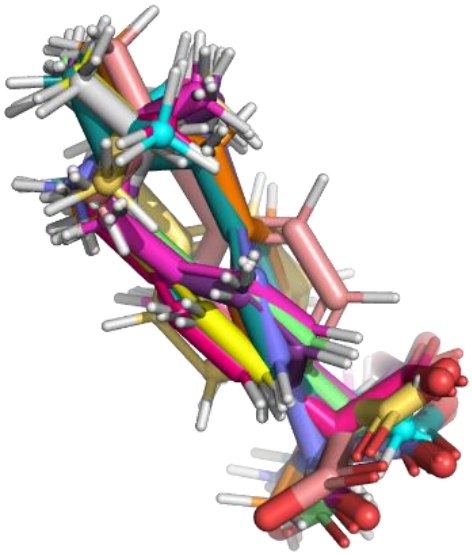
QUANTUM FUTURE
Scientific Software

Web: *QFSciences.com*
Support: Support@qfsciences.com
CEO: Laszlo@qfsciences.com
Phone: 505-570-1150

Introducing QuantumFuture's Scientific Database for Drug Repurposing

Using FDA Approved Drugs and Natural Compounds



1. Extensive Conformational Library: Explore low-energy conformations of approximately 1600 FDA-approved drugs and about 3900 natural compounds. All conformations are ranked with accurate DFT-D4 revSCAN, def2-TZVP *ab initio* calculations.

2. Thermodynamic Insights: Access comprehensive quantum mechanical thermodynamic calculations for each drug molecule's conformations as well as for the conformational ensembles.

3. The database can be used in various ways in drug repurposing projects. One example is screening for new protein targets with a specialized docking program, utilizing the precomputed and physically sound conformations, and considering accurate *ab initio* DFT-D4 based conformation strain energies in the docking hit selection processes.

4. Accessibility and Appreciation: Enjoy complimentary access for special QF customers and Microsoft-affiliated research groups, courtesy of Azure credits. A nominal fee is charged for others to help cover our operation expenses.

5. 2024 Software Release: Unlock new drug possibilities with our 2024 QF software, capable of handling similar projects involving tens of thousands or even hundreds of thousands of private drug-like compounds.

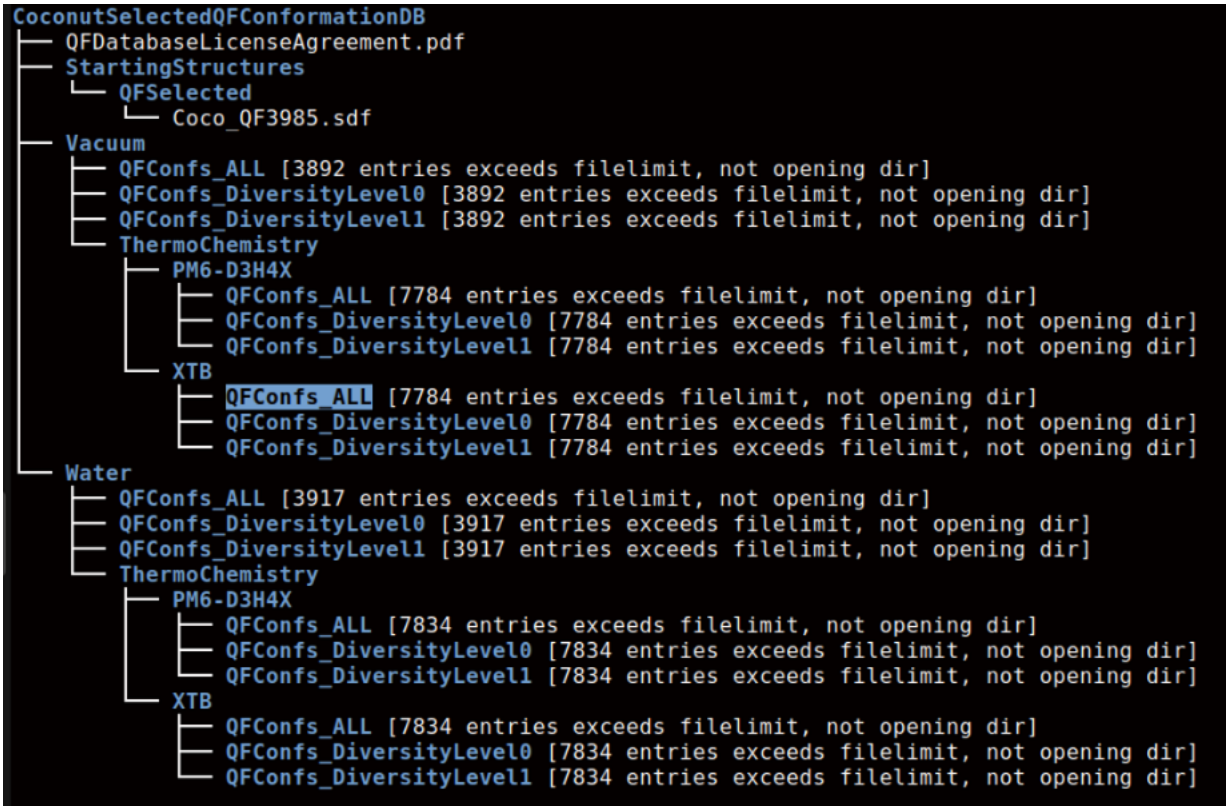
6. Project Outsourcing: For those seeking to outsource similar projects, we offer our expertise and capacity to meet your research needs.

Details of the QF Drug Repurposing Conformational Database

The summary of the data structure can be shown conveniently with the tree command under Linux by using the –filelimit option which indicates how many files each directory has. The summary is shown below for the FDA approved drug database in the picture and a more detailed explanation follows.

```
FDAApprovedQFConformationDB
├── StartingStructures
│   ├── Downloaded
│   │   └── e-Drug3D_2083.sdf
│   └── QFSelected
│       └── FDA_QF1641.sdf
├── Vacuum
│   ├── QFConfs_ALL [1600 entries exceeds filelimit, not opening dir]
│   ├── QFConfs_DiversityLevel0 [1599 entries exceeds filelimit, not opening dir]
│   ├── QFConfs_DiversityLevel1 [1599 entries exceeds filelimit, not opening dir]
│   └── ThermoChemistry
│       ├── PM6-D3H4X
│       │   ├── QFConfs_ALL [3198 entries exceeds filelimit, not opening dir]
│       │   ├── QFConfs_DiversityLevel0 [3198 entries exceeds filelimit, not opening dir]
│       │   └── QFConfs_DiversityLevel1 [3198 entries exceeds filelimit, not opening dir]
│       └── XTB
│           ├── QFConfs_ALL [3198 entries exceeds filelimit, not opening dir]
│           ├── QFConfs_DiversityLevel0 [3198 entries exceeds filelimit, not opening dir]
│           └── QFConfs_DiversityLevel1 [3198 entries exceeds filelimit, not opening dir]
└── Water
    ├── QFConfs_ALL [1610 entries exceeds filelimit, not opening dir]
    ├── QFConfs_DiversityLevel0 [1609 entries exceeds filelimit, not opening dir]
    ├── QFConfs_DiversityLevel1 [1609 entries exceeds filelimit, not opening dir]
    └── ThermoChemistry
        ├── PM6-D3H4X
        │   ├── QFConfs_ALL [3218 entries exceeds filelimit, not opening dir]
        │   ├── QFConfs_DiversityLevel0 [3218 entries exceeds filelimit, not opening dir]
        │   └── QFConfs_DiversityLevel1 [3218 entries exceeds filelimit, not opening dir]
        └── XTB
            ├── QFConfs_ALL [3218 entries exceeds filelimit, not opening dir]
            ├── QFConfs_DiversityLevel0 [3218 entries exceeds filelimit, not opening dir]
            └── QFConfs_DiversityLevel1 [3218 entries exceeds filelimit, not opening dir]
```

A similar tree structure for the natural compounds database is shown below.



1. **Input data preparation steps with python script using RDKit**

   We have downloaded the 2083 starting structures from https://chemoinfo.ipmc.cnrs.fr/MOLDB/index.php site and applied the so-called Weber filter allowing only maximum 10 rotatable bonds and limit the topological polar surface area values below 140 Angstrom^2. This filter has reduced the number of drug molecules to 1642. We ran our qfLowerLevel.x application to optimize the geometries just to make sure that we have valid quantities in all sdf files. (*This step is not strictly needed for the further steps*.) We have filtered out one more structure due to a sdf problem resulting in 1641 drug molecules. These sdf files were the starting inputs for this project.

   For the selection of natural compounds, we have started with the Coconut database which has over 400K natural compounds (as of January 2024). We have selected about 1% of the compounds for this database by utilizing the Lipinski, Ghose, Veber, Reos, and drug_like_filters and limiting the number of atoms to be in the 28 and 62 range. This selection resulted in 3987 compounds and 3985 of them had acceptable sdf files for our current sdf reader.

2. Finding the conformations with the qfconfsearchDFT.x application

We ran our qfconfsearchDFT.x application with default parameters that utilizes revSCAN functional with the new D4 dispersion correction parameters and def2-TZVP basis set in all DFT-D4 calculations. This functional/basis set combination has been found very accurate in both our intermolecular interaction benchmarks using S66x8 standard set as well as in our study to compare calculated vacuum conformations to high quality experimental geometries. For more details about those benchmark studies please look at our corresponding articles on LinkedIn. We have selected 1600 successful results in vacuum and 1610 results in aqueous continuum for the FDA approved drugs and 3892 and 3917 for the natural compounds in vacuum and in aqueous continuum respectively. The relatively small number of the rest of the calculations have shown some unwanted behaviors where the original bonds from the original sdf file have been changed during the QM semi-empirical geometry optimizations and we did not want to allow such drastic change in the structures since we are looking for conformations of the same molecules in this project and not looking to investigate tautomers or chemical reactions.

The QFConfs_ALL directories have all the results. All sdf files have all important conformations that we had found with our application and the DFT-D4 (revSCAN, def2-TZVP) energies as well as the relative energies compared to the lowest energy conformation (conformational strain energies) are attached as standard sdf tags. Note, that the numbering in the file names have no meaning whatsoever, it was just the order of the files that we used during the calculations. The sdf file for a given drug or natural compound can be located by grep the corresponding name or the CAS number.

3. Further deduplications of the conformations using the qfdeduplicate.x application

We have introduced two new directories under the name of QFConfs_DiversityLevel0 and the QFConfs_DiversityLevel1. Data files in those directories require some explanation. The qfconfsearchDFT.x application uses RMSD pair wise alignment-based deduplications via the C++ APIs of RDKit only for molecules with no more than 40 atoms by default. The problem is that we want to include hydrogen atoms during the RMSD alignments and deduplications because we do not want to lose important conformations where the orientation of the hydrogen atoms is different. This makes the RMSD calculations very expensive scaling very poorly in molecular size. We have some test examples showing that the RMSD based deduplications are more expensive than all *ab initio* DFT-D4 calculations. Therefore, we limit the RMSD based deduplications up to 40 atoms and we have developed an alternative scheme for larger molecules. This alternative scheme is new and very experimental at this point and with the current parameters it is on the conservative side which means that sometimes it keeps structures which are very close to each other. This conservative approach is obviously preferable since we can always eliminate some

duplicates later much easier than finding missing conformations that we had eliminated by mistake because that would require repeating some expensive calculations.

Thus, the QF_DiversityLevel0 and the QF_DiversityLevel1 directories have such post processing deduplicated conformations. We used 0.3 Angsrom RMSD limit in the QF_DiversityLevel0 directory and 0.6 Angstrom RMSD in the QF_DiversityLevel1 directory. Note, that if the energies of two conformations differ more than 0.3 Kcal/mol than we keep both regardless of the aligned RMSD difference in the geometries. Obviously, further deduplications or different deduplications starting from the conformations in the QFConfs_ALL directories are also possible, and users could do that easily by themselves. Unfortunately one structure failed to be processed so we have 1599 instead of 1600 and 1609 instead of 1610 structures for vacuum and in aqueous continuum respectively for the FDA approved drugs database.

4. **Going beyond *ab initio* DFT-D4 energies and strain energies by using qfLowerLevel.x and qfensemble.x applications.**

Besides accurate *ab initio* DFT-D4 energy orders and strain energies of the conformations, this database provides additional quantum mechanical information. We have performed statistical thermodynamic calculations with the GFN-XTB and the PM6-D3H4X quantum mechanical semi-empirical methods for all conformations of all drug molecules by using our qfLowerLevel.x application. Having all necessary thermochemistry quantities for all conformations we have also calculated the Gibbs free energies for the conformational ensemble of all drug molecules using our qfensamble.x program. The Gibbs free energies for the conformational ensembles are written to individual log file for each sdf files by adding a simple .log file extension to each sdf files since this value belongs to the whole conformational ensemble of the given drug molecules. For this reason, the number of files is doubled in all directories of thermochemistry calculations.

We wish all users as much enjoyment as we had in creating this database, and we hope for great success in your drug repurposing endeavors! Please share your success stories with us and don't hesitate to reach out if you have any questions!