Review

Is isolation of comprehensive human plasma peptidomes an achievable quest?☆,☆☆

S. Mahboob [a], A. Mohamedali [b], S.B. Ahn [a], P. Schulz-Knappe [c], E. Nice [d], M.S. Baker [a,*]

[a] Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, NSW 2109, Australia
[b] Department of Chemistry and Biomolecular Sciences, Faculty of Science, Macquarie University, NSW 2109, Australia
[c] Protagen AG, 44227 Dortmund Germany
[d] Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia

ARTICLE INFO

ABSTRACT

The low molecular weight (LMW; <10 kDa)* plasma peptidome has been considered a source of useful diagnostic biomarkers and potentially therapeutic molecules, as it contains many cytokines, peptide hormones, endogenous peptide products and potentially bioactive fragments derived from the parent proteome. The small size of the peptides allows them almost unrestricted vascular and interstitial access, and hence distribution across blood–brain barriers, tumour and other vascular permeability barriers. Therefore, the peptidome may carry specific *signatures* or *fingerprints* of an individual's health, wellbeing or disease status. This occurs primarily because of the advantage the peptidome has in being readily accessible in human blood and/or other biofluids.

However, the co-expression of highly abundant proteins (>10 kDa) and other factors present inherently in human plasma make direct analysis of the blood peptidome one of the most challenging tasks faced in contemporary analytical biochemistry. A comprehensive compendium of extraction and fractionation tools has been collected concerning the isolation and micromanipulation of peptides. However, the search for a reliable, accurate and reproducible single or combinatorial separation process for capturing and analysing the plasma peptidome remains a challenge.

This review outlines current techniques used for the separation and detection of plasma peptides and suggests potential avenues for future investigation.

This article is part of a Special Issue entitled: HUPO 2014.
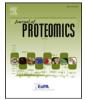
© 2015 Elsevier B.V. All rights reserved.

Contents

☆ For the purpose of avoiding confusion and for no other chemical or biochemical reason, we arbitrarily use the term peptidome to refer to amino acid-containing biomolecules that are ≤10 kDa, whilst small proteins are defined as being >10 kDa in molecular weight.
☆☆ This article is part of a Special Issue entitled: HUPO 2014.
* Corresponding author at: Department of Biomedical Sciences, Faculty of Medicine and Health Science, F10A, 2 Technology Place, Macquarie University, NSW 2109, Australia.
   E-mail address: mark.baker@mq.edu.au (M.S. Baker).

1. Introduction

From a clinical perspective, human blood plasma, the pale yellowish proteinaceous solution that usually holds all blood cells suspended, is one of the most informative and clinically important human proteomes in existence. The most recent version of the Human Proteome Organisation's (HUPO) Plasma Proteome Database [1] reports that

10,546 proteins have been detected in either human serum or plasma and that amongst these 3784 proteins have been reported in two or more mass spectrometry (MS)-based studies [2]. This has been realised not only because 20–25% of all proteins encoded by the human genome are known to be secreted into plasma [3], but also because plasma is one of the most readily available and accessible biofluids for research study. Human plasma is by far the most commonly sampled diagnostic biospecimen and the vast majority of patients are prepared to offer these biospecimens for clinical analysis.

Plasma is composed of high levels of water, proteins, glycoproteins, phosphoproteins, peptides, lipids, fatty acids, carbohydrates and ions [4]. Whilst the liver-synthesised protein albumin is by far the most abundant in plasma (~60 mg/mL) [5], many other less abundant plasma proteins come from sources like normal physiological tissue leakage and/or release during disease conditions [4]. Within the complex mixtures of biomolecules found in plasma, peptides can be simply defined as those amino acid sequences present in blood with a M_r less than or equal to 10 kDa. Plasma peptides such as hormones, cytokines and growth factors are an essential part of the homeostatic regulatory mechanism of many physiological processes [6]. Many years of research has revealed physiologically important peptides (e.g., amyloid peptides, vasoactive intestinal peptide (VIP); peptide histidine isoleucine (PHI), neuropeptide Y and thymosin β4 (Tβ4) amongst many others). These play critical roles in the regulation of biological processes in mammals (man, rat, guinea-pig, cat, pig, monkey, dog and even polar bear [7–10]). It is therefore not surprising that monitoring the profiles of endogenous plasma peptides (including neuropeptides, peptide hormones, growth factors and intracellular/extracellular products of proteolysis) can effectively assist in diagnosis and prognosis, as well as indicate treatment options for many diseases (e.g., cancer, cardiac, infection, inflammation, gastrointestinal complications and hormone therapy to name but a few) [11,12].

Following individual attempts by several industry and clinical proteomics/peptidomics research groups, the HUPO Plasma Proteome Project [13–15] was established in 2002. It aimed to comprehensively determine the protein/peptide constituents of human plasma or serum. The project involved 35 laboratories from around the globe and initially identified 3020 plasma proteins [16] with high stringency (two or more peptides) where most laboratories focussed exclusively on plasma proteins and contributed to one of the major aims, namely that of creating a plasma proteome inventory. The identification of smaller molecular weight protein "degradome" products and endogenous circulating peptides was seen as more of a degradative by-product, with only a few labs contributing to this aspect of the HUPO Plasma Proteome Project [17]. More recently, many proteomics studies of low molecular weight (LMW ≤10 kDa) blood constituents have aimed to identify novel biomarkers for a range of disorders [18]. Unfortunately, there has been only limited success in establishing reliable, reproducible and comprehensive extraction/separation procedures that allow analysis of both endogenous and disease-related peptides using global shotgun proteomic MS approaches [19]. This has led to continuous calls for improvements in analytical technologies for the identification of the LMW "degradome" and equally the endogenous peptidome [20–23] by both HUPO and the research community at large.

It has been more than a decade since researchers focussed on developing peptidomic technologies to isolate peptides (0.5–10 kDa) from human body fluids [24] (e.g., blood, urine and saliva), with a view to identifying potential diagnostic and disease biomarkers [25]. As modern MS technologies have excellent sensitivity and dynamic range, current routine proteomic approaches can successfully detect and identify several hundred plasma proteins in a single experiment. However, a surprisingly low success rate has been seen for the capture, isolation, measurement and identification of plasma peptides. Possible reasons for this blind-spot in our knowledge include; (i) the broad dynamic abundance range (i.e., over 10 orders of magnitude) between the least abundant (e.g., interleukins, cytokines etc.) and most abundant (e.g.,

albumin, IgG) protein constituents [26] occupying the peptide discovery space, (ii) the very low absolute levels of endogenous peptides compared to these proteins, (iii) large amounts of other endogenous compounds (e.g., fatty acids, amino acids, steroids, hormones), and (iv) the fact that peptides inherently bind to other high abundance proteins (HAPs) in human plasma in a manner that is difficult to disrupt [27].

An essential part of peptidomics sample preparation is the preservation of the integrity of the in vivo peptidome by preventing all types of proteolytic degradation during all separation and analytical procedures. Peptidase activity plays an important role in several phases of the peptide lifecycle, including the production, activation, inactivation and degradation of bioactive peptides [28]. Some of the peptide–peptidase pairings are well explained, however there are still a vast number of bioactive peptides whose in vivo regulation by peptidases is not well characterized [28]. Isolated peptide samples are traditionally frozen at a very low temperature as soon as possible and are often subsequently freeze–dried [29]. Unfortunately, such strategies do not always appear adequate (rapid enough) to prevent artifactual peptide degradation, and here it should be noted that several peptidases "survive" repeated freeze–thaw cycles [30]. Collectively, these issues present significant challenges during work-up of any peptide-containing complex biological sample [31].

The concept of peptidomics was launched when the composition of the peptide fraction in the human blood plasma database of circulating human peptides was published in 1999 [32]. The actual term peptidome/peptidomics was first "coined" simultaneously in three papers in 2001 [6,33,34]. Since 2001, the science of peptidomics has made slow and steady progress alongside other more rapidly evolving 'omics'. Peptidomics can be defined as the comprehensive, qualitative and quantitative study of all native peptides in a biofluid sample in a defined space at a defined time under defined biological conditions [35]. The peptidome also represents a collection of substrates and products derived from proteolysis (i.e., specific and/or non-specific) that are dynamically linked to the proteome of the same sample [35].

Several strategies for enrichment of peptidomes have been investigated in the search to establish a reproducible pipeline for the separation of comprehensive peptidomes from plasma [36,37]. Most are based upon finding a reproducible, cheap and efficient means by which one can remove the vast majority of protein from a biofluid without inadvertent removing the peptide components. These methods include protein removal by liquid–liquid extraction, protein precipitation (e.g., using acetonitrile [38], acetone [39] or methanol [40] to name but a few useful solvents), membrane-based molecular weight filtration cut-off methods (e.g., centrifugal ultrafiltration [3]), affinity or immunodepletion of a significant proportion of the protein complement [41], solid phase hydrophobic (C_8 or C_{18}) peptide extraction [42–44], on-column chromatography (i.e., ion exchange [45], reversed phase supports [46] or magnetic beads [47,48]), stable isotope labelling multiple reaction monitoring (MRM) or selected reaction monitoring (SRM)-based techniques [49], and 1D SDS-PAGE [50] or other electrophoretic methods [51]. Each of the methods has been developed bearing in mind the diverse physicochemical and structural characteristics exhibited by peptides (e.g., solubility, hydrophobicity, molecular weight, isoelectric point, protein binding and specific epitopes recognised by affinity chromatography).

Not surprisingly, each technique has been reported to have specific advantages and limitations. This review will discuss each of the major methods (i.e., solvent precipitation, centrifugal ultrafiltration, affinity-based immunodepletion, solid phase extraction electrophoresis and other proteomics-based techniques) attempted in the quest to achieve isolation of comprehensive human plasma peptidomes. Each method centres primarily on removal of a high percentage of abundant proteins that would allow enrichment of endogenous peptides, with a long term view of revealing these important potential biomarkers [52]. Fig. 1 lists the names for the most commonly used methods and their modifications along with attempted multidimensional combinations, whilst