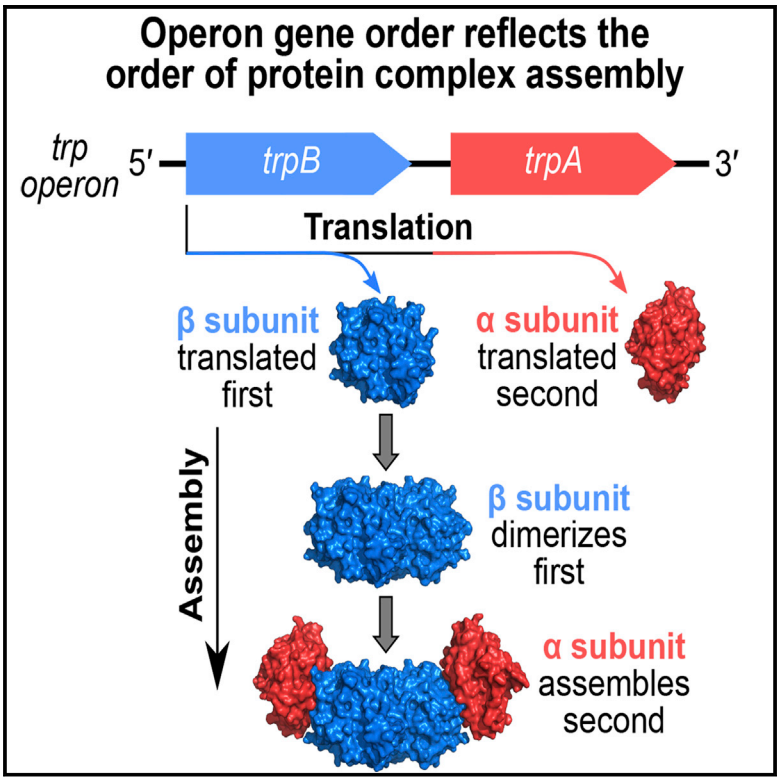
Cell Reports

Report

Operon Gene Order Is Optimized for Ordered Protein Complex Assembly

Graphical Abstract



Highlights

- Operon-encoded subunits tend to be encoded by neighboring genes and form large interfaces
- Operon gene order is often optimized for the order of protein complex assembly
- Exceptions are mostly highly expressed proteins for which assembly is less stochastic

Authors

Jonathan N. Wells,
L. Therese Bergendahl, Joseph A. Marsh

Correspondence

joseph.marsh@igmm.ed.ac.uk

In Brief

Many prokaryotic protein complexes are operon-encoded, so that subunits of the same complex will be translated from the same mRNA. Wells et al. show that the order in which genes are arranged in operons tends to be optimized for the order in which protein complex subunits assemble.

Wells et al., 2016, Cell Reports *14*, 679–685
February 2, 2016
http://dx.doi.org/10.1016/j.celrep.2015.12.085

CellPress

Cell Reports
Report

Operon Gene Order Is Optimized for Ordered Protein Complex Assembly

Jonathan N. Wells,[1] L. Therese Bergendahl,[1] and Joseph A. Marsh[1,*]
[1]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom
*Correspondence: joseph.marsh@igmm.ed.ac.uk
http://dx.doi.org/10.1016/j.celrep.2015.12.085

SUMMARY

The assembly of heteromeric protein complexes is an inherently stochastic process in which multiple genes are expressed separately into proteins, which must then somehow find each other within the cell. Here, we considered one of the ways by which prokaryotic organisms have attempted to maximize the efficiency of protein complex assembly: the organization of subunit-encoding genes into operons. Using structure-based assembly predictions, we show that operon gene order has been optimized to match the order in which protein subunits assemble. Exceptions to this are almost entirely highly expressed proteins for which assembly is less stochastic and for which precisely ordered translation offers less benefit. Overall, these results show that ordered protein complex assembly pathways are of significant biological importance and represent a major evolutionary constraint on operon gene organization.

INTRODUCTION

The assembly of proteins into complexes is integral to a wide range of biological processes. Although we now have extensive knowledge of the diverse quaternary structures formed by protein complexes (Goodsell and Olson, 2000; Janin et al., 2008; Marsh and Teichmann, 2015; Ahnert et al., 2015), much less is known about how they assemble and how assembly is regulated. In recent years, advances in electrospray mass spectrometry techniques have provided major new insights into in vitro assembly, allowing the assembly and disassembly pathways of protein complexes with diverse quaternary structure topologies to be elucidated in detail (Hernández and Robinson, 2007). In homomers, formed from the self-assembly of a single type of polypeptide chain, experimentally identified assembly intermediates often correspond to putative evolutionary precursors, so that the evolutionary history of a complex is reflected in its assembly pathway (Levy et al., 2008). Heteromers, formed from multiple distinct subunits, also tend to assemble and disassemble via ordered pathways that have a strong tendency to be evolutionarily conserved (Marsh et al., 2013). Although these experiments can be time-consuming, ordered assembly pathways can usually be predicted with very good accuracy from the known three-dimensional structure of a complex (Levy et al., 2008; Marsh et al., 2013). Given the many thousands of protein complex structures that are now available, this enables the study of assembly on a larger scale using computationally predicted assembly pathways.

Within the cell, assembly is much more complex and stochastic than in vitro, particularly in heteromers where multiple protein-coding genes must first be transcribed to mRNA and translated into protein, and those proteins must then find each other and assemble. Assembly is especially difficult for lowly expressed proteins, for which the stochastic variations in relative subunit concentrations are greater and the probability of interaction is lower (Kovács et al., 2009; Swain et al., 2002). How do cells cope with this? Does assembly within the cell follow similar ordered pathways as those observed in vitro and predicted computationally? Where does assembly occur within the cell? Has the regulation of gene expression been optimized for protein complex assembly order, as appears to be the case for the large multi-subunit bacterial flagella (Kalir et al., 2001)? Here we were able to address all of these questions by considering the relationship between protein complex assembly and gene organization in prokaryotic operons.

RESULTS

Operon-Encoding of Protein Complexes Is Likely to Enhance the Efficiency of Assembly

Many operons contain genes encoding different subunits of the same protein complex (Dandekar et al., 1998; Mushegian and Koonin, 1996) that can then be transcribed onto the same polycistronic mRNA. We first searched for heteromeric protein complexes of known structure from all prokaryotic organisms where at least two of the subunits are encoded by different genes from the same operon. In total, we identified 368 non-redundant pairs of subunits from the same heteromer encoded by different genes from the same operon (Figure 1A, left) from 70 different bacterial and archaeal species. This compares to 711 pairs encoded by different transcriptional units (*i.e.* translated from different mRNAs) from the same species (Figure 1A, right).

It has been suggested previously that a major advantage of operon-encoded complexes is their more efficient assembly because of smaller stochastic fluctuations in relative concentration than would occur if separate transcription steps were required for each subunit (Shieh et al., 2015; Sneppen et al.,