

Molar Concentration Welcomes Avogadro in Postgenomic Analytics

Andrey V Lisitsa*, Elena A Ponomarenko, Olga I Kiseleva, Ekaterina V Poverennaya and Alexander I Archakov

Institute of Biomedical Chemistry, 119121, Pogodinskaya Street 10, Moscow, Russian Federation

Abstract

The researchers working with high-throughput methods of genomics, transcriptomics, and proteomics reconsider the concept of concentration and evaluate the data obtained in the number of copies of biomacromolecules. Measurement of copy number reflects a steady trend in increasing the sensitivity of postgenomic analytical methods, up to the level of a single molecule. In this paper we review the physical meaning of the terms "molar concentration" and "Avogadro's number" to establish a relationship between them. The relationship between the molar concentration and the number of copies of that same macromolecule in a certain volume is set through the reverse Avogadro's number, the value of which (10^{-24} M) characterizes the molar concentration of a single molecule in 1 liter. Using the reverse Avogadro's number, we deal with situations in analyzing homogeneous biological solutions and heterogeneous cellular material.

Keywords: Avogadro's number; Protein copy number; Blood plasma; Cell line

Introduction

At the end of the 18th century the physicist Amedeo Avogadro formulated the law [1], which subsequently allowed establishing one of the fundamental constants called Avogadro's number [2]. The Avogadro's number defined the number of molecules in one gram-molecule. The Avogadro's constant (N_A) is expressed in the unit mol^{-1} and used in the International System of Units (SI), instead of the dimensionless Avogadro's number, which counts for the number of molecules.

For a century the Avogadro's number and molar concentration existed in parallel. These two terms were rarely used together to characterize the number of molecules of any substance in a solution. In the early 21st century the situation changed due to the emergence of genomics and its derivatives – transcriptomics and proteomics. Postgenomic science aroused a necessity to characterize the concentration of DNA (for genome), RNA (for transcriptome), and proteins (for proteome) in the copies of molecules in a biological sample. Combining the concepts of the Avogadro's number and molecular concentration it was possible to revisit the notion of protein content in a solution (and cells) using number of molecules instead of chemistry-imposed concentration units [3]. The reverse Avogadro's number was introduced for recalculating the concentration of substances in a solution to the copy numbers, i.e. the determination of the number of entities of a certain biomacromolecule in a cell volume or in a biofluid, for example, in blood plasma.

In postgenomic biology copy numbers seem to be more important than concentrations, as genes and proteins could be definitely identified. Postgenomic technologies are used to investigate heterogeneous cellular material and more homogenous in composition of biological fluids. Each cell is a heterogeneous formation, to which the concept of volume is not applicable due to the existence of biological membranes. Membranes create the boundaries of the partitions inside a cell and impede free diffusion, which is a mandatory requirement to a solution [4,5].

The homogeneity of a biological fluid, at first glance, raises no doubts. However, in the case of blood plasma, taking into account the maximum soluble concentration of the plasma main component – albumin 10^{-3} M [6], the obvious question that arises is: how many other proteins can physically have room in the micro volumes loaded into analytical equipment? Analyzing the protein component of plasma it is necessary to consider, what the minimum volume of a sample is

that can typically represent all 4-5 liters of blood in human body. At the same time, the sensitivity of an analytical method should be taken into account, since the more sensitive method, the lesser amount of blood is required for an analysis. In genomics and transcriptomics the analytical sensitivity of technology is not so urgent due to the possibility of multiplying molecules by the PCR method. So, relatively to the proteins here we focus at the link of the copy numbers of biomolecules and the problem of heterogeneity in the sample.

Historical Background

The essentiality of concentration has been adopted by biochemists to define the amount of substance in a sample for over a hundred of years. At the beginning of the 19th century, biochemists described amounts per cubical centimeters or normalized them to the mass units, e.g. per grams of sample [7]. Molar concentration entered biochemistry more than half a century later, synchronizing the terminology with the standards established by organic chemistry [8]. Molar concentration is convenient because it simultaneously gives the amount of substance in relation to the volume in which the substance is dissolved.

At the dawn of, who deals with the dynamic nature of biological processes, taking place within living cells, widely used masses and percentages to characterize the containing of substances, but not molarity [7]. These units maintained hegemony in measurement system till 1960s, when molarity finally stepped forward [9-12]. The metamorphosis of the domination can be easily explained: the calculation of molarity requires volume and amount of substance of interest. Due to simplicity of composition of inorganic and elementary organic molecules calculation of molar weight and, as a consequence, molarity literally was not rocket science—in contradistinction from study of bulky and undeciphered biological molecules. With all reliable technics of fractionation and weighing, and even ambitious,

***Corresponding author:** Andrey V Lisitsa, Institute of Biomedical Chemistry, 119121, Pogodinskaya Street 10, Moscow, Russian Federation, Tel: (+7) 499 2463731; E-mail: lisitsa063@gmail.com

Received: October 30, 2015; **Accepted:** November 05, 2015; **Published:** November 07, 2015

Citation: Lisitsa AV, Ponomarenko EA, Kiseleva OI, Poverennaya EV, Archakov AI (2015) Molecular Sociology: Postgenome Science Upsets the Avogadro's Number. Biochem Anal Biochem 4: 216. doi:10.4172/2161-1009.1000216

Copyright: © 2015 Lisitsa AV et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

but unsuccessful efforts to invent a universal formula for all proteins [13], it took a great while to reveal the atomic composition of these biomolecules.

At the beginning of the 20th century Emil Fischer found that protein molecules contain long sequences of amino acids [14]. Works of Sanger (who determined the exact sequence in which the 51 amino acids in the molecule of insulin are linked together) and Anfinsen (who used paper chromatography to reveal the structure of ribonuclease) spawned a new phase in biochemical research [15,16]. Furthermore, well-known Lowry and Bradford assays for determining the total level of protein in solution were proposed in 1951 and 1976, respectively [17,18]. Evolving fractionation technics coupled with emerged approaches of protein sequence analysis allowed to estimate exact molecular weight, which is crucial for calculation of amount (v) and concentration of biomacromolecules in traditional for chemists molar units.

Reverse Avogadro's Number

What is the sense of using the reverse Avogadro's number instead of the canonical Avogadro's number (aka Avogadro's constant)? Reversing the Avogadro's number stipulates the transition from concentration units to the number of copies of biomolecules. Indeed, Avogadro's number is dimensionless and expressed in units of molecules, so it is incorrect to apply it in the formulas in the SI metrics. The Avogadro's constant is expressed in inverse moles (mol^{-1}), the units that do not have a physical meaning. At the same time, the reverse Avogadro's number characterizes the lowest concentration value having physical sense corresponding to one molecule per one liter [19]. Reverse Avogadro's number can be easily imagined as a single particular molecule floating in the one liter of the fluid.

The Avogadro's number, characterizing the number of molecules in 1 mole of gas, was calculated on the basis of Avogadro's law [1]. Avogadro's law applies to ideal gases. However, assumptions made by Avogadro extended to weak solutions, in which analyte concentration is substantially less than the concentration of solvent molecules. The logic of calculations using Avogadro's number (and, thus, the N_A constant) has been adopted in biochemistry in the first half of the 20th century. The initial restriction of Avogadro's law, namely that it only applies to ideal gases, was transferred to biological systems where intermolecular interactions are considerably less important than the kinetic energy of individual molecules [20].

Deviations from Avogadro's law could be said unimportant, but for ever-increasing sensitivity of postgenomic technologies. Using standard measuring instruments (spectrophotometer, fluorimeter, potentiometer, etc.) in a biochemical experiment the concentrations of the molecules analyzed is around 10^{-6} M. In this case, we deal with the stochastic distribution of the distances between molecules. The distribution, on one hand, reflects the concept of concentration, but on the other hand – is not truly consistent with Avogadro's law, since the intermolecular interactions are not taken into account in this law. Increasing the sensitivity by 10 orders of magnitude, up to 10^{-15} M and below, theoretically – up to 10^{-24} M, the concept of concentration completely loses its meaning. Instead of concentration, the counting of a number of molecules in a given sample becomes an issue. The approach to ultra-low concentrations is seemingly consistent with the concept of an ideal gas, however, it ignores the fact that molecules interact with the environment – either within a cell or within blood plasma. The vision of reverse Avogadro's number would be complete if a single certain biomolecule in one liter is seen in surrounding of billions of different molecules, which all together makes up what is called “biomaterial”.

We propose to combine the concepts of “concentration” and “copy number”. Using the reverse Avogadro's number the situation in a cell and in blood plasma is analyzed. Have to be comprehended, that the reverse Avogadro's number determines: if a gram-molecule of a compound contains 6.022×10^{23} of molecules (ions), then the mole fraction of each molecule in one gram mole of a substance is approximately 10^{-24} of molecules per mole. Since one mole, dissolved in one liter is a 1 M of concentration, then one molecule in the one liter corresponds to 10^{-24} moles/liter. This value has a meaning as a physical constant in the SI, since it represents the lowest unit of naturally achievable concentration of a substance in *one* liter.

The concept of concentration virtually characterizes the distances between molecules [21]. The smaller the distance between adjacent molecules, the greater the possibility of their interaction, and therefore, the more the deviation from Avogadro's law applicable for an ideal gas is expressed (Figure 1).

If the number of molecules decreases, then the physical regularity established by Avogadro is observed up to a certain limit: modern biochemistry is built up on this point. Considering the reverse Avogadro's number, we should proceed from the fact that the sensitivity of some instruments could be as low as one molecule per liter, because nanotechnologies approached these detection limits [3]. Within ultralow concentration range the Avogadro's law may be relevant only in a probabilistic formulation. In this case, it is not the volume, which is one mole of an ideal gas, but the volume, in which a single molecule can be detected with the specified probability, provided that we take a certain amount of the substance for analytical investigation.

Blood Plasma Homogeneity

Imagine myriads of individual molecules, comprising a concentration of plasma proteins – 10^{-3} M. It is unthinkable how such enormous number of individual protein molecules can fit into 1-10 μL microvolume selected for the analysis using highly sensitive postgenomic technologies. Is it possible for such a tiny volume of plasma (after a series of hard dilutions) to contain at least several copies of each protein present in the total volume of blood?

What could be the number of protein molecules that are present in the certain analytical volume, and what is the *minimum volume* of plasma that resembles the composition of the whole blood? It is assumed that both the high- and low-copy number proteins are evenly distributed over the volume, and that the distribution of proteins by their copy number follows the extreme value distribution at the log-scale [22].

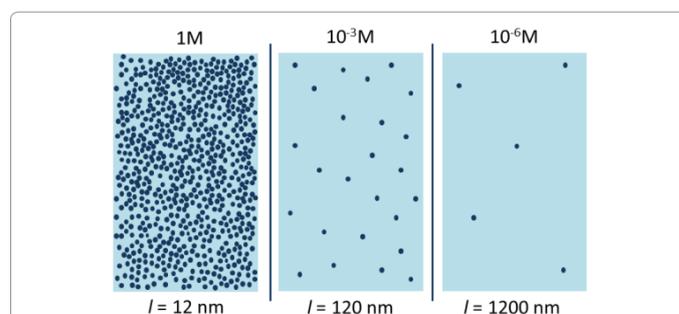


Figure 1: Concentration is measure of the distance between the molecules. Visualized from the data of Table 1 “Distance between molecules as a function of concentration” from ref [21]. Approximated using the equation $l = N^{-1/3} \times f(V/R^3)$, where l is the distance between the molecules, N is the number of molecules, V is the volume of solution, R is the radius of smooth sphere, that contains a protein, and function $f(x)$ has to be defined for the case if $V/R^3 \gg 1$.

First, let's estimate the maximum number of protein molecules that could be physically accommodated within a given volume. For the albumin, the Stokes radius is at least 3 nm [23], thus the volume would be ca. 10 nm^3 ($4/3\pi R^3$). A protein molecule has the same density as Plexiglas, almost incompressible under normal conditions [21]. Therefore, neglecting the ellipsoidal form of albumin, we will divide 1 μl of plasma (i.e. 1 mm^3) in the volume of a molecule – approximately 113 nm^3 . The result is that the spatial package allows placing 10^{17} of albumin molecules in 1 μl volume.

Thus, 1 μl can physically embrace up to 10^{17} average protein molecules. That is orders of magnitude greater than the estimate of the human proteome width (from 500 thousand to several million proteoforms taking into account post-translational modifications, splice variants and single amino acid polymorphisms [24,25]). Dividing the volume into the maximal number of proteoforms, we obtain that every proteoform has at least 10^3 copies of molecules (assuming, until proven different, that all the proteoforms are present in equal concentrations). However, the dynamic range of proteins in blood plasma is 10 orders of magnitude [6]. So it is impossible to settle all protein species into 1 μl , since high-copied proteins will statistically displace low-copied molecules out of the volume.

In addition to the above, the dynamic range is actually determined by the sensitivity of the analytical methods of protein detection. Therefore the presence of low-copied proteins can be judged only within the capabilities of analytical technologies. If the sensitivity limit of mass spectrometric methods equals to concentration of 10^{-18} M [26], then the dynamic range exceeds 15 orders of magnitude: from 10^{-3} M to 10^{-18} M . Applying the depletion methods of highly-abundant proteins, the range of concentration can be reduced by no more than two orders of magnitude [27], i.e. the ratio between the values of the highest-copied protein and the lowest-copied one will be 10^{13} . At the same time, the analysis of the human chromosome 18 protein products showed that the average concentration of one protein is 10^8 copies in 1 μl of plasma [28].

Analyzing the protein content of blood, it should be thought over the minimum sampled volume, which represents the entire of this tissue. So, the blood volume in the human body is about 4-5 liters, while just a tiny part of this amount is randomly picked up during blood sampling; even smaller subfraction is loaded into the analytical instrument.

Calculation of the minimum representative sample volume of plasma may be performed taking into account that concentration of one molecule in one liter is 10^{-24} M . In the volume six orders of magnitude smaller – 1 μl , the concentration of one molecule will be 10^{-18} M . This means that with a uniform molecule distribution within the volume, 1 μl of blood is representative for the proteins in a concentration of 10^{-18} M . At lower concentration, e.g., 10^{-19} M , the probability that 1 μl will contain a molecule is 1%, and so on, as the concentration of target analyte lowers down.

In cells, dynamic range of protein concentrations is believed to be significantly lower than in blood: deep proteome analysis resolved about 2 thousand proteins in plasma [29], versus 10-20 thousand gene products potentially observed in the cells [30,31]. At the same time, heterogeneity is observed both inside cells and between cells. This is an obstacle for the analysis of cellular biological material; the obstacle no less difficult than the problem of dynamic range in plasma proteomics.

Cell Heterogeneity

Cells are heterogeneous in their molecular composition, which means that each cell is to some extent different in the profile of expressed genes

and synthesized proteins from the adjacent one. For example, the level of mRNA expression of GAPDH, averaged for a pool of cells is shown not to reflect the situation with any single cell in the pool [32]. The degree of cell heterogeneity is unknown: we cannot generally say that hepatocytes differ in their molecular profiles, for example, on 0.1%, while neurons – on 10%. Moreover, it is unclear whether the cells differ only in quantity or different proteins that can be produced inside them [33].

In regard to the problem of heterogeneity of a cellular material, it should be taken into account that for the mass spectrometric analysis one normally takes not a single cell but plenty of them, in most cases, greater than 1 million. Indeed, high-throughput analysis of a single cell is not feasible; instead, we should apply the concept of an average cell. The average cell does not exist in reality; it reflects the uniform properties of cell population.

There are two sides in the problem of cell heterogeneity. The first was already mentioned above and refers to the sensitivity limit of mass spectrometer; the second is related to the volume of the investigated cells. The concentration of protein material in a cell is a function of the cell volume (Table 1). For a bacterium, the volume of which is 10^{-16} liters, the concentration of a single molecule will be $3.3 \times 10^{-8} \text{ M}$ – the detection limit achieved by ion trap mass detectors recently [19]. In the connective tissue, lymphocytic mass and in liver sample the concentrations of one molecule are 10^{-11} , 10^{-12} , and 10^{-13} M , respectively. The measurements of these concentrations are well-consistent with the detection limits of modern biological mass spectrometry.

The problem of cellular material analysis is that it is necessary to work with a sensitivity that allows detecting 1 protein copy not in a quantity but in a *particular* single cell. Indeed, if the detection limit of a method allows determining two or more copies of the protein per cell, then it is not clear whether these two (or more) molecules are in one particular cell or distributed across multiple cells. For a mass spectrometric analysis it is necessary to take a sample containing not a single cell but thousands, usually millions of cells to obtain reliable registration of peptide ions. For instance, the wide-spread Orbitrap™ mass detectors allow detecting a signal coming from at least 100-1000 molecular ions, in such cases analytical sensitivity required for molecular detection should be several orders of magnitude higher, that demonstrated in Table 1 [34].

Conclusion

Our vision is based on the rethinking the role of Avogadro's law, which was formulated about 200 years ago. For biomacromolecules this law became a regular fixture much later, after determining the number of molecules in one mole. For almost two centuries the "Avogadro's law" and "Avogadro's number" existed separately, linked just because of the name of the sagacious scientist and celebration of the "Moles Day".

Cell type	Cell volume, L	Concentration of a protein in a cell, M	Sensitivity**, M
Bacterium	5.0×10^{-16}	10^{-8}	10^{-14}
Connective tissue cell*	6.3×10^{-14}	2.7×10^{-10}	10^{-16}
Lymphocyte*	5.0×10^{-13}	3.3×10^{-11}	10^{-17}
Hepatocyte*	4.0×10^{-12}	4.2×10^{-12}	10^{-18}

*data is shown for human cells.

**analytical sensitivity, sufficient to detect 1 molecule in a sample, consisting of 1 million of cells.

Table 1: The analytical sensitivity of mass spectrometric method necessary to determine 1 copy of a protein per average cell. Sample contains 10^6 cells.

Avogadro considered an ideal gas, i.e. the environment where the energy of particle motion is much higher than the energy of interactions between them. Historically formed interpretation of Avogadro's law made Avogadro's number and Avogadro's constant equal, wherein the first one is measured in molecules and the second one is expressed in reverse concentrational units. Avogadro's number is applied to the objects with ordered structure – cells. Avogadro's constant is applied to blood plasma, in which the concentration of protein 10^{-3} M is extremely high. At such concentrations, the molecules do not form any weak solution or especially an ideal gas, since they start interacting among themselves. Proceeding from what has been set forth above, postgenomic technologies dealing with ultralow-abundant proteins and complex biological matrix should switch from measuring concentrations to counting the number of copies.

Thus, it is necessary to combine the common concept of molar concentration with the arrangements applying the reverse Avogadro's number – the connecting link between Avogadro's law and Avogadro's number essential for postgenomic research.

Acknowledgements

This work was funded by the Russian Science Foundation, grant no. 14-25-00132.

References

1. Avogadro A (1811) Essai d'une maniere de determiner les masses relatives des molecules elementaires des corps, et les proportions selon lesquelles elles entrent dans ces combinaisons. *J. Phys. Chim. D'histoire Nat. Des Arts* 73: 58-76.
2. Slabaugh WH (1965) Determination of Avogadro's number by Perrin's law. *J Chem Educ* 42: 471.
3. Archakov AI, Ivanov YD, Lisitsa AV, Zgoda VG (2007) AFM fishing nanotechnology is the way to reverse the Avogadro number in proteomics. *Proteomics* 7: 4-9.
4. SJ Singer, GL Nicolson (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175: 720-731.
5. Archakov A, Bachmanova G (1990) Cytochrome P-450 and active oxygen. Taylor & Francis, London, New York, Philadelphia.
6. Polanski M, Anderson NL (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomark. Insights* 1: 1-48.
7. Stewart GN (1899) Extracts of Removal of the Proteids. *J Physiol* 24 460-463.
8. McNaught AD, Wilkinson A (1997) Compendium of Chemical Terminology. The Gold Book (4th edn), Blackwell Science.
9. Macleod CM, Avery OT (1941) The Occurrence during Acute Infections of a Protein not Normally Present in the Blood: II. Isolation and Properties of the Reactive Protein. *J Exp Med* 73: 183-190.
10. Bueding E (1950) Carbohydrate metabolism of schistosoma mansoni. *J Gen Physiol* 33:475-95.
11. Deichmiller MP, Dixon FJ (1960) The metabolism of serum proteins in neonatal rabbits. *J Gen Physiol* 43: 1047-1059.
12. Sweet C (1970) The binding of serum albumin to phospholipid liposomes. *Biochim Biophys Acta - Biomembr* 219: 253-262
13. Mulder GJ (1839) Ueber die Zusammensetzung einiger thierischen Substanzen. *J Fur Prakt Chemie* 16: 129-152.
14. Fischer E (1906) Untersuchungen über Aminosäuren, Polypeptide und Proteine. *Berichte Der Dtsch Chem Gesellschaft* 39: 530-610
15. Sanger F (1949) The terminal peptides of insulin. *Biochem J* 45: 563-574
16. Katz AM, Dreyer WJ, Anfinsen CB (1959) Peptide Separation by Two-dimensional Chromatography and Electrophoresis. *J Biol Chem* 234: 2897-2900
17. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein Measurement with the Folin Phenol Reagent. *J Biol Chem* 193: 265-275
18. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.
19. Archakov A, Zgoda V, Kopylov A, Naryzhny S, Chernobrovkin A, et al. (2012) Chromosome-centric approach to overcoming bottlenecks in the Human Proteome Project. *Expert Rev Proteomics* 9: 667-676.
20. Bonetta L (2010) Protein-protein interactions: Interactome under construction. *Nature* 468: 851-854.
21. Erickson HP (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* 11: 32-51.
22. Bokes P, Singh A (2015) Protein copy number distributions for a self-regulating gene in the presence of decoy binding sites. *PLoS One* 10: 0120555.
23. Zhang X, Poniewierski A, Hou S, Sozański K, Wisniewska A, et al. (2015) Tracking structural transitions of bovine serum albumin in surfactant solutions by fluorescence correlation spectroscopy and fluorescence lifetime analysis. *Soft Matter* 11: 2512-2518
24. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1: 845-867.
25. Archakov A, Ivanov Y, Lisitsa A, Zgoda V (2009) Biospecific irreversible fishing coupled with atomic force microscopy for detection of extremely low-abundant proteins. *Proteomics* 9: 1326-1343.
26. Kopylov AT, Zgoda VG, Lisitsa AV, Archakov AI (2013) Combined use of irreversible binding and MRM technology for low- and ultralow copy-number protein detection and quantitation. *Proteomics* 13: 727-742.
27. Kiseleva O, Romashova Y, Moskaleva N, Petushkova N, Teryaeva N, et al. (2015) Plasma preparation to measure FDA-approved protein markers by selected reaction monitoring. *Clin Transl Med* 4: 32.
28. Ponomarenko EA, Kopylov AT, Lisitsa AV, Radko SP, Kiseleva YY, et al. (2014) Chromosome 18 Transcriptome of Liver Tissue and HepG2 Cells and Targeted Proteome Mapping in Depleted Plasma: Update 2013. *J Proteome Res* 13: 183-190.
29. Jones KA, Kim PD, Patel BB, Kelsen SG, Braverman A, et al. (2013) Immunodepletion plasma proteomics by TripleTOF 5600 and Orbitrap Elite/LTQ-Orbitrap Velos/Q Exactive mass spectrometers. *J Proteome Res* 12: 4351-4365.
30. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, et al (2014) A draft map of the human proteome. *Nature* 509: 575-581.
31. Wilhelm M, Schlegel J, Hahne H, Moghaddas Gholami A, Lieberenz M, et al. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509: 582-587.
32. Wang D, Bodovitz s (2010) Single cell analysis: the new frontier in "omics". *Trends Biotechnol* 28: 281-290.
33. Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, et al. (2014) From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res* 24: 496-510.
34. Scigelova M, Makarov A (2009) Advances in bioanalytical LC-MS using the OrbitrapTM mass analyzer. *Bioanalysis* 1: 741-754.