Protein's 14%

The proteins contain primitive and mysterious fragments, investigated by scientists from Gliwice and Warsaw

Protein is usually associated with egg white or with the description of nutritional value on food packaging. Fortunately, it's true, there are indeed proteins inside but of course they are not
a white mass of a size of the chicken's egg but small, usually invisible neck-chains. These chains consist of beads, whose chemical name is amino acids. Nature is able to use 20 types of amino acids of which it arranges various chains. In some proteins there is more of a particular amino acid or another one, but in general all 20 amino acids are in use in almost every protein on Earth. Of course, there would be nothing particularly interesting about it, if not for the fact that we do not really understand how these chains form a 3D structure, and then often we do not know how they work. A lot of work ahead for scientists. One of the amazing features of proteins are their special fragments (or fragments of these chains) that look completely different than other fragments. It turns out that there are regions of proteins in which beads are of only one kind. Then instead of the amazing diversity we have a fragment that looks like a boring chain consisting of exactly the same beads. There are also other "abnormal" fragments in which there are only 2 or 3 types of beads appearing consecutively in several or several dozens of repetitions. There are 14% of those in total. So instead of using all the possibilities, some specific protein uses only a few. Why? What is this "primitive" piece of protein for? In some cases, we know that these fragments, which we call low complexity regions (LCRs), decide about the attachment of this protein to another protein. We also know cases in which these fragments regulate some specific process in the cell. But there are millions of proteins and we only know a few dozen such cases. Other LCRs are shrouded in secrecy. Out of this curiosity this project was born. We want to understand what the other LCRs are doing. For this we must first find all LCRs in proteins, and then we want to put them in separate 'baskets'. In each 'basket' the most similar LCRs will be located. We already know from preliminary research that quite often there are very similar LCRs in proteins (chains) quite unlike each other. We assume that although they are in different proteins, they may have similar tasks. And that's what we want to exploit to understand LCRs better. As shown previously, we also are now counting on the fact that in each 'basket' with very similar LCRs there will be known functions of a few of them. In this way, we will be able to assume that others can play a similar role. When we accomplish that, we want to make our work available in the form of a publicly available database that you can look into. You will be able to look not only into the 'baskets', but also you could check the protein you are interested in, whether it has LCRs or not, and which 'basket' it is the most similar to. We invite you to have a look in four years.