English lessons for *E. coli*

Helene Klug*, Florian Christoph Sigloch

PolyQuant GmbH, Bad Abbach, Germany

The amino acid sequence of protein, written in the one letter code, seems a little like a word search puzzle. If you look at the sequence, from time to time you notice a word. Have you then also wondered if you can build a whole sentence using these amino acids? And if you can teach *E. coli* to express this sentence? The Young Proteomics Investigator Club (YPIC) wanted to address these questions, and PolyQuant stepped in to teach *E. coli* it's very first lesson in English.

Concerning the phrase to teach, there are limitations since not all letters of the English alphabet are reflected in the amino acid code. An entire sentence using only the one letter code should not include the letters B, J, O, U, X and Z. Short phrases like "Merry Christmas" are possible but it is much more difficult to write a whole sentence. Therefore, we decided to work around that problem by avoiding B, J, X, Z and K in the sentence and replace O and U by K. The latter complicated the final deciphering to some extent, but had the benefit to shorten the resulting tryptic peptides, which are easier to identify in first place.

The questions we wanted *E. coli* to ask were: "Have you ever wondered what the most fundamental limitations in life are? Is there a structure to respect when it comes to what you can produce in a cell?"

To enable our student *E. coli* asking the questions, we added a short peptide to initiate protein synthesis at the N-terminus (MAGR). However, if we perceive each protein as a word, *E. coli* is constantly babbling in its own language by expressing its native proteins. Among this gibberish, it would be hard to hear the questions. To be able to discriminate the questions from *E. coli*'s background babble, we added a protein "hashtag": a common His tag at the C-terminus (LAAALEHH-HHHH).

After these modifications the whole sentence looked like this:

MAGRHAVEYKKEVERWKNDEREDWHATTHEMKSTFKNDAMENT-ALLIMITATIKNSINLIFEAREISTHEREASTRKCTKRETKRESPECTWHENI-TCKMESTKWHATYKKCANPRKDKCEINACELLLAAALEHHHHHH

This "Protein English "was already a bit difficult to read but still understandable for humans. As a next step, we translated the amino acid sequence into a language *E. coli* is able to understand (DNA) and gave it to *E. coli* to read. To motivate our student, we wrapped our
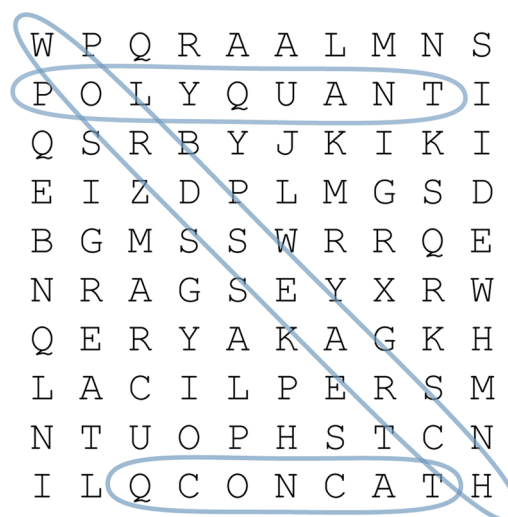

Fig. 1. PolyQuant word game, on paper here.

* Corresponding author at: PolyQuant GmbH, Bad Abbach, Germany.
  E-mail address: sigloch@polyquant.com (F.C. Sigloch).

questions in a bacterial expression vector.

Then we were hoping that our student would understand the lesson. And yes, it did: already before purification, we could see the results of our effort. All tested clones clearly expressed the questions as a protein of 16.7 kDa. To facilitate the deciphering for the YPIC members, we purified the expressed question via immobilized metal ion chromatography (IMAC).

In the end, we had achieved our goal: we had taught *E. coli* to ask questions in English. The proteomic word search puzzle was ready to be solved by the participants of the second YPIC challenge (Fig. 1).