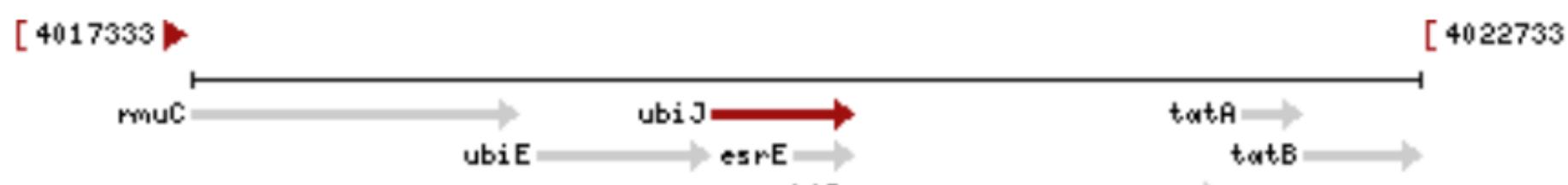


# Р.Раевский, М.Тутукина, С.Петрова, З.Червонцева ubiJ или не ubiJ?

## Введение

В аннотации даже таких хорошо изученных геномов, как *E.coli*, есть "белые пятна".



Показано, что в аэробных условиях синтез убихинона не происходит без этого участка, а в анаэробных - снижается, причём критично сохранение только последних 150 нуклеотидов. Одной группой показано, что в этом участке находится белок-кодирующий ген *ubiJ* [1], а другая группа показала наличие нкРНК, закодированной геном *esre* [2].

## Цели

- Определить, кодирует ли этот участок генома белок или нкРНК.
- Если это нкРНК - предсказать вторичную структуру.
- Если белок - предсказать доменную организацию и функцию.

## Выходы

- Это белок. Гена *esre* не существует
- Ген *ubiJ* кодирует цитоплазматический белок, состоящий из двух доменов: стерол-связывающий SCP2 и BMFP-подобного домена.

## Материалы и методы

- Поиск гомологов: BlastX, BlastP, PSI-Blast
- Поиск доменов: Pfam
- Выравнивание: AliView, ClustalX
- Структурное выравнивание: TM-Align
- Визуализация: PyMOL
- Подсчет Kn/Ks: CBU Kn/Ks Calculation tool
- Предсказание вторичных структур РНК: RNAfold, The UNAFold Web Server,
- Предсказание вторичной структуры белка: REPROFSeq, PSIPRED
- Предсказание трансмембранных доменов: TMHMM
- Предсказание сигнальных пептидов: SignalP
- Предсказание ДНК-связывающих доменов: GYM, 2ZIP, DBD HMM search
- Предсказание промоторов: PlatProm

## Результаты

1. В выравнивании гомологов гена *ubiJ* прослеживается большое количество синонимичных замен ( $K_n/K_s=0.02$ ).

Alignment of *ubiJ* homologs has many synonymous substitutions ( $K_a / K_s = 0.02$ ).

2. Белок предположительно цитоплазматический. Трансмембранных сегментов и сигнальных пептидов не обнаружено.

The protein is likely cytoplasmic. Transmembrane segments and signal peptides were not found.

3. Белок *UbiJ* состоит из двух доменов: стерол-связывающего SCP2 и BMFP-подобного домена (см. Рис.1)

*UbiJ* protein consists of two domains: the sterol-binding domain SCP2 and a BMFP-like domain (see Fig.1)

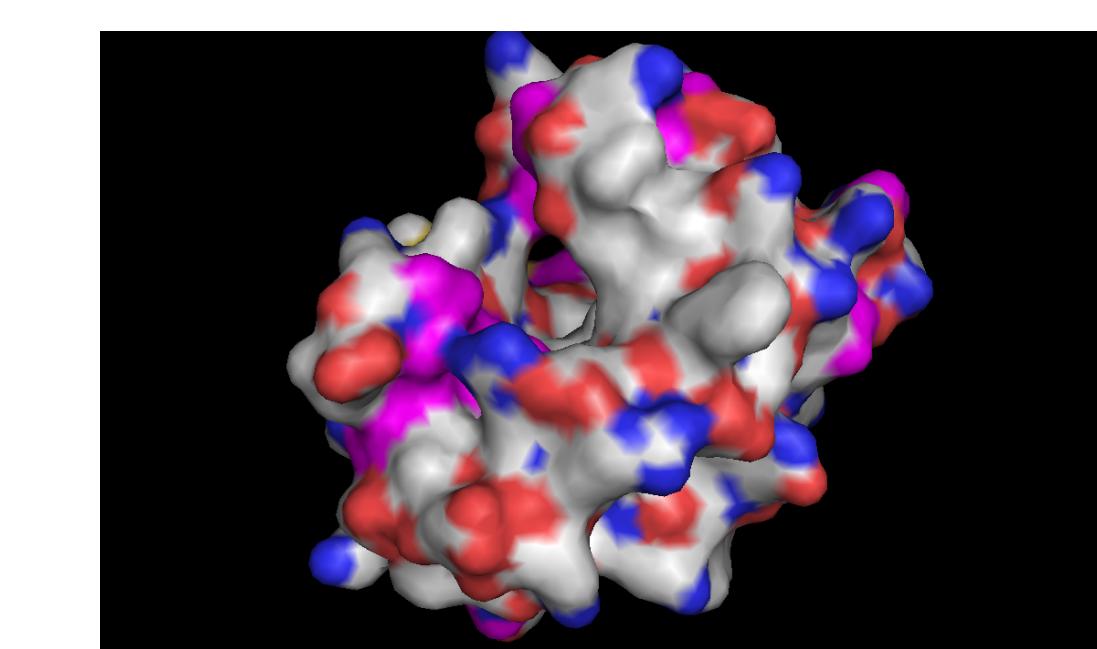
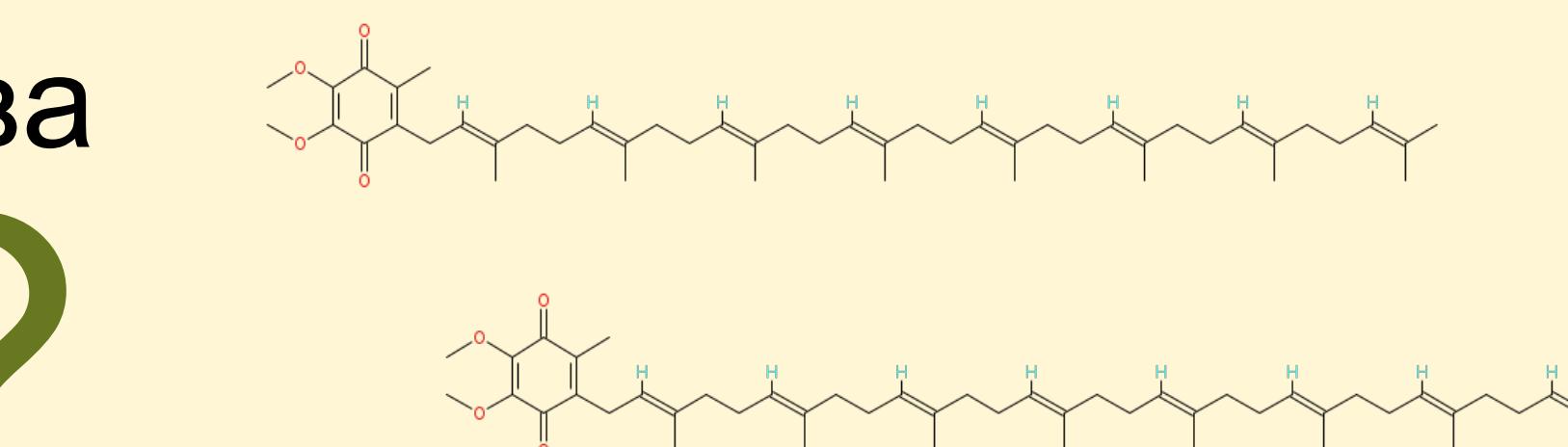


Рис.1. Структура белка.  
А. Предсказание вторичной структуры программой PSIPRED. В. Домен SCP2, консервативные аминокислоты отмечены розовым.

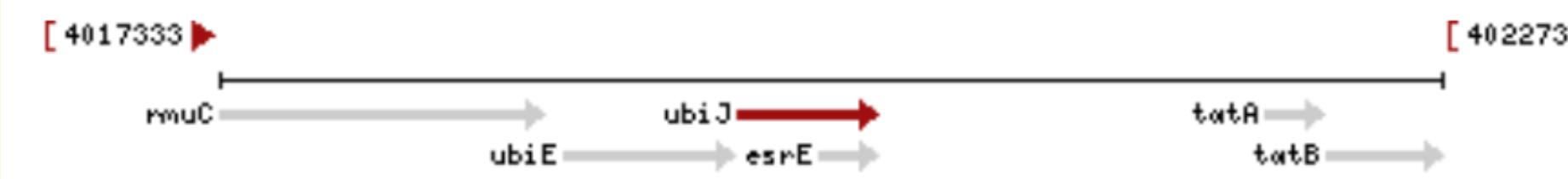
Fig.1. Protein structure.  
A. Secondary structure prediction by PSIPRED.  
B. SCP2 domain with conserved a.a. colored pink.

**Acknowledgements:**  
We are grateful to Peter Vlasov for consultations.



## Introduction

Annotation of the even well studied genomes, such as *E.coli*, contains "white spots".



The analyzed region is located within an operon between two genes responsible for the synthesis of ubiquinone. This site is known to be essential for the ubiquinone synthesis under aerobic conditions. Only the last 150 nucleotides are crucial. One group annotated this site as a protein-coding gene *ubiJ* [1], while another group annotated it as a sRNA-coding gene *esre* [2].

## Results

1. Предсказание вторичной структуры РНК *esre* из статьи [2] не воспроизводимо (Рис.2).

RNA secondary structure prediction of *esre* [2] is not reproducible (Fig.2).

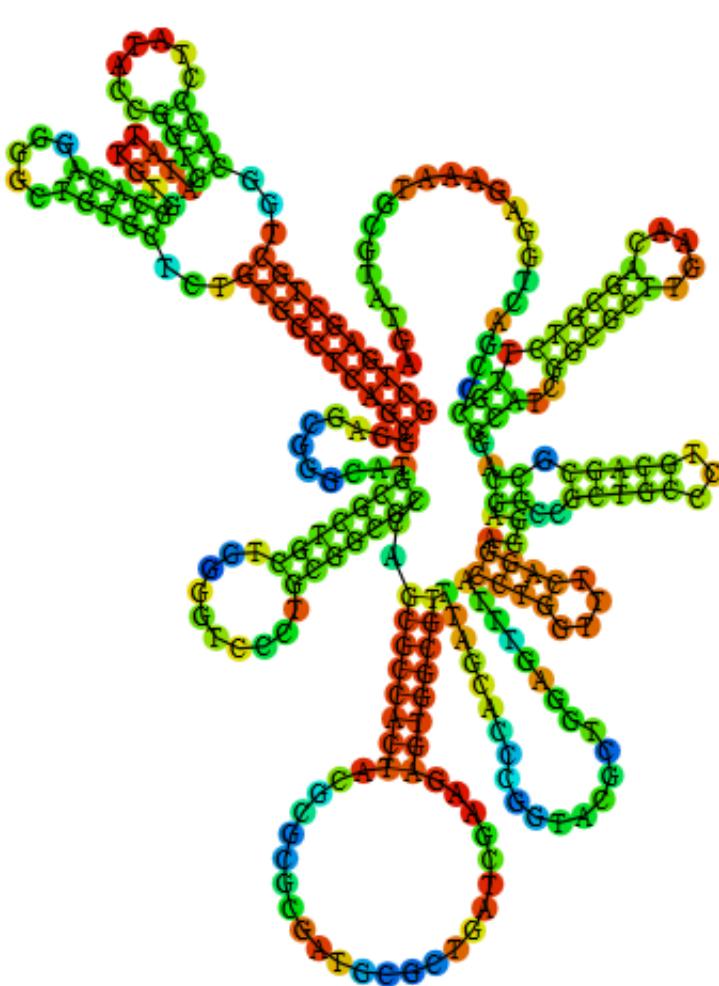


Рис.2. Вторичная структура РНК *esre*, предсказанные программой RNAFold.

Fig.1. RNA secondary structure of *esre* by RNAFold.

2. Промоторной последовательности перед геном *esre* не обнаружено.

There is no promoter sequence upstream the *esre* gene.

3. Паттерна комплементарных замен в множественном выравнивании гомологов *esre* не обнаружено.

There is no pattern of complementary substitutions in the multiple alignment of *esre* homologs.

4. Обнаружена ошибка в методике исследования [2] - при проведении RACE не был добавлен фермент TAP, обеспечивающий селективность.

A flaw in the crucial experiment of [2] was found: while carrying out RACE, the TAP enzyme providing selectivity was not added.

## Aims

- To determine whether this region encodes protein or sRNA.
- If it is sRNA, to predict the secondary structure
- If it is a protein, to predict the domain organization and function

## Conclusions

- It is a protein. The *esre* gene does not exist.
- The *ubiJ* gene encodes a cytoplasmic protein that contain two domains: a sterol-binding domain SCP2 and a coiled coil domain.

## Materials and methods

- Search for homologs: BlastX, BlastP, PSI-Blast
- Domain search: Pfam
- Alignment: AliView, ClustalX
- Structural alignment: TM-Align
- Visualization: PyMOL
- Ka/Ks calculation: CBU Ka/Ks Calculation tool
- Prediction of RNA secondary structure: RNAfold, The UNAFold Web Server,
- Prediction of protein secondary structure: REPROFSeq, PSIPRED
- Prediction of transmembrane domains: TMHMM
- Prediction of signal peptides: SignalP

## References:

1. Aussel L. et al. J Bacteriol. 2014;196(1):70-9.
2. Chen Z. et al. FEBS Lett. 2012;586(8):1195-200.