

**Differential Expression of Photosynthesis Genes in *R. capsulata*...**  
**Belasco JG et al (1985). Cell 40:171-181**  
A tour

**I. Why read this article?**

When you pick up a novel, usually it is without too many preconceptions or demands as to content. Maybe you were looking for a spy novel, but beyond that, you just want something that's good. This is the wrong way to go at a research article. Very, very few pass the test of great literature. If you're looking for a good read, you will be disappointed.

But your efforts may well be repaid if you go in hoping to answer a particular question. This different goal demands a different way of reading. With a novel, you would probably not start by skimming the first 100 pages in a couple of minutes, jumping to the middle, and when you find something interesting, leafing backwards to try to find out who are the main characters. But this is precisely the way to read a research article. In a novel, you rely on the author to tell a story. With a research article, you should generally keep control yourself, as your purposes in reading the article may be very different from the purposes of the authors in writing it.

You should have in your hands (or on the screen) the article by Belasco et al. Why? Well, most likely it's there because you trusted me and were following my suggestion to get it. That's not enough to get you through the article. You need a real reason. But since this is an exercise designed to help you learn how to read research articles, let me provide you with a suitable back story:

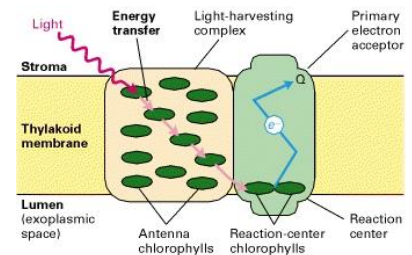
*You are studying plant biology and are struck by a remarkable feature of photosynthesis. You understand that photosynthesis requires that the cell capable of it absorb light energy and then channel the energy from a large collection of antenna chlorophyll molecules to a special chlorophyll molecule in a reaction center, where complicated electron-transfer reactions occur.*

*It makes sense that the antenna structure should be much larger than the reaction center, just as a satellite dish is much larger than the electronics at its center that eventually gets the signal. As you consider this, the following chain of reasoning occur to you:*

- Any cell that does photosynthesis should have both antenna and reaction centers, while any cell that does not should lack both.
- The genes that encode the proteins of the antenna and reaction centers must be coregulated: active in photosynthetic cells and inactive in other cells.
- But genes that encode the proteins of the antenna have to be much more active – producing much more protein – than the genes encoding the reaction center.

***What mechanism ensures that the two sets of genes are linked in when they are expressed but not in how much they are expressed?***

*You hit the literature to find an answer to that question and come up with Belasco et al.*



Transfer of electrons from light-harvesting complex (antenna) to reaction center. Lodish et al (2000). [Molecular Cell Biology](#).

**SQ1. Close your eyes and explain in words your mother could understand what question you are trying to answer and what led you to it. (If your mother happens to be a molecular biologist, choose someone else)**

## II. First pass through article

### II.A. Title and Abstract

The title and abstract of an article may be used to give you an idea if the article is worth spending time on. Don't expect them to give you much insight, however. Abstracts are generally written under severe word constraints. They're also often written for the computer, with an eye to putting in good words that will be found by search engines. There's generally not much space in which to be comprehensible. Of course the space constraints in a title are even worse.

**SQ2. What in the title leads you to think that the article might address your question?**

This title has very good words in it: "Differential expression of photosynthesis genes". That sounds exactly like what you're looking for. Differential expression... a difference in expression, yes, that's what you want. The rest may be gobbledy gook, but never mind. The first sentence of the abstract clinches it: "*We report that the light-harvesting and reaction center genes in [something] of R. capsulata are contained within [something] and that their differential expression results predominantly from [something].*" So long as "*R. capsulata*" is relevant to the question...

*As a student of plant biology and photosynthesis, you are quite familiar with the organisms in which photosynthesis has been studied, not only plants but also photosynthetic bacteria, most notably cyanobacteria and the purple bacteria within Rhodospseudomonas (also called Rhodobacter). The photosynthetic apparatus of the latter is somewhat different from that of plants, but it must solve the same problem of the expression of proteins related to the light harvesting apparatus and reaction center. [Here](#) is a link to an anatomically correct picture of the two reaction center (RC) proteins of Rhodospseudomonas surrounded by light-harvesting complex I (LH-I) and II (LH-II). You are delighted to have found an article about Rhodospseudomonas, because the mechanism of achieving proper gene regulation will almost surely be simpler than the one used by plants.*

### II.B. Introduction

Bla-bla-bla. All you're interested in is the answer to your question.

**SQ3. ...which was what again?**

The one thing pertinent I see in the first paragraph is a translation tool. You're familiar with terms like "light-harvesting complex" and "reaction center". They've connected those to the more specific terms they use. I've collected them (and others that appear in the next paragraph) in the table to the right.

The second paragraph is considerably more interesting.

**SQ4. What do you take away from the first sentence of the second paragraph ("*Mutations at the...*")? Don't get stopped by words you don't understand. Just replace**

| Synonyms                          |               |
|-----------------------------------|---------------|
| <u>Light harvesting complexes</u> |               |
| LH-I (two subunits)               | B870          |
| 1 <sup>st</sup> subunit           | B870 $\alpha$ |
| 2 <sup>nd</sup> subunit           | B870 $\beta$  |
| LH-II                             | B800          |
| <u>Reaction center</u>            |               |
| 1 <sup>st</sup> subunit           | L             |
| 2 <sup>nd</sup> subunit           | M             |

**them with thingie and move on. Don't replace B870 with thingie. You now know what that means.**

This is more and more looking like the right article! Evidently single mutations -- single changes in the DNA of the organism -- causes proteins in *both* the light harvesting complex and the reaction center to be lost. That simplifies matters! Coming into this, for all we knew genes encoding the light harvesting complex and those of the reaction center were on opposite sides of the genome and there were separate mechanisms to keep both sets expressed. But it sounds like there's just one mechanism controlling both. It's not at all clear what this *rxcA* locus thing is, but odds are we're going to find out soon.

**SQ5. What do you make of the second sentence (“A DNA fragment...”) ? What previously mentioned possibility does it seem to exclude?**

Hmmm. If “DNA fragment” just means “DNA molecule”, then this is a very uninteresting sentence. *R. capsulata*, like *Anabaena variabilis*, probably has only one chromosome, and so of *course* all those genes are on the same molecule. Where else could they be? But “fragment” in common use means a piece of DNA in a test tube. The routine processing of DNA shears DNA to pieces smaller than 50,000 nucleotides, i.e. ~2% of a bacterial genome. Saying that genes are on the same fragment says that they are in relatively close physical proximity to each other. If they're close together, that makes it more likely that their expression is regulated together. More good news!

**SQ6. What do you make of the third sentence (“The coordinate regulation...”) ?**

You may not be familiar with the term “operon”, and ordinarily, I'd advise bleeping it out and moving on. But that term is the whole point of the sentence. OK, ordinarily I'd advise bleeping out the whole sentence and moving on. But this paragraph has been so uplifting, saying exactly the sort of things I was looking for, that I would be emboldened to pay a trip to Google to find out what an operon is.

**SQ7. What is an operon? How does the concept of an operon add to this discussion?**

**SQ8. Rewrite the next sentence (“However, the B870 complexes...”) in a language you are more comfortable with. Does it correspond with what you already know about light harvesting complexes and reaction centers?**

And finally, the climactic last sentence of this paragraph...

**SQ9. What do you make of the last sentence (“If these photosynthesis gene products...”) ? Change the words to those that feel more comfortable.**

**SQ10. What does this sentence mean: “If these streams of water come in fact from the same hose (one with multiple ends, like a hydra), then control of the flows of water must occur at a level beyond the simple modulation of a single spigot.”**

**SQ11. Suppose that the control of gene expression for these photosynthesis genes is controlled exclusively by the initiation of transcription. How could that be squared with what we learned earlier in the paragraph?**

A well written Introduction typically ends in one of two ways: (A) [Storyteller mode] It presents the central mystery addressed in the article, as a logical conclusion of the reasoning and

observations that preceded it; or (B) [Lawyer mode] It presents the solution to the central mystery, claiming that it will be shown to be correct.

**SQ12. What can you get from the last paragraph of the Introduction?**

If your answer is “Not much”, that’s OK. I generally don’t expect too much from an Introduction and don’t spend much time worrying over it. But it’s useful to keep in mind the question delineated in the final paragraph. If the article is well written, everything that follows should in some way refer back to it.

II.C. Results

Understand that you are a finite resource. You need to pace yourself. If you try to understand every detail of every experiment, you will be carried away in a box. That’s not why you came to the article. You came to answer a question. If you stay in control, asking the article “What (if anything) can you tell me about my question”, then you can prioritize, focusing on areas closely related to your interests and chucking the rest (probably the majority of the article).

**SQ13. So it’s rather important to have your question firmly in mind. ...What was it again?**

(If you’ve forgotten, a trip back to the first page of this tour might be called for)

**SQ14. How would you restate that question, using the specifics of the genes described in this article?**

You want to do a preliminary interrogation of the article, to see how it can help you answer your question. I’d suggest making a quick survey of the Results section, especially the tables and figures, trying to the extent possible to identify the aim of each part. At this point, never mind the techniques. For example, I’d gravitate first to Figure 1. From the legend, I’d note that the figure relates messages (mRNA) to boxes (which from the labels are evidently genes). This is a figure I’m going to mark down as something worth understanding. I always feel more comfortable with a map, a graphical representation of the agents involved in the phenomenon. Where (besides the figure legend) is this figure described? A quick scan indicates that the first subsection (*Identification of the rxcA message...*) of the Results mentions Figure 1.

What else does this first subsection talk about. Skimming the beginnings of the paragraphs, I see that it is interested in where “...both messages originate...” and whether “...the two messages... share a common 5’ terminus.”)

**SQ15. What is meant by 5’ terminus? Hint: choose between “beginning” and “end”. Why?**

**SQ16. Outline the article (without spending too much time figuring it out), giving perhaps one sentence per subsection, identifying which subsection discusses which figures, and giving a summary (in general terms, if possible) what each figure might be talking about.**

**SQ17. Which sections and figures do you think relate closest to your motivating question?**

**SQ18. Sorry to be repetitive, but what was your question again? (Can’t ask this too often)**

II.D. Discussion (and Experimental Procedures)

I generally don’t waste much time on the Discussion section at this point and none at all on the Experimental Procedures. Save your strength for when it’s needed.

### III. A plan

Since we're doing this together, and since you're not here at the moment (my moment), I have to make some unilateral decisions. It seems to me that the second section ("*Determination of the Relative Steady State Concentrations of the rxcA Messages*") is the section of the Results most pertinent to our question. I read provocative things that suggest that there are multiple messages and that they are found in very different amounts.

**SQ19. What in this section would lead me to think so?**

**SQ20. What genes does each of the multiple message cover? The proteins encoded by these genes encode which photosynthetic structures?**

The section refers to two experiments. One seems to be a crude preliminary experiment while the second experiment gives the ultimate answer.

**SQ21. What specific question does the second experiment answer? Which figure shows the results of each of the two experiments?**

OK. I've decided that if I have to go through the trouble of figuring out an experiment, it might as well be the more informative one, so I'm (we're) going to focus on the second experiment – how was it done and what the results shown in Fig. 2 (right) mean, and along the way, we'll also need to figure out Figure 1, at least enough to understand the second experiment. The first section ("*Identification of the rxcA Messages...*") is also interesting – I'd like to know why they think there are multiple mRNAs covering this region – but I think the different *abundance* of the messages is more critical to our question.

**SQ22. Why is the abundance of these messages pertinent to your central question?**

The determination of termini ("*3'-End Mapping...*" and Figure 3) I'm going to declare to be a detail of lesser importance to me. Maybe the hairpin structures shown in Figure 4 are also a detail, but hey – it's a sequence! -- and I have a soft spot for connections between DNA sequences and function. So I'm going to try to understand what that sequence has to do with regulation of gene expression. The rest of the paper I declare to be for some other time, not something to concern us right now.

*Note that I have ruthlessly dismissed maybe two-thirds of the results in this article, despite all the work the authors put into writing it. This is well within my rights. I am not in the service of the article. Rather, the article exists to serve me, and I take from it what I can and what I want. I spare myself the effort of learning about plasmid construction, pulse-chase experiments, and  $\beta$ -galactosidase assays – all techniques extraneous to the experiment I've defined as of interest to me. You must develop a sense of what is important to you and what is not, otherwise you'll get bogged down in the first article you pick up.*

While I've thrown out a good deal of this article, it is imperative that I understand as well as I can the part I've retained. In particular, it is important that I understand how the results shown in Figure 2 (right) were obtained and what they imply. It is impossible to understand the implications of an experiment without understanding how the experiment was done, as the conditions of any experiment limits conclusions that can be drawn from it.

*I will venture the opinion that the separation of conclusions from the experiments or observations on which they are based is one of the primary sources of nonsense that infects science and society..*

We need to identify what is necessary to understand in order to make sense out of the chosen experiment.

**SQ23. Reconstruct the steps required to produce the picture shown in Figure 2 (right). If there are steps you don't understand (and I'd be amazed if there weren't), note these.**

No doubt one or more steps you identified relates to "S1 analysis" or "S1 protection experiments". Now, for the first time, the moment has arrived to look at the **Experimental Procedure** section. Find the section on S1 protection experiments. Understand that the section was written for those already familiar with the basics of the technique who wants to reproduce the results in their own labs. It wasn't written for you. Nonetheless, read the section just to get an idea of the elements of the technique.

**SQ24. But what does "S1 protection" mean? Look around on the web (Wikipedia is often a helpful initial resource).**

**SQ25. Integrate what you found about S1 protection with details in the legend to Figure 2, Figure 1, and the Experimental Procedures section.**

**SQ26. Identify what you deem to be important concepts that still don't make sense regarding the experiment whose results are shown in Figure 2.**

*(Tour to be continued)*