



# RDKit UGM 2021

virtual....

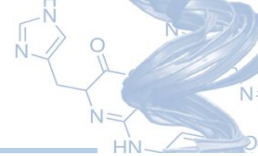
[https://github.com/rdkit/UGM\\_2021](https://github.com/rdkit/UGM_2021)

<https://discord.gg/WXWjR8SQNH>

[!\[\]\(d66ff64371a51729ac8c1cdaa685ba6f\_img.jpg\) #RDKitUGM2021](#)

# Introductions

---



## Who am I?

Greg Landrum @dr\_greg\_landrum

ETH Zurich

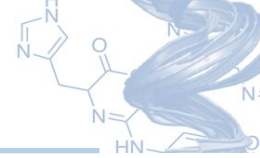
T5 Informatics GmbH

KNIME AG

Who are you?

# Introductions

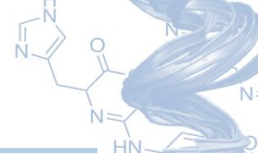
---



Who am I?

**Who are you?**

# Wow!



Manage my events

## Virtual 10th RDKit UGM 2021

Thu, Oct 14, 2021 8:00 AM

✓ Basic Info

✓ Details

3 Online Event Page

## Tickets

Admission

Add-ons

Promo codes

Holds

Settings

Add tickets



### General Admission

• On Sale • Ends tomorrow at 7:00 PM

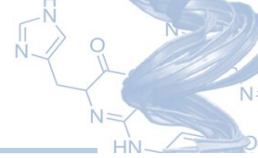
771 / 850

Free



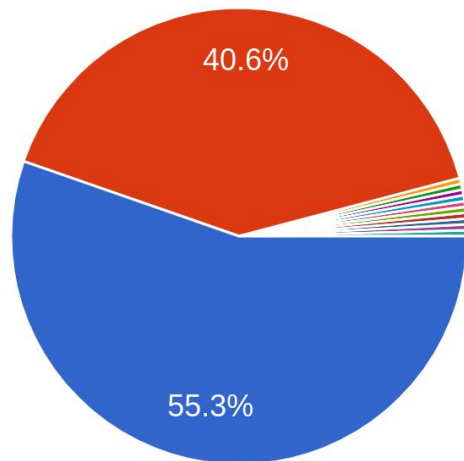
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



Where do you work?

244 responses

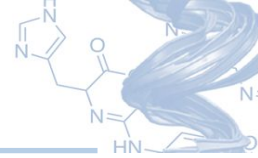


- Industry
- Academia
- Freelancer
- Consultant working with academia and industry
- Graduate Student / Self-Employed
- Both (Full-time industrial + part-time P...
- Government research center
- Both, actually. I work for a commercial...

▲ 1/2 ▼

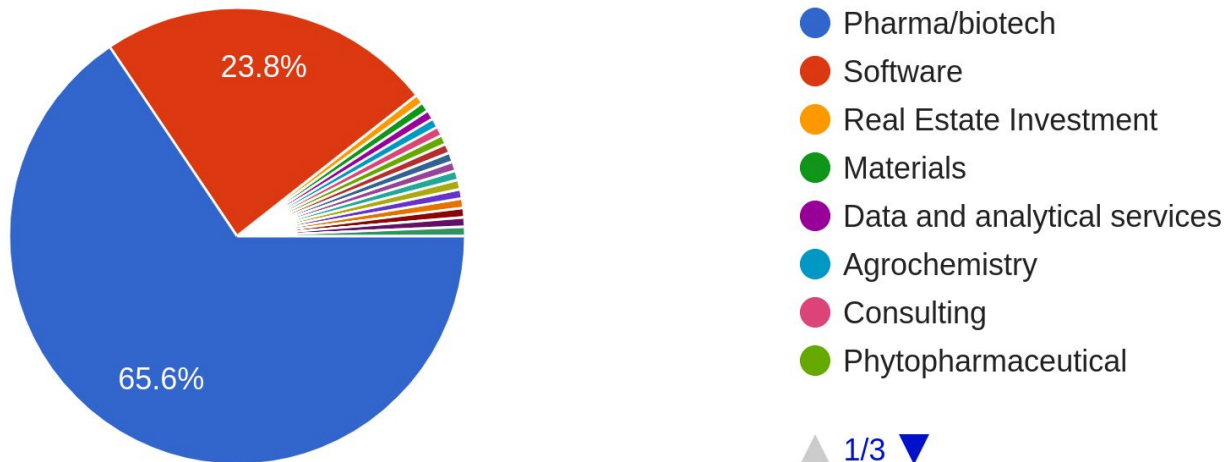
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



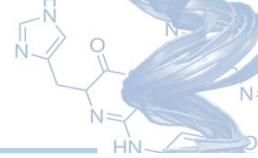
If you're in industry, what kind?

151 responses



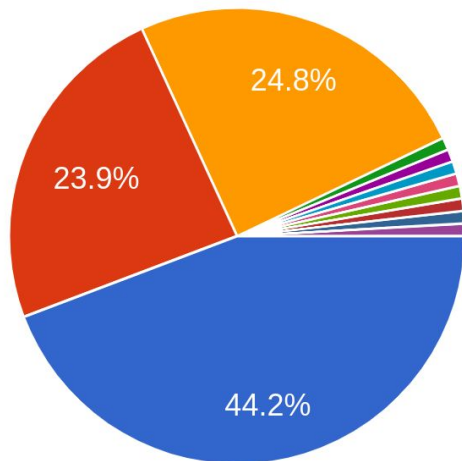
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



If you're in academia, what do you do?

113 responses

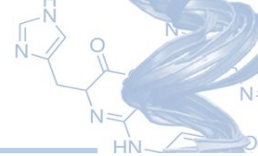


- Student
- Postdoc
- Faculty/Staff
- Head of AI and Cheminformatics
- Consultant but I do support a MRes c...
- Researcher, unit head
- Ph.D. student
- Visiting Scientist

▲ 1/2 ▼

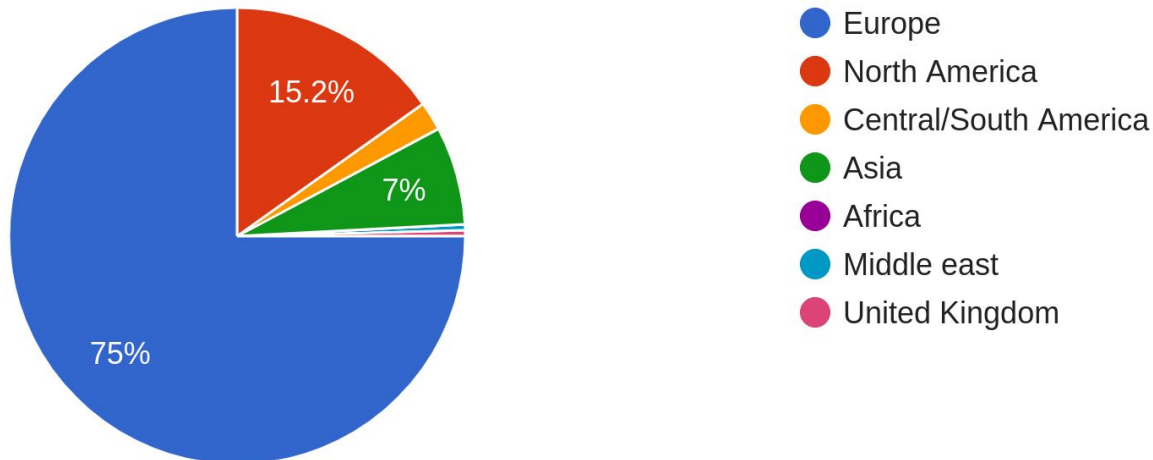
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



Where are you?

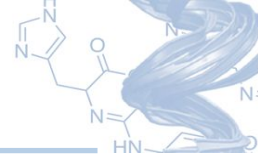
244 responses





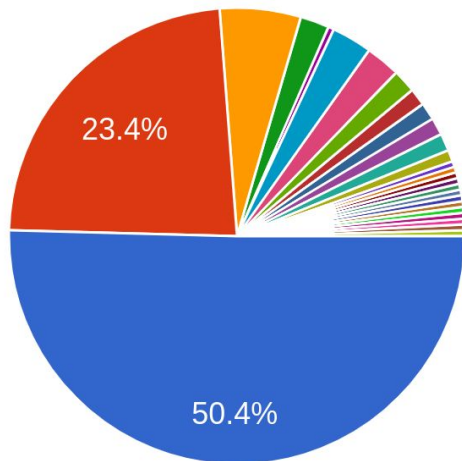
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



How many RDKit UGMs have you attended?

244 responses

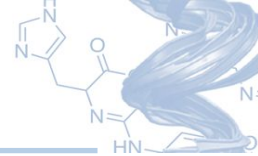


- This is my first one!
- Three or more
- Five or more
- All of them except one
- All of them
- 1
- 2
- one

▲ 1/4 ▼

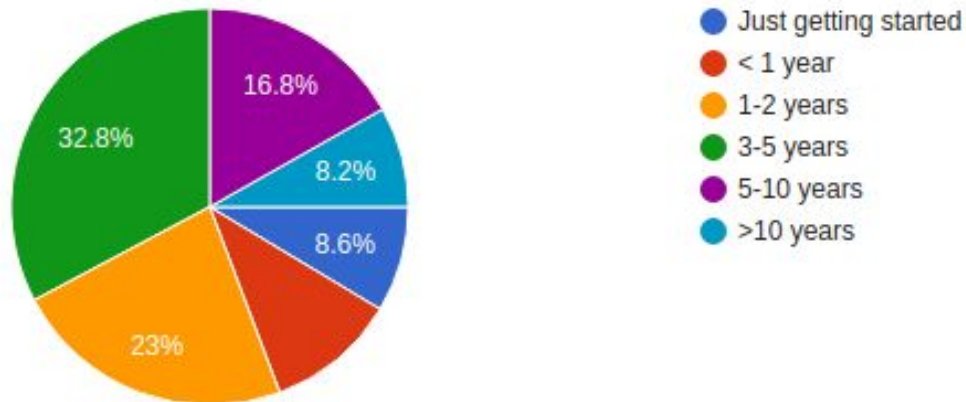
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



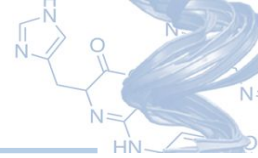
How long have you been using the RDKit?

244 responses



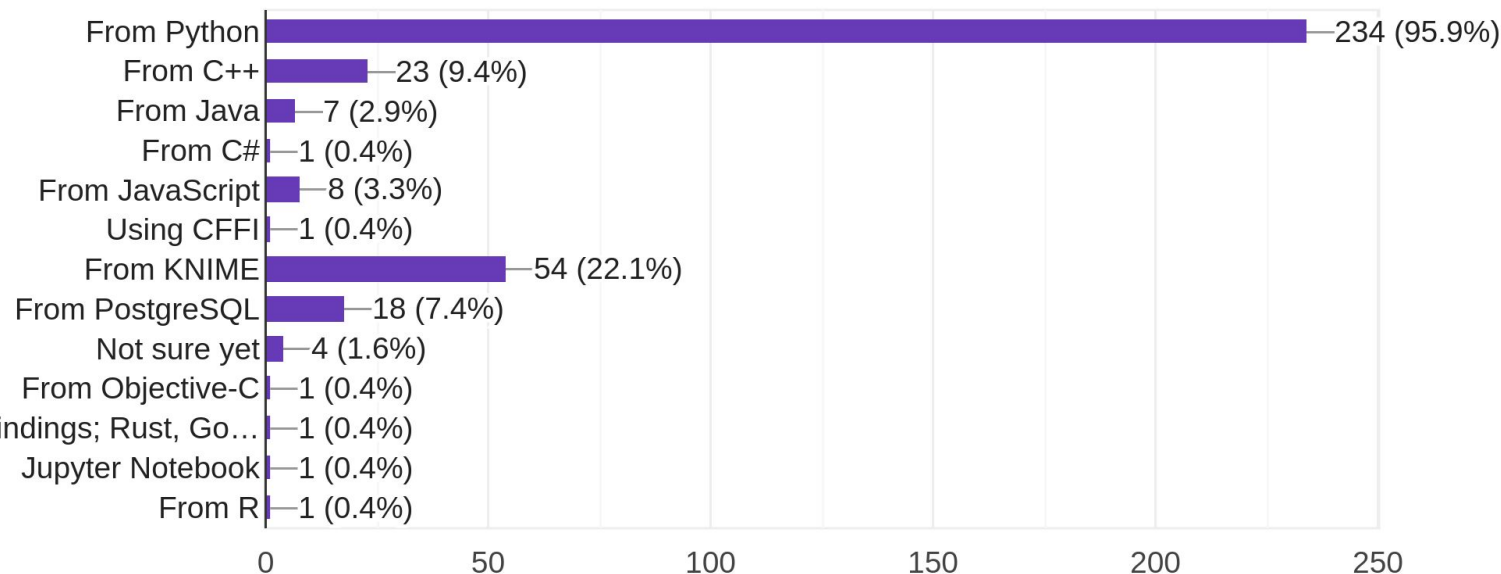
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



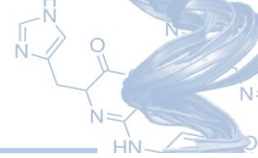
## How are you using the RDKit?

244 responses



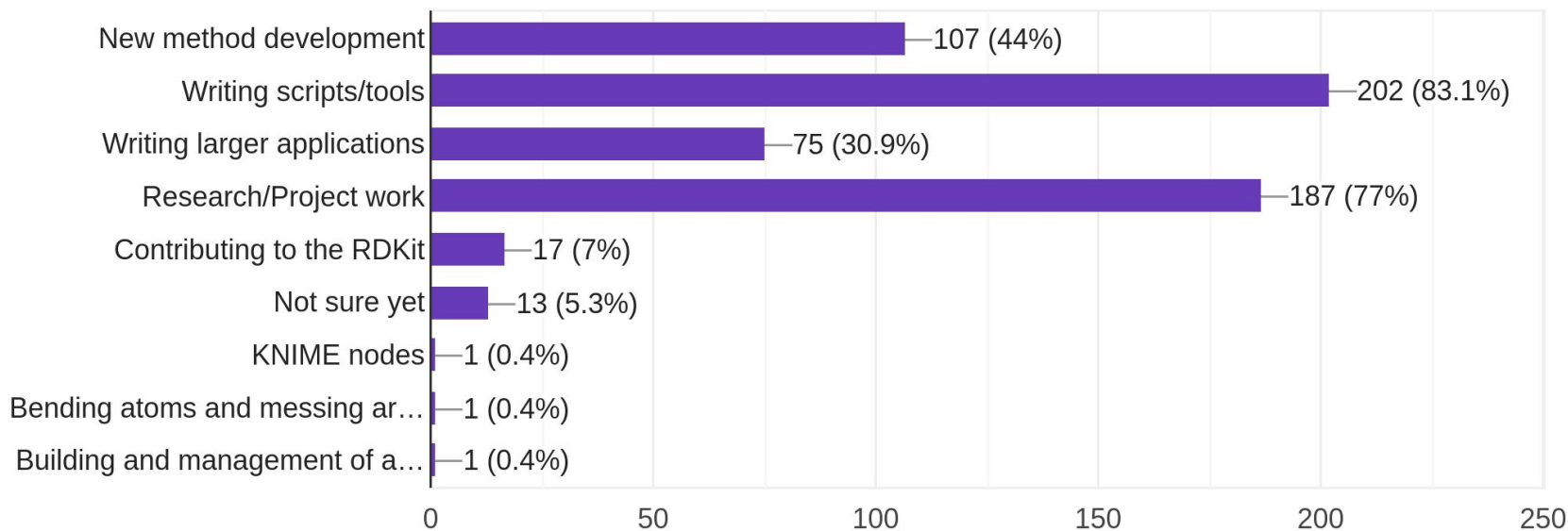
# Who are you?

<https://tinyurl.com/RDKit-survey2021>



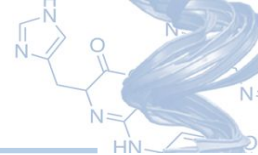
## What are you doing with the RDKit?

243 responses



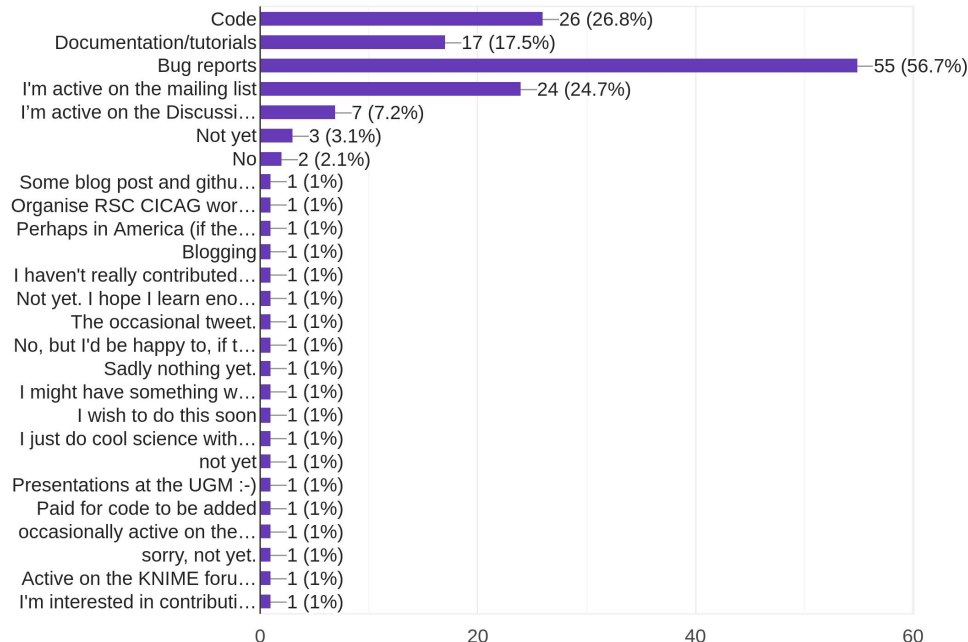
# Who are you?

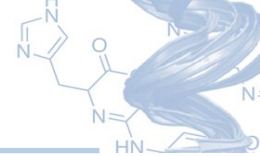
<https://tinyurl.com/RDKit-survey2021>



Have you contributed to the RDKit and/or the community?

97 responses





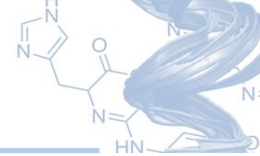
## Aside: Past meetings/organizers

---

1. London (ICR): Nathan Brown
2. Hinxton (EMBL): George Papadatos
3. Darmstadt (Merck): Paul Czodrowski
4. Zurich (ETH): Sereina Riniker
5. Basel (Novartis/Roche): Nadine Schneider, Nik Stiefl, Christian Kramer
6. Berlin (Charité Berlin/FU Berlin): Andrea Volkamer, Gerhard Wolber
7. Cambridge (Cambridge University): Andreas Bender
8. Hamburg (Uni Hamburg): Emanuel Ehmki
9. Virtual: Floriane Montanari, Djork-Arné Clevert (Bayer)

# “Frequent Hitters”

---

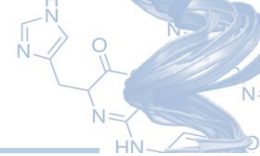


- Ten UGMs:
  - Roger S
  - Grégori G
- Nine UGMs:
  - Sereina R
  - Paul C
  - Noel

If you should be on here and aren't: please  
let me know!

# About the meeting

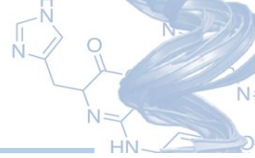
---



- Friendly
- Diverse
  - People
  - Topics
  - Degrees of technicality
- By/for the community
  - Organized and run by volunteers

Please join the discord chat: <https://discord.gg/WXWjR8SQNH>



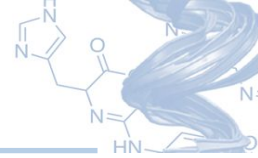


---

# **A quick aside about toolkits**

# What *is* a toolkit anyway?

---



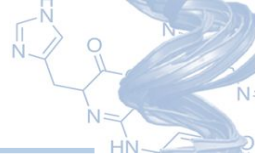
A collection of tools for building things

some simple

some not

you'll use some all the time

some you'll never use and may not even  
know about



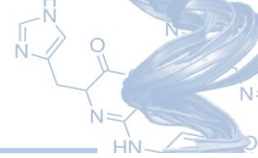
---

**Next topic: community**

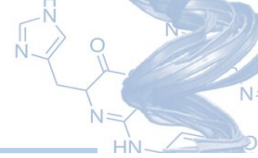
# Community

---

The heart of any  
*successful* open-source  
project



# Community



- Mailing lists: ~600 messages to rdkit-discuss from 2020.10.11- 2021.10.12
- Google scholar: >1200 hits for "rdkit" in 2020, >1400 so far in 2021
- Searching github for `from rdkit import Chem` returns >30000 code results
- Each of the last eight UGMs at capacity with 40-100+ attendees

## Virtual 10th RDKit UGM 2021

October 14, 2021 at 8:00 AM CEST

Online event

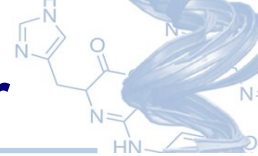
\$0.00 Gross

● On Sale

754 / 850

# Contributions to github issue tracker in the last year

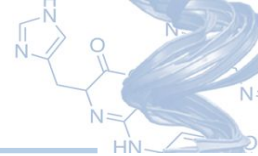
---



greglandrum jasondbiggs bp-kelley ricrogz ptosco UnixJunkie IchiruTake mbanck  
yurivict tdudgeon paconius gmseabra evansenter e-kwsm bzoracler SPKorhonen  
MichelML ErikCVik zhandn ys0123 wopozka wjm41 sirbiscuit shayakhmetov  
sargelavoie ryszard314159 rvianello rmmg rexdwyer rd0102 qq87258580 pschwillr  
nicogam nbehrnd nanhuayu nadzhou mwojcikowski msudhanshu10 mrcblt mjlw99  
mily33 mayankBIL matthewyclark malteseunderdog maksbotan macraild ljn917  
likhangy kienerj khoivan88 kheyer kekoburgo kazuyaujihara kapilaGIT kangway  
jw-feng jones-gareth jinpan jhjensen2 iwatobipen imxj i-tub hsiaoyi0504 hgarrereyn  
hadim goldwind-ting ghiander eloyfelix eguidotti d-b-w chrisbutch chmnk  
chaoyan1037 cchang373 cbouy brilee bokertof autodataming andt88 amintavakol  
adelenelai ZontaNicola Zax0NL XFaIT RobinFrcd Plancalkuele MahaveerSatra  
LivC182 LanceKnight KramerChristian JinghanH IngvarLa DavidACosgrove  
ChinzoD AustinApple Andy-Wilkinson AIDossetter

That's 97 different people

# Community (update)




⋮

#

General

Welcome to RDKit Discussions!

 greglandrum

New

Top: All ▾

Answered

Unanswered

Label ▾

New discussion


Q


Search all discussions


Categories


✎


∞ View all

 Development

 General


 Ideas


 Q&A

 Show and tell


Most helpful


↑ 1

 Bug in InChI parsing?  
simonmb asked 23 hours ago in Q&A · Answered


 4


↑ 1

 Is this a SMARTS bug? Trying to find a "S" atom with at least one hydrogen  
simonmb asked 23 hours ago in Q&A · Answered


 2


↑ 0

 Explicit Hs in a SMILE do not get an atom Id  
chymotrypsin13 asked 6 days ago in Q&A · Unanswered

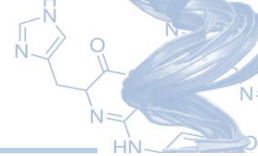
 1

↑ 1

 Wavy Bonds on unlabeled chiral centers  
sudoPete asked 4 days ago in Q&A · Unanswered

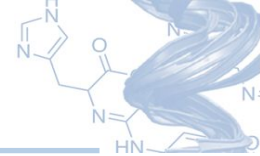
 1

# Usage in other open-source projects (updated 2021)



- Shape-IT - shape-based alignment
- ChEMBL Structure Pipeline - ChEMBL protocols used to standardise and salt strip molecules.
- FPSim2 - Simple package for fast molecular similarity searches.
- DockOnSurf - high-throughput code to find stable geometries for molecules on surfaces
- <https://datamol.io/> - A Python library to intuitively manipulate molecules.
- Scopy - Python library for desirable HTS/VS database design
- stk (docs, paper) - a Python library for building, manipulating, analyzing and automatic design of molecules.
- OpenFF - Open source approach for better force fields
- gpusimilarity - GPU implementation of fingerprint similarity searching
- Samson Connect - Software for adaptive modeling and simulation of nanosystems
- mol\_frame - Chemical Structure Handling for Dask and Pandas DataFrames
- mmpdb 2.0 - matched molecular pair database generation and analysis
- CheTo - Chemical topic modeling
- OCEAN - web-tool for target-prediction of chemical structures which uses ChEMBL as datasource
- Coot - software for macromolecular model building, model completion and validation
- DeepChem - deep learning toolkit for drug discovery
- sdf2ppt - Reads an SDF file and displays molecules as image grid in powerpoint/openoffice presentation.
- chemfp
- PYPL - Simple cartridge that lets you call Python scripts from Oracle PL/SQL.
- WONKA - Tool for analysis and interrogation of protein-ligand crystal structures
- OOMMPPAA - Tool for directed synthesis and data analysis based on protein-ligand crystal structures
- chemicalite - SQLite integration for the RDKit
- django-rdkit - Django integration for the RDKit
- ... more ...



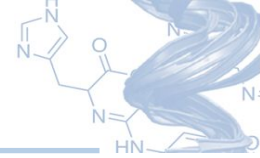


# Usage in commercial tools

---

- Cresset Software
- Dalke Scientific Software
- NextMove Software
- Schrödinger
- SCM
- Wolfram Research

Disclaimer: this info is from public statements made by people from those companies.  
I almost certainly have forgotten someone



# Usage in online tools/resources

---

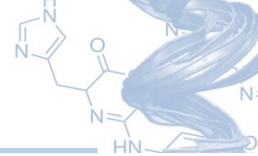
- ChEMBL
- ZINC
- Google Patents
- PDBe
- Enamine
- TeachOpenCADD

Disclaimer: this info is from public statements made by people associated with those projects. I almost certainly have forgotten someone

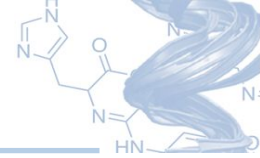
# Support

---

- Web searches
- Mailing list
- Github discussions
- Commercial support



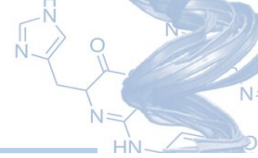
# Sustainability: the bus problem



[https://commons.wikimedia.org/wiki/File:Postauto\\_susten.jpg](https://commons.wikimedia.org/wiki/File:Postauto_susten.jpg)

# Sustainability: the bus problem

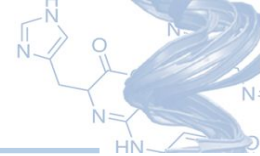
---



RDKit maintainers:

- Greg
- Brian Kelley (Relay Therapeutics)
- Ricardo Rodriguez (Schrödinger)
- Paolo Tosco (Novartis)

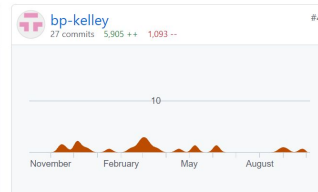
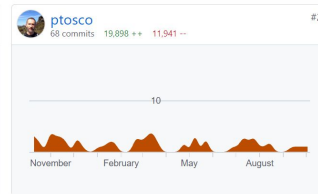
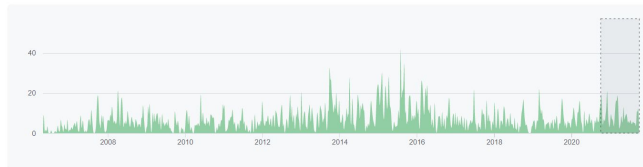
# Most frequent code contributors in the last year



Oct 5, 2020 – Oct 6, 2021

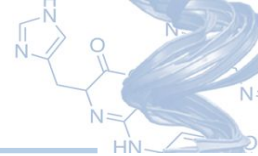
Contributions: Commits

Contributions to master, excluding merge commits and bot accounts



# Roadmap

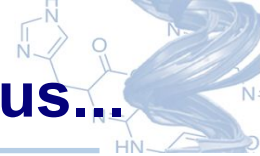
---



Future work tends to be determined by what's needed for active projects or requests that come out of the community. So there's not much of a roadmap.

# Still, some parts of the way forward are pretty obvious...

---



Making sure all the pieces required  
to build a good compound  
registration system are there

Making sure all the pieces required  
to build a good corporate chemical  
database are there

Better support for polymers and  
organometallics

Performance improvements

Ongoing refactoring and code  
cleanup



# Enjoy the meeting!

---

