Mining enzyme data in UniProtKB using Rhea - streamed

Trainers: **Anne Morgat, Marie-Claude Blatter, Parit Bansal, Elisabeth Gasteiger**

Please use this document to exchange information/questions with the instructors and organizer of the course.
It can be any type of question (from “Why do you use ChEBI?” to “Which button is it again?” to “I can’t hear, please speak up”) and one of our instructors will answer them.

Important Links

- Rhea website: https://www.rhea-db.org/
- UniProt website: https://www.uniprot.org/
- Course material: http://education.expasy.org/cours/SIB_Rhea_UniProt_2022/

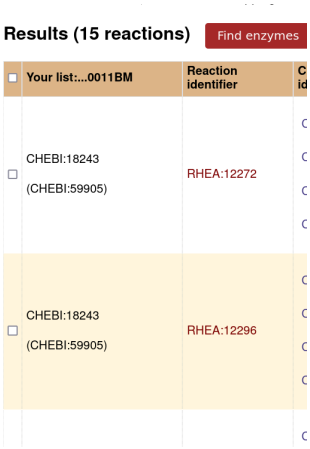
March 4 2022 - Practicals

Q: How to extract annotation information (e.g. Catalytic activity, GO function, PTM, domains) through API? (Hao Wang)
A: Thanks. I think we answer this one once Elisabeth presents the solution.
https://www.uniprot.org/help/api
https://www.uniprot.org/uniprot/?query=gene_exact:gba&format=tab&limit=10&columns=id,entry%20name,reviewed,comment(CATALYTIC%20ACTIVITY)&sort=score (Hao: thanks for the answer)

Q: It seems there could be multiple ids (in different protonation states) existing in ChEBI for a given chemical. How to quickly figure out which id (pH = 7.3) was actually used by Rhea? (Hao Wang)
A: *Does the below explanation answer your question Hao?* **(Hao: yes indeed)**
This is exactly what was in the last part. So, if you take CHEBI:18243 and perform a “Id-Mapping” functionality. In the result table you will get Rhea ids and the chebi’s used in Rhea for Chebi:18243.

For Reference:
https://www.rhea-db.org/rhea/?query=job:M20220304D4A9756708BA4FF95ED4AEC2A0276BA

300011BM&columns=yourlist(M20220304D4A9756708BA4FF95ED4AEC2A0276BA300011BM),rhea-id,chebi-id,equation,enzyme-class,uniprot

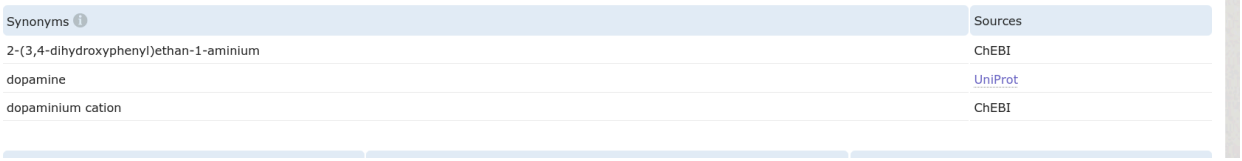


Also, On the chebi page of the molecule, if it has a uniprot synonym it is mostly used in Rhea.

Ref: https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:18243

No Uniprot Synonym hence not used in UniProtKB/Rhea

https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:59905
UniProt Synonym hence used in UniProtKb/Rhea



Similarly, if you perform an id mapping using CHEBI:18243 (dopamine) you'll get **chebi:18243 (CHEBI:59905) - major species 7.3 used in Rhea.**
But you don't have to bother about protonation states, Rhea does it for you

Q: Sorry for being so naive, but what is the difference between InChi-key and SMILES? Thanks
A: There could be multiple SMILES for a molecule but Inchi is unique. Both are textual

representations for the molecule. Just different ways. Therefore the chance of finding your molecule correctly across databases is high if you use Inchikey for search compared to the SMILES string.

February 24 2022 - Lecture

Course description: https://www.sib.swiss/training/course/20220224_RHEAL

Ask your questions here!

Q: How does Rhea find similar reactions? By EC? (Aaron)
A: On the reaction page there is a section “Related reactions”, about more general form or more specific form of reactions. That’s done “manually” by biocurators. So, inherently there is a “is_a”/hierarchical relationship among Rhea reactions. I think that’s what the question is about.
You can also play with similar compounds (Structural search: substructure or similarity).
It is in our to-do list to provide search for similar reactions as a whole (so coming soon, hopefully).

Q: How can we search the effect of one small molecule on a specific enzyme, such as the effect of anisomycin (CHEBI:338412) on p38alpha enzyme (UniProtKB - Q16539)? Can we also search the selected effects (activation) of multiple molecules on the same enzyme (p38alpha)? How can we also find the activator molecules of an enzyme on Rhea? (Name: Murat Delman)
A: question for UniProt later
You can use ChEMBL (https://www.ebi.ac.uk/chembl/) to find chemicals that inhibit/activate enzymes. ChEMBL provides links to CHEBI and UniProt

Q: For the accession ID/mapping service in Uniprot, Pfam accessions have been removed now. How can I get Pfam mappings from Uniprot now? Thanks (Name: Kumar Saurabh Singh)
A: See https://www.uniprot.org/help/uploadlists, last remark. The IDmapping service favours mappings that are more or less 1:1, not 1:hundreds or even 1:thousands (for performance reasons). You could try to address the issue programmatically, or try queries like *Pfam:PFxxxxx or Pfam:PFyyyyy*. Don’t hesitate to contact us at the helpdesk (https://www.uniprot.org/contact) if you would like to discuss a specific use case, in more detail.
Example URL:
https://www.uniprot.org/uniprot/?query=database:(type:pfam%20pf08563)%20OR%20database:(type:pfam%20pf00870)%20OR%20database:(type:pfam%20pf07710)&format=tab&columns=id,entry%20name,database(Pfam)&sort=score
Entry Entry name Cross-reference (Pfam)

Q8SPZ3	P53_DELLE	PF00870;PF08563;PF07710;
Q00366	P53_MESAU	PF00870;PF08563;PF07710;
P02340	P53_MOUSE	PF00870;PF08563;PF07710;
P10361	P53_RAT	PF00870;PF08563;PF07710;
Q9TTA1	P53_TUPBE	PF00870;PF08563;PF07710;PF18521;
Q9WUR6	P53_CAVPO	PF00870;PF08563;PF07710;
P67939	P53_BOVIN	PF00870;PF08563;PF07710;
P56423	P53_MACFA	PF00870;PF08563;PF07710;PF18521;
P56424	P53_MACMU	PF00870;PF08563;PF07710;PF18521;
Q95330	P53_RABIT	PF00870;PF08563;PF07710;

…..

Q:
A:

3 March 2022 - Practicals

Course description: https://www.sib.swiss/training/course/20220303_RHEAP

Feedback form: https://edu.sib.swiss/course/view.php?id=574 (please use username and pw sent in mail with practical information)

Ask your questions here!

Q:
A:

Q:
A:

Q:
A:

BREAK -> back at

Q:
A:

Q:
A:

Q:
A:

4 March 2022 - Practicals

Course description: https://www.sib.swiss/training/course/20220304_RHEAP

Feedback form: https://edu.sib.swiss/course/view.php?id=574 (please use username and pw sent in mail with practical information)

Ask your questions here!

Q: is the total entry for "Rhea- Annotated reactions database" https://www.rhea-db.org/rhea/13269 will be 4,919 proteins + 1,941 proteins?
A:

Q:
A:

Q:
A:

BREAK -> back at

Q:
A:

Q:
A:

Q:
A:

End of Training

Put the questions you have after the course here:

Q:
A:

Q:
A:

Q:
A: