REVEALING MACROMOLECULES PRIMARY STRUCTURE BY LC: FROM POLYMERS TO BIOPOLYMERS

Alexander V. Gorshkov[1], Viktor V. Evreinov[1], Mikhail V. Gorshkov[2]

[1]N.N. Semenov Institute of Chemical Physics, Russian Academy of Sciences, Kosygina 4, 119991, Moscow, Russia
[2]Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, 38 Leninsky Pr., Bld.2, 119334, Moscow, Russia

ABSTRACT

Primary structure or sequence of monomers in a chain plays an important role in properties of macromolecules. Sequence of synthetic polymers has a statistical meaning, and these macromolecules are characterized by one - or multidimensional distribution functions - *MMD*, *FTD* or *CCD*. On the contrary, sequence of biopolymers has an absolute value. The structure and the way the proteins are functioning are governed by the unique order of amino acids (*AA*) in their chains. A deviation in sequence of a particular protein as a result of mutations, rearrangements of *AA* residues, or post-translational modifications is related to a wide scope of problems in the Life Science.

Liquid chromatography (LC) coupled with a mass spectrometer (MS) is commonly used to study macromolecules and their heterogeneities. In this combination LC brings its fundamental property as a method revealing the structure of a chain through the on-going adsorption process. Binding of monomers into a chain reduces translation entropy of their independent motion that results in collective, or phase like character of their adsorption. It also results in existence of so-called critical conditions under which the separation process becomes dependent on the macromolecule's sequence. Here, one deals with three modes of macromolecule separation, specifically adsorption (*LAC*), exclusion (*SEC*), and critical (*LCCC*) with the latter considered as a transition one between the first two and processing a number of unique features. The most profound one in case of synthetic polymers is that under *LCCC* the *MMD* disappears and various types of chain heterogeneity become visible. Important for biopolymers, *LCCC* region is characterized by a balance between entropy and energy changes in the thermodynamic function that depends on the biopolymer's sequence. The features of the biopolymer separation have been described by the authors recently as the *BioLCCC* model. One of the consequences of this model is that in case of a gradient elution the large enough macromolecules migrate along the column within the solvent composition range corresponding to the critical point of the macromolecule's adsorption. Therefore, the separation takes place in accordance with their adsorption critical points and the retention volume becomes dependent on the chain sequence: biopolymers having the same *AA* compositions but different *AA* order will have different retention volumes. The *BioLCCC* model allows both qualitatively and quantitatively describe this fundamental dependence of retention on the sequence.

In this presentation we describe the *BioLCCC* approach to extract information about the biopolymers' sequence from LC data. Specifically, we will focus the attention on the following questions: (1) the mechanism of biopolymers separation that results in sequence dependent retention volume; (2) prediction of the biopolymers' retention volumes based on their sequence; and (3) the utility of prediction algorithms for proteins and peptides identifications.