Minutes of

The 13th MIRAGE Meeting

July 27 & 28, 2022
12-3 pm CEST
Online

September 23, 2022
Dr. Carsten Kettner

Table of Conents

List of Participants

Participants (day 1) - Please add your name**:** Kiyoko F. Aoki-Kinoshita, Nicki Packer, René Ranzinger, Ten Feizi, Yan Liu, Cathy Costello, Maria Angeles Canales, Joe Zaia, Carsten Kettner, Guinevere Lageveen-Kammeijer (only July 27), Yukie Akune, Weston Struwe (July 27 only), Akul Mehta, Catherine Hayes, Niclas Karlsson, Erdmann Rapp, Joshua Klein, Daniel Kolarich, William Hackett, Pauline Rudd

Participants (day 2) - Please add your name: Carsten Kettner, Erdmann Rapp, Yan Liu, Akul Mehta, Will Hackett, Pauline Rudd, Daniel Kolarich, Niclas Karlsson (Jackal), Rene Ranzinger, Frédérique Lisacek, Kiyoko F. Aoki-Kinoshita (llama), Cathy Costello, Erdmann Rapp, Ten Feizi, Joe Zaia, Yukie Akune

Excused: Jim Paulson, Catherine Hayes (day 2), Weston Struwe (day 2), Guinevere Lageveen-Kammeijer (day 2), Maria Angeles Canales (day 2)

The participants allowed the recording of the meeting (for protocol purposes only). The recording is deleted as soon as the purpose is fulfilled.

Agenda and results

Outreach and project management

- the MIRAGE landscape
    - extension by additional projects (e.g. Glygen, GlySpace etc.)?

RESULT: Yes – but needs to be defined in some details as the relationships are not that easy to present in the landscape graphics
(see e.g. https://www.beilstein-institut.de/index.php?rex_media_type=tinymcewysiwyg&rex_media_file=mirage_spinoffs_20190708.png).

- CE Guidelines published (see below).
- MIRAGE website updated

- - more example publications or even templates with a data example appreciated,
  - recommendations for databases and repositories - any updates required? → link added for CarbArrayART

- Update MIRAGE@FAIRSharing.org
  - all MIRAGE guidelines are part of the ISO 20691 collection (Requirements for data formatting and description in the life sciences)
  - the registered guidelines are easily to reach on FAIRSharing.org: either search in FAIRSharing.org for MIRAGE or click on the corresponding link („Registered with FAIRSharing.org) on the guidelines page of MIRAGE (https://www.beilstein-institut.de/en/projects/mirage/guidelines/)

- MIRAGE flyers sent out – who's received them?

RESULT: check for Pauline, René, Kyoko, Cathy. Niklas and Nicki will check at their offices and send a note.

  - Update needed?

RESULT: Yes – for NMR and CE Guidelines, QR code appreciated

- organization of glycoproteomics guidelines meetings

Overview presentations

- Kiyoko: Integrating GlycoPOST and UniCarbDR
- Frédérique: The GlySpace-MIRAGE interaction
- Catherine: Tool for converting a number of GWS strings (from GlycoWorkBench/GlycanBuilder/GRITS) into GlycoCT, and using this to assign IDs from GlyConnect and GlyTouCan (Catherine Hayes)

Guidelines

MS Guidelines

Discussion

- there seem to be some issues regarding the MIRAGE guidelines that are intended to support authors gathering and reporting all relevant data in publications and those applications that collect data in compliance with the MIRAGE guidelines. For some applications the MIRAGE recommendations are too strong and burdening for users that submit their data, for other applications the guidelines do not provide all details or are too unspecific in order to describe an experiment comprehensively.
- minimum information standards for repositories possible to create? What would be then 'minimum'?
- Where should be the burden be placed: in the manuscript or in the data submission – or both?

CONSENSUS: guidelines support authors to report their data comprehensively and completely in a publication. Intention of the guidelines was not to support software development. However, concerns were raised that if the entire information is left exclusively with the paper, there is no control over the metadata (this targets to the use of unambiguous terminologies and ontologies).

- require repository input and allow "Reporting" tool to provide information for manuscripts - as supplementary.
- How to assess "MIRAGE Compliance" of a repository/database?
    - background: sometimes not entirely true for some tools although claimed but very close to compliance
    - should be done by a third-party
    - based on trust that the members of the Commission do know what they're doing → software reflects guidelines as accurately as possible and necessary.

CONSENSUS: As long as no doubts are raised about this, we can live with this pragmatic solution

Suggestion: Update of guidelines to simplify work of data reporters and software developers and review guidelines

ACTION POINT: Reassessment of MS guidelines.
- ○ Contributors: **Niclas (lead)**, Guinevere, Cathy, Daniel, Weston, Nicki, Frederique?, Kiyoko
- ○ Tasks:
    - ■ define the borders, how far to go, what to include and what to leave out
    - ■ start with a core document that is applicable to any MS experiments
    - ■ creation of variants based on the core document on
        - ● intact glycoproteins
        - ● glycoproteomics
        - ● N- and O-glycans
        - ● glycoconjugates

Suggestion: Publication: how to apply the MIRAGE guidelines, retrospective on all MIRAGE guidelines, note on the revision/update of the MS guidelines, making the recommending journals aware that there is something new,

ACTION POINT: draft of a manuscript covering the topics listed above.
Contributors: **Niclas (lead)**, Daniel, Frédérique, Cathy, Carsten

Guidelines for developers (suggestion from René)

Background: the description in the MIRAGE guidelines leave a lot of space for interpretation which may lead to different sets of "MIRAGE compliant" datasets in different databases and software tools. Especially, if that interpretation is done by software developers that are not experts in the corresponding fields. Developer guidelines or let's call them developer recommendations can give software developers information/blueprint in hand that they can easily implement without the need for interpretation.
We went through this exercise for the array guidelines:
https://docs.google.com/spreadsheets/d/1SdVLUxb45mC3EantKIIixW4YkIqoT1XJllrVKJyY13E/edit#gid=0

Please keep in mind that this is a working document to give you an idea of what René is talking about, not meant as a standard template or something. The grey rows come one to one from the array guidelines. The rows below each of these MIRAGE rows show the list of "atomistic information" we agreed is needed to be stored/handled by software tools and databases for the corresponding section.

As a developer that document gives the idea what needs to be stored, e.g. the structure of the data (column B-D), what is mandatory or optional (column F), how often certain information can be provided (column G), the data type (column H), what is currently required by MIRAGE (column I) and some additional conditions between individual information (column J).

René believes that this is a useful document and should best be provided by experts that do a common interpretation of the guidelines, rather than individual developers. He suggests to create such a document for any of the MIRAGE guidelines.

However, this is a lot of work and it is fine to say that it is outside of the scope of MIRAGE.

Frédérique, Niclas and Kiyoko agreed that such a document is very relevant for database/repository maintainers and mentions that this is at least in parts done for glycoproteomics guidelines and GlycoPOST and UniCarbDR as well.

René adds that this descriptive document is not only relevant for databases and repositories (as it is doubtful that we need multiple databases for e.g. MS or array data) but also for developers of tools such as for data submission in order to use a consistent metadata schema.

CONSENSUS: These documents should be published on the MIRAGE website

ACTION POINT: **Provide the guidelines for developers** (for MIRAGE MS Guidelines and Array Guidelines) to CK who will put them on the website under CC-BY 4.0 license.

INTACT Glycoprotein Guidelines

Modularity (for binding and identification, as suggested at the previous meeting) possible? deposition of raw data in PRIDE possible?
next steps:
1) drafting the manuscript (postponed) and
2) incorporating feedback from the MIRAGE members and colleagues

RESULT, ACTION POINT: The INTACT Glycoprotein Guidelines will become part of the MS guidelines as a variant, maybe there is also some overlap with the Glycoproteomics Guidelines which is to be figured out. Check this potential overlap.

Glycan array guidelines

https://docs.google.com/presentation/d/1zp3p1wgVlwOJQjvC1_x-LB1NWfnbmTOm7aPs3dMZhnE/edit?usp=sharing

Yukie:

- the MIRAGE array guideline template is a Word file that is flexible and allows a lot of free text but the content is not ready to transfer into software tools and databases. Here, she presents a first draft of an interactive Excel template file for the collection of array data - discuss issues with using Excel.
- templates are for free text entries, provide too much flexibility and liberty to add many binding sample data, abbreviations, full names, references to other documents but not machine-readable (not suitable to transfer into software).
- headings for metadata are needed, implemented as Excel file.
- still some issues to be resolved: large amount of information (assay conditions, glycan binding samples, glycan libraries), complexity of parameters (mandatory vs. optional fields, multiple entries, missing taxonomy, conditional information (commercial/non-commercial), active scripting (!) need for consistency check and validation (security problems, excludes other Excel-like software).
- software application needed to implement these requirements → creating a repository which allows data export as Excel file.

René:

- we were considering an amended version of the Glycan array MIRAGE document which is suitable for ready download by bioinformaticians, also a smooth linkup with the Glycan Array Repository in which we are including metadata entries that are mandatory for the MIRAGE document.
- Use cases/motivation: users not willing to enter data in a web form but prefer locally stored files (e.g. Excel file), file can be used as supporting information for the corresponding manuscript, useful for data exchange.
- Repository already provides web forms that only show that fields that are required/requested,
- export of data from repository as Excel file or machine readable JSON format. The Excel file contains links to the information pages with the metadata and optional separate sheets with the information

Ten:

- update of developments with Glycan Array Repository relate to Jim's and Ten's interactions with the *Nature* Journals. *Nature* editors are interested ot have glycan array data reported according to the MIRAGE standards but require the deposition of data in a repository. This discussion will be resumed once this repository is publicly available.

- ○ perspectives: building a repository by fall, pre-populated with CFG data,
- ○ users then can upload their data using the Excel file but uploaded data won't be curated,
- ○ *Nature* editors indicated that they require the unprocessed data (raw scans), images and annotated processed data which appear in the manuscript.
- ○ this data will be stored along with the metadata in that repository.

DISCUSSION
- can the Excel file be adopted and used for other guidelines, including CE, MS etc? → not really, it needs to be customized for each guideline. Define the content, create a useful logic (e.g. commercial vs. non-commercial with subsequently opening corresponding fields),
- methods, materials and techniques need to be better described, such an Excel file would be helpful,
- Niclar points to the Excel spreadsheet he's developed for UniCarb which fulfills most of the demands for MS data (see Rojas-Macias MA, Mariethoz J, Anderson P, et al., 2019, *Nat. Commun*. doi:10.1038/s41467-019-11131-x), not sure how much efforts is required to include CE data (Niclas and Kiyoko will discuss this),
- Kiyoko: Excel sheet for Sample Preparation is already available,
- the MIRAGE relevant part from GlycoPOST maybe separated and reused for other MIRAGE guidelines.
- additional support via an Excel file to be filled out in compliance with the MIRAGE guidelines would be welcome
- further meeting required to organize acquisition form in Excel
- testers welcome for the input form

CONSENSUS and ACTION POINTS
templates for LC and other guidelines needed as well: **Kiyoko (lead)**, René, Niclas, Erdmann, Pauline, Cathy, Weston, Nicki, Daniel. Kiyoko sends out an email.

Meeting requested to discuss about the structured files already existing to get to a consistent and consolidated common solution for (at least most of) the MIRAGE guidelines.

Glycoproteomics Guidelines

- presentation of draft and progress report
- https://docs.google.com/spreadsheets/d/1KRvALkOcvg8Ixcr7vpKDgbvVYGAv5E9A/edit#gid=499472310
  - ○ homogenization of the comments and instructions,

- ○ bioinformatics resources available and suggested that are relevant to glycoproteomics in order to meet more standardized requirements,
- ○ after review from HGI people (sometimes too verbose), revised and much less verbose now.
- ○ new aspects: procedure output, interpretation & validation and quantification

DISCUSSION:

Niclas: How do these guidelines support GlyConnect? Frédérique: Assumed there is a template for software developers to output the results (part of the HGI competition, where every participant is going to create a template that is filled in accord to the guidelines).

- • Attempt to make Bionic files and the guideline template consistent and the other way round,

Carsten: is this still discussed with HGI? Frédérique: yes, it is still an ongoing process. One result of these discussions is that a template is required.

ACTION POINT(S)

- ● ALL: request for final review by MIRAGE group, deadline for comments: end of year,
- ● CK: publish on MIRAGE Guidelines page, indicated as draft,
- ● next steps:
  - ○ creation of a template that allows data acquisition based on the guidelines (,
  - ○ refer to other relevant MIRAGE guidelines (e.g. MS, HPLC, etc.)

NMR guidelines

- ● no progress so far due to new commitments. More action will be taken soon..

CE Guidelines

- ● paper published
- ● waiting for example publications to be published on the MIRAGE website

LC Guidelines

- ● template and completed example still missing. This was Matthew's task but with his leave of academia it is not very likely that he will be able to finalize this. Who'd be willing to take over?

- Current Version of the LC Guidelines: Guidelines - Beilstein-Institute, https://www.beilstein-institut.de/en/projects/mirage/guidelines/#liquid_chromatography

ACTION POINT: DK and PR: provide completed template(s) with real data, exemplar publications, by the end of the year.

DISCUSSION:

Nicki: GlycoStore: There is a lot of and very valuable data in this database which should be still accessible to the public. What about its future? Can it move to another place to be permanently accessible? We need to avoid losing data that has been deposited. Curation and technical development may be a future issue to keep the database alive. Pauline: Future of GlycoStore is not finally clear. There are some parties interested, maybe, GlycoStore will move under the umbrella of BTI.

Glycan binding protein specificity Guidelines

- JP has identified a number of scientists interested in this topic
- Feedback from Lara: she's received minor modifications from Ten but did not have any time to look into it. She's published a ML-based set of annotations for the commercially available lectins and called it guide to lectin binding (https://doi.org/10.1021/acschembio.1c00689).
- it's a different kind of guide as we have in mind….

ACTION POINTS
- CK: Lara is requested to provide the guidelines in an editable format (not PDF).
- Kiyoko will meet with Lara, Jim and Hiroaki in September. Good chance to restart the discussion.

Promotion

Journals

Nature - glycan array guidelines?

PNAS, JBC, Science, JACS, PLOS?
No news from these journals.

Manufacturers

We were in contact with some manufacturers in the past but the outcomes were very vague. So, we'd need to define what we'd expect from the manufacturers and what we can offer. Why should the manufacturers be interested in adopting the MIRAGE guidelines? What would be the benefit for both sides?

DISCUSSION
- approach software producers rather than hardware producers (e.g. Bionic, PEAKS Glycanfinder https://www.bioinfor.com//)
- Joe: bioinformatics solutions (CAN): adapt software for proteomics, recently issued a glycoproteomics version, i.e. PEAKS,
- Cathy: they should be involved before guidelines and templates (seems to be too late now!)
- William: glycoproteomics guidelines can be suggested their software provider

Data exchange and ontologies

mzIdentML (J. Klein and J. Zaia)

- mzIdentML: The PSI has started a new process for mzIdentML to support both the glycopeptide and cross-linking extensions. Right now, they're reasonably happy with the shape things are in, I just need to rewrite my proposal in the new format. I was also asked at the last PSI workshop to check how PSI-MOD and UNIMOD modification controlled vocabularies (CV) should be interpreted so that glycans from all three controlled vocabularies are handled correctly, but this seems like a synonym handling issue that we can solve either in those CVs or in GNOme itself before we devolve to counting monosaccharides.
- There are notes from Nathan Edwards I want to have a more principled discussion about with Andy Jones, but finding time for writing out the discussion points is difficult right now. Nathan is of the opinion that we should minimize the number of new PSI-MS CV terms we create to support this work, but I am introducing several concepts to handle different search strategies that may reside inside the same software, which translate to different meanings for different scores. Andy has said in the past that there are generic terms we could use, but during the workshop we found some ambiguity in defining the scope of those generic terms.

Metadata

Ontology for glycan data on GitLab (https://gitlab.com/m708/mirage-ontologies). The site is now open for any interested in this project, just log on the GitLab site.

GlycoPOST and UniCarbDR follow glycan ontology and allow entering terms according to OLS for other terms

Save the date(s)

4th AustralAsian Glycoscience Symposium and 9th Warren Workshop for Glycoanalytics - Sea World Resort, Gold Coast, QLD, 22-25th of November 2022, http://ags-2022.w.kamevents.currinda.com/

Beilstein GlycoBioinformatics Symposium 2023 - Unparalleled Diversity and Functionality of the Glycome. June 20-22, 2023, Limburg (Germany). Scientific Organizers: C. Creuzenet, E. Fadda, B. Schumann, C. Kettner.
Applications for presentations from the Commission are welcome!
https://www.beilstein-institut.de/en/symposia/glyco-bioinformatics/

14th MIRAGE Meeting: June 19, 2023, Limburg. For accommodation and travel expenses, there will be the same regulations in place as for the previous MIRAGE meetings that take place directly related to the Beilstein Symposium.
https://www.beilstein-institut.de/en/projects/mirage/meetings/

Joint Notes

- review MIRAGE website Recommendation for repositories and databases as many of the entries are not related with MIRAGE (but glyco-related) or are not suitable to submit data which is misleading in the context of recommending sites for data. However, a central page for any users (submitters and data users) would be helpful for the community. A clear guidance for tools and repositories is required

- **Preference/ACTION POINT**: on guidelines pages: provide additional info on tools and repositories

- the vast number of acronyms and logos for repositories, databases and tools seems to be very confusing. Therefore, a compendium of acronyms (UniCarDR, -KB, ….) would be helpful. Maybe this is info that could be provided on the MIRAGE project website?
- How many visitors do we see at the MIRAGE project webpages?
- what can be/must be done to make the experimental data readily available for data analysts and bioinformaticians?

GWS Parser

- takes as input a GlycoWorkBench Structure (GWS) file - a semi-colon separated list of structures in gws format
- output is an EXCEL file with GWS string, GlycoCT string, GlyConnect ID, GlyTouCan_composition ID (if available) and GlyTouCan_Structure ID (if available)
- Windows executable is available at https://bitbucket.org/kate1711/gwsparser/
  - Click on file
  - Click on "view raw"
  - You will be asked to download the file
- For MAC or Linux users, please email me (catherine.hayes@unige.ch) & I can share the scripts
- **IMPORTANT** - Composition output is not complete (it is missing some types of residues) - if you find that the output is incorrect please contact me & I will input the missing residues
- **IMPORTANT** - sometimes the converter throws a "GlycoCT validation Error" - if this happens please sent me the relevant structure for testing