# Peptidomics for Cancer Diagnosis: Present and Future

## Eleftherios P. Diamandis*

*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada*

*Abstract:* Analysis of peptides in biological fluids by mass spectrometry holds promise of providing diagnostic and prognostic information for cancer and other diseases. Before this technology is used clinically, it is important to understand its advantages and limitations. These are summarized, along with proposals, on how to proceed in the future.

"Peptidomics" is the field that deals with the comprehensive qualitative and quantitative analysis of peptides in biological samples.[1] These peptides are either intact small molecules, such as hormones, cytokines, and growth factors, peptides that are released form larger protein precursors during protein processing or they may represent degradation products of proteolytic activity. Thus, in biological fluids, peptides represent protein synthesis, processing, and degradation. Since the amount and repertoire of peptides in the circulation change dynamically according to the physiological or pathological state of an individual, it is possible that comprehensive peptide analysis (i.e., exploitation of the "peptidome") may lead to discovery of novel biomarkers or to new diagnostic approaches.[2]

Already, a number of peptides are used for diagnostic purposes through individual assays, usually ELISAs. Examples include insulin and C-peptide for diabetes, parathyroid hormone, calcitonin, and collagen fragments for osteoporosis, pro-brain-type naturetic peptide (pro-BNP) for congestive heart failure, pro-gastrin releasing peptide (pro-GRP) for small cell lung carcinoma, $\beta$-amyloid 1−42 for Alzheimer's disease and angiotensin II for hypertension.

Recently, powerful analytical technologies and, especially, mass spectrometry, allowed identification of numerous and previously unknown peptides in the circulation. The extraordinary power of mass spectrometry in identifying and quantifying peptides in complex biological mixtures offers opportunities for developing novel technologies for diagnosis of cancer and other diseases. Such technologies include the following important sub-components:

1. Development of optimized methods for sample collection and processing, peptide extraction, chromatographic separation and analytical detection.

2. Development of appropriate bioinformatic tools which would allow multiparametric analysis of potentially thousands of informative peptides and differential approaches for comparing normal peptide patterns with those found in disease states.

3. Development of methods for positive peptide identification through sequence determination, to derive clues for their origin and their possible biological function.

In one of the first attempts to use serum proteomic (including peptidomic) profiling, in combination with surface-enhanced laser desorbtion−ionization time-of-flight mass spectrometry (SELDI-TOF−MS), Petricoin et al. reported outstanding sensitivities and specificities for early ovarian cancer diagnosis by using peaks of unknown identity.[3] The $m/z$ values used as diagnostic peaks in this paper (534, 989, 2111, 2251, and 2465) suggest that they likely represent peptides of relatively low molecular weight (i.e., < 50 amino acids). Subsequently, others developed similar technologies for diagnosing numerous other cancer types, with claimed sensitivities and specificities that far exceed those achieved with the classical cancer biomarkers (reviewed in ref 4). It was thus postulated that a "new era" in cancer diagnostics had emerged, in which serum proteomic/peptidomic profiles would fulfill the long-sought goal of early cancer detection. Unfortunately, other groups soon identified major methodological and bioinformatic artifacts and biases, which questioned the validity of the published results.[5−8] It is now clear that sample collection, storage, and processing procedures can produce proteomic patterns that could overshadow those generated by the presence of disease.[9,10] Naturally, the enthusiasm of using these technologies for diagnostics has declined, in part due to the inability of other groups to reproduce or validate the originally published data.[11,12]

In 2003, Marshall et al. claimed that peptides from the sera of normal individuals and patients who suffered myocardial infarction (MI) can produce MALDI-TOF patterns that provide an accurate diagnosis of myocardial infarction.[13] These authors have shown that the spectral patterns mainly originated from the cleavage of complement C3 alpha chain to release the C3f peptide, and from cleavage of fibrinogen to release peptide A. The fibrinogen peptide A and complement C3f peptides were in turn progressively truncated by aminopeptidases to produce two families of fragments that formed the characteristic spectral pattern of MI. These authors have shown that the peptide patterns in serum reflected the balance of disease-specific protease and aminopeptidase activity ex-vivo. Around the same time, Liotta et al. postulated that serum and/or plasma contains a large repertoire of different peptides which are bound to high abundance proteins such as albumin and are thus protected

---

* To whom correspondence should be addressed. Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Tel: (416) 586-8443. Fax: (416) 586-8628. E-mail: ediamandis@mtsinai.on.ca.
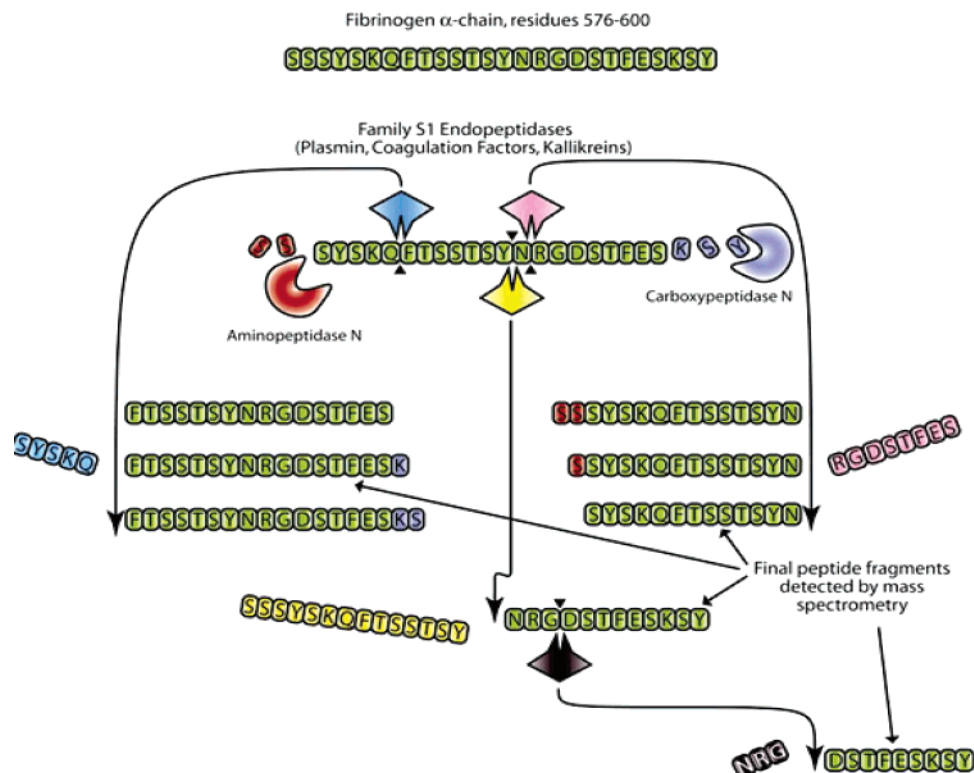
**Figure 1.** Mechanisms of generation of multiple peptides from a parent cleavage fragment of fibrinogen α-chain, used as an example. Initially, this fragment is proteolyzed by endoproteases such as plasmin, coagulation factors, and kallikreins. Subsequently, the original fragment and other fragments are further processed by aminopeptidases and carboxypeptidases, as shown. At the end, a large family of peptides is generated from the original fragment. Amino acids are shown with single letter code.

from clearance by the kidneys.[14] They further hypothesized that these peptides (the serum "peptidome", "fragmentome" or "degradome") may have important diagnostic value.

Apparently, there are various classes of peptides in serum, some of them of relatively high abundance (originating from degradation of high abundance proteins) and others of relatively low or very low abundance (e.g., cytokines, hormones, etc.). It is possible that low molecular weight peptides may carry as yet unrealized diagnostic information. The diagnostic potential of low molecular weight peptides has recently been explored by Lopez et al. for diagnosis of Alzheimer's disease[15] and by Lowenthal et al. for diagnosis of ovarian carcinoma.[16]

Recently, Koomen et al. provided important information on the generation of the serum peptidome by using peptide extraction, fractionation, and characterization by liquid chromatography coupled to MALDI tandem mass spectrometry.[17] These authors were able to identify more than 250 peptides in plasma and demonstrated that they originated from a surprisingly small number of proteins (∼20) which were all common and of high abundance, including fibrinogen, complement components, anti-proteases, apoliproproteins, acute phase reactants, and carrier proteins. The mechanism of generation of multiple peptides from a small number of abundant proteins is shown in Figure 1. It is postulated that initial endoproteolytic cleavages of these abundant proteins occur by common enzymes such as thrombin, plasmin, and complement proteins, followed by aminopeptidase and carboxypeptidase exoprotease processing.

Recently, Villanueva et al. suggested a novel way of diagnosing cancer by using peptidomic analysis, combined with MALDI-TOF mass spectrometry.[18] The method received enthusiastic endorsement by some.[19] The procedure of Villanueva et al. has many technical similarities to the one originally proposed by Petricoin et al.[3] but it is based on a different hypothesis. Villanueva et al. postulate that their informative diagnostic peptides, identified by mass spectrometry, originate after ex-vivo exoproteolytic processing of high abundance protein fragments, primarily generated by the coagulation and complement enzymatic cascades. The novel aspect of this approach is that the candidate discriminatory peptides (the biomarkers) are not present in serum at the time of sample collection but they are generated after the coagulation and complement cascades are activated. Once high abundance fragments from proteins of the coagulation and complement cascades are generated, it is postulated that tumor-specific (but currently unknown) exoproteases generate series of peptides with diagnostic potential. This method introduces for the first time, the concept that cancer biomarkers, which are not present in the sample at the time of blood collection, are generated ex-vivo. The diagnostic sensitivity of this method for prostate cancer was claimed to be 100%. Rather surprisingly, these ex-vivo generated peptides were able to discriminate not only normal individuals from cancer patients, but various cancer types from each other (breast, prostate, bladder).

What are the realistic prospects that these technologies can become practical ways of diagnosing cancer?

First, let us examine if the so-called "serum peptidome" exists. Indeed, Koomen et al. positively identified over 250 peptides, Villanueva et al. over 650 peptides and Lowenthal et al., more than 1200 peptides in human serum.[16−18] Others have already assembled large databases (unpublished or proprietary) of serum peptides as well. It is thus clear that this large repertoire of peptides exists, can be isolated from serum or

plasma by using robust extraction techniques[15−18] and can be identified efficiently by mass spectrometry. Villanueva et al. have further shown that not all peptides are necessary for accurate classification of various types of cancers. The 68 most informative peptides, used in their final analysis, are almost identical to those identified by Koomen et al.[17]

Second, it seems that there is also unanimous agreement that most of these peptides are generated ex-vivo after the coagulation and complement cascades are initiated. While Koomen et al. consider this a disadvantage for using these peptides for diagnostic purposes, Villanueva et al. believe that this is the key to the success of their method. A few years ago, van Hensbergen et al. showed that in both malignant effusions and intratumoral fluid, as well as in plasma, there is increased aminopeptidase N activity which correlates with tumor load, suggesting that this enzyme may be originating from tumor cells.[20] Similar observations for carboxypeptidase N activity increase in myocardial infarction have also been published by Zaninotto et al.[21] Thus, as Villanueva et al. suggest, there is a chance that what they see may correlate with exoprotease activities originating from tumors.

Did others identify proteolytic protein fragments as candidate cancer biomarkers? Zhang et al. reported a transthyretin fragment (truncated transthyretin) that appears to be down-regulated in ovarian cancer.[22] The down-regulation is in conflict with the hypothesis of increased endoproteolytic or exoproteolytic activity in cancer. The same group identified a cleavage fragment of inter-alpha-trypsin inhibitor heavy chain H4 which was apparently up-regulated in serum of patients with ovarian cancer. The same peptide was identified in plasma by Koomen et al. and by Villanueva et al. as a putative bladder but not prostate or breast cancer biomarker (increased intensity in patients). Furthermore, Li et al. attempted to validate three previously identified breast cancer biomarkers, of which two of them represented truncated forms of C3a.[23] Series of peptides from complement C3 were also observed by both Koomen et al. and Villanueva et al. Unfortunately, validation of these truncated forms of C3 has shown that their diagnostic value is minimal or nonexistent.[24] More recently, fragmented proteins such as eosinophil-derived neurotoxin and COOH-terminal osteopontin were proposed as ovarian cancer biomarkers in urine[25] but have not as yet been validated. In short, none of the available data on protein fragments as diagnostic markers is either strong enough or has passed validation.

On the basis of the available data, what is the possible diagnostic value of "peptidomics" in cancer?

[1] There is now convincing evidence that a large variety of peptides can be found in biological fluids and these can be extracted efficiently and identified by mass spectrometry. Comparisons between serum and plasma confirm that the vast majority of high abundance peptides in serum originate ex-vivo by the activity of the coagulation and complement cascades. A major question that begs an answer is if serum or plasma are the preferred specimens for biomarker discovery studies. In my view, and as also proposed by Koomen et al. and others, use of serum for peptidomic analysis introduces a serious complication. It seems preferable to avoid, as much as possible ex-vivo generation of peptides before peptidomic analysis in plasma. Specialized tubes, additives, and optimized procedures for arresting ex-vivo peptide generation should be developed and validated.

[2] On the basis of the available data, the small group of protein fragments, originating from high abundance proteins

| Complement C3f fragments in serum | Prostate | Bladder | Breast |
|---|---|---|---|
| SSKITHRIHWESASLL | Increase | Increase | Decrease |
| SKITHRIHWESASLL | Increase | Increase | No change |
| KITHRIHWESASLL | Increase | Increase | No change |
| ITHRIHWESASLL | No change | No change | Decrease |
| THRIHWESASLL | Increase | Increase | No change |
| HRIHWESASLL | No change | No change | No change |
| RIHWESASLL | No change | No change | No change |
| IHWESASLL | Increase | Increase | No change |
| HWESASLL | No change | Increase | Decrease |

**Figure 2.** Series of peptides originating from complement C3f fragment in serum after proteolytic processing by an aminopeptidase. This series, reported by Koomen et al.[17] and Villanueva et al.,[18] was shown by the latter group to have diagnostic value in prostate, bladder and breast cancer. The intensity of the identified peptides (increase or decrease) is shown for the three cancer types. For more discussion, see text. Amino acids are shown with single letter code.

and identified by SELDI-TOF mass spectrometry, do not appear upon reevaluation by confirmatory techniques such as ELISA to have much value in diagnostics.

As postulated earlier,[5−7] putative new cancer biomarkers, including peptides, are likely to be present in biological fluids at extremely low concentrations, necessitating identification by quantitative, highly sensitive and reproducible techniques. Convincing data for identifying low abundance, tumor-derived peptides for diagnostic purposes are currently lacking.

We have indicated earlier that the paper by Villanueva et al. suffers from important design biases which put their findings in serious question.[26] For example, these authors selected young men and women as the control group (median age 31 years) in contrast to the much older patient groups (median ages of 66−67 years for prostate cancer, 49 years for breast cancer and 67 years for bladder cancer). It is thus entirely possible that their findings are related more to the age of the patients than to the presence of cancer. Despite the caution by Ransohoff in designing studies of this kind,[11] high-profile papers like this one are still compromised by avoidable bias.

[3] Villanueva et al. postulated that tumor-related exoprotease activity may be responsible for the observed findings. However, this hypothesis is inconsistent with their data. For example, a series of peptides was informative for prostate, bladder and breast cancer (Figure 2). These peptides originate from one parental peptide that is further processed by a putative aminopeptidase. If this aminopeptidase was increased due to tumor presence, then one would expect that the whole series of daughter peptides would increase in all types of cancer. However, while a number of peptides increase (and others do not change), in prostate and bladder cancer, three of these peptides actually decreased in breast cancer. This puts into question the hypothesis that tumor-derived aminopeptidases are likely generating the diagnostic information.

It is now time to approach proteomic and peptidomic profiling for cancer diagnosis with increased scrutiny, to avoid biases and move to the next step (i.e., beyond the "proof-of-principle" stage). Some proposals are summarized below:

1. Since these methods are significantly affected by sample collection and storage conditions,[9,10] it should be mandatory that in future publications, these parameters are thoroughly examined.

2. Studies should be designed carefully so that biological and bioinformatic biases are avoided as much as possible, as previously described.[11]

3. The effects of variables such as age, gender, common drug ingestion, ethnicity, dietary habits, exercise, etc. should be

## communications

studied, to establish if the methods are robust enough for routine testing.

4. Since some of the methods are dependent on ex-vivo proteolysis and on pathways such as complement and coagulation, the effects of coagulopathies and inflammatory conditions should be reported.

5. When deriving clinical sensitivities and specificities, patients with early stage, in addition to late stage disease, and age-matched controls with nonmalignant inflammatory or other related conditions should be included in the study.

6. The original hypothesis should not be conflicting with the data provided.

7. In cases where putative proteolytic activity is implicated as the discriminatory factor, it is important to characterize the proteolytic activity, to strengthen the hypothesis.

8. The identified biomarkers or panels should be examined with other standard parameters, i.e., correlation with tumor burden and other clinicopathological variables, change of biomarker concentration after treatment or post- surgery, etc.

9. All cancer biomarkers used at the clinic today are increased in the circulation since they are derived from tumor cells. Biomarker concentration decrease with tumor burden increase should be interpreted with caution and explained biologically, where possible.

10. Last but not least, these new diagnostic technologies must be evaluated by using the principles proposed by Pepe et al.[27]

These new diagnostic methods will be accepted at the clinic only when it is shown in large and well-designed validation studies that they have value for patient care. In such studies, the originally derived algorithms and multiparametric schemes proposed by the authors should be applied to patients who are assessed blindly and with samples originating from different institutions under standardized sample collection and storage conditions. It will be highly important that validation studies with negative results are published, so that the scientific literature is cleaned from data that do not represent real advances in the field.

### References

(1) Schulte, I.; Tammen, H.; Selle, H.; Schulz-Knappe, P. Peptides in body fluids and tissues as markers of disease. *Expert Rev. Mol. Diagn.* **2005**, *5*, 145−157.

(2) Schrader, M.; Schulz-Knappe, P. Peptidomics technologies for human body fluids. *Trends Biotechnol.* **2001**, *19*, S55−60.

(3) Petricoin, E. F., III; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002**, *359*, 572−577.

(4) Wulfkuhle, J. D.; Liotta, L. A.; Petricoin, E. F. Proteomic applications for the early detection of cancer. *Nat. Rev. Cancer* **2003**, *3*, 267−275.

(5) Diamandis, E. P. Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics? *Clin. Chem.* **2003**, *49*, 1272−1278.

(6) Diamandis, E. P. Analysis of serum proteomic patterns for early cancer diagnosis: Drawing attention to potential problems. *J. Natl. Cancer Inst.* **2004**, *96*, 353−356.

(7) Diamandis, E. P.; van der Merwe, D.-E. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin. Cancer Res.* **2005**, *11*, 963−965.

(8) Baggerly, K. A.; Morris, J. S.; Edmonson, S. R.; Coombes, K. R. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J. Natl. Cancer Inst.* **2005**, *97*, 307−309.

(9) Banks, R. E.; Stanley, A. J.; Cairns, D. A.; Barrett, J. H.; Clarke, P.; Thompson, D.; et al. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin. Chem.* **2005**, *51*, 1637−1649.

(10) Karsan, A.; Eigl, B. J.; Flibotte, S.; Gelmon, K.; Switzer, P.; Hassell, P.; et al. Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin. Chem.* **2005**, *51*, 1525−1528.

(11) Ransohoff, D. F. Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer* **2005**, *5*, 152−149.

(12) Ransohoff, D. F. Lessons from controversy: ovarian cancer screening and serum proteomics. *J. Natl. Cancer Inst.* **2005**, *97*, 315−319.

(13) Marshall, J.; Kupchak, P.; Zhu, W.; Yantha, J.; Vrees, T.; Furesz, S.; et al. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J. Proteome Res.* **2003**, *2*, 361−372.

(14) Liotta, L. A.; Ferrari, M.; Petricoinm E. Clinical proteomics: written in blood. *Nature* **2003**, *425*, 905.

(15) Lopez, M. F.; Mikulskis, A.; Kuzdzal, S.; Bennett, D. A.; Kelly, J.; Golenko, E.; et al. High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific, carrier-protein-bound mass signatures. *Clin. Chem.* **2005**, *51*, 1946−1954.

(16) Lowenthal, M. S.; Mehta, A. I.; Frogale, K.; Bandle, R. W.; Araujo, R. P.; Hood, B. L.; et al. Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin. Chem.* **2005**, *51*, 1933−1945.

(17) Koomen, J. M.; Li, D.; Xiao, L.-C.; Liu, T. C.; Coombes, K. R.; Abbruzzese, J.; et al. Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. *J. Proteome Res.* **2005**, *4*, 972−981.

(18) Villanueva, J.; Shaffer, D. R.; Philip, J.; Chaparro, C. A.; Erdjument-Bromage, H.; Olshen, A. B.; et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest.* **2006**, *116*, 271−284.

(19) Liotta, L. A.; Petricoin, E. F. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J. Clin. Invest.* **2006**, *116*, 25−30.

(20) van Hensbergen, Y.; Broxterman, H. J.; Hanemaaijer, R.; Jorna, A. S.; van Lent, N. A.; Verheul, H. M.; et al. Soluble aminopeptidase N/CD13 in malignant and nonmalignant effusions and intratumoral fluid. *Clin. Cancer Res.* **2002**, *8*, 3747−3754.

(21) Zaninotto, M.; Altinier, S.; Lachin, M.; Plebani, M. Carboxypeptidase N and creatine kinase-MB isoforms in acute myocardial infarction. *Eur. J. Clin. Chem. Clin. Biochem.* **1997**, *35*, 291−295.

(22) Zhang, Z.; Bast, R. C., Jr.; Yu, Y.; Li, J.; Sokoll, L. J.; Rai, A. J.; et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **2004**, *64*, 5882−5890.

(23) Li, J.; Orlandi, R.; White, C. N.; Rosenzweig, J.; Zhao, J.; Seregni, E.; et al. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clin. Chem.* **2005**, *51*, 2229−2235.

(24) Diamandis, E. P. Validation of breast cancer biomarkers identified by mass spectrometry. *Clin. Chem.* **2006**, *52*, 771−772.

(25) Ye, B.; Skates, S.; Mok, S. C.; Horick, N. K.; Rosenberg, H. F.; Vitonis, A.; et al. Proteomic-based discovery and characterization of glycosylated eosinophil-derived neurotoxin and COOH-terminal osteopontin fragments for ovarian cancer in urine. *Clin. Cancer Res.* **2006**, *12*, 432−441.

(26) Diamandis, E. P.; Kulasingam, V.; Sardana, G. Letter to the Editor about Differential exoprotease activities confer tumor-specific serum peptidome. *J. Clin. Invest.* **2006**; http://www.jci.org/cgi/eletters/116/1/271 #464.

(27) Pepe, M. S.; Etzioni, R.; Feng, Z.; Potter, J. D.; Thompson, M. L.; Thornquist, M.; et al. Phases of biomarker development for early detection of cancer. *J. Natl. Cancer Inst.* **2001**, *83*, 1054−1061.