

Marcadores séricos e espectrometria de massa no diagnóstico do câncer

Primeira submissão em 25/01/06
Última submissão em 28/11/06
Aceito para publicação em 29/11/06
Publicado em 20/12/06

Serum markers and mass spectrometry in the diagnosis of cancer

Paulo Costa Carvalho¹, Juliana de Saldanha da Gama Fischer², Wim M. Degraeve³, Maria da Gloria da Costa Carvalho⁴

unitermos	resumo
Espectrometria de massa	Esta revisão abrange as principais técnicas, limitações e utilidades da espectrometria de massa aplicada à análise de fluidos biológicos para buscar biomarcadores com potencialidade de diagnóstico médico. Atualmente esse método é capaz de discernir, em segundos, padrões moleculares diferencialmente expressos entre indivíduos controles e com câncer. Resultados da literatura apontam a espectrometria de massa como metodologia promissora no futuro do diagnóstico.
Proteoma	
Biomarcador	
Câncer	

abstract key words

This manuscript reviews mass spectrometry methods and limitations for analysis of biological fluids in the search for biomarkers that can aid medical diagnosis. Currently, mass spectrometry has the ability to discriminate differentially expressed molecular patterns among cancer patients and control subjects. Results in the literature point mass spectrometry as having a major role in the future of medical diagnosis.

*Mass spectrometry
Proteomics
Biomarker
Cancer*

Biomarcadores

As primeiras referências sobre diagnóstico de enfermidades por marcadores datam de 560 anos a.C., quando Hipócrates propôs que modificações em líquidos orgânicos, entre eles o sangue, estavam associadas a patologias. A partir dessa época, as doenças foram mais bem caracterizadas pela busca de alterações em nossos *elementos constitutivos*. Ao final do século XX e início do XXI, foram iniciadas pesquisas para definir a nossa *identidade molecular normal* e as suas modificações devidas a doenças, visando aumentar o poder do diagnóstico. Cunhou-se o termo biomarcador para

qualquer molécula ou característica biológica que possa ser detectada e medida revelando os processos biológicos normais, patogênicos ou a resposta farmacológica após intervenção terapêutica. Assim, tal termo deve indicar alterações do estado fisiológico normal para o patológico ou de mudanças no ambiente corporal interno ou no meio ambiente. No campo molecular, as pesquisas focalizaram a atenção sobre nossos genes, RNAs, proteínas e moléculas ligadas ao metabolismo. Conseqüentemente, o conjunto dessas moléculas foi denominado, respectivamente, genoma (todo o material genético contido na célula), transcriptoma (todos os transcritos de RNA), proteoma (conjunto de proteínas

1. Mestre em Biologia Celular e Molecular/Bioinformática pela Fundação Oswaldo Cruz (FIOCRUZ); doutorando pelo Programa de Engenharia de Sistemas e Computação da Coordenação dos Programas de Pós-Graduação de Engenharia (COPPE) da Universidade Federal do Rio de Janeiro (UFRJ).

2. Mestra em Medicina pela Faculdade de Medicina da UFRJ; doutoranda pelo Instituto de Química da UFRJ.

3. Doutor em Ciências; chefe do Laboratório de Genômica Funcional e Bioinformática da FIOCRUZ.

4. Doutora em Ciências; chefe do Laboratório do Controle da Expressão Gênica do Instituto de Biofísica Carlos Chagas Filho da UFRJ.

expressas) e metaboloma para relacionar os metabólitos presentes em determinadas situações. Essas definições podem ser usadas em relação a organismo, tecido, fluido corporal ou célula.

No ano de 2001 foi concluída e publicada a primeira versão da seqüência do genoma humano⁽²⁹⁾. Ao mesmo tempo, inúmeros laboratórios iniciaram a busca para identificar as proteínas expressas em organismos, tecido, fluido corporal ou célula humana e atribuir a essas moléculas uma função. A pretensão de conhecer os genes, mRNAs, metabólitos e proteínas da célula em determinadas condições constitui um dos maiores desafios das ciências biológicas.

Sabe-se atualmente que a expressão gênica pode ser transcricional ou pós-transcricionalmente regulada. Durante as últimas décadas a análise quantitativa de transcritos gênicos mostrou não haver proporcionalidade direta entre o nível destes e o da proteína presente na célula em resposta a um estímulo específico ou estado patológico⁽¹⁰⁾. Isso se deve à existência de numerosos mecanismos de controle da expressão gênica que operam durante a tradução^(8, 11, 16, 22). Em outras palavras, os quadros transcricionais qualitativo e quantitativo da célula podem não corresponder ao encontrado para proteínas. Muitas espécies de mRNAs são sintetizadas em resposta a perturbações específicas e possuem vida média curta; outras populações podem estar presentes, porém não estar sendo traduzidas⁽²⁶⁾. Outros problemas de mesma natureza decorrem das modificações pós-transcricionais, experimentadas pelos RNAs (*splicing*), e pós-traducionais, sofridas por muitas proteínas. Na comparação direta entre o genoma e o proteoma a situação é ainda mais complexa, considerando-se que os mecanismos de *splicing* alternativo (método pelo qual íntrons são removidos e éxons são unidos para formar o RNA) aumentam bastante a capacidade de codificação de um determinado gene⁽³⁾. Conseqüentemente, as pesquisas foram direcionadas para conhecer o proteoma celular padrão ou situações em que as funções celulares foram perturbadas ou encontram-se em situações patológicas específicas.

Assim como o conhecimento do genoma deve-se praticamente ao advento da engenharia genética e às metodologias de seqüenciamento de DNA, o do proteoma foi alavancado principalmente graças ao aparecimento da eletroforese bidimensional (2D) de alta resolução e às técnicas de espectrometria de massa (EM), capazes de ionizar uma biomolécula sem degradá-la⁽⁷⁾. A eletroforese 2D⁽²¹⁾ combina uma focalização isoeletrica na primeira dimensão com a eletroforese em gel de poli(acrilamida), contendo dodecil sulfato de sódio (SDS) na segunda. A partir daí, a

técnica foi aperfeiçoada em vários aspectos, como no poder resolutivo, tornando-se operacionalmente simples devido a gradientes de pH, comercialmente disponíveis, imobilizados em suportes. Dependendo das condições de fracionamento e da espécie biológica investigada, essa técnica permite observar centenas de proteínas simultaneamente⁽³¹⁾. No campo do estudo de proteomas, essa metodologia permitiu identificar a presença e a localização de proteínas nas células, caracterizando marcadores de doença e alvos de medicamentos.

A identificação de marcadores pela técnica do gel 2D consiste em contrastar diferenças no gel de indivíduos portadores da patologia em estudo em relação ao de indivíduos saudáveis. Após coloração por prata ou azul de coomassie, os *spots* no gel com intensidades diferentes podem representar os biomarcadores, nesse caso, proteínas diferencialmente expressas, isoladas de acordo com o seu ponto isoeletrico e sua massa molecular. A limitação inicial na identificação da maioria dos sinais protéicos revelados pela segunda dimensão foi modificada devido ao auxílio dos espectrômetros de massa e dos computadores. Para que uma molécula seja analisada por espectrometria de massa, ela primeiramente deve ser ionizada. Com a criação de metodologias capazes de ionizar biomoléculas sem degradá-las, a caracterização delas pelo respectivo perfil de massas resultante de uma digestão enzimática tornou-se possível. À medida que essas máquinas evoluíram, foi possível utilizar programas mais sofisticados para tal finalidade (exs.: MS-FIT, em <http://prospector.ucsf.edu>, e MASCOT, em www.matrixscience.com).

Em recente trabalho, mostramos a aplicabilidade de tal método em pacientes com câncer de pulmão do Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro (HUCFF/UFRJ), onde cinco proteínas diferencialmente expressas com potencialidade de uso para diagnóstico foram caracterizadas⁽²⁰⁾. Contudo, tal metodologia possui limitações. A de análise 2D é um processo demorado, de alto custo, laborioso, restrito geralmente à observação de proteínas desnaturadas em uma faixa de pH entre 3,5 e 11,5 e com peso molecular variando de 7 ~ 150kDa. Essa tecnologia também se torna inviável para estudos de proteomas em larga escala, uma vez que seria necessária a realização de diversos géis bidimensionais para estudar uma população. A utilização de *pools* de fluidos biológicos torna-se uma alternativa eficiente para avaliar diversos materiais biológicos de uma só vez. Contudo, ao estudar *pools*, não é possível obter a freqüência das proteínas diferencialmente expressas, perdendo-se a indi-

vidualidade molecular dos pacientes⁽¹⁾. A necessidade do conhecimento dessas frequências torna-se imprescindível na criação de modelos probabilísticos para viabilizar painéis de marcadores a serem utilizados no diagnóstico e prognóstico personalizados. Por essas razões, a abordagem para o estudo de proteomas complexos busca novas técnicas de análise para complementar a técnica da eletroforese.

Introdução à espectrometria de massa

O espectrômetro de massa é um instrumento analítico capaz de converter moléculas neutras em íons na forma gasosa e separá-las de acordo com a sua razão massa/carga (m/z), utilizando para isso campos eletromagnéticos. Esse equipamento atua como uma balança de íons de altíssima precisão e, em sua grande maioria, é composto por uma fonte ionizante, analisador(es), e detector(es), conforme esquematizado na **Figura 1**. Como resultado é emitido um gráfico onde o eixo y representa a intensidade do sinal dos íons e o eixo x, a razão m/z destes.

Apesar de a EM ser uma tecnologia usada há bastante tempo no campo das engenharias e da física, era pouco utilizada nas ciências da saúde por degradar biomoléculas ao ionizá-las. Recentemente foram desenvolvidas duas técnicas capazes de ionizar biomoléculas de alto peso molecular, como oligopeptídeos e proteínas⁽³⁰⁾. Estas técnicas também são capazes de alterar o analito de sua fase sólida ou líquida para gasosa, estado necessário para realização da análise da EM sem degradá-lo. Tais técnicas foram denominadas *matrix assisted laser desorption ionization* (MALDI)⁽¹⁵⁾ e *electron spray ionization* (ESI) aperfeiçoada para estudos de macromoléculas biológicas por John Fenn *et al.* em 1989⁽⁷⁾.

Reconhecimento de padrões moleculares por EM para auxílio no diagnóstico médico

Em 2002, o grupo de Lance Liotta, do National Institutes of Health (NIH), publicou metodologia inovadora para o

diagnóstico do câncer de ovário⁽²³⁾. Os autores têm como premissa que proteínas secretadas por tumores podem ser observadas, em concentrações femtomolares, por metodologias estado-da-arte em EM, analisando padrões complexos diretamente no soro, sem isolamento prévio das proteínas. Neste trabalho, foi usada uma variante da técnica de ionização MALDI, chamada *surface enhanced laser desorption ionization – time of flight* (SELDI-TOF MS), que dá preferência ao estudo de peptídeos hidrofóbicos de baixo peso molecular ($< \sim 10$ kDa). Na ocasião foi analisado o perfil proteômico do soro de 50 portadoras de câncer de ovário e de 50 mulheres não afetadas. Os resultados permitiram discriminar picos no espectro de massa correspondente a peptídeos específicos para cada grupo. Para validar o método, um grupo adicional de 116 amostras foi analisado. Baseando-se apenas nos picos diferencialmente intensificados oriundos de seu modelo prévio, foi possível diagnosticar os casos de câncer de ovário em 100% e de mulheres não afetadas pela doença em 95%. Posteriormente foi mostrado que o padrão proteômico encontrado era específico de câncer de ovário e diferente dos obtidos em soro de pacientes com neoplasia benigna ou maligna de próstata. Por essas razões foi concluído que a técnica era capaz também de fornecer diagnóstico diferencial⁽²⁴⁾. Este trabalho impulsionou inúmeros outros similares em que diferentes grupos aplicaram a metodologia em outros tipos de tumores e tecidos^(17, 33).

Diversas críticas aos trabalhos iniciais do estudo de proteoma por EM diferencial foram realizadas pela comunidade científica, tendo como principais argumentos a dificuldade de reprodutibilidade dos dados e as limitações na resolução do SELDI-TOF utilizado na época^(5, 25). Atualmente empregam-se espectrômetros de massa de maior resolução, como o Micromass Ultima Q-TOF, dotado de fonte ionizante do tipo *electrospray*. A remoção de albumina do soro e o uso de cromatografia acoplada a EM são práticas utilizadas para reduzir o efeito de supressão de ionização e saturação do detector. Outra razão para a remoção de proteínas abundantes é dada pela supressão do sinal detectado no EM de



Figura 1 – Esquema simplificado de um espectrômetro de massa

proteínas expressas em baixa quantidade que também possuem vir a ser os biomarcadores^(6, 35). A utilização de padrões internos com quantidade e massa previamente conhecidos também é técnica que permite um melhor controle perante a reprodutibilidade dos espectros. É importante ressaltar que, mesmo em experimentos nos quais a amostra é analisada por infusão direta no espectrômetro, apenas uma parte das biomoléculas é ionizada. A caracterização das biomoléculas, em alguns casos, pode ser realizada pelo uso de EM em *tandem*. Essa técnica permite evidenciar a seqüência de aminoácidos, no caso, de peptídeos, para posterior identificação, contudo está limitada a peptídeos de baixo peso molecular (<~ 5kDa).

A metodologia de EM diferencial, em conjunto com técnicas de reconhecimento de padrões (redes neurais, árvores de decisão, etc.), tem sido utilizada extensamente para o estudo de diversas neoplasias^(4, 9, 18). Recentemente demonstramos a efetividade da técnica de EM diferencial em pacientes com a doença de Hodgkin (DH)⁽²⁾ do HUCFF/UFRJ. A DH é uma neoplasia que ocorre em um linfonodo, eventualmente evoluindo para linfonodos mais próximos. Isso caracteriza a diferença entre seus estágios inicial e avançado⁽¹²⁾. Foram obtidos espectros de massa (Micromass Ultima ESI Q-TOF) do soro de 30 pacientes com DH antes do tratamento e de 30 indivíduos controles. Um algoritmo foi elaborado com base no princípio de minimização de risco estrutural⁽²⁸⁾ apontando oito picos com intensidades diferenciadas entre os controles e os pacientes com DH. A **Figura 2** mostra uma região dos espectros de massa contendo picos considerados representativos de um possível marcador. Com base no método de validação cruzada *leave-one-out*⁽¹⁴⁾ e nos oito picos selecionados, foi possível classificar corretamente todos os espectros de massa como pertencendo a doente ou controle. Para verificar quanto à reprodutibilidade do experimento, um segundo *set* de espectros oriundos dos mesmos indivíduos foi utilizado e classificado corretamente. De acordo com o modelo estatístico, estimou-se um erro de 8% de generalização para futuras amostras. A vantagem desse método sobre os atuais foi ser ele o primeiro a embasar-se no princípio de minimização de risco estrutural para análise de dados de EM. Esse princípio visa uma solução que generaliza eficientemente para futuras amostras, superando inclusive a técnica de redes neurais^(13, 19, 34). Outra vantagem do algoritmo apresentado foi o fato de aproveitar-se de dados intrínsecos do espectro como informações da distribuição isotópica do marcador para minimizar o risco de conter picos não provenientes de peptídeos na solução final.

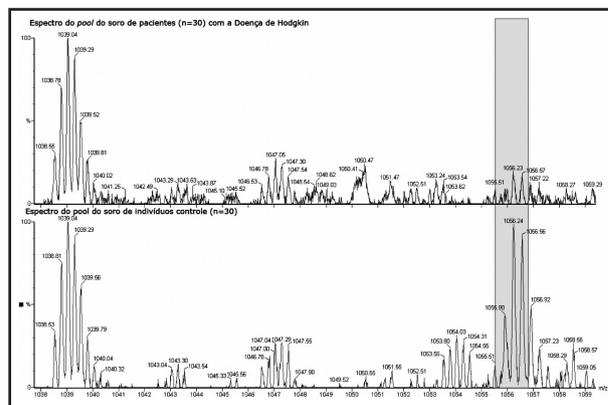


Figura 2 – Espectro de massa dos pools de pacientes com a doença de Hodgkin (superior) e indivíduos controles (inferior). O eixo y representa intensidade do sinal e o eixo x, a razão massa/carga (m/z) dos peptídeos ionizados. A janela indica os picos com intensidade diferenciada de razão m/z de 1.056 indicando a presença de um possível marcador

Considerações finais

A existência de padrões específicos no proteoma e, mais recentemente, no perfil de lipídios (lipidoma)^(27, 32) reforça a hipótese de que biomoléculas podem originar-se de microambientes do tumor hospedeiro. Recentemente também foi mostrado que diversas alterações moleculares indiretamente ligadas à patologia podem ser detectadas por EM para aumentar a certeza do diagnóstico⁽²⁵⁾. O fato estabelece um novo paradigma, no qual a tecnologia dos *omas*, aliada à bioinformática, poderá trazer impacto direto na clínica médica. A técnica introduzida por Liotta é inovadora por capacitar estudos em grandes populações amostrais e apontar um caminho no qual diagnósticos poderão basear-se no nível da expressão de diversas biomoléculas simultaneamente, aumentando suas especificidade e sensibilidade. Devido à nossa individualidade molecular, são necessários arranjos de marcadores para compor a base de modelos probabilísticos de diagnósticos diferenciados e personalizados⁽¹⁾. A caracterização dos biomarcadores selecionados é condição necessária para a utilização dos padrões moleculares no auxílio de diagnósticos, prognósticos e também para fornecer maiores conhecimentos, ao nível molecular, sobre a patologia em questão.

Agradecimentos

Fundação Ary Frauzino/Educacional Charles Darwin, Programa Estratégico de Apoio à Pesquisa em Saúde da Fundação Oswaldo Cruz (PAPES/FIOCRUZ), Programa de Desenvolvimento Tecnológico em Insumos para Saúde (PDTIS), Fundação Carlos Chagas de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

1. CARVALHO, P. C. et al. Detection of potential serum molecular markers for Hodgkin's disease. *J Bras Patol Med Lab*, v. 41, n. 3, p. 99-103, 2005.
2. CARVALHO, P. C. *Reconhecimento de padrões proteômicos e genômicos por aprendizagem de máquinas para o diagnóstico médico*. 2006. Tese (Mestrado) – Departamento de Biologia Celular e Molecular do Instituto Oswaldo Cruz, Rio de Janeiro.
3. CHAND, H. S.; NESS, S. A.; KISIEL, W. Identification of a novel human tissue factor splice variant that is upregulated in tumor cells. *Int J Cancer*, v. 118, n. 7, p. 1713-20, 2005.
4. CONRADTS, T. P. et al. Cancer diagnosis using proteomic patterns. *Expert Rev Mol Diagn*, v. 3, n. 4, p. 411-20, 2003.
5. DIAMANDIS, E. P.; VAN DER MERWE, D. E. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin Cancer Res*, v. 11, n. 3, p. 963-5, 2005.
6. ECHAN, L. A. et al. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics*, v. 5, n. 13, p. 3292-303, 2005.
7. FENN, J. B. et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, v. 246, n. 4926, p. 64-71, 1989.
8. FILIPOWICZ, W. et al. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol*, v. 15, n. 3, p. 331-41, 2005.
9. GRONBORG, M. et al. Biomarker discovery from pancreatic cancer secretome using a differential proteomics approach. *Mol Cell Proteomics*, v. 5, p. 157-71, 2005.
10. HACK, C. J. Integrated transcriptome and proteome data: the challenges ahead. *Brief Funct Genomic Proteomic*, v. 3, n. 3, p. 212-9, 2004.
11. HODGETTS, R. Eukaryotic gene regulation by targeted chromatin re-modeling at dispersed, middle-repetitive sequence elements. *Curr Opin Genet Dev*, v. 14, n. 6, p. 680-5, 2004.
12. HOSTER, H. A. Hodgkin's disease. *Am J Roentgenol Radium Ther Nucl Med*, v. 64, n. 6, p. 913-8, 1950.
13. HUANG, Y. L.; CHEN, D. R. Support vector machines in sonography: application to decision making in the diagnosis of breast cancer. *Clin Imaging*, v. 29, n. 3, p. 179-84, 2005.
14. JOACHIMS, T. Estimating the generalization performance of an SVM efficiently. In: 17th International Conference on Machine Learning, 1999, San Francisco. *Proceedings of the 17th International Conference on Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc., 2000. p. 431-8.
15. KARAS M. et al. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes*, v. 78, p. 53-68, 1987.
16. KRAMER, D. A. Commentary: gene-environment interplay in the context of genetics, epigenetics, and gene expression. *J Am Acad Child Adolesc Psychiatry*, v. 44, n. 1, p. 19-27, 2005.
17. LI, J. et al. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem*, v. 48, n. 8, p. 1296-304, 2002.
18. LIU, J. J. et al. Multiclass cancer classification and biomarker discovery using GA-based algorithms. *Bioinformatics*, v. 21, n. 11, p. 2691-7, 2005.
19. LU, W. Z.; WANG, W. J. Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*, v. 59, n. 5, p. 693-701, 2005.
20. MARCIEL, C. M. et al. Differential proteomic serum pattern of low molecular weight proteins expressed by adenocarcinoma lung cancer patients. *JETO*, v. 5, p. 21-28, 2005.
21. O'FARRELL, P. H. High resolution two dimensional electrophoresis of proteins. *J Biol Chem*, v. 25, n. 4007-21, 1975.
22. PANDEY, R. R. et al. NF-Y regulates the antisense promoter, bidirectional silencing, and differential epigenetic marks of the Kcnq1 imprinting control region. *J Biol Chem*, v. 279, n. 50, p. 52685-93, 2004.
23. PETRICOIN, E. F. et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, v. 359, n. 9306, p. 572-7, 2002.
24. PETRICOIN, E. F., III et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst*, v. 94, n. 20, p. 1576-8, 2002.
25. PETRICOIN, E., III; LIOTTA, L. A. Counterpoint: the vision for a new diagnostic paradigm. *Clin Chem*, v. 49, n. 8, p. 1276-8, 2003.
26. SOUFLA, G. et al. Transcriptional deregulation of VEGF, FGF2, TGF-beta1, 2, 3 and cognate receptors in breast tumorigenesis. *Cancer Lett*, v. 221, p. 105-18, 2005.
27. UMEZU-GOTO, M. et al. Lysophosphatidic acid production and action: validated targets in cancer? *J Cell Biochem*, v. 92, n. 6, p. 1115-40, 2004.
28. VAPNIK, V. N. *The nature of statistical learning theory*. New York: Springer-Verlag, 1995.
29. VENTER, J. C. et al. The sequence of the human genome. *Science*, v. 291, n. 5507, p. 1304-51, 2001.
30. VESTAL, M. L. Methods of ion generation. *Chem Rev*, v. 101, n. 2, p. 361-75, 2001.
31. WEINGARTEN, P. et al. Application of proteomics and protein analysis for biomarker and target finding for immunotherapy. *Methods Mol Med*, v. 109, p. 155-74, 2005.
32. WENK, M. R. The emerging field of lipidomics. *Nat Rev Drug Discov*, v. 4, n. 7, p. 594-610, 2005.
33. YASUI, Y. et al. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, v. 4, n. 3, p. 449-63, 2003.

34. ZHAO, C.Y. *et al.* Application of support vector machine (SVM) for prediction toxic activity of different data sets. *Toxicology*, v. 217, n. 2-3, p. 105-119, 2006.
35. ZOLOTARJOVA, N. *et al.* Differences among techniques for high-abundant protein depletion. *Proteomics*, v. 5, n. 13, p. 3304-13, 2005.

Endereço para correspondência

Maria da Glória da Costa Carvalho
Instituto de Biofísica Carlos Chagas Filho
CCS - Bloco G - UFRJ
Ilha do Fundão
CEP: 21949-900 - Rio de Janeiro-RJ
e-mail: paulo@buscario.com.br