Moreira (Orientador)**  
Universidade Federal do Ceará (UFC)



**Profa. Dra. Ana Cristina de Oliveira Monteiro Moreira (Co-Orientadora)**  
Universidade de Fortaleza (UNIFOR)



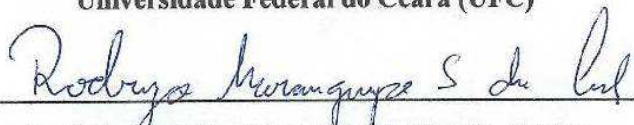
**Prof. Dr. Bruno Anderson Matias da Rocha**  
Universidade Federal do Ceará (UFC)



**Profa. Dra. Cláudia do Ó Pessoa**  
Universidade Federal do Ceará (UFC)



**Prof. Dr. Hélio Vitoriano Nobre Júnior**  
Universidade Federal do Ceará (UFC)



**Prof. Dr. Rodrigo Maranguape Silva da Cunha**  
Universidade Estadual Vale do Acaraú (UEVA)

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca de Ciências e Tecnologia

- 
- L784a Lobo, Marina Duarte Pinto.  
Análise proteômica de plasma de pacientes com câncer de mama utilizando lectinas vegetais e *Label-Free* LC-MS<sup>E</sup> / Marina Duarte Pinto Lobo. – 2016.  
157 f. : il., color., enc. ; 30 cm.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Departamento de Bioquímica e Biologia Molecular, Programa de Pós-Graduação em Bioquímica, Fortaleza, 2016.  
Área de Concentração: Bioquímica Vegetal.  
Orientação: Prof. Dr. Renato de Azevedo Moreira.  
Coorientação: Profa. Dra. Ana Cristina de Oliveira Monteiro-Moreira.
1. Mamas - Câncer. 2. Lectinas. 3. Glicoproteínas. I. Título.

---

CDD :

*A Deus e  
a minha amada família.*

## AGRADECIMENTOS

A *Deus*. Por estar ao meu lado em todos os momentos da minha vida, me dando força, e me fazendo sentir Sua presença mesmo nos dias difíceis. Sem Ele, nada sou.

Ao meu querido orientador Prof. Dr. *Renato de Azevedo Moreira*. Pelo acolhimento em seu laboratório, pela orientação, pelas discussões científicas sempre engrandecedoras e estimulantes, e pelo grande apoio, presença constante e suporte oferecido durante todos esses anos de doutorado.

A minha querida co-orientadora Profa. Dra. *Ana Cristina de Oliveira Monteiro Moreira*. Por ter me recebido tão bem em seu grupo de pesquisa, pela confiança depositada e por ter acreditado em mim e no meu trabalho. Pela orientação, pela amizade construída ao longo desses anos, pela boa risada contagiante e por todos os momentos de injeção de ânimo que, muitas vezes, manteve acesa em mim a “chama da pesquisa”, mesmo quando as coisas não iam a nosso favor. Serei sempre grata a você e ao Prof. Renato, que juntos formam um casal fenomenal!

Aos demais membros da banca, Prof. Dr. *Bruno Anderson Matias da Rocha*, Profa. Dra. *Cláudia do Ó Pessoa*, Prof. Dr. *Hélio Vitoriano Nobre Júnior* e Prof. Dr. *Rodrigo Maranguape Silva da Cunha*. Pela participação relevante em minha banca, trazendo vasta experiência científica para discussão e melhorias do trabalho.

À Profa. Dra. *Sara Maria Lima Verde*. Pela realização das coletas das amostras utilizadas nesse trabalho e pela sempre disponibilidade em ajudar.

À *equipe do Hospital Geral de Fortaleza*, pelo auxílio na coleta das amostras de plasma das mulheres participantes.

A todos os *integrantes do Laboratório F66, F51 e F52*. Pela colaboração quando foi necessária.

Ao amigo e companheiro brilhante de laboratório, *Felipe Sousa*. Pelas inúmeras conversas divertidas entre um experimento e outro e pelas palavras confortantes e animadoras sempre em momentos oportunos.

Aos meus amados pais, *Walter e Margaret*. Por todo esforço em me proporcionarem uma educação de qualidade e uma vida digna e honesta. Pelo amor sem medidas dedicado à família todos esses anos. Família essa que é o meu porto-seguro.

As minhas irmãs *Raquel e Nathália*, meu meio-irmão *Rilke* e meu cunhado *Walder*. Pela presença constante em minha vida, proporcionando tantos momentos felizes.

Aos amores da minha vida *Danielle e Matheus*. Por encherem nossa casa de alegria, e bagunça, conseguindo me fazer sorrir mesmo nos momentos de cansaço e ansiedade.

Ao meu amado marido *Frederico*. Por todo carinho, apoio, incentivo, atenção, ajuda e compreensão ao longo desses anos. Muito obrigada, de todo coração, pelas doses de amor e paciência diárias, que tornaram esse percurso mais leve.

A todas as mulheres que, mesmo sem conhecer, aceitaram em participar dessa pesquisa, algumas em um momento tão difícil da vida que é enfrentar um câncer. A elas meu mais sincero agradecimento.

As agências de financiamento, *Conselho Nacional de Pesquisa e Desenvolvimento (CNPq)*, a *Fundação Estadual de Pesquisa e Desenvolvimento (FUNCAP)* e *FINEP* pelo apoio financeiro.

A *Universidade Federal do Ceará (UFC)*, em especial a todo *corpo docente do Departamento de Bioquímica e Biologia Molecular*, pela estrutura necessária para realização de experimentos e pela formação acadêmica de excelência oferecida durante os 11 anos que estive como discente na instituição.

A *Universidade de Fortaleza (UNIFOR)*, pela infraestrutura primorosa, a total disposição, que possibilitou a realização de toda a parte experimental desta pesquisa.

E a todas as outras pessoas, que pequei em esquecer, que contribuíram de alguma forma para a realização desse trabalho.

*"Não cruzarás o mesmo rio duas vezes, porque  
outras são as águas que correm nele."*

*Heráclito de Efeso*

## RESUMO

O câncer de mama representa o tipo de câncer mais comum entre as mulheres no mundo e no Brasil, depois do câncer de pele não melanoma. Caracterizando-se como uma doença clínica e biologicamente heterogênea, a detecção precoce do câncer de mama, uma caracterização fidedigna da doença e um diagnóstico preciso são de fundamental importância para redução da mortalidade. Assim, o objetivo do presente trabalho foi realizar análises qualitativas e quantitativas de proteínas plasmáticas de mulheres saudáveis e mulheres com câncer de mama tipo ductal, em diferentes estágios. Para isso, inicialmente, as amostras de plasma foram depletadas de albumina e IgG e depois reunidas em *pools* representativos para cada grupo em estudo. Os *pools* foram então fracionados por meio de cromatografias de afinidade em matrizes de Sepharose 4B imobilizadas com as lectinas vegetais  $\alpha$ -galactose-ligante de *Artocarpus incisa* - Frutalina e glucose/manose-ligante de *Dioclea altíssima* – DAL. Posteriormente, tanto os *pools* como as frações cromatográficas obtidas foram digeridas e submetidas à análise independente de dados (MS<sup>E</sup>) e quantificadas (*label-free*) por espectrometria de massas. A utilização das cromatografias de afinidade com lectinas vegetais *a priori* da análise por espectrometria de massas foi fundamental para reduzir a complexidade das amostras e estender o intervalo dinâmico de proteínas, e frutalina apresentou os melhores resultados de fracionamento. Um somatório de todas as análises realizadas revelou a identificação de cerca de 445.000 peptídeos e mais de 30.000 proteínas, com redundância entre isoformas do mesmo grupo e entre grupos. Ademais, além de fracionar, as cromatografias de afinidade com lectinas vegetais foram eficientes em isolar glicofomas específicas que refletem um padrão de glicosilação alterado associado à doença. Diversas proteínas diferencialmente expressas, algumas delas com mudanças na glicosilação, foram identificadas, tais como apolipoproteína A2, apolipoproteína C3, fatores do sistema complemento (C3, C4b e C4A), clusterina,  $\alpha$ -1-2-ácido glicoproteína, haptoglobina, hemopexina, paraoxonase arilesterase sérica, proteína plasmática de ligação ao retinol, plasminogênio e vitronectina. Dentre as proteínas identificadas, algumas estão relacionadas com a decomposição da matriz extracelular, metabolismo lipídico, estresse oxidativo e outras caracterizam-se como proteínas de fase aguda. Finalmente, os dados de expressão diferencial e possíveis alterações de glicosilação das proteínas contribuem para o desenvolvimento de um perfil protéico, alvo para posterior validação, que apresenta uma associação com o desenvolvimento e a caracterização do câncer de mama em seus diferentes estágios.

## ABSTRACT

Title: PROTEOMIC ANALYSIS OF PLASMA FROM PATIENTS WITH BREAST CANCER USING PLANT LECTINS AND LABEL FREE MS<sup>E</sup>

The breast cancer presents one of the most commonly diagnosed types of cancer among women worldwide and in Brazil, after nonmelanoma skin cancer. It's a clinically and biologically heterogeneous disease. Therefore, early detection of breast cancer, a reliable characterization of the disease and an accurate diagnosis are crucial to reducing mortality. The aim of this work was perform a qualitative and quantitative analyzes of plasma proteins from healthy women and from women with ductal breast cancer in different stages. For this, initially, plasma samples were depleted of IgG and albumin and then pooled into samples representative for each study group. The pools were then fractionated by immobilized plant lectins (frutalin and DAL)-affinity chromatography. Subsequently, the pools and chromatographic fractions were digested and submitted to and data-independent label-free mass spectrometric analysis. The use of affinity chromatography with plant lectins prior to analysis by mass spectrometry was essential to reduce sample complexity and extend the dynamic range of proteins, and frutalin showed the best results from fractionation. A sum of all analyzes performed revealed the identification of about 445,000 peptides and more than 30,000 proteins, with redundancy between isoforms of the same group and between groups. Furthermore, in addition to fractionation, the chromatography of affinity with plant lectins were effective in isolating specific glycoforms that reflect an altered glycosylation pattern associated with the disease. Several differentially expressed proteins, some of which changes in glycosylation were identified, such as apolipoprotein A2, apolipoprotein C3, complement factors (C3, C4b and C4A), clusterin, 1-2- $\alpha$ -acid glycoprotein, haptoglobin, hemopexin, serum paraoxonase arylesterase, plasma retinol binding protein, plasminogen and vitronectin. Among the proteins identified, some are related to the breakdown of the extracellular matrix, lipid metabolism, oxidative stress and other characterized as acute phase proteins. Finally, the data of differential expression and possible changes in the glycosylation patterns of proteins contributes to the development of a protein profile, subject to later validation, presenting an association with the development and characterization of breast cancer in its different stages.

## LISTA DE FIGURAS

<b>Figura 1</b>	Curva universal de reposta-sinal para quantificação absoluta de proteínas padrão (ou referência)	36
<b>Figura 2</b>	Fluxograma representativo da estratégia experimental utilizada no presente estudo	42
<b>Figura 3</b>	Purificação das lectinas D-galactose-ligante das sementes de <i>Artocarpus incisa</i> (frutalina) e glucose/manose-ligante de <i>Dioclea altissima</i> (DAL)	55
<b>Figura 4</b>	Análise por SDS-PAGE das lectinas purificadas	56
<b>Figura 5</b>	Espectro de massa intacta deconvoluído de Frutalina	58
<b>Figura 6</b>	Espectro de massa intacta deconvoluído de DAL	59
<b>Figura 7</b>	Depleção de albumina e IgG das amostras de plasma humano	61
<b>Figura 8</b>	Cromatografias de afinidade em coluna de Frutalina-Sepharose	59
<b>Figura 9</b>	Cromatografias de afinidade em coluna de DAL-Sepharose	61
<b>Figura 10</b>	Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração não retida (Pico 1) obtida após cromatografia dos <i>pools</i> em coluna de Frutalina-Sepharose	68
<b>Figura 11</b>	Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração retida (Pico 2) obtida após cromatografia dos <i>pools</i> em coluna de Frutalina-Sepharose	69
<b>Figura 12</b>	Análise por nanoUPLC-HDMS <sup>E</sup> das frações obtidas após cromatografia dos <i>pools</i> em matriz de Frutalina-Sepharose	71
<b>Figura 13</b>	Intervalo Dinâmico das proteínas identificadas e quantificadas nas frações cromatográficas obtidas após cromatografia dos <i>pools</i> em matriz de Frutalina-Sepharose	74
<b>Figura 14</b>	Valores de razão (Câncer de mama:Controle) identificados para algumas proteínas encontradas nos picos 1 e 2 após cromatografia de afinidade em FTL-Sepharose, nos diferentes estágios da doença	76
<b>Figura 15</b>	Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração não retida (Pico 1) obtida após cromatografia dos <i>pools</i> em coluna de DAL-Sepharose	83
<b>Figura 16</b>	Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração retida (Pico 2) obtida após cromatografia dos <i>pools</i> em coluna de DAL-Sepharose	85
<b>Figura 17</b>	Análise por nanoUPLC-HDMS <sup>E</sup> das frações obtidas após cromatografia dos <i>pools</i> em matriz de DAL-Sepharose	87
<b>Figura 18</b>	Intervalo Dinâmico das proteínas identificadas e quantificadas nas frações cromatográficas obtidas após cromatografia dos <i>pools</i> em matriz de DAL-Sepharose	89
<b>Figura 19</b>	Valores de razão (Câncer de mama:Controle) identificados para algumas proteínas encontradas nos picos 1 e 2 após cromatografia de afinidade em DAL-Sepharose, nos diferentes estágios da doença	107
<b>Figura 20</b>	Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos do <i>pool</i> (controle) sem fracionamento prévio em coluna de lectina vegetal imobilizada	100
<b>Figura 21</b>	Análise por nanoUPLC-HDMS <sup>E</sup> de plasma humano total depletado ( <i>pools</i> )	101
<b>Figura 22</b>	Intervalo Dinâmico das proteínas identificadas nos <i>pools</i> sem fracionamento prévio em colunas de lectinas vegetais imobilizadas	102
<b>Figura 23</b>	Comparação do número total de peptídeos e proteínas identificadas, de proteínas quantificadas e de proteínas diferentes incluídas nas análises de expressão diferencial, entre as 3 análises realizadas (Fracionamento com Frutalina, Fracionamento com DAL e sem fracionamento com lectina vegetal <i>a priori</i> da análise por espectrometria de massas)	107
<b>Figura 24</b>	Diagramas de Venn para proteínas identificadas e quantificadas por LC-MS <sup>E</sup> obtidas sem e com fracionamento prévio em colunas de frutalina e DAL imobilizadas	108
<b>Figura 25</b>	Interatoma elaborado a partir da lista de proteínas que apresentaram diferenças de expressão e/ou no padrão de glicosilação no plasma de mulheres com câncer de mama, comparado ao de mulheres saudáveis a 0.5 de confiança	118

## ANEXOS

- Tabela 1 -** Lista de proteínas plasmáticas identificadas na fração não retida em matriz de Frutalina-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis
- Tabela 2 -** Lista de proteínas plasmáticas identificadas na fração não retida em matriz de Frutalina-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis
- Tabela 3 -** Lista de proteínas plasmáticas identificadas na fração não retida em matriz de Frutalina-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio III (avançado) comparada com mulheres saudáveis
- Tabela 4 -** Lista de proteínas plasmáticas identificadas na fração retida em matriz de Frutalina-Sepharose 2) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis
- Tabela 5 -** Lista de proteínas plasmáticas identificadas na fração retida em matriz de Frutalina-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis
- Tabela 6 -** Lista de proteínas plasmáticas identificadas na fração retida em matriz de Frutalina-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio III (avançado) comparadas com mulheres saudáveis
- Tabela 7 -** Lista de proteínas plasmáticas identificadas na fração não retida em matriz de DAL-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis
- Tabela 8 -** Lista de proteínas plasmáticas identificadas na fração não retida em matriz de DAL-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis
- Tabela 9 -** Lista de proteínas plasmáticas identificadas na fração não retida em matriz de DAL-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio III (avançado) comparadas com mulheres saudáveis
- Tabela 10 -** Lista de proteínas plasmáticas identificadas na fração retida em matriz de DAL-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis
- Tabela 11 -** Lista de proteínas plasmáticas identificadas na fração retida em matriz de DAL-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis
- Tabela 12 -** Lista de proteínas plasmáticas identificadas na fração retida em matriz de DAL-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio III (avançado) comparadas com mulheres saudáveis
- Tabela 13 -** Lista de proteínas plasmáticas identificadas no *pool* de plasma de mulheres com câncer de mama ductal I, estágio precoce, comparadas com mulheres saudáveis sem fracionamento prévio em coluna de lectina vegetal imobilizada
- Tabela 14 -** Lista de proteínas plasmáticas identificadas no *pool* de plasma de mulheres com câncer de mama ductal II, estágio intermediário, comparadas com mulheres saudáveis sem fracionamento prévio em coluna de lectina vegetal imobilizada
- Tabela 15 -** Lista de proteínas plasmáticas identificadas no *pool* de plasma de mulheres com câncer de mama ductal III, estágio avançado comparadas com mulheres saudáveis sem fracionamento prévio em coluna de lectina vegetal imobilizada
- Tabela 16 -** Lista de proteínas plasmáticas identificadas nas diferentes análises (com fracionamento em co Frutalina imobilizada, com fracionamento em coluna de DAL imobilizada e sem fracionamento em coluna de lectina imobilizada) e seus respectivos valores de razão (Câncer de Mama : Controle) encontrados.

## LISTA DE ABREVIATURAS

<b>ADH</b>	Alcool dehydrogenase de levedura
<b>AUC</b>	Método de quantificação baseado na mensuração da área do pico
<b>DAL</b>	Lectina de <i>Dioclea altissima</i>
<b>DDA</b>	Aquisição dependente de dados
<b>DO<sub>280</sub></b>	Densidade óptica a 280 nm
<b>DTT</b>	Dithithreitol
<b>ESI</b>	Ionização por eletrospray
<b>FDR</b>	<i>False discovery rate</i> – Taxa de falsa descoberta
<b>FTL</b>	Lectina de <i>Artocarpus incisa</i>
<b>FPLC</b>	<i>Fast protein liquid chromatography</i> – Cromatografia líquida rápida de proteína
<b>HER2</b>	Receptor 2 do fator de crescimento epidérmico humano
<b>IAA</b>	Iodoacetamida
<b>ICAT</b>	Marcador isotópico de afinidade
<b>iTRAQ</b>	Marcador isobárico para quantificação absoluta e relativa
<b>kDA</b>	Quilodalton
<b>LC</b>	<i>Liquid chromatography</i> – Cromatografia líquida
<b>MS</b>	Espectrometria de massas
<b>MS1</b>	Espectro de massas do íon precursor
<b>MS2(MS/MS)</b>	Espectro de massas do íon fragmento
<b>N- terminal</b>	Amino-terminal
<b>PHOSP-B</b>	Proteína rábica fosforilase B
<b>SDS</b>	Dodecil sulfato de sódio
<b>SDS-PAGE</b>	Eletroforese em gel de poliacrilamida na presença de SDS
<b>SILAC</b>	Marcação isotópica em aminoácidos em cultura de células
<b>TOF</b>	<i>Time of flight</i> – Tempo de voo
<b>TNBC</b>	<i>triple negative breast cancer</i> – cancer de mama triplo-negativo
<b>TNM</b>	Classificação de Tumores Malignos

## SUMÁRIO

<b>1. JUSTIFICATIVA</b> .....	17
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	18
2.1. Câncer de mama.....	18
2.2. Biomarcadores .....	21
2.3. A glicosilação de proteínas .....	24
2.4. Lectinas .....	27
2.5. Abordagens proteômicas na busca por biomarcadores.....	31
2.6. A abordagem proteômica quantitativa label-free LC-MS <sup>E</sup> utilizada no estudo.....	33
<b>3. HIPÓTESES</b> .....	39
<b>4. OBJETIVOS</b> .....	40
4.1. Objetivo Geral.....	40
4.2. Objetivos Específicos .....	40
<b>5. ESTRATÉGIA EXPERIMENTAL</b> .....	41
<b>6. MATERIAIS E MÉTODOS</b> .....	43
6.1. Materiais .....	43
6.1.1. Grupo de estudo .....	43
6.1.1.1. Critérios de inclusão.....	43
6.1.1.2. Critérios de exclusão .....	44
6.1.1.3. Coleta das amostras de sangue .....	44
6.1.2. Reagentes e outros materiais analíticos .....	45
6.2. Métodos .....	45
6.2.1. Isolamento e Imobilização das lectinas vegetais .....	45
6.2.1.1. Isolamento da lectina D-galactose-ligante das sementes de Artocarpus incisa (frutalina).....	45
6.2.1.2. Isolamento da lectina glucose/manose-ligante de Dioclea altíssima (DAL).....	46
6.2.1.3. Avaliação do grau de pureza das lectinas vegetais isoladas.....	47

6.2.1.4. Imobilização das lectinas em matriz de Sepharose 4B .....	48
6.2.3. Cromatografias de depleção.....	49
6.2.4. Cromatografias de afinidade em colunas de lectinas imobilizadas .....	49
6.2.5. Digestão de proteínas para LC-MS <sup>E</sup> .....	51
6.2.6. NanoUPLC e aquisição independente de dados (MSE) livre de marcação (label-free) por espectrometria de massas .....	52
6.2.7. Processamento de dados e análise de expressão diferencial.....	53
6.2.8. Interatoma .....	54
<b>7. RESULTADOS E DISCUSSÃO .....</b>	<b>55</b>
7.1. Isolamento das lectinas .....	55
7.2. Imobilização das lectinas .....	60
7.3. Cromatografias de depleção de Albumina e IgG.....	60
7.4. Cromatografias de afinidade em colunas contendo lectinas vegetais imobilizadas .....	63
7.5. Análise quali e quantitativa das proteínas fracionadas por meio de cromatografia com Frutalina-Sepharose por espectrometria de massas .....	67
7.5.1. Identificação e quantificação, por espectrometria de massas, das proteínas plasmáticas presentes nas frações após cromatografias de afinidade em matriz de frutalina imobilizada .....	67
7.5.2. Relevância biológica dos dados proteômicos quantitativos.....	75
7.5.2.1. Análise proteômica da fração que não interagiu com a lectina $\alpha$ -galactose ligante frutalina (Pico 1).....	77
7.5.2.2. Análise proteômica da fração que interagiu com a lectina $\alpha$ -galactose ligante frutalina (Pico 2).....	78
7.6. Análise quali e quantitativa das proteínas fracionadas por meio de cromatografia com DAL-Sepharose por espectrometria de massas .....	82
7.6.1. Identificação e quantificação, por espectrometria de massas, das proteínas plasmáticas presentes nas frações após cromatografias de afinidade em matriz de DAL imobilizada .....	82
7.6.2. Relevância biológica dos dados proteômicos quantitativos.....	90
7.6.2.1. Análise proteômica da fração que não interagiu com a lectina manose-ligante de Dioclea altíssima - DAL (Pico 1).....	90

7.6.2.2. Análise proteômica da fração que interagiu com a lectina manose-ligante de Dioclea altíssima - DAL (Pico 2).....	93
7.7. Análise quali e quantitativa das proteínas plasmáticas totais, sem fracionamento prévio utilizando lectinas vegetais, por espectrometria de massas .....	99
7.7.1. Identificação e quantificação, por MS, das proteínas plasmáticas presentes nas amostras sem fracionamento prévio em colunas de lectinas vegetais imobilizadas .....	99
7.7.2. Relevância biológica dos dados proteômicos quantitativos obtidos após análise, por espectrometria de massas, das amostras sem fracionamento prévio com lectinas vegetais .....	104
7.8. Compilado entre as três análises de expressão diferencial realizadas neste trabalho ..	107
7.9. Interatoma .....	118
<b>8. CONCLUSÃO</b> .....	122
<b>10. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	123
<b>11. ANEXOS</b> .....	135

## 1. JUSTIFICATIVA

O câncer de mama é a causa mais frequente de morte por câncer em mulheres e sua incidência é de aproximadamente 1,7 milhão de casos novos ao ano. Para o Brasil, em 2016, são esperados 57.960 casos novos de câncer de mama.

Portanto, a detecção precoce do câncer de mama, uma caracterização fidedigna da doença e um diagnóstico preciso, são de fundamental importância na redução da mortalidade. No entanto, o diagnóstico do câncer de mama, por exemplo, é dificultado pela falta de um método de detecção acurado e biomarcadores válidos.

Assim, esse contexto aponta para a necessidade da condução incessante de investigações que visem à prospecção e avaliação de proteínas alvos mais sensíveis e específicas, que possam ser utilizadas na detecção precoce, na previsão da evolução do câncer de mama, no seu estadiamento e na melhor compreensão da patogênese da doença.

O objetivo do presente estudo foi identificar proteínas únicas e expressas diferencialmente, e possíveis alterações em seus padrões de glicosilação, utilizando o fracionamento por lectinas vegetais, de modo a contribuir para o desenvolvimento de um perfil de proteínas que sugerem uma associação com o desenvolvimento e a caracterização do câncer de mama em diferentes estágios.

Essa proposta, associada as demais técnicas utilizadas hoje, visa uma diminuição nos custos associados com a obtenção, processamento e interpretação de exames, bem como aqueles referentes ao reexame decorrente de amostras inadequadas. Tendo sempre em mente que melhorias na acurácia tendem a eliminar a vulnerabilidade envolvida nos erros de subjetividade de interpretação dos resultados, tornando a caracterização e o diagnóstico do câncer de mama mais definitivo, conciso e correto.

## **2. FUNDAMENTAÇÃO TEÓRICA**

### ***2.1. Câncer de mama***

Câncer é o nome dado ao conjunto de doenças que têm em comum o crescimento desordenado de células que invadem os tecidos e órgãos, podendo espalhar-se para outras regiões do corpo, caracterizando uma metástase. As seis características do câncer - capacidades distintivas e complementares que permitem o crescimento do tumor e disseminação metastática – são: manutenção de sinais proliferativos, mecanismos de escape de supressores de crescimento, ativação de mecanismos de invasão e metástase, imortalidade replicativa, indução da angiogênese e resistência à morte celular (HANAHAN; WEINBERG, 2011).

O desenvolvimento de tumores é um processo de múltiplos passos envolvendo fatores genéticos e ambientais. Assim, as causas de câncer são variadas, podendo ser externas ou internas ao organismo, estando ambas inter-relacionadas. As causas externas relacionam-se ao meio ambiente e aos hábitos ou costumes próprios de um ambiente social e cultural. As causas internas são, na maioria das vezes, geneticamente pré-determinadas, estão ligadas à capacidade do organismo de se defender das agressões externas. Esses fatores causais, físicos, químicos ou genéticos, podem interagir de várias formas, aumentando a probabilidade de transformações malignas nas células normais (INCA, 2015).

Apesar da boa notícia de que a incidência do câncer e suas taxas de mortalidade entre homens e mulheres estão declinando em países desenvolvidos, projeta-se que o câncer tornar-se-á líder de causa de morte em todo o mundo nos próximos anos, superando as doenças cardíacas (SIEGEL; MILLER; JEMAL, 2015).

Entre os diferentes tipos de neoplasia, o câncer de mama apresenta-se como a segunda forma mais comum de câncer, sendo a principal causa de morte por câncer entre mulheres, com estimativa anual de 1,4 milhões de novos casos e mais de 450 mil mortes no ano. Mais de 60 mil casos são esperados para o ano de 2016 nos Estados Unidos (SIEGEL; MILLER; JEMAL, 2015).

O câncer de mama, também conhecido como neoplasia maligna da mama, é um tumor originário do revestimento interno dos ductos de leite ou dos lóbulos que suprem os ductos de leite. Dependendo do local do tumor, caracteriza-se como carcinoma ductal ou carcinoma

lobular, respectivamente. Alguns sintomas do câncer de mama são alterações na pele que recobre a mama, secreções no mamilo e nódulo na mama (KUMAR *et al.*, 2012).

A grande maioria das massas de mama detectadas por palpação e/ou pela mamografia são lesões epiteliais, que incluem mudança fibrocística benigna, hiperplasia, carcinoma *in situ* e carcinoma infiltrado mamário. Apesar de vários tipos e subtipos histológicos de carcinomas mamários existirem, mais de 95% ou são carcinomas ductais ou carcinomas lobulares, com a maioria (75% a 80%) de carcinomas mamários sendo os carcinomas ductais (YODER; WILKINSON; MASSOLL, 2007).

A origem da maior parte dos casos de câncer de mama não é conhecida. No entanto, muitos fatores de risco já foram identificados, incluindo o gênero feminino, o aumento da idade da mulher, histórico familiar de câncer de mama em idade precoce, menarca precoce, menopausa tardia, gravidez com idade avançada, terapia de reposição hormonal prolongada, exposição à irradiação terapêutica na região peitoral, doença proliferativa benigna da mama e mutações genéticas em genes tais como BRCA1/2 (CARLSON *et al.*, 2007).

O câncer de mama é uma doença heterogênea, consistindo de subtipos moleculares diferentes, cada um desses com características biológicas e clínicas também diferentes. Clinicamente, é categorizado em três grupos terapêuticos. O grupo positivo para o receptor de estrogênio (ER) é o mais numeroso e diverso, existindo vários testes genômicos para auxiliar no acompanhamento do tratamento de pacientes. O segundo grupo, positivo para o fator de crescimento epidérmico humano (HER2 ou ERBB2), é um grande sucesso clínico por causa do direcionamento terapêutico eficaz de HER2, o que tem levado a intensos esforços para caracterizar outros números de cópias de DNA aberrantes. Cânceres de mama triplo-negativos (*triple negative breast cancer* - TNBCs), caracterizados pela falta de expressão de ER, de HER2 e de receptor de progesterona (PR), também conhecido como câncer de mama tipo basal, formam um grupo com apenas opções quimioterápicas de tratamento (KOBOLDT *et al.*, 2012).

A *American Joint Committee on Cancer* (AJCC) e a União Internacional de Controle do Câncer (UICC) utilizam o sistema de classificação TNM (*Classification of Malignant Tumours*) como uma ferramenta para os médicos estadiarem diferentes tipos de câncer com base em determinadas normas. Ele é atualizado a cada 6 a 8 anos para incluir os avanços na compreensão de uma doença como o câncer. No sistema TNM, a cada tipo de câncer é atribuída uma letra ou número para avaliar o estágio do câncer: o próprio tumor, os linfonodos regionais ao redor do tumor, e se o tumor se espalhou para outras partes do corpo (metástase).

O carcinoma ductal *in situ*, também conhecido como carcinoma intraductal, é considerado não invasivo ou câncer de mama pré-invasivo. A diferença entre o carcinoma ductal *in situ* e carcinoma invasivo é que as células não se espalharam através dos ductos para o tecido mamário adjacente. Já os tumores infiltrantes ou invasivos menores e mais localizados são classificados em estágio I. Tumores invasivos com mais de 2 cm ou havendo comprometimento de gânglio axilar são classificados em estágio II. No estágio III, encontram-se tumores localmente avançados que muitas vezes precisam de tratamento antes da cirurgia para tornarem-se operáveis. No estágio IV já foi observada metástase.

Existe ainda uma classificação do câncer de mama baseada no padrão diferenciado de expressão gênica, que representam a heterogeneidade entre os carcinomas mamários. Portanto, com base em características gênicas particulares, quatro grandes subtipos de câncer de mama foram inicialmente classificados: luminal, HER2, subtipo *basal-like* e subtipo mama *normal-like*. O subtipo luminal é caracterizado pela expressão de marcadores epiteliais luminais e três grupos são reconhecidos atualmente: luminal A, luminal B/HER2-negativo e luminal B/HER2-positivo. O câncer de mama *basal-like* é um grupo heterogêneo de tumores compreendendo diferentes características histológicas, que expressam marcadores epiteliais basais. O subtipo mama *normal-like*, foi localizado em um *cluster* contendo amostras normais da mama e tumores epiteliais benignos que mostraram superexpressão de genes relacionados ao tecido adiposo e tipos celulares não epiteliais em estudos de validação inicial e subsequentes. Este último subtipo citado pode ser um artefato técnico devido à baixa celularidade tumoral. Assim, o subtipo mama *normal-like* foi muitas vezes ignorado e, por conseguinte, mal caracterizado (PEROU *et al.*, 2000; CIRQUEIRA *et al.*, 2011).

O câncer de mama é tratado de forma mais eficaz quando a malignidade ainda não se espalhou a outros órgãos, isto é, em estágios precoces. O método mais comum para a detecção precoce do câncer de mama é o uso da mamografia. A introdução dessa técnica está associada com uma progressiva redução da taxa de mortalidade por câncer de mama durante as duas últimas décadas. O sucesso da mamografia é devido a sua capacidade de detectar formas precoces de câncer além do desenvolvimento concomitante de diversas opções de tratamento para o câncer de mama. Logo, a mamografia é provavelmente a melhor ferramenta para rastreio de uma população em geral, mas está longe de ser perfeita: de 100 mulheres com câncer da mama, cerca de 20 não serão diagnosticadas através da mamografia. Dessa maneira, faz-se necessário o investimento em pesquisas na busca por métodos alternativos para complementar a mamografia, reduzindo seus falsos negativos e falsos positivos (SCHUMMER; BEATTY; URBAN, 2010).

Muitas moléculas e mecanismos moleculares por trás da complexidade e heterogeneidade do câncer de mama e de seu comportamento clínico permanecem desconhecidos, e os avanços na pesquisa biomédica são esperados para desvendar esses mistérios. A proteômica clínica está experimentando atualmente rápidos avanços tecnológicos que permitem novos meios para melhorar o diagnóstico precoce e a resposta ao tratamento do câncer de mama (ZEIDAN *et al.*, 2014).

Além disso, investigações em curso utilizando abordagens complementares baseadas na proteômica do câncer, poderão fornecer novos *insights* em processos aberrantes entre subtipos do câncer de mama, podendo identificar proteínas adicionais ou perfis de proteínas para refinar as atuais classificações de câncer de mama (LAM; JIMENEZ; BOVEN, 2014).

## **2.2. Biomarcadores**

Como já dito anteriormente, o câncer de mama é tradicionalmente considerado como uma doença bastante heterogênea, com características biológicas e clínicas diversas para os diferentes subtipos. Perfis moleculares do câncer de mama estabelecidos por meio de estudos de expressão gênica tem fornecido uma ferramenta importante para discriminar uma série de subtipos, subtipos esses associados com dados clínicos e resposta ao tratamento. A fim de esclarecer as consequências funcionais da expressão gênica alterada relacionada a cada subtipo de câncer de mama, as tecnologias proteômicas podem proporcionar mais conhecimento por meio da identificação e quantificação de proteínas. Nos últimos anos, as tecnologias proteômicas têm amadurecido a tal ponto que elas podem fornecer expressões de todo o proteoma em diferentes materiais clínicos. Esta tecnologia pode ser aplicada para a identificação de proteínas ou de perfis proteicos para refinar mais subtipos do câncer de mama ou mesmo para a descoberta de novos biomarcadores proteicos que apontem para potencial metastático ou mesmo resistência à terapia (LAM; JIMENEZ; BOVEN, 2014).

O intenso e crescente interesse na descoberta de biomarcadores é um reflexo da necessidade clínica para desenvolvimento de testes com elevado grau de sensibilidade e especificidade para o diagnóstico e prognóstico de doenças, bem como para monitoração das respostas à terapia e recorrência da doença (DRAKE *et al.*, 2012).

Os biomarcadores caracterizam-se como substâncias que se apresentam em níveis alterados no sangue, na urina ou outros fluidos biológicos, e até mesmo em tecidos tumorais, podendo ser hormônios, proteínas, peptídeos, etc., específicos ou não, que podem ser mensuradas e avaliadas como indicadores de processos biológicos normais, processos

patológicos ou de respostas farmacológicas a intervenções terapêuticas (KUMAR *et al.*, 2012).

Biomarcadores do câncer clinicamente úteis proporcionariam diagnóstico precoce e mais preciso, facilitando o estadiamento, o entendimento da patogênese do câncer e permitindo tratamentos que sejam iniciados em fases precoces da doença, nas quais a chance de serem mais eficazes aumenta (BLONDER *et al.*, 2011).

Para entender a função de todos os genes em um organismo, é necessário conhecer não só quais genes são expressos, quando e onde, mas também quais são os produtos da expressão e em que condições esses produtos (proteínas) são sintetizados. Assim sendo, um estudo proteômico apresenta o propósito de descrever o conjunto completo de proteínas presentes em determinado tecido, ou fluido, que são expressas sob determinadas condições, sejam elas normais ou atípicas. É válido ressaltar que o estudo do proteoma, que, em contraste com o genoma, é dinâmico e variável, depende de uma combinação de numerosos fatores internos e externos, como, por exemplo, estado fisiológico, comportamento alimentar, estresse, doenças e tratamentos médicos (BINNECK, 2004).

As diferenças entre células cancerosas e células sadias são, em grande parte, determinadas por expressão diferencial de proteínas, assim, é razoável estudar os mecanismos da patogênese do câncer em nível proteômico. Juntamente com a aplicação extensiva da teoria proteômica e as relevantes técnicas de pesquisas na área do câncer, a oncoproteômica tem sido uma área ativa de pesquisa médica, fazendo com que alguns autores caracterizem o câncer como uma doença proteômica (GUOQING *et al.*, 2011).

Diferentes fluidos biológicos têm sido explorados como fontes potenciais de biomarcadores, tais como urina, fezes, fluido cérebro-espinhal e sangue, alguns se apresentando como excelentes fontes de proteínas, padrões de expressão, que refletem de forma autêntica a extensão da doença. Para a maioria dos estudos de descoberta de biomarcadores, no entanto, o sangue (soro ou plasma) é a fonte primária para pesquisa devido a sua riqueza em proteínas secretadas (FANAYAN; HINCAPIE, HANCOCK, 2012).

Dessa forma, uma caracterização abrangente do proteoma sérico em estados saudáveis e doentes pode facilitar muito a detecção de biomarcadores. As chances de encontrar um novo biomarcador aumentam com o número de proteínas identificadas, no entanto, a caracterização do proteoma do soro ou plasma humano é uma tarefa muito difícil, pois as dez principais proteínas plasmáticas mais abundantes são responsáveis por aproximadamente 90% do teor proteico total, enquanto outras proteínas estão presentes em um intervalo muito dinâmico,

abrangendo mais de 10 ordens de grandeza em termos de concentração (LIUMBRUNO *et al.*, 2010).

E, por esse motivo, em quase todos os estudos do proteoma de soro ou plasma, o primeiro passo é a imunodepleção de proteínas de alta abundância, um passo fundamental para a detecção de proteínas de baixa abundância, que podem revelar-se como biomarcadores informativos. Diversos estudos sobre a eficiência, a reprodutibilidade e utilização de colunas de imunoafinidade tem sido publicados, pois só fazendo uso dessa técnica de imunodepleção os pesquisadores tem sido capazes de detectar um número substancial de proteínas de baixa abundância (100 ng/mL) (MILLIONI *et al.*, 2011).

Um dos grandes benefícios do investimento nos estudos da proteômica do câncer ao longo das últimas décadas tem sido a identificação de dezenas de biomarcadores potenciais para prognóstico e diagnóstico do câncer baseado em mecanismos de estudos da função de proteínas em células epiteliais normais e células cancerosas. No entanto, utilizar uma única proteína que combine especificidade e sensibilidade necessária para ser uma ferramenta clinicamente útil em ensaio é insuficiente. Atualmente, o antígeno prostático específico, mais conhecido como PSA, é considerado o melhor dos marcadores que consistem em uma única proteína, mas, por outro lado, uma única leitura de PSA pode ser difícil de interpretar, particularmente no que diz respeito à distinção entre a hiperplasia prostática benigna e o câncer de próstata, devido à elevada variabilidade nos níveis basais de PSA entre os indivíduos (RODLAND, 2004).

Tendo em vista que vários dos marcadores tumorais são também produzidos por tecidos não tumorais, como no caso dos níveis altos de PSA em doenças benignas da próstata, como foi dito anteriormente, e da alfa-fetoproteína (AFP), que é um biomarcador utilizado no diagnóstico de carcinoma hepatocelular, mas tem utilização limitada visto que níveis séricos normais de AFP são encontrados mesmo em alguns pacientes com o câncer, faz-se necessária a pesquisa não só para identificar proteínas singulares, mas também perfis de expressão que aumentem, de maneira significativa, a sensibilidade e especificidade do diagnóstico utilizando biomarcadores (AHMED, 2009).

Para o câncer de mama, já existem alguns biomarcadores teciduais, tais como receptores de estrógeno e progesterona, HER2/neu, p53, Ki-67/MIB-1 e fator de crescimento endotelial vascular (VEGF) e biomarcadores circulantes, tais como antígeno carcinoembriogênico (CEA), CA 15.3, CA549, CA27-29, TPA, antígeno mucina-like associado a carcinoma (MCA), catepsina D e chaperona 14-3-3 (ZENG *et al.*, 2010).

Apesar de vários biomarcadores tumorais terem sido identificados para o câncer de mama, a falta de sensibilidade e especificidade para o diagnóstico precoce é a principal desvantagem. Conseqüentemente, os marcadores disponíveis não se mostram úteis em qualquer rastreio ou diagnóstico do câncer de mama em estágio inicial. Estes marcadores tumorais se tornam úteis apenas quando somados a resultados de outros testes, tais como biópsia, mamografia, ultra-som e ressonância magnética da mama (KUMAR; KUMAR; SRIVASTAVA, 2012).

Claramente, novas estratégias têm de serem desenvolvidas, tanto para identificar as proteínas individuais envolvidas no início ou progressão do câncer, bem como novas abordagens que não se concentrem apenas em proteínas individuais, mas em padrões de co-expressão de proteínas que representem um sistema biológico característico do câncer de mama. E, para tal, técnicas como a espectrometria de massa e ferramentas de estatística e informática, necessárias para a interpretação de dados, abrem a perspectiva de um novo caminho para a identificação de marcadores tumorais (RODLAND, 2004).

### ***2.3. A glicosilação de proteínas***

O estabelecimento do câncer não está só correlacionado com fenótipos patológicos, sintomas clínicos e prognóstico, mas principalmente com alterações proteômicas, como foi dito anteriormente. E, além da expressão diferencial de muitas proteínas, as modificações pós-traducionais, tais como glicosilação, fosforilação e ubiquinação, também desempenham importantes papéis na gênese e desenvolvimento do câncer (GUOQING *et al.*, 2011).

Dentre as modificações pós-traducionais que ocorrem em proteínas, a glicosilação é uma das mais importantes, afetando a carga, a conformação e a estabilidade das proteínas, e desempenhando papéis cruciais nos processos biológicos, tais como reconhecimento celular, interação célula-célula, sinalização, desenvolvimento embrionário e reconhecimento de hormônios e toxinas (FANAYAN; HINCAPIE; HANCOCK, 2012).

Cerca de 60% das proteínas em células de mamíferos são glicosiladas em algum momento da vida, com um número estimado em 3.000 diferentes estruturas de glicanos registradas (sem incluir todas as variantes resultantes das diferenças nas ligações glicosídicas e anômeros), que podem variar em diferentes níveis, baseados em diferentes tecidos, tipo celular e estágio da doença (KUZMANOV; KOSANAM; DIAMANDIS, 2013).

Proteínas glicosiladas compreendem enzimas, anticorpos, hormônios, citocinas, receptores e várias proteínas estruturais. Sabe-se que 13 monossacarídeos e 8 aminoácidos

estão envolvidos em ligações glicoproteicas, enquanto que pelo menos 41 pontes, incluindo configurações anoméricas, ligações fosfoglicosil, C-manosilação, bem como GPI (âncoras glicosilfosfatidilinositol) podem ser formadas. Entre os monossacarídeos básicos para construção de glicanos maiores, os mais comuns são hexopiranosídeos, açúcares acetamidos, 6-deoxi-hexoses e ácido siálico (JANKOVIC, 2011).

A glicosilação não é um processo dirigido por um “*template*” e estima-se que o número de genes diretamente envolvidos na montagem de glicanos corresponda a cerca de 250 a 500, representado 1 a 2% do número total de genes humanos, codificando muitos tipos diferentes de enzimas: glicosiltransferases, sulfotransferases, oligossacariltransferases, glicosidases, lectinas, proteínas necessárias para um bom funcionamento do Complexo de Golgi, quinases, mutases, isomerases e demais enzimas que atuam sob carboidratos (SCHACHTER; FREEZE, 2009).

Existem diferentes tipos de glicosilação, incluindo N-glicosilação, O-glicosilação, C-glicosilação e S-glicosilação (encontrada apenas em bactérias). A N-glicosilação ocorre em resíduos de asparagina no motivo Asn-X-Ser/Thr (e ocasionalmente Cys) em que X pode ser qualquer resíduo de aminoácido com exceção da prolina. É iniciada no lado citoplasmático do retículo endoplasmático rugoso, e o glicano precursor mais bem conhecido é  $\text{Glc}_3\text{Man}_9\text{GlcNac}_2$ , e os N-glicanos podem ser categorizados pelo tipo e posição dos resíduos de monossacarídeos adicionados a esse precursor. Já resíduos de serina ou treonina podem ser O-glicosilados pela adição de N-acetilglucosamina, manose, fucose, glucose, N-acetilgalactosamina ou xilose. Por fim, a C-glicosilação envolve  $\alpha$ -manose C-ligada a um resíduo de triptofano. Vale ressaltar que muitos glicanos apresentam ainda outras modificações como sulfonação e fosforilação, o que o demonstra a complexidade e dinamicidade desse tipo de modificação pós-traducional (KIM; MISEK, 2011).

Modificações nos carboidratos anexados às proteínas são importantes em interações patógeno-hospedeiro, inflamação, desenvolvimento celular e malignidade. A glicosilação aberrante pode resultar em mudanças anormais na atividade/função biológica da proteína, no dobramento da mesma e no reconhecimento molecular na doença (KIM; MISEK, 2011).

A glicosilação anormal de proteínas já tem sido correlacionada com várias doenças tais como desordens hereditárias, alguns tipos de leucemia e de doenças hematopoiéticas, doenças por deficiência de carboidratos, diabetes, doenças inflamatórias, osteoartrite, trombose, doenças neurodegenerativas, doenças cardiovasculares e o câncer, e, dessa forma, proteínas glicosiladas tem sido alvos úteis para o desenvolvimento de novas vacinas, bem

como utilizadas como marcadores para a detecção precoce e diagnóstico de uma doença (LAZAR *et al.*, 2010).

O distúrbio metabólico celular consequência de uma doença pode resultar em perda ou ganho de sítios de glicosilação, que levam a uma mudança no seu número por proteína, à glicosilação alterada no sítio, tipo de glicanos presentes e as suas distribuições em sítios específicos, ou a quantidade de glicoproteínas específicas. Glicoproteínas derivadas de células tumorais, dependendo do contexto da doença, carregam formas truncadas de O-glicanos, níveis aumentados de ramificações de N-glicanos e elevada sialilação, sulfonação e fucosilação, indicando uma possível associação entre a doença e padrões irregulares de glicosilação (CHANDLER; GOLDMAN, 2013).

Já existem vários relatos na literatura sobre glicosilação e o câncer, mas seu papel e os mecanismos regulatórios complexos acerca da biossíntese desses glicanos alterados ainda não foram completamente elucidados. Geralmente, um padrão oligossacarídico relacionado ao câncer pode ser associado tanto com a expressão como com a secreção de carboidratos inapropriados, ou moléculas glicosiladas de forma incompleta, e, isso, conseqüentemente, gera novos antígenos. Muitas alterações acompanham a malignidade: mudanças na sialilação, fucosilação, galactosilação ou formas truncadas de N e O-glicanos. De maneira geral, uma atividade aumentada de sialitransferases e galactosiltransferases séricas é observada no câncer (JANKOVIC, 2011).

A grande maioria dos biomarcadores clínicos e alvos terapêuticos no câncer são glicoproteínas, tais como CA-125 no câncer de ovário, HER2/neu no câncer de mama e antígeno prostático específico (PSA) no câncer da próstata. O fator de crescimento epidérmico humano 2 (HER2/neu) é uma glicoproteína transmembranar e sua superexpressão parece ser um fator-chave na transformação maligna e é preditivo de um mau prognóstico no câncer de mama. Já CA-125 é uma mucina comumente utilizada como marcador de diagnóstico para o câncer epitelial de ovário. Embora o CA-125 já tenha sido usado como um marcador do câncer de ovário por um longo período de tempo, muitas das estruturas O- e N-glicanos só foram caracterizadas recentemente. O PSA é secretado primariamente pelas células epiteliais da próstata no plasma seminal e constitui-se de um dos melhores exemplos já caracterizados de uma glicoproteína secretada utilizada para diagnósticos de câncer, e as suas glicofomas têm sido descritas. As alterações na glicosilação da proteína consequência da variação da heterogeneidade de sítios de glicosilação ou mudanças na estrutura dos glicanos, tanto de proteínas da superfície celular como de proteínas presentes em fluidos corporais, estão correlacionadas com o desenvolvimento ou a progressão de câncer (MISEK; KIM, 2011).

Entre os diferentes tipos de modificações de oligossacarídeos associados ao câncer, a fucosilação é um dos tipos mais importantes. O carboidrato fucose (6-desoxi-L-galactose) é um monossacarídeo encontrado em glicoproteínas e glicolípidos presentes nos vertebrados, invertebrados, plantas e bactérias. A fucosilação, que compreende a transferência de um resíduo de fucose a oligossacarídeos e proteínas, é regulada por vários tipos de moléculas, incluindo fucosiltransferases, enzimas difosfato de guanosina (GDP)-fucose e transportadores GDP-fucose. Níveis de fucosilação em condições normais são relativamente baixos, mas aumentam durante a carcinogênese, e a maioria das glicoproteínas fucosiladas são proteínas secretadas ou proteínas de membrana na superfície da célula (MIYOSHI, *et al.* 2012).

Assim, os crescentes estudos que demonstram a clara associação entre glicosilação aberrante e doenças, entre elas o câncer, juntamente com a busca por biomarcadores, tem proporcionado o surgimento de uma subárea da proteômica caracterizada como glicoproteômica ou glicômica que visa à elucidação estrutural e caracterização de todas as glicofórmulas encontradas na natureza, e, de forma mais aprofundada, a dinâmica das mudanças espaço-temporal que estão associadas com processos biológicos (GUPTA; SUROLIA; SAMPATHKUMAR, 2010).

Com a finalidade de fracionar, ou mesmo isolar glicoproteínas ou glicoproteomas a partir de amostras biológicas complexas, muitos métodos tem sido desenvolvidos, entre eles o uso de lectinas, método químico baseado em hidrazida ou ácido borônico, cromatografia líquida de interação hidrofílica ou dióxido de titânio imobilizado, cada um com suas vantagens e limitações. Porém, a característica singular das lectinas de ligarem-se reversivelmente a especificamente a glicanos tem tornado essa classe de proteínas uma excelente ferramenta no estudo de glicoproteomas nas pesquisas biológica e médica.

#### **2.4. Lectinas**

Lectinas, palavra derivada do latim *legere*, que significa “escolher” ou “selecionar”, são proteínas ou glicoproteínas de origem não imune que se ligam especificamente e reversivelmente a carboidratos ou glicoconjugados, sem alterá-los quimicamente, e estão amplamente distribuídas na natureza, desempenhando diferentes papéis e funções em processos biológicos tais como reconhecimento de moléculas, proteínas de reserva e até mesmo em mecanismos de defesas em plantas. A interação com carboidratos se dá, principalmente, através de ligações de hidrogênio e interações de van der Waals, secundariamente por interações hidrofóbicas, e ocorrendo com menor frequência, por

interações eletrostáticas (PEUMANS, VAN DAMME, 1995; VAN DAMME *et al.*, 1998; RUDIGER, GABIUS 2001; FRANCIS *et al.*, 2011; VAN DAMME *et al.*, 2011).

As lectinas clássicas contêm dois ou mais sítios de ligação a carboidratos e, portanto, a sua interação com os açúcares presentes na superfície de eritrócitos resulta na reticulação de várias células do sangue e subsequente precipitação. Esse fenômeno, conhecido como aglutinação, é um atributo importante da atividade de lectinas e tem sido utilizado rotineiramente e classicamente para a detecção e caracterização dessa classe proteínicas, e pode ser inibido pela adição do carboidrato para o qual a lectina é específica (AMBROSI *et al.*, 2005).

De acordo com o monossacarídeo ligante para o qual as lectinas apresentam maior afinidade, as lectinas podem ser classificadas em cinco grupos, ligantes a: manose, galactose/N-acetilgalactosamina, N-acetilglucosamina, fucose e ácido N-acetilneuramínico. Com a exceção de fucose, todos estes açúcares estão na configuração D. Além disso, com algumas poucas exceções, nem lectinas específicas para glicose ligam-se a galactose, nem aquelas específicas para manose ligam-se a galactose. Curiosamente, certas variações na posição do C-2 do anel piranose podem ser toleradas, de modo que a maioria das lectinas que se ligam a galactose podem também interagir com N-acetilgalactosamina. Algumas lectinas interagem preferencialmente, ou com o anômero  $\alpha$  ou com o anômero  $\beta$ , enquanto em outras não há essa especificidade anomérica. Apesar de bem estabelecido e praticamente útil, esta categorização baseada na afinidade ao açúcar vem se tornando obsoleta. Pois, em primeiro lugar, existem diferenças marcantes nas finas especificidades de lectinas dentro de uma única categoria, e, em segundo lugar, um número crescente de lectinas que não apresentam alta afinidade para sacarídeos simples já foi identificado (AMBROSI; CAMERON; DAVIS, 2005).

Existe uma diversidade de lectinas de origem vegetal já isolada, caracterizadas e disponível comercialmente. Por muito tempo, as lectinas de plantas foram as únicas estruturalmente caracterizadas, e, especificamente, as lectinas de leguminosas, têm sido consideradas modelos para o estudo de interações proteína-carboidrato. O conhecimento detalhado da estrutura da proteína e da interação estabelecida com os carboidratos é um pré-requisito para a melhor compreensão dos processos biológicos e físicos de reconhecimento e para o desenvolvimento de processos biotecnológicos e aplicações terapêuticas (ARNAUD; AUDFRAYZ; IMBERTY, 2013).

As lectinas de plantas são divididas em 12 diferentes famílias de acordo com suas diferentes especificidades de interação com carboidratos e similaridade de sequências: (1)

homólogas da aglutinina de *Agaricus bisporus*, (2) Amarantinas, (3) Homólogas e Quitinases de classe V com atividade lectínica, (4) Família das Cianovirinas, (5) Família EEA, (6) Família GNA, (7) Proteínas com domínios heveínicos, (8) Jacalinas, (9) Proteínas com domínios lectínicos de leguminosas, (10) Domínio LysM, (11) Família Nictaba (formalmente lectinas do floema de Cucurbitaceae) e (12) Família Ricina-B. Entre as famílias de lectinas mencionadas, membros da família Ricina-B, proteínas com domínios lectínicos de leguminosas e da família GNA tem sido amplamente reportados demonstrando associação com diferentes processos patológicos, dentre eles o câncer (VAN DAMME *et al.*, 2008; LIU *et al.*, 2010).

A frutalina, uma lectina tetramérica isolada a partir de sementes de *Artocarpus incisa* ou “fruta-pão”, apresenta afinidade por resíduos de  $\alpha$ -D-galactose envolvidos em estruturas complexas de carboidratos, pertencendo à família de lectinas Jacalina-relacionadas, mais precisamente à subfamília de galactose-ligantes. Alguns estudos comprovam que a frutalina pode desempenhar diferentes aplicações biomédicas, por ser capaz de induzir a migração de neutrófilos em ratos *in vivo*, ser uma molécula quimiotática potente para neutrófilos e estimulador de linfócitos humanos, além de apresentar efeitos citotóxicos sob células tumorais em cultura *in vitro* (MOREIRA *et al.*, 1998; OLIVEIRA *et al.*, 2009; OLIVEIRA *et al.*, 2011).

A DAL, lectina isolada a partir de sementes de *Dioclea altissima*, uma leguminosa, apresenta afinidade para resíduos de D-glicose ou D-manose, e cátions divalentes são necessários para sua completa atividade. Estudos preliminares também mostram que a DAL apresenta citotoxicidade seletiva para algumas linhagens celulares cancerígenas (MOREIRA *et al.* 1997).

Lectinas tem sido utilizadas por décadas na glicobiologia, e, nos dias atuais, na glicoproteômica, para reconhecer e isolar motivos estruturais de glicanos específicos. A habilidade das lectinas de detectarem variações suaves nas estruturas de carboidratos encontrados em glicoproteínas e glicolipídios de superfície celular tem feito dessa classe de proteínas um paradigma para o reconhecimento proteína-carboidrato (FANAYAN; HINCAPIE; HANCOCK, 2012).

Dessa forma, aproveitando-se da habilidade das lectinas de interagirem com carboidratos, essas proteínas podem ser utilizadas para capturar diferentes glicoproteomas, e, assim, estudos recentes utilizando, em uma primeira etapa, arranjos de lectinas e cromatografia de afinidade, e posteriormente, cromatografia líquida acoplada a espectrometria de massas, avançam no uso das lectinas para a prospecção de novos biomarcadores.

Com essa finalidade, as lectinas podem ser imobilizadas em diferentes suportes, incluindo agarose, sepharose, polímero polihidroxilado (POROS) e sílica por meio de imobilização covalente. Muitas estratégias experimentais utilizam lectinas imobilizadas disponíveis comercialmente em matrizes de agarose para uso de cromatografias de bancada, sob baixas condições de pressão. As principais desvantagens dessa matriz, contudo, são as condições cromatográficas restritas devido ao fluxo utilizado e limitações de pressão, o que pode aumentar significativamente o tempo de análise e manipulação da amostra levando a perdas amostrais e podendo ter consequências na análise proteômica *a posteriori*. Quando um maior grau de reprodutibilidade e de velocidade é desejado, lectinas podem ser conjugadas a matrizes compatíveis a HPLC (*High-performance liquid chromatography*), que consiste em pequenas e rígidas partículas feitas com base em sílica ou polímeros sintéticos, capazes de suportar o alto fluxo e as altas pressões, que são características dos sistemas de HPLC. Embora essa abordagem ofereça muitas vantagens, tais como perfis de eluição mais nítidos, análises mais rápidas, e a possibilidade de interface com vários sistemas de alta pressão, o que permite alto rendimento e a automatização, ela apresenta algumas desvantagens, tais como as limitações de pH e interações não-específicas, o que limita a sua utilidade em certas aplicações (FANAYAN; HINCAPIE, HANCOCK, 2012).

Ferreira e colaboradores (2011), por exemplo, desenvolveram uma plataforma para capturar glicoproteínas baseada em nanossondas magnéticas revestidas com três lectinas diferentes: concanavalina A (ConA), aglutinina de gérmen de trigo (WGA) e lectina de *Maackia amurensis* (MA), visando capturar, de forma seletiva, glicoproteínas de fluidos corporais humanos. Condições operacionais ótimas, como temperatura e tempo, e as capacidades máximas de ligação foram determinadas para cada lectina, utilizando fetuína como referência. Estudos comparativos usando as nanossondas desenvolvidas e as lectinas imobilizadas em Sepharose demonstraram um desempenho superior da nanoplataforma, exibindo uma afinidade 5 vezes maior, quando comparado com ConA-Sepharose com a mesma quantidade de lectina imobilizada.

Portanto, sabendo que o desenvolvimento e progressão do câncer estão intimamente relacionados com expressão anormal de proteínas, modificações pós-traducionais alteradas e interação entre proteínas e que alterações aberrantes nas proteínas podem levar diretamente a características malignas e transformação oncogênica, a utilização de diferentes abordagens proteômicas aplicadas ao estudo dessas alterações e, conseqüentemente, nas vias de sinalização, podem ajudar os cientistas a compreender melhor os mecanismos da

tumorigênese, identificar novos marcadores diagnósticos e alvo de drogas, além de fornecer uma base clínica para diagnóstico e tratamento precoce (GUOQING *et al.*, 2011).

### **2.5. Abordagens proteômicas na busca por biomarcadores**

Tendo em vista que a proteômica pode fazer a ponte entre as alterações genéticas subjacentes ao câncer e a fisiologia celular, muito se espera das análises proteômicas principalmente utilizando a espectrometria de massas, para a prospecção de biomarcadores (GAST; SCHELLENS; BEIJNEN, 2008).

A espectrometria de massas tem sido usada como ferramenta de diagnóstico em laboratórios clínicos por muitas décadas. Essa tecnologia acoplada à cromatografia gasosa foi muito utilizada, com sucesso, para identificação e quantificação de moléculas, relativamente pequenas (com massa molecular < 1.000 Da), altamente informativas, em programas de rastreamento de recém-nascidos, aplicações toxicológicas e forenses, identificação de erros congênitos do metabolismo, detectar *dopping* em atletas, etc. E, ao longo de 15 anos, tem-se percebido a ressurgência dessa tecnologia para estudo de moléculas maiores, como as proteínas. Novas aplicações tornaram esse estudo possível, devido ao desenvolvimento de novas metodologias que volatizam e ionizam efetivamente as proteínas, por meio do uso de matrizes químicas e lasers. A habilidade de mensurar com elevada acurácia a razão massa/carga, fornecer um espectro de alta resolução e o desenvolvimento da espectrometria de massas em *tandem*, ou MS/MS, a fim de obter o sequenciamento *de novo* de uma proteína tem aumentado de força significativa a aplicação dessa tecnologia na proteômica. Atualmente, o uso combinado da cromatografia líquida acoplada ao espectrômetro de massa expandiu ainda mais o poder discriminatório desse método, caracterizando-o como uma das ferramentas mais poderosas da proteômica (DIAMANDIS, 2004).

Pesquisadores pioneiros no uso da espectrometria de massas como ferramenta de diagnóstico sugerem que essa abordagem representa uma mudança de paradigma no diagnóstico do câncer, baseado nas diferenças complexas entre padrões proteômicos do soro de paciente com e sem câncer, identificadas por bioinformática. Eles baseiam-se na premissa de que não importa qual a natureza dessas moléculas, mas o potencial de discriminação entre duas condições deve ser explorado a fundo. A hipótese central dessa abordagem é a seguinte: proteínas ou fragmentos proteicos produzidos por células cancerosas ou nos microambientes nos quais essas células estão inseridas podem eventualmente entrar na circulação sistêmica. E

assim, a concentração dessas proteínas ou desses fragmentos, pode ser analisada por espectrometria de massas e usada para fins de diagnóstico (PETRICOIN III *et al.*, 2002).

Portanto, os recentes avanços técnicos em espectrometria de massas oferecem oportunidade para encontrar biomarcadores em diferentes amostras biológicas. E apesar de o soro humano ser o fluido biológico mais comum para diagnóstico, pois reflete a doença em diferentes órgãos e tecidos através de mudanças no perfil de proteínas secretadas, ainda é desafiante a caracterização de glicoproteínas séricas, o "glicoproteoma" do soro (YANG & HANCOCK, 2004, PIETROWSKA *et al.*, 2009).

O desenvolvimento de uma abordagem analítica para o estudo da variação de padrões de glicosilação em amostras de soro tem sido dificultado pela heterogeneidade estrutural desse tipo de modificação pós-traducional e da complexidade do proteoma do soro. Yang e Hancock (2004) tem proposto a utilização de lectinas para reconhecimento de motivos específicos de glicosilação diferentes a fim de desenvolver uma forma de capturar diversas glicoproteínas séricas. Em uma investigação preliminar, foi avaliada a capacidade de cinco lectinas vegetais, comumente utilizadas, imobilizadas, para capturar glicoproteínas de soro humano e os pesquisadores observaram, por análise em SDS-PAGE, que cada lectina utilizada foi capaz de reunir um subconjunto diferente do glicoproteoma sérico.

Baseada nessa abordagem proteômica utilizando lectinas vegetais para prospecção de biomarcadores, Drake e colaboradores (2011) descreveram uma estratégia experimental que consistia na cromatografia de afinidade utilizando a aglutinina de *Sambucus nigra* (SNA) e a lectina de *Aleuria aurantia* (AAL), lectinas que apresentam afinidade de ligação a estruturas contendo o ácido siálico e a fucose, respectivamente, e posterior análise por espectrometria de massas. As amostras proteicas foram previamente digeridas com tripsina e os glicopeptídeos isolados nessas cromatografias foram então deglicosilados, e em seguida, os peptídeos analisados por espectrometria de massa. Um total de 122 glicoproteínas de plasma humano, contendo 247 glicosítios únicos, foi identificado exclusivamente no meio de cultura de células do câncer de mama de um tipo mais agressivo.

Apesar de algumas limitações, abordagens glicoproteômicas com foco em glicopeptídeos vêm ganhando bastante popularidade. As análises por espectrometria de massa de glicopeptídeos ou glicoproteínas são gratificantes, pois têm, intrinsecamente, o potencial de relacionar estruturas de glicanos específicos para um determinado sítio de uma proteína específica. Esta informação é, muitas vezes, de extrema importância, já que o principal papel dos glicanos está em modular as propriedades (tais como a função, atividade, estabilidade) das proteínas que os carregam (WUHRER, 2013).

De fato, glicoproteínas já tem sido usadas como biomarcadores para diagnóstico e prognóstico e, hoje em dia, os pesquisadores estão cientes de que glicanos são únicos por sua potencialidade bioinformativa *in vivo* e, portanto, constiuem-se de alvos para testes de diagnósticos originais e sensíveis, que incluem a utilização isolada ou dentre um painel de marcadores tumorais em diferentes plataformas. A descoberta de biomarcadores é um processo de múltiplas etapas e dados glicômicos mais atuais estão principalmente relacionados com a chamada fase de descoberta, para, só em seguida, proceder-se às etapas de validação e padronização. Como mudanças estruturais e funcionais relacionadas às doenças já vem sendo bem definidas, o número de técnicas visando o uso na química clínica tende a se expandir. Portanto, atualmente, existe uma necessidade constante para a cooperação entre cientistas da glicobiologia e glicotecnologia almejando o desenvolvimento de uma plataforma rápida e de baixo custo que seja aplicável ao cotidiano de práticas laboratoriais na detecção de biomarcadores (JANKOVIC, 2011).

## **2.6. A abordagem proteômica quantitativa label-free LC-MS<sup>E</sup> utilizada no estudo**

O número de potenciais proteínas marcadoras que pode ser usado para o diagnóstico, detecção precoce, caracterização da doença ou monitoramento do tratamento, continua a crescer rapidamente na literatura científica, no entanto, a dificuldade de aplicação clínica é ainda grande, por causa dos baixos níveis de sensibilidade e especificidade de algumas delas. E a técnica de quantificação por espectrometria de massas (MS) auxilia bastante na descoberta de biomarcadores, tendo em vista que biomarcadores muito provavelmente não são proteínas específicas isoladas, mas sim um conjunto de proteínas ou peptídeos que tem um padrão de expressão único em resposta a um estímulo ou condição.

E, nesse contexto, há alguns anos o campo da proteômica foi além da simples identificação de proteínas... Agora é levado a sério quanto à precisão e confiabilidade para quantificar as diferenças, mesmo que mínimas, na abundância de diferentes proteínas em um organismo, célula ou tecido em um determinado momento ou sob uma determinada condição. Dessa forma, a proteômica quantitativa é atualmente dividida em duas grandes abordagens: livre de marcação, ou mais comumente chamada de *label-free*, ou com utilização de marcação (NEILSON *et al.*, 2011).

As técnicas de marcação envolvem a modificação de peptídeos com etiquetas isobáricas para quantificação relativa e absoluta – iTRAQ (ROSS *et al.*, 2004), marcação de

proteínas com etiquetas isotópicas de afinidade – ICAT (GYGI *et al.*, 1999) ou proteínas metabolicamente marcadas pela incorporação de isótopos em aminoácidos inseridos na cultura celular – SILAC (ONG *et al.*, 2002).

Estratégias de marcação são, muitas vezes, consideradas mais precisas na quantificação de proteínas, entretanto, estas técnicas exigem sondas caras, *softwares* específicos, e *expertise* para analisar os dados. Além disso, algumas estratégias de marcação não podem ser facilmente aplicadas a todos os tipos de amostras (NEILSON *et al.*, 2011).

Nos últimos anos, a observação de uma correlação entre abundância de proteínas e áreas de picos cromatográficos ou número de espectros MS/MS, dos peptídeos pertencentes àquelas proteínas, ampliou a possibilidade de escolha do procedimento analítico no campo da proteômica quantitativa. Sendo assim, uma abordagem acurada e livre de marcação, por isso sendo denominada *label-free*, para MS quantitativa, envolve a medição integrada das áreas dos picos cromatográficos para cada peptídeo isolado por cromatografia líquida acoplada ao espectrômetro de massas em associação com as intensidades dos peptídeos oriundos da digestão (CHELIUS; BONDARENKO, 2002). Foi essa a técnica utilizada no presente trabalho.

Explicando melhor, a mensuração da área de pico cromatográfico (também chamada pela sigla AUC) é linearmente proporcional à concentração do peptídeo, isso para peptídeos que estejam em um intervalo ideal de 10 fmol a 100 pmol, demonstrado por alguns estudos que relatam boa detecção de proteínas, previamente selecionadas, em misturas complexas em um intervalo de concentração de 10 fmol a 1000 pmol. O processo geral envolve também a medição da abundância (intensidade) de íons em tempos de retenção específicos para os peptídeos ionizados sem o uso de marcação isotópica, o que é frequentemente referido como contagem de íons (*Total ion counting* - TIC) (PODWOJSKI *et al.*, 2010).

Embora o método seja relativamente simples conceitualmente, várias considerações devem ser levadas em conta: preocupações com a resolução da cromatografia, pois sinais peptídicos podem apresentar-se espalhados ao longo do tempo de retenção causando sobreposição com peptídeos co-eluídos naquele mesmo tempo; variações biológicas que resultam em múltiplos sinais para o mesmo peptídeo (diferentes modificações pós-traducionais, por exemplo), bem como variações técnicas (do próprio equipamento naquele dia de corrida) no tempo de retenção ou na intensidade dos picos, e ruído de fundo por interferentes químicos na amostra, o que requer a necessidade de “limpar” computacionalmente os dados de LC-MS para que seja possível o alinhamento de tempos de

retenção e a supressão de ruído permitindo, assim, a avaliação comparativa dos peptídeos ao longo de todo experimento (LISTGARTEN; EMILI, 2005).

Logo, um desafio quando se utiliza essa técnica AUC, ou seja, de mensuração da área de pico cromatográfico para correlacionar com abundância peptídica, é encontrar um equilíbrio na energia de colisão para obter quantificação precisa e maximizar a identificação de proteínas. Quando se usa uma aquisição dependente de dados, mais comumente chamada de DDA na espectrometria de massas, há uma alternância entre modos MS e MS/MS no equipamento e, para isso, há muitas vezes o “sacrifício” do íon precursor ou mesmo de íons fragmentos, pois naquele momento que o equipamento está promovendo a colisão do íon precursor (MS/MS), para gerar íons fragmentos, alguns outros íons precursores podem ser perdidos (RADULOVIC *et al.*, 2004).

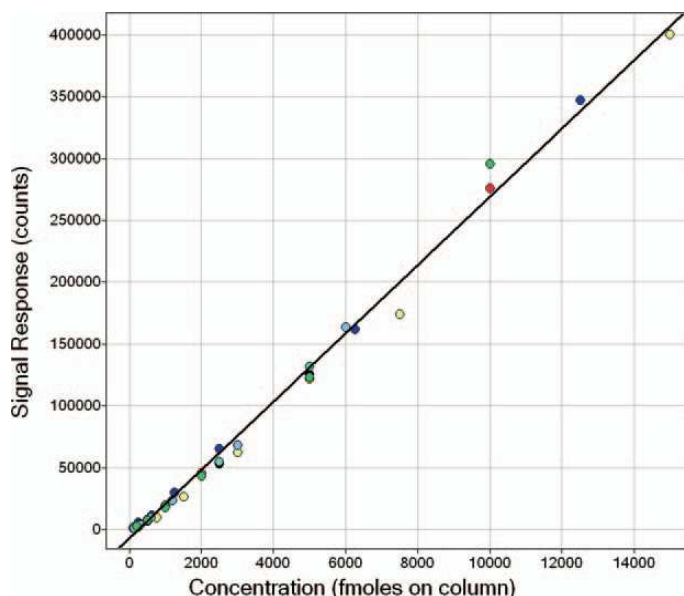
Dessa forma, uma maneira de superar isso é executar múltiplas análises de uma amostra em modos MS e MS/MS separadamente e subsequentemente corresponder à massa exata e o tempo de retenção à identificação de peptídeos (SILVA *et al.*, 2005). Alternativamente, dados quantitativos e qualitativos de proteínas podem ser obtidos a partir de uma única aquisição independente de dados promovida pela alternância entre condições de baixa e alta energia de colisão no espectrômetro de massa através do qual todos os peptídeos são fragmentados sem seleção de íon precursor. Isto foi denominado MS<sup>E</sup>, onde esse “E” indica energia elevada, e está disponível como um modo de aquisição de dados exclusivo de instrumentos Waters Q-TOF, espectrômetro de massas utilizado nesse trabalho (SILVA *et al.*, 2006; GEROMANOS *et al.*, 2009).

Tem sido demonstrado que LC-MS<sup>E</sup> é a técnica *label-free* mais precisa para estimar a abundância absoluta, que utiliza a média da intensidade dos três peptídeos trípticos mais abundantes por proteína (técnica Hi-3 ou Top-3), que é diretamente proporcional à quantidade da proteína (SILVA *et al.*, 2006).

A técnica Hi-3 baseia-se na quantificação absoluta de proteínas com o uso de um ou mais peptídeos externos de referência para gerar uma curva-resposta de calibração para polipeptídeos específicos a partir dessa proteína. A figura 1 ilustra uma correlação linear entre os três peptídeos trípticos mais intensos de diferentes proteínas e suas concentrações absolutas correspondentes, no trabalho desenvolvido por Silva e colaboradores (2006), no qual é descrito o método Hi-3 para quantificação.

Nesse sentido, a quantificação absoluta de uma dada proteína é determinada a partir do sinal observado para um polipeptídeo específico da amostra em relação ao gerado na curva de calibração. Se a quantificação absoluta de um número de diferentes proteínas deve ser

Figura 1 – Curva universal de resposta-sinal para quantificação absoluta de proteínas padrão (ou referência).



Resposta média dos sinais para os três peptídeos trípticos mais intensos para cada uma das seis proteínas estudadas. Um único gráfico de dispersão da resposta média dos sinais (eixo Y) e a concentração correspondente da proteína (eixo X) foi produzido para todas as proteínas e encontrado sendo linear para mais de 2 ordens de magnitude. As esferas coloridas seguem o seguinte código: vermelho, álcool desidrogenase de levedura (ADH); azul escuro, albumina de soro bovino; amarelo, enolase; preto,  $\alpha$ -hemoglobina bovina; azul claro,  $\beta$ -hemoglobina bovina; verde,  $\alpha/\beta$ -hemoglobina bovina; e cinza, fosforilase B rábica. O ajuste da curva linear foi calculado para todo o conjunto de dados ( $y = 27,6 \times (X) - 7401$ ,  $R^2 0,9939$ ). Fonte: SILVA *et al.*, 2006.

determinada, curvas de calibração separadas são necessárias para cada peptídeo específico da proteína padrão (ou referência externa) para cada proteína (SILVA *et al.*, 2006).

E, assim, a quantificação absoluta permite não só determinar alterações entre duas condições, como também permite comparações quantitativas de proteínas dentro de uma mesma amostra.

Devido ao modo de aquisição de dados  $MS^E$ , todos os peptídeos trípticos que apresentem intensidade suficiente irão produzir dados de fragmentação dos íons produtos que poderão, então, ser utilizados para identificação estrutural do referido peptídeo e sua proteína parental correspondente. A partir desta análise, os dados de MS dos peptídeos trípticos de uma proteína e as informações de sequência associadas a eles, a partir dos dados de MS/MS para cada peptídeo, irá gerar uma análise abrangente para cada proteína presente na mistura (SILVA *et al.*, 2006).

Portando, o método *label-free* descrito por Silva e colaboradores (2006), e que foi o método utilizado no presente trabalho, é ideal para a determinação da concentração absoluta de proteínas presentes em amostras complexas. A técnica Hi-3 acima descrita usufrui o máximo possível do modo de aquisição  $MS^E$  e sua capacidade de reduzir, exaustivamente,

dezenas de milhares de detecções de íons a uma lista simples de peptídeos precursores juntamente com o seus íons produtos em seus tempos específicos de retenção.

A especificidade conferida pelas medições acuradas de massa, tipicamente menor que 5 ppm, de ambos os íons precursores e íons fragmentos associados, permite a capacidade de identificar, com alta confiança, um grande número de proteínas com alta cobertura de sequência. A capacidade de coletar os dados de MS ao longo de toda a largura do pico cromatográfico, para todos os peptídeos acima do limite de detecção do instrumento, permite a quantificação precisa de peptídeos/proteínas a partir das intensidades de sinal deconvoluídas. Portanto, esses atributos da aquisição de dados LC-MS<sup>E</sup>, em associação com a correlação entre a média do sinal-resposta MS dos três peptídeos ionizantes mais abundantes de uma proteína padrão, fornecem um meio para determinar a concentração absoluta de qualquer proteína bem caracterizada presente numa amostra (SILVA *et al.*, 2006).

Um estudo comparativo realizado por Turck e colaboradores (2006) com 52 grupos participantes teve o objetivo de avaliar metodologias disponíveis para quantificar oito proteínas diferentes em uma mistura proteica. Os grupos participantes receberam amostras contendo concentrações variáveis de proteínas padrão, mais uma quantidade conhecida de albumina sérica bovina (BSA) para a normalização de dados, e foram capazes de escolher o seu próprio método de preferência para análise. Dentre os participantes, 37% escolheram técnicas que utilizaram eletroforese bidimensional (2D), 42% escolheram a técnica *shotgun* utilizando marcação e 21% optaram pela técnica *shotgun* sem marcação (*label-free*).

O estudo demonstrou uma variabilidade considerável nos resultados de quantificação e métodos baseados em gel exibiram a maior inconsistência de erro. Razões obtidas por *label-free* MS AUC, ou métodos baseados na contagem de espectros, estavam mais próximos dos valores esperados, em comparação com as proporções obtidas por métodos de marcação isotópica. Os resultados com marcação isotópica exibiram uma distribuição uniforme da percentagem de erro entre todas as proteínas, exceto para a proteína com a concentração mais baixa, enquanto que a AUC e método de contagem de espectros mostraram uma distribuição igual de erro percentual para todas as proteínas. A disparidade entre as razões previstas e observadas foi encontrada sendo dependente tanto do método como para proteínas específicas.

A realização de análises em triplicata garante melhores resultados, no entanto, nenhum aumento significativo na precisão dos dados foi observado quando se aumentou o número de replicatas para além de 3. E as triplicatas das amostras que utilizaram marcação isotópica exibiram maior consistência em comparação entre replicatas da abordagem *label-free*.

Em outro estudo, realizado por Ryu e colaboradores (2008), foi avaliado o desempenho dos métodos utilizando marcação (ICAT) e *label-free* AUC e *label-free* por contagem de espectros através da análise de três misturas proteicas contendo sete proteínas padrão em várias concentrações, variando de proteínas com concentrações similares a diferença de duas ou quatro vezes. Os resultados mostraram que ambos os métodos *label-free* eram tão precisos como ICAT em detectar as proteínas padrão. No entanto, o método ICAT falhou na detecção precisa de proteínas com diferença de quatro vezes em concentração.

Outro estudo selecionou três técnicas proteômicas quantitativas: *label-free* AUC, *label-free* LC-MS<sup>E</sup>, e iTRAQ, para identificação e quantificação relativa de proteínas de uma bactéria metanotrófica cultivada em dois substratos diferentes. Como resultado, o número total de proteínas identificadas em cada método foi semelhante com 49% de proteínas comuns a todas as técnicas. No entanto, LC-MS<sup>E</sup> proporcionou maior cobertura de sequência e um maior número de peptídeos médios identificados por proteínas em comparação com os outros dois métodos, que foram bastante semelhantes entre si no que diz respeito a estas medidas (PATEL *et al*, 2009).

A variabilidade é um componente intrínseco e, portanto, inevitável, de cada experimento e um conceito central para as análises estatísticas. Amostras biológicas têm variabilidade genética e ambiental natural. Além do que, métodos de preparação de amostras e aquisição de dados são fatores dependentes da acurácia instrumental e da própria manipulação humana. Dessa forma, para ter acesso à significância da alteração na quantidade de uma proteína mensurada, a análise estatística deve ser procedida. Assim sendo, é fato que a proteômica quantitativa *label-free* produz um grande volume de dados que exigem avaliação estatística. A estrutura dos dados brutos (*raw data*) adquiridos pela técnica MS<sup>E</sup> não é padrão e, portanto, requer *software* feito sob medida para processá-lo (EYERS; GASKELL, 2014).

Atualmente, o único *software* capaz de processar esses dados é o *IdentityE*, incluso no *ProteinLynx global Server* (PLGs, Waters Corp., Milford MA, EUA), que foi o programa utilizado na metodologia de processamento de dados do presente estudo, como será comentado adiante. O *software* utiliza um algoritmo de detecção de pico chamado Apex3D e um algoritmo de busca em banco de dados chamado *Ion Accounting*, ambos os quais são descritos em pormenores por Geromanos e colaboradores (2011).

Diante do exposto, a técnica escolhida *label-free* LC-MS<sup>E</sup>, juntamente com o pacote de *softwares* apropriados utilizados no presente estudo, ofereceu todo o suporte técnico e analítico para produzir resultados estatisticamente confiáveis.

### 3. HIPÓTESES

Considerando a habilidade das lectinas de interagirem de forma específica e reversível com carboidratos/glicoconjugados e de relatos na literatura que fazem uso dessa classe de proteínas para capturar outras que estejam diretamente relacionadas ao câncer, o presente trabalho visou confirmar a seguinte hipótese:

Diferentes (glico)proteínas plasmáticas de pacientes com câncer de mama podem ser isoladas por meio de cromatografias de afinidade em colunas de lectinas vegetais imobilizadas ( $\alpha$ -D-galactose-ligante de *Artocarpus incisa* – Frutalina e glucose/manose-ligante de *Dioclea altissima* - DAL), o que é útil para redução da complexidade da amostra e posterior análise por espectrometria de massas e, algumas dessas proteínas isoladas, identificadas e quantificadas, estão diretamente relacionadas com o câncer de mama, em diferentes estágios da doença, apresentando-se como proteínas alvo para subsequente validação.

## **4. OBJETIVOS**

### **4.1. Objetivo Geral**

Identificar e quantificar proteínas plasmáticas de pacientes portadoras de câncer de mama do tipo ductal em diferentes estágios (I, II e III), utilizando espectrometria de massas, fracionadas por meio de cromatografias de afinidade em duas matrizes de lectinas vegetais imobilizadas.

### **4.2. Objetivos Específicos**

- Fracionar proteínas plasmáticas de mulheres sem e com câncer de mama, através de cromatografia de afinidade em duas diferentes colunas de lectinas imobilizadas: FTL-Sepharose e DAL-Sepharose;
- Identificar e quantificar as proteínas plasmáticas presentes nas amostras totais de plasma;
- Identificar e quantificar as proteínas plasmáticas presentes nas frações após cromatografias de afinidade em FTL-Sepharose e DAL-Sepharose;
- Avaliar a necessidade e importância do uso das lectinas vegetais para o fracionamento das amostras e redução da complexidade;
- Avaliar a expressão de proteínas relacionando-as com diferentes níveis de estadiamento do câncer de mama;
- Discutir o envolvimento das proteínas super ou sub expressas, ou com alteração no padrão de glicosilação, no desenvolvimento do câncer de mama.

## 5. ESTRATÉGIA EXPERIMENTAL

O câncer de mama é tradicionalmente considerado como uma doença bastante heterogênea, com características biológicas e clínicas diversas para os diferentes subtipos. Perfis moleculares do câncer de mama tem fornecido uma ferramenta importante para discriminar uma série de subtipos, subtipos esses associados com dados clínicos, diagnóstico precoce e resposta ao tratamento.

Dessa forma, visando à identificação de novas proteínas alvos, relacionadas ao câncer de mama em diferentes estágios, foi realizada uma análise quali e quantitativa, por espectrometria de massas, de proteínas plasmáticas, fracionadas por meio de cromatografia de afinidade em colunas de lectinas vegetais imobilizadas, de mulheres com câncer de mama em diferentes estágios da doença.

Para isso, inicialmente, as lectinas  $\alpha$ -D-galactose-ligante das sementes de *Artocarpus incisa* (frutalina) e glucose/manose-ligante de *Dioclea altissima* (DAL) foram isoladas por meio de cromatografia de afinidade em colunas de D-galactose-Agarose e Sephadex G5015, respectivamente. A confirmação da purificação das lectinas foi confirmada por meio de eletroforese em gel de acrilamida (SDS-PAGE) e análise por espectrometria de massa. Em seguida, as proteínas purificadas foram imobilizadas em matriz de Sepharose-4B ativada com brometo de cianogênio.

Para etapa de imunodepleção de cada amostra individual, foi utilizada a coluna *HiTrap IgG&Albumin Depletion*, acoplada a um sistema de FPLC *ÄKTApurifier 10*, para retirada de IgG e albumina das amostras de plasma.

Após a etapa de depleção, e anteriormente à etapa de cromatografias de afinidade em colunas de lectinas vegetais, as amostras foram, individualmente, quantificadas por espectrofotometria e, em seguida, foram preparados *pools*, com contribuições homogêneas de massa proteica de cada amostra individual, correspondentes a cada grupo em estudo (Controle, Câncer de Mama Ductal I/Estágio Precoce, Câncer de Mama Ductal II/ Estágio Intermediário e Câncer de Mama Ductal III/Estágio Avançado).

Em seguida, para fracionamento das proteínas presentes no plasma de mulheres sem câncer de mama e com câncer de mama, os *pools* de amostras depletadas foram aplicados, separadamente, nas diferentes colunas de afinidade com lectinas imobilizadas: FTL-Sepharose e DAL-Sepharose, acopladas a um sistema de FPLC *ÄKTApurifier 10*.

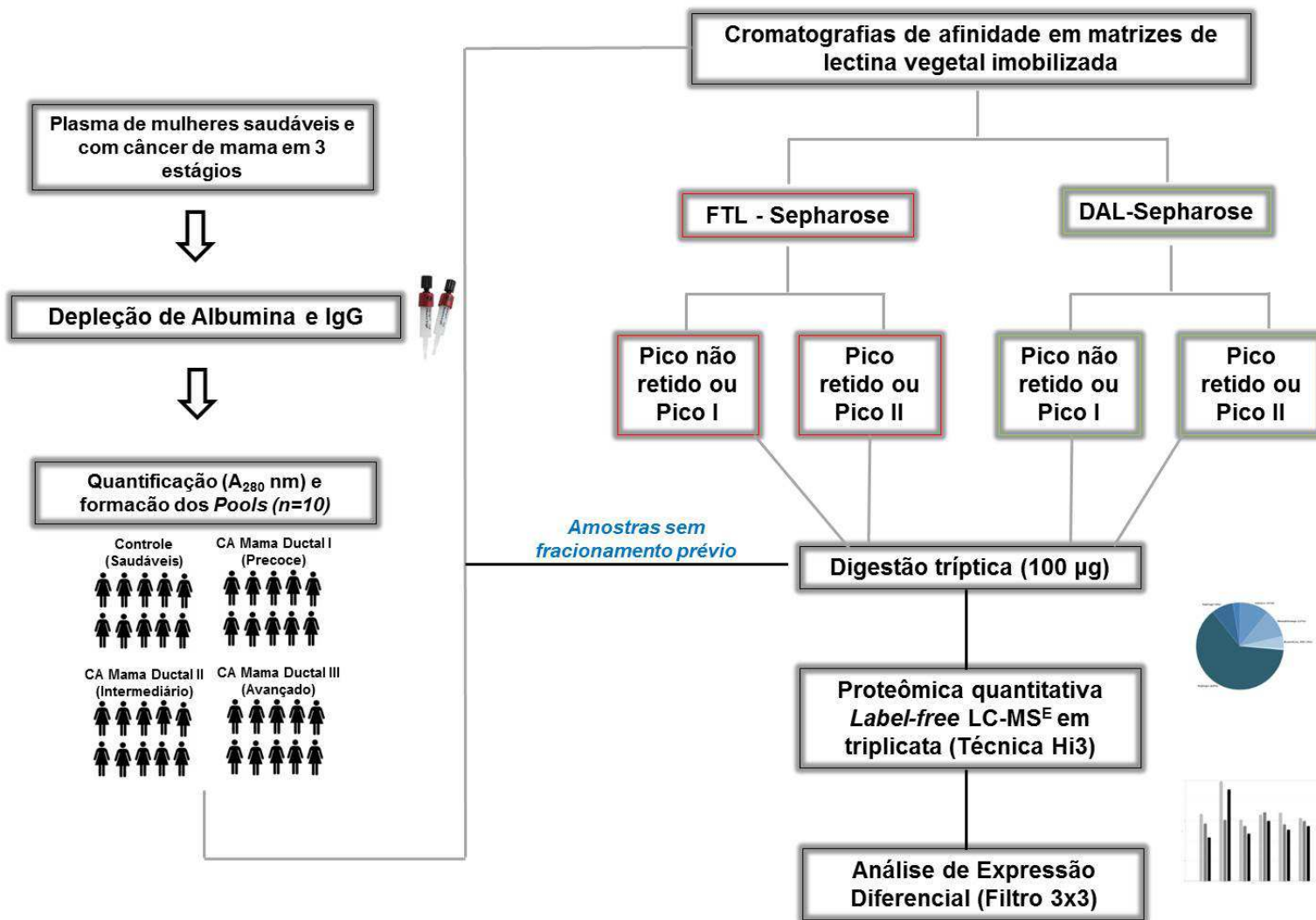
As frações cromatográficas obtidas, após cromatografias de afinidade em colunas de lectinas vegetais imobilizadas, foram então dialisadas, concentradas e submetidas à reação de digestão com a enzima tripsina e, posteriormente, os dados de massas MS e MS/MS adquiridos em espectrômetro de massas nanoESI, *Synapt HDMS* (Waters, Manchester, UK), acoplado a um sistema de cromatografia nanoUPLC.

Em paralelo, amostras dos *pools*, sem fracionamento prévio em colunas de lectinas vegetais imobilizadas, foram também submetidas à reação de digestão e aplicadas no espectrômetro de massas.

A análise dos dados, que consiste na identificação de proteínas (qualitativo) e quantificação absoluta e relativa (quantitativo), foi realizada com o uso de *softwares* próprios e busca em banco de dados espécie-específicos. Os níveis de expressão das proteínas identificadas foram avaliados utilizando o pacote *ProteinLynxGlobalServer* (PLGs) v.2.4. contendo o programa *Expression*<sup>E</sup> v.2.4., que realizou a análise estatística dos resultados, e posteriormente foram relacionadas com diferentes níveis de estadiamento do câncer de mama.

De maneira resumida, a Figura 2 ilustra a estratégia experimental utilizada no presente trabalho.

Figura 2 – Fluxograma representativo da estratégia experimental utilizada no presente estudo



## **6. MATERIAIS E MÉTODOS**

### ***6.1. Materiais***

Este trabalho envolveu a coleta de sangue humano, e, portanto, foi submetido e aprovado pelo Comitê de Ética em Pesquisa do Hospital Geral de Fortaleza (050507/10 – Adendo aprovado em Out/12). A coleta de sangue e de dados foi realizada somente após obtenção da assinatura do termo de consentimento livre e esclarecido pelas mulheres.

Essa pesquisa foi desenvolvida de forma a incorporar, sob a ótica do indivíduo e das coletividades os quatro referenciais básicos da bioética: autonomia, não maleficência, beneficência e justiça; assegurando os direitos e deveres que dizem respeito aos pesquisadores e aos sujeitos da pesquisa, e todos os procedimentos de obtenção e divulgação de informações seguiram as normas estabelecidas pelo Conselho Nacional de Saúde, item de Ética em Pesquisas com Humanos, resolução 466/12.

#### ***6.1.1. Grupo de estudo***

A triagem de pacientes foi realizada no Hospital Geral de Fortaleza - HGF (Fortaleza, Ceará).

##### ***6.1.1.1. Critérios de inclusão***

Um total de 40 mulheres, de 33 a 78 anos, que concordaram em participar da pesquisa, foram agrupadas em grupo controle (10 mulheres) e grupo caso (30 mulheres), cujo tumor foi enviado para biópsia no Laboratório Biopse, no Hospital Geral de Fortaleza, e diagnosticada portadora de carcinoma ductal da mama. As amostras, que consistiram em sangue total, foram obtidas junto ao laboratório, fiel depositário e responsável pela coleta de materiais biológicos e diagnóstico inicial da doença, no período de maio de 2011 a dezembro de 2012.

Para formação do grupo CONTROLE, foram consideradas elegíveis mulheres sem diagnóstico de câncer de mama ou qualquer outra neoplasia, avaliadas por meio da história do prontuário e de exame clínico realizado por médico colaborador do estudo.

Já para o grupo CASO, foram consideradas elegíveis as mulheres com recém-diagnóstico clínico e anatomopatológico de neoplasia mamária tipo ductal, estadiamento I, II e III, sem neoplasias associadas, sem tratamento antineoplásico prévio e com índice de

Karnofsky acima de 70. O índice de Karnofsky é uma escala que classifica o paciente de acordo com o grau das suas inaptidões ou deficiências funcionais que tem sido amplamente utilizado como “padrão-ouro” na avaliação de pacientes oncológicos (KARNOFSKY *et al.*, 1948). Caracteriza-se por uma avaliação subjetiva que segue uma escala que varia de 0 (zero) a 100 (cem), onde valores de 0 – 40 indicam que o paciente está inapto para cuidar de si mesmo; de 50 – 70 inapto para trabalhar e, de 80 – 100 indicam que o paciente está apto para atividades rotineiras e cuidar de si mesmo.

Dados clínico-patológicos das pacientes foram adquiridos por meio da análise dos prontuários, onde foi registrada a idade no momento do diagnóstico, peso, altura, história clínica e familiar, estado menopausal, comorbidades e estadiamento clínico do tumor.

#### *6.1.1.2. Critérios de exclusão*

Não foram incluídas no estudo mulheres com doenças crônicas não transmissíveis não controladas, ou transmissíveis (portadoras do HIV), que estivessem sob uso de terapia farmacológica, além de alcoolistas, fumantes, com problemas neurológicos ou psiquiátricos, ou ainda os casos cujos dados do prontuário estivessem incompletos quanto aos itens necessários para o estudo.

#### *6.1.1.3. Coleta das amostras de sangue*

Após jejum de 12 h das mulheres que concordaram em participar da pesquisa foi coletada uma alíquota de sangue (5 ml). A coleta foi realizada por profissional capacitado (técnico ou auxiliar de enfermagem) e em local reservado e equipado de modo a conferir segurança laboratorial ao indivíduo doador e coletor. Todo material necessário para a coleta de sangue foi utilizado pela primeira vez para este fim e teve seu descarte efetuado dentro dos padrões de segurança laboratorial. O sangue retirado foi coletado por punção venosa periférica, as amostras foram centrifugadas a  $3.500 \times g$  por 5 minutos e o sobrenadante, ou seja, o plasma obtido, foi armazenado a  $-20\text{ }^{\circ}\text{C}$  para uso posterior.

### **6.1.2. Reagentes e outros materiais analíticos**

As colunas de cromatografia, para isolamento das lectinas vegetais, D-galactose Agarose e Sephadex G5015 foram adquiridas da *Thermo Fisher Scientific* (Rockford, IL, USA) e GE Healthcare (Sunnyvale, CA, USA), respectivamente.

O marcador de massa molecular de proteínas que foi utilizado nas eletroforeses em gel de poliacrilamida foi o *Low Molecular Weight* (GE Healthcare - Sunnyvale, CA, USA).

A matriz de Sepharose 4B ativada com brometo de cianogênio, para imobilização das lectinas vegetais, e a coluna *HiTrap Albumin&IgG Depletion*, para depleção de IgG e Albumina, foram também adquiridas da GE Healthcare (Sunnyvale, CA, USA).

Os reagentes e colunas cromatográficas para análise em espectrômetro de massas nanoESI, *Synapt HDMS*, acoplado a um sistema de cromatografia nanoUPLC, foram adquiridos da *Waters Corporation* (Manchester, UK).

Todos os demais reagentes usados neste trabalho foram de grau analítico.

## **6.2. Métodos**

### **6.2.1. Isolamento e Imobilização das lectinas vegetais**

#### **6.2.1.1. Isolamento da lectina D-galactose-ligante das sementes de *Artocarpus incisa* (frutalina)**

O protocolo de extração e purificação seguiu a metodologia descrita por Moreira e colaboradores (1998).

A extração das proteínas solúveis totais foi realizada inicialmente ressuspendendo-se a farinha de sementes de *Artocarpus incisa* em solução de NaCl 0,15 M, na proporção 1:10 (m/v) e mantida em agitação contínua por 1 hora, à temperatura ambiente.

Decorrido esse tempo, a suspensão foi centrifugada a 10.000 x g, por 30 minutos, a 4 °C em centrífuga Eppendorf 5810R. Após a centrifugação, o sobrenadante foi filtrado em papel de filtro comum e aplicado a uma coluna de D-galactose-Agarose, previamente equilibrada com solução salina 0,15 M.

Para purificação da frutalina (FTL), as proteínas que não interagiram com a matriz (pico não retido ou Pico 1) foram eluídas com o tampão de equilíbrio (NaCl 0,15 M), até que

a  $A_{280}$  do eluato atingisse 0,020, após o que, as proteínas retidas (Pico retido ou Pico 2) foram, então, eluídas com adição de D-galactose 0,2 M em NaCl 0,15 M.

Durante todo o processo cromatográfico, foram coletadas frações de 2,5 mL, sob um fluxo de eluição de 0,5 mL/min e todas as alíquotas foram analisadas por medida da absorvância a 280 nm ( $A_{280}$ ) em espectrofotômetro.

Várias cromatografias seguindo a metodologia acima foram realizadas, visando acumular proteína suficiente para as etapas seguintes do trabalho, e os picos retidos obtidos após realização de cada cromatografia foram exaustivamente dialisados contra água destilada e posteriormente liofilizados.

#### 6.2.1.2. Isolamento da lectina glucose/manose-ligante de *Dioclea altissima* (DAL)

O protocolo de extração e purificação seguiu a metodologia descrita por Moreira e colaboradores (1997).

A extração das proteínas solúveis totais foi realizada inicialmente ressuspendendo-se a farinha de sementes de *Dioclea altissima* em solução de NaCl 0,15 M, na proporção 1:10 (m/v) e mantida em agitação contínua por 3 horas, à temperatura ambiente.

Decorrido esse tempo, a suspensão foi centrifugada a 10.000 x g, por 30 minutos, a 4 °C em centrífuga Eppendorf 5810R. Após a centrifugação, o sobrenadante foi filtrado em papel de filtro comum, foi adicionada à amostra  $Ca^{2+}$  e  $Mn^{2+}$ , para concentração final de 5 mM, centrifugada a 12.000 x g por 10 minutos e o sobrenadante aplicado à coluna de Sephadex G5015, previamente equilibrada com solução salina 0,15 M contendo também  $Ca^{2+}$  e  $Mn^{2+}$ , na concentração final de 5 mM.

Para obtenção da lectina DAL pura, proteínas que não interagiram com a matriz de Sephadex G5015 (pico não retido ou Pico 1) foram eluídas com o tampão de equilíbrio (NaCl 0,15 M contendo  $Ca^{2+}$  e  $Mn^{2+}$ , para concentração final de 5 mM), até que, mais uma vez, a  $A_{280}$  do eluato atingisse 0,020, e as proteínas retidas (pico retido ou Pico 2) foram, então, eluídas com tampão glicina 0,1 M pH 2,6.

Durante todo o processo cromatográfico, foram coletadas frações de 2,5 mL, sob um fluxo de eluição de 0,5 mL/min e todas as alíquotas foram analisadas por medida da absorvância a 280 nm ( $A_{280}$ ) em espectrofotômetro.

E, mais uma vez, várias cromatografias seguindo a metodologia acima foram realizadas, visando o acúmulo da lectina DAL em quantidade suficiente para as etapas

seguintes do trabalho, e os picos retidos obtidos após realização de cada cromatografia foram exaustivamente dialisados contra água destilada e posteriormente liofilizados.

### 6.2.1.3. Avaliação do grau de pureza das lectinas vegetais isoladas

A pureza das lectinas isoladas foi avaliada por eletroforese em gel de poliacrilamida em presença de SDS e  $\beta$ -mercaptoetanol e por espectrometria de massas.

Uma alíquota de cada lectina (1 mg/mL, em água ultrapura) foi misturada ao tampão de amostra 4 vezes concentrado [Tris-HCl 0,0625 M, pH 6,8; SDS 2% (m/v); 2-mercaptoetanol 5% (v/v) e azul de bromofenol 0,001% (m/v)], fervidas a 100 °C por 7 minutos e analisadas em eletroforese em gel de poliacrilamida 15%, em condições desnaturantes, como descrito por Laemmli (1970), adaptado para géis montados em placas de vidro.

A placa foi montada com gel de aplicação a 3,5 % de acrilamida em tampão Tris-HCl 0,5 M pH 6,8 e SDS 1,0 % e o gel de separação a 15 % em tampão Tris-HCl 3 M pH 8,8 e SDS 1,0%. A corrida eletroforética foi realizada a voltagem constante de 120 V, 40 mA, em tampão de corrida Tris-HCl 0,025 M pH 8,3, contendo glicina 0,192 M e SDS 0,1% (m/v). Após a corrida, o gel foi corado com Coomassie Brilliant Blue G-250 (Amresco) para detecção das bandas de proteínas, e descorado com água. O marcador de proteínas *Low Molecular Weight* (GE Healthcare - USA) foi utilizado na corrida eletroforética como padrão.

Alternativamente, para verificar o grau de pureza das lectinas isoladas, amostras das mesmas foram submetidas à análise por espectrometria de massa. Para isso, as amostras foram solubilizadas em ácido acético 10% em água ultrapura, centrifugadas a 12.000  $\times g$  por 10 minutos e o sobrenadante foi aplicado por infusão direta a uma concentração de 10  $\mu\text{g/mL}$ .

A aquisição de dados (MS) foi realizada com um espectrômetro de massas *Synapt HDMS* (Waters, Manchester, UK), programado com uma  $\text{RF}_{\text{offset}}$  (perfil MS) ajustada de forma que os dados de MS fossem efetivamente adquiridos de 1000-4000 m/z, o que possibilitou a obtenção das massas distribuídas multiplamente carregadas. Para todas as medições, o espectrômetro de massas foi operado no modo 'V', com uma potência de resolução de, pelo menos, 12.000. Todas as análises foram realizadas utilizando ionização por electrospray no modo ESI (+) e a coleta de dados foi realizada utilizando o programa *MassLynx v.4.1* (Waters – USA) e os espectros de distribuição de cargas foram deconvoluídos para análise dos resultados.

#### 6.2.1.4. Imobilização das lectinas em matriz de Sepharose 4B

As lectinas purificadas a partir de semente de *Artocarpus incisa* (frutalina - FTL) e de *Dioclea altíssima* (DAL) foram imobilizadas, separadamente, em resina de Sepharose 4B ativada com brometo de cianogênio (GE Healthcare - USA).

A metodologia consistiu, essencialmente, na incubação das lectinas, na presença de seus respectivos açúcares inibidores (D-galactose 0,2 M, para frutalina, e D-manose 0,2 M, contendo  $\text{Ca}^{2+}$  e  $\text{Mn}^{2+}$  para concentração final de 5 mM, para DAL), seguindo as instruções fornecidas pelo fabricante.

Para isso, 1 g de resina foi ressuspenso em 100 mL de HCl 1 mM e mantido em repouso durante 30 minutos, para total intumescimento e reidratação. Após este período, foram realizadas três lavagens consecutivas, com cinco volumes da mesma solução de reidratação, e o precipitado remanescente foi mantido em HCl 1 mM, a fim de preservar os sítios ativos da resina.

Concomitantemente, as lectinas liofilizadas, cerca de 15 mg de cada, foram dissolvidas em tampão  $\text{NaHCO}_3$  0,1 M, pH 8,3, contendo NaCl 0,5 M e os respectivos açúcares inibidores na concentração final de 0,2 M.

As suspensões, contendo as lectinas na presença dos seus açúcares inibidores em meio bicarbonato 0,1 M, pH 8,3, contendo NaCl 0,5 M e o gel de Sepharose 4B ativado com brometo de cianogênio hidratado com HCl 1,0 mM foram então mantidas em contato por 2 horas, sob agitação pendular discreta, à temperatura ambiente.

Após esta etapa, os sobrenadantes tiveram suas absorbâncias monitoradas a 280 nm, a fim de observar se todo ligante (lectinas) foi devidamente acoplado à resina de Sepharose 4B. Na sequência, os sobrenadantes foram descartados e as matrizes, cuidadosamente, lavadas cinco vezes com cinco volumes de tampão bicarbonato 0,1 M, pH 8,3 contendo NaCl 0,5 M.

Em seguida, foi adicionado tampão Tris-HCl 0,1 M, pH 8,0, contendo NaCl 0,5 M e mantido em repouso por 2 horas, à temperatura ambiente, a fim de bloquear os sítios ativos remanescentes na resina. Depois deste período, foi realizado um ciclo de seis lavagens, alternando entre os tampões acetato 0,1 M, pH 4,0 e Tris-HCl 0,1 M, pH 8,0, ambos contendo NaCl 0,5 M, iniciando com o primeiro e terminando com o segundo, respectivamente.

Ao fim do processo, as matrizes foram estocadas, separadamente, em solução salina contendo 0,02% de azida, a 8 °C.

### 6.2.3. Cromatografias de depleção

Inicialmente, as amostras de plasma foram retiradas do freezer – 80 °C e foram depletadas. Esse procedimento objetivou a depleção de proteínas abundantes, tais como albumina e IgG, para enriquecimento de proteínas menos abundantes presentes no plasma.

Assim, em um primeiro momento, uma alíquota de 500 µL de plasma foi filtrado em membrana 0,22 µM e um volume de 150 µL de plasma filtrado foi reservado para aplicação na coluna *HiTrap Albumin & IgG Depletion* de 1 mL (GE Healthcare - USA) acoplada a um sistema de FPLC *ÄKTApurifier 10* (GE Healthcare - USA).

Inicialmente a coluna foi equilibrada com cinco volumes de tampão de equilíbrio Tris-HCl 20 mM, pH 7,4, contendo NaCl 0,15 M, e, em seguida, foram aplicados 150 µL de plasma filtrado.

Para lavagem de proteínas que não interagiram com a matriz, ou seja, as proteínas que caracterizam o plasma depletado, que posteriormente foram aplicadas nas cromatografias de lectinas imobilizadas, foram adicionados à coluna oito volumes de tampão de equilíbrio, e, para eluição da fração proteica que interagiu com a matriz, ou seja, fração rica em albumina e IgG, foram adicionados 5 volumes de tampão Glicina-HCl, 0,1 M pH 2,7.

As cromatografias foram realizadas a um fluxo constante de 1 mL/min, com coletas de 200 µL e monitoramento da eluição de proteínas por absorvância a 216, 280 e 290 nm.

Para cada 500 µL de amostra de plasma filtrado foram realizadas 3 cromatografias de depleção consecutivas e as alíquotas de plasma depletado foram concentradas a vácuo em *Concetrator plus* (Eppendorf – DE) para experimentos posteriores, como abaixo descrito. A fração retida, rica em albumina e IgG, foi estocada.

### 6.2.4. Cromatografias de afinidade em colunas de lectinas imobilizadas

Para a realização das cromatografias de afinidade em colunas diferentes de lectinas vegetais imobilizadas, as amostras de plasma depletadas e concentradas, obtidas como descrito acima, foram, inicialmente, individualmente quantificadas por absorvância a 280 nm usando um espectrômetro *NanoVue Plus* (GE Healthcare), para posteriormente serem preparados os *pools* correspondentes aos grupos em estudo.

Para compor os *pools* referentes a cada grupo em estudo (Controle, Câncer de Mama Ductal I/Estágio Precoce, Câncer de Mama Ductal II/ Estágio Intermediário e Câncer de Mama Ductal III/Estágio Avançado) utilizou-se contribuições equivalentes, em massa de

proteínas, de 10 amostras individuais depletadas, para compor um volume final de 1 mL, e totalizando 4 *pools*.

A normalização dos dados é de essencial importância para quantificação *label-free*, na qual as amostras são processadas de forma totalmente independente, o que potencialmente gera bastante variação técnica quando comparada com estratégias baseadas em marcação.

Portando, considera-se essa etapa de quantificação por espectrofotometria, após depleção, para então formação de *pools*, anterior às cromatografias em matrizes de lectinas vegetais imobilizadas, o primeiro passo para normalização das amostras, uma vez que reunir quantidades iguais, em massa de proteínas, de amostra de mulheres diversas para formação de uma única amostra que represente cada grupo em estudo minimiza o erro na aplicação de massas desiguais nas cromatografias subsequentes.

As matrizes de lectinas imobilizada, Frutalina-Sepharose e DAL-Sepharose foram então montadas em colunas vazias de 1 mL e acopladas a um sistema de FPLC *ÄKTApurifier 10* (GE Healthcare - USA) para os procedimentos cromatográficos visando o fracionamento das proteínas plasmáticas.

Um volume de 1 mL de amostra, ou seja, o *pool* representativo para cada foi aplicado a cada coluna de lectina imobilizada.

O método de cromatografia foi padronizado para lavagem da coluna com 3,5 volumes de tampão de equilíbrio, seguido da aplicação de 1 mL de amostra, lavagem de proteínas não retidas, ou seja, proteínas que não interagiram com a matriz (Pico 1), com 5 volumes de tampão de equilíbrio, e para eluição da fração proteica que interagir com a matriz, ou seja, a fração retida (Pico 2), a qual representa principalmente as glicoproteínas pelas quais a lectina imobilizada tem afinidade, foram adicionados 5 volumes da solução dos respectivos açúcares ligantes em Tris 0,1 M pH 7,5, contendo NaCl 0,15 M: galactose 0,2 M (para coluna de Frutalina-Sepharose) e manose 0,2 M (para coluna de DAL-sepharose).

As cromatografias foram realizadas a um fluxo constante de 1 mL/min, com coletas de 1 mL e a eluição das glicoproteínas foi acompanhada por absorvância a 216 e 280 nm.

As frações contendo as proteínas fracionadas por cromatografia de afinidade nas 2 colunas propostas foram dialisadas contra água ultra pura e concentradas por ultrafiltração, em concentradores *Amicon* (Millipore-USA) com limite de exclusão (*cutoff*) de 5 kDA.

Obeve-se, após essa etapa, duas frações para cada grupo em estudo para cada coluna de afinidade, ou seja, a fração não retida (Pico 1) e a fração retida na coluna de lectina imobilizada (rica em glicoproteínas com afinidade por aquela lectina – Pico 2). Portanto, levando em consideração que se trata de quatro grupos de estudo, para cada lectina utilizada

nesse trabalho foi obtido um total de 08 frações. Consequentemente, 16 frações no total foram obtidas após as cromatografias de afinidade nas duas matrizes de lectina vegetal imobilizada propostas.

#### **6.2.5. Digestão de proteínas para LC-MS<sup>E</sup>**

As frações dialisadas e concentradas contendo as proteínas, fracionadas a partir das cromatografias nas diferentes colunas com lectina vegetal imobilizada, como acima descrito, ou mesmo sem fracionamento prévio (*pools* totais) foram inicialmente submetidas à digestão com tripsina.

Com esta finalidade, os picos obtidos por meio de cromatografias de afinidade em lectinas imobilizadas, foram, separadamente, tratados com 0,2% da solução do surfactante *RapiGest SF* (Waters, Milford, USA), em uma solução de  $\text{NH}_4\text{HCO}_3$  50 mM, adicionados a 100  $\mu\text{g}$  de fração proteica. A mistura foi incubada a 80 °C, por 15 min. Em seguida, foram adicionados às amostras 5  $\mu\text{L}$  de ditionitrotol (DTT) 100 mM, para abertura da proteínas, o que facilita o acesso para alquilação e digestão. Após um *vortex*, o tubo foi colocado em chapa aquecedora a 60 °C. Após incubação de 30 min, foram adicionados 5  $\mu\text{L}$  de iodoacetamida (IAA) 300 mM à mistura, para alquilação das cisteínas e, mais uma vez, foi dado um *vortex*. As amostras foram então colocadas em local escuro, à temperatura ambiente, para que a reação de alquilação ocorresse por 30 min. Decorrido esse tempo, 1  $\mu\text{g}$  da enzima tripsina (Promega) foi adicionado às amostras e a reação de digestão ocorreu em estufa, a 37 °C, por aproximadamente 16 h. Por fim, a amostra foi centrifugada a 13000 x g, por 30 min, a 6 °C, e o sobrenadante foi transferido para frascos apropriados para análise por espectrometria de massas, denominados *vials*.

Aos frascos contendo as proteínas digeridas foram adicionados peptídeos trópticos da proteína fúngica álcool-dehidrogenase (ADH; UniProt: P00330) ou peptídeos trópticos da proteína rábica fosforilase B (PhospB; UniProt: P00489), para serem utilizadas como proteínas padrão conhecidas e essenciais para quantificação absoluta de cada amostra (GEROMANOS *et al.*, 2011; MBEUNKUI *et al.*, 2011).

### **6.2.6. NanoUPLC e aquisição independente de dados (MSE) livre de marcação (label-free) por espectrometria de massas**

A aquisição de dados de espectrometria de massas foi realizada através de um espectrômetro de massas NanoESI-Q-TOF, acoplado a um sistema de cromatografia nanoUPLC.

Dessa forma, inicialmente, os peptídeos trípticos foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa C18 HSS T3 com dimensões 1,8  $\mu\text{m}$ , 75  $\mu\text{m}$  x 20 mm, por 110 min utilizando um gradiente de 0% a 40% durante 90 min, 40% a 85% durante 5 min, em seguida, a coluna foi reequilibrada durante 15 min a 35 °C. O fluxo foi de 0,35 mL/min e as fases móveis A e B continham 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente.

Os peptídeos, à medida que iam sendo eluídos no gradiente mencionado acima, eram submetidos à ionização em fonte nanoESI em capilar de sílica de 22  $\mu\text{m}$  através de spray em forma de cone. O spray fora espacialmente ajustado a fim de obter-se uma distribuição ideal de amostra no cone com a obtenção do maior número de íons por MS (TIC) em uma aquisição de 1,5 segundos.

Para todas as medições, o espectrômetro de massas foi operado no modo 'V', em uma resolução de, pelo menos, 12.000. Todas as análises foram realizadas utilizando ionização por electrospray no modo ESI (+) através da fonte *NanoLockSpray*. O canal de coleta da amostra analisada foi fechado a cada 30 seg para passagem do íon de referência.

O espectrômetro de massa foi calibrado com uma solução de GFP ([Glu]-*fibrinogen peptide*) 500 fmol/mL injetado através do spray de referência da fonte *NanoLockSpray*. O íon duplamente carregado ( $[M + 2H]^{2+}$ ) foi usado para a calibração de um ponto inicial ( $L_{\text{teff}}$ ), e a fragmentação de íons MS/MS do GFP foi usada para obter a calibração final do instrumento.

O espectrômetro foi programado para alternar automaticamente entre MS padrão (3 eV) e energias elevadas de colisão em rampa MSE (15-50 eV) na célula de colisão com gás argônio. A aquisição de dados após colisão foi realizada em faixa de 50-2000 m/z. A  $RF_{\text{offset}}$  (perfil MS) foi ajustada de forma que os dados de LC/MS foram efetivamente adquiridos de 300-2000 m/z, o que garante que qualquer massa observada nos dados de LC/MS<sup>E</sup> com menos 300 m/z surgiram de dissociações na célula de colisão.

Após o ajuste do processo cromatográfico, *spray* e configurações do equipamento, uma única corrida para cada amostra foi realizada. O objetivo da realização dessa única

corrida de cada fração, antes da aquisição total de dados em replicata, foi obter dados para quantificação absoluta, pelo método Hi3, a fim de mensurar a quantidade total de proteínas, em nanogramas, para cada fração e, em seguida, normalizar o volume de amostra a ser injetado para o processo de aquisição de dados através de NanoUPLC/MS. Seguindo esse procedimento inicial, foi injetada no espectrômetro de massas uma mesma quantidade, em massa de peptídeos, para todas as frações analisadas nesse trabalho.

Uma vez determinado o volume a ser injetado para cada fração conforme acima descrito, a aquisição de dados de todas as frações fora feita em triplicata. As corridas foram monitoradas quanto à separação cromatográfica, intensidade média dos picos cromatográficos e resolução dos mesmos.

Em um total de 8 frações (4 para pico não retido ou pico 1 e 4 para pico retido ou pico 2) obtidas para cada lectina utilizada, obtivemos 24 aquisição de LS/MS<sup>E</sup>. Somando-se, portanto, às amostras sem fracionamento prévio, nesse estudo foram realizadas 60 aquisições de dados.

#### **6.2.7. Processamento de dados e análise de expressão diferencial**

Para coleta, processamento e análise de dados foram utilizados os programas *MassLynx* v4. 1 e *ProteinLynx* v2.4 e, para busca no banco de dados adequado, foi utilizado o pacote *ProteinLynxGlobalServer* (PLGs) v.2.4. contendo o programa *Expression*<sup>E</sup> v.2.4. Os bancos de dados UniProtKB/Swiss-Prot 57.1 e UniProtKB/TrEMBL 40.1 foram utilizados e as condições de pesquisa foram baseadas na taxonomia [*Homo sapiens* (humano)].

Para busca de espectros e comparação no banco de dados, foram utilizados os parâmetros *default* do PLGs: máximo de uma clivagem triptica perdida, carbamidometilação como modificação fixa e oxidação da metionina como modificação variável (SILVA *et al.*, 2005; SILVA *et al.*, 2006).

A quantificação absoluta para cada corrida foi calculada de acordo com a média da intensidade dos três peptídeos mais intensos para cada proteína (*label-free* HI3) utilizando uma concentração conhecida de peptídeos da ADH ou PhospB como padrão interno e marcador para cálculo das concentrações de todas as proteínas presentes nas diferentes frações analisadas (SILVA *et al.*, 2006).

Após a identificação e quantificação absoluta das proteínas de cada fração, algumas proteínas foram selecionadas como proteínas internas (*housekeeping*) para normalizar os

níveis de expressão entre os grupos em estudo e realizar a quantificação relativa, *label-free*, de proteínas (LI *et al.*, 2009).

O critério para a seleção das *housekeepings* foi que essas proteínas fossem detectadas em todas as corridas de todas as frações naquela análise e apresentassem o menor coeficiente de variância, de acordo com as contagens quantitativas para as seus respectivos peptídeos HI3. Ademais, a normalização feita utilizando-se o PLGS utilizou tanto o *housekeeping* escolhido como também os EMRTs (*Elution masses retention time*) também detectados em todas as corridas e em tempos diferentes de separação.

Da lista final de proteínas identificadas com confiança ( $0,05 \geq p \geq 0,95$ ), aquelas que apresentaram razão (Câncer de mama: Controle) igual ou superior a 1,5 foram consideradas superexpressas ou *up-regulated* e aquelas que apresentaram razão igual ou inferior 0,66 foram consideradas subexpressas ou *down-regulated*. Proteínas com razões de expressão entre os valores acima estabelecidos foram consideradas sem expressão diferencial ou *unchanged*.

Os valores de razão acima estabelecidos para super ou sub expressão foram determinados com base na expressão, aumentada ou diminuída, de, pelo menos, 50% da proteína identificada.

#### **6.2.8. Interatoma**

Uma vez listadas as proteínas que apresentaram diferenças de expressão e/ou alteração no padrão de glicosilação no plasma de mulheres com câncer de mama, comparado ao de mulheres saudáveis, foi realizada a construção de uma rede interatômica.

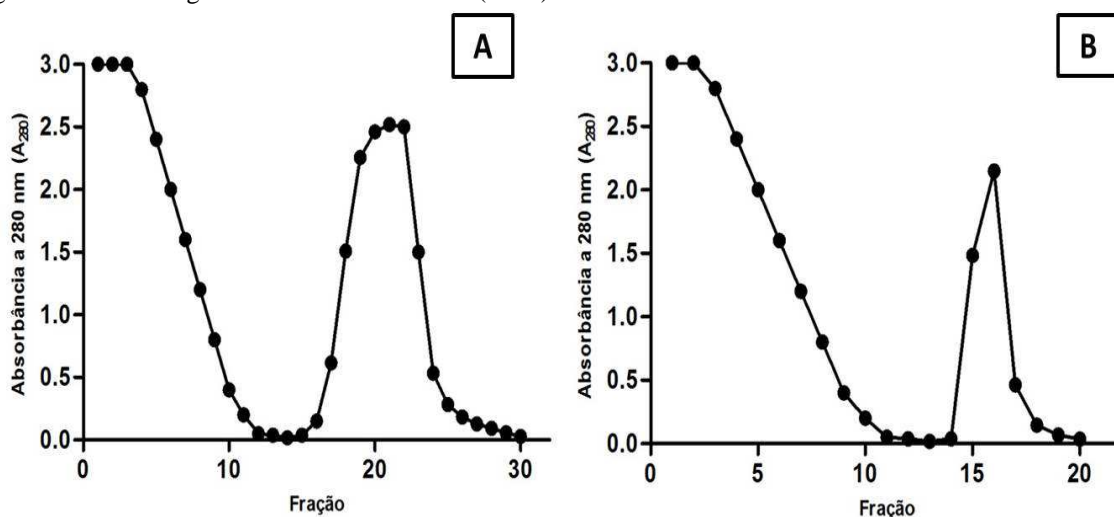
Para isso, os códigos de acesso das proteínas foram submetidos ao portal STRING (*Search Tool for the Retrieval of Interacting Genes/Proteins*) - disponível em <http://string.embl.de/> - o qual consiste em bancos de dados sobre predições e interações de proteínas conhecidas, envolvendo associações físicas e funcionais (SZKLARCZYK *et al.*, 2015).

## 7. RESULTADOS E DISCUSSÃO

### 7.1. Isolamento das lectinas

A purificação de ambas as lectinas em estudo se deu de maneira simples, necessitando de apenas um passo cromatográfico, utilizando colunas de afinidade. Todas as alíquotas cromatográficas foram analisadas através da medida da absorbância a 280nm ( $A_{280}$ ) ao longo de todo processo cromatográfico, como representado na Figura 3-A, para purificação de frutalina, lectina D-galactose-ligante das sementes de *Artocarpus incisa*, e na Figura 3-B, para purificação da lectina glucose/manose-ligante de *Dioclea altíssima* (DAL).

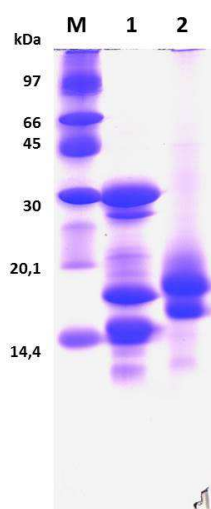
Figura 3 - Purificação das lectinas D-galactose-ligante das sementes de *Artocarpus incisa* (frutalina) e glucose/manose-ligante de *Dioclea altíssima* (DAL)



Perfis cromatográficos do processo de purificação da Frutalina em coluna de D-galactose (A) e de DAL em coluna de Sephadex G5015 (B). Para purificação de frutalina, lectina D-galactose-ligante das sementes de *Artocarpus incisa*, o extrato foi aplicado a uma coluna de D-galactose-Agarose, previamente equilibrada com solução salina 0,15 M. As proteínas que não interagiram com a matriz (pico não retido – Pico 1) foram eluídas com o tampão de equilíbrio (NaCl 0,15 M), até que a  $A_{280}$  do eluato atingisse 0,020, após o que, as proteínas retidas (Pico 2) foram, então, eluídas com adição de D-galactose 0,2 M em NaCl 0,15 M. Já para purificação da lectina glucose/manose-ligante de *Dioclea altíssima* (DAL) o processo cromatográfico consistiu na aplicação do extrato em coluna de Sephadex G5015, previamente equilibrada com solução salina 0,15 M contendo  $Ca^{2+}$  e  $Mn^{2+}$ , na concentração final de 5 mM. As proteínas que não interagiram com a matriz de Sephadex G5015 (pico não retido – Pico 1) foram eluídas com o tampão de equilíbrio, até que, mais uma vez, a  $A_{280}$  do eluato atingisse 0,020, e as proteínas retidas (Pico 2) foram, então, eluídas com tampão glicina 0,1 M pH 2,6.

Alíquotas dos picos retidos, oriundos de vários processos cromatográficos, que apresentaram as maiores leituras em  $A_{280}$ , foram reunidas e dialisadas exaustivamente contra água. As amostras das lectinas purificadas foram então liofilizadas, ressuspendidas em água ultrapura e submetidas à eletroforese em gel de poliacrilamida 15% em condições desnaturantes (com SDS) e redutoras (na presença de  $\beta$ -mercaptoetanol), como mostrado na Figura 4.

Figura 4 - Análise por SDS-PAGE das lectinas purificadas



No poço M, marcador de peso molecular *Low Molecular Weight* (GE Healthcare). No poço 1, amostra de DAL após diálise e liofilização dos picos obtidos em coluna de Sephadex G5015, e no poço 2, amostra de frutalina obtida após diálise e liofilização dos picos obtidos em coluna de D-galactose Sepharose. Foram aplicados 25  $\mu$ g de cada amostra.

Alternativamente, para verificar o grau de pureza das lectinas isoladas, amostras das mesmas foram submetidas à análise por espectrometria de massa. As Figuras 5 e 6 ilustram os espectros deconvoluídos das lectinas D-galactose-ligante das sementes de *Artocarpus incisa* (Frutalina - FTL) e glucose/manose-ligante de *Dioclea altíssima* (DAL), respectivamente. Os dados de massa intacta foram processados usando a técnica de máxima entropia (*maximum-entropy technique* - MaxEnt) para obtenção do espectro deconvoluído (FERRIGE *et al.*, 1991).

A frutalina é expressa em diferentes isoformas, que refletem, principalmente, diferenças nas modificações pós-traducionais, tais como glicosilação. O perfil eletroforético mostra um padrão de massas típico de uma *jacalina-like* com bandas entre 20 e 14 kDa (Figura 4), que corresponde à fração glicosilada e à fração pouco ou sem glicosilação, respectivamente (OLIVEIRA *et al.*, 2008). No espectro de massas deconvoluído é possível

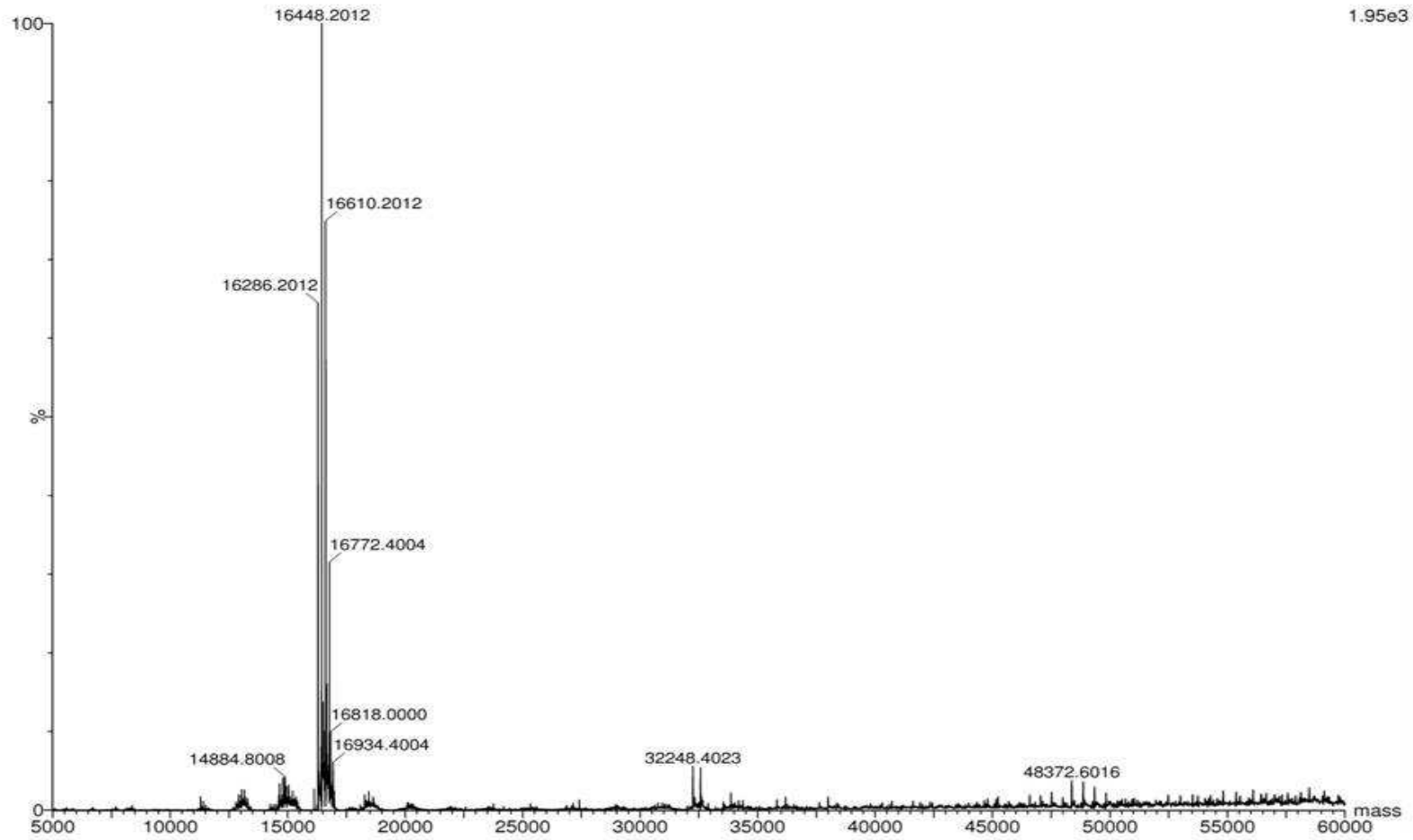
visualizar diferentes massas em torno de 16,5 kDa, que é consistente com a presença de isoformas de monômeros semelhantes (MONTEIRO-MOREIRA *et al.*, 2015).

A lectina de sementes de *Dioclea altíssima* (DAL) apresentou um perfil eletroforético com 3 bandas evidentes de massas moleculares aparentes de 30, 18 e 14 kDa e variações. Moreira e colaboradores (1997), em experimentos de filtração em gel em coluna de Superose 12R, em intervalos de pH de 2,3 a 8,3, encontraram diferentes padrões eletroforéticos da DAL. Enquanto um pico foi encontrado tanto em pH 2,6 (MM de 26 kDa, monômero) e pH 7 (MM de 50 kDa, dímero), em pH 8,3 foram encontrados dois picos de massa molecular de 50 kDa (73% de dímero) e 100 ou mais kDa (27% de tetrâmero).

Moreira e colaboradores (1997) realizaram testes com DAL usando o método do ácido fenol-sulfúrico, porém, diferentemente de frutalina, não foi detectado qualquer açúcar, indicando que a lectina de *Dioclea altíssima*, à semelhança de outras lectinas da subtribo *Diocleinae*, não é uma glicoproteína.

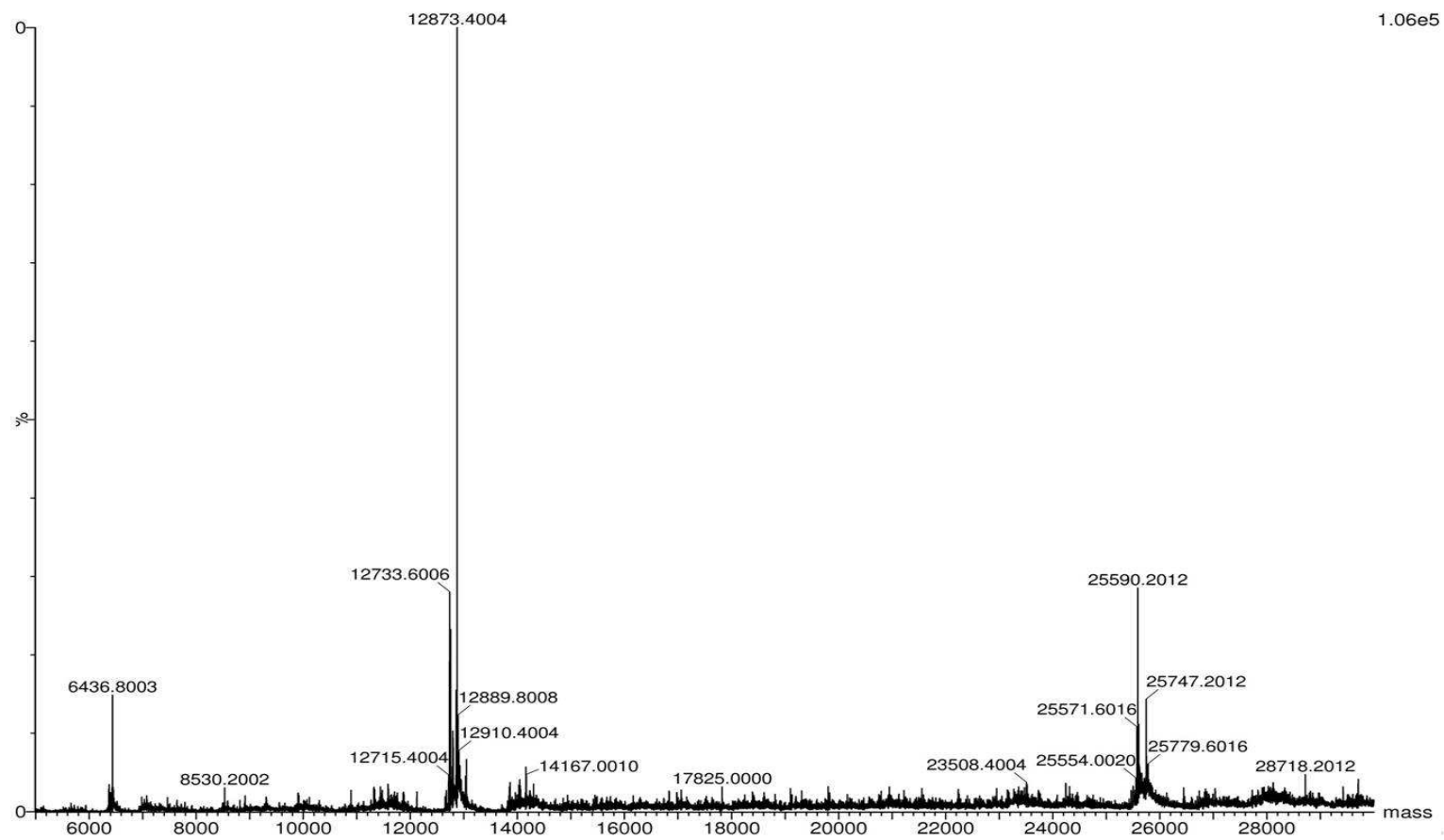
No espectro de massas deconvoluído de DAL é possível visualizar massas em 12 kDa e 25 kDa, referentes às cadeias isoladas e à proteína completa (monômero), respectivamente. Esse perfil resulta das diferentes formas de processamento pós-traducional da proteína, na qual uma fração de DAL produzida apresenta as cadeias unidas covalentemente, e em outra fração, as cadeias unem-se por interações fracas para formar o monômero. A estrutura tridimensional da proteína ainda não foi elucidada.

Figura 5 - Espectro de massa intacta deconvoluido de Frutalina



Espectro de massa intacta deconvoluido da lectina  $\alpha$ -D-galactose-ligante das sementes de *Artocarpus incisa* (Frutalina - FTL) apresentando massas em torno de 16 kDa.

Figura 6 - Espectro de massa intacta deconvoluído de DAL



Espectro de massa intacta deconvoluído da lectina glucose/manose-ligante de *Dioclea altissima* (DAL) apresentando massas em torno de 12 e 25 kDa.

## **7.2. Imobilização das lectinas**

Previamente à realização de cromatografias utilizando as matrizes com lectinas imobilizadas, um indício de que as lectinas foram imobilizadas com sucesso na resina de Sepharose 4B, baseia-se nas leituras espectrofotométricas realizadas com as amostras de lectinas solubilizadas em tampão de acoplamento, a 280 nm, antes e após incubação com a matriz ativada com brometo de cianogênio.

As amostras aplicadas à matriz apresentaram ODs de 1,398, para DAL, e de 1,575, para frutalina. Após seguir as instruções do fabricante e decorrido o período de 2 horas de incubação com a matriz de Sepharose 4B ativada, o sobrenadante, após decantação da matriz, apresentou ODs de 0,011 para DAL, e de 0,021 para frutalina.

A diminuição significativa dos valores de absorbância a 280 nm no tampão de acoplamento após incubação com a resina sugere uma alta eficiência de imobilização das lectinas em matriz de Sepharose 4B.

## **7.3. Cromatografias de depleção de Albumina e IgG**

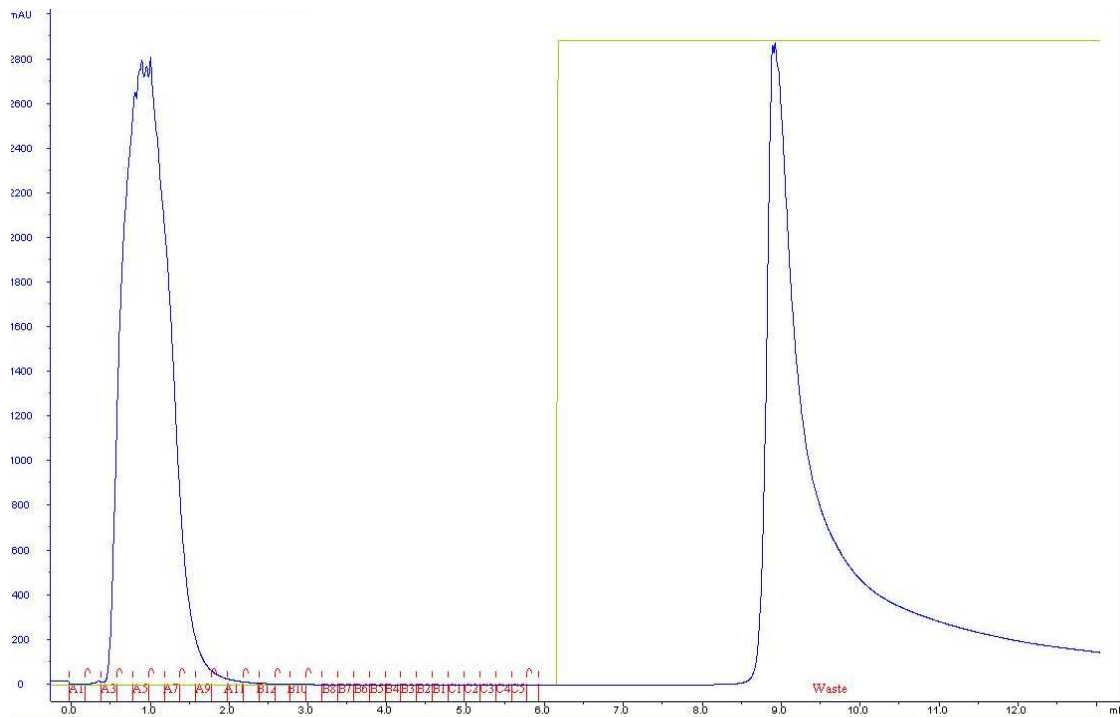
Inicialmente, uma alíquota de 500 µL de plasma de cada mulher participante do estudo foi filtrada em membrana 0,22 µM e, para cada amostra, foram realizadas 3 cromatografias subsequentes, a partir da aplicação de 150 µL de plasma filtrado à coluna *HiTrap Albumin & IgG Depletion* de 1 mL (GE Healthcare - USA) acoplada a um sistema de FPLC *ÄKTApurifier 10* (GE Healthcare - USA).

Todas as alíquotas foram analisadas através da medida da absorbância a 216 e 280nm ao longo de todo processo cromatográfico, como representado na Figura 7.

Inicialmente, as amostras de plasma aplicadas à coluna apresentavam uma concentração média de aproximadamente 70 mg/mL, após o processo cromatográfico de depleção as frações isentas de albumina e IgG, que caracterizam o pico não retido (pico 1) ou plasma depletado, apresentaram concentrações médias de cerca de 4 mg/mL, o que demonstra a eficiência da depleção das proteínas abundantes albumina e IgG.

Ao fim das cromatografias de depleção referentes a cada participante do estudo, as frações de plasma depletado foram concentradas à vácuo em *Concetrator plus* (Eppendorf – DE) para posterior cromatografia em colunas de lectinas vegetais imobilizadas.

Figura 7 – Depleção de albumina e IgG das amostras de plasma humano



Perfil cromatográfico do processo de depleção de albumina e IgG das amostras de plasma humano utilizando coluna *HiTrap Albumin & IgG Depletion* de 1 mL (GE Healthcare - USA) acoplada a um sistema de FPLC *ÄKTApurifier 10* (GE Healthcare - USA). As proteínas que não interagiram com a matriz, ou seja, as proteínas que caracterizam o plasma depletado, foram eluídas com tampão de equilíbrio (Tris-HCl 20 mM, pH 7,4, contendo NaCl 0,15 M) e, para eluição da fração proteica que interagiu com a matriz, ou seja, fração rica em albumina e IgG, foi adicionado tampão Glicina-HCl, 0,1 M pH 2,7 (representado pela linha verde). Todas as alíquotas foram analisadas através da medida da absorbância a 280nm ao longo de todo processo cromatográfico, como ilustrado pela linha azul.

O plasma humano contém de 60 a 80 mg de proteína por mililitro, somadas a outras várias moléculas, que incluem sais, lipídeos, eletrólitos e pequenos metabólitos, como peptídeos e aminoácidos (CHAN *et al.*, 2004).

A caracterização de proteínas séricas, assim como a prospecção de biomarcadores no plasma sanguíneo, é consideravelmente dificultada pelo vasto intervalo dinâmico, ou *dynamic range*, do proteoma sérico. Com concentrações de proteína medindo cerca de 12 ordens de grandeza, a análise por espectrometria de massa convencional permite a detecção de apenas algumas proteínas de baixa abundância. Dessa forma, a depleção prévia de proteínas de alta abundância da amostra pode aumentar consideravelmente a profundidade analítica, tornando-se, portanto, uma prática amplamente utilizada (JAROS *et al.*, 2013).

Existe ainda bastante controvérsia sobre a estratégia de depleção mais adequada para análise proteômica do plasma humano. Devido aos desafios analíticos de mensurar

precisamente amostras contendo um grande intervalo dinâmico de proteínas em diferentes concentrações, a maioria dos protocolos atuais tem empregado depleção prévia de proteínas altamente abundantes tais como albumina e imunoglobulinas.

Atualmente, kits de depleção para a remoção simultânea de albumina e IgG estão disponíveis comercialmente. A recente tendência na análise proteômica apoia o uso de matrizes de imunoafinidade, que são constituídas de matrizes com anticorpos diferentes covalentemente ligados contra a maioria das proteínas abundantes. Alguns estudos comparativos sobre as várias abordagens de depleção vêm sendo realizados (BELLEI *et al.*, 2011).

Yocum e colaboradores (2005) analisaram soro humano com e sem depleção, por imunoafinidade, de seis proteínas abundantes em massa por cromatografia líquida multidimensional em *tandem* e espectrometria de massas. Duas replicatas de cada experimento foram realizadas e comparadas umas com as outras. Em ambas replicatas, da amostra depletada e não-depletada, houve uma sobreposição de 73% e 72% de peptídeos identificados e sobreposição de 64% e 78% de proteínas identificadas, respectivamente. Das 262 proteínas únicas identificadas nas quatro corridas, 82 foram encontradas em todas as quatro amostras, 142 exclusivas para o soro depletado e 38 exclusivas para o soro não-depletado. Logo, os autores concluíram que, embora a depleção sérica de proteínas altamente abundantes tenha aumentado significativamente o número de proteínas identificadas, tanto o grau de complexidade da amostra e seu método de depleção resultaram em perda não seletiva de algumas outras proteínas.

Infelizmente, a probabilidade de remoção de algumas proteínas de baixa abundância, juntamente com as moléculas abundantes, é um verdadeiro problema associado com qualquer método de separação de proteínas. A albumina é a proteína mais abundante presente no plasma, mas, ao mesmo tempo, é uma excelente proteína transportadora, ligando-se a vários compostos, incluindo hormônios, lipídeos e aminoácidos. Por este motivo, a remoção da albumina a partir do plasma também pode resultar na perda de alguns peptídeos de baixa abundância ou pequenas proteínas de interesse, tais como citocinas (GRANGER *et al.*, 2005).

Proteínas transportadoras, tais como a albumina, atuam como ímãs para acumular e amplificar as proteínas de menor abundância, tais como bilirrubina, funcionando como uma esponja molecular. Portanto, em qualquer ponto, a concentração de uma proteína de baixo peso molecular ou mesmo um peptídeo é em função da sua entrada e saída do sangue e se ela tem afinidade de ligação a uma proteína transportadora. Portanto, se um biomarcador é uma proteína de baixo peso molecular que tem uma elevada afinidade de ligação com uma proteína

transportadora, sua existência é então dependente da sua taxa de produção, taxa de depuração/excreção e taxa de depuração/excreção da proteína transportadora. Visto que uma proteína transportadora tem uma meia-vida mais longa do que o biomarcador não ligado, a concentração total do biomarcador pode tornar-se elevada devido à sua associação com a proteína transportadora. Esta amplificação pode, em teoria, provocar um aumento de várias ordens de grandeza (YOCUM *et al.*, 2005).

#### ***7.4. Cromatografias de afinidade em colunas contendo lectinas vegetais imobilizadas***

Inicialmente, foram preparados quatro *pools*, referentes a cada grupo em estudo (controle, câncer de mama ductal estágio I, câncer de mama ductal estágio II e câncer de mama ductal estágio III), a partir de 10 amostras individuais depletadas, pertencentes a cada grupo.

A utilização de *pools* para a análise proteômica se baseia no fato de que, ao reunir amostras individuais, para formação de uma única amostra referente a um grupo de estudo, aquelas proteínas que apresentam diferenças entre as amostras e que não tenham relação com a doença sofrerão diluição, o que implica em baixa identificação. Já aquelas proteínas que forem comuns às amostras individuais que compõem o *pool* e, conseqüentemente, apresentam relação com a doença, se evidenciam, sendo mais fácil de serem identificadas. Além disso, quando se trata do câncer de mama, a formação de *pools* acaba também por reduzir a heterogeneidade entre amostras de mulheres que se encontram em um mesmo estágio da doença.

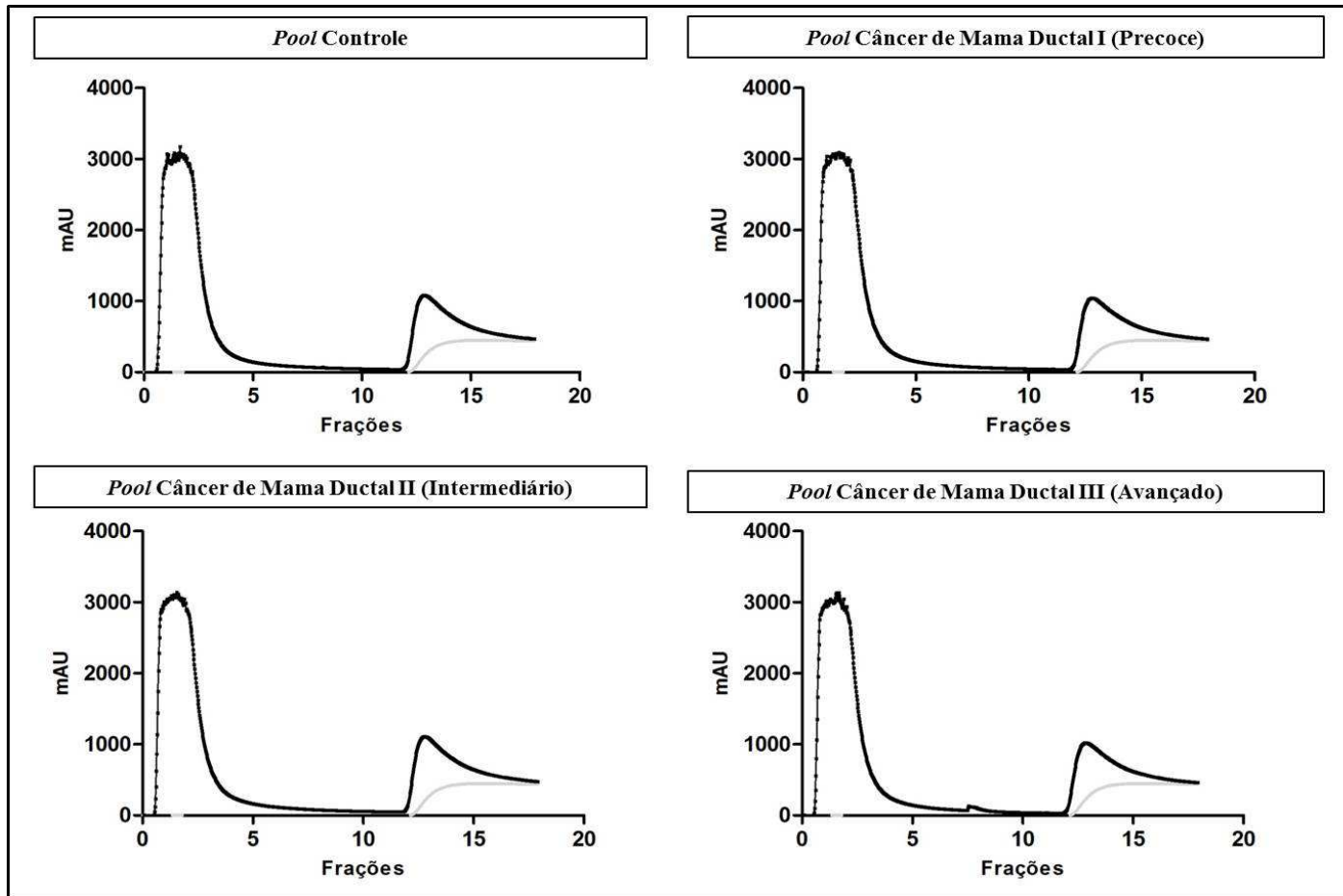
Dessa forma, os *pools*, em um volume final de 1 mL para cada grupo, foram submetidos, separadamente, a cromatografias de afinidade em 2 colunas diferentes de lectinas vegetais imobilizadas: coluna de Frutalina-Sepharose e DAL-Sepharose.

O método de cromatografia foi padronizado para equilíbrio da coluna com Tris 0,1 M pH 7,5, contendo NaCl 0,15 M, seguido da aplicação de 1 mL de amostra, lavagem de proteínas não retidas, ou seja, proteínas que não interagiram com a matriz (Pico 1) com o mesmo tampão de equilíbrio, e para eluição da fração proteica que interagiu com a matriz, ou seja, a fração retida (Pico 2), a qual representa as glicoproteínas pelas quais a lectina imobilizada tem afinidade, foram adicionados os respectivos açúcares ligantes em Tris 0,1 M pH 7,5, contendo NaCl 0,15 M: galactose 0,2 M (para coluna de Frutalina-Sepharose) e manose 0,2 M (para coluna de DAL-sepharose).

As cromatografias foram realizadas a um fluxo constante de 1 mL/min, com coletas de

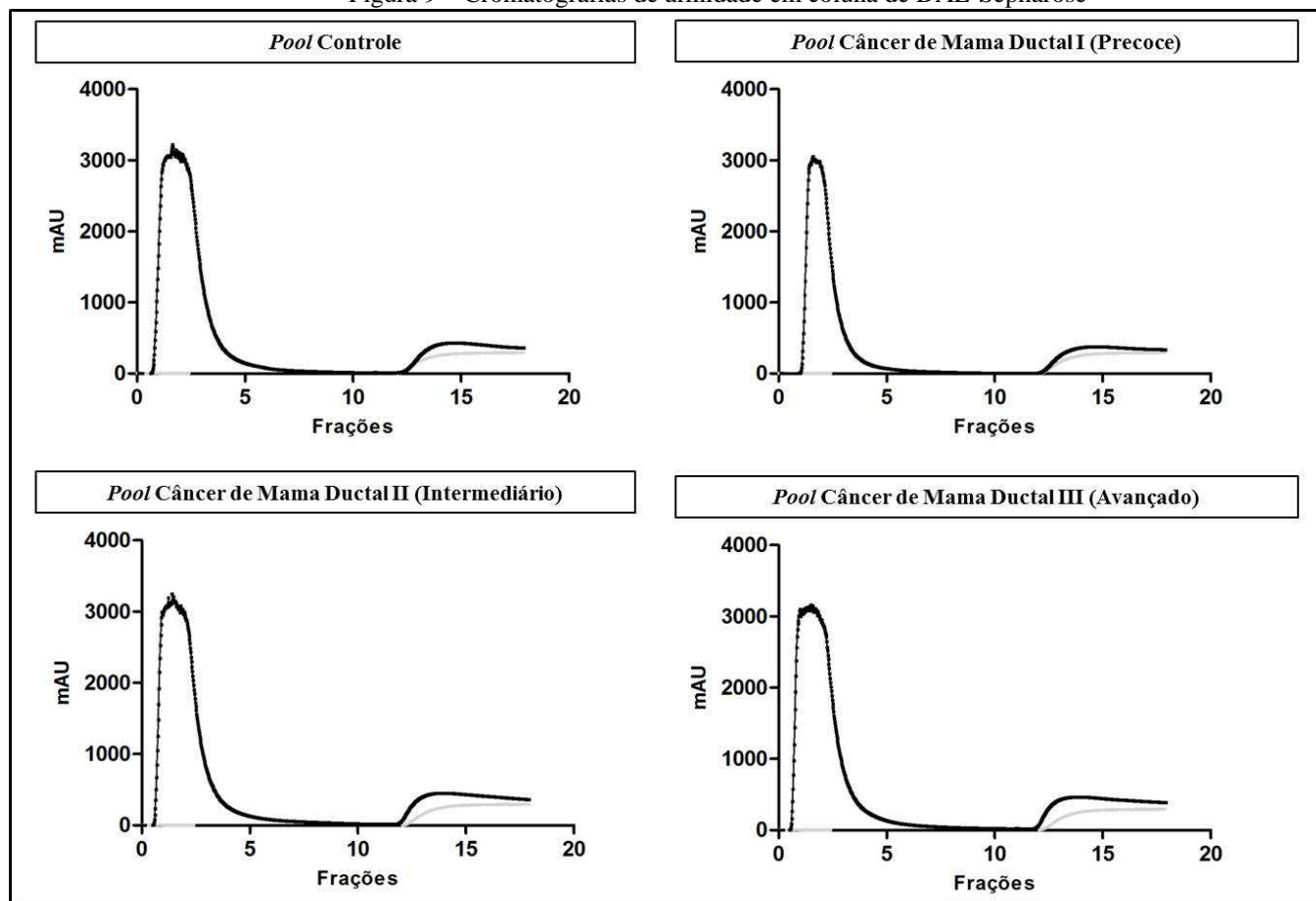
1 mL e a eluição das proteínas foi acompanhada por absorvância a 216 e 280 nm, como representado na Figura 8 e 9. Os picos obtidos foram então dialisados contra água ultra pura e concentradas por ultrafiltração, em concentradores *Amicon* (Millipore-USA) com limite de exclusão (*cutoff*) de 5 kDA.

Figura 8 – Cromatografias de afinidade em coluna de Frutalina-Sepharose



Perfis cromatográficos do processo de fracionamento dos *pools* referentes aos grupos em estudo, utilizando coluna de Sepharose4B imobilizada com a lectina  $\alpha$ -D-galactose-ligante das sementes de *Artocarpus incisa* (Frutalina - FTL) acoplada a um sistema de FPLC *ÄKTApurifier 10* (GE Healthcare - USA). As proteínas que não interagiram com a lectina, ou seja, a fração proteica correspondente ao Pico 1, foram eluídas com tampão de equilíbrio (Tris-HCl 20 mM, pH 7,4, contendo NaCl 0,15 M) e, para eluição da fração proteica que interagiu com a lectina, ou seja, a fração proteica correspondente ao Pico 2, foi adicionada uma solução de galactose 0,2M em Tris 0,1 M pH 7,5, contendo NaCl 0,15 M. Todas as alíquotas foram analisadas através da medida da absorbância a 216nm ao longo de todo processo cromatográfico, como ilustrado pelas linhas. A linha cinza corresponde às leituras de absorbância a 216 nm dos tampões utilizados na cromatografia.

Figura 9 – Cromatografias de afinidade em coluna de DAL-Sepharose



Perfis cromatográficos do processo de fracionamento dos *pools* referentes aos grupos em estudo, utilizando coluna de Sepharose4B imobilizada com a lectina glucose/manose-ligante de *Dioclea altissima* (DAL) acoplada a um sistema de FPLC *ÅKTApurifier 10* (GE Healthcare - USA). As proteínas que não interagiram com a lectina, ou seja, a fração proteica correspondente ao Pico 1, foram eluídas com tampão de equilíbrio (Tris-HCl 20 mM, pH 7,4, contendo NaCl 0,15 M) e, para eluição da fração proteica que interagiu com a lectina, ou seja, a fração proteica correspondente ao Pico 2, foi adicionada uma solução de manose 0,2M em Tris 0,1 M pH 7,5, contendo NaCl 0,15 M (representado pela linha verde). Todas as alíquotas foram analisadas através da medida da absorbância a 216nm ao longo de todo processo cromatográfico, como ilustrado pelas linhas. A linha cinza corresponde às leituras de absorbância a 216 nm dos tampões utilizados na cromatografia.

Embora as cromatografias tenham sido realizadas em equipamento de excelente desempenho e reprodutibilidade satisfatória, não foi possível observar, em nível de detecção U.V., nenhum padrão cromatográfico diferenciado entre os diferentes grupos em estudo (*pools*). Isso evidencia que diferenças de expressão, muitas vezes sutis, são imperceptíveis por detecção U.V., requerendo, portanto, uma técnica mais analítica, tal como o LC/MS, para constatação de expressão diferencial de proteínas.

Ademais, observa-se que os picos retidos (picos 2) em coluna de afinidade de Frutalina foram mais abundantes, em termo de absorção a 216 nm, do que os picos retidos em coluna de DAL. Dessa forma, o que se pode concluir, até esta etapa, é que houve um fracionamento útil para os experimentos de LC/MS, e este ocorreu de maneira diferenciada entre as lectinas. A mesma constatação fora feita na análise quali e quantitativa, vista por espectrometria de massas, que será discutida em seções seguintes.

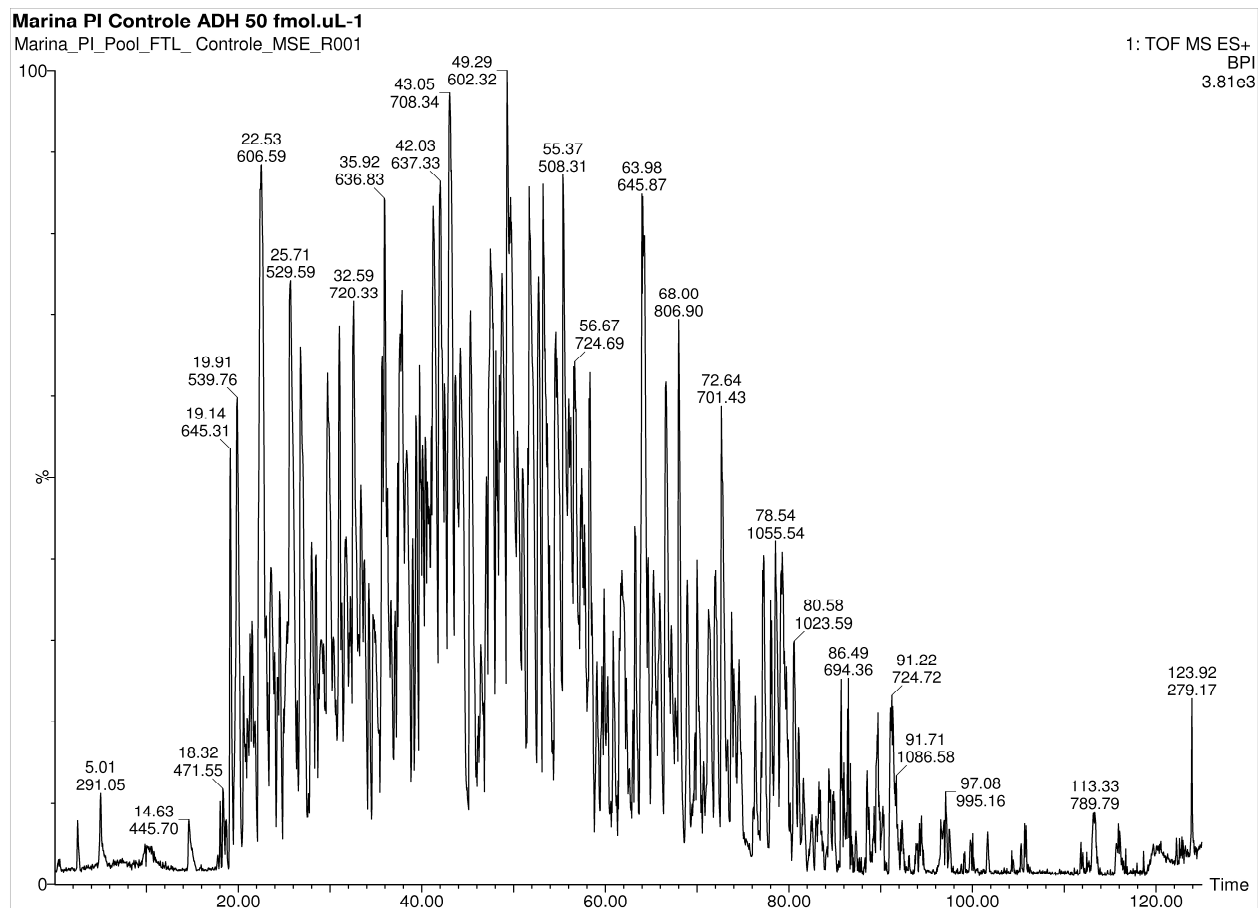
### ***7.5. Análise quali e quantitativa das proteínas fracionadas por meio de cromatografia com Frutalina-Sepharose por espectrometria de massas***

#### ***7.5.1. Identificação e quantificação, por espectrometria de massas, das proteínas plasmáticas presentes nas frações após cromatografias de afinidade em matriz de frutalina imobilizada***

As frações, retida (Pico 1) e não retida (Pico 2), obtidas após cromatografias dos *pools* (referentes a cada grupo em estudo) em coluna de frutalina imobilizada, foram submetidas ao processo de digestão enzimática com tripsina e então, os produtos da digestão, os peptídeos trípticos, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa.

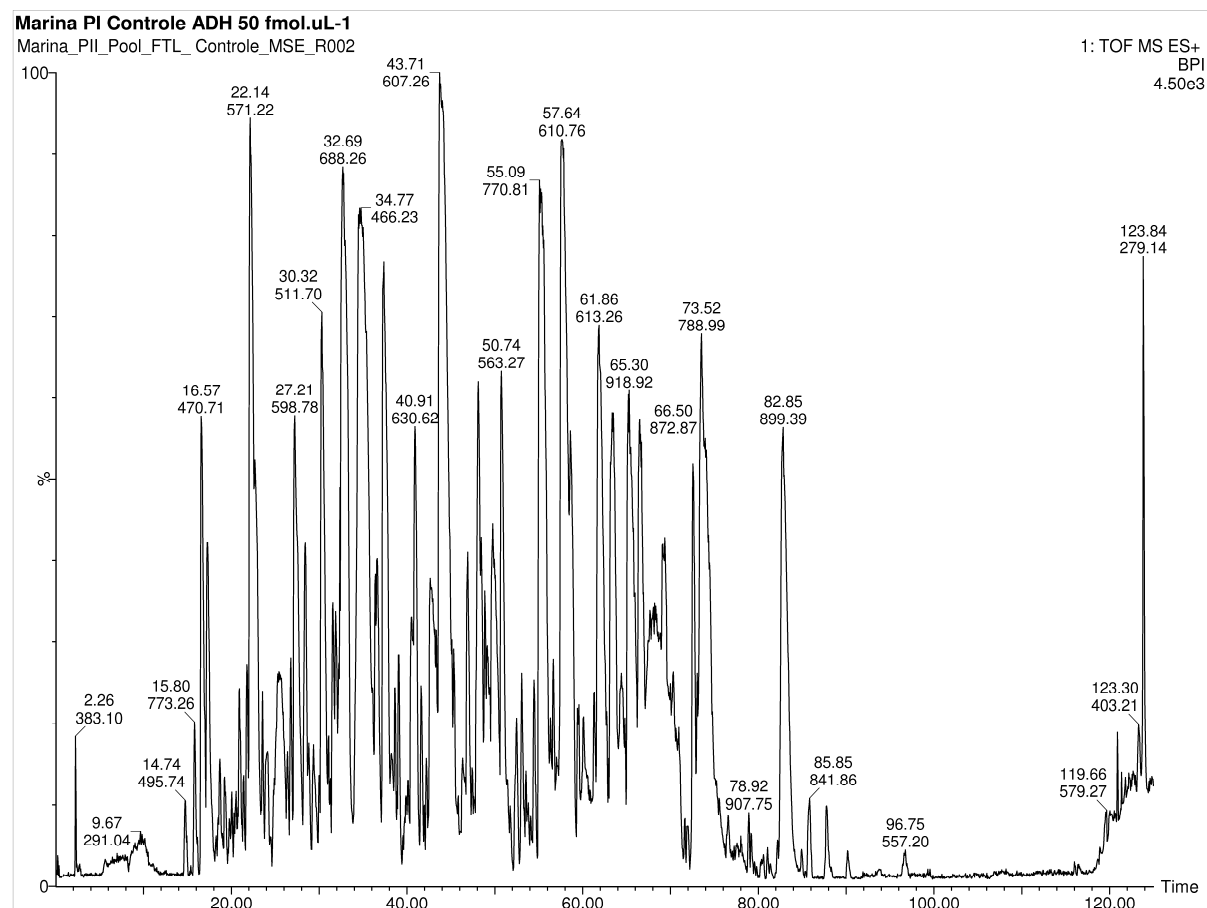
Os peptídeos foram eluídos ao longo de um gradiente de 0% a 40% durante 90 min e 40% a 85% durante 5 min, sob fluxo de 0,35 mL/min, utilizando as fases móveis A e B contendo 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente. Cromatogramas das frações retida (Pico 1) e não retida (Pico 2) após o processo cromatográfico para separação dos peptídeos estão representados pelas Figuras 10 e 11, respectivamente.

Figura 10 – Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração não retida (Pico 1) obtida após cromatografia do *pool controle* em coluna de Frutalina-Sepharose



Os peptídeos trípticos resultantes da digestão da fração não retida (Pico 1), obtida após fracionamento dos *pools* em coluna de FTL-Sepharose, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa C18 HSS T3 com dimensões 1,8 um, 75 um x 20 mm, por 110 min utilizando um gradiente de 0% a 40% durante 90 min, 40% a 85% durante 5 min, em seguida, a coluna foi reequilibrada durante 15 min a 35 °C. O fluxo foi de 0,35 mL/min e as fases móveis A e B continham 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente.

Figura 11 – Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração retida (Pico 2) obtida após cromatografia do *pool controle* em coluna de Frutalina-Sepharose



Os peptídeos trípticos resultantes da digestão da fração retida (Pico 2), obtida após fracionamento dos *pools* em coluna de FTL-Sepharose, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa C18 HSS T3 com dimensões 1,8  $\mu$ m, 75  $\mu$ m x 20 mm, por 110 min utilizando um gradiente de 0% a 40% durante 90 min, 40% a 85% durante 5 min, em seguida, a coluna foi reequilibrada durante 15 min a 35 °C. O fluxo foi de 0,35 mL/min e as fases móveis A e B continham 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente.

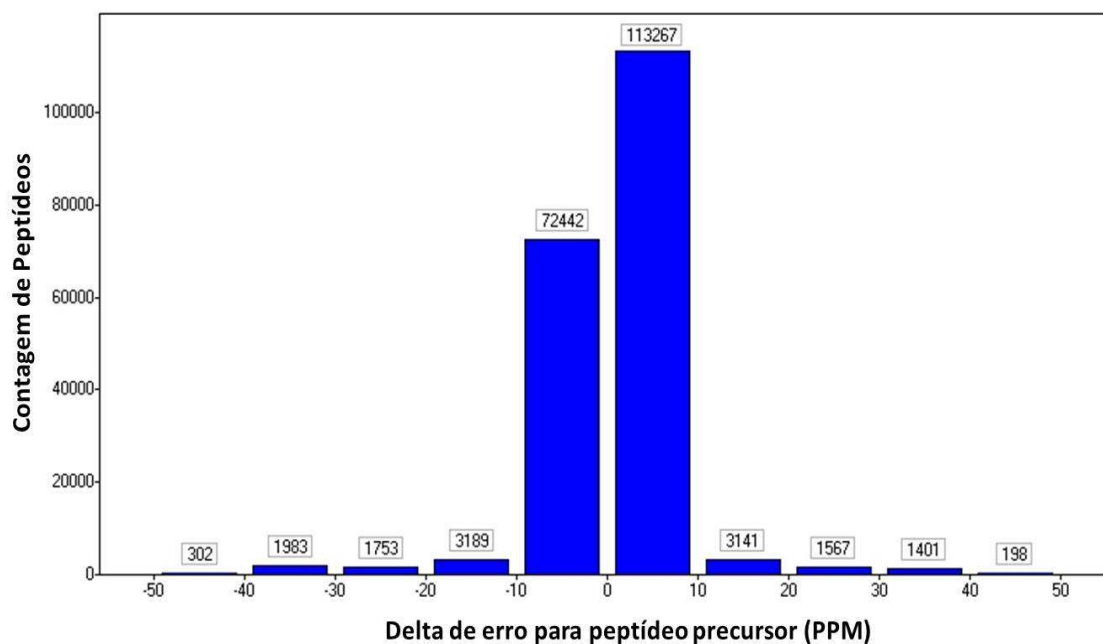
Analisando o perfil cromatográfico dos peptídeos trípticos da fração não retida (Pico 1), observa-se que a eluição dos peptídeos se inicia em torno de 20 minutos após o início do processo cromatográfico e perdura até os 90 minutos. Nesse intervalo de tempo de eluição peptídica, são observados diversos picos de alta intensidade e boa resolução, o que demonstra uma riqueza e qualidade na separação de peptídeos na amostra.

Já na análise do perfil cromatográfico dos peptídeos trípticos da fração retida (Pico 2), observa-se que a eluição dos peptídeos ocorreu em um intervalo de tempo ligeiramente menor, entre 20 e 80 minutos, após início do processo cromatográfico. São observados também picos de alta intensidade e boa resolução, porém em número reduzido quando comparado ao perfil cromatográfico dos peptídeos trípticos da fração não retida (Pico 1), o que indica um número menor de peptídeos presentes na fração retida à coluna de FTL. Fato esse já esperado, tendo em vista a afinidade da lectina especificamente por glicoproteínas com resíduos de  $\alpha$ -D-galactose. Além disso, nota-se claramente que diversos picos são bem mais largos que os vistos no cromatograma do Pico 1. Fato esse justificado pela existência de modificações pós-traducionais, tais como a própria glicosilação que, por muitas vezes, podem interferir no tempo de eluição de um peptídeo.

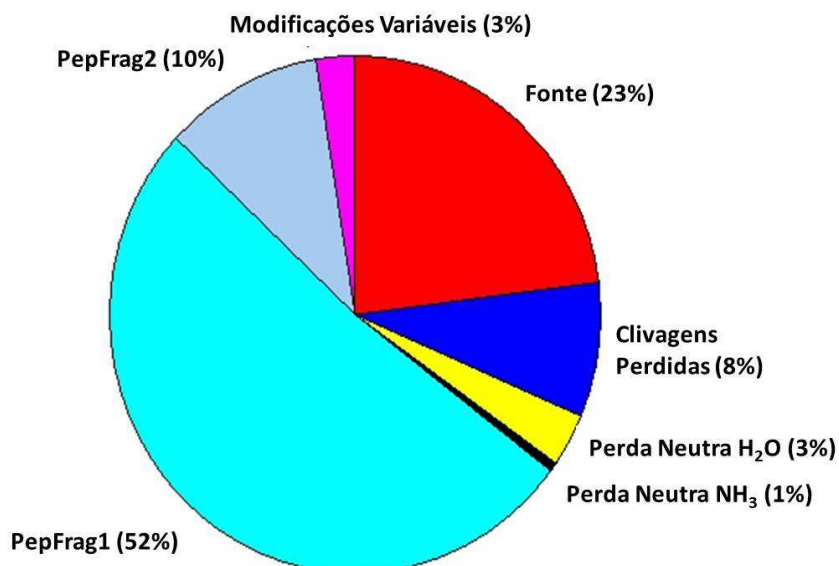
À medida que os peptídeos iam sendo eluídos da coluna de fase reversa, os mesmos iam sendo ionizados e analisados por espectrometria de massas. Portanto, a partir da análise por LC-MS<sup>E</sup>, observou-se a identificação de um total de 195.788 peptídeos (com redundância de peptídeos dentro de um mesmo grupo e entre grupos) nas frações obtidas após cromatografia de afinidade em matriz de FTL-Sepharose para todos os grupos em estudo – controle, câncer de mama ductal estágios I, II e III, incluindo as replicatas.

Esses peptídeos apresentaram uma distribuição normal de erro de massa para precursor ilustrado na Figura 12-A, apresentando um máximo de  $\pm 10$  ppm de erro de massa para cerca de 93% dos peptídeos. Normalmente, existem dois conjuntos de tolerâncias de massa que são setadas para uma pesquisa de banco de dados: tolerância de massa para o peptídeo precursor e tolerância de massa para o fragmento. A tolerância de massa do peptídeo precursor baseia-se na precisão de massa do analisador usado para medir as massas dos íons precursores na varredura MS (*scans*). Em diferentes ferramentas de busca, isto pode ser referido como a tolerância para peptídeo, erro de massa do peptídeo ou mesmo tolerância para precursor. A tolerância de massa para fragmento é baseada na precisão do analisador usado para medir as massas dos íons fragmentos, ou filhos, ou seja, na varredura MS/MS. Este parâmetro, portanto, pode ser referido como tolerância de MS/MS, erro de massa do fragmento, ou tolerância do íon produto (ENG *et al.*, 2011).

Figura 12 - Análise por nanoUPLC-HDMS<sup>E</sup> das frações obtidas após cromatografia dos *pools* em matriz de Frutalina-Sepharose



A - Distribuição de massas de fragmentos e a precisão de massa exata calculada para os íons precursores demonstrando 93% a um máximo de  $\pm 10$  ppm.



B - Distribuição de peptídeos para determinação da qualidade de fragmentação e digestão.

Logo, no presente estudo, essa porcentagem de 93% dos peptídeos identificados apresentando um erro de massa de  $\pm 10$  ppm favoreceu a busca no banco de dados, e conseqüentemente, isso implica em boa taxa de identificação. Haja vista que o parâmetro de tolerância de erro de massa para peptídeo, geralmente dependente da precisão dos espectros de massa MS, é de  $\pm 10$  ppm para instrumentos com analisador TOF (tempo de voo). Portanto, conclui-se que, uma maior tolerância para erro de massa do precursor pode resultar em uma sensibilidade de identificação diminuída, pois pode haver correspondência entre peptídeos de forma aleatória, ou seja, ao acaso (ENG *et al.*, 2011).

Os peptídeos identificados nessa análise apresentaram um FDR (*false discovery rate*) de 1,73%. A crescente busca por padrões de qualidade dentro da área da proteômica fez crescer a importância do chamado FDR (*False Discovery Rate*), ou taxa de falsa descoberta. O FDR é estimado baseado no uso da função “*decoy*” do banco de dados. Essa abordagem envolve a confrontação de espectros MS/MS experimentais com o banco de dados sem sentido e equivalente em tamanho ao banco de dados de sequências de proteínas usado. Na prática, esse banco de dados falso ou sem sentido (*decoy*) é usualmente o reverso do banco de dados usado originalmente. Qualquer correspondência para o banco de dados reverso deve ocorrer puramente ao acaso, sendo então esse número de correspondências usado para calcular o FDR do experimento. Ou seja, muitas vezes o FDR é resultado do número de correspondências falsas dividido pelo número de correspondências totais.

Dos peptídeos identificados nessa análise, apenas 23% dos peptídeos dupla ou triplamente carregados foram produzidos por fragmentação na fonte e uma porcentagem de 8% foi encontrada como resultado de clivagens perdidas (Figura 12-B).

A seletividade e especificidade de uma pesquisa em banco de dados são comprometidas por clivagens perdidas e fragmentação na fonte. Por exemplo, um banco de dados pode exponencialmente aumentar de tamanho de 1,2 milhões para mais de 30 milhões de peptídeos-teóricos sem que se adeque a especificidade da enzima utilizada, a possibilidade de clivagens perdidas e de modificações como oxidação da metionina ou carbamidometilação, pois tudo isso influi na massa detectada pelo equipamento no momento da análise, e que será confrontada com massa teóricas no banco de dados (LI *et al.*, 2009).

As clivagens perdidas acontecem normalmente quando o tempo de digestão *a priori* análise por espectrometria de massa não é adequado, podendo ser curto ou longo demais. Pode acontecer então que a enzima clive em pontos não usuais de clivagem. No caso da

tripsina, enzima usualmente utilizada na espectrometria de massas, ela realiza a hidrólise de polipeptídeos, com elevada especificidade, na região C-terminal de arginina e lisina, se esses resíduos não forem seguidos por uma prolina, gerando peptídeos de 6 a 35 aminoácidos. No entanto, se a tripsina não clivar todos os sítios em potencial com igual probabilidade, obter-se-ão alguns peptídeos experimentais com resíduos internos de arginina ou lisina e, conseqüentemente, valores de massa diferentes do peptídeo teórico presente no banco de dados, gerado após digestão *in silico*. E, por fim, fragmentação na fonte pode resultar na aquisição de espectros MS/MS de peptídeos semi-trípticos que não serão bem identificados (ENG *et al.*, 2011).

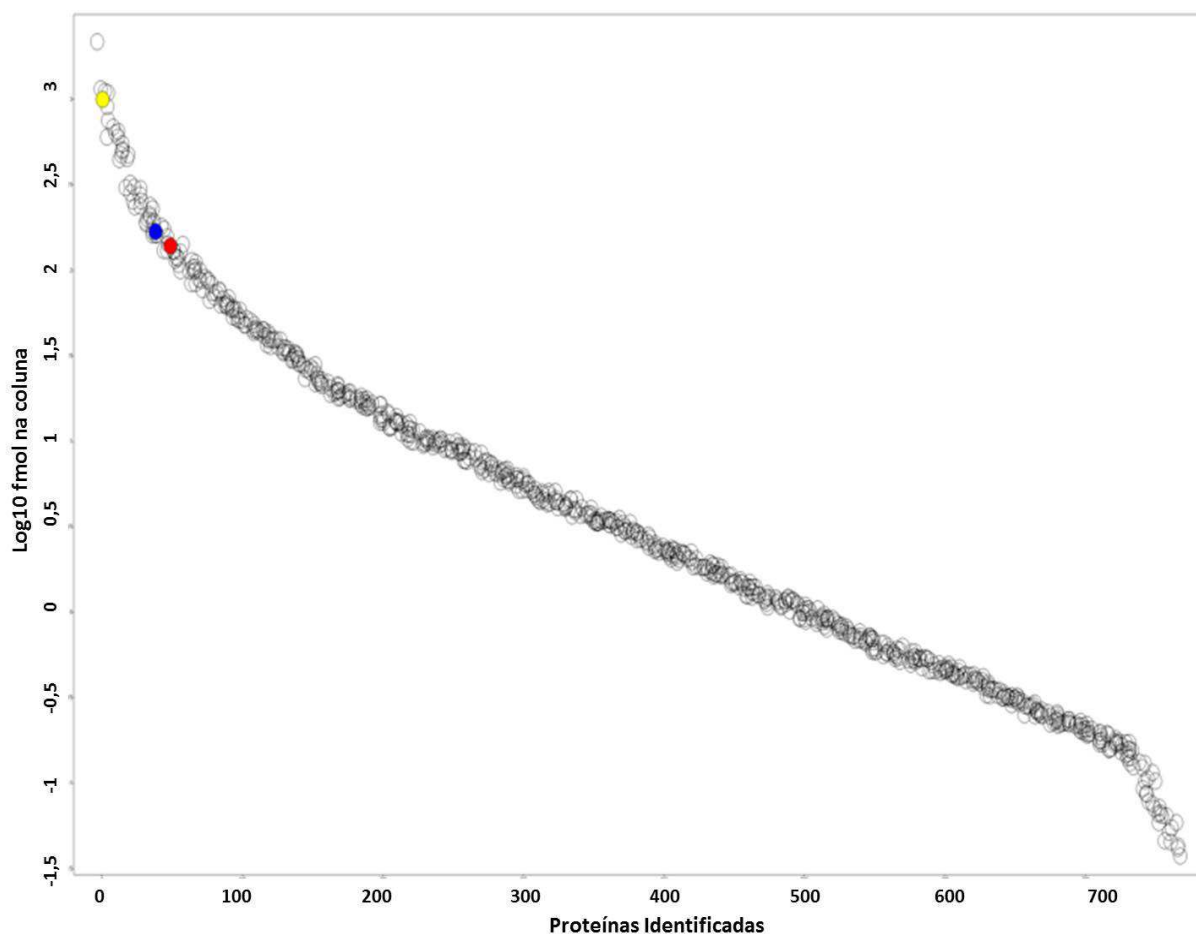
Os peptídeos então identificados foram associados a 14.899 proteínas (com redundância entre isoformas do mesmo grupo e entre grupos), obtendo-se uma média de aproximadamente 13 peptídeos por proteína. Das proteínas identificadas, 8.410 e 6.489 proteínas foram identificadas na fração cromatográfica que não interagiu com a matriz de Frutalina-Sepharose (Pico 1) e na fração que interagiu com a lectina (Pico 2), respectivamente, incluindo as replicatas.

Embora as quantidades de peptídeos (195.788) e proteínas (14.899) sejam elevadas, esses números sofrem redução após os critérios de quantificação, uma vez que diversas isoformas são identificadas, entretanto apenas uma única isoforma para cada proteína é de fato quantificada. Ou seja, dentre todas as identificações de proteínas acima, 765 foram quantificadas e utilizadas para análise quantitativa.

Das proteínas identificadas e quantificadas através do método Hi-3 *label-free*, todas estão em um intervalo de quantificação de 0,1 fmol a  $10^3$  fmol. Ou seja, uma diferença de massa em fmol numa ordem de magnitude 4, como é possível observar no intervalo dinâmico das proteínas identificadas, ilustrado na Figura 13.

Após a identificação e quantificação absoluta das proteínas de cada corrida usando peptídeos trípticos de ADH (UniProt: P00330) como padrão interno, as proteínas alfa-1-antiquimotripsina (UniProt: P01011) e alfa-1-antitripsina (UniProt: P01009), presente em todas as amostras e com baixo coeficiente de variação de massa em fmol, foram selecionadas como proteínas (*housekeeping*) para fins de normalizar os níveis de expressão e realizar a quantificação relativa *label-free*, das proteínas presentes no Pico 1 e Pico 2, respectivamente.

Figura 13 – Intervalo Dinâmico das proteínas identificadas e quantificadas nas frações cromatográficas obtidas após cromatografia dos *pools* em matriz de Frutalina-Sepharose



Após a identificação e quantificação absoluta das proteínas, utilizando álcool desidrogenase de levedura (ADH) (esfera vermelha) como padrão interno, as proteínas alfa-1 antiqumotripsina (AACT\_HUMAN) e alfa-1-antitripsina (A1AT\_HUMAN), representadas pelas esferas azul e amarelo, respectivamente, foram selecionadas como *housekeeping* para normalizar os dados e realizar a quantificação relativa *label-free* de proteínas presentes no pico 1 e pico 2, respectivamente.

Após normalização, obteve-se uma tabela de resultados de expressão que consiste em valores de razão, que refletem os níveis de expressão entre proteínas de um grupo em comparação com outro. Dentre as proteínas identificadas, aquelas que estavam presentes em diferentes níveis nas amostras de plasma de mulheres com estadiamento I, II, e III do câncer ductal de mama foram comparadas com as amostras de mulheres saudáveis.

Vale salientar que, através do PLGS, é gerada esta lista de proteínas descrita, e que foram consideradas para análise de expressão diferencial apenas as proteínas identificadas nas três corridas das replicatas (filtro 3 x 3) e com valores de confiança para  $0,05 \geq p \geq 0,95$ .

Somente proteínas com valores de razão Câncer de mama: Controle acima de 1,5 foram consideradas superexpressas (*up-regulated*) e aquelas com valores de razão Câncer de mama: Controle abaixo de 0,66 foram consideradas subexpressas (*down-regulated*). As proteínas que apresentaram valores de razão entre 0,66 e 1,5 foram consideradas sem expressão diferencial (*unchanged*).

Dessa forma, seguindo os critérios acima estabelecidos foram observadas 102 e 94 proteínas identificadas e quantificadas nas frações correspondentes ao Pico 1 e Pico 2, respectivamente (Tabelas de 1 a 6 - ANEXO).

### 7.5.2. Relevância biológica dos dados proteômicos quantitativos

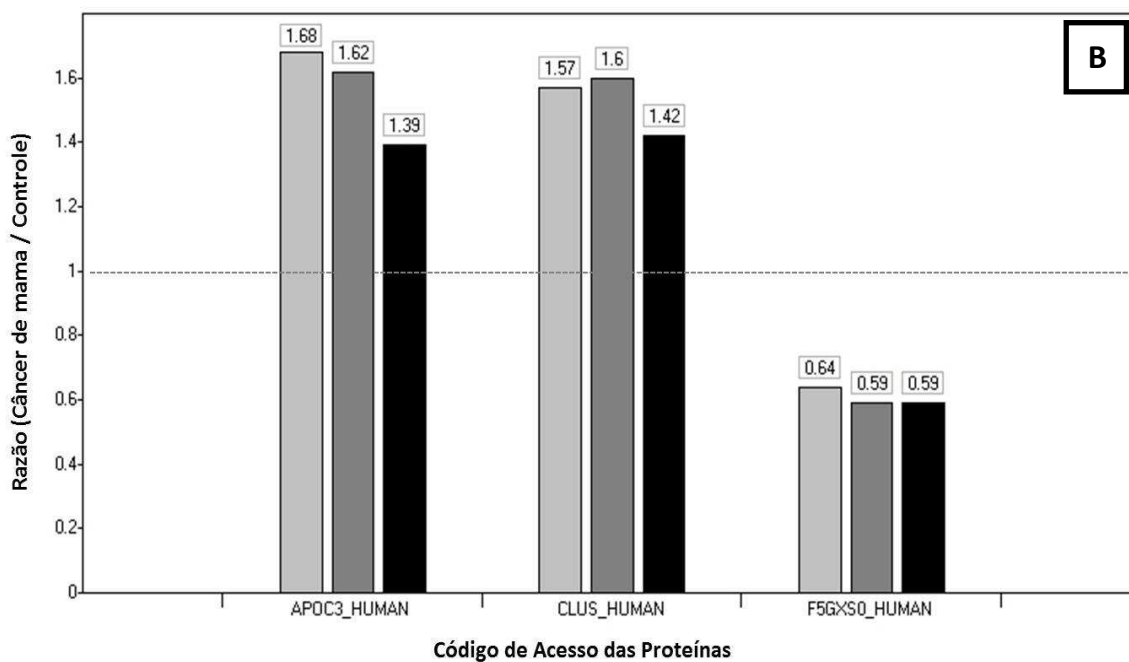
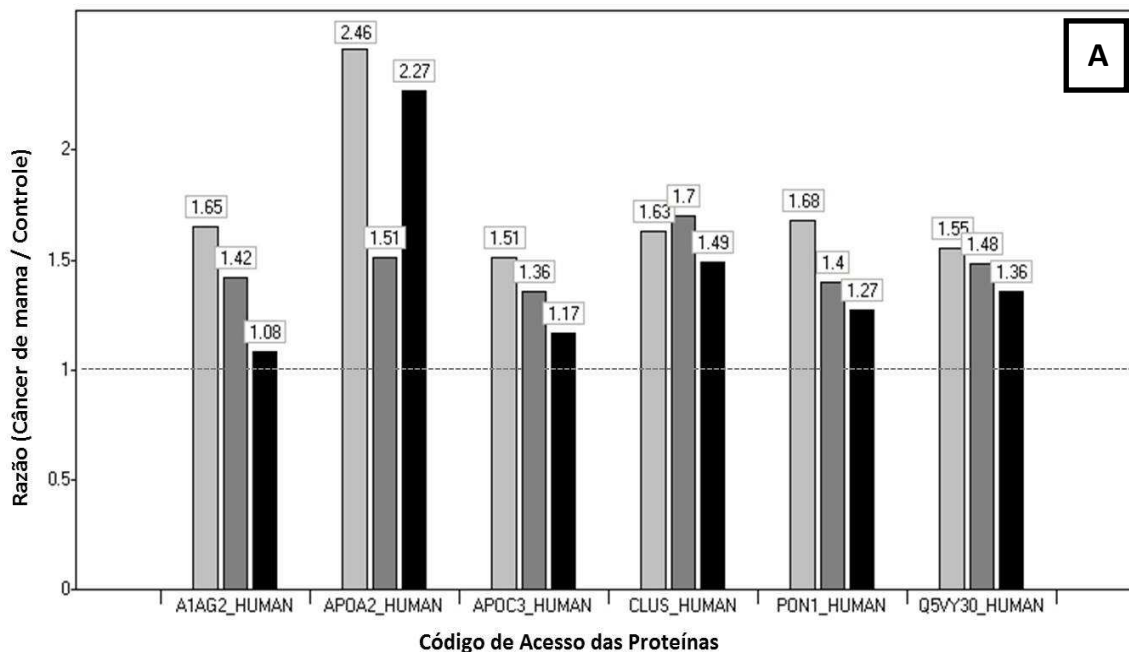
O desenvolvimento de um tumor é um processo de várias etapas, que envolve fatores genéticos, biológicos e ambientais, todos inter-relacionados. A tumorigênese correlaciona-se com fenótipos patológicos, sintomas clínicos e prognóstico, bem como com alterações proteômicas (LI *et al.*, 2011).

Dessa forma, durante este trabalho, em todas as análises proteômicas realizadas, diversas proteínas foram identificadas, proteínas essas que estão relacionadas com processos biológicos comuns, tais como imunidade, hematopoese, inflamação, metabolismo de lipídeos e diferenciação celular.

É digno de nota que algumas proteínas, tais como imunoglobulinas não depletadas e queratina, foram excluídas da discussão da tese haja vista a possibilidade de algumas representarem contaminantes. Portanto, nas discussões que se seguem foram escolhidas proteínas que potencialmente estão envolvidas direta ou indiretamente com o desenvolvimento do câncer de mama.

A Figura 14 sumariza as razões (Câncer de mama: Controle) de algumas proteínas identificadas nos picos 1 e 2, após cromatografia em matriz de frutalina imobilizada, nos diferentes estágios da doença, que serão discutidas a seguir.

Figura 14– Valores de razão (Câncer de mama:Controle) identificados para algumas proteínas encontradas nos picos 1 e 2 após cromatografia de afinidade em FTL-Sepharose, nos diferentes estágios da doença



Valores de razão para as proteínas Q5VY30\_HUMAN (proteína plasmática de ligação ao retinol), PON1\_HUMAN (paraoxonase arilesterase sérica), A1AG2\_HUMAN (glicoproteína alfa-1-2-ácida), APOA2\_HUMAN (apolipoproteína AII), CLUS\_HUMAN (clusterina), APOC3\_HUMAN (apolipoproteína C III) e F5GXS0\_HUMAN (C4b B) detectadas no Pico 1 (A) e no Pico 2 (B) de pacientes com câncer de mama e controles saudáveis. Os diferentes estágios do câncer de mama estão representados da seguinte forma: cinza claro (ductal I - estágio precoce), cinza escuro (ductal II – estágio intermediário) e preto (ductal III – estágio avançado).

### 7.5.2.1. Análise proteômica da fração que não interagiu com a lectina $\alpha$ -galactose ligante frutalina (Pico 1)

A análise proteômica dessa fração revelou um aumento nos níveis da proteína plasmática de ligação ao retinol (Q5VY30\_HUMAN) no estágio inicial do câncer de mama. Esta proteína liga-se a vitamina A e é a principal transportadora específica de vitamina A no sangue. Alguns estudos sugerem que a atividade da vitamina-A no câncer pode ser comprometida em nível do metabolismo do retinol. Além disso, as interações entre os retinóides e os fatores de crescimento tais como fator de crescimento transformante- $\beta$  (TGF- $\beta$ ) estão envolvidos na regulação da proliferação e diferenciação celular (KUPPUMBATTI *et al.*, 2000; FARIAS *et al.*, 2000). Berry, Levi e Noy (2014) demonstraram que a expressão da proteína plasmática de ligação ao retinol e seu receptor *Strab* é acentuadamente regulada em tumores de mama e cólon, e que esse complexo promove propriedades oncogênicas, induzindo fibroblastos a sofrer uma transformação oncogênica, tornando-os altamente tumorigênicos.

Altos níveis da enzima sérica paraoxonase/arilesterase (PON1\_HUMAN) foram detectados principalmente em amostras da fase inicial do câncer de mama. O estresse oxidativo ocorre em muitos tipos de células cancerosas e está associado com a proliferação de células tumorais (BEHREND; HENDERSON; ZWACKA, 2003). A peroxidação lipídica caracteriza dano oxidativo das membranas celulares, lipoproteínas, e outras estruturas contendo lipídeos. A superprodução de subprodutos da peroxidação lipídica e alterações no sistema de defesa antioxidante, bem como nas atividades paraoxonase e arilesterase séricas estão implicadas na patogênese do câncer (CAMUZCUOGLU *et al.*, 2009; BALCI *et al.*, 2012). Assim, a paraoxonase/arilesterase sérica superexpressa na presente análise poderia ser uma estratégia biológica para acelerar a eliminação de produtos da peroxidação lipídica, diminuindo o estresse oxidativo gerado.

Recentemente, Carioca e colaboradores (2015) avaliaram a associação entre adiposidade e marcadores de estresse oxidativo com o estadiamento clínico do câncer de mama. Como resultado, percebeu-se que os níveis de antioxidantes foram maiores em mulheres com estadiamento clínico precoce e intermediário. A adiposidade e estadiamento clínico dos pacientes foram associados positivamente com estresse oxidativo (CARIOCA *et al.*, 2015).

Um elevado nível da  $\alpha$ -1-2-ácido glicoproteína - AAG (A1AG2\_HUMAN) foi identificado no estágio precoce do câncer da mama e o seu nível apresentou-se ligeiramente aumentado na fase intermediária do câncer de mama. A alfa-1-2-ácido glicoproteína é uma proteína de fase aguda do plasma, com uma proporção elevada de resíduos de aminoácidos glicosilados. Os níveis observados de AAG nesse estudo, 1,65 vezes mais elevados no plasma de mulheres com câncer da mama em comparação com os de mulheres saudáveis, são corroborados por achados que relatam a concentração média de AAG plasmática de, normalmente, duas vezes maior em pacientes com câncer de mama, pulmão e ovário, quando comparada com a média de pessoas saudáveis (DUCHÉ *et al.*, 2000). Além disso, é observado um aumento de 5 vezes na concentração de AAG em várias respostas de fase aguda, tais como a inflamação, o stress, a gravidez, e infarto do miocárdio (HOCHEPIED *et al.*, 2003).

Altos níveis de apolipoproteína A-II (APOA2\_HUMAN) foram detectados e apresentaram-se significativamente elevados nos estágios precoce e avançado da doença. A síntese de lipídeos sendo resultado integrado de atividade genética, epigenética e fatores ambientais, parece promover o crescimento e a sobrevivência de células cancerosas. Algumas apolipoproteínas desempenham um papel central no metabolismo de lipídeos e alterações na expressão de ApoA2 ocorrem em outros cânceres além do câncer de mama (MALIK *et al.*, 2005; LIU *et al.*, 2013).

Podzielinski e colaboradores (2013) sugeriram que o metabolismo de lipoproteínas é desregulado no câncer de ovário e que merece uma investigação mais aprofundada no intuito de utilizar ApoA2 como um marcador para tumores de ovário.

Além disso, há uma clara associação entre a obesidade e o câncer de mama. O microambiente do tumor, particularmente o tecido adiposo, e o desenvolvimento de câncer de mama, parecem ter uma intrincada relação (SAUTER *et al.*, 2008; SCHAUB *et al.*, 2009).

#### 7.5.2.2. *Análise proteômica da fração que interagiu com a lectina $\alpha$ -galactose ligante frutalina (Pico 2)*

O nível de apolipoproteína C-III (APOC3\_HUMAN) apresentou-se aumentado principalmente no estágio precoce do câncer de mama, em comparação com o controle (razão de 1,5), no pico 1, porém níveis ainda mais elevados dessa proteína foram encontrados no

pico 2 de mulheres com câncer de mama, tanto no estágio precoce como intermediário da doença.

APOC3 é uma proteína constituinte da lipoproteína de muito baixa densidade (VLDL) que é sintetizada no fígado e no intestino. Um alto nível de APOC3 é comumente correlacionado com hipertrigliceridemia e frequentemente associada à doença coronariana e aterosclerose (BATAL *et al*, 2000; ONAT *et al*, 2003).

Já é descrito na literatura o padrão de glicosilação da APOC3 (BREWER *et al.*, 1974; VAITH, P.; ASSMANN, G.; UHLENBRUCK, G. 1978). É uma proteína glicosilada em treonina (74) e há três glicofomas da proteína que compartilham um *core* tipo mucina O-glicosilada (glicano  $\beta$ 1-3 galactose) ligados a resíduo de *N*-acetilgalactosamina (Gal $\beta$ 1-3GalNAc). Duas dessas três isoformas ocorrem com um glicano modificado com um ou dois resíduos de ácido siálico (Gal $\beta$ 1-3[NeuAc $\alpha$ 2-6]GalNAc e NeuAc $\alpha$ 2-3Gal $\beta$ 1-3[NeuAc $\alpha$ 2-6]GalNAc).

Um fato curioso é que APOC3 foi detectada em ambas as frações cromatográficas utilizando matriz imobilizada com frutalina. Porém, diferencialmente expressa principalmente no pico 2, ou seja, uma fração de APOC3 presente nas amostras de câncer de mama, interagiu, reversivelmente com a frutalina imobilizada. Vale ressaltar que frutalina apresenta afinidade por resíduos de  $\alpha$ -D-galactose e, de forma anomérica-específica a carboidratos complexos que contenham glicanos do tipo Gal $\alpha$ 1-3 (MOREIRA *et al*, 1998; TEIXEIRA *et al*, 2007; OLIVEIRA *et al.*, 2009; NOBRE *et al*, 2010).

Tendo em vista que APOC3 é glicosilada principalmente com um conteúdo de glicanos do tipo Gal $\beta$ 1-3, e os resultados encontrados neste trabalho mostram uma super expressão dessa proteína em mulheres com câncer de mama, quando comparados com mulheres saudáveis, principalmente no pico 2, suspeita-se, portanto, que haja uma mudança no padrão de glicosilação de APOC3 durante o desenvolvimento do câncer de mama, fato esse que justifica essa proteína ter apresentado capacidade de interagir com a frutalina imobilizada. Dessa forma, parece haver uma substituição de resíduos galactose  $\beta$ 1-3 para galactose  $\alpha$ 1-3 em APOC3 durante a tumorigênese. Essa possível mudança na glicosilação é intrigante e corroborada pela forte relação entre as alterações na maquinaria de glicosilação e a biologia do câncer.

APOC3 é extensivamente estudada para utilização pelo seu potencial uso como biomarcador, porque as alterações na razão entre as suas diferentes glico-isoformas ocorrem

na obesidade, em doenças renais, doenças hepáticas e sepse (JIAN *et al.*, 2013). No entanto, até o momento, não há informações publicadas indicando uma associação de alteração no padrão de glicosilação da APOC3 com o câncer de mama.

Os níveis de clusterina (CLUS\_HUMAN) apresentaram-se aumentados no plasma de mulheres com câncer de mama. Diferentemente dos resultados encontrados para APOC3, os níveis de clusterina foram mais elevados em ambas as frações (pico 1 e pico 2), nos estágios precoce e intermediário da doença.

Clusterina é uma chaperona extracelular envolvida no reparo do DNA, na regulação do ciclo celular, na morte celular por apoptose e na tumorigênese (ROSENBERG; SILKENSEN, 1995). É intensamente glicosilada e cerca de 25% da sua massa compreende diferentes glicanos ligados a asparagina (STEWART *et al.*, 2007) o que pode justificar sua identificação em ambas as frações.

Redondo e colaboradores (2000) avaliaram a expressão da clusterina em uma série de amostras teciduais, entre elas lesões benignas e carcinomas da mama, por imuno-histoquímica e hibridização *in situ*. Como resultado, foi observado que células epiteliais normais foram negativas para a expressão de clusterina e apenas 19% das lesões benignas apresentaram coloração positiva. Por fim, a expressão da clusterina foi associada, de forma proporcional, ao tamanho do tumor e com a progressão do carcinoma primário para o carcinoma metastático nos nódulos linfáticos. Comparando os resultados de Redondo e colaboradores com os encontrados neste trabalho, parece não haver essa relação proporcional entre o avanço do câncer de mama e os níveis de clusterina quando se trata de amostras de plasma.

A expressão de clusterina apresenta-se alterada em diferentes cânceres, incluindo o câncer de mama (NAKAMURA *et al.*, 2006; TOUSI *et al.*, 2012; MU *et al.*, 2013), a qual está associada com o tamanho do tumor, status negativo para receptores de progesterona e progressão de carcinoma primário para metástase nos linfonodos (WANG *et al.*, 2012).

O silenciamento da expressão de clusterina usando pequenos RNAs de interferência (siRNA) diminuiu a invasão e a migração de células do câncer da mama e inibiu o crescimento celular e a progressão metastática em um modelo ortotópico, indicando que a clusterina pode contribuir para os efeitos inibidores de crescimento no microambiente tumoral da mama (NIU *et al.*, 2012).

Uma diminuição dos níveis do fator 4b do sistema complemento - C4b\_B (F5GXS0\_HUMAN) foi observado no pico 2 das amostras de câncer de mama, em todos os

estágios, em relação às amostras controle. Sabe-se que o sistema complemento desempenha um papel importante na imunidade inata e imunidade a tumores e, embora alguns estudos discutam a relação do processo inflamatório no microambiente imunológico dos tumores, ainda é um pouco controversa a influência dos componentes do sistema complemento na tumorigênese.

Sabendo que essa proteína é glicosilada, o decréscimo observado em seus níveis no plasma de mulheres com câncer de mama, em relação ao plasma de mulheres saudáveis, pode ser explicado por uma possível redução na glicosilação dessa proteína durante o desenvolvimento da doença, haja vista que os valores de razões encontrados apresentam uma tendência a sofrer redução à medida que o câncer evolui. A glicosilação anormal de proteínas, seja ela aumentada ou diminuída, está correlacionada com várias doenças, incluindo câncer (LAZAR *et al*, 2011).

## ***7.6. Análise quali e quantitativa das proteínas fracionadas por meio de cromatografia com DAL-Sepharose por espectrometria de massas***

### ***7.6.1. Identificação e quantificação, por espectrometria de massas, das proteínas plasmáticas presentes nas frações após cromatografias de afinidade em matriz de DAL imobilizada***

As frações retida (Pico 1) e não retida (Pico 2), obtidas após cromatografias dos *pools* (referentes a cada grupo em estudo) em coluna de DAL imobilizada, foram submetidas ao processo de digestão enzimática e então, os produtos da digestão, os peptídeos trípticos, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa.

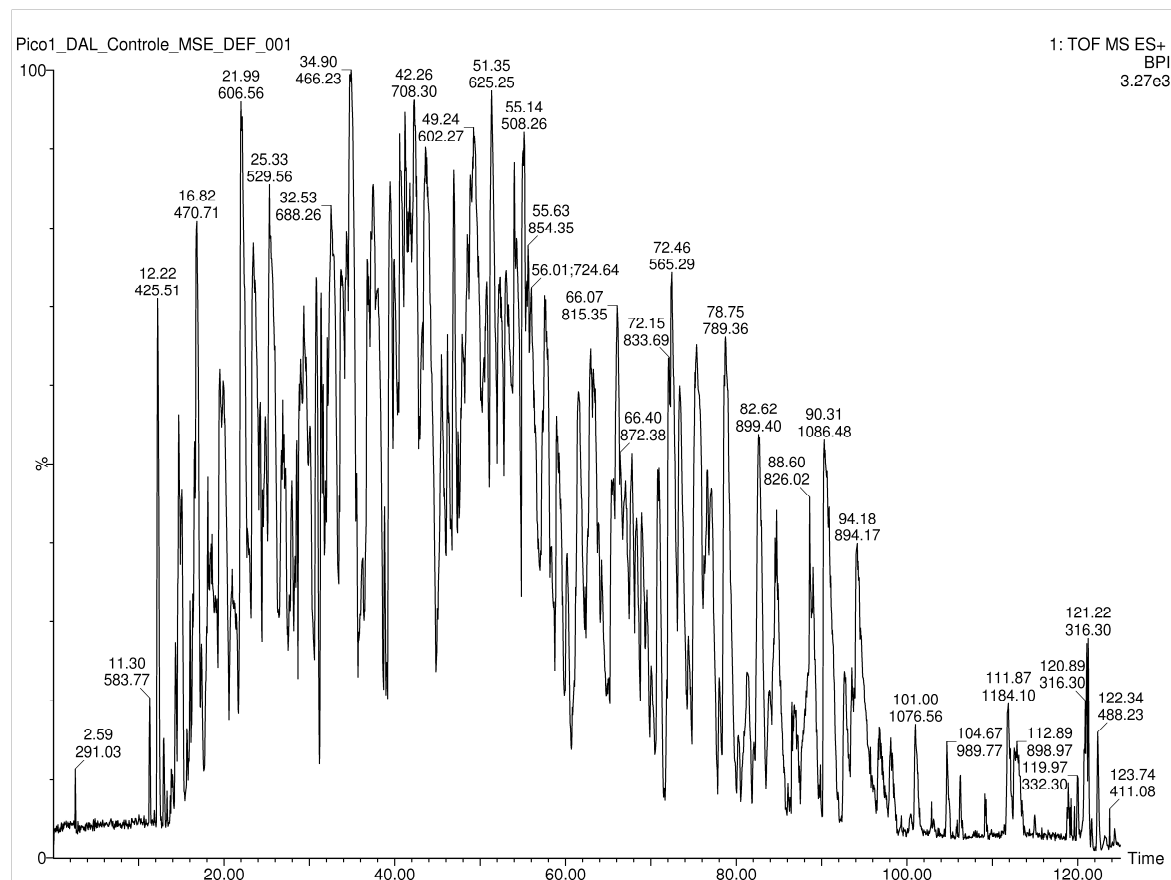
Os peptídeos foram eluídos ao longo de um gradiente de 0% a 40% durante 90 min e 40% a 85% durante 5 min, sob fluxo de 0,35 mL/min, utilizando as fases móveis A e B contendo 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente. Cromatogramas das frações retida (Pico 1) e não retida (Pico 2) após o processo cromatográfico para separação dos peptídeos estão representados pelas Figuras 15 e 16, respectivamente.

Em uma análise do perfil cromatográfico dos peptídeos trípticos da fração não retida (Pico 1), observa-se que a eluição dos peptídeos se inicia em torno de 10 minutos após o início do processo cromatográfico e perdura até os 100 minutos, portanto apresentando um intervalo de tempo de eluição peptídica maior do que foi observado para o Pico 1 de Frutalina.

Já na análise do perfil cromatográfico dos peptídeos trípticos da fração retida (Pico 2), obtida após fracionamento em coluna de DAL imobilizada, observa-se que a eluição dos peptídeos ocorreu também em um intervalo de tempo ligeiramente menor, entre os 20 e 80 minutos, após início do processo cromatográfico, assim como observado no cromatograma com as amostras de Pico 2 resultantes do fracionamento com frutalina. Mais uma vez, são observados picos de alta intensidade, melhor resolução e em número reduzido, quando comparado ao perfil cromatográfico dos peptídeos trípticos da fração não retida (Pico 1).

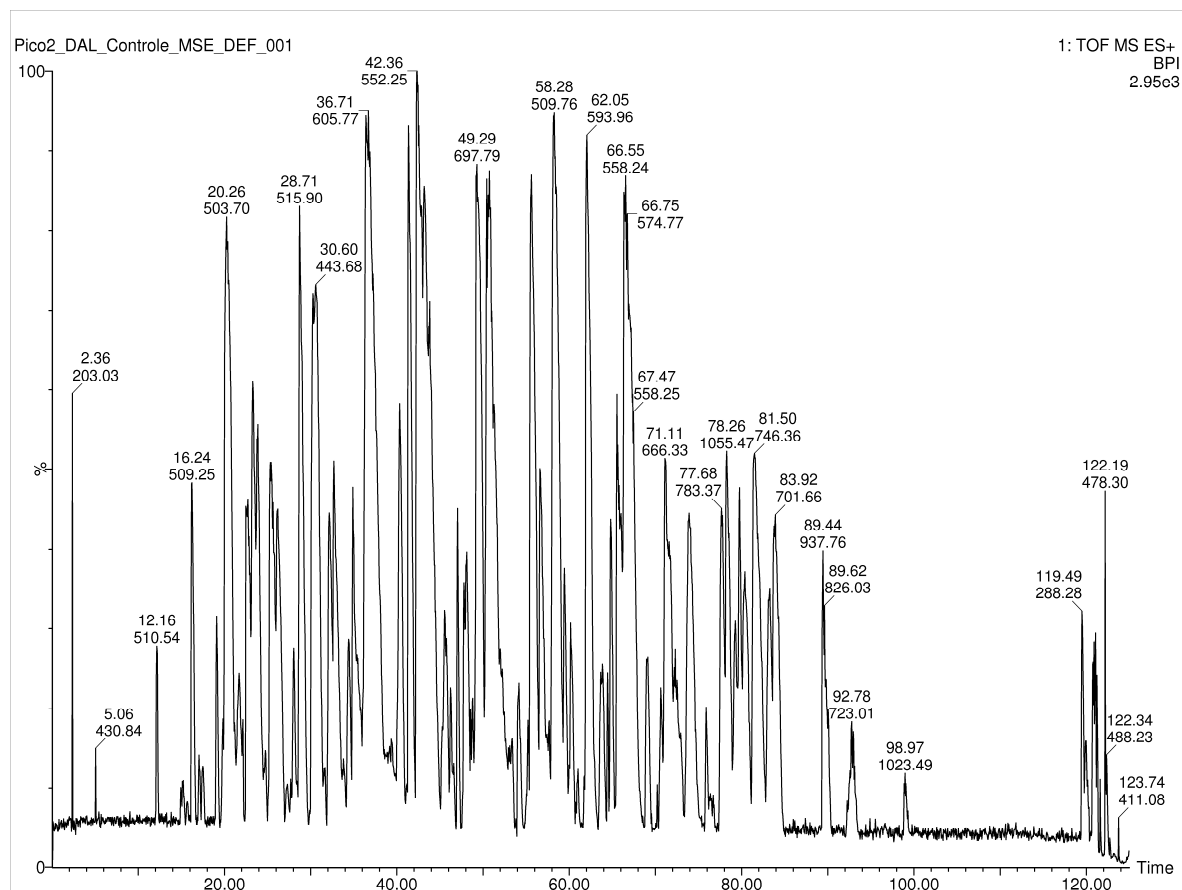
Comparando-se, portanto, os perfis cromatográficos dos peptídeos trípticos das frações obtidas após fracionamento nas diferentes colunas de lectinas vegetais imobilizadas, infere-se que houve um menor fracionamento quando utilizada a coluna de DAL-Sepharose, haja vista a presença de picos com menor resolução, quando comparados àqueles observados

Figura 15 – Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração não retida (Pico 1) obtida após cromatografia do *pool controle* em coluna de DAL-Sepharose



Os peptídeos trípticos resultantes da digestão da fração não retida (Pico 1), obtida após fracionamento dos *pools* em coluna de DAL-Sepharose, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa C18 HSS T3 com dimensões 1,8 um, 75 um x 20 mm, por 110 min utilizando um gradiente de 0% a 40% durante 90 min, 40% a 85% durante 5 min, em seguida, a coluna foi reequilibrada durante 15 min a 35 °C. O fluxo foi de 0,35 mL/min e as fases móveis A e B continham 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente.

Figura 16 – Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos da fração retida (Pico 2) obtida após cromatografia do *pool controle* em coluna de DAL-Sepharose



Os peptídeos trípticos resultantes da digestão da fração retida (Pico 2), obtida após fracionamento dos *pools* em coluna de DAL-Sepharose, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa C18 HSS T3 com dimensões 1,8  $\mu$ m, 75  $\mu$ m x 20 mm, por 110 min utilizando um gradiente de 0% a 40% durante 90 min, 40% a 85% durante 5 min, em seguida, a coluna foi reequilibrada durante 15 min a 35 °C. O fluxo foi de 0,35 mL/min e as fases móveis A e B continham 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente.

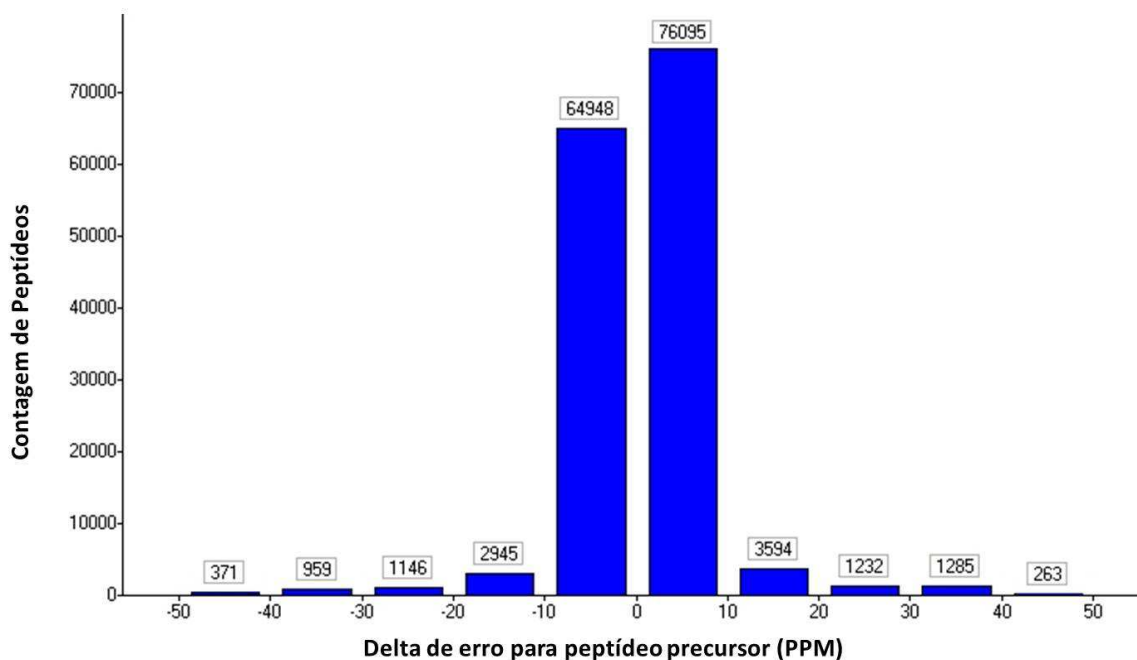
nas amostras obtidas após cromatografia em coluna de FTL-Sepharose, o que reflete abundância maior de peptídeos, e o que, conseqüentemente, pode dificultar identificação e quantificação subsequentes.

À medida que os peptídeos iam sendo eluídos da coluna de fase reversa, os mesmos iam sendo ionizados e analisados por espectrometria de massas. Portanto, a partir da análise por LC-MS<sup>E</sup>, observou-se a identificação de um total de 150.541 peptídeos (com redundância de peptídeos dentro de um mesmo grupo e entre grupos) nas frações obtidas após cromatografia de afinidade em matriz de DAL-Sepharose para todos os grupos em estudo – controle, câncer de mama ductal estágios I, II e III, incluindo as replicatas.

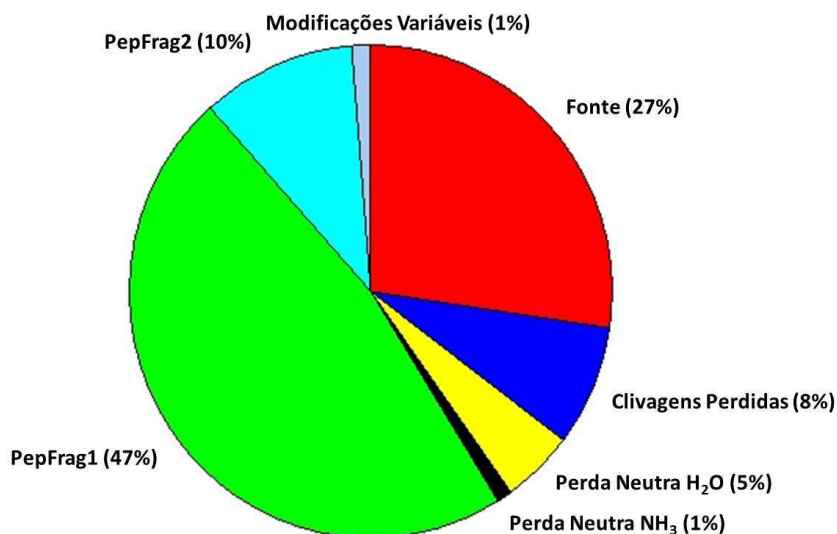
Esses peptídeos identificados apresentaram uma distribuição normal de erro de massa para precursor ilustrado na Figura 17-A, apresentando um máximo de  $\pm 10$  ppm de erro de massa para cerca de 93% dos peptídeos e um FDR (*false discovery rate*) de 1,56%.

Um total de 27% de peptídeos dupla ou triplamente carregados foram produzidos por fragmentação na fonte e 8% dos peptídeos resultaram de clivagens perdidas (Figura 17-B).

Figura 17 - Análise por nanoUPLC-HDMS<sup>E</sup> das frações obtidas após cromatografia dos *pools* em matriz de DAL-Sepharose



A - Distribuição de massas de fragmentos e a precisão de massa exata calculada para os íons precursores demonstrando 93% a um máximo de  $\pm 10$  ppm.



B - Distribuição de peptídeos para determinação da qualidade de fragmentação e digestão.

Os peptídeos então identificados foram associados a 9.974 proteínas (com redundância entre isoformas do mesmo grupo e entre grupos), obtendo-se uma média de aproximadamente 15 peptídeos por proteína. Das proteínas identificadas, 7.487 e 2.127 proteínas foram identificadas na fração cromatográfica que não interagiu com a matriz de DAL-Sepharose (Pico 1) e na fração que interagiu com a lectina (Pico 2), respectivamente, incluindo as replicatas.

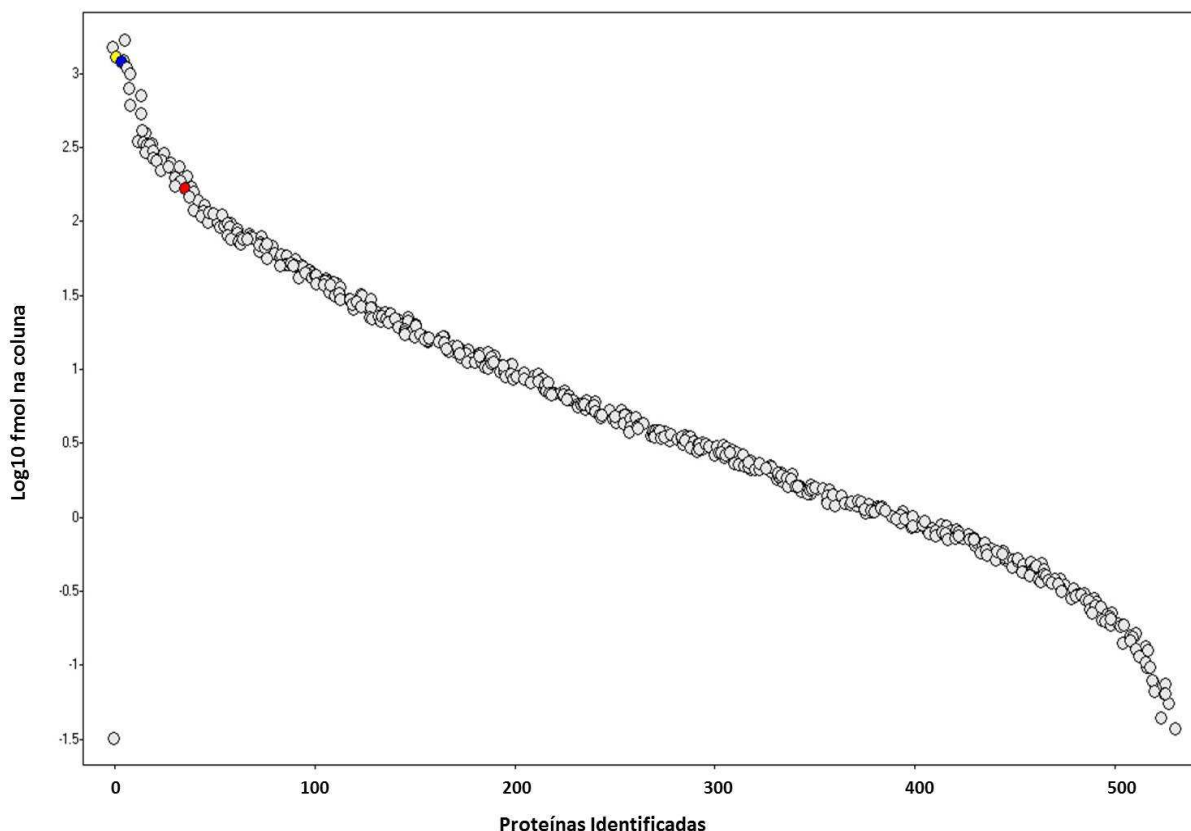
Mais uma vez, embora as quantidades de peptídeos (150.541) e proteínas (9.974) sejam elevadas, esses números sofreram redução após os critérios de quantificação uma vez que diversas isoformas são identificadas, entretanto apenas uma única isoforma para cada proteína é de fato quantificada. Ou seja, dentre todas as identificações de proteínas acima, 529 foram quantificadas e utilizadas para análise quantitativa.

Das proteínas identificadas e quantificadas através do método Hi-3 *label-free*, todas estão em um intervalo de massas de 0,1 fmol a 10<sup>3</sup> fmol. Ou seja, uma diferença de massa em fmol numa ordem de magnitude 4, como é possível observar o intervalo dinâmico das proteínas identificadas, ilustrado na Figura 18.

Após a identificação e quantificação absoluta das proteínas de cada corrida usando peptídeos trípticos de ADH (UniProt: P00330) como padrão interno, alfa-1-antitripsina (UniProt: P01009) e alfa-2-macroglobulina (UniProt: P01023) foram selecionadas como

proteínas *housekeeping* para fins de normalizar os níveis de expressão e realizar a quantificação *label-free*, das proteínas presentes no Pico 1 e Pico 2, respectivamente.

Figura 18 – Intervalo Dinâmico das proteínas identificadas e quantificadas nas frações cromatográficas obtidas após cromatografia dos *pools* em matriz de DAL-Sepharose



Após a identificação e quantificação absoluta das proteínas, utilizando álcool desidrogenase de levedura (ADH) (esfera vermelha) como padrão interno, as proteínas alfa-1-antitripsina (A1AT\_HUMAN) e alfa-2-macroglobulina (A2MG\_HUMAN) representadas pelas esferas azul e amarelo, respectivamente, foram selecionadas como *housekeeping* para normalizar os dados e realizar a quantificação *label-free* de proteínas presentes no pico 1 e pico 2, respectivamente.

Seguindo o mesmo padrão de análise proteômica utilizado com as amostras resultantes das cromatografias utilizando frutalina, dentre as proteínas identificadas, aquelas que estavam presentes em diferentes níveis nas amostras de plasma de mulheres com estadiamento I, II, e III do câncer ductal de mama foram comparadas com as amostras de mulheres saudáveis.

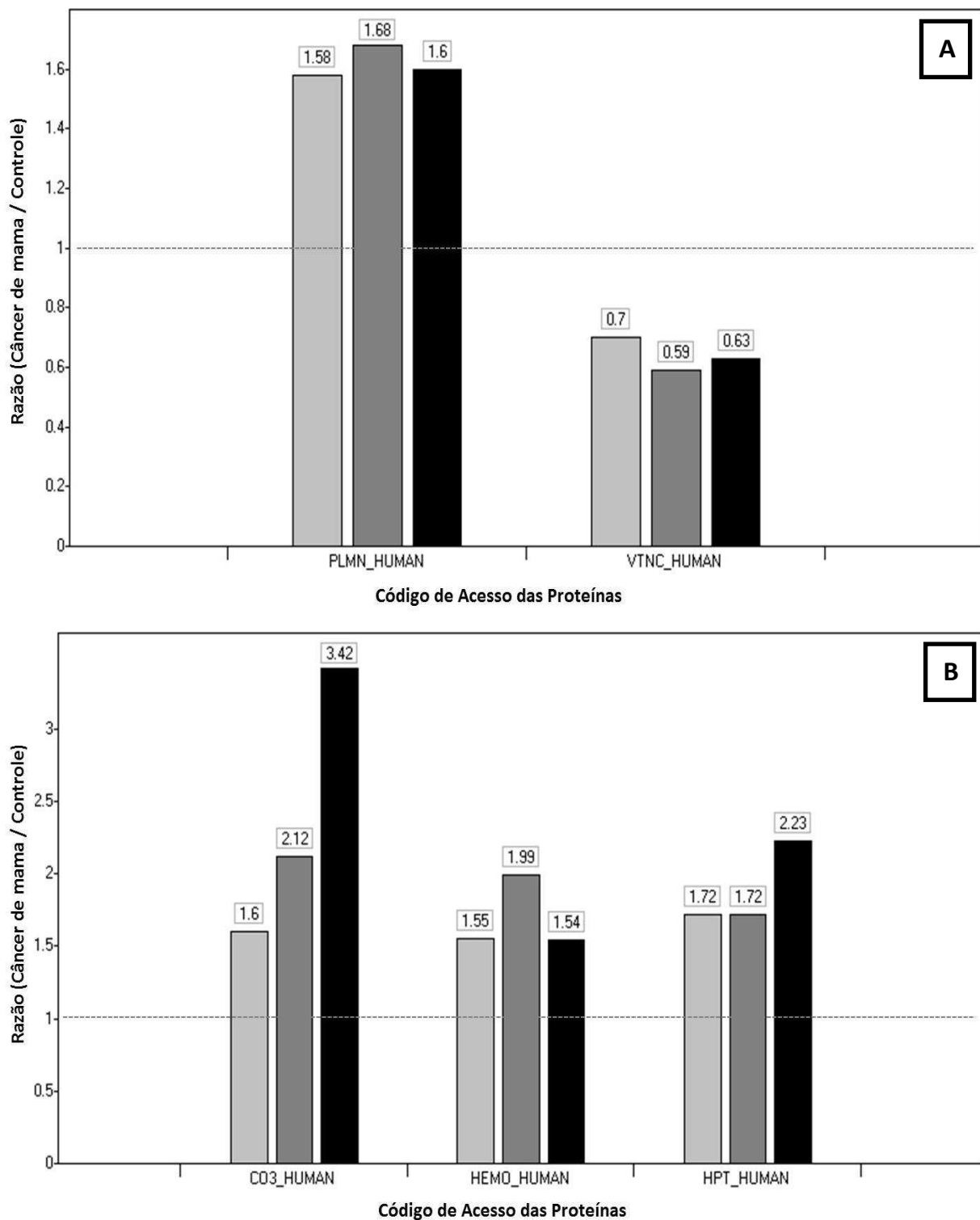
Mais uma vez reforça-se que, a partir da lista de proteínas gerada pelo PLGS, foram consideradas para análise de expressão diferencial somente as proteínas identificadas nas três corridas das replicatas (filtro 3 x 3) e com valores de confiança para  $0,05 \geq p \geq 0,95$ .

E, mais uma vez, somente proteínas com valores de razão Câncer de mama: Controle acima de 1,5 foram consideradas superexpressas (*up-regulated*) e aquelas com valores de razão Câncer de mama: Controle abaixo de 0,66 foram consideradas subexpressas (*down-regulated*). As proteínas que apresentaram valores de razão entre 0,66 e 1,5 foram consideradas sem expressão diferencial (*unchanged*).

Portanto, nos critérios acima estabelecidos, foram observadas 78 e 10 proteínas identificadas e quantificadas nas frações correspondentes ao Pico 1 e Pico 2, respectivamente (Tabelas de 7 a 12).

A Figura 19 sumariza as razões (Câncer de mama: Controle) de algumas proteínas identificadas nos picos 1 e 2, nos diferentes estágios da doença, que serão discutidas a seguir.

Figura 19 – Valores de razão (Câncer de mama:Controle) identificados para algumas proteínas encontradas nos picos 1 e 2 após cromatografia de afinidade em DAL-Sepharose, nos diferentes estágios da doença



Valores de razão para as proteínas plasminogênio (PLMN\_HUMAN), vitronectina (VTNC\_HUMAN), hemopexina (HEMO\_HUMAN), hatoglobina (HPT\_HUMAN) e C3 (CO3\_HUMAN) detectadas no Pico 1 (A) e no Pico 2 (B) de pacientes com câncer de mama e controles saudáveis. Os diferentes estágios do câncer de mama estão representados da seguinte forma: cinza claro (ductal I - estágio precoce), cinza escuro (ductal II – estágio intermediário) e preto (ductal III – estágio avançado).

## 7.6.2. Relevância biológica dos dados proteômicos quantitativos

### 7.6.2.1. Análise proteômica da fração que não interagiu com a lectina manose-ligante de *Dioclea altissima* - DAL (Pico 1)

Dentre as proteínas identificadas e quantificadas na fração cromatográfica que não interagiu com a matriz de DAL-Sepharose, o plasminogênio (PLMN\_HUMAN) apresentou-se superexpresso em todos os estágios do câncer de mama, em comparação com o controle (valores de razão entre 1,58 e 1,68).

Hoje se tem uma imagem clara, embora complexa, dos mecanismos gerais subjacentes a degradação proteolítica da matriz extracelular no câncer. A decomposição da matriz extracelular é crucial para a invasão do câncer e, conseqüentemente, metástases. Isto é alcançado através da ação combinada de várias proteases, incluindo a protease serínica plasmina e inúmeras metaloproteinases. A geração de plasmina envolve a pró-enzima plasminogênio, resultado da cascata de reações proteolíticas do sistema uroquinase de ativação do plasminogênio (uPA - *urokinase plasminogen activator*), que pertence à família de proteases serínicas, que inclui o ativador de plasminogênio do tipo uroquinase (*urokinasetype plasminogen activator* - uPA), um receptor celular ancorado na membrana glicolipídica (*glycolipid-anchored cell membrane receptor for the uPA* - uPAR) e inibidores da ativação do plasminogênio (*plasminogen activator inhibitors* – PAIs) (DANO *et al.*, 2005).

O plasminogênio, zimogênio central nesse processo, é produzido no fígado e secretado como uma proteína de cadeia única e glicosilada com um N-terminal contendo um resíduo de ácido glutâmico (Glu), cinco regiões contendo sítios de ligação à lisina que regulam a ligação e ativação do plasminogênio, para sua conversão em plasmina, e no C-terminal apresenta um domínio proteolítico (PONTING; MARSHALL; CEDERHOLM-WILLIAMS, 1992).

Levando em consideração que cerca de 25 a 40% dos cânceres de mama evoluem para metástases, mulheres com algumas das proteínas desse sistema alteradas são sinalizadoras de mau prognóstico no câncer de mama primário (TANG; HAN, 2013). O ativador de plasminogênio do tipo uroquinase (uPA) e um inibidor da ativação do plasminogênio tipo I (PAI-1) são fatores biológicos de prognóstico tumoral confirmados a respeito de sua utilidade clínica no câncer de mama (HARBECK *et al.*, 2002).

Como um resultado semelhante ao encontrado no presente estudo, em que os níveis de plasminogênio apresentaram superexpressos em todos os três estágios do câncer, porém com razão maior no plasma de mulheres em estágio intermediário, uPA e PAI-1 são mais elevados em níveis medianos no câncer de mama tipo ductal do que em tumores lobulares, como encontrado por Descotes e colaboradores (2008).

Como um marcador para câncer de mama, uPA pode ser uma importante variável independente para identificar a taxa de recorrência da doença e é mais interessante do que a maior parte dos fatores de prognóstico tradicionais, especialmente no subtipo que inclui linfonodos negativos. Além disso, uPA e PAI-1 desempenham um papel chave na seleção de terapias adequadas para pacientes com câncer de mama. Elevadas concentrações de uPA e PAI-1 em mulheres com câncer de mama linfonodo-negativos poderiam se beneficiar de quimioterapia adjuvante, enquanto aquelas com baixas concentrações de ambas as proteínas podem ser poupadas dos efeitos colaterais e os custos deste tratamento. Alguns achados sugerem que os pacientes com aumento das concentrações de ambos uPA e PAI-1 não conseguem responder à terapia hormonal em um estágio avançado da doença (DUFFY, 2002).

Além disso, uPA e PAI-1 podem classificar cerca de metade das pacientes com câncer de mama linfonodo-negativos como de baixo risco, nas quais níveis baixos dessas proteínas têm um prognóstico muito bom, e metade como de alto risco, visto que altos níveis delas estão correlacionados com intervalo livre de doença reduzido e baixa sobrevida global (JÄNICKE *et al*, 2001; HARBECK *et al*, 2004).

Assim, suspeita-se que os elevados níveis de plasminogênio encontrados no plasma de mulheres com câncer de mama, nos diferentes estágios da doença, no presente estudo, sejam consequência do remodelamento descontrolado da matriz extracelular durante a doença.

Em contra partida, os níveis de vitronectina (VTNC\_HUMAN) apresentaram-se subexpressos no plasma de paciente com câncer de mama, nos estágios intermediário e avançado, em relação ao plasma de mulheres sem câncer de mama.

A vitronectina é uma glicoproteína multifuncional, produzida primariamente no fígado, e presente no sangue e na matriz extracelular, sendo reconhecida como um substrato adesivo por células que expressam pelo menos um dos diferentes receptores para vitronectina. A interação celular mediada por vitronectina pode induzir espalhamento e migração, e apresenta efeitos no crescimento e diferenciação celular de processos biológicos específicos, tais como cicatrização de

feridas, reabsorção óssea, crescimento de tumores e metástase (FELDING-HABERMANN; CHERESH, 1993).

A vitronectina se liga a glicosaminoglicanos, colágeno e também ao plasminogênio e receptor de uroquinase, estabilizando a conformação ativa do inibidor-1 da ativação do plasminogênio. Pela sua localização na matriz extracelular e a sua ligação ao inibidor-1 da ativação do plasminogênio, a vitronectina pode potencialmente regular a degradação proteolítica da matriz. Além disso, a vitronectina liga-se ao sistema complemento, à heparina e aos complexos trombina-antitrombina, implicando a sua participação na resposta imune e na regulação da formação de coágulos (SCHVARTZ; SEGER; SHALTIEL, 1998).

O pequeno decréscimo nos níveis de vitronectina observados neste trabalho foi também encontrado em um estudo realizado por Kim e colaboradores (2009), no qual os níveis de vitronectina em soro de mulheres com câncer de mama apresentaram-se significativamente reduzidos, quando comparados ao soro de mulheres saudáveis.

Por outro lado, essa redução da vitronectina não é corroborada por achados de Kadowaki e colaboradores (2011), os quais afirmam que níveis aumentados de vitronectina podem ser utilizados como marcadores do câncer de mama precoce.

Em sua pesquisa, Kadowaki e colaboradores usaram amostras de soro de pacientes com carcinoma ductal de mama *in situ* e controles normais, que foram submetidas a uma análise proteômica que consistia em 3 etapas: imunodepleção das proteínas mais abundantes; fracionamento usando cromatografia de fase reversa; e, por fim, separação de proteínas usando eletroforese bidimensional (2D). Diferenças de expressão foram também confirmadas por *Western blotting*, imuno-histoquímica, com amostras teciduais, e ensaios de imunoabsorção enzimática (ELISA).

A análise imunohistoquímica de amostras de câncer de mama revelou alta expressão de vitronectina em paredes de pequenos vasos que envolvem as células cancerosas e na matriz extracelular. Além disso, as concentrações séricas de vitronectina, conforme medido por ELISA, foram significativamente maiores em pacientes com câncer de mama avançado em comparação com os de controles normais (KADOWAKI *et al.*, 2011).

Já foi relatado, em estudos com câncer de ovário, que a metaloproteinase de matriz MMP-2 é secretada por células tumorais e degrada vitronectina, produzindo fragmentos de vitronectina, que são mais potentes do que a forma íntegra da proteína em promover a adesão e migração de células cancerosas (KENNY *et al.*, 2008).

Estranhamente, fragmentos da vitronectina são encontrados em níveis elevados no soro de doentes com carcinoma hepatocelular, mas a expressão do mRNA de vitronectina é paradoxalmente diminuída em tecido do tumor (PARADIS *et al.*, 2005).

Dessa forma, dado o envolvimento contraditório da vitronectina em diferentes tipos de câncer, e mesmo em estudos com câncer de mama, a diminuição dos seus níveis no plasma, observados nesse trabalho, poderia ser reflexo do maior volume de *turnover* desta proteína pelas células tumorais, mas a confirmação dessa hipótese requer investigações.

#### 7.6.2.2. *Análise proteômica da fração que interagiu com a lectina manose-ligante de Dioclea altíssima - DAL (Pico 2)*

Na análise da fração proteica que interagiu com a lectina DAL, os níveis da proteína hemopexina (HEMO\_HUMAN) apresentaram-se superexpressos nas amostras de mulheres com câncer de mama. Níveis esses que foram similares e ligeiramente aumentados no estágio precoce e avançado do câncer. Porém, no estágio intermediário do câncer de mama foi encontrado o maior valor de razão (1,99) para a superexpressão de hemopexina quando comparada à amostra controle.

Hemopexina é uma proteína do plasma que desempenha um papel biológico bem estabelecido no sequestro de heme que é liberado para o plasma a partir da hemoglobina e mioglobina, como resultado de hemólise intravascular ou extravascular, bem como a partir de trauma do músculo esquelético ou doença neuromuscular. Nos últimos anos, uma grande variedade de atividades biológicas adicionais tem sido atribuída à hemopexina como, por exemplo, atividade de hialuronidase, atividade de protease serínica, atividade pró-inflamatória e anti-inflamatória e inibição da adesão celular. Uma das atividades biológicas adicionais atribuídas à hemopexina que chama atenção no presente estudo é a ligação à íons metálicos divalentes, e, portanto, a hemopexina caracterizar-se como um domínio adicional em metaloproteinases da matriz (MMPs) (MAUK; SMITH; MAUK, 2011).

A maioria das MMPs também possuem domínios acessórios adicionais que atuam na modulação da atividade catalítica, reconhecimento do substrato e localização e localização celular. Todas as MMPs humanas, com exceção da MMP-7, -23 e -26A possuem um domínio C-terminal hemopexina (PEX), unido ao domínio catalítico através de um ligante flexível. Tem sido demonstrado que esse domínio hemopexina atua auxiliando o reconhecimento de substratos e facilita a clivagem pelo domínio catalítico (LAUER-FIELDS *et al.*, 2009).

Metaloproteinases de matriz (MMPs) desempenham funções chaves no desenvolvimento embrionário e na morfogênese da glândula mamária e apresentam-se superexpressos no câncer de mama, onde estimulam tumorigênese, a invasão de células cancerosas e metástase. MMPs têm sido investigadas quanto a potenciais alvos para terapia do câncer, mas estudos clínicos com inibidores de MMP de amplo espectro apresentaram resultados decepcionantes, em parte devido à falta de especificidade em relação às MMPs individuais e às fases específicas do desenvolvimento do tumor. A transição epitélio-mesenquimal (EMT) é um processo de desenvolvimento em que as células epiteliais assumem as características de células mesenquimais invasivas, e a ativação de EMT tem sido implicada na progressão do tumor. E, assim, relatos recentes na literatura mostram que as MMPs podem estar implicadas como promotoras e mediadoras do desenvolvimento patogênico do processo de transição epitélio-mesenquimal na mama (RADISKY; RADISKY, 2010).

MMPs tem sido associada com EMT na progressão do câncer através de três mecanismos distintos: (a) níveis elevados das MMPs no microambiente tumoral pode diretamente induzir EMT em células epiteliais, (b) células cancerosas que sofrem EMT podem produzir mais MMPs, facilitando a invasão celular e metástases, e (c) EMT pode gerar células *stroma-like* que impulsionam a progressão do câncer via produção aumentada de MMPs (RADISKY; RADISKY, 2010).

Em células epiteliais mamárias, interações específicas entre domínio hemopexina e integrinas ou outros receptores de superfície celular parecem facilitar a ativação de MMP, o reconhecimento de locais de proteólise pericelular e *turnover* (PICCARD; VAN DEN STEEN; OPDENAKKER, 2007).

A superexpressão de hemopexina já foi relatada em um estudo de Wang e colaboradores (2012), no qual amostras de efusão pleural de pacientes com câncer de pulmão foram analisadas, com razões semelhantes às encontradas no presente trabalho.

A hemopexina possui cinco sítios de glicosilação, onde oligossacarídeos de glucosamina estão ligados (TAKAHASHI; TAKAHASHI; PUTNAM, 1984). Essa proteína também foi identificada na fração que não interagiu com a matriz de DAL imobilizada, porém com valores de razão que variaram entre 0,90 e 1,06, e, portanto, não havendo diferença significativa de expressão entre as amostras identificadas no Pico 1. Portanto, a superexpressão encontrada nas amostras de câncer de mama no Pico 2 sugere uma elevada glicosilação com resíduos de glucose/manose, durante a doença, em parte das moléculas de hemopexina produzida.

A proteína haptoglobina (HPT\_HUMAN) é uma  $\alpha$ -2-sialoglicoproteína de fase aguda com propriedades imunomodulatórias, que se liga a hemoglobina livre, prevenindo a perda do ferro, e exibe atividade antioxidante. Ela foi identificada na fração proteica que interagiu com DAL de amostras de mulheres com câncer de mama com razão de 1,72 no estágio precoce e intermediário e 2,23 no estágio avançado da doença, quando comparada à amostra controle de mulheres sem câncer de mama.

A haptoglobina apresenta quatro sítios de glicosilação, cada um desses sítios podem levar a uma gama de N-glicanos (sendo conhecidos mais de 100 tipos no soro humano), fazendo com que o número total de glicofomas diferentes possíveis seja muito grande. A composição e proporção de todas essas diferentes glicofomas não são aleatórias, no entanto, são notavelmente constantes ao longo do tempo em cada indivíduo saudável, e também variam pouco entre a maioria dos indivíduos de uma população, sugerindo uma regulação fisiológica e funcional (GORNİK *et al.*, 2009).

Elevados níveis de haptoglobina são encontrados em diferentes tipos de cânceres, entre eles o de mama. Tabassum, Reddy e Mukherjee (2012) analisaram o soro de pacientes com câncer de mama triplo-negativo (TNBC), ou seja, aquele definido como negativo para receptores de hormônio e HER2/*neu* negativo, muitas vezes associado com prognóstico ruim e difícil tratamento, e demonstraram que há um aumento da expressão de haptoglobina em pacientes triplo-negativos acompanhado dos graus e, por essa razão, podendo ser um indicador de prognóstico útil em pacientes com TNBC.

Dessa forma, no presente estudo, os valores crescentes de razão encontrados para haptoglobina em mulheres com câncer de mama nos três estágios corroboram com os achados na literatura, em que nota-se um aumento crescente da expressão dessa proteína concomitante à evolução da doença.

Carlsson e colaboradores (2011) analisaram os soros de 25 pacientes com câncer de mama metastático e 25 controles saudáveis submetidos à cromatografia de afinidade usando galectina-1 imobilizada. As principais glicoproteínas que interagiram com a galectina foram 2- $\alpha$ -macroglobulina, IgM e haptoglobina. Ambas as concentrações de IgM e haptoglobina foram semelhantes no câncer, em comparação com soros controle, mas o percentual ligado a galectina-1 foi menor para IgM e superior para a haptoglobina: cerca de 50% no soro de pacientes com câncer e cerca de 30% em soros saudáveis. Os N-glicanos de cada fração de haptoglobina, que interagiram ou não com a galectina, foram analisados por espectrometria de massas e as diferenças estruturais e mutantes de galectina-1 foram usadas para identificar possíveis locais de interação com a lectina.

As frações de haptoglobina ligadas e não ligadas à galectina também foram analisadas quanto à função da proteína.

Como resultado, haptoglobina de ambas as frações foram semelhantes quanto ao comportamento de formar complexos com a hemoglobina e mediar sua absorção em macrófagos alternativamente ativados. No entanto, depois da fagocitose, havia uma diferença dramática no direcionamento intracelular, a fração não ligada à galectina-1 era direcionada aos lisossomos, enquanto que a fração de haptoglobina que interagiu com a galectina-1 formou grânulos positivos maiores. Com isso, os autores concluem que um tipo específico de glicofoma de haptoglobina, e, possivelmente, outras glicoproteínas, apresenta uma função diferente após absorção nas células (CARLSSON *et al.*, 2011).

Assim como no estudo citado acima, no presente trabalho a haptoglobina também foi identificada na fração que não interagiu com a lectina (DAL), nos três estágios da doença, porém com valores de razão entre 1,06 a 1,14, ou seja, sem expressão diferencial. Assim sendo, esses resultados apoiam fortemente a ideia de que há aumento na glicosilação e expressão diferencial entre algumas glicofomas da haptoglobina durante a tumorigênese da mama.

A proteína C3 do sistema complemento (CO3\_HUMAN), identificada apenas na fração que interagiu com a lectina DAL, apresentou-se superexpressa nos três estágios do câncer analisados, com valores crescentes de razão (de 1,6 a 3,4).

As proteínas do sistema complemento fazem parte do grupo de proteínas de fase aguda (APPs), as quais são abundantes no plasma humano. APPs geralmente constituem uma resposta rápida para processos inespecíficos tais como inflamação/infecção, dano tecidual, cirurgia, infarto do miocárdio ou a presença de tumores (DOWLING *et al.*, 2012).

Rutkowski e colaboradores (2010) relataram que as proteínas do complemento podem promover a carcinogênese, facilitando a desregulação de vias de sinalização mitogênicas; estimular a angiogênese; induzir a resistência à apoptose; sustentar a proliferação celular, invasão e metástase.

A relação entre as APPs e o câncer tem sido bem documentada na literatura com numerosas investigações sobre os níveis alterados de várias APPs e diferentes tipos de câncer. Além disso, pensava-se que as APPs não seriam derivadas do tumor, e assim representariam epifenômenos do câncer ao invés de serem proteínas diretamente relacionadas à doença. Dessa forma, por algum tempo, sugeriu-se que as APPs não eram específicas para determinados tipos de câncer, sendo esperado que seus níveis estivessem elevados em todas as malignidades e em doenças inflamatórias em geral.

No entanto, os progressos recentes na proteômica, com estudos baseados nos perfis proteômicos do plasma de pacientes com câncer e indivíduos saudáveis, indicam que a expressão alterada de APPs pode ser diferente para diversos tipos, subtipos e, até mesmo, estágios de câncer. Uma variedade de citocinas, quimiocinas e fatores de crescimento são produzidos pelas células tumorais e por outras células presentes no microambiente tumoral. Esse padrão de moléculas ativas difere entre os tipos de tumores e estágios da doença resultando na estimulação diferencial de APPs e que, portanto, podem refletir o tipo e estágio do tumor. E, assim, embora APPs individuais podem não ser biomarcadores satisfatórios, em combinação ou com outros biomarcadores séricos, podem permitir testes mais sensíveis e específicos (DOWLING *et al.*, 2012).

Ainda em 1992, Niculescu e colaboradores analisaram a presença de algumas proteínas do complemento em 17 amostras de câncer de mama e em 6 amostras de tumores benignos da mama utilizando anticorpos policlonais e monoclonais. Em todos os tecidos de amostras de carcinoma, de cada um dos estágios TNM, foram encontrados depósitos de C5b-9 sobre as membranas de células tumorais, grânulos finos em restos celulares e depósitos difusos em áreas de necrose. Já os depósitos de C5b-9 estavam ausentes em todas as amostras de lesões benignas. E depósitos da proteína C3 foram identificados apenas em amostras de carcinoma. Os autores concluíram que a presença desses depósitos de proteínas do complemento é indicativa de intensa ativação do complemento e seus efeitos patológicos subsequentes no câncer de mama.

Recentemente foi demonstrado superexpressão do fator C3 do sistema complemento no plasma de pacientes com neuroblastoma quando comparado a indivíduos saudáveis (KIM *et al.*, 2014).

Para os critérios estabelecidos para expressão diferencial utilizados nesse trabalho, a apolipoproteína B (Q7Z7Q0\_HUMAN) foi encontrada como proteína única para o câncer de mama, em estágio avançado, sendo identificada e quantificada apenas no Pico 2 após cromatografia em coluna de DAL imobilizada.

Apolipoproteína B (apoB) é um componente estrutural essencial de partículas de lipoproteínas ricas em triglicerídeos (TGs) e desempenha um importante papel na manutenção da homeostase do colesterol. Já existem relatos da associação da apoB com processos inflamatórios, infecções virais e desenvolvimento tumoral e metástase (LIU *et al.*, 2013).

Como já comentado em discussões anteriores, vários estudos epidemiológicos têm mostrado uma forte ligação entre obesidade e câncer de mama, pois o metabolismo lipídico parece estar envolvido no desenvolvimento da doença. A síntese de ácidos graxos aparece cedo na oncogênese e aumenta à medida que as células tornam-se malignas, promovendo a transição de lesões pré e de

alto risco para cânceres invasivos. Alguns estudos apontam evidências de uma associação positiva entre triglicerídeos (TGs) e o risco de câncer de mama (LLANOS *et al.*, 2012).

A apoB sérica já tem sido caracterizada como um marcador em potencial para diagnóstico e prognóstico de câncer de ovário, vesícula e mama, este último sendo aplicado apenas à população chinesa (HAN *et al.*, 2008).

Sabe-se que as concentrações de lipídios circulantes em uma pessoa podem ser influenciadas por sua herança genética. Vários polimorfismos de um único nucleotídeo (SNPs) são associados com concentrações lipídicas séricas aberrantes. Embora já se relate associação positiva entre polimorfismos no gene da ApoB e o câncer de mama, os mecanismos exatos dessa associação ainda não estão claros (LIU *et al.*, 2013).

## ***7.7. Análise quali e quantitativa das proteínas plasmáticas totais, sem fracionamento prévio utilizando lectinas vegetais, por espectrometria de massas***

### ***7.7.1. Identificação e quantificação, por MS, das proteínas plasmáticas presentes nas amostras sem fracionamento prévio em colunas de lectinas vegetais imobilizadas***

A fim de inferir possíveis alterações na glicosilação de algumas proteínas discutidas anteriormente, identificadas nas frações que interagiram ou não com as lectinas utilizadas nesse estudo, bem como confrontar as taxas de identificação e quantificação totais sem que houvesse o fracionamento dos *pools*, e dessa forma avaliar a necessidade e importância do uso das lectinas vegetais no delineamento experimental, foi realizada uma análise das proteínas totais presentes no plasma de mulheres com e sem câncer de mama, ou seja, apenas dos *pools* depletados de albumina e IgG, seguindo os mesmos métodos antes descritos para análise das amostras fracionadas com lectinas vegetais.

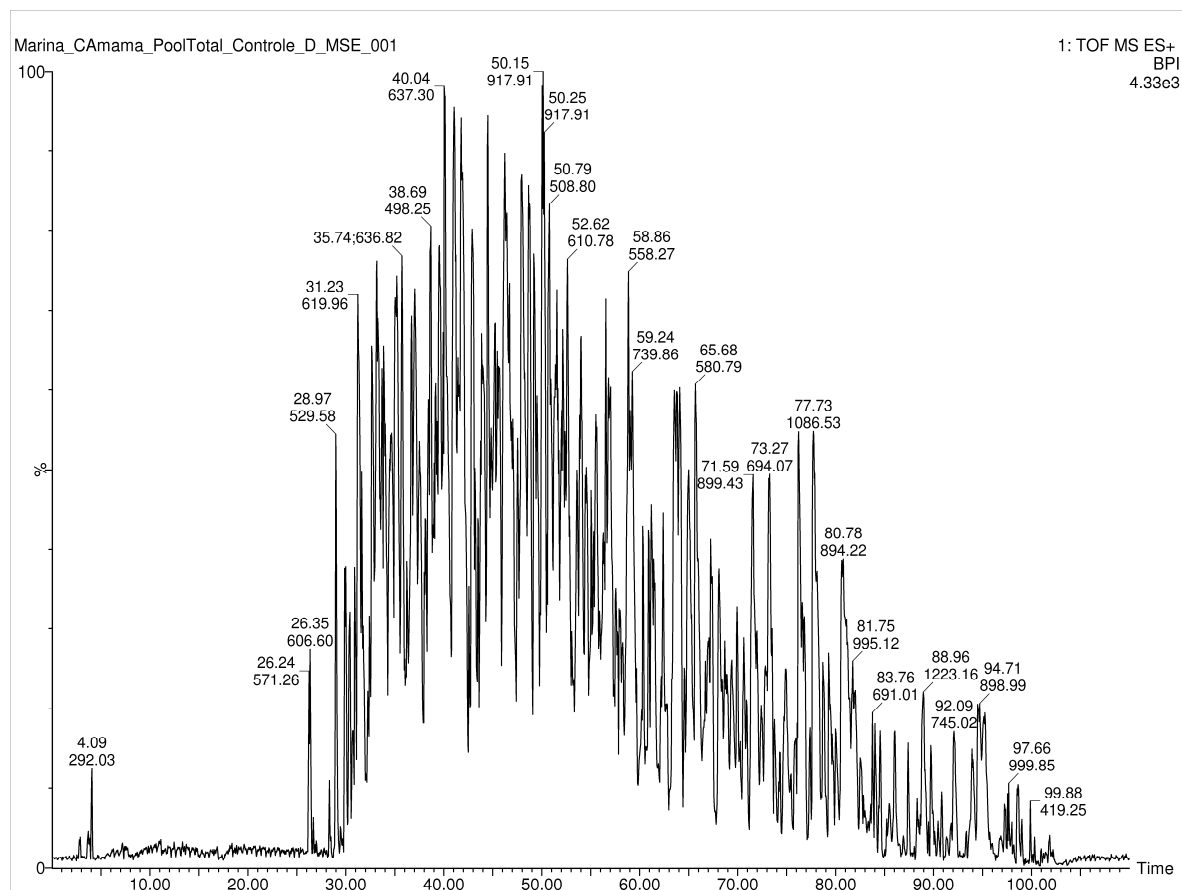
Assim sendo, os *pools* (referentes a cada grupo em estudo) foram submetidos ao processo de digestão enzimática e então, os produtos da digestão, os peptídeos trípticos, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa. Um cromatograma após esse processo cromatográfico para separação dos peptídeos está representado pela Figura 20.

Analisando o perfil cromatográfico dos peptídeos trípticos das amostras totais, ou seja, dos *pools* sem serem submetidos a fracionamento prévio em colunas de lectinas vegetais imobilizadas, observa-se que a eluição dos peptídeos se inicia em torno de 30 minutos após o início do processo cromatográfico e perdura até os 100 minutos, apresentando um intervalo de tempo de eluição peptídica diferente do que foi observado para as amostras fracionadas.

São observados diversos picos de alta intensidade, muito próximos uns aos outros e com sobreposição, demonstrando, portanto, separação inferior em relação aos perfis cromatográficos dos peptídeos trípticos das amostras fracionadas. Essa observação reflete a quantidade de peptídeos nessas amostras totais, sem fracionamento prévio, ser bem superior às fracionadas.

Assim, à medida que os peptídeos iam sendo eluídos da coluna de fase reversa, os mesmos iam sendo ionizados e analisados por espectrometria de massas. Portanto, a partir da análise por LC-MS<sup>E</sup> observou-se a identificação de um total de 98.216 peptídeos (com redundância de peptídeos dentro de um mesmo grupo e entre grupos), incluindo as replicatas.

Figura 20 – Cromatografia de fase reversa, em sistema nanoACQUITY UPLC, dos peptídeos trípticos do *pool* (controle) sem fracionamento prévio em coluna de lectina vegetal imobilizada

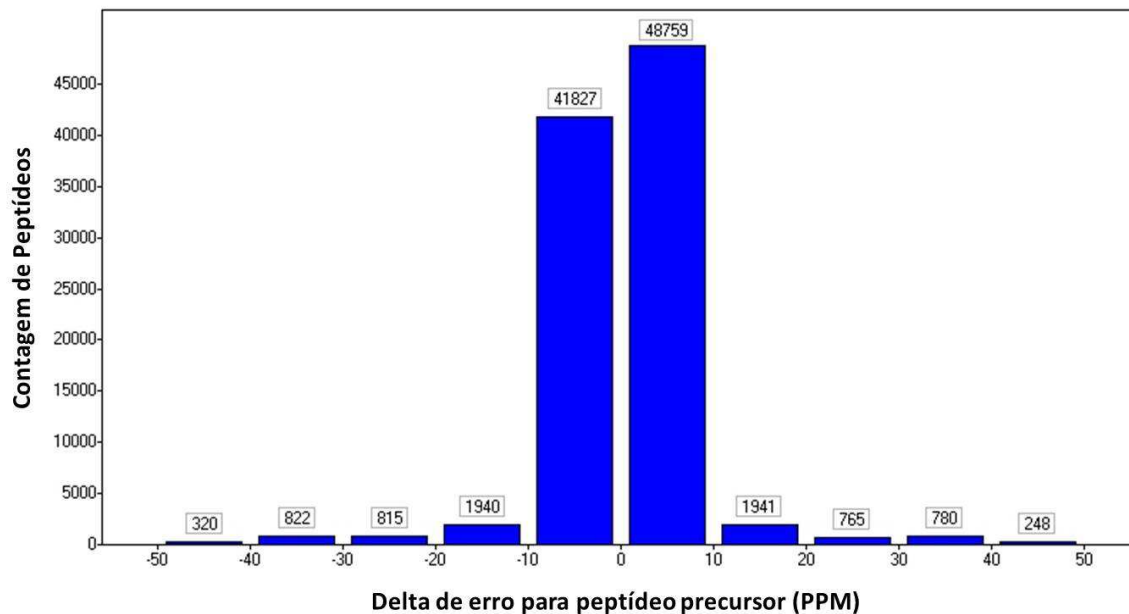


Os peptídeos trípticos resultantes da digestão do *pool* (controle), sem fracionamento prévio em coluna e lectina vegetal imobilizada, foram submetidos a um processo cromatográfico utilizando o sistema nanoACQUITY UPLC (Waters) equipado com uma coluna de fase reversa C18 HSS T3 com dimensões 1,8  $\mu$ m, 75  $\mu$ m x 20 mm, por 110 min utilizando um gradiente de 0% a 40% durante 90 min, 40% a 85% durante 5 min, em seguida, a coluna foi reequilibrada durante 15 min a 35 °C. O fluxo foi de 0,35 mL/min e as fases móveis A e B continham 0,1% de ácido fórmico em água e ácido fórmico a 0,1% em acetonitrila, respectivamente.

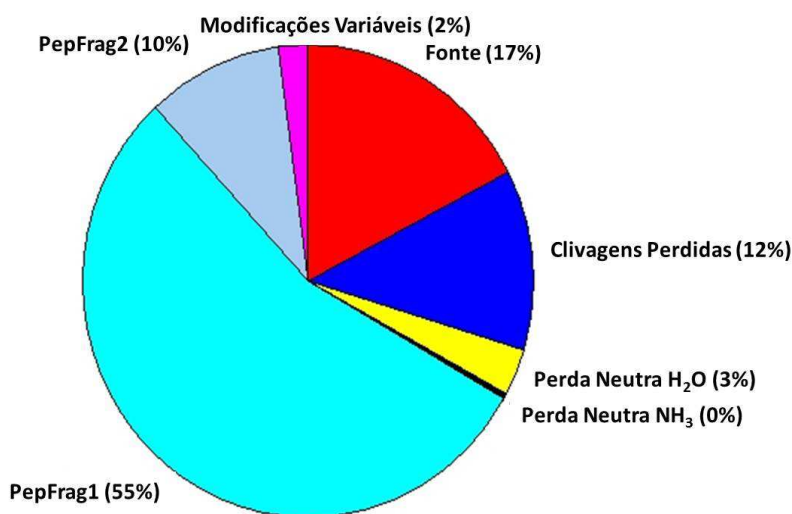
Esses peptídeos identificados apresentaram uma distribuição normal de erro de massa para precursor ilustrado na Figura 21-A, apresentando um máximo de  $\pm 10$  ppm de erro de massa para cerca de 93% dos peptídeos.

Um total de 17% de peptídeos dupla ou triplamente carregados foram produzidos por fragmentação na fonte e 12% dos peptídeos resultaram de clivagens perdidas (Figura 21-B).

Figura 21 - Análise por nanoUPLC-HDMS<sup>E</sup> de plasma humano total depletado (*pools*)



A - Distribuição de massas de fragmentos e a precisão de massa exata calculada para os íons precursores demonstrando 93% a um máximo de  $\pm 10$  ppm.



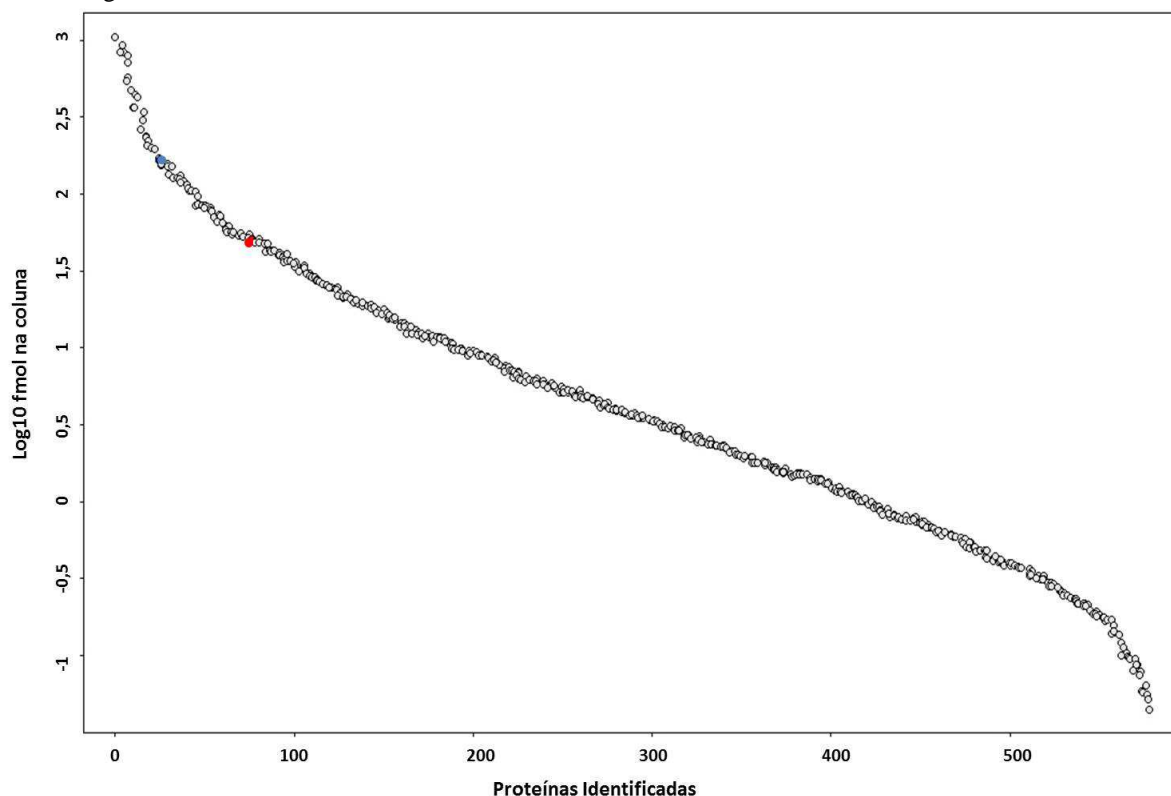
B - Distribuição de peptídeos para determinação da qualidade de fragmentação e digestão.

Os peptídeos então identificados foram associados a 8.534 proteínas (com redundância entre isoformas do mesmo grupo e entre grupos, incluindo as replicatas), obtendo-se uma média de aproximadamente 13 peptídeos por proteína.

Da mesma forma que nas análises anteriores, embora as quantidades de peptídeos (98.216) e proteínas (8.534) sejam elevadas, esses números sofreram redução após os critérios de quantificação, uma vez que diversas isoformas são identificadas, entretanto apenas uma única isoforma para cada proteína é de fato quantificada. Assim sendo, dentre todas as identificações de proteínas acima, 580 foram quantificadas e utilizadas para análise quantitativa.

Dentre as proteínas identificadas e quantificadas através do método Hi-3 *label-free*, todas estão em um intervalo de massas de 0,1 fmol a  $10^3$  fmol. Ou seja, uma diferença de massa em fmol numa ordem de magnitude 4, como é possível observar o intervalo dinâmico das proteínas identificadas, ilustrado na Figura 22.

Figura 22 – Intervalo Dinâmico das proteínas identificadas nos *pools* sem fracionamento prévio em colunas de lectinas vegetais imobilizadas



Após a identificação e quantificação absoluta das proteínas, utilizando a proteína rábica fosforilase B (PHOSP\_B) (esfera vermelha) como padrão interno, a glicoproteína alfa-1-2-ácido (A1AG2\_HUMAN) representada pela esfera azul foi selecionada como *housekeeping* para normalizar os dados e realizar a quantificação relativa *label-free* de proteínas totais presentes nas amostras analisadas.

Após a identificação e quantificação absoluta das proteínas de cada corrida usando peptídeos trípticos de fosforilase-B rábica (UniProt: P00489) como padrão interno, a  $\alpha$ -1-2-ácido glicoproteína (UniProt: P19652) foi selecionada como proteína *housekeeping* para fins de normalizar os níveis de expressão e realizar a quantificação relativa *label-free* das proteínas totais identificadas.

Dentre as proteínas identificadas, aquelas que estavam presentes em diferentes níveis nas amostras de plasma de mulheres com estadiamento I, II, e III do câncer de mama ductal foram comparadas com as amostras de mulheres saudáveis.

A análise de expressão diferencial foi conduzida da mesma forma que em análises anteriores, ou seja, a partir da lista de proteínas gerada pelo PLGS, foram consideradas apenas as proteínas identificadas nas três corridas das replicatas (filtro 3 x 3) e com valores de confiança para  $0,05 \geq p \geq 0,95$ .

E, mais uma vez, somente proteínas com valores de razão Câncer de mama: Controle acima de 1,5 foram consideradas superexpressas (*up-regulated*) e aquelas com valores de razão Câncer de mama: Controle abaixo de 0,66 foram consideradas subexpressas (*down-regulated*). As proteínas que apresentaram valores de razão entre 0,66 e 1,5 foram consideradas sem expressão diferencial (*unchanged*).

Portanto, nos critérios acima estabelecidos, nesta análise das amostras sem fracionamento prévio em cromatografias em lectinas vegetais imobilizadas, foram observadas 90 proteínas, com redundância entre os grupos (Tabelas de 13 a 15).

Em uma análise de expressão diferencial das amostras, apenas depletadas de albumina e IgG, não submetidas ao fracionamento prévio em coluna de lectina vegetal imobilizada, algumas proteínas se apresentaram super ou subexpressas e essas serão discutidas a seguir.

### 7.7.2. Relevância biológica dos dados proteômicos quantitativos obtidos após análise, por espectrometria de massas, das amostras sem fracionamento prévio com lectinas vegetais

Os valores de razão encontrados para o fator C4 A do sistema complemento (CO4A\_HUMAN), para amostras de plasma de mulheres com câncer de mama em comparação com plasma de mulheres saudáveis, foram superiores a 1,5 em todos os estágios do câncer em estudo, demonstrando, portanto, uma superexpressão dessa proteína do complemento durante o processo de tumorigênese.

Como já discutido anteriormente, a contribuição do sistema complemento no processo inflamatório agudo é bem estabelecida, assim como o consumo contínuo das proteínas do complemento em estados inflamatórios crônicos. A descoberta de que proteínas do complemento tais como C3, C4, e C5a podem ajudar o crescimento tumoral através de imunossupressão sugere uma relação insidiosa e previamente não reconhecida entre o sistema complemento e o câncer (MARKIEWSKI *et al.*, 2008).

Durante décadas, o sistema complemento foi reconhecido com um mecanismo efetor do sistema imune que contribuiria para a destruição de células tumorais. Na verdade, muitas estratégias terapêuticas foram sendo propostas seguindo esse raciocínio e, portanto, baseadas na intensificação das respostas mediadas pelo sistema complemento contra tumores. No entanto, estudos recentes têm desafiado esse paradigma, demonstrando um papel de promoção do tumor por meio do sistema complemento. As células cancerosas parecem ser capazes de estabelecer um equilíbrio conveniente entre a ativação e inibição do sistema complemento, aproveitando-se do início da ativação do sistema complemento sem sofrer os seus efeitos deletérios. A ativação do complemento pode apoiar a inflamação crônica, promover um microambiente imunossupressor, induzir a angiogênese e ativar vias de sinalização relacionadas ao câncer. E assim, neste contexto, a inibição da ativação do complemento poderia agora ser uma opção terapêutica para o tratamento do câncer. Porém, esse conceito é relativamente novo e merece mais atenção (PIO; AJONA; LAMBRIS, 2013).

Ainda nesta análise das amostras de plasma sem fracionamento prévio em coluna de lectina vegetal imobilizada, uma apolipoproteína B 100 mutante foi encontrada como proteína única nas amostras de mulheres com câncer de mama estágio precoce, não sendo identificada nas demais amostras e em nenhuma outra análise.

Lipoproteínas de muito baixa densidade (VLDL) são um dos principais produtos de secreção do fígado. Elas servem para o transporte de lipídeos sintetizados endogenamente, principalmente triglicerídeos (mas também colesterol e ésteres de colesterol), para os tecidos

periféricos. VLDL é também precursor da LDL, a lipoproteína plasmática carreadora de colesterol, e ApoB100 é absolutamente necessária para a montagem de VLDL e sua secreção. A quantidade de VLDL-triglicerídeos secretada pelo fígado depende da quantidade de triglicerídeos carregada em cada partícula lipoprotéica, bem como o número de partículas. Cada VLDL tem uma molécula de ApoB100, tornando a disponibilidade dessa lipoproteína um fator determinante do número de partículas de VLDL, e, portanto, dos triglicerídeos que podem ser secretados pelas células hepáticas. Surpreendentemente, o *pool* de ApoB100 no fígado não é regulada pelo seu nível de síntese, o qual é relativamente constante, mas sim pelo seu nível de degradação. Reconhece-se agora que existem múltiplas possibilidades para a célula hepática para interceptar moléculas ApoB100 e a encaminhá-los para distintos processos degradativos (FISHER, E.A.; 2012).

A Hipercolesterolemia Familiar (HF) é uma doença genética do metabolismo das lipoproteínas cujo modo de herança é autossômico codominante e que se caracteriza por níveis muito elevados do colesterol da lipoproteína de baixa densidade (LDL-c), e pela presença de sinais clínicos característicos, como xantomas tendíneos e risco aumentado de doença arterial coronariana. O fenótipo clínico de HF é geralmente decorrente de defeitos genéticos, como por exemplo, defeitos no gene APOB, que codifica a apolipoproteína B-100 (Apo B-100) (PEREIRA *et al.*, 2012).

Dessa forma, a apoB100 mutante identificada no presente trabalho pode ser resultado de alterações genéticas e estar implicada em uma maior síntese de VLDL e, conseqüentemente, LDL.

Recentemente, Coticchia e colaboradores (2015) estudaram a relação entre o colesterol da dieta e o câncer de mama *in vivo* utilizando modelos animais utilizando uma dieta isocalórica inovadora e uso ou não de drogas que bloqueiam absorção de colesterol. Após alcançar os níveis desejados de colesterol sérico entre os grupos, os ratinhos foram então implantados ortotopicamente com células mamárias tumorais e o crescimento tumoral foi monitorado a cada 3 dias até ao sacrifício, quando os tumores foram coletados. Como resultado, foi observado que os tumores de animais alimentados com a dieta rica em colesterol exibiram progressão mais rápida. Embora tenham sido observadas diferenças significativas nos níveis de colesterol sérico entre os grupos de dieta, não foram observadas diferenças nos níveis de colesterol intratumorais. Os tumores dos animais hipercolesterolêmicos exibiram significativamente menos apoptose e, além disso, os tumores de ratinhos

hipercolesterolêmicos tinham áreas de vasos que eram 70% maiores, em média, em comparação com os tumores dos outros grupos. Assim, os pesquisadores concluíram que a hipercolesterolemia induz a angiogênese e acelera o crescimento de tumores da mama *in vivo*.

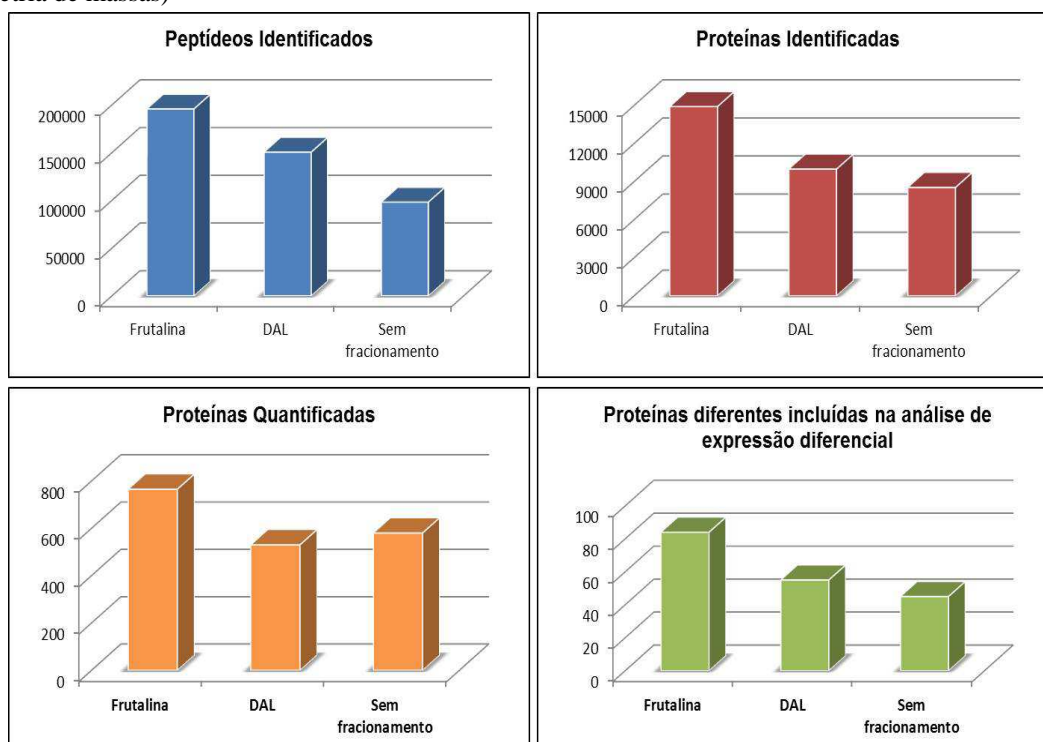
Portanto, a identificação da ApoB100 exclusivamente em mulheres com câncer de mama pode ser um indício da relação dessa proteína com os níveis de colesterol, e que, conseqüentemente, podem influenciar na angiogênese e no crescimento de tumores da mama.

### 7.8. Compilado entre as três análises de expressão diferencial realizadas neste trabalho

Como dito anteriormente, a análise por espectrometria de massas das amostras totais utilizadas nesse estudo, ou seja, os *pools* depletados, sem fracionamento prévio utilizando cromatografias de afinidade com lectinas vegetais imobilizadas, teve como objetivo obter dados dos quais fosse possível inferir informações acerca de alterações no padrão de glicosilação de algumas proteínas, bem como analisar a influência da etapa de fracionamento com lectinas vegetais *a priori* da análise por espectrometria de massas na taxa de identificação e quantificação das proteínas.

Inicialmente, comparando-se o número total de peptídeos e proteínas identificadas nas 3 análises realizadas (Fracionamento com Frutalina, Fracionamento com DAL e sem fracionamento com lectina vegetal *a priori* da análise por espectrometria de massas), observa-se que os números sofrem uma redução perceptível à medida que a complexidade da amostra aumenta (Figura 23). Ou seja, foi identificado um número bem maior de peptídeos e proteínas quando as amostras (*pools*) foram submetidas ao fracionamento, por meio de cromatografias de afinidade em colunas de lectina vegetal imobilizada, previamente à análise por espectrometria de massas. Além disso, esse ganho em identificação foi maior ainda quando a coluna de frutalina foi utilizada para fracionamento das amostras.

Figura 23 – Comparação do número total de peptídeos e proteínas identificadas, de proteínas quantificadas e de proteínas diferentes incluídas nas análises de expressão diferencial, entre as 3 análises realizadas (Fracionamento com Frutalina, Fracionamento com DAL e sem fracionamento com lectina vegetal *a priori* da análise por espectrometria de massas)

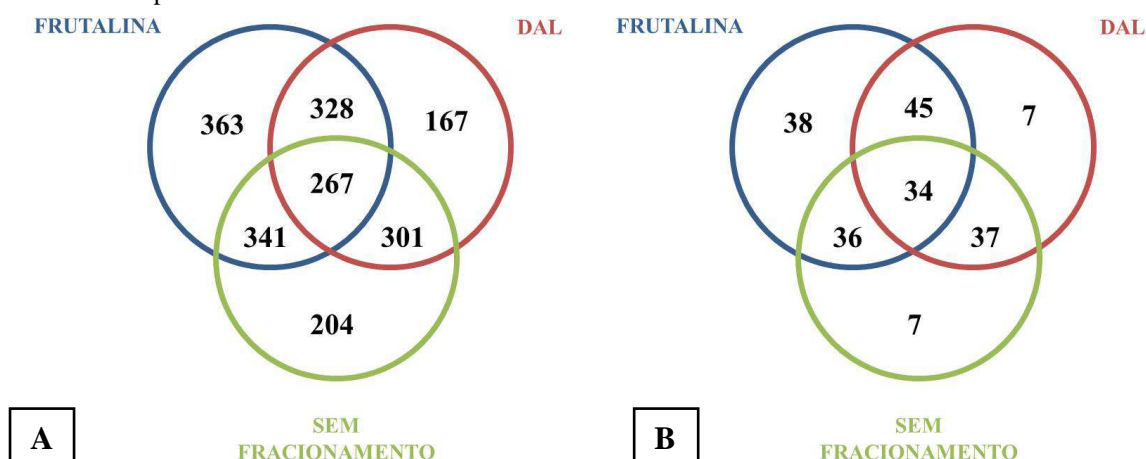


Analisando o Diagrama de Venn construído com base no número de proteínas identificadas e quantificadas (aquelas incluídas no intervalo dinâmico das amostras), ilustrado na figura 24-A, nota-se que o maior número encontrado para proteínas exclusivas foi quando se utilizou o fracionamento em coluna de frutalina imobilizada. Além disso, nas amostras sem fracionamento foi encontrado um número maior de proteínas exclusivas quando comparada às amostras fracionadas em coluna de DAL imobilizada. Porém, vale ressaltar que, embora tenha sido constatada essa diferença, o número de proteínas que ficaram dentro do intervalo de confiança ( $0,05 \geq p \geq 0,95$ ) utilizado para escolha das proteínas que entraram na análise de expressão diferencial foi maior nas amostras fracionadas em coluna de DAL.

Ademais, sob os critérios já descritos em seções anteriores (filtro 3 x 3 e  $0,05 \geq p \geq 0,95$ ), em relação ao número de proteínas incluídas nas análises de expressão diferencial, foi observado um valor de 84 proteínas diferentes após fracionamento com frutalina, 55 proteínas diferentes após fracionamento com DAL e apenas 45 proteínas diferentes foram utilizadas na análise com as amostras sem fracionamento. Entre essas, algumas eram comuns às análises, como é possível observar no Diagrama de Venn ilustrado na figura 24-B.

Portanto, esse número bem reduzido de proteínas com expressão diferencial observado na análise das amostras sem fracionamento prévio demonstra que diferenças sutis de expressão são mais bem evidenciadas quando há fracionamento (ou enriquecimento) da amostra utilizando as lectinas vegetais.

Figura 24 – Diagramas de Venn para proteínas identificadas e quantificadas por LC-MS<sup>E</sup> obtidas sem e com fracionamento prévio em colunas de frutalina e DAL imobilizadas



(A) Diagrama de Venn para valores totais de proteínas identificadas e quantificadas, sem redundância.

(B) Diagrama de Venn para valores de proteínas identificadas e quantificadas dentro dos critérios de filtro 3 x 3 e  $0,05 \geq p \geq 0,95$ .

A natureza complexa do soro e do plasma humano e a presença de um número modesto de proteínas em níveis de mg/ml, por exemplo, 0,1 a 401 mg/mL, fazem com que a detecção de marcadores de doenças, de baixa abundância, seja muito desafiadora. Apesar do plasma rico em proteínas ser utilizado como ferramentas de diagnóstico ao longo de décadas, ainda há menos de 1000 proteínas diferentes identificadas, e apenas uma pequena parte destas proteínas apresentam um potencial de diagnóstico conhecido (ECHAN *et al.*, 2005).

Com o objetivo de investigar o glicoproteoma sérico, Yang e Hancock (2004) utilizaram cromatografias em coluna de afinidade contendo diferentes lectinas vegetais: ConA, WGA e jacalina. Os resultados mostraram que o processo de fracionamento foi específico, eficiente e reprodutível. No estudo foi demonstrado, com sucesso, que a coluna de multi-lectinas removeu uma grande porção da fração de albumina sérica e resultou em maior identificação de proteínas, além de um número médio de peptídeos por proteína também maior. Portanto, assim como observado no presente trabalho, Yang e Hancock concluíram que o uso de colunas de afinidade utilizando lectinas vegetais é bastante útil para estender o intervalo dinâmico das proteínas presentes no plasma humano, além de ser útil também para a captura específica de glicoproteínas.

Em um estudo realizado por Selvaraju e Rassi (2013), uma plataforma proteômica foi desenvolvida para capturar glicoproteínas fucosiladas, presentes no soro de pacientes com câncer de mama, utilizando a lectina *Aleuria aurantia* (AAL) e a aglutinina de *Lotus tetragonolobus* (LTA). Notou-se que a depleção/enriquecimento/fracionamento alcançado com o uso das cromatografias em lectinas vegetais imobilizadas permitiu a detecção e identificação de diferentes proteínas de baixa abundância.

Assim sendo, ficou evidente que, para estudos com o plasma sanguíneo humano, a utilização das cromatografias de afinidade utilizando lectinas vegetais *a priori* da análise por espectrometria de massas é fundamental para reduzir a complexidade das amostras e estender o intervalo dinâmico de proteínas, permitindo uma melhor identificação e quantificação de proteínas, inclusive daquelas presentes em concentrações muito baixas ou pouco abundantes.

Bom, a proposta agora é compilar os resultados de relevância biológica extraídos das 3 análises de expressão diferencial realizadas no presente trabalho. Para isso, será discutida, de agora em diante, a identificação das proteínas super ou subexpressas nas diferentes análises (com fracionamento em coluna de Frutalina imobilizada, com fracionamento em coluna de DAL imobilizada e sem fracionamento prévio em coluna de lectina imobilizada). Portanto, a Tabela 16 reúne todas as proteínas discutidas até o momento e seus respectivos valores de razão (Câncer de Mama: Controle) encontrados em todas as análises realizadas neste trabalho.

Vale a pena ratificar que, ao longo de toda discussão realizada até então, somente proteínas com valores de razão Câncer de mama: Controle acima de 1,5 foram consideradas superexpressas (*up-regulated*) e aquelas com valores de razão Câncer de mama: Controle abaixo de 0,66 foram consideradas subexpressas (*down-regulated*). As proteínas que apresentaram valores de razão entre 0,66 e 1,5 foram consideradas sem expressão diferencial (*unchanged*).

No entanto, embora os valores de “corte” acima determinados tenham sido utilizados para guiar a escolha das proteínas que foram discutidas nas seções anteriores, nesta seção será considerada também a tendência de expressão das proteínas, por meio da observação dos valores de razão encontrados nas demais análises em que elas, por acaso, não tenham apresentado valores de razão dentro do estipulado.

A apolipoproteína A2, por exemplo, foi identificada no Pico 1 após cromatografia de afinidade em matriz de Frutalina imobilizada, apresentando-se superexpressa dos 3 estágios do câncer. Mas foi também identificada no Pico 1 após cromatografia de afinidade em matriz de DAL imobilizada e na análise das amostras sem fracionamento, apresentando, nessas últimas análises, uma tendência à superexpressão, porém com valores de razão inferiores a 1,5. Provavelmente, o fracionamento em coluna de Frutalina imobilizada auxiliou no enriquecimento da amostra de tal forma que essa diferença de expressão ficou mais evidenciada.

Já a apolipoproteína C3 foi identificada em ambos os picos após cromatografia em coluna de Frutalina imobilizada, porém apresentando expressão diferencial apenas nas amostras de Pico 2, nos estágios precoce e intermediário do câncer. Sendo também identificada no Pico 1 após cromatografia de afinidade em matriz de DAL imobilizada e na análise sem fracionamento, porém sem expressão diferencial dentro dos valores de razão pré-determinados. Esses resultados corroboram com a hipótese, anteriormente discutida,

TABELA 16 – Lista de proteínas plasmáticas identificadas nas diferentes análises (com fracionamento em coluna de Frutalina imobilizada, com fracionamento em coluna de DAL imobilizada e sem fracionamento prévio em coluna de lectina imobilizada) e seus respectivos valores de razão (Câncer de Mama : Controle) encontrados

PROTEINAS	FTL PICO 1			FTL PICO 2			DAL PICO 1			DAL PICO 2			SEM FRACIONAMENTO		
	Duc I	Duc II	Duc III	Duc I	Duc II	Duc III	Duc I	Duc II	Duc III	Duc I	Duc II	Duc III	Duc I	Duc II	Duc III
Apolipoproteína A2	2,46	1,51	2,27	-	-	-	1,34	1,23	0,91	-	-	-	1,43	1,43	1,34
Apolipoproteína C3	1,51	1,36	1,17*	1,68	1,62	1,39	1,13*	1,38	1,43	-	-	-	1,00*	0,76	0,89*
C3	0,97	0,88	0,90	-	-	-	-	-	-	1,60	2,12	3,42	0,84	0,86	0,87
C4b B	1,13	0,99*	1,04*	0,64	0,59	0,59	-	-	-	-	-	-	-	-	-
C4 A	-	-	-	-	-	-	-	-	-	-	-	-	1,97	1,58	1,72
Clusterina	1,63	1,70	1,49	1,57	1,60	1,42	1,01*	1,32	1,08*	-	-	-	0,84	0,99*	0,84
$\alpha$ -1-2-ácida glicoproteína	1,65	1,42	1,08	-	-	-	1,04*	1,20	1,06*	-	-	-	-	-	-
Haptoglobina	1,13	1,13	1,20	-	-	-	1,06	1,03	1,14	1,72	1,72	2,23	0,91	0,94	1,09
Hemopexina	1,09	1,03*	1,02*	0,93	0,85	0,90	1,00*	1,06	0,91	1,55	1,99	1,54	0,80	0,88	0,90
Paraoxonase arilesterase sérica	1,68	1,40	1,27	-	-	-	1,02*	1,12*	0,93*	-	-	-	0,84	0,89*	0,74
Proteína plasmática de ligação ao retinol	1,55	1,48	1,36	-	-	-	1,25*	1,34	1,20*	-	-	-	0,93*	1,04*	0,98*
Plasminogênio	1,21	1,00*	0,95*	-	-	-	1,58	1,68	1,60	-	-	-	-	-	-
Vitronectina	1,20	0,98*	0,88	0,85	0,84	1,11	0,70	0,59	0,63	-	-	-	0,75	0,85	0,84

\* Valores fora do intervalo de  $0,05 \geq p \geq 0,95$ ; Dentro do critério estabelecido, proteínas com valores de razão Câncer de mama:Controle acima de 1,5 foram consideradas super expressas (*up-regulated*) e aquelas com valores de razão Câncer de mama:Controle abaixo de 0,66 foram consideradas sub expressas (*down-regulated*). As proteínas que apresentaram valores de razão entre 0,66 e 1,5 foram consideradas sem expressão diferencial (*unchanged*). Portanto, valores de razão considerados para **superexpressão** estão realçados em verde e valores de razão considerados para **subexpressão** estão realçados em amarelo.

de que houve alteração no padrão de glicosilação desta proteína, e este fato ainda não foi associado com o câncer de mama até o momento na literatura científica.

A proteína C3 do sistema complemento apresentou-se superexpressa nas amostras de plasma de mulheres com câncer de mama, nos três estágios do câncer, apenas na análise após fracionamento em coluna de DAL imobilizada, com valores crescentes de razão à medida que o estadiamento do câncer aumentava. Embora também tenha sido identificada nas demais análises, os valores de razão encontrados foram próximos de 1,00, portanto, não apresentando tendência a superexpressão.

Essa proteína (C3 do sistema complemento) apresenta três sítios de N-glicosilação ao longo de sua composição aminoacídica e é altamente manosilada, fato esse que justifica a afinidade apenas pela lectina de *Dioclea altíssima* (LIU *et al.*, 2005). Em um estudo realizado por Ahn e colaboradores (2012), no qual foi monitorada, por espectrometria de massas, a afinidade de diferentes glicoproteínas plasmáticas a diferentes lectinas vegetais, observou-se que a proteína C3 do sistema complemento apresentou seus maiores níveis, ou seja, maior contagem de peptídeos, na fração que interagiu com a lectina concanavalina A (Con A) que, assim como DAL, apresenta afinidade por resíduos de manose, quando comparada às frações das demais lectinas utilizadas.

Ademais, existem relatos de que a quimioterapia em pacientes com câncer de mama induz mudança no padrão de glicosilação no fator C3 do sistema complemento (OEHLER *et al.*, 2012).

Reunindo essas observações, é possível sugerir que, durante o câncer de mama, parece haver um aumento na taxa de glicosilação, com resíduos de glucose/manose, da proteína C3 do sistema complemento.

Já a proteína C4 A do sistema complemento foi identificada apenas na análise das amostras sem fracionamento prévio, apresentando-se superexpressa no plasma de mulheres em diferentes estágios do câncer de mama. Dadas às inúmeras alterações genéticas e epigenéticas associadas com a carcinogênese, a transformação neoplásica pode ser acompanhada por um aumento da capacidade das células malignas em ativar o sistema complemento. Isto é apoiado por dados clínicos que demonstram intensa ativação do sistema complemento em pacientes com câncer (PIO, R.; 2006). Assumindo que essa proteína tem cinco sítios de glicosilação, uma possível justificativa para a mesma não ter sido identificada nas demais análises, quando se utilizou o fracionamento das amostras, é que ela não tenha sido eluída das matrizes de lectina vegetal imobilizadas sob as condições de eluição utilizadas nesse estudo.

A clusterina apresentou-se superexpressa, com valores de razão semelhantes, em ambos os picos após cromatografia de afinidade em matriz de Frutalina imobilizada, nos diferentes estágios do câncer. Sendo também identificada no Pico 1 após cromatografia de afinidade em matriz de DAL imobilizada e na análise das amostras sem fracionamento, porém com valores de razão que não tendenciam à expressão diferencial.

Kapron e colaboradores (1997) identificaram e caracterizaram os sítios de glicosilação da clusterina plasmática humana. Foram identificados seis sítios de N-glicosilação e sete diferentes tipos possíveis de estruturas de oligossacacídeos, que incluíam estruturas com ácido siálico e fucose.

Portanto, esses elevados valores de razão, encontrados apenas nos picos de frutalina, apoiam a utilização do fracionamento em coluna de Frutalina, que tornou mais evidente as diferenças de expressão entre os grupos de estudo, além de sugerir o possível aumento na taxa de glicosilação dessa proteína com resíduos de  $\alpha$ -galactose durante o câncer de mama.

A superexpressão da  $\alpha$ -1-2-ácido-glicoproteína foi observada nos estágios precoce e intermediário do câncer da mama no Pico 1 após cromatografia em coluna de frutalina imobilizada. Além disso, ela também foi identificada no Pico 1 após cromatografia em coluna de DAL imobilizada, porém apresentando razões próximas a 1,0. Essa proteína, curiosamente, foi utilizada como *housekeeping*, para normalização dos dados na análise das amostras sem fracionamento prévio, o que corrobora, mais uma vez, com a conclusão de que o fracionamento evidencia diferenças sutis de expressão ao favorecer uma melhor identificação e quantificação das proteínas.

Além disso, um fato intrigante foi a presença dessa proteína ( $\alpha$ -1-2-ácido-glicoproteína) apenas nas frações que não interagiram com as lectinas testadas no presente trabalho, pois esta proteína é altamente glicosilada. Essa observação pode ser explicada por relatos na literatura de que a porção glicídica dessa proteína plasmática está modificada no câncer e durante reações de fase aguda. Observa-se uma diminuição na porção glicídica e um aumento, de 2 a 3 vezes, em oligossacarídeos contendo fucose em alguns tipos de câncer, com exceção da leucemia (BATT, T.; 2000).

Essas alterações na glicosilação também tem uma importância funcional, pois já foi percebido que cadeias de carboidratos purificadas a partir da  $\alpha$ -1-2-ácido-glicoproteína, de diferentes tipos de câncer, suprimem a proliferação mais do que os glicanos normalmente encontrados nessa proteína (SHIYAN; BOVIN, 1997).

A superexpressão da proteína haptoglobina foi observada apenas na fração proteica que interagiu (Pico 2) com a lectina DAL. Essa proteína foi também identificada no Pico 1 de frutalina, no Pico 1 de DAL e nas análises de amostras sem fracionamento, porém com valores de razão próximos de 1,0. Como já comentado em discussões anteriores, a haptoglobina apresenta diferentes sítios de glicosilação, onde cada um desses sítios podem levar a uma gama de N-glicanos, fazendo com que o número total de glicofomas diferentes possíveis seja muito grande. Portanto os resultados de quantificação dessa proteína, nas diferentes análises realizadas nesse trabalho, corroboram com a ideia de que há expressão diferencial entre glicofomas da haptoglobina durante a tumorigênese.

A hemopexina revelou-se superexpressa nos três estágios do câncer apenas no pico 2 de DAL. Essa proteína foi também identificada em todas as outras amostras, tanto fracionadas em coluna de frutalina imobilizada como nas amostras sem fracionamento, nos 3 estágios do câncer de mama, porém apresentando valores de razão que não tendenciam à expressão diferencial. Esses dados comparativos apoiam um aumento aberrante na glicosilação da hemopexina. Selvaraju e Rassi (2013) também encontram hemopexina superexpressa, neste caso, fucosilada, sendo capturada pela aglutinina de *Lotus tetragonulobus* (LTA).

Proteínas aberrantemente glicosiladas podem ser secretadas para a corrente sanguínea a partir das células por meio das vias de secreção convencionais ou simplesmente soltar-se das membranas celulares como resultado de aumento na atividade proteolítica. Assim, glicofomas aberrantes caem na corrente sanguínea e podem refletir estados anormais de células malignas, caracterizando-se como atraentes candidatas a biomarcadores do câncer (AHN *et al.*, 2012).

A enzima sérica paraoxonase/arilesterase apresentou-se superexpressa no pico 1 de frutalina, principalmente no estágio precoce do câncer, sendo também identificada no Pico 1 de DAL e na análise das amostras sem fracionamento prévio, porém com valores de razão que não tendenciam a expressão diferencial. Mais uma vez, o fracionamento em coluna de frutalina imobilizada evidenciou diferenças sutis de expressão.

A proteína de ligação ao retinol foi identificada com ligeira superexpressão no Pico 1 das amostras de mulheres com câncer de mama após cromatografia em coluna de frutalina imobilizada, porém também no Pico 1 de DAL, com valores de razão entre 1,20 e 1,34, o que sugere uma tendência à expressão aumentada nas amostras de plasma de mulheres com câncer de mama. Essa proteína foi também identificada nas amostras sem fracionamento prévio,

porém com razões em torno de 1,0. Assim como visto para outras proteínas anteriormente citadas nessa discussão, o fracionamento em coluna de frutalina imobilizada evidenciou diferenças sutis de expressão dessa proteína.

No entanto, alterações não significativas nos níveis tanto de vitamina A como da proteína de ligação ao retinol foram encontradas entre mulheres com e sem câncer de mama, em diferentes estágios da doença, em um estudo realizado em 2001 por Ray e Husain. Porém, vale ressaltar que a metodologia utilizada foi detecção por ELISA e não foi realizada nenhuma abordagem proteômica baseada em espectrometria de massas.

De forma geral, a concentração plasmática de proteína de ligação ao retinol (RBP) não é alterada na maioria dos cânceres, no entanto, uma diminuição na concentração de RBP é detectada durante a displasia do colo do útero e, notavelmente, no câncer cervical (BATT, T.; 2000). Portanto, até o presente momento, não há relatos na literatura da associação de níveis aumentados da proteína de ligação ao retinol plasmática com o câncer de mama, utilizando uma abordagem proteômica com espectrometria de massas.

O plasminogênio apresentou-se superexpresso nos 3 estágios do câncer de mama no Pico 1 de DAL, sendo também identificado no Pico 1 de Frutalina, porém com valores de razão que denotam ausência de expressão diferencial significativa. A descrição na literatura das estruturas de *N*- e *O*-glicanos de plasminogênio humano afirma um conteúdo glicídico de *N*-acetil-galactosamina, com sialilação com ácido *N*-acetilneuramínico na galactose (MARTI *et al.*, 1988). Isso justifica a proteína ter sido identificada apenas nas frações que não interagiram com as lectinas imobilizadas. Por alguma razão, ainda não muito clara, o sutil aumento de expressão foi percebido apenas na análise com DAL.

A vitronectina foi identificada em todas as análises, em todos os estágios do câncer, com exceção apenas do Pico 2 de DAL. Porém, os valores de razão encontrados apresentaram-se abaixo de 0,66 apenas no Pico 1 de DAL. Analisando as razões obtidas nas diferentes análises parece haver uma tendência à subexpressão da vitronectina no câncer de mama. Porém, em um estudo realizado por Zeng e colaboradores (2010) com glicoproteínas presentes do soro de pacientes com câncer de mama, verificou-se que não há alteração nem de glicosilação nem em seus níveis quando o soro de mulheres com e sem câncer de mama foram comparados.

Como comentado anteriormente, fragmentos de vitronectina, resultado da atividade da metaloproteinase de matriz MMP-2, são mais potentes do que a forma íntegra da proteína em

promover a adesão e migração de células cancerosas (KENNY *et al.*, 2008). Tendo em mente que o tamanho de uma proteína determina o quão rápido ela é eliminada do sangue, proteínas de baixo peso molecular de fluxo livre e peptídeos, ou seja, menores de 30 kDa, são normalmente rapidamente eliminadas pelo rins. Proteínas abundantes de alto peso molecular acima da faixa de filtração dos rins resistirão no soro até que sejam clivadas proteoliticamente em fragmentos menores e, assim, excretadas (YOCUM *et al.*, 2005). Sendo assim, a tendência à subexpressão da vitronectina no câncer de mama, percebida nesse trabalho, pode ser resultado dessa atividade proteolítica durante a doença.

Em resumo, diferentes proteínas quantificadas neste trabalho apresentaram-se superexpressas e com alteração no padrão de glicosilação no plasma de mulheres com câncer de mama, em diferentes estágios.

Li e colaboradores, em 2011, já afirmavam que o câncer é uma doença proteômica, e que, além das diferenças de expressão, modificações pós-traducionais de proteínas, tais como glicosilação, fosforilação e ubiquitinação, desempenham papéis importantes na gênese e no desenvolvimento de câncer.

A glicosilação, por exemplo, é uma modificação importante que afeta a carga, conformação e estabilidade da proteína e desempenha um papel crucial em processos tais como o reconhecimento de células, a própria sinalização célula-célula, desenvolvimento embrionário e a ligação a hormônios e toxinas (FANAYAN; HINCAPIE; HANCOCK, 2012).

As alterações na glicosilação de proteínas são importantes em muitos processos biológicos e já está bem evidente que uma glicosilação proteica aberrante está associada com o desenvolvimento do câncer (MISEK; KIM, 2011).

O câncer tem efeitos profundos sobre a expressão gênica, incluindo a maquinaria de glicosilação de uma célula. Assim, tumores produzem glicoproteínas que transportam oligossacarídeos com estruturas que são marcadamente diferentes a partir da mesma proteína produzida por uma célula normal. Uma única proteína pode ter muitos locais de glicosilação que amplificam os sinais que eles geram em comparação com os seus esqueletos proteicos. Os estudos baseados em biomarcadores glicosilados têm aumentado de forma maciça. Em todas as fases, desde a descoberta através da verificação até implantação em análises clínicas, a glicosilação deve ser considerada uma leitura primária ou uma forma de aumentar a sensibilidade e especificidade das análises proteômicas (DRAKE *et al.*, 2010).

É relatada na literatura uma alta variabilidade biológica de N-glicanos no plasma humano em nível da população mundial. As variações observadas eram maiores do que alterações relatadas em associação com algumas doenças. Assim, Gornik e colaboradores (2009) examinaram a constância temporal do N-glicoma humano antes que houvesse alterações na glicosilação. Amostras de plasma foram retiradas de 12 indivíduos saudáveis e o sangue foi colhido em sete momentos durante 5 dias. N-glicanos liberados das proteínas plasmáticas foram separados por HPLC em 16 grupos e, posteriormente, quantificados. Os resultados mostraram uma variação muito pequena em todos os grupos de glicanos, indicando uma estabilidade temporal muito boa do N-glicoma em um único indivíduo. Dessa forma, a razoável estabilidade do N-glicoma em plasma de indivíduos saudáveis implica que a glicosilação está sob controle genético significativo. E, portanto, as alterações observadas em perfis de glicanos são consequência de influências ambientais e respostas fisiológicas e, assim, apresentam um potencial de diagnóstico significativo.

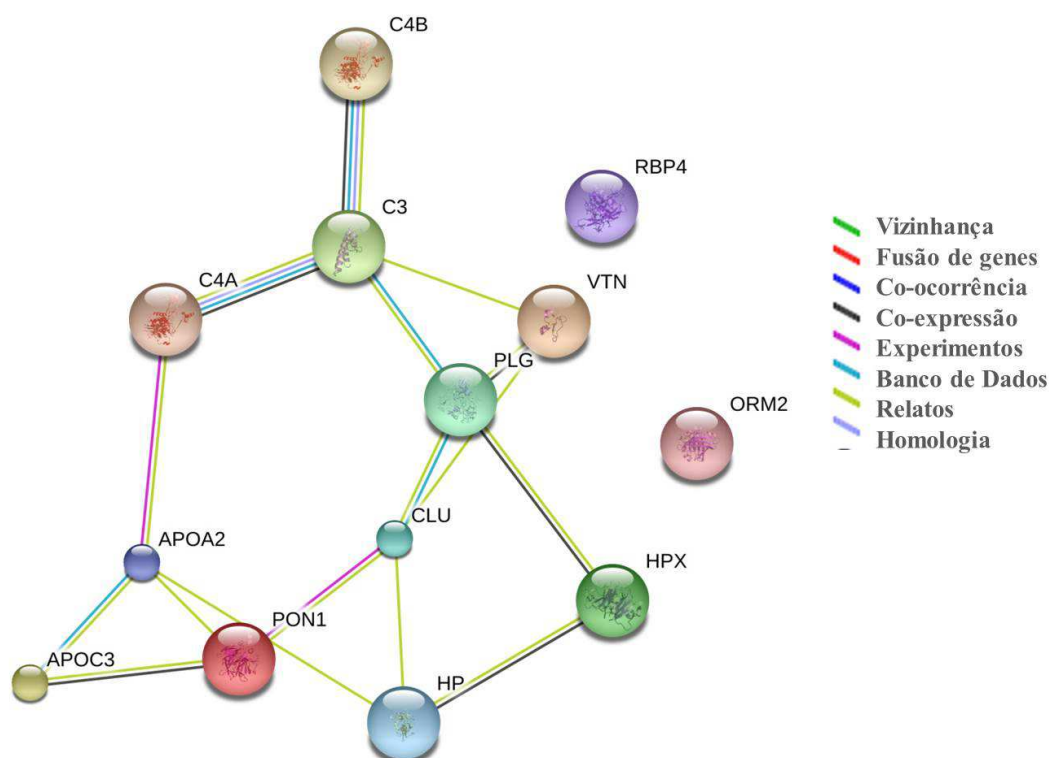
As estruturas de glicanos bem como o perfil de glicoformas diferentes para determinadas proteínas são conhecidas há muito tempo como fatores alterados no câncer. Isto tem estimulado um esforço crescente para utilizar glicoformas particulares como biomarcadores séricos para alguns tipos de câncer, estudo esse facilitado por abordagens proteômicas que associam lectinas vegetais e análise por espectrometria de massa (DRAKE *et al.*, 2010).

### 7.9. Interatoma

Uma vez listadas as proteínas que apresentaram diferenças de expressão e/ou no padrão de glicosilação no plasma de mulheres com câncer de mama, comparado ao de mulheres saudáveis, foi realizada a construção de uma rede interatômica (Figura 25). Para execução dessa atividade, foi utilizada a ferramenta STRING, a qual consiste em bancos de dados que contêm interações previstas entre proteínas.

As interações são estabelecidas utilizando-se diversos parâmetros, que incluem as associações diretas (físico) e indiretas (funcional), provenientes de quatro fontes: o contexto genômico, experimentos, co-expressão e relatos da literatura acerca das proteínas. Atualmente, o STRING possui informações sobre mais de 2 mil organismos e mais de cinco milhões de proteínas (SZKLARCZYK *et al.*, 2015).

Figura 25 - Interatoma elaborado a partir da lista de proteínas que apresentaram diferenças de expressão e/ou no padrão de glicosilação no plasma de mulheres com câncer de mama, comparado ao de mulheres saudáveis a 0.5 de confiança



APOA2 - Apolipoproteína A2; APOC3 - Apolipoproteína C3; C3 – Fator C3 do sistema complemento; C4B – Fator C4b do sistema complement; C4A – Fator C4 do sistema complemento; CLU - Clusterina; ORM2 - Glicoproteína alfa-1-2-ácida; HP - Haptoglobina; HPX - Hemopexina; PON1 - Paraoxonase arilesterase sérica; RBP4 - Proteína plasmática de ligação ao retinol; PLG - Plasminogênio; VTN – Vitronectina.

As diferentes interações e “parcerias” funcionais que ocorrem entre as proteínas estão no cerne do processamento celular e sua caracterização sistemática ajudam a contextualizar essas proteínas no sistema molecular biológico (SZKLARCZYK *et al.*, 2015).

Apolipoproteína A2 e Apolipoproteína C3, superexpressas no câncer de mama no presente estudo e esta última com alteração no padrão de glicosilação, apresentam interação entre si, bem como com a enzima paraoxonase arilesterase sérica. É evidente o papel dessas apolipoproteínas com o transporte de lipídeos no sangue e a relação da ApoA2 com a enzima paraoxonase arilesterase sérica se justifica no fato de que esta enzima está envolvida no remodelamento de HDL, relatado por Gugliucci, A. (2014) e da co-expressão de ApoC3 e paraoxonase arilesterase sérica. Já a interação funcional da Apo2 com a haptoglobina (HP) se baseia apenas em relatos científicos nos quais as duas proteínas são mencionadas.

Alteração no metabolismo lipídico durante o câncer de mama já é bem documentada e foi discutida em seções anteriores deste trabalho. De modo geral, a obesidade está relacionada com vários distúrbios metabólicos, tais como diabetes mellitus tipo 2, doença cardíaca coronariana e hipertensão, bem como está associada com o desenvolvimento do câncer em diversos tecidos incluindo os do cólon, da próstata e da mama, e por essa razão, já ter sido claramente demonstrado que a obesidade é um fator de risco para o desenvolvimento de câncer de mama em mulheres pós-menopáusicas (LORINCZ; SUKUMAR, 2006).

A relação entre as proteínas C3, C4a e C4b é bem evidente, haja vista todas fazem parte do sistema complemento e, portanto, sendo comum a co-expressão e a co-ocorrência das mesmas. Não há nenhuma evidência direta que suporte a afirmação de que o sistema complemento possa eliminar tumores incipientes. No entanto, considerando que o sistema complemento é projetado para o reconhecimento de elementos não próprios, é razoável supor que as mudanças na composição das membranas das células tumorais fazem dessas células alvo de reconhecimento. Consistente com esta suposição, uma série de estudos clínicos tem reportado uma ativação intensa do sistema complemento em pacientes com câncer, fato esse também observado no presente estudo para as proteínas C3 e C4a.

Uma evidência indireta para o papel do sistema complemento na imunovigilância também vem do fato de que as células cancerosas desenvolvem uma variedade de estratégias para evitar danos mediados pelo complemento. Seu mecanismo de fuga mais conhecido é a superexpressão de proteínas reguladoras do sistema complemento. De acordo com a hipótese de imunoedição, esta superexpressão é um indicativo da pressão seletiva desenvolvida pela

ativação do complemento no microambiente tumoral que modela as células cancerosas de forma a contornar os efeitos nocivos do sistema complemento (PIO; AJONA; LAMBRIS, 2013).

Para a clusterina, apenas evidências de relatos na literatura com algumas proteínas são observadas na rede interatômica, nenhum envolvimento funcional direto com as demais proteínas listadas foi identificado.

No interatoma gerado, a haptoglobina está relacionada com a clusterina, com a enzima paraoxonase arilesterase sérica e com a hemopexina apenas de acordo com achados em publicações nos quais as proteínas são co-mencionadas. Nenhuma relação direta é evidenciada, a não ser o fato de que possíveis homólogas da haptoglobina e da hemopexina sejam co-expressas em outras espécies.

Da mesma forma, a relação existente entre a hemopexina e o plasminogênio, e desta última proteína com a vitronectina, se justifica no fato de que possíveis homólogas dessas proteínas sejam co-expressas em outras espécies. Assim como ambas são, muitas vezes, co-mencionadas em trabalhos que associam a degradação da matriz extracelular com a progressão do câncer (FARINA; MACKAY, 2014).

As duas únicas proteínas que não apresentaram relação física, nem funcional, com nenhuma outra foi a  $\alpha$ -1-2-ácido glicoproteína (ORM2) e a proteína plasmática de ligação ao retinol (RBP4). Ambas são lipocalinas, atuando como proteínas transportadoras de moléculas hidrofóbicas pequenas.

De maneira geral, quando se analisa um interatoma, as proteínas identificadas como membros de mais de uma “comunidade proteica” são de particular interesse. Em geral, cada “comunidade proteica” representa um processo celular distinto, por conseguinte, aquela proteína que, por exemplo, é membro de várias “comunidades” pode estar participando de múltiplos processos, e pode ser considerada como sendo a “interface” de distintos, mas adjacentes, processos celulares (JONSSON; BATES, 2006).

Comparando-se a população de proteínas presentes no câncer com a população de proteínas saudáveis, foi observado que as proteínas relacionadas câncer residem nas interfaces da comunidade para uma extensão muito maior do que a suas homólogas não-cancerosas. Em resumo, foi demonstrado que as proteínas humanas envolvidas no câncer exibem uma topologia de rede diferente daquela de proteínas não documentadas como sendo mutadas no câncer. A propriedade mais marcante de proteínas associadas ao câncer é o aumento da

frequência de interações que elas participam, apresentando uma elevada proporção de domínios altamente promíscuos, em termos do número de proteínas diferentes com os quais elas interagem. Esta observação indica uma pressão evolutiva subjacente a qual os genes do câncer, como genes de importância central, são submetidos (JONSSON; BATES, 2006).

Em 2009, Sun e Zhao exploraram as características de redes interatômicas de proteínas codificadas por genes do câncer e, portanto, proteínas do câncer, no interatoma humano. Foi observado que a topologia de rede das proteínas cancerosas era muito diferente da das proteínas codificadas pelos genes essenciais (proteínas essenciais) ou genes de controle (proteínas controle). Em relação às proteínas essenciais ou proteínas controle, proteínas cancerosas apresentaram uma tendência a ter maior interposição e coeficiente de agrupamento menor no interatoma humano.

Em resumo, algumas das proteínas aqui identificadas e associadas com o desenvolvimento do câncer de mama não mostraram tantas relações diretas entre si, como se, durante a doença diferentes redes proteicas fossem afetadas ao mesmo tempo.

E, curiosamente, algumas proteínas apresentaram níveis alterados principalmente nas fases iniciais da doença, como reflexo de uma tentativa emergente de “resolver o problema” e, por outro lado, ao mesmo tempo, ganhar espaço (quando se trata daquelas relacionadas à destruição da matriz extracelular e invasão), parecendo normalizar à medida que a doença avança, como se o corpo se “acostumasse” com a presença do câncer.

Portanto, fica aqui a reflexão, com as palavras de Siddhartha Mukherjee retiradas do livro “O Imperador de todos os males: uma biografia do câncer”:

*“...esta imagem – do câncer como nosso desesperado, malévolo e contemporâneo menecma – é tão perturbadora por ser parcialmente verdadeira. Uma célula cancerosa é uma espantosa perversão da célula normal. O câncer é um invasor e colonizador fenomenalmente bem sucedido, em parte, porque explora as mesmas características que nos tornam bem sucedidos como espécie ou organismo.”*

## 8. CONCLUSÃO

Conclui-se que, para estudos com o plasma sanguíneo humano, a utilização das cromatografias de afinidade utilizando lectinas vegetais *a priori* da análise por espectrometria de massas é fundamental para reduzir a complexidade das amostras e estender o intervalo dinâmico de proteínas, permitindo uma melhor identificação e quantificação de proteínas. No presente estudo, e para as condições estabelecidas, a lectina que apresentou melhores resultados de fracionamento foi Frutalina.

Além disso, a partir dos dados obtidos por espectrometria de massas foi gerada uma lista de proteínas diferencialmente expressas durante o desenvolvimento do câncer de mama, algumas delas com mudanças no padrão de glicosilação, tais como Apolipoproteína A2, Apolipoproteína C3, Fatores do sistema complemento (C3, C4b e C4A), Clusterina,  $\alpha$ -1-2-ácido Glicoproteína, Haptoglobina, Hemopexina, Paraoxonase arilesterase sérica, Proteína plasmática de ligação ao retinol, Plasminogênio e Vitronectina.

Finalmente, os dados de expressão diferencial e possíveis alterações de glicosilação aqui apresentados contribuem para o desenvolvimento de um perfil ou painel proteico, alvo para posterior validação, que sugere uma associação com o desenvolvimento e a caracterização do câncer de mama em seus diferentes estágios.

## 10. REFERÊNCIAS BIBLIOGRÁFICAS

- AHN *et al.* A multiplex lectin-channel monitoring method for human serum glycoproteins by quantitative mass spectrometry. **Analyst**, v. 137, p. 691–703, 2012.
- AHMED, E.F. Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. **Journal of Separation Science**, v. 32, p. 771 – 798, 2009.
- ALLEN , H.J.; JOHNSON E.A.Z. A simple procedure for the isolation of L-fucose-binding lectins from *Ulex europaeus* and *Lotus tetragonolobus*. **Carbohydrate Research**, v. 58, p. 253-265, 1977.
- AMBROSI, M.; CAMERON, N.R.; DAVIS, B.J. Lectins: tools for the molecular understanding of the glycode. **Organic & Biomolecular Chemistry**, v. 3, p. 1593-1608, 2005.
- ARNAUD, J.; AUDFRAYZ, A.; IMBERTY, A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. **Chemical Society Reviews**, v. 42, n. 11, p. 4798-7813, 2013.
- BALCI, H. *et al.* Serum lipid hydroperoxide levels and paraoxonase activity in patients with lung, breast, and colorectal cancer. **Journal of Clinical Laboratory Analysis**, v. 26, p. 155–160, 2012.
- BATAL *et al.* Plasma kinetics of apoC-III and apoE in normolipidemic and hypertriglyceridemic subjects. **The Journal of Lipid Research**, v. 41, p. 706–718, 2000.
- BATT, T. Lipocalins and cancer. **Biochimica et Biophysica Acta**, v. 1482, p. 318-326, 2000.
- BEHREND, L.; HENDERSON, G.; ZWACKA, R.M. Reactive oxygen species in oncogenic transformation. **Biochemical Society Transactions**, v. 31, p. 1441–1444, 2003.
- BELLEI, E. *et al.* High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins. **Amino Acids**, v. 40, n. 1, p. 145-56, 2011.
- BERRY, D.C.; LEVI, L.; NOY, N. Holo-retinol-binding protein and its receptor strA6 drive oncogenic transformation. **Cancer Research**, v. 74, n. 21, p. 6341–6351, 2014.
- BINNECK, E. As ômicas: integrando a bioinformação. **Biotecnologia, Ciência e Desenvolvimento**, n. 32, p. 28-37, 2004.
- BLONDER, J.; ISSAQ, H.J.; VEENSTRA, T.D. Proteomic biomarker discovery: It's more than just mass spectrometry. **Electrophoresis**, n. 32, p. 1541–1548, 2011.

BREWER *et al.* The complete amino acid sequence of alanine apolipoprotein (apoCIII), an apolipoprotein from human plasma very low density lipoproteins. **The Journal of Biological Chemistry**, v. 249, p. 4975–4984, 1974.

CARLSON, R.W.; ANDERSON, B.O.; BURSTEIN, H.J. *et al.* Invasive breast cancer: clinical practice guidelines in oncology. **Journal of the National Comprehensive Cancer Network**, v. 5, n. 3, p. 246–312, 2007.

CARLSSON *et al.* Galectin-1-Binding Glycoforms of Haptoglobin with Altered Intracellular Trafficking, and Increase in Metastatic Breast Cancer Patients. **PLoS ONE**, v. 6, n. 10, 2011.

CARIOCA *et al.* Association of oxidative stress biomarkers with adiposity and clinical staging in women with breast cancer. **European Journal of Clinical Nutrition**, v. 69, p. 1256-1261, 2015.

CAMUZCUOGLU, H. *et al.* Serum paraoxonase and arylesterase activities in patients with epithelial ovarian cancer. **Gynecologic Oncology**, v. 112, p. 481–485, 2009.

CIRQUEIRA *et al.* Subtipos moleculares do câncer de mama. **Femina**, v. 39, n. 10, p. 499-503, 2011.

CHAN, K.C.; LUCAS, D.A.; HISE, D.; SCHAEFER, C.F.; XIAO, Z.; JANINI, G.M.; BUETOW, K.H.; ISSAQ, H.J.; VEENSTRA, T.D.; CONRADS, T.P. Analysis of the human serum proteome. **Clinical Proteomics**, v. 1, p. 101–225, 2004.

CHELIUS, D.; BONDARENKO, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. **Journal of Proteome Research**, v. 1, p. 317–323, 2002.

COTICCHIA *et al.* Hypercholesterolemia induces angiogenesis and causes accelerated growth of breast tumors *in vivo*. **Cancer Research**, v. 73, p. 173-176, 2015.

DANO *et al.* Plasminogen activation and cancer. **The Journal of Thrombosis and Haemostasis**, v. 93, n. 4, p. 676-681, 2005.

CHANDLER, K.; GOLDMAN, R. Glycoprotein Disease Markers and Single Protein-omics. **Molecular and Cell Proteomics**, v. 12, n. 4, p. 836-845, 2013.

DESCOTES *et al.* Plasminogen activator inhibitor type 1 is the most significant of the usual tissue prognostic factors in node-negative breast ductal adenocarcinoma independent of urokinase-type plasminogen activator. **Clinical Breast Cancer**, v. 8, p. 168–177, 2008.

DIAMANDIS, E.P. Mass Spectrometry as a Diagnostic and a Cancer Biomarker Discovery Tool. **Molecular & Cellular Proteomics**, v. 3, p. 367–378, 2004.

DOWLING *et al.* Analysis of acute-phase proteins, AHSR, C3, CLI, HP and SAA, reveals distinctive expression patterns associated with breast, colorectal and lung cancer. **International Journal of Cancer**, v. 131, p. 911–923, 2012.

DRAKE *et al.* Sweetening the pot: adding glycosylation to the biomarker discovery equation. **Clinical Chemistry**, v. 56, p. 223–236, 2010.

DRAKE *et al.* A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma, **Analytical Biochemistry**, n. 408, p. 71–85, 2011.

DRAKE *et al.* Lectin Chromatography/Mass Spectrometry Discovery Workflow Identifies Putative Biomarkers of Aggressive Breast Cancers. **Journal of Proteome Research**. v. 11, p. 2508–2520, 2012.

DUCHÉ, J.C. *et al.* Expression of the Genetic variants of human alpha-1-acid glycoprotein in cancer. **Clinical Biochemistry**, v. 33, p. 197–202, 2000.

DUFFY, M.J. Urokinase plasminogen activator, its inhibitor. PAI-1, as prognostic markers in breast cancer: from pilot to level 1 evidence studies. **Clinical Chemistry**, v. 48, p. 1194–1197, 2002.

ECHAN *et al.* Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. **Proteomics**, v. 5, p. 3292–3303, 2005.

ENG *et al.* A Face in the Crowd: Recognizing Peptides Through Database Search. **Molecular & Cellular Proteomics**, v. 10, n. 11, p. 111, 2011.

EYERS, C.E.; GASKELL, S. **Quantitative Proteomics**, Royal Society of Chemistry, p. 129, 2014.

FANAYAN, S.; HINCAPIE, M.; HANCOCK, W.S. Using lectins to harvest the plasma/serum glycoproteome. **Electrophoresis**, v. 33, p. 1746–1754, 2012.

FARIAS *et al.* Cellular retinol-binding protein I, a regulator of breast epithelial retinoic acid receptor activity, cell differentiation, and tumorigenicity. **Journal of the National Cancer Institute**, v. 97, p. 21–29, 2005.

FARINA, A.R.; MACKAY, A.R. Gelatinase B/MMP-9 in Tumour Pathogenesis and Progression. **Cancers**, v. 6, n. 1, p. 240–296, 2014.

FELDING-HABERMANN, B.; CHERESH, D.A. Vitronectin and its receptors. **Current Opinion in Cell Biology**, v. 5, n. 5, p. 864–868, 1993.

FERREIRA *et al.*, Synthesis and Optimization of Lectin Functionalized Nanoprobes for the Selective Recovery of Glycoproteins from Human Body Fluids. **Analytical Chemistry**, v. 83, p. 7035–7043, 2011.

FERRIGE *et al.* Maximum entropy deconvolution in electrospray mass spectrometry. **Rapid Communications in Mass Spectrometry**, v. 5, p. 374–377, 1991.

FISHER, E.A. The degradation of apolipoprotein B100: Multiple opportunities to regulate VLDL triglyceride production by different proteolytic pathways. **Biochimica et Biophysica Acta**, v. 1821, p. 778–781, 2012.

FRANCIS, F. *et al.* Purification of a new fungal mannose-specific lectin from *Penicillium chrysogenum* and its aphicidal properties. **Fungal Biology**, n. 115, p. 1093-1099, 2011.

GAST, M.C.W.; SCHELLENS, J.H.M; BEIJNEN, J.H. Clinical proteomics in breast cancer: a review. **Breast Cancer Research and Treatment**, n. 116, p. 17–29, 2009.

GORNIK *et al.* Stability of Nglycan profiles in human plasma. **Glycobiology**, v. 19, p. 1547–1553, 2009.

GRANGER, J.; SIDDIQUI, J.; COPELAND, S.; REMICK, D. Albumin depletion of human plasma also removes low abundance proteins including the cytokines. **Proteomics**, v. 5, p. 4713–4718, 2005.

GEROMANOS *et al.*, The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. **Proteomics**, v. 9, p. 1683–1695, 2009.

GEROMANOS, S.J. *et al.* Simulating and validating proteomics data and search results. **Proteomics**, v. 11, p. 1189–1211, 2011.

GUOQING *et al.* Cancer: A proteomic disease. **Science China Life Sciencers**, v. 54 n.5, p. 403–408. 2011.

GUPTA, G.; SUROLIA, A.; SAMPATHKUMAR, S.G. Lectin Microarrays for Glycomic Analysis. **Journal of Integrative Biology**, v. 14, n. 4, p. 419-437, 2010.

GUGLIUCCI, A. Activation of paraoxonase 1 is associated with HDL remodeling ex vivo. **Clinica Chimica Acta**, v. 429, p. 3-8-45, 2014.

GYGI *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. **Nature Biotechnology**, v. 17, p. 994–999, 1999.

HAN *et al.* Associations among lipids, leptin, and leptin receptor gene Gin223Arg polymorphisms and breast cancer in China. **Biological Trace Element Research**, v. 126, p. 38–48, 2008.

HARBECK *et al.* Clinical utility of urokinase-type plasminogen activator and plasminogen activator inhibitor-1 determination in primary breast cancer tissue for individualized therapy concepts. **Clinical Breast Cancer**, v. 3, p. 196–200, 2002.

HARBECK *et al.* Urokinase-type plasminogen activator (uPA) and its inhibitor PAI-I: novel tumor-derived factors with a high prognostic and predictive impact in breast cancer. **Journal of Thrombosis and Haemostasis**, v. 91, p. 450–456, 2004.

HANAHAN, D.; WEINBERG, R.A. Hallmarks of Cancer: The Next Generation. **Cell**, v. 144, p. 646-674, 2011.

HOCHEPIED, T. *et al.*  $\alpha$ 1-Acid glycoprotein: an acute phase protein with inflammatory and immunomodulating properties. **Cytokine & Growth Factor Reviews**, v. 14, p. 25–34, 2003.

INSTITUTO NACIONAL DE CÂNCER. **Estimativa – Incidência de Câncer no Brasil**. Disponível em: <<http://www.inca.gov.br/estimativa/2014/sintese-de-resultados-comentarios.asp>>. Acesso em: 11 de agosto de 2015.

JANKOVIC, M. Glycans as Biomarkers: Status and Perspectives. **Journal of Medical Biochemistry**, v. 30, p. 213–223, 2011.

JÄNICKE *et al.* Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. **The Journal of the National Cancer Institute**, v. 93, p. 913–920, 2001.

JAROS, J.A.J.; GUEST, P.C.; BAHN, S.; MARTINS-DE-SOUZA, D. Affinity Depletion of Plasma and Serum for Mass Spectrometry-Based Proteome Analysis. **Proteomics for Biomarker Discovery Methods in Molecular Biology**, v. 1002, p. 1-11, 2013.

JIAN *et al.* Relative quantitation of glycoisoforms of intact apolipoprotein C3 in human plasma by liquid chromatography-high-resolution mass spectrometry. **Analytical Chemistry**, v. 85, p. 2867–2874, 2013.

JONSSON, P.F.; BATES, P.A. Global topological features of cancer proteins in the human interactome. **Systems biology**, v. 22, n. 18, p. 2291–2297, 2006.

KAPRON, J.T. *et al.* Identification and characterization of glycosylation sites in human serum clusterin. **Protein Science**, v. 6, p. 2120-2133, 1997.

KARNOFSKY, D.A. *et al.* The use of nitrogen mustard in the palliative treatment of carcinoma. **Cancer**, v. 1, p. 634-656, 1948.

KADOWAKI *et al.* Identification of vitronectin as a novel serum marker for early breast cancer detection using a new proteomic approach. **Journal of Cancer Research and Clinical Oncology**, v. 137, p. 1105-1115, 2011.

KENNY *et al.* The initial steps of ovarian cancer cell metastasis are mediated by MMP-2 cleavage of vitronectin and fibronectin. **The Journal of Clinical Investigation**, v. 118, p.1367-1379, 2008.

KIM, E.H.; MISEK, D.E. Glycoproteomics-Based Identification of Cancer Biomarkers. **International Journal of Proteomics**, v. 2011, p. 1-10, 2011.

KIM *et al.* Identification of plasma Complement C3 as a potential biomarker for neuroblastoma using a quantitative proteomic approach. **Journal of Proteomics**, v. 96, p. 1-12, 2014.

KOBOLDT *et al.* (The Cancer Genome Atlas Network). Comprehensive molecular portraits of human breast tumours. **Nature**, 2012.

KUMAR, R.; KUMAR, A.N.; SRIVASTAVA, A. Breast cancer tumor markers. **Journal of Solid Tumors**, v. 2, n. 1, p. 43-46, 2012.

KUPPUMBATTI *et al.* Cellular retinol-binding protein expression and breast cancer. **Journal of the National Cancer Institute**, v. 92, p. 475–480, 2000.

KUZMANOV, U.; KOSANAM, H.; DIAMANDIS, E.P. The sweet and sour of serological glycoprotein tumor biomarker quantification. **BMC Medicine**, v. 11, n. 31, 2013.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 22, p. 680-685, 1970.

LAM, S.W.; JIMENEZ, C.R.; BOVEN, E. Breast cancer classification by proteomic technologies: Current state of knowledge. **Cancer Treatment Reviews**, v. 40, p. 129–138, 2014.

LAZAR *et al.* Recent advances in the MS analysis of glycoproteins: Theoretical considerations. **Electrophoresis**, n. 32, p. 3-13, 2011.

LAUER-FIELDS *et al.* Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity. **Journal of Biological Chemistry**, v. 284, n. 36, p. 24017–24024, 2009.

LI *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. **Proteomics**, n. 9, p. 1696–1719, 2009.

LI *et al.* Cancer: A proteomic disease. **Science China Life Sciences**, v. 54, p. 403–408, 2011.

LISTGARTEN, J., EMILI, A., Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. **Molecular and Cellular Proteomics**, v. 4, p. 419–434, 2005.

LIU *et al.* Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. **Journal of Proteome Research**, v. 4, p. 2070-2080, 2005.

LIU, B.; BIAN, H.J.; BAO, J.K. Plant lectins: potential antineoplastic drugs from bench to clinic. **Cancer Letters**, v. 287, p. 1–12, 2010.

- LIU, X. *et al.* Associations of polymorphisms of rs693 and rs1042031 in apolipoprotein b gene with risk of breast cancer in Chinese. **Japanese Journal of Clinical Oncology**, v. 43, p. 362–368, 2013.
- LIUMBRUNO *et al.* Blood-related proteomics. **Journal of Proteomics**, v.73, p. 483–507, 2010.
- LLANOS *et al.* Cholesterol, lipoproteins, and breast cancer risk in African American women. **Ethnicity & Disease**, v. 22, p. 281-287, 2012.
- LORINCZ, A.M.; SUKUMAR, S. Molecular links between obesity and breast cancer. **Endocrine-Related Cancer**, v. 13, p. 279-292, 2006.
- MALIK, G. *et al.* Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer. **Clinical Cancer Research**, v. 11, p. 1073–85, 2005.
- MALLICK *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. **Nature Biotechnology**, v. 25, p. 125-131, 2007.
- MARKIEWSKI *et al.* Modulation of the antitumor immune response by complement. **Nature Immunology**, v. 9, p. 1225–35, 2008
- MARTI *et al.* The N- and O-linked carbohydrate chains of human, bovine and porcine plasminogen: species specificity in relation to sialylation and fucosylation patterns. **FEBS**, v. 173, p. 57-63, 1988.
- MAUK, M.R.; SMITH, A.; MAUK, G. An alternative view of the proposed alternative activities of hemopexin. **Protein Science**, v. 20, n. 5, p.791-805, 2011.
- MBEUNKUI, F.; GOSHE, M.B. Investigation of solubilization and digestion methods for microsomal membrane proteome analysis using data-independent LC-MSE. **Proteomics**, v. 11, p. 898–911, 2011.
- MILLIONI, *et al.* High Abundance Proteins Depletion vs Low Abundance Proteins Enrichment: Comparison of Methods to Reduce the Plasma Proteome Complexity. **PLoS ONE**, v. 6, n. 5, p. 1-7, 2011.
- MISEK, D.E.; KIM, E.H. Protein Biomarkers for the Early Detection of Breast Cancer. **International Journal of Proteomics**, v. 2011, p. 1-9, 2011.
- MIYOSHI *et al.* Fucosylation Is a Promising Target for Cancer Diagnosis and Therapy. **Biomolecules**, v. 2, p. 34-45, 2012.
- MONTEIRO-MOREIRA *et al.* Crystallization and preliminary X-ray diffraction studies of frutalin, an  $\alpha$ -D-galactose-specific lectin from *Artocarpus incisa* seeds. **Acta Crystallographica Section F**, v. 71, p. 1282-1285, 2015.

MOREIRA *et al.* Isolation and characterization of *Dioclea altissima* var. megacarpa seed lectin. **Phytochemistry**, v. 46, n. 1, p. 139-144, 1997.

MOREIRA, *et al.* Isolation and partial characterization of a lectin from *Artucarpus incisa* l. seeds. **Phytochemistry**, v. 47, n. 7, p. 1183-1188, 1998.

MORENO *et al.* Identification of a new quaternary association for legume lectins. **Journal of Structural Biology**, v. 161, p. 133–143, 2008.

MU *et al.* Identification of o-glycosylated proteins that are aberrantly excreted in the urine of patients with early stage ovarian cancer. **International Journal of Molecular Sciences**, v. 14, p. 7923–7931, 2013.

NAKAMURA *et al.* Clusterin is a secreted marker for a hypoxia-inducible factor-independent function of the von Hippel-Lindau tumor suppressor protein. **The American Journal of Pathology**, v. 168, p. 574–584, 2006.

NEILSON *et al.* Less label, more free: Approaches in label-free quantitative mass spectrometry. **Proteomics**, v. 11, p. 535–553, 2011.

NICULESCU *et al.* Persistent Complement Activation on Tumor Cells in Breast Cancer. **American Journal of Pathology**, v. 140, n. 5, 1992.

NIU *et al.* Small interfering RNA targeted to secretory clusterin blocks tumor growth, motility, and invasion in breast cancer. **Acta Biochimica et Biophysica Sinica**, v. 44, p. 991–998, 2012.

NOBRE *et al.* The specificity of frutalin lectin using biomembrane models. **Biochimica et Biophysica Acta**, v. 1798, p. 1547–1555, 2010.

OEHLER, *et al.* Chemotherapy of breast cancer patients induces a shift in the glycosylation pattern of plasma complement components C3 and C4. **Immunobiology**, v. 11, n. 217, p. 1150, 2012.

OLIVEIRA *et al.* cDNA Cloning and Functional Expression of the  $\alpha$ -D-Galactose-Binding Lectin Frutalin in *Escherichia coli*. **Molecular Biotechnology**, v. 43, p. 212-220, 2009.

OLIVEIRA *et al.* Cytotoxic Effects of Native and Recombinant Frutalin, a Plant Galactose-Binding Lectin, on HeLa Cervical Cancer Cells. **Journal of Biomedicine and Biotechnology**, v. 2011, p. 1-9, 2011.

ONAT *et al.* Apolipoprotein C-III, a strong discriminant of coronary risk in men and a determinant of the metabolic syndrome in both genders. **Atherosclerosis**, v. 168, p. 81–89, 2003.

ONG *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. **Molecular and Cellular Proteomics**, v. 1, p. 376–386, 2002.

PARADIS *et al.* Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. **Hepatology**, v. 41, p. 40-47, 2005.

PATEL *et al.* A comparison of labeling and label-free mass spectrometry-based proteomics approaches. **Journal of Proteome Research**, v. 8, p. 3752–3759, 2009.

PEREIRA, AC *et al.* I Diretriz Brasileira de Hipercolesterolemia Familiar (HF). **Arquivos Brasileiros de Cardiologia**, v. 99, n.2, p. 1-28, 2012.

PEROU *et al.* Molecular portraits of human breast tumours. **Nature**, v. 406, p. 747–52, 2000.

PETRICOIN III. *et al.*, Use of proteomic patterns in serum to identify ovarian cancer. **Lancet**, v. 359, p. 572–575, 2002.

PEUMANS, W.J.; VAN DAMME, E.J.M. Lectins as plant defense proteins. **Plant Physiology**, n. 109, p. 347-352, 1995.

PICCARD. H.; VAN DEN STEEN, P.E.; OPDENAKKER, G. Hemopexin domains as multifunctional liganding modules in matrix metalloproteinases and other proteins. **Journal of Leukocyte Biology**, v. 81, n. 4, p. 870–892, 2007.

PIETROWSKA *et al.* Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. **Journal of Translational Medicine**, v. 7, n. 60, p. 1-13, 2009.

PIO, R. Control of complement activation by cancer cells and its implications in antibody-mediated cancer immunotherapy. **Imunología**, v. 25, n. 3, p. 173-187, 2006.

PIO, R.; AJONA, D.; LAMBRIS, J.D. Complement inhibition in cancer therapy. **Seminars in Immunology**, v. 25, n. 1, p. 54-64, 2013.

PODZIELINSKI *et al.* Apolipoprotein concentrations are elevated in malignant ovarian cyst fluids suggesting that lipoprotein metabolism is dysregulated in epithelial ovarian cancer. **Cancer Investigation**, v. 31, p. 258–272, 2013.

PODWOJSKI *et al.* Peek a peak: a glance at statistics for quantitative label-free proteomics. **Expert Review of Proteomics**, v. 7, p. 249–261, 2010.

PONTING, C.P.; MARSHALL, J.M.; CEDERHOLM-WILLIAMS, S.A. Plasminogen: a structural review. **Blood Coagulation & Fibrinolysis**, v. 3, p. 605-614, 1992.

RADISKY, E.S.; RADISKY, D.C. Matrix Metalloproteinase-Induced Epithelial-Mesenchymal Transition in Breast Cancer. **Journal of Mammary Gland Biology and Neoplasia**, v. 15, p. 201–212, 2010.

RADULOVIC *et al.*, Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. **Molecular and Cellular Proteomics**, v. 3, p. 984–997, 2004.

RAY, G.; HUSAIN, S.A. Role of lipids, lipoproteins and vitamins in women with breast cancer. **Clinical Biochemistry**, v. 34, p. 71–76, 2001.

REDONDO *et al.* Overexpression of Clusterin in Human Breast Carcinoma. **The American Journal of Pathology**, v. 157, n. 2, p. 393-399, 2000.

RODLAND, K.D. Proteomics and cancer diagnosis: the potential of mass spectrometry. **Clinical Biochemistry**, n. 37, p. 579-583, 2004.

ROSENBERG, M.E.; SILKENSEN, J. Clusterin: physiologic and pathophysiologic considerations. **The International Journal of Biochemistry & Cell Biology**, v. 27, p. 633–645, 1995.

ROSS *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. **Molecular and Cellular Proteomics**, v. 3, p. 1154–1169, 2003.

RUDIGER, H.; GABIUS, H.J. Plant lectins: occurrence, biochemistry, functions and applications. **Glycoconjugate Journal**, n.18, p. 589-613, 2001.

RUTKOWSKI *et al.* Cancer and the complement cascade. **Molecular Cancer Research**, v. 8, p. 1453–1465, 2010.

RYU *et al.* Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. **Cancer Informatics**, v. 6, p. 243–255, 2008.

SAUTER *et al.* Biomarkers associated with breast cancer are associated with obesity. **Cancer Detection and Prevention**, v. 32, p. 149–155, 2008.

SCHAUB *et al.* Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. **Journal of the American College of Surgeons**, v. 208, p. 970–980, 2009.

SCHACHTER, H.; FREEZE, H.H. Glycosylation diseases: Quo vadis?. **Biochimica et Biophysica Acta**, v. 1792, p. 925–930, 2009.

SHIYAN, S.D.; BOVIN, N.V. Carbohydrate composition and immunomodulatory activity of different glycoforms of  $\alpha$ -1-acid glycoprotein. **Glycoconjugate Journal**, v. 14, p. 631-638, 1997.

SCHUMMER, M.; BEATTY, B.; URBAN, N. Breast cancer genomics: normal tissue and cancer markers. **Annals of The New York Academy of Sciences**, v. 1210, p. 78-85, 2010.

SCHVARTZ, I.; SEGER, D.; SHALTIEL, S. Molecules in focus: Vitronectin. **The International Journal of Biochemistry & Cell Biology**, v. 31, p. 539-544, 1999.

SELVARAJU, S.; RASSI, Z.E. Targeting human serum fucose by an integrated liquid-phase multicolumn platform operating in “cascade” to facilitate comparative mass spectrometric analysis of disease-free and breast cancer sera. **Proteomics**, v. 13, p. 1701–1713, 2013.

SIEGEL, R.L.; MILLER, K.D.; JEMAL, A. Cancer Statistics, 2015. **Cancer Journal for Clinicians**, v. 65, p. 5–29, 2015.

SILVA, J.C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. **Analytical Chemistry**, n. ;77, p. 2187–2200, 2005.

SILVA, J.C.; GORENSTEIN, M.V.; LI, G.Z.; VISSERS, J.P.; GEROMANOS, S.J. Absolute quantification of proteins by LCMS<sup>E</sup>: a virtue of parallel MS acquisition. **Molecular and Cellular Proteomics**, n. 5, p. 144–156, 2006.

STEWART *et al.* Effects of glycosylation on the structure and function of the extracellular chaperone clusterin. **Biochemistry**, v. 46, p. 1412–1422, 2007.

SZKLARCZYK *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. **Nucleic Acids Research**, v. 43, p. 447–452, 2015.

TABASSUM, U.; REDDY, O.; MUKHERJEE, G. Elevated Serum Haptoglobin is Associated with Clinical Outcome in Triple-Negative Breast Cancer Patients. **Asian Pacific Journal of Cancer Prevention**, v. 13, p. 4541–4544, 2012.

TAKAHASHI, N.; TAKAHASHI, Y.; PUTNAM, F.W. Structure of human hemopexin: O-glycosyl and N-glycosyl sites and unusual clustering of tryptophan residues. **Proceedings of the National Academy of Sciences of the United States of America (PNAS)**, v. 81, p. 2021–2025, 1984.

TANG, L.; HAN, X. The urokinase plasminogen activator system in breast cancer invasion and metastasis. **Biomedicine & Pharmacotherapy**, v. 62, n. 2, p. 179–182, 2013.

TEIXEIRA *et al.* Spondias purpurea Exudate polysaccharide as affinity matrix for the isolation of a galactose-binding-lectin. **Carbohydrate Polymers**, v. 70, p. 369–377, 2007.

TOUSI *et al.* Multidimensional liquid chromatography platform for profiling alterations of clusterin N-glycosylation in the plasma of patients with renal cell carcinoma. **Journal of Chromatography A**, v. 1256, p. 121–128, 2012.

TURCK *et al.* The association of biomolecular resource facilities proteomics research group 2006 study: relative protein quantitation. **Molecular and Cellular Proteomics**, v. 6, p. 1291–1298, 2007.

VAITH, P.; ASSMANN, G.; UHLENBRUCK, G. Characterization of the oligosaccharide side chain of apolipoprotein C-III from human plasma very low density lipoproteins. **Biochimica et Biophysica Acta**, v. 541, p. 234–240, 1978.

VAN DAMME, E.J.M. *et al.* **Handbook of Plant Lectins: properties and biomedical applications.** John Wiley and Sons, Oxford, p. 452, 1998.

VAN DAMME, E.J.; LANNOO, N.; PEUMANS, W.J. Plant lectins. **Advances in Botanical Research**, v. 48, p. 107–209, 2008.

VAN DAMME *et al.* Novel Concepts About the Role of Lectins in the Plant Cell. **The Molecular Immunology of Complex Carbohydrates-3**, v. 705. Ed. A.M. Wu, New York, Springer, p. 295-324, 2011.

WANG *et al.* Clusterin confers resistance to TNF-alpha induced apoptosis in breast cancer cells through NF- $\kappa$ B activation and Bcl-2 overexpression. **Journal of Chemotherapy**, v. 24, p. 348–357, 2012.

WANG *et al.* Differential proteome profiling of pleural effusions from lung cancer and benign inflammatory disease patients. **Biochimica et Biophysica Acta**, n. 1824, p. 692-700, 2012.

WUHRER, M. Glycomics using mass spectrometry. **Glycoconjugate Journal**, v. 30, n. 1, p. 11-22, 2013.

YANG, Z.; HANCOCK, W.S. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. **Journal of Chromatography A**, n.1053, p. 79–88, 2004.

YOCUM *et al.* Effect of Immunoaffinity Depletion of Human Serum during Proteomic Investigations. **Journal of Proteome Research**, v. 4, p. 1722-1731, 2005.

YODER, B.J.; WILKINSON, E.J.; MASSOLL, N.A. Molecular and morphologic distinctions between infiltrating ductal and lobular carcinoma of the breast. **Breast Journal**, v. 13, n. 2, p. 172–179, 2007.

ZEIDAN, B.A.; TOWNSEND, P.A.; GARBIS, S.D.; COPSON, E.; CUTRESS, R.I. Clinical proteomics and breast cancer. **The Surgeon, Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland**, v. X, p. 1-8, 2015.

ZENG *et al.* The Development of an Integrated Platform to Identify Breast Cancer Glycoproteome Changes in Human Serum. **Journal of Chromatography A**, v. 1217, n. 19, p. 3307–3315, 2010.

## **11. ANEXOS**

TABELA 1 – Lista de proteínas plasmáticas identificadas na fração não retida em matriz de Frutalina-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis

Pico I				
Mulheres com câncer de mama ductal estágio I versus Mulheres sem câncer de mama				
ACESSO	DESCRIÇÃO	ESCORE	RAZÃO	NÍVEL DE EXPRESSÃO
K2C1_HUMAN	<i>Keratin type II cytoskeletal 1</i>	726.51	0.51	DOWN
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	79323.34	0.93	UNCHANGED
ANT3_HUMAN	<i>Antithrombin III</i>	22472.41	0.94	UNCHANGED
CO3_HUMAN	<i>Complement C3</i>	103367.10	0.97	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	13404.46	1.03	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	46364.00	1.03	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	188508.50	1.05	UNCHANGED
A8K5A4_HUMAN	<i>cDNA FLJ76826 highly similar to Homo sapiens ceruloplasmin ferroxidase CP mRNA</i>	21073.75	1.06	UNCHANGED
A2AP_HUMAN	<i>Alpha 2 antiplasmin</i>	2201.38	1.09	UNCHANGED
ZA2G_HUMAN	<i>Zinc alpha 2 glycoprotein</i>	4409.95	1.09	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	13339.87	1.09	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	40213.04	1.11	UNCHANGED
B4E1Z4_HUMAN	<i>Complement factor B</i>	35020.26	1.12	UNCHANGED
HPT_HUMAN	<i>Haptoglobin</i>	88686.21	1.13	UNCHANGED
F5GXS0_HUMAN	<i>C4b B</i>	35537.98	1.13	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	20667.82	1.14	UNCHANGED

A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	12497.08	1.15	UNCHANGED
C1S_HUMAN	<i>Complement C1s subcomponent</i>	1624.61	1.20	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	5679.42	1.20	UNCHANGED
PLMN_HUMAN	<i>Plasminogen</i>	2504.17	1.21	UNCHANGED
E7ETN3_HUMAN	<i>Uncharacterized protein</i>	23794.11	1.22	UNCHANGED
SAMP_HUMAN	<i>Serum amyloid P component</i>	4090.96	1.23	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	10936.37	1.25	UNCHANGED
HEP2_HUMAN	<i>Heparin cofactor 2</i>	1564.95	1.26	UNCHANGED
CO9_HUMAN	<i>Complement component C9</i>	294.21	1.27	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	44612.15	1.28	UNCHANGED
Q53H26_HUMAN	<i>Transferrin variant Fragment</i>	206300.41	1.34	UNCHANGED
THRB_HUMAN	<i>Prothrombin</i>	7760.71	1.35	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	18547.61	1.40	UNCHANGED
IGJ_HUMAN	<i>Immunoglobulin J chain</i>	6534.17	1.45	UNCHANGED
CD5L_HUMAN	<i>CD5 antigen like</i>	1318.69	1.46	UNCHANGED
APOC3_HUMAN	<i>Apolipoprotein C III</i>	2621.36	1.51	UP
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	12106.20	1.54	UP
Q5VY30_HUMAN	<i>Plasma retinol binding protein 1</i>	1483.52	1.55	UP
CPN2_HUMAN	<i>Carboxypeptidase N subunit 2</i>	145.72	1.58	UP
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	33661.36	1.58	UP

CLUS\_HUMAN Clusterin 12211.18 1.63 UP

TABELA 2 – Lista de proteínas plasmáticas identificadas na fração não retida em matriz de Frutalina-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis

Pico I				
Mulheres com câncer de mama ductal estágio II versus Mulheres sem câncer de mama				
ACESSO	DESCRIÇÃO	ESCORE	RAZÃO	NÍVEL DE EXPRESSÃO
CO6_HUMAN	<i>Complement component C6</i>	2184.70	0.78	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	46364.00	0.83	UNCHANGED
CBG_HUMAN	<i>Corticosteroid binding globulin</i>	538.17	0.85	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	79323.34	0.87	UNCHANGED
CO3_HUMAN	<i>Complement C3</i>	103367.10	0.88	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	13404.46	0.90	UNCHANGED
A1AT_HUMAN	<i>Alpha 1 antitrypsin</i>	109062.60	0.91	UNCHANGED
AFAM_HUMAN	<i>Afamin</i>	1777.66	0.92	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	188508.50	1.04	UNCHANGED
A8K5A4_HUMAN	<i>cDNA FLJ76826 highly similar to Homo sapiens ceruloplasmin ferroxidase CP mRNA</i>	21073.75	1.06	UNCHANGED
THRB_HUMAN	<i>Prothrombin</i>	7760.71	1.09	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	18547.61	1.11	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	20667.82	1.11	UNCHANGED
ANT3_HUMAN	<i>Antithrombin III</i>	22472.41	1.11	UNCHANGED

HPT_HUMAN	<i>Haptoglobin</i>	88686.21	1.13	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	10936.37	1.14	UNCHANGED
SAMP_HUMAN	<i>Serum amyloid P component</i>	4090.96	1.17	UNCHANGED
CD5L_HUMAN	<i>CD5 antigen like</i>	1318.69	1.21	UNCHANGED
IGJ_HUMAN	<i>Immunoglobulin J chain</i>	6534.17	1.21	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	44612.15	1.23	UNCHANGED
CO9_HUMAN	<i>Complement component C9</i>	294.21	1.26	UNCHANGED
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	12106.20	1.36	UNCHANGED
APOC3_HUMAN	<i>Apolipoprotein C III</i>	2621.36	1.36	UNCHANGED
PON1_HUMAN	<i>Serum paraoxonase arylesterase I</i>	1968.45	1.40	UNCHANGED
A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	12497.08	1.19	UNCHANGED
Q5VY30_HUMAN	<i>Plasma retinol binding protein 1 182</i>	1483.52	1.48	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	4007.37	1.51	UP
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	33661.36	1.51	UP
CLUS_HUMAN	<i>Clusterin</i>	12211.18	1.70	UP

TABELA 3 – Lista de proteínas plasmáticas identificadas na fração não retida em matriz de Frutalina-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio III (avançado) comparadas com mulheres saudáveis

Pico I				
Mulheres com câncer de mama ductal estágio III versus Mulheres sem câncer de mama				
ACESSO	DESCRIÇÃO	ESCORE	RAZÃO	NIVEL DE EXPRESSÃO
Q53H26_HUMAN	<i>Transferrin variant Fragment</i>	206300.41	0.83	UNCHANGED

AFAM_HUMAN	<i>Afamin</i>	1777.66	0.86	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	5679.42	0.88	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	79323.34	0.89	UNCHANGED
CO3_HUMAN	<i>Complement C3</i>	103367.10	0.90	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	188508.50	0.92	UNCHANGED
A8K5A4_HUMAN	<i>cDNA FLJ76826 highly similar to Homo sapiens ceruloplasmin ferroxidase CP mRNA</i>	21073.75	0.94	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	13404.46	0.95	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	40213.04	0.95	UNCHANGED
A1AT_HUMAN	<i>Alpha 1 antitrypsin</i>	109062.60	1.04	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	20667.82	1.05	UNCHANGED
APOA4_HUMAN	<i>Apolipoprotein A IV</i>	6071.76	1.08	UNCHANGED
E7ETN3_HUMAN	<i>Uncharacterized protein</i>	23794.11	1.09	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	46364.00	1.12	UNCHANGED
C1S_HUMAN	<i>Complement C1s subcomponent</i>	1624.61	1.13	UNCHANGED
CD5L_HUMAN	<i>CD5 antigen like</i>	1318.69	1.14	UNCHANGED
THRB_HUMAN	<i>Prothrombin</i>	7760.71	1.14	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	10936.37	1.14	UNCHANGED
SAMP_HUMAN	<i>Serum amyloid P component</i>	4090.96	1.15	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	44612.15	1.17	UNCHANGED
HEP2_HUMAN	<i>Heparin cofactor 2</i>	1564.95	1.17	UNCHANGED

IGJ_HUMAN	<i>Immunoglobulin J chain</i>	6534.17	1.20	UNCHANGED
HPT_HUMAN	<i>Haptoglobin</i>	88686.21	1.20	UNCHANGED
CO9_HUMAN	<i>Complement component C9</i>	294.21	1.21	UNCHANGED
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	12106.20	1.21	UNCHANGED
PON1_HUMAN	<i>Serum paraoxonase arylesterase 1</i>	1968.45	1.27	UNCHANGED
CPN2_HUMAN	<i>Carboxypeptidase N subunit 2</i>	145.72	1.30	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	18547.61	1.32	UNCHANGED
Q5VY30_HUMAN	<i>Plasma retinol binding protein 1</i>	1483.52	1.36	UNCHANGED
CLUS_HUMAN	<i>Clusterin</i>	12211.18	1.49	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	33661.36	1.55	UP
APOA2_HUMAN	<i>Apolipoprotein A II</i>	4007.37	2.27	UP

TABELA 4 – Lista de proteínas plasmáticas identificadas na fração retida em matriz de Frutalina-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis

<b>Pico II</b>				
<b>Mulheres com câncer de mama ductal estágio I versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>ESCORE</b>	<b>RAZÃO</b>	<b>NÍVEL DE EXPRESSÃO</b>
Q6N092_HUMAN	<i>Putative uncharacterized protein DKFZp686K18196 Fragment</i>	36802.04	0.41	DOWN
Q9NPP6_HUMAN	<i>Immunoglobulin heavy chain variant Fragment</i>	17866.64	0.53	DOWN
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	86719.59	0.60	DOWN
F5GXS0_HUMAN	<i>C4b B</i>	41296.95	0.64	DOWN

---

B2R8I2_HUMAN	<i>cDNA FLJ93914 highly similar to Homo sapiens histidine rich glycoprotein HRG mRNA</i>	8247.17	0.68	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	14800.35	0.73	UNCHANGED
Q6MZQ6_HUMAN	<i>Putative uncharacterized protein DKFZp686G11190</i>	2964.75	0.73	UNCHANGED
B7ZLE5_HUMAN	<i>FN1 protein</i>	14423.90	0.76	UNCHANGED
KV101_HUMAN	<i>Ig kappa chain V I region AG</i>	7683.56	0.80	UNCHANGED
A8K3E4_HUMAN	<i>cDNA FLJ78367 highly similar to Homo sapiens fibrinogen A alpha polypeptide FGA transcriptvaria</i>	9077.80	0.82	UNCHANGED
Q8N355_HUMAN	<i>IGL protein</i>	89642.59	0.82	UNCHANGED
Q8NCL6_HUMAN	<i>cDNA FLJ90170 fis clone MAMMA1000370 highly similar to Ig alpha 1 chain C region</i>	36698.85	0.84	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	1051.50	0.85	UNCHANGED
Q6GMW4_HUMAN	<i>IGL protein</i>	82101.37	0.87	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	45971.04	0.93	UNCHANGED
IGJ_HUMAN	<i>Immunoglobulin J chain</i>	21557.39	1.09	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	1082.37	1.15	UNCHANGED
Q6MZV6_HUMAN	<i>Putative uncharacterized protein DKFZp686L19235</i>	28265.69	1.16	UNCHANGED
Q6ZVX0_HUMAN	<i>cDNA FLJ41981 fis clone SMINT2011888 highly similar to Protein Tro alpha1 H myeloma</i>	21264.71	1.16	UNCHANGED
AMBP_HUMAN	<i>Protein AMBP</i>	31516.27	1.17	UNCHANGED
E9PHK0_HUMAN	<i>Tetranectin</i>	2878.24	1.23	UNCHANGED
B4DPP8_HUMAN	<i>cDNA FLJ53075 highly similar to Kininogen 1</i>	8802.97	1.26	UNCHANGED

APOA1_HUMAN	<i>Apolipoprotein A I</i>	5594.32	1.32	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	1302.46	1.35	UNCHANGED
Q96K68_HUMAN	<i>cDNA FLJ14473 fis clone MAMMA1001080 highly similar to Homo sapiens SNC73 protein SNC73 mRNA</i>	34989.58	1.43	UNCHANGED
CLUS_HUMAN	<i>Clusterin</i>	1846.24	1.57	UP
APOC3_HUMAN	<i>Apolipoprotein C III</i>	4164.40	1.68	UP
Q567P1_HUMAN	<i>IGL protein</i>	85322.29	1.92	UP
Q6GMX0_HUMAN	<i>Uncharacterized protein</i>	130309.10	3.03	UP

TABELA 5 – Lista de proteínas plasmáticas identificadas na fração retida em matriz de Frutalina-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis

Pico II				
Mulheres com câncer de mama ductal estágio II versus Mulheres sem câncer de mama				
ACESSO	DESCRIÇÃO	ESCORE	RAZÃO	NIVEL DE EXPRESSÃO
Q6N092_HUMAN	<i>Putative uncharacterized protein DKFZp686K18196 Fragment</i>	36802.04	0.27	DOWN
Q9NPP6_HUMAN	<i>Immunoglobulin heavy chain variant Fragment</i>	17866.64	0.52	DOWN
B7ZLE5_HUMAN	<i>FN1 protein</i>	14423.90	0.55	DOWN
F5GXS0_HUMAN	<i>C4b B</i>	41296.95	0.59	DOWN
KV116_HUMAN	<i>Ig kappa chain V I region Roy</i>	6105.97	0.63	DOWN
IGHM_HUMAN	<i>Ig mu chain C region</i>	1082.37	0.70	UNCHANGED
B4E1Z4_HUMAN	<i>Complement factor B</i>	5692.44	0.70	UNCHANGED

B2R8I2_HUMAN	<i>cDNA FLJ93914 highly similar to Homo sapiens histidine rich glycoprotein HRG mRNA</i>	8247.17	0.72	UNCHANGED
Q9HCC1_HUMAN	<i>Single chain Fv Fragment</i>	7583.12	0.75	UNCHANGED
KV101_HUMAN	<i>Ig kappa chain V I region AG</i>	7683.56	0.79	UNCHANGED
Q6GMW4_HUMAN	<i>IGL protein</i>	82101.37	0.84	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	1051.50	0.84	UNCHANGED
Q8NCL6_HUMAN	<i>cDNA FLJ90170 fis clone MAMMA1000370 highly similar to Ig alpha 1 chain C region</i>	36698.85	0.84	UNCHANGED
Q8N355_HUMAN	<i>IGL protein</i>	89642.59	0.84	UNCHANGED
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	86719.59	0.84	UNCHANGED
PGRP2_HUMAN	<i>N acetylmuramoyl L alanine amidase</i>	2414.54	0.85	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	45971.04	0.85	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	14800.35	0.85	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	1302.46	0.86	UNCHANGED
Q6P089_HUMAN	<i>IGH protein</i>	16000.74	0.86	UNCHANGED
CO9_HUMAN	<i>Complement component C9</i>	1057.90	0.86	UNCHANGED
Q6MZV6_HUMAN	<i>Putative uncharacterized protein DKFZp686L19235</i>	28265.69	0.87	UNCHANGED
B4DPP8_HUMAN	<i>cDNA FLJ53075 highly similar to Kininogen 1</i>	8802.97	0.91	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	5594.32	0.93	UNCHANGED
IGJ_HUMAN	<i>Immunoglobulin J chain</i>	21557.39	0.94	UNCHANGED
Q6ZVX0_HUMAN	<i>cDNA FLJ41981 fis clone SMINT2011888 highly similar to Protein Tro alpha1 H myeloma</i>	21264.71	1.27	UNCHANGED
A0N5G3_HUMAN	<i>Rheumatoid factor G9 light chain Fragment</i>	1640.68	1.31	UNCHANGED

B2R582_HUMAN	<i>cDNA FLJ92374 highly similar to Homo sapiens C type lectin domain family 3 member B CLEC3B mRNA</i>	2838.22	1.38	UNCHANGED
Q96K68_HUMAN	<i>cDNA FLJ14473 fis clone MAMMA1001080 highly similar to Homo sapiens</i>	34989.58	1.39	UNCHANGED
CLUS_HUMAN	<i>Clusterin</i>	1846.24	1.60	UP
APOC3_HUMAN	<i>Apolipoprotein C III</i>	4164.40	1.62	UP
Q567P1_HUMAN	<i>IGL protein IGL PE 2 SV 1</i>	85322.29	1.84	UP
Q6GMX0_HUMAN	<i>Uncharacterized protein</i>	130309.10	2.56	UP

TABELA 6 – Lista de proteínas plasmáticas identificadas na fração retida em matriz de Frutalina-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio III (avanzado) comparadas com mulheres saudáveis

<b>Pico II</b>				
<b>Mulheres com câncer de mama ductal estágio III versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>ESCORE</b>	<b>RAZÃO</b>	<b>NÍVEL DE EXPRESSÃO</b>
Q9HCC1_HUMAN	<i>Single chain Fv Fragment</i>	7583.12	0.35	DOWN
Q9NPP6_HUMAN	<i>Immunoglobulin heavy chain variant Fragment</i>	17866.64	0.48	DOWN
Q6P089_HUMAN	<i>IGH protein</i>	16000.74	0.49	DOWN
Q6N092_HUMAN	<i>Putative uncharacterized protein DKFZp686K18196 Fragment</i>	36802.04	0.50	DOWN
Q8NCL6_HUMAN	<i>cDNA FLJ90170 fis clone MAMMA1000370 highly similar to Ig alpha 1 chain C region</i>	36698.85	0.58	DOWN
F5GXS0_HUMAN	<i>C4b B</i>	41296.95	0.59	DOWN
KV116_HUMAN	<i>Ig kappa chain V I region Roy</i>	6105.97	0.68	UNCHANGED
B4E1Z4_HUMAN	<i>Complement factor B</i>	5692.44	0.68	UNCHANGED
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	86719.59	0.70	UNCHANGED
B7ZLE5_HUMAN	<i>FN1 protein</i>	14423.90	0.72	UNCHANGED

IGHM_HUMAN	<i>Ig mu chain C region</i>	1082.37	0.73	UNCHANGED
Q6GMW4_HUMAN	<i>IGL protein</i>	82101.37	0.74	UNCHANGED
Q8N355_HUMAN	<i>IGL protein</i>	89642.59	0.77	UNCHANGED
KV101_HUMAN	<i>Ig kappa chain V I region AG</i>	7683.56	0.77	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	1302.46	0.85	UNCHANGED
E9PHK0_HUMAN	<i>Tetranectin</i>	2878.24	0.87	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	45971.04	0.90	UNCHANGED
CO9_HUMAN	<i>Complement component C9</i>	1057.90	0.90	UNCHANGED
Q6MZV6_HUMAN	<i>Putative uncharacterized protein DKFZp686L19235</i>	28265.69	0.91	UNCHANGED
IGJ_HUMAN	<i>Immunoglobulin J chain</i>	21557.39	0.92	UNCHANGED
AMBP_HUMAN	<i>Protein AMBP</i>	31516.27	1.06	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	1051.50	1.11	UNCHANGED
B2R8I2_HUMAN	<i>cDNA FLJ93914 highly similar to Homo sapiens histidine rich glycoprotein HRG mRNA</i>	8247.17	1.11	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	5594.32	1.11	UNCHANGED
Q567P1_HUMAN	<i>IGL protein</i>	85322.29	1.22	UNCHANGED
Q96K68_HUMAN	<i>cDNA FLJ14473 fis clone MAMMA1001080 highly similar to Homo sapiens SNC73 protein SNC73 mRNA</i>	34989.58	1.30	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	14800.35	1.34	UNCHANGED
A8K3E4_HUMAN	<i>cDNA FLJ78367 highly similar to Homo sapiens fibrinogen A alpha polypeptide FGA transcriptvaria</i>	9077.80	1.35	UNCHANGED
APOC3_HUMAN	<i>Apolipoprotein C III</i>	4164.40	1.39	UNCHANGED

Q6MZQ6_HUMAN	<i>Putative uncharacterized protein DKFZp686G11190</i>	2964.75	1.63	UP
Q6GMX0_HUMAN	<i>Uncharacterized protein</i>	130309.10	2.64	UP

TABELA 7 – Lista de proteínas plasmáticas identificadas na fração não retida em matriz de DAL-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis

<b>Pico I</b>				
<b>Mulheres com câncer de mama ductal estágio I versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>ESCORE</b>	<b>RAZÃO</b>	<b>NIVEL DE EXPRESSÃO</b>
B4E1Z4_HUMAN	<i>Complement factor B</i>	19890.07	0.07	DOWN
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	42650.55	0.66	DOWN
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	28074.99	0.68	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	8387.47	0.70	UNCHANGED
B7ZLE5_HUMAN	<i>FN1 protein</i>	7228.38	0.74	UNCHANGED
PEDF_HUMAN	<i>Pigment epithelium derived factor</i>	292.91	0.77	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	7962.47	0.79	UNCHANGED
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	7441.61	0.83	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	252660.80	0.90	UNCHANGED
CERU_HUMAN	<i>Ceruloplasmin</i>	27963.75	0.91	UNCHANGED
A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	18594.76	1.05	UNCHANGED
HPT_HUMAN	<i>Haptoglobin</i>	97704.73	1.06	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	28993.84	1.07	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain FGB PE 1 SV 2</i>	32246.21	1.12	UNCHANGED

TRFE_HUMAN	<i>Serotransferrin</i>	311585.00	1.15	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	15517.56	1.15	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	28535.33	1.17	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	40912.72	1.20	UNCHANGED
THRB_HUMAN	<i>Prothrombin</i>	6530.39	1.21	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	11999.16	1.34	UNCHANGED
Q6N041_HUMAN	<i>Putative uncharacterized protein DKFZp686O16217 Fragment</i>	2825.36	1.36	UNCHANGED
Q53H26_HUMAN	<i>Transferrin variant Fragment</i>	293171.31	1.36	UNCHANGED
PLMN_HUMAN	<i>Plasminogen</i>	2209.08	1.58	UP

TABELA 8 – Lista de proteínas plasmáticas identificadas na fração não retida em matriz de DAL-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis

<b>Pico I</b>				
<b>Mulheres com câncer de mama ductal estágio II versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>ESCORE</b>	<b>RAZÃO</b>	<b>NÍVEL DE EXPRESSÃO</b>
VTNC_HUMAN	<i>Vitronectin</i>	8387.47	0.59	DOWN
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	42650.55	0.60	DOWN
Q6N041_HUMAN	<i>Putative uncharacterized protein DKFZp686O16217 Fragment</i>	2825.36	0.67	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	7962.47	0.75	UNCHANGED
B7ZLE5_HUMAN	<i>FN1 protein</i>	7228.38	0.80	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	28074.99	0.90	UNCHANGED
HPT_HUMAN	<i>Haptoglobin</i>	97704.73	1.03	UNCHANGED

CERU_HUMAN	<i>Ceruloplasmin</i>	27963.75	1.05	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	53209.75	1.06	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	28993.84	1.09	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	40912.72	1.09	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	252660.80	1.14	UNCHANGED
THRB_HUMAN	<i>Prothrombin</i>	6530.39	1.15	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	28535.33	1.15	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	15517.56	1.15	UNCHANGED
A1AG2_HUMAN	<i>Alpha 1 acid glycoprotein 2</i>	9028.49	1.20	UNCHANGED
HEP2_HUMAN	<i>Heparin cofactor 2</i>	649.24	1.21	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	31658.60	1.21	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	11999.16	1.23	UNCHANGED
TRFE_HUMAN	<i>Serotransferrin</i>	311585.00	1.25	UNCHANGED
A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	18594.76	1.26	UNCHANGED
A2AP_HUMAN	<i>Alpha 2 antiplasmin</i>	5520.93	1.27	UNCHANGED
PGRP2_HUMAN	<i>N acetylmuramoyl L alanine amidase</i>	870.24	1.32	UNCHANGED
CLUS_HUMAN	<i>Clusterin</i>	6560.15	1.32	UNCHANGED
Q5VY30_HUMAN	<i>Plasma retinol binding protein 1 182</i>	797.45	1.34	UNCHANGED
ZA2G_HUMAN	<i>Zinc alpha 2 glycoprotein</i>	6252.03	1.34	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	6125.10	1.36	UNCHANGED

CO9_HUMAN	<i>Complement component C9</i>	604.96	1.36	UNCHANGED
APOC3_HUMAN	<i>Apolipoprotein C III</i>	4601.90	1.38	UNCHANGED
APOA4_HUMAN	<i>Apolipoprotein A IV</i>	2646.20	1.38	UNCHANGED
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	7441.61	1.42	UNCHANGED
PLMN_HUMAN	<i>Plasminogen</i>	2209.08	1.68	UP

TABELA 9 – Lista de proteínas plasmáticas identificadas na fração não retida em matriz de DAL-Sepharose (Pico 1) de mulheres com câncer de mama ductal estágio III (avançado) comparadas com mulheres saudáveis

<b>Pico I</b>				
<b>Mulheres com câncer de mama ductal estágio III versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>ESCORE</b>	<b>RAZÃO</b>	<b>NÍVEL DE EXPRESSÃO</b>
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	42650.55	0.59	DOWN
VTNC_HUMAN	<i>Vitronectin</i>	8387.47	0.63	DOWN
B7ZLE5_HUMAN	<i>FN1 protein</i>	7228.38	0.74	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	7962.47	0.79	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	252660.80	0.79	UNCHANGED
AFAM_HUMAN	<i>Afamin</i>	2245.67	0.79	UNCHANGED
CERU_HUMAN	<i>Ceruloplasmin</i>	27963.75	0.85	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	28074.99	0.89	UNCHANGED
B4E1Z4_HUMAN	<i>Complement factor B</i>	19890.07	0.90	UNCHANGED
Q53H26_HUMAN	<i>Transferrin variant Fragment</i>	293171.31	0.91	UNCHANGED

HEMO_HUMAN	<i>Hemopexin</i>	53209.75	0.91	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	11999.16	0.91	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	28993.84	0.93	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	31658.60	1.07	UNCHANGED
THRB_HUMAN	<i>Prothrombin</i>	6530.39	1.09	UNCHANGED
HPT_HUMAN	<i>Haptoglobin</i>	97704.73	1.14	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	28535.33	1.21	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	15517.56	1.22	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	32246.21	1.27	UNCHANGED
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	7441.61	1.40	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	6125.10	1.43	UNCHANGED
APOC3_HUMAN	<i>Apolipoprotein C III</i>	4601.90	1.43	UNCHANGED
PLMN_HUMAN	<i>Plasminogen</i>	2209.08	1.60	UP

TABELA 10 – Lista de proteínas plasmáticas identificadas na fração retida em matriz de DAL-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio I (inicial) comparadas com mulheres saudáveis

<b>Pico II</b>				
<b>Mulheres com câncer de mama ductal estágio I versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>SCORE</b>	<b>RAZÃO</b>	<b>NIVEL DE EXPRESSÃO</b>
HEMO_HUMAN	<i>Hemopexin</i>	1048.35	1.55	UP
CO3_HUMAN	<i>Complement C3</i>	3896.21	1.60	UP

HPT_HUMAN	<i>Haptoglobin</i>	3446.98	1.72	UP
-----------	--------------------	---------	------	----

TABELA 11 – Lista de proteínas plasmáticas identificadas na fração retida em matriz de DAL-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio II (intermediário) comparadas com mulheres saudáveis

<b>Pico II</b>				
<b>Mulheres com câncer de mama ductal estágio II versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>SCORE</b>	<b>RAZÃO</b>	<b>NIVEL DE EXPRESSÃO</b>
PZP_HUMAN	<i>Pregnancy zone protein</i>	4735.05	1.23	UNCHANGED
HPT_HUMAN	<i>Haptoglobin</i>	3446.98	1.72	UP
HEMO_HUMAN	<i>Hemopexin</i>	1048.35	1.99	UP
CO3_HUMAN	<i>Complement C3</i>	3896.21	2.12	UP

TABELA 12 – Lista de proteínas plasmáticas identificadas na fração retida em matriz de DAL-Sepharose (Pico 2) de mulheres com câncer de mama ductal estágio III (avançado) comparadas com mulheres saudáveis

<b>Pico II</b>				
<b>Mulheres com câncer de mama ductal estágio III versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>SCORE</b>	<b>RAZÃO</b>	<b>NIVEL DE EXPRESSÃO</b>
HEMO_HUMAN	<i>Hemopexin</i>	1048.35	1.54	UP
HPT_HUMAN	<i>Haptoglobin</i>	3446.98	2.23	UP
CO3_HUMAN	<i>Complement C3</i>	3896.21	3.42	UP

TABELA 13 – Lista de proteínas plasmáticas identificadas no *pool* de plasma de mulheres com câncer de mama ductal I, estágio precoce, comparadas com mulheres saudáveis sem fracionamento prévio em coluna de lectina vegetal imobilizada

Mulheres com câncer de mama ductal estágio I versus Mulheres sem câncer de mama				
ACESSO	DESCRIÇÃO	ESCORE	RAZÃO	NIVEL DE EXPRESSÃO
Q8NCL6_HUMAN	<i>cDNA FLJ90170 fis clone MAMMA1000370 highly similar to Ig alpha 1 chain C region</i>	23685.91	0.02	DOWN
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	181692.20	0.46	DOWN
A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	16212.86	0.66	DOWN
Q6MZV6_HUMAN	<i>Putative uncharacterized protein DKFZp686L19235</i>	13150.50	0.74	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	7171.00	0.75	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	82695.92	0.76	UNCHANGED
CBG_HUMAN	<i>Corticosteroid binding globulin</i>	506.39	0.77	UNCHANGED
A1AT_HUMAN	<i>Alpha 1 antitrypsin</i>	62695.45	0.77	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	21473.18	0.79	UNCHANGED
A2AP_HUMAN	<i>Alpha 2 antiplasmin</i>	5365.25	0.79	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	27053.98	0.80	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	8265.72	0.80	UNCHANGED
AFAM_HUMAN	<i>Afamin</i>	1240.27	0.81	UNCHANGED
B7ZLE5_HUMAN	<i>FN1 protein</i>	9573.30	0.82	UNCHANGED
ZA2G_HUMAN	<i>Zinc alpha 2 glycoprotein</i>	4052.91	0.82	UNCHANGED
CO3_HUMAN	<i>Complement C3</i>	113929.00	0.84	UNCHANGED
CERU_HUMAN	<i>Ceruloplasmin</i>	20491.83	0.84	UNCHANGED

CLUS_HUMAN	<i>Clusterin</i>	12679.21	0.84	UNCHANGED
PON1_HUMAN	<i>Serum paraoxonase arylesterase 1</i>	12591.48	0.84	UNCHANGED
HEP2_HUMAN	<i>Heparin cofactor 2</i>	589.23	0.86	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	19993.47	0.88	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	29726.03	0.91	UNCHANGED
TRFE_HUMAN	<i>Serotransferrin GN TF PE 1 SV 3</i>	226601.20	0.93	UNCHANGED
FIBG_HUMAN	<i>Fibrinogen gamma chain</i>	45288.43	1.13	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain OS Homo sapiens GN FGB PE 1 SV 2</i>	23058.88	1.15	UNCHANGED
Q86TT1_HUMAN	<i>Full length cDNA clone CS0DD006YL02 of Neuroblastoma of Homo sapiens human</i>	17396.79	1.16	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	22597.56	1.17	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	10855.73	1.22	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	12117.33	1.28	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	6359.78	1.43	UNCHANGED
C4BPA_HUMAN	<i>C4b binding protein alpha chain</i>	1314.16	1.43	UNCHANGED
CO4A_HUMAN	<i>Complement C4 A</i>	28340.88	1.97	UP
Q5FWF9_HUMAN	<i>IGL protein</i>	30582.23	2.20	UP

---

TABELA 14 – Lista de proteínas plasmáticas identificadas no *pool* de plasma de mulheres com câncer de mama ductal II, estágio intermediário, comparadas com mulheres saudáveis sem fracionamento prévio em coluna de lectina vegetal imobilizada

Mulheres com câncer de mama ductal estágio II versus Mulheres sem câncer de mama				
ACESSO	DESCRIÇÃO	ESCORE	RAZÃO	NIVEL DE EXPRESSÃO
Q8NEJ1_HUMAN	<i>Uncharacterized protein</i>	37741.15	0.41	DOWN
Q6MZV6_HUMAN	<i>Putative uncharacterized protein DKFZp686L19235</i>	13150.50	0.73	UNCHANGED
Q8NCL6_HUMAN	<i>cDNA FLJ90170 fis clone MAMMA1000370 highly similar to Ig alpha 1 chain C region</i>	23685.91	0.75	UNCHANGED
APOC3_HUMAN	<i>Apolipoprotein C III</i>	14762.00	0.76	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	8265.72	0.77	UNCHANGED
AFAM_HUMAN	<i>Afamin</i>	1240.27	0.81	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	82695.92	0.82	UNCHANGED
B7ZLE5_HUMAN	<i>FN1 protein FN1 PE 2 SV 1</i>	9573.30	0.82	UNCHANGED
A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	16212.86	0.82	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	7171.00	0.85	UNCHANGED
CO3_HUMAN	<i>Complement C3</i>	113929.00	0.86	UNCHANGED
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin</i>	181692.20	0.86	UNCHANGED
A1AT_HUMAN	<i>Alpha 1 antitrypsin</i>	62695.45	0.87	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	27053.98	0.88	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	19993.47	0.89	UNCHANGED
CERU_HUMAN	<i>Ceruloplasmin</i>	20491.83	0.94	UNCHANGED

HPTR_HUMAN	<i>Haptoglobin related protein</i>	29726.03	0.94	UNCHANGED
TRFE_HUMAN	<i>Serotransferrin</i>	226601.20	1.11	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	10855.73	1.15	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	23058.88	1.15	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	22597.56	1.16	UNCHANGED
PGRP2_HUMAN	<i>N acetylmuramoyl L alanine amidase</i>	1996.47	1.22	UNCHANGED
FIBG_HUMAN	<i>Fibrinogen gamma chain</i>	45288.43	1.22	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	12117.33	1.35	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	6359.78	1.43	UNCHANGED
CO4A_HUMAN	<i>Complement C4 A</i>	28340.88	1.58	UP

TABELA 15 – Lista de proteínas plasmáticas identificadas no *pool* de plasma de mulheres com câncer de mama ductal III, estágio avançado comparadas com mulheres saudáveis sem fracionamento prévio em coluna de lectina vegetal imobilizada

<b>Mulheres com câncer de mama ductal estágio III versus Mulheres sem câncer de mama</b>				
<b>ACESSO</b>	<b>DESCRIÇÃO</b>	<b>ESCORE</b>	<b>RAZÃO</b>	<b>NÍVEL DE EXPRESSÃO</b>
Q6MZV6_HUMAN	<i>Putative uncharacterized protein DKFZp686L19235</i>	13150.50	0.05	DOWN
Q8NCL6_HUMAN	<i>cDNA FLJ90170 fis clone MAMMA1000370 highly similar to Ig alpha 1 chain C region</i>	23685.91	0.21	DOWN
B4E1B2_HUMAN	<i>cDNA FLJ53691 highly similar to Serotransferrin 2 SV 1</i>	181692.20	0.61	DOWN
B7ZLE5_HUMAN	<i>FN1 protein</i>	9573.30	0.73	UNCHANGED
PON1_HUMAN	<i>Serum paraoxonase arylesterase 1</i>	12591.48	0.74	UNCHANGED
IGHM_HUMAN	<i>Ig mu chain C region</i>	10855.73	0.75	UNCHANGED

AFAM_HUMAN	<i>Afamin</i>	1240.27	0.76	UNCHANGED
CFAH_HUMAN	<i>Complement factor H</i>	8265.72	0.76	UNCHANGED
A2AP_HUMAN	<i>Alpha 2 antiplasmin</i>	5365.25	0.79	UNCHANGED
A2MG_HUMAN	<i>Alpha 2 macroglobulin</i>	82695.92	0.83	UNCHANGED
AMBP_HUMAN	<i>Protein AMBP</i>	3553.69	0.83	UNCHANGED
VTNC_HUMAN	<i>Vitronectin</i>	7171.00	0.84	UNCHANGED
CLUS_HUMAN	<i>Clusterin</i>	12679.21	0.84	UNCHANGED
CO3_HUMAN	<i>Complement C3</i>	113929.00	0.87	UNCHANGED
CERU_HUMAN	<i>Ceruloplasmin</i>	20491.83	0.87	UNCHANGED
D6RF35_HUMAN	<i>Vitamin D binding protein</i>	19993.47	0.88	UNCHANGED
A1AG1_HUMAN	<i>Alpha 1 acid glycoprotein 1</i>	16212.86	0.88	UNCHANGED
HEMO_HUMAN	<i>Hemopexin</i>	27053.98	0.90	UNCHANGED
A1BG_HUMAN	<i>Alpha 1B glycoprotein</i>	21473.18	0.90	UNCHANGED
FIBG_HUMAN	<i>Fibrinogen gamma chain</i>	45288.43	0.91	UNCHANGED
FIBB_HUMAN	<i>Fibrinogen beta chain</i>	23058.88	0.91	UNCHANGED
A1AT_HUMAN	<i>Alpha 1 antitrypsin</i>	62695.45	0.94	UNCHANGED
TRFE_HUMAN	<i>Serotransferrin</i>	226601.20	0.95	UNCHANGED
APOA1_HUMAN	<i>Apolipoprotein A I</i>	22597.56	1.08	UNCHANGED
HPTR_HUMAN	<i>Haptoglobin related protein</i>	29726.03	1.09	UNCHANGED
APOA2_HUMAN	<i>Apolipoprotein A II</i>	6359.78	1.34	UNCHANGED
D9IWP9_HUMAN	<i>Beta 2 glycoprotein I Fragment</i>	12117.33	1.54	UP

CO4A_HUMAN	<i>Complement C4 A</i>	28340.88	1.72	UP
Q5FWF9_HUMAN	<i>IGL protein</i>	30582.23	2.64	UP

---