MIM workshop: General principles of scientific writing

13./4. Sep. 2023

Shinichi Sunagawa

First things first…

Course website/material

https://sunagawalab.ethz.ch/MIM_SW/HS-2023

Requirements for credit

Participation + report (due on 24. September 2023)

ETH zürich

S01xE01: Why am I here, what do I expect

Collected answers HS2023

- improve clarity of writing / communication
- write more efficiently, consicely, coherently
- understand WHY sentences read more correct or not
- input from others
- how to target specific audience / simplify scientific language
- manuscript structure /
- phrasing
- scientific writing tools, also AI, google docs, chatGPT
- how to get inspiration for writing
- improve on story telling / boarding
- how to write specific sections (abstract)
- how to get started (first sentences…)
- manuscripts vs grant writing
- improve writing style / order content and place them into right sections
- tips /
- consolidate different writing styles

S01xE02: Alignment of writing and audience

- Please read the abstract of Helfrich et al. 2018; highlight expressions, terminology, etc. that may be difficult to understand for a layperson (5 min).

- Next, please read the corresponding press release (Helfrich et al. 2018-press_release) and compare the language used (definitions, jargon, etc.), and identify what has been paraphrased.

S01xE02: Alignment of writing and audience

Plants are colonized by phylogenetically diverse microorganisms that affect plant growth and health. Representative genome-sequenced culture collections of bacterial isolates from model plants, including *Arabidopsis thaliana*, have recently been established. These resources provide opportunities for systematic interaction screens combined with genome mining to discover uncharacterized natural products. Here, we report on the biosynthetic potential of 224 strains isolated from the *A. thaliana* phyllosphere. Genome mining identified more than 1,000 predicted natural product biosynthetic gene clusters (BGCs), hundreds of which are unknown compared to the MIBiG database of characterized BGCs. For functional validation, we used a high-throughput screening approach to monitor over 50,000 binary strain combinations. We observed 725 inhibitory interactions, with 26 strains contributing to the majority of these. A combination of imaging mass spectrometry and bioactivity-guided fractionation of the most potent inhibitor, the BGC-rich *Brevibacillus* sp. Leaf182, revealed three distinct natural product scaffolds that contribute to the observed antibiotic activity. Moreover, a genome mining-based strategy led to the isolation of a *trans*-acyltransferase polyketide synthase-derived antibiotic, macrobrevin, which displays an unprecedented natural product structure. Our findings demonstrate that the phyllosphere is a valuable environment for the identification of antibiotics and natural products with unusual scaffolds.

- reduce redundancy of the 2nd and 3rd sentence
- improve structuring to better see what was done when (first, then, then)
- split long sentences into two or more
- get difficult from “Genome mining identified …”
- what is the MIBiG database?
- first two sentences are ok, then relevance is lost until concluding sentence
- **some words are not needed for content**
- **jargon** / targeted towards microbiologists
- why is it important to discover unusual structures?

S01xE02: Alignment of writing and audience

A team of ETH researchers led by Julia Vorholt and Jörn Piel have discovered new antibiotic substances in bacteria that colonise the leaf surfaces of a local wild plant.

A wide variety of different microorganisms, such as bacteria and fungi, live on the leaves of plants. Although they offer few nutrients, leaf surfaces are densely populated. In an effort to keep the competition at bay, many of the leaf dwellers turn to chemical warfare: they develop antibiotic substances that prevent the growth and reproduction of their fellow occupants.

During a systematic search of the leaves of thale cress (Arabidopsis thaliana), a group of researchers led by the ETH professors Julia Vorholt and Jörn Piel from the Institute of Microbiology have now discovered a remarkably chemically productive bacterium: Brevibacillus sp. Leaf 182. In experiments, it inhibited half of the 200 strains that the researchers had isolated from the leaf surfaces. The bacterium produces and secretes at least four antibiotic chemical compounds. Two of these compounds were already known, while a substance called macrobrevin presented a previously unknown chemical structure.

S01xE02: Alignment of writing and audience

"Using bioinformatic methods, we looked for groups of genes that generally control the production of substances and could thus have effects on other bacteria," explains Vorholt. At the same time, the researchers performed laboratory tests to determine which of these strains have an antibiotic effect on other strains, ensuring th at certain bacteria can no longer reproduce. In total, they discovered over 700 such antibiotic interaction s between different microbial strains.
The aim of the project, which is funded through SNSF and ERC grants, was to find new antibiotics in a previously unexplored habitat. "Until now, research has focused particularly on soil as a habitat; however, we keep finding the same substances," says Vorholt.

The search for new antibiotics is becoming more and more difficult, with Piel speaking of the antibiotics crisis: "We hardly have any antibiotics now that at least one pathogen is not resistant to." He says that companies have more or less suspended the search for new substances because they are not considered profitable enough.
With their project, the ETH researchers are tapping into a new reservoir with great potential. "We will now determine whether macrobrevin and other newly discovered substances are also effective against bacteria that cause diseases in humans," says Piel. But in his opinion, the even greater achieve ment is having shown that there are still many natural antibiotic substances waiting to be discovered in the microcosms of leaf surfaces, which until now have not been thoroughly investigated. "This incredibly diverse ecosystem can most definitely still offer medicine many new leads. Our results confirm that it is worth expanding the search for antibiotics in nature."

S01: Take home message I

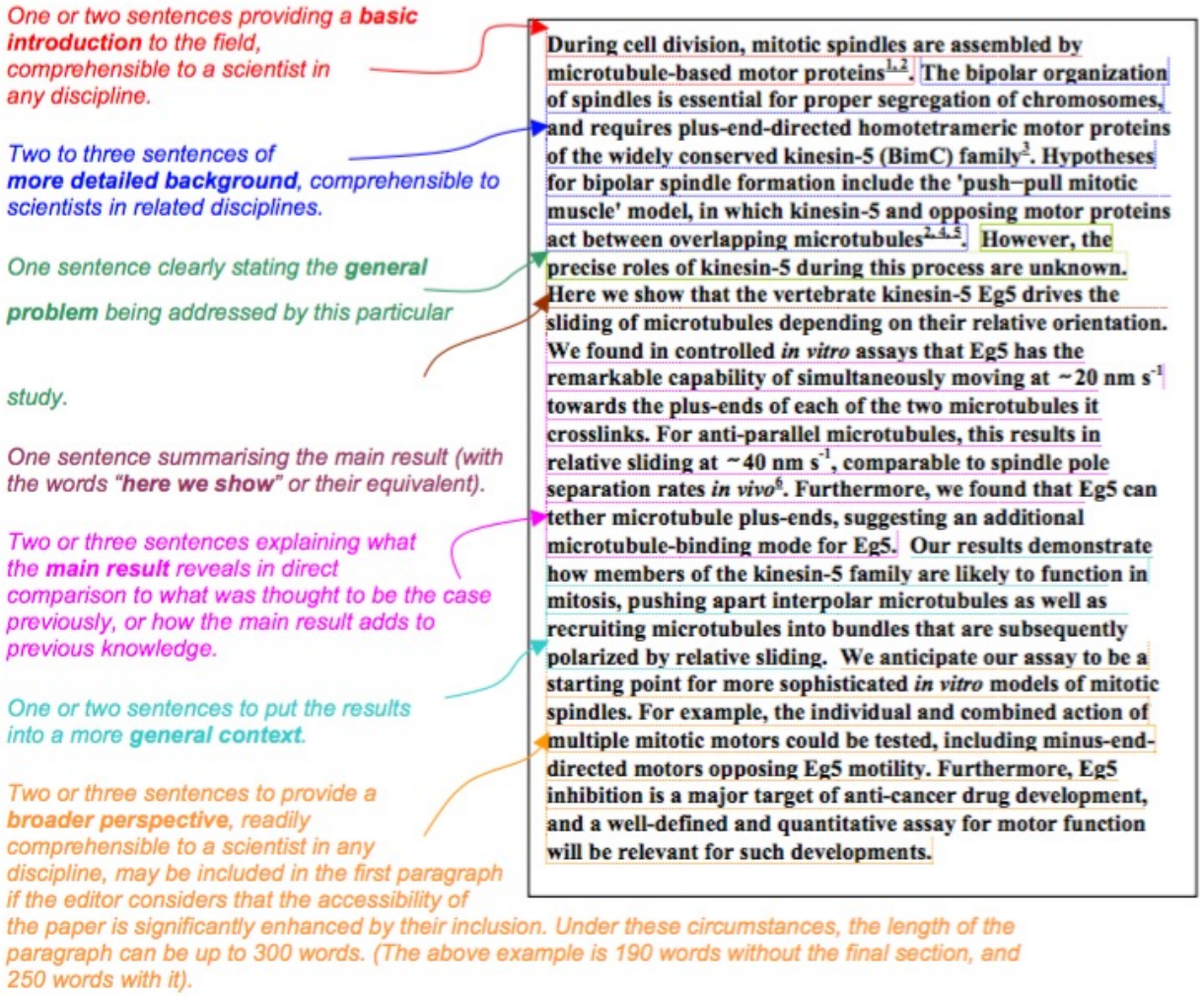
- **Appropriate targeting of audience is key to scientific writing (and oral presentations, too!)**

- **Tool: Try as hard as possible to read your own writing from the perspective of a well-educated, but non-informed reader**

S02xE01: writing an abstract

- Individually, prepare a 1 min elevator pitch on the background, content, latest results and significant of your research.
- In pairs, explain your research to each other within 1 min (sharp). At this stage, do not ask questions.
- After that, ask follow up questions on details that were not clear in the first round (2 x 3min).
- Write down, **as bullet points**, your improved elevator pitch

The Abstract – template by Nature



- **Why did I start?**

→ **Background/Significance**

- **What did I do? / How did I do it?**

→ **Material/Methods**

- **What did I find?**

→ **Results**

- **What does it mean?**

→ **Discussion/Conclusion**

https://sunagawalab.ethz.ch/MIM_SW/HS-2023/2-Annotated_Nature_abstract.pdf

S02xE01: writing an abstract

- Use the summary paragraph for *Nature* and **write full sentences** for a draft abstract of your own work

S02: Take home messages II

- **Brainstorming and explaining to others, getting feedback and reviewing your ideas are powerful steps that help writing well-structured, concise abstracts directed towards a specific target audience.**
- **Abstract should provide answers to 4 questions:**
  - **Why did I start? -> Background/Significance**
  - **What did I do? / How did I do it? -> Material/Methods**
  - **What did I find? -> Results**
  - **What does it mean? -> Discussion/Conclusion**

Tool: General strategy for efficient writing:

→ build skeleton first, add meat, do cosmetics last

Lunch break until 13:00

Introduction: why did I start?

- Provide the reader with the necessary background to understand the very **specific question** or hypothesis you are aiming to investigate, **and only with that background**
  - → typical mistake: review-type introductions
  - → better: foreshadow alignment to results
  - indicate why the research is important: “so what?”
  - indicate the need to extend previous work: “what do we miss if not?”
  - announce the experimental procedure and general findings:
    - whet appetite and let reader find out how you got there

Materials and Methods: what did I do?

- Accurate description of materials and methods used to generate the data of your work
- Necessary and sufficient information for reproducing all results
- Pay attention to use of units, vendor details, software version used, etc.

Results: what did I find?

- Write results “around” your data, i.e., figures and tables
- Methods can be mentioned, but only as much detail as needed to understand data

Discussion: what does it mean?

- Reference to the main purpose or hypothesis of the study

  → Re-read Introduction and Discussion (skip Methods and Results)
- Turn summary of the most important findings into implications

  Instead of starting with: “We found that A impact B (summary repeated)”

  → “Our finding that A impacts B shows …”
- Put results into context by comparing them to state-of-the-art

  → “Our study corroborates…”; “Our results contrasts previous findings…”
- State limitations of the study in a forward-looking way:

  Instead of: “Our study was limited to …, but could show that (summary repeated)”;

  → “To generalize our findings beyond …, future studies should […]”
- Explain wider implications of the study and future directions

General rules

- Space, that is, length of text, number of figures/tables, number of citations, is limited. Therefore, scientific writing is accurate, structured, clear, non-redundant and concise.

- General rule: Write your text as long as required to understand in as few words as possible.
  Example: *“Due to the fact of more run-off into the water, the end result is more bacteria in the water.”*
  Revision: *“Higher levels of bacteria are caused by increased run-off.”*

- Avoid:
  - non-quantitative adjectives (many/lots, some, little, very)
  - ambiguous wording, grammar

  “Using multiple-regression techniques, the animals in Experiment I …”

Analytical reading

- Is the state of the art well described? Is prior work referenced?
- Is the specific knowledge/method gap identified (usually by question and/or hypothesis: Q and H)?
- Is the relevance of Q and H described?
- Are the experiments/analyses aligned to Q and H?
- What evidence do the data exactly provide? Are limitations / alternative explanations adequately discussed?
- Are claims aligned with questions? Do they close the specific gaps?
- Are claims overstated/-sold?
- Methodology: adequate, timely, solid
- Is the presentation (text, display items) accessible (i.e., easy to understand)?
- How large is the impact/advance?
- Check for errors/negligence/sloppiness

S03: Take home message I

- **Analytical reader → better writer**

Tool: Read manuscripts like a robot (i.e., without any pre-assumptions) focusing on structure rather than content

- **Relevance: a good research question has significant consequences if it remains unanswered. Ask the question: "So what?"**

S03xE01: Power of position

- *Fleming, in 1929, discovered penicillin after a bacterial plate he was culturing became contaminated with a spore of the fungus Penicillium.*

- **Re-order the sentence in as many ways as possible to put emphasis on different aspects.**

S03xE02: Cohesion vs Coherence

Compare the following passages A) and B). Which one appears to be more coherent? Why?

A) Consistent ideas toward the beginnings of sentences, especially in their subjects, help readers understand what a passage is generally about. A sense of coherence arises when a sequence of topics comprises a narrow set of related ideas. But the context of each sentence is lost by seemingly random shifts of topics. Unfocused paragraphs result when that happens.

B) Readers understand what a passage is generally about when they see consistent ideas toward the beginning of sentences, especially in their subjects. They feel a passage is coherent when they read a sequence of topics that focuses on a narrow set of related ideas. But when topics seem to shift randomly, readers lose the context of each sentence. When that happens, they feel they are reading paragraphs that are unfocused end even disorganized.

S03xE02: Cohesion vs Coherence

Compare the following passages A) and B). Which one appears to be more coherent? Why?

A) Consistent ideas toward the beginnings of sentences, especially in their subjects, help readers understand what a passage is generally about. A sense of coherence arises when a sequence of topics comprises a narrow set of related ideas. But the context of each sentence is lost by seemingly random shifts of topics. Unfocused paragraphs result when that happens.

B) Readers understand what a passage is generally about when they see consistent ideas toward the beginning of sentences, especially in their subjects. They feel a passage is coherent when they read a sequence of topics that focuses on a narrow set of related ideas. But when topics seem to shift randomly, readers lose the context of each sentence. When that happens, they feel they are reading paragraphs that are unfocused end even disorganized

S03: Take home messages II

- **Use power of position to keep writing structured and logical to help reader follow easily.**
  - **Within a sentence**
  - **Within a paragraph: → topic sentence**
- **Sentences are cohesive if they share a topic/idea. A passage/paragraph may, however, not be coherent if it contains too many topics/ideas.**
- **Each paragraph aims at communicating a specific topic/idea. Coherence is achieved when this topic/idea is followed throughout.**
- **For clarity, sentences need to be cohesive and paragraphs coherent.**

S04: Analytical reader II

→ Rather than reading a piece of writing to understand the content it aims to convey, read analytically:

→ What is the topic of this paragraph?
→ What is the paragraph trying to convey?

→ Are paragraphs coherent?
→ Is the “power of position” used efficiently?

→ Is the writing clear? Can sentences be divided to improve clarity?
→ Are there grammatical errors?

→ Figures and Tables: does the information provided convey a message? Is all required information presented?

S04xE01: Analytical reading of an article

- Download the manuscript:
- https://sunagawalab.ethz.ch/MIM_SW/HS-2023/cfDNA_paper.pdf

- Read through **analytically** and make notes of what you find

- Take a look at the figures and captions and make suggestions, how they could be improved

Tables

- Use clear and informative titles

Table 1. Height after treatment

Group	light	5 days	10 days
control	12	70.3±2	90±10.5
test	12	60.4±1.5*	78±7.9*
control	16	75.7±8.	100±3
test	16	52.2±2	81±6.7

*$P<0.05$.

Table 1. Exposure to salinity reduces the growth of wheat plants.

Group	light	5 days	10 days
control	12	70.3±2	90±10.5
test	12	60.4±1.5*	78±7.9*
control	16	75.7±8.	100±3
test	16	52.2±2	81±6.7

*$P<0.05$.

Tables

- Use informative row and column titles, units, error values and sample sizes

Table 1. Exposure to salinity reduces the growth of wheat plants

		Height, cm (mean ± S.E.M)	
Group (n = 5 each)	Light/day (h)	5 days exposure	10 days exposure
Control group (0 mM NaCl)	12	70.3±2	90±10.5
50 mM NaCl	12	60.4±1.5*	78±7.9*
Control group (0 mM NaCl)	16	75.7±8.	100±3
50 mM NaCl	16	52.2±2	81±6.7

*$P<0.05$.

S04: Take home messages I

- Convey a message in the title of figure captions and table headers (rather than describing the obvious)

S04xE03: Tips for writing

- being familiar with the requirements / structure, who the audience is
- start with bullet points / a plan; clarity about what to present
- have guidelines / draft / reference material
- evenings/weekend rather than in office or busy lab environment
- rather write down too much and then shorten
- DEADLINE x 5
- focus, set apart time for writing task accounting for preparations
- early morning x2 or evening when no people are around x2, when nothing else is to do
- need free blocks of time (2-3 hours) x 2
- divide tasks into smaller parts
- start with a fresh mind (after leaving the work for some time)
- music
- during the night when there is no distraction
- get feedback / help
- detailed plan of what gets done after what amount of time

S04xE03: Tips for writing

- **Brainstorm factors that help you or prevent you from writing**
  - quiet environment
  - **deadline:** https://www.ted.com/talks/tim_urban_inside_the_mind_of_a_master_procrastinator
  - follow up quickly / not right after experiments
  - in the lab (not at home); in-between work
  - morning hours (better focus)
  - no distractions
  - structure (time) / rather no time constraints

→ Optimal conditions differ from person to person

Time management & Pareto

The Eisenhower Matrix

	URGENT	NOT URGENT
IMPORTANT	**QUADRANT 1** IMPORTANT AND URGENT	**QUADRANT 2** IMPORTANT BUT NOT URGENT
NOT IMPORTANT	**QUADRANT 3** URGENT BUT NOT IMPORTANT	**QUADRANT 4** NOT IMPORTANT AND NOT URGENT

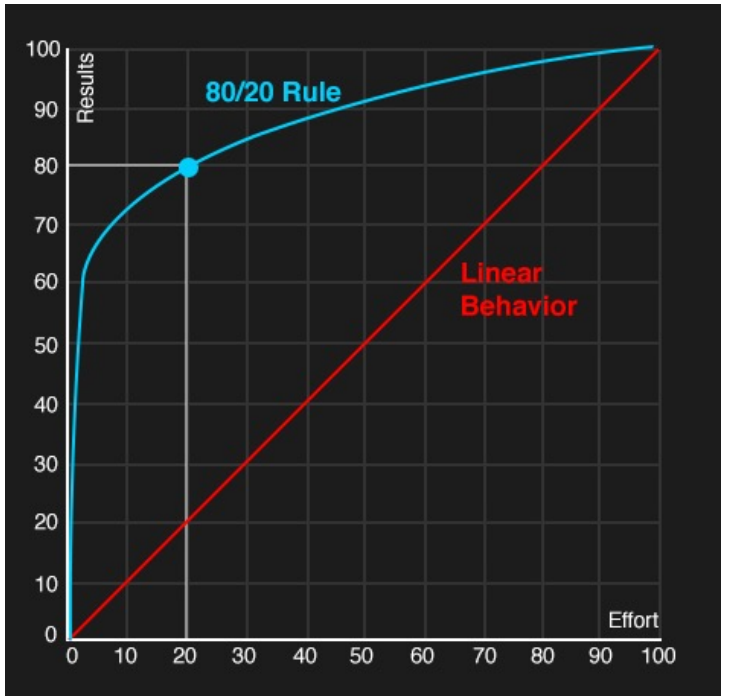
waitbutwhy.com

The Eisenhower Matrix

	URGENT	NOT URGENT
IMPORTANT	**Q1** DO NOW	**Q2** DECIDE WHEN TO DO IT
NOT IMPORTANT	**Q3** DELEGATE IT AWAY	**Q4** DELETE IT

waitbutwhy.com

The Pareto principle



David Lindsay: A guide to scientific writing.

S04: Take home messages

- Set realistic goals with timelines, meet them.
- If you do not meet them, revise accordingly.
- Eisenhower matrix + Pareto's principle can be helpful in managing your time.

S05: Better reader = better writer

- **S05xEx01 - Recap and homework**
- Summarize the tools you learned today and apply the tools to improve it.
- For next day, analytically read the “Draft manuscript”:
  - what is the topic each paragraph? what could have been the initial bullet points for each paragraph?
  - are the sentences cohesive? Is the paragraph coherent?
  - what is the problem?
  - is it clear why I should care?
- Edit this document by identifying issues, making suggestions, pointing out missing information, and importantly being constructive!