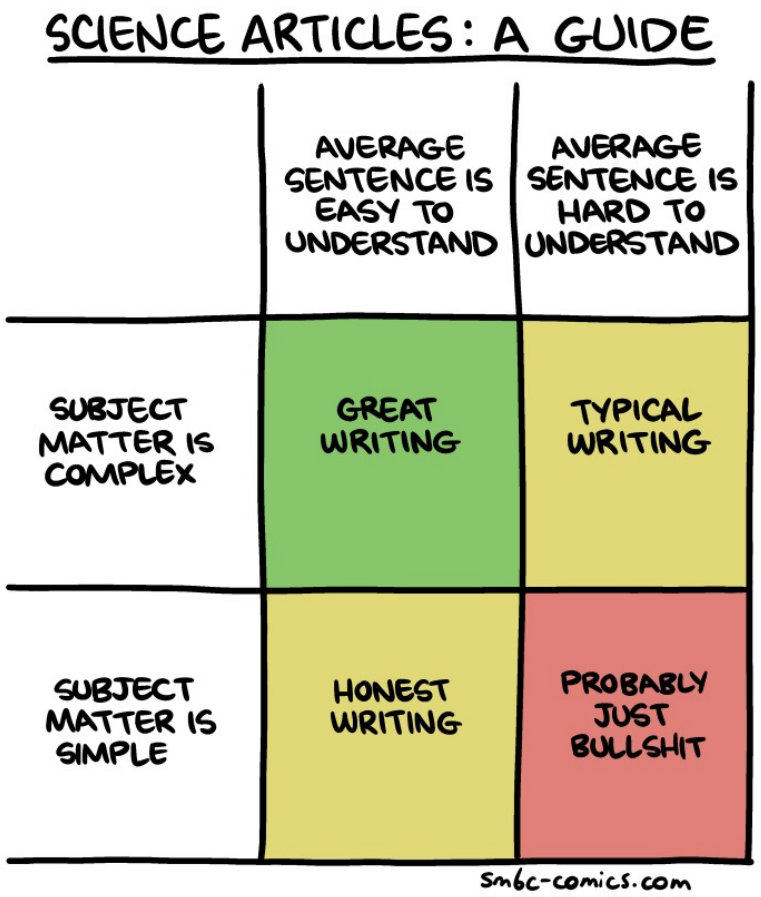
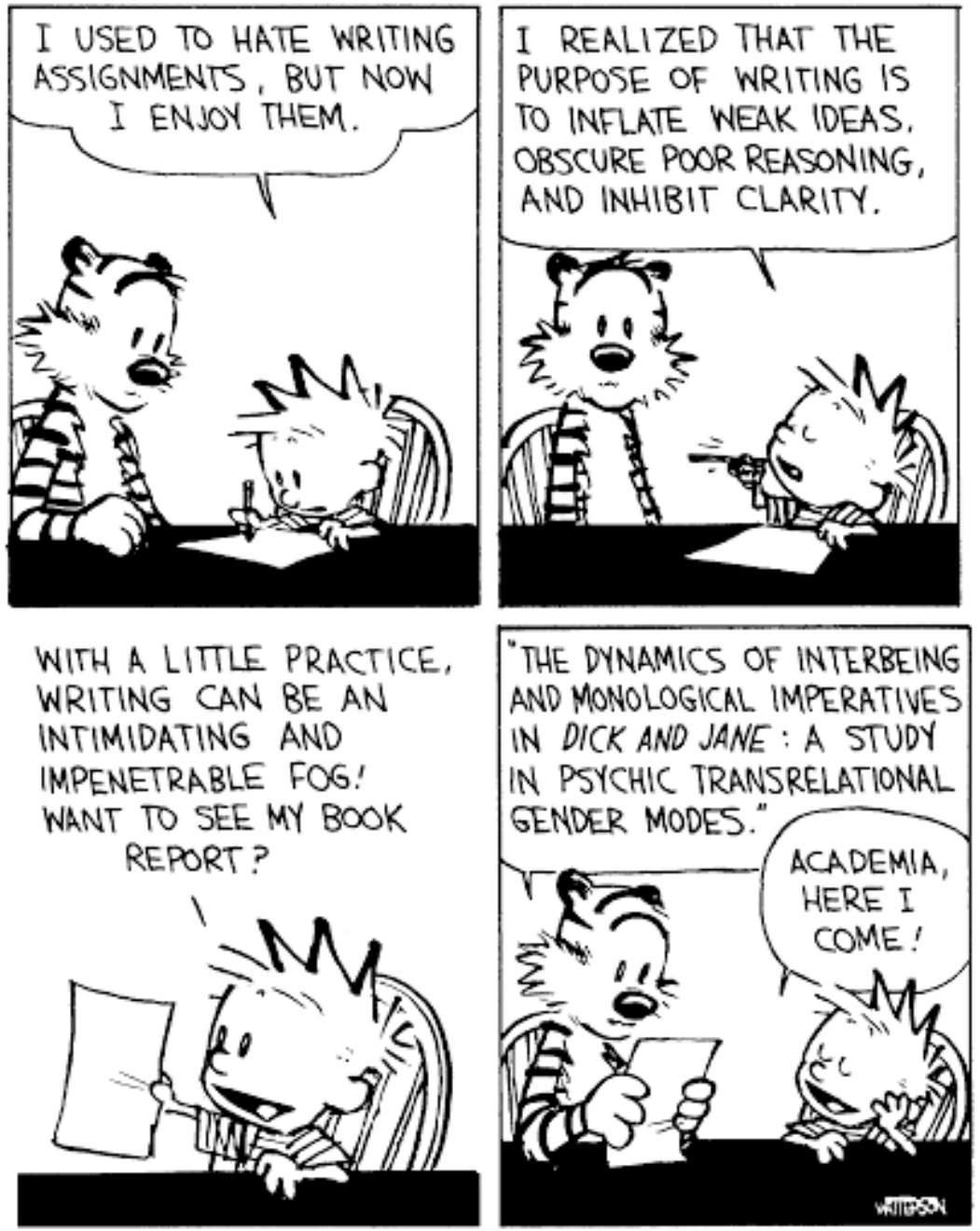
How to write a scientific paper **well**

style, grammar, and other tips

Dr. Morgan Feeney
morgan.feeney@strath.ac.uk
October 2021

Why bother?


SCIENCE ARTICLES: A GUIDE
AVERAGE SENTENCE IS EASY TO UNDERSTAND
AVERAGE SENTENCE IS HARD TO UNDERSTAND
SUBJECT MATTER IS COMPLEX
GREAT WRITING
TYPICAL WRITING
SUBJECT MATTER IS SIMPLE
HONEST WRITING
PROBABLY JUST BULLSHIT
smbc-comics.com


I USED TO HATE WRITING ASSIGNMENTS, BUT NOW I ENJOY THEM.
I REALIZED THAT THE PURPOSE OF WRITING IS TO INFLATE WEAK IDEAS, OBSCURE POOR REASONING, AND INHIBIT CLARITY.
WITH A LITTLE PRACTICE, WRITING CAN BE AN INTIMIDATING AND IMPENETRABLE FOG! WANT TO SEE MY BOOK REPORT?
"THE DYNAMICS OF INTERBEING AND MONOLOGICAL IMPERATIVES IN *DICK AND JANE*: A STUDY IN PSYCHIC TRANSRELATIONAL GENDER MODES."
ACADEMIA, HERE I COME!
WATTERSON

Part 1: what is good scientific writing?

Exercise 1: evaluating scientific writing

3 questions:

1. Is it well-written?
2. Why (or why not?)
3. How can it be improved?

The RAG endonuclease initiates *Igh* V(D)J assembly in B cell progenitors by joining D segments to J_H segments, before joining upstream V_H segments to DJ_H intermediates[1]. In mouse progenitor B cells, the CTCF-binding element (CBE)-anchored chromatin loop domain[2] at the 3′ end of *Igh* contains an internal subdomain that spans the 5′ CBE anchor (IGCR1)[3], the D_H segments, and a RAG-bound recombination centre (RC)[4]. The RC comprises the J_H-proximal D segment (DQ52), four J_H segments, and the intronic enhancer (iEμ)[5].

The fundamental role of chromatin loop extrusion in physiological V(D)J recombination.

Zhang et al (2019) Nature 573, 600–604.

The nucleotide second messenger c-di-GMP nearly ubiquitously promotes bacterial biofilm formation, with enzymes that synthesize and degrade c-di-GMP being controlled by diverse N-terminal sensor domains. Here, we describe a novel class of widely occurring c-di-GMP phosphodiesterases (PDE) that feature a periplasmic "CSS domain" with two highly conserved cysteines that is flanked by two transmembrane regions (TM1 and TM2) and followed by a cytoplasmic EAL domain with PDE activity. Using PdeC, one of the five CSS domain PDEs of *Escherichia coli* K-12*,* we show that DsbA/DsbB-promoted disulfide bond formation in the CSS domain reduces PDE activity.

Transmembrane redox control and proteolysis of PdeC, a novel type of c-di-GMP phosphodiesterase.

Herbst et al (2018) EMBO J. 37(8). pii: e97825. doi: 10.15252/embj.201797825.

Transition metals serve as an important class of micronutrients that are indispensable for bacterial physiology but are cytotoxic when they are in excess. Bacteria have developed exquisite homeostatic systems to control the uptake, storage, and efflux of each of biological metals and maintain a thermodynamically balanced metal quota. However, whether the pathways that control the homeostasis of different biological metals cross talk and render cross resistance or sensitivity in the host-pathogen interface remains largely unknown.

Zinc excess increases cellular demand for iron and decreases tolerance to copper in Escherichia coli.

Xu et al (2019) J Biol Chem.. doi: 10.1074/jbc.RA119.010023.

Tip #1: Read broadly & pay attention to how papers are written

A few examples of papers that I consider to be exceptionally well-written:

- Tschowri N, Schumacher MA, Schlimpert S, Chinnam NB, Findlay KC, Brennan RG, Buttner MJ. Tetrameric c-di-GMP mediates effective transcription factor dimerization to control Streptomyces development. Cell. 2014 Aug 28;158(5):1136-1147.
- Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S., Rogulja, D., Nickels, B., . . . Severinov, K. (2002). A Role for Interaction of the RNA Polymerase Flap Domain with the σ Subunit in Promoter Recognition. *Science, 295*(5556), 855-857.
- Chen, X., Schauder, S., Potier, N., Van Dorsselaer, A., Pelczer, I., Bassler, B. L., & Hughson, F. M. (2002). *Structural identification of a bacterial quorum-sensing signal containing boron. Nature, 415(6871), 545–549.* doi:10.1038/415545a

Part 2: how can you write well?

Exercise 2: describe the results in this figure

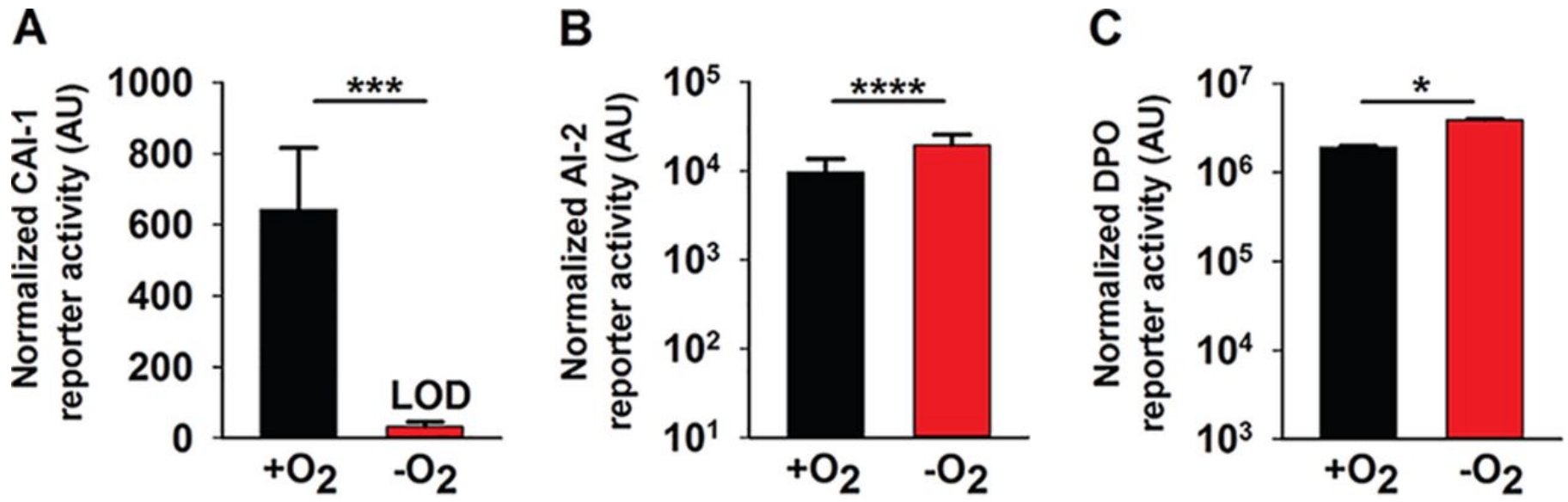


Figure 2. Oxygen deprivation modulates V. cholerae AI production. (A) 80%, (B) 25%, or (C) 30% cell-free culture fluids prepared from WT V. cholerae grown in the presence or absence of O_2 were provided to *V. cholerae* reporter strains that produce bioluminescence in response to exogenous (A) CAI-1, (B) AI-2, or (C) DPO. Data represent the average values from biological replicates ($n = 3$), and error bars represent SD. Statistical significance was calculated using a two-tailed Student *t* test. Asterisks are as follows: * denotes $P < 0.05$, *** denotes $P < 0.001$, and **** denotes $P < 0.0001$. LOD, limit of detection; AU, arbitrary units.

Ameya A. Mashruwala, and Bonnie L. Bassler mBio 2020; doi:10.1128/mBio.01572-20

Tip #2: Pay attention to the mechanics of writing -- Use a style guide

Rules on the following pages all come from

Strunk & White

section II: Principles of Composition

Available http://www.gutenberg.org/ebooks/37134 or at your friendly university library

Strunk & White:
Principles of Composition

8. Make the paragraph the unit of composition: one paragraph to each topic.

Strunk & White:
Principles of Composition

9. As a rule, begin each paragraph with a topic sentence, end it in conformity with the beginning.

(a) the topic sentence comes at or near the beginning;

(b) the succeeding sentences explain or establish or develop the statement made in the topic sentence; and

(c) the final sentence either emphasizes the thought of the topic sentence or states some important consequence.

Ending with a digression, or with an unimportant detail, is particularly to be avoided.

Example of paragraph structure: Tschowri et al 2014

BldD sits at the top of the regulatory cascade controlling development, serving to repress expression of sporulation genes during vegetative growth (den Hengst et al., 2010). In *Streptomyces coelicolor*, BldD controls the expression of at least 167 genes, including 42 genes (~25% of the regulon) that encode regulatory proteins (Elliot et al., 2001, den Hengst et al., 2010). Among these BldD targets are many genes known to play critical roles in *Streptomyces* development, including other bld regulators (e.g., *bldA, bldC, bldH/adpA, bldM*, and *bldN*), several whi (white) regulators required for the differentiation of aerial hyphae into spores (e.g., *whiG* and *whiB*), and genes encoding critical components of the cell division and chromosome segregation machineries such as FtsZ, SsgA, SsgB, and the DNA translocase SffA (den Hengst et al., 2010, McCormick, 2009). How BldD activity is regulated, however, has been unknown.

Strunk & White: Principles of Composition

10. Use the active voice.

Examples from Tschowri et al 2014

Having shown that BldD binds c-di-GMP, we tested the effect of c-di-GMP on BldD DNA binding.

Using global chromatin immunoprecipitation-microarray analysis (ChIP-chip), we previously identified the complete BldD regulon in S. coelicolor, showing that it encompasses ~167 transcription units

Strunk & White: Principles of Composition

10. Use the active voice.

EXCEPTION: methods sections!

Determination of c-di-GMP Binding to Proteins by Differential Radial Capillary Action of Ligand Assay
Radiolabeled c-di-GMP was synthesized in vitro using [γ32-P]-GTP and the purified diguanylate cyclase PleD∗ as described (Paul et al., 2004). The DRaCALA assays (Roelofs et al., 2011) were performed using 2 μg of His6-BldD or its N-terminally His-tagged domains that were incubated with ~11 nM 32P-c-di-GMP in DGC buffer (250 mM NaCl, 25 mM Tris pH 8, 10 mM MgCl2, 5 mM β-mercaptoethanol). For competition experiments, 266 μM cold c-di-GMP or GTP were added to the reaction. After a 5 min incubation at room temperature, 5 μl of the binding sample were spotted onto nitrocellulose membrane and the dried membranes were analyzed using a Phosphorimager.

Strunk & White: Principles of Composition

11. Put statements in positive form.

Make definite assertions. Avoid tame, colorless, hesitating, non-committal language. Use the word *not* as a means of denial or in antithesis, never as a means of evasion.

Example from Tschowri et al 2014:

Strikingly, the *bldD* null mutant formed small colonies lacking aerial hyphae, but—when examined by SEM—even young colonies of the *bldD* mutant were found to contain spore chains embedded in an excess of extracellular matrix

Strunk & White: Principles of Composition

12. Use definite, specific, concrete language.

Prefer the specific to the general, the definite to the vague, the concrete to the abstract.

Example from Tschowri et al 2014:

Arg114 is the only CTD residue that makes contacts to both intercalated c-di-GMP dimers. Arg114 hydrogen bonds to the guanines contacted by Asp128, as well as the O6 atoms of the adjacent guanines of the other c-di-GMP dimer (Figures 5A and 5B).

Strunk & White:
Principles of Composition

13. Omit needless words.

Vigorous writing is concise. A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts. This requires not that the writer make all his sentences short, or that he avoid all detail and treat his subjects only in outline, but that he make every word tell.

Strunk & White: Principles of Composition

15. Express co-ordinate ideas in similar form.
This principle, that of parallel construction, requires that expressions of similar content and function should be outwardly similar. The likeness of form enables the reader to recognize more readily the likeness of content and function.

Strunk & White:
Principles of Composition

17. In summaries, keep to one tense.

Strunk & White:
Principles of Composition

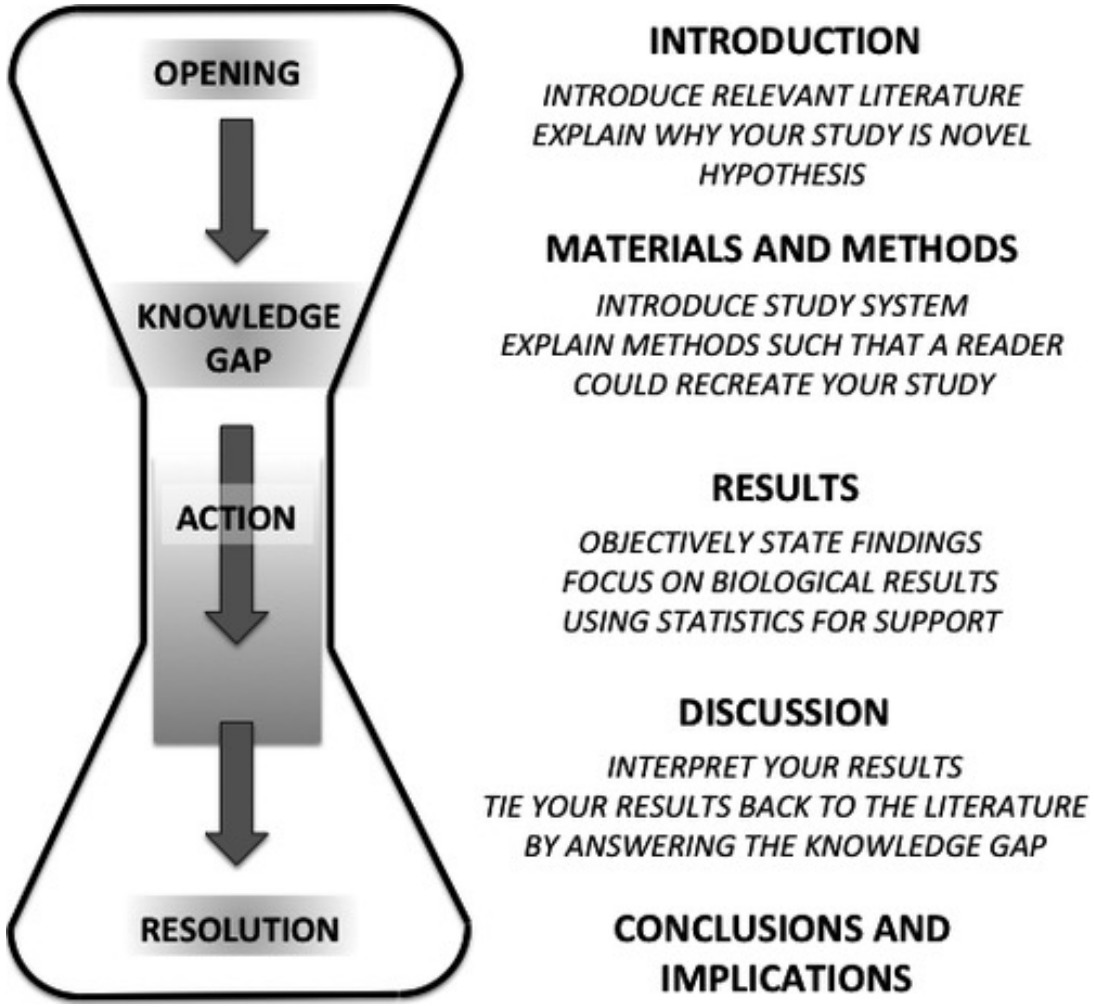
18. Place the emphatic words of a sentence at the end.

The proper place in the sentence for the word, or group of words, which the writer desires to make most prominent is usually the end.

Exercise 3: Write a paragraph

- Keep in mind the rules we just discussed
- Write a topic sentence, elaborate on it (2-3 sentences), and a concluding sentence
- Suggested topics: the weather, the nutritional content of the last meal you ate, something you learned about in class recently, the microbe you're focusing on for your project….

Part 3: The structure of scientific writing



Turbek et al (2016) Scientific Writing Made Easy: A Step-by-Step Guide to Undergraduate Writing in the Biological Sciences

Structuring a scientific paper

Introduction

- Start broad ("In all domains of life…")
- Give all necessary background information your reader will need
  - Edit later if necessary
- End with your hypothesis and aims

Make your train of thought logical and clear

- **To describe an experiment:**
  - 1. What you were trying to do (aim)
  - 2. How
  - 3. Results
  - 4. Significance (what it means)
- You should follow this format for every experiment (in writing your thesis or giving a formal/informal presentation)

Example:

The opposing effects of the overexpression of the DGC CdgB and the PDE YhjH suggested that high levels of c-di-GMP retard sporulation and low levels of c-di-GMP accelerate sporulation. Because the BldD-(c-di-GMP) complex serves to keep sporulation genes shut off during vegetative growth, loss of BldD should have a similar effect on *Streptomyces* development as depletion of c-di-GMP levels. To test this hypothesis, we deleted *bldD* from the *S. venezuelae* chromosome. Strikingly, the *bldD* null mutant formed small colonies lacking aerial hyphae, but—when examined by SEM—even young colonies of the *bldD* mutant were found to contain spore chains embedded in an excess of extracellular matrix (Figure 3A).

rationale

hypothesis

experiment

result

Example:

To gain insight into the cellular processes controlled by c-di-GMP in streptomycetes, we overexpressed either the active DGC CdgB from *S. coelicolor* (Tran et al., 2011) or the active PDE YhjH from *E. coli* (Pesavento et al., 2008). Strikingly, overexpression of both CdgB and YhjH blocked the generation of aerial mycelium by *S. venezuelae* (Figure 1B). However, scanning electron micrographs (SEMs) revealed that, whereas overexpression of CdgB blocked development, resulting in a classical bald phenotype, overexpression of the PDE YhjH in fact promoted sporulation, but the colonies appeared bald to the naked eye because aerial mycelium formation had been bypassed (Figure 1C). As judged by heat resistance, the spores made by the YhjH overexpression strain were as robust as those of the wildtype (WT) (Figure S1A available online). Moreover, overexpression of catalytically inactive versions of YhjH or CdgB had no effect on *S. venezuelae* development (Figure S1B). These data suggest that intracellular levels of c-di-GMP influence the timing of development. In particular, they suggest that increased c-di-GMP levels delay differentiation, arresting the colonies in the vegetative growth stage, whereas decreased levels of the second messenger accelerate development, favoring sporulation.

Tip #3: Make your train of thought logical and clear

- Use transitional words and phrases (long list @ http://writing2.richmond.edu/writing/wweb/trans1.html)
- However
- Moreover
- In spite of
- Although
- For example
- Therefore

Example:

c-di-GMP is monomeric in solution at physiological concentrations (Gentner et al., 2012). However, intercalated c-di-GMP dimers have been observed in crystal structures of the nucleotide alone and in complexes with effector proteins. Higher order c-di-GMP structures such as tetramers and octamers have thus far only been inferred from NMR and spectroscopic studies and require very high c-di-GMP concentrations (up to 30 mM) and monovalent cations (Zhang et al., 2006). These higher order structures are characterized by G-quartet interactions with a centrally bound potassium ion coordinated by four guanines. There are minimal base contacts and no base stacking interactions in these structures (Figure S4A) (Zhang et al., 2006; Gentner et al., 2012). By sharp contrast, the BldD-bound tetrameric c-di-GMP is a tightly packed structure that is not secured by ions. Rather, the c-di-GMP molecules are closely spaced and optimally positioned for interbase pairing, leading to the formation of a multistranded, base-stacked structure with top, middle, and bottom layers (Figures 5D and andS4A).S4A). There are 12 hydrogen bonds between the two intercalated dimers within the c-di-GMP tetramer, including contacts between the N3 atoms and exocyclic NH_2 amides of an adjacent base (Figures 5C and andS4B).S4B). Such contacts could not be formed with c-di-AMP due to its lack of an exocyclic NH_2 atom. Therefore, in addition to contacts from motifs 1 and 2, guanine-guanine base hydrogen bonds serve to specify c-di-GMP tetramer binding to BldD. Notably, formation of the c-di-GMP tetramer buries 24% of the total surface area (buried surface area [BSA]) of the c-di-GMP molecules (Figure S4B). By comparison, in most protein oligomers the BSA between protomers is ~15% (Wang et al., 2009). Finally, the interface between the intercalated c-di-GMP dimers that forms the tetramer is remarkably complementary in shape (Figure S4B). Thus, the combination of multiple contacts between the c-di-GMP moieties along with its extensive BSA and molecular shape complementarity lead to the creation of a compact and highly specific c-di-GMP tetramer. However, BldD is necessary to stabilize this tetramer and template its formation.

Exercise 4: Describe an experiment

- Write a paragraph describing an experiment
- Keep in mind the rules we just discussed
- Explain the point of the experiment, how it was done, what the results were, and what the results mean
- Suggested topics: dropping a pen to test the theory of gravity, flicking a light switch or changing a light bulb, attempting to delete *ftsZ* to show that it is an essential gene in *E. coli*, performing an experiment that you plan to do during your project…

Part 3: Edit, edit, edit…

Edit for clarity

- Does it make logical sense?
- Does it mean what you intend it to mean?
- Can a colleague/friend understand it?

Remember our 3 questions:

1. Is it well-written?
2. Why (or why not?)
3. How can it be improved?

Edit for grammar; proofread

- Put it in a drawer
- Ask Dr. Google (or a style guide) – which versus that, affect versus effect, etc.
- Details matter – italicize species names, get chemical formulae correct, fix typos
- Ask for help

Exercise 5: Edit, edit, edit

Can you improve this paragraph?

Staphylococcus aureus is notoriously known for its rapid development of resistance to conventional antibiotics. *S. aureus* can alter its membrane composition to reduce the killing effect of antibiotics and antimicrobial peptides (AMPs). To obtain a more complete picture, this study identified the resistance genes of *S. aureus* in response to human cathelicidin LL-37 peptides by screening the Nebraska Transposon Mutant Library.

Golla, R., Mishra, B., Dang, X., Lakshmaiah Narayana, J., Li, A., Xu, L., & Wang, G. (2020). *Resistome of Staphylococcus aureus in Response to Human Cathelicidin LL-37 and Its Engineered Antimicrobial Peptides. ACS Infectious Diseases.*

Exercise 5: Edit, edit, edit

Can you improve this paragraph?

Corynebacterium pseudotuberculosis is a pathogen of great veterinary and economic importance, since it affects livestock, mainly sheep and goats, worldwide, together with reports of its presence in camels in several Arabic, Asiatic, and East and West African countries, as well as Australia. In this article, we report the genome sequence of *Corynebacterium pseudotuberculosis* strain Cp162, collected from the external neck abscess of a camel in the United Kingdom.

Hassan et al. Whole-genome sequence of Corynebacterium pseudotuberculosis strain Cp162, isolated from camel. J Bacteriol. 2012 Oct;194(20):5718-9. doi: 10.1128/JB.01373-12. PMID: 23012291; PMCID: PMC3458653.

Exercise 6: Edit your own writing

- Exercise 2 (figure description)
- Exercise 3 (your own paragraph)
- Exercise 4 (describe an experiment)

Exercise 2

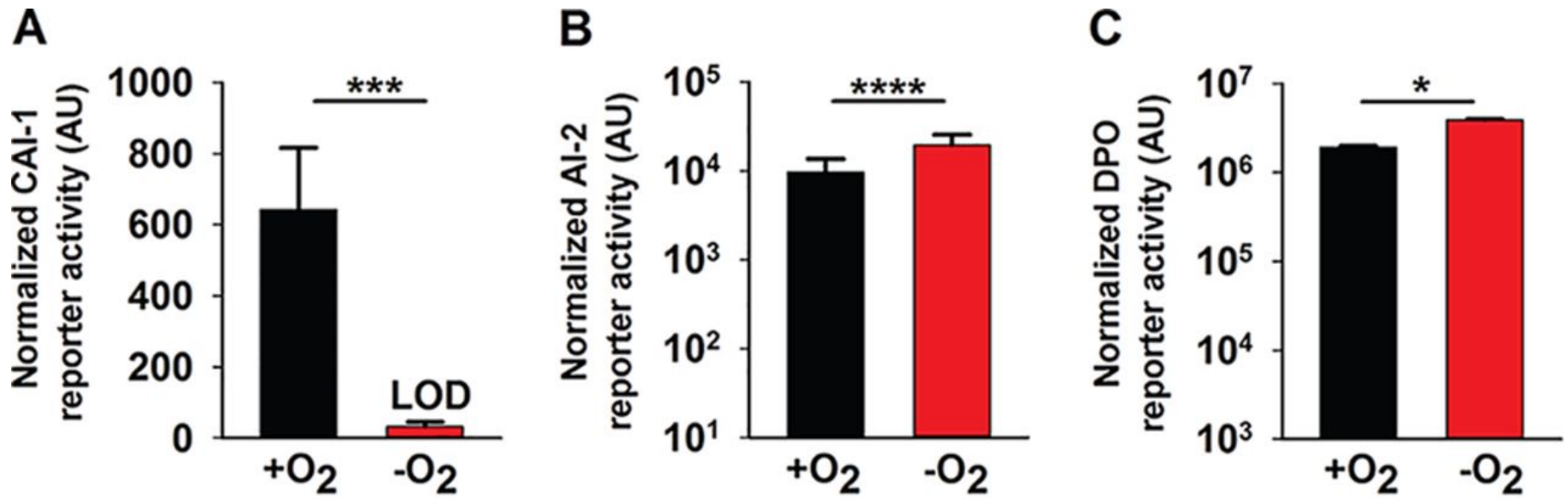


Figure 2. Oxygen deprivation modulates V. cholerae AI production. (A) 80%, (B) 25%, or (C) 30% cell-free culture fluids prepared from WT V. cholerae grown in the presence or absence of O_2 were provided to *V. cholerae* reporter strains that produce bioluminescence in response to exogenous (A) CAI-1, (B) AI-2, or (C) DPO. Data represent the average values from biological replicates ($n = 3$), and error bars represent SD. Statistical significance was calculated using a two-tailed Student *t* test. Asterisks are as follows: * denotes $P < 0.05$, *** denotes $P < 0.001$, and **** denotes $P < 0.0001$. LOD, limit of detection; AU, arbitrary units.

Ameya A. Mashruwala, and Bonnie L. Bassler mBio 2020; doi:10.1128/mBio.01572-20

Exercise 2: Compare with the authors' description

***V. cholerae* does not produce the CAI-1 QS AI under anaerobic conditions.** To our knowledge, *V. cholerae* QS has been studied only under aerobic conditions. We know that the marine-human host life cycle demands that *V. cholerae* transition between environments containing widely varying oxygen levels (17, 18). Moreover, QS is crucial in both *V. cholerae* habitats. Thus, we sought to investigate whether oxygen modulates *V. cholerae* QS. First, we assessed the relative levels of the three known QS AIs from *V. cholerae* C6706 Sm^r (here wild type [WT]) following aerobic and anaerobic growth. AI activity in cell-free culture fluids was measured using a set of three bioluminescent *V. cholerae* strains, each of which exclusively reports on one QS AI (either AI-2, CAI-1, or DPO) when it is supplied exogenously.

Unlike *V. cholerae* cultured in the presence of oxygen (here $+O_2$), *V. cholerae* grown in the absence of oxygen (here $-O_2$) produced no CAI-1 (Fig. 2A). Twice as much AI-2 and DPO accumulated in *V. cholerae* cultured $-O_2$ as in $+O_2$ (Fig. 2B and C). We note that the dynamic ranges for the CAI-1 and DPO assay are ~1,000- and ~4-fold, respectively, while that for the AI-2 assay is ~100,000-fold (2, 11). Thus, we consider the changes in CAI-1 and DPO to be physiologically relevant, whereas that for AI-2 is likely not, so we do not consider AI-2 further in this work. Additionally, *V. cholerae* cultured $-O_2$ grew to a lower final cell density than when grown $+O_2$ (see Fig. S1A in the supplemental material). We controlled for the reduced cell growth that occurs under the $-O_2$ conditions; nonetheless, no CAI-1 could be detected (Fig. S1B). Beyond lacking O_2, our culture medium lacked an alternative terminal electron acceptor. Thus, we also considered the possibility that *V. cholerae* cultured under $-O_2$ conditions was unable to respire and therefore unable to drive CAI-1 generation. However, supplementation of the *V. cholerae* $-O_2$ cultures with the alternative terminal electron acceptor fumarate, which is readily consumed by *V. cholerae* (29), did not rescue CAI-1 production (Fig. S1B). Collectively, these data suggest that production of CAI-1 and DPO by *V. cholerae* is affected by oxygen levels. In the remainder of this study, we focus on the functioning of the DPO-VqmA QS circuit under different conditions that are predicted to be encountered in the host. We address possible ramifications of our results concerning CAI-1 and AI-2 in Discussion.

Ameya A. Mashruwala, and Bonnie L. Bassler mBio 2020; doi:10.1128/mBio.01572-20

Tip #4: Practice, practice, practice

- Write a little every day (anything)