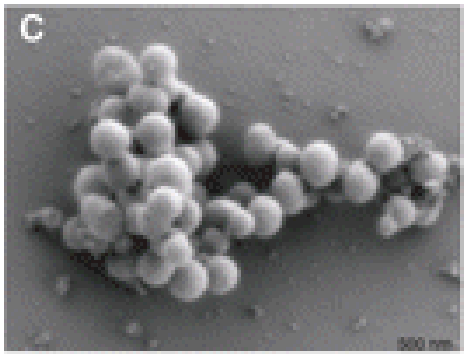
BNFO301: Introduction to Bioinformatics

Companion to Fuller RS et al (1984): The dnaA protein complex with... oriC... Cell 38:889-900.

In 2010, Gibson et al (2010)[1] reported that they had created a cell that reproduced under the direction of completely synthesized DNA (summarized by Pennisi, 2010[2]). While the synthetic genome was 99.99% identical to that of a naturally occurring bacterium (except for a few sequences added for positive identification), one can readily imagine more radical uses of the technology, e.g. the creation of artificial algae specialized for biofuel production[3] or of bacteria that use an alternative genetic code and non-standard amino acids, expanding the range of possible proteins.[4]



Electron micrograph of cells created from the modified genome of *Mycoplasma mycoides* inserted into cells of *Mycoplasma capricolum* (ref. 1).

The primary limit on useful artificial cells is not the technology – how to make artificial cells -- but the biology – knowing what changes to make in natural genomes.[5] How do the hundreds of proteins that are essential for the viability of even the smallest bacterium work together over the course of the cell cycle and under different environmental conditions? In some sense, the synthesis of proteins is the most understood element of the system. Our knowledge of how information goes from DNA to protein is relatively mature.

We know far less about information in DNA that is not directly related to encoding proteins. The regulatory systems embedded in DNA that control when genes are transcribed are quite imperfectly understood. There are other DNA functions as well. One is the information in DNA that determines where and ultimately when the DNA is replicated. Since DNA replication is essential for the viability of any cell – natural or artificial – it is important to understand what DNA sequences are required for it to take place in an orderly fashion. Fuller et al (1984) sought to elucidate the role of certain 9-basepair DNA sequences in the cell's decision to initiate DNA replication.

While you won't be making artificial cells (any time soon), you ***will*** be looking for small DNA sequences that might be important in the physiology of the cell. You're no doubt comfortable with the idea that a gene's-worth of DNA – a few hundred nucleotides – can determine the structure of a protein, but how can 9-basepairs (just three codons!) do anything useful? Learning about this particular small DNA sequence may increase your repertoire of examples and give you a better idea how to find new ones.

With that in mind, now would be a good time to get the full text of Fuller et al (1984), so you can see what they found and how they found it.

---

[1] Gibson DG, et al (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52-56.
[2] Pennisi E (2010). Synthetic genome brings new life to bacterium. Science 328:958-959.
[3] Varughese A (2011). Venter's new hope: Synthetic algae for biofuels. BioTechniques 11/07/2011
[4] Lajoie, MJ (2013). Genomically recoded organisms expand biological functions. Science 342:357-360.
[5] Baker M (2011). Synthetic genomes: The next step for the synthetic genome. Nature 473:403-408.

[An aside to those new to reading research articles...

- *Everyone is confused by research articles. If you are too, that's a good sign.*
- *However, research articles are as close as you'll come to scientific truth, except by doing experiments yourself. So they're worth it.*
- *They're worth it so long as you don't try to get too much. You are under no obligation to understand everything in an article. In fact, trying to do so is generally a bad idea.*
- *It's better to go in with a goal, find the small fraction of the article that addresses your goal, squeeze that fraction for what its worth, then get out.]*

SQ1. What is your goal in reading Fuller et al (1984)? (If you don't know, try the last couple of paragraphs on the previous page)

SQ2. How might you go about finding the fraction of Fuller et al (1984) that addresses your goal? What might you do to avoid burning out in the process?

I would skim the **Summary**. Often they are impenetrable with jargon. Those summaries I would skim very fast and forget about them. Summaries are sometimes frustrating, because they don't make clear what was already known and what is the contribution of the work presented in the article. **Introductions** are often more comprehensible. I would skim this section as well, looking for something that is immediately interesting (and not worrying about things that are not comprehensible) and for anything that might help me understand what was already known by the authors concerning my area of prime interest – the nature and function of the 9-basepair sequence. What implicates it in the regulation of DNA replication? How might it work?

Let's pause for a moment, considering just the **Summary** and the **Introduction.** One thing that pops up immediately from both sections is that the authors think quite highly of something called DnaA protein.

SQ3. What characteristics of DnaA protein, already known at the time of the study, point to its importance in DNA replication and to how it works? Don't worry about what you don't understand. Focus on what you can understand.

SQ4. The word "oriC" appears over a hundred times in this article. We're clearly not going to make much headway until we understand what it is. What is it? (Don't just replace oriC with a different term. Describe a picture that you build in your mind related to what this thing is and does.) (Don't have a picture? Go out and get one.)

SQ5. Then there's the 9-basepair sequence. What is the sequence? If you don't understand how it's written,... wait. What picture do you have about what was known about this sequence prior to the work presented in the article (i.e. prior to 1984)?

SQ6. What evidence is there that the 9-basepair sequence has anything to do with DnaA? With DNA replication?

Pretty weak, actually. Suppose that Sherlock Holmes observed that the last four murders in London were committed by someone with a scar. Does that mean it was the same person? Maybe not. Maybe people who commit murders tend to get scars.

That's what confronted Fuller et al. Can they find more compelling evidence to link the 9-basepair sequence to DnaA protein and therefore to DNA replication?

Results section – first pass

If there's any truth to be had from this article, it will be in the **Results** section. That section, however, looks pretty imposing! I'm pretty sure I don't want to wade through 13 complicated figures plus two tables. Therefore, I propose we skim the **Results** section, gaining no more than a broad view of what's there and noting which subsections might be most pertinent to our goal.

SQ7. And what is our goal?

DnaA Protein Binds to oriC and Other Sites Containing the 9 bp Sequence

The title certainly sounds pertinent, no? The first paragraph, however, quickly gets into experimental minutiae that may leave you far behind. Never mind, for now. I notice that it has to do with oriC, and that's good enough. I count this paragraph as Potentially Important. Maybe the next paragraph as well. From that point on, the section loses its connection to oriC and focuses on plasmids called pPM28, pBF120, pBR322,.... I think I'm going to focus for now on what is more comprehensible -- the *E. coli* origin of replication, oriC -- and let the rest go.

Footprint of dnaA protein at oriC

oriC, DnaA protein, 9-bp sequence,... this paragraph has it all. But it's also pretty much goo, and the paragraph ends with the statement that the authors find it difficult to attribute the results to a specific interaction of DnaA with the 9-bp sequence. With that in mind, I'm going to mark this Read Only If Necessary.

Footprint of dnaA protein at sites containing one copy of the 9 bp sequence

No oriC here, so no read.

Electron microscopy of dnaA protein-DNA complexes

This paragraph is tied to Figure 12, which is totally different in nature from any other figure in the article. I gather that they tried a different approach to assess binding of DnaA protein to DNA. If the first approach is understandable to me, then I might be inclined to skip wading through the second. Not sure.

- - - - - -

In brief, I come up with one paragraph, maybe two I need to read. If that doesn't work out, I may go on to the first footprinting subsection or the electron microscopy subsection. Otherwise, I might attain satisfaction with the first paragraph and stop there.

Discussion section

A well written **Discussion** section can help you navigate the results. A poorly written **Discussion** section merely shovels more confusion on the heap. I often don't bother even skimming the **Discussion** section until I've tried my hand understanding at least one result.

Experimental Procedures section

No way I'm even going to touch this section, unless I have to!

Results section – second pass

I've determined through skimming the article that I may find what I'm looking for in the first paragraph of the **Results** section. I therefore will try to understand thoroughly the experiment described there, but anything else in the article I'm free to ignore. Even without understanding that paragraph, it's possible to organize it to see what task lay ahead, and that's the next step.

SQ8. Look over the first paragraph and decide (a) what part talks about experimental protocol, (b) what part talks about primary results, and (c) what part provides some analysis of those results.

So long as your confusion shields are on high, this isn't so bad. The first part of the paragraph (up to "(Experimental Procedures; Figure 2A)" is clearly experimental protocol, even though I may not understand a word of it. Unfortunately, I'm going to have to figure much of this out, but not now.

The next part of the paragraph (starting with "Only three fragments...") seems to be talking about results, but where are they? They claim all sorts of things here, but where's the evidence? "...three fragments..."?

SQ9. Is there some table or figure that identifies the three fragments the authors mention here?

They just referred to Figure 2A. Anything there that connects to the description of the three fragments in the text? Reflecting on this, I'm convinced that I'm going to have to understand what Figure 2A is about.

The rest of the paragraph concerns Table 1 and Figure 2B. Table 1 compares the 9-bp sequence found in oriC with other sequences. That's interesting but off topic at the moment. So forget it. Figure 2B I may have to look into.

So at the top of my list is to understand Figure 2A and the experiment that led to it. Then maybe I go to Figure 2B, maybe not.

Results section – third pass

Well, can't put it off any longer. Now comes the hard part – understanding the experiment. However, the process we went through has the beneficial outcome of focusing our attention on just two sentences of a 12-page article. I suppose that's progress.

SQ10. What do you make of the first phrase of the first sentence of the first paragraph of the Results (up to "...other sequences")? "Specific binding" of what? (contextual clues point to what?). Is the goal stated by the author one that overlaps with your own goal for your reading of the article?

The next phrase begins "M13*oriC26* RFI DNA". I guess we need to find out what that is. The usual places to look are (1) the first mention of the term, and (2) the **Experimental Procedure**. Using Adobe Reader's search facility, I find the first mention... no help there (except a reference to Fuller and Kornberg, 1983, which I note in case all fails and I have to look it up).

SQ11. What does the Experimental Procedure have to say about M13*oriC26*?

All seems to have failed. The article gives me no help at all to understand this... entity... so I've no better choice than to start looking up articles they reference. I'll save you the trip. In brief,

Fuller et al (1984) references Fuller and Kornberg (1983),[6] which also says nothing useful about M13*oriC26* but instead references Kaguni et al (1981),[7] which also says nothing useful about M13*oriC26* but instead references three articles, one of which is Kaguni et al (1979),[8] which ***finally*** describes what the thing is! (Please think about this painful journey when you are tempted to provide a merely convenient reference rather than a reference that really supports the statement it's attached to)

Here's what M13*oriC26* RFI turns out to be. M13 is a bacteriophage (a bacterial virus) that can infect *E. coli* as a single strand of DNA. Inside the cell, it synthesizes a second strand. The resulting double-stranded, circular DNA molecule is now able to replicate and is therefore known as the replicative form, or "RF". M13 therefore replicates like a plasmid, a small circle of DNA.

Kaguni et al (1979) cloned into a M13 derivative a 3.5 KB (kilobasepair) DNA fragment containing *oriC* from *E. coli*. The fragment was cloned into a *XhoI* restriction site on the phage. You can see the result in Figure 1.

SQ12. From Fig. 1, what symbolizes the DNA derived from the *E. coli* chromosome? What is the significance of the X that flanks that region? Using the scale at the top of the figure, how big is the chromosomal fragment? What symbolizes the DNA derived from phage M13?

SQ13. What is the significance of the triangles sitting under the map of M13oriC26?

Back to the first paragraph of the **Results**. Still on the first sentence.

SQ14. What does "restriction endonuclease mean? If you don't know, find out.

SQ15. What does "restriction endonuclease TaqI" mean? If you don't know, find out. Google is your friend.

SQ16. Why were 20 fragments generated? (Confused? Click here for help)

We're told that the fragments were labeled in some way. I'm content to trust that the DNA polymerase was able to work some magic to accomplish this.

SQ16. What's the significance of ^{32}P? If you don't know, find out.

We've reached the end of the first sentence. Thank the gods that we have only two! The second sentence has two issues: (a) "...fragments that bound to or flowed through nitrocellulose filters..." and (b) polyacrylamide gel electrophoresis. "Fragments" clearly refer to the fragments mentioned in the previous sentence. Evidently when the DNA fragments produced by digestion by restriction enzyme TaqI are incubated with DnaA protein, they either can bind to a nitrocellulose filter (whatever that may be) or flow through it. No more clues here, except we're pointed to the dread **Experimental Procedures** section. No use trying to avoid it.

But where? Which of the five subsections of the **Experimental Procedures** section will tell us about the nitrocellulose filter experiment? None of the titles of the subsections are obvious tells. I

---

[6] Fuller RS, Kornberg (1983). Purified dnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. Porc Natl Acad Sci USA 80:5817-5821.

[7] Kaguni LS, Kaguni JM, Ray DS (1981). Replication of M13 *oriC* bacteriophages in an *Escherichia coli rep* mutant is dependent on the cloned *Escherichia coli* replication origin. J Bacteriol 145:974-979.

[8] Kaguni J, LaVerne LS, Ray DS (1979). Cloning and expression of the Escherichia coli replication origin in a single-stranded DNA phage. Proc Natl Acad Sci USA 76:6250-6254.

guess the thing to do is to read a little (as little as possible) of each subsection to see if anything sounds familiar.

SQ17. Which subsection of the Experimental Procedures section gives the protocol for the experiment involving nitrocellulose filters?

So long as you're not put off by the mindnumbing details, it's not too difficult to find. And about those details... there is absolutely no point in spending more than a microsecond on things like the concentration of Tris-HCl or the temperature of incubation or the lists and lists of chemicals. They're provided for those who are trying to ***replicate*** the experiment in the lab (and are highly appreciated by them). For those, like you, who merely want to understand the principle behind the experiment, they should be put behind you as fast as possible. For example, I would read the first paragraph like this:

> *DNA to be analyzed was digested with TaqI. Ends were labeled (^{32}P).*

That's it.

SQ18. Try translating the next paragraph.

SQ19. Where does the "nitrocellulose filter" come into the picture?

Maybe when we look at the results it will become clear which DNA binds to the filter and which flows through it.

There's still one piece of work to be done to understand the experiment: polyacrylamide gel electrophoresis. What's that? There are many resources on the web that can help you understand the principles behind gel electrophoresis. This video[9] (the first part) is one of them (though you should bear in mind that agarose is not polyacrylamide and fluorescence is not ^{32}P).

SQ20. Take a look at Figure 2A and form a theory as to what you're looking at. What does TD, B, and FT mean? Why is the lanes below TD and FT full of black blobs (bands) while the lanes below B largely empty? What distinguishes the lanes below "0" from the lanes below "220"?

SQ21. Comparing the lanes of Figure 2A for yourself, what do you see that changes? What do you think those changes mean? Then (and only then) compare your observations with those of the authors in the first paragraph of the Results.

SQ22. Which fragments shown in Figure 1 correspond to the bands visible in the lanes marked B in Figure 2A? Why do you think just those?

I think I'm just about satisfied. There is certainly more to be gained from this article, but I think I got what I came for.

You might want a more global, visual picture of the connection between DnaA protein, the 9-basepair sites, and the initiation of DNA replication. There are many review articles that cover this question. The newer ones are certainly informative, but reality has turned out to be rather complicated. If you want a picture that is much simpler (though still basically in agreement with what we know now, you might take a look at Bramhill and Kornberg (1988).[10]

---

[9] Sadava et al (2010). Separating fragments of DNA by gel electrophoresis. From *Life: The Science of Biology*, Ninth Edition, Sinauer Associates. http://www.sumanasinc.com/webcontent/animations/content/gelelectrophoresis.html

[10] Bramhill D and Kornberg A (1988). A Model for Initiation at Origins of DNA Replication. Cell 54:915-918.