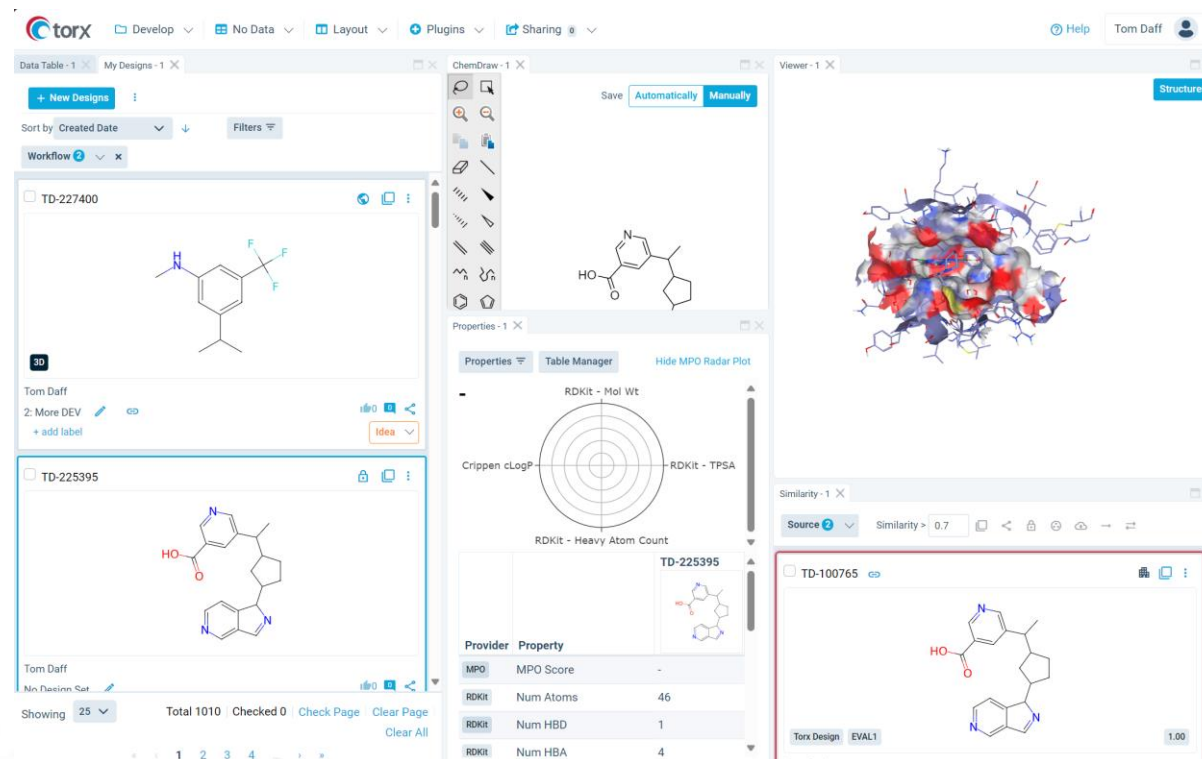




RDKit on RDS

Tom Daff
Web Development, Torx

- Torx is a small molecule DMTA platform
- RDKit is used:
 - > Structure aware database storage and searching from Django
 - > Properties
 - > Structure rendering
- Users typically work with hundreds to thousands of structures at any time
- “Cloud native” software-as-a-service



The screenshot displays the Torx DMTA software interface. The main window shows a workflow view with two chemical structures: TD-227400 and TD-225395. The interface includes a sidebar with navigation options like 'New Designs', 'Filters', and 'Workflow'. The central panel shows the chemical structure of TD-225395, which is a complex molecule with multiple rings and functional groups. The right sidebar contains a 'Properties' panel with a 'Table Manager' and a 'Hide MPO Radar Plot' option. Below this is a 'Similarity' panel showing a similarity score of 0.7. The bottom right corner features a 'Torx Design' panel with a 'TORX EVAL1' button.

Provider	Property	Value
MPO	MPO Score	-
RDKit	Num Atoms	46
RDKit	Num HBD	1
RDKit	Num HBA	4

Dockerfile 3.08 KiB

Blame Edit Lock Replace Delete

```
1 # Build of RDKit database cartridge into the official Postgres Docker
2 # image. Modified in the Torx fork to use specific versions of Postgres
3 # and RDKit. Debian preferred as Alpine reduces structure search speed
4 # by ~30%.
5 ARG RDKIT_VERSION=Release_2023_09_5
6 ARG POSTGRES_VERSION=14.11-bookworm
7
8 FROM postgres:${POSTGRES_VERSION} AS rdkit-postgres-build-env
9
10 # hadolint ignore=DL3008
11 RUN apt-get update \
12     && apt-get install -y --no-install-recommends \
13         build-essential \
14         ca-certificates \
15         cmake \
16         git \
17         libboost-dev \
18         libboost-iostreams-dev \
19         libboost-serialization-dev \
20         libboost-system-dev \
21         libeigen3-dev \
22         postgresql-server-dev-14 \
23         wget \
24         zlib1g-dev \
25     && apt-get clean \
26     && rm -rf /var/lib/apt/lists/*
27
28 # Refresh arg from outer scope
29 ARG RDKIT_VERSION
30 RUN wget --quiet https://github.com/rdkit/rdkit/archive/${RDKIT_VERSION}.tar.gz \
31     && tar xf ${RDKIT_VERSION}.tar.gz \
32     && mv rdkit-${RDKIT_VERSION} rdkit \
33     && rm ${RDKIT_VERSION}.tar.gz \
34     && mkdir /rdkit/build
35 WORKDIR /rdkit/build
36
37 # Disabled Freetype as it fails to configure.
38 RUN cmake -Wno-dev \
39     -D CMAKE_BUILD_TYPE=Release \
40     -D RDK_BUILD_FREETYPE_SUPPORT=OFF \
41     -D RDK_BUILD_INCHI_SUPPORT=ON \
42     -D RDK_BUILD_PYTHON_WRAPPERS=OFF \
43     -D RDK_BUILD_PGSQL=ON \
44     -D PostgreSQL_ROOT=/usr \
45     -D PostgreSQL_TYPE_INCLUDE_DIR=/usr/include/postgresql/14/server/ \
46     .. \
47     && make -j "$(nproc)"
48
49 # Run the tests
50 ENV RDBASE=/rdkit/
51
52 # Postgres tests need a working database and write access
53 # for the postgres user to the test directory
54 RUN cp /rdkit/build/Code/PostgreSQL/rdkit/rdkit--*.sql /usr/share/postgresql/14/extension \
55     && cp /rdkit/build/Code/PostgreSQL/rdkit/rdkit.control /usr/share/postgresql/14/extension \
56     && cp /rdkit/build/Code/PostgreSQL/rdkit/librdkit.so /usr/lib/postgresql/14/lib/rdkit.so \
57     && chmod 777 /rdkit/build/Code/PostgreSQL/rdkit/ \
58     && chmod 777 /rdkit/build/Testing/Temporary/
```

Extensions supported for RDS for PostgreSQL 16

The following table shows PostgreSQL extensions for PostgreSQL version 16 that are currently supported on Amazon RDS. For more information on PostgreSQL extensions, see [Packaging related objects into an extension](#).

Extension	16.4	16.3- R2	16.3	16.2- R3	16.1- R4
rdkit	4.5.0 (Release 2024_03_5)	4.4.0	4.4.0	4.4.0	4.4.0

Tuning performance

- Noticed when testing a “security” update to move from Debian to Alpine containers there was an over 20% *decrease* in performance
- Highlighted that our current cloud setup is not well optimised
- Huge choice of options in AWS

Instance types (1/760+) [Get advice](#)

DB instance size

☐ Production

db.r6g.xlarge
4 vCPUs
32 GiB RAM
500 GiB
1.180 USD/hour

☐ Dev/Test

db.r6g.large
2 vCPUs
16 GiB RAM
100 GiB
0.269 USD/hour

☒ Free tier

db.t3.micro
2 vCPUs
1 GiB RAM
20 GiB
0.023 USD/hour

- Loosely followed the 10M compound benchmarks on the RDKit blog

	Self-managed	Managed small	Managed large	Managed xlarge
	t3a.medium	db.t3.micro	db.m5.large	db.m6g.xlarge
Cost / month	£17	£10	£140	£200
Create, index and fingerprint	4 hr		2 hr	
Similarity / s	32	29	18	18
Substructure / s	2.9	3.5	2.8	3.3

- Cheapest managed option better than untuned self-managed
- Databases scale better for concurrency than vertically for the cost
- Next up: “serverless” and specialised systems