




Windows version installation and quick run guide
Command line & GUI

OPERA Github repo:

 **NIEHS / OPERA**
forked from kmansouri/OPERA

Watch 2 Star 8 Fork 11

Code Pull requests 0 Actions Projects 0 Wiki Security Insights Settings


Free and open-source application (command line and GUI) providing QSAR models predictions as well as applicability domain and accuracy assessment for physicochemical properties, environmental fate and toxicological endpoints.
=====>Download the latest compiled version from the "releases" tab and run the executable installer.







[Manage topics](#)

176 commits 1 branch 0 packages **25 releases** 1 contributor MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

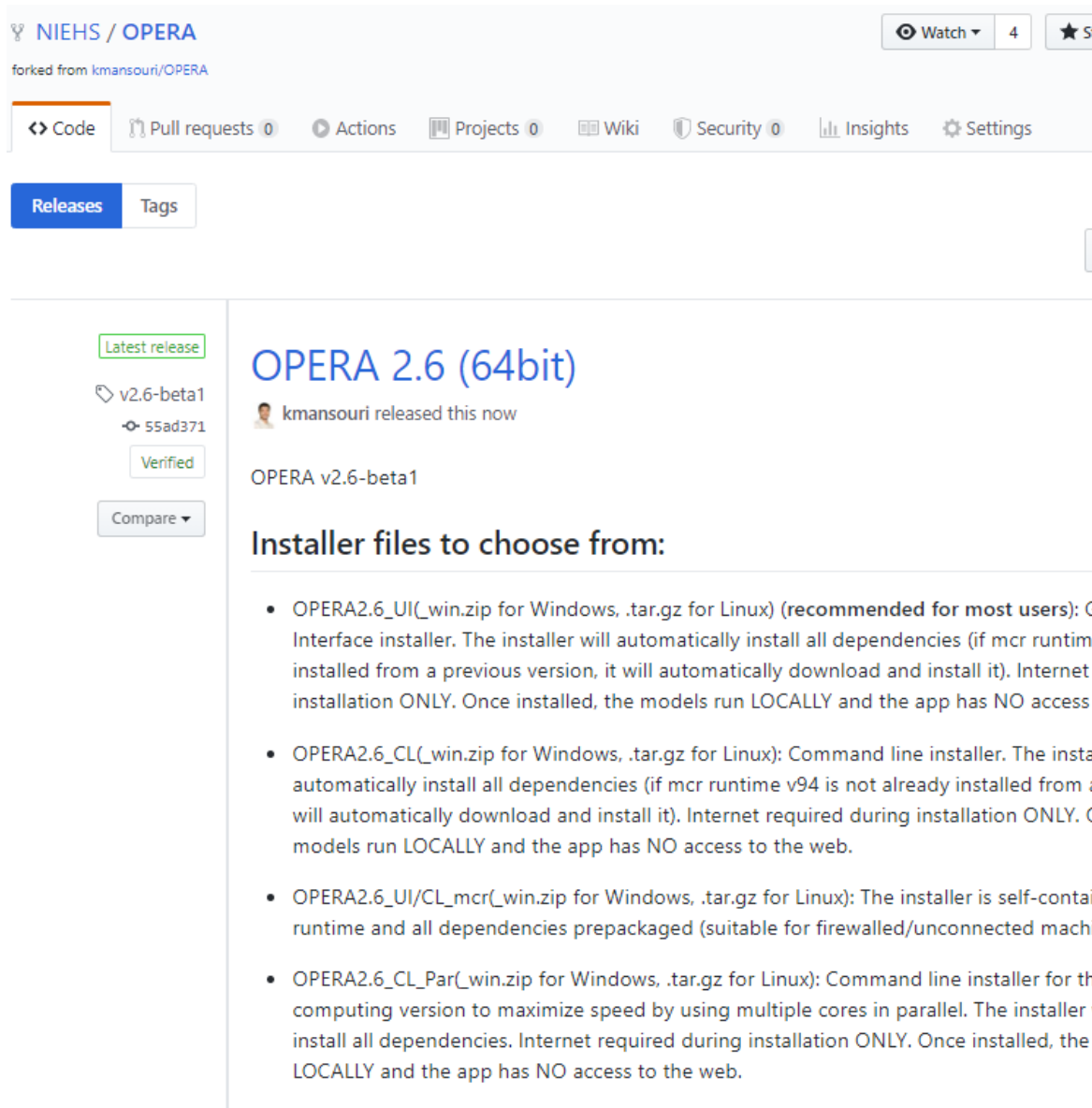
This branch is 24 commits ahead of kmansouri:master. Pull request Compare

 kmansouri Merge pull request #22 from kmansouri/master Latest commit 930a3cb 6 minutes ago

 OPERA_Source_code	v2.5-beta1	1 hour ago
 Icon.png	OPERA 1.2 icon	3 years ago
 Install_guide.pdf	Install and quick run guide	10 months ago
 LICENSE	Initial commit	3 years ago
 Logo.png	Added logo and icon	3 years ago
 OPERA1.5_Source_code.zip	MATLAB source code for OPERA1.5	13 months ago

Go to the releases to download the latest version

Download the installer from “Releases”



NIEHS / OPERA
forked from kmansouri/OPERA

Watch 4 Star

Code Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

Releases Tags

Latest release

v2.6-beta1
55ad371
Verified

Compare

OPERA 2.6 (64bit)

kmansouri released this now

OPERA v2.6-beta1

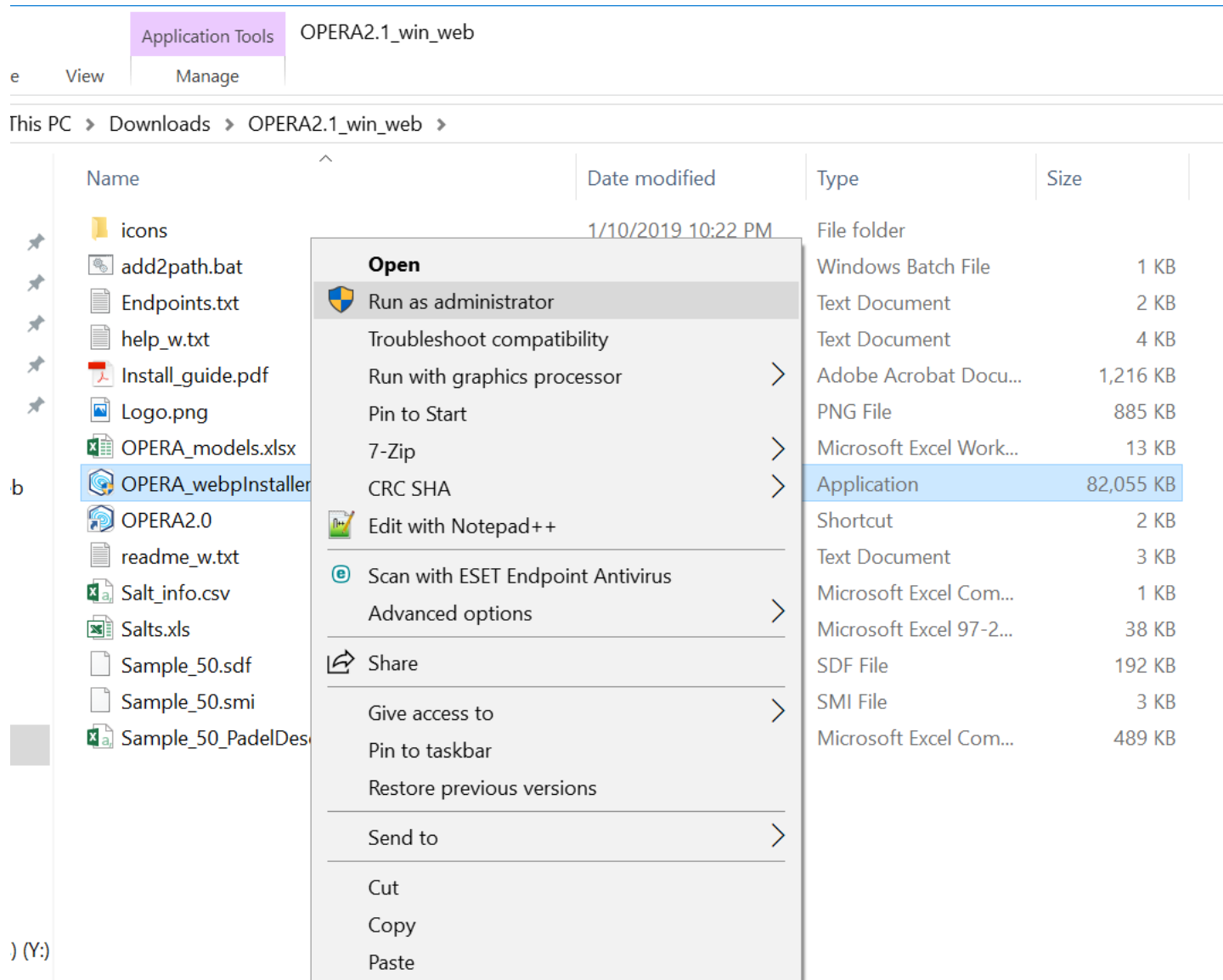
Installer files to choose from:

- OPERA2.6_UI_win.zip for Windows, .tar.gz for Linux) (**recommended for most users**): Graphical Interface installer. The installer will automatically install all dependencies (if mcr runtime is installed from a previous version, it will automatically download and install it). Internet installation ONLY. Once installed, the models run LOCALLY and the app has NO access to the web.
- OPERA2.6_CL_win.zip for Windows, .tar.gz for Linux): Command line installer. The installer will automatically install all dependencies (if mcr runtime v94 is not already installed from a previous version, it will automatically download and install it). Internet required during installation ONLY. Once installed, the models run LOCALLY and the app has NO access to the web.
- OPERA2.6_UI/CL_mcr_win.zip for Windows, .tar.gz for Linux): The installer is self-contained and includes the mcr runtime and all dependencies prepackaged (suitable for firewalled/unconnected machines).
- OPERA2.6_CL_Par_win.zip for Windows, .tar.gz for Linux): Command line installer for the parallel computing version to maximize speed by using multiple cores in parallel. The installer will automatically install all dependencies. Internet required during installation ONLY. Once installed, the models run LOCALLY and the app has NO access to the web.

Read the provided information about the different installers available and select one of the installer options:

- OS: windows, Linux
- Command line, GUI
- Online installer, offline installer
- **New since v2.6: QSAR-ready standardization workflow included!**

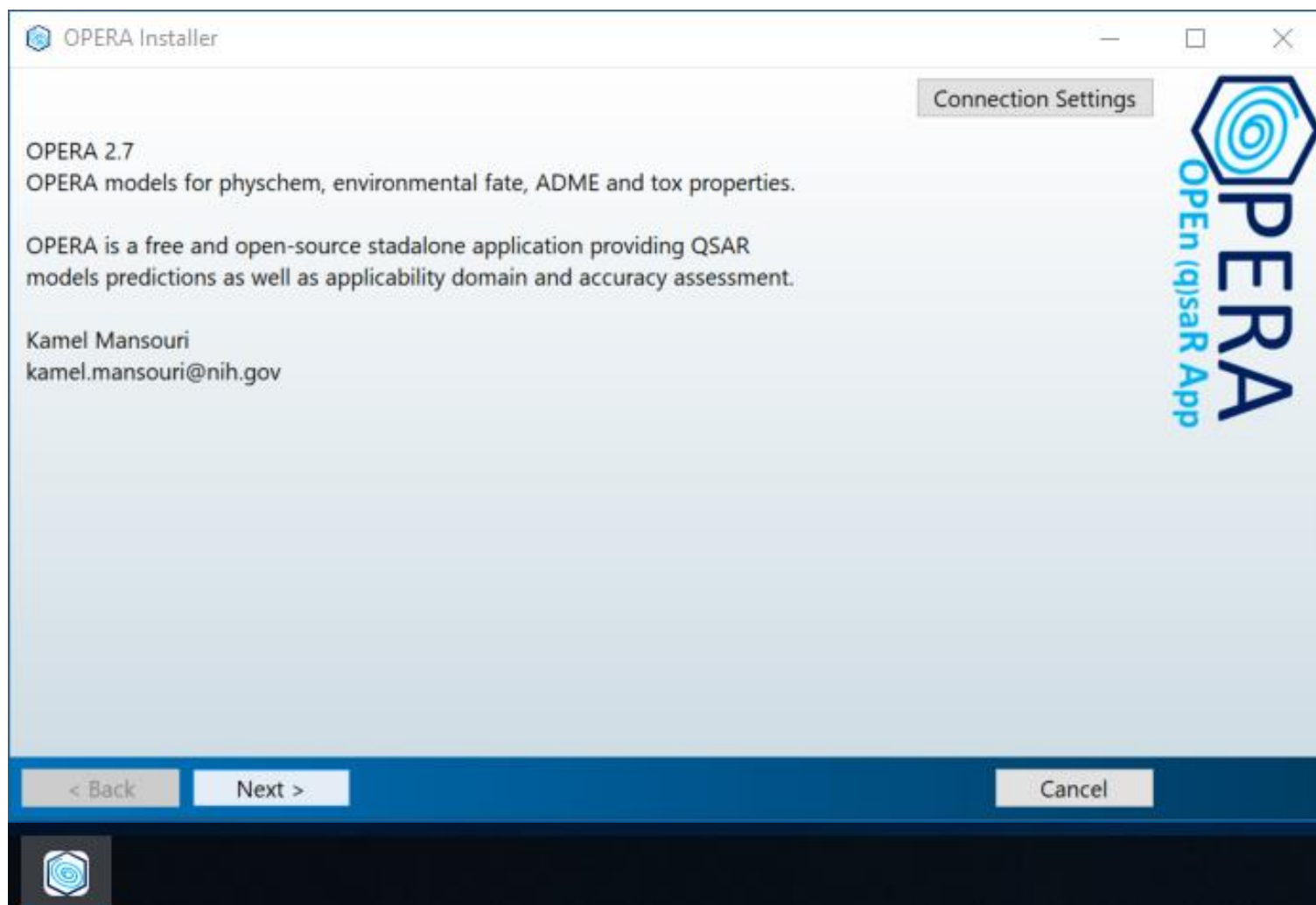
Unzip and run OPERA_Installer as administrator



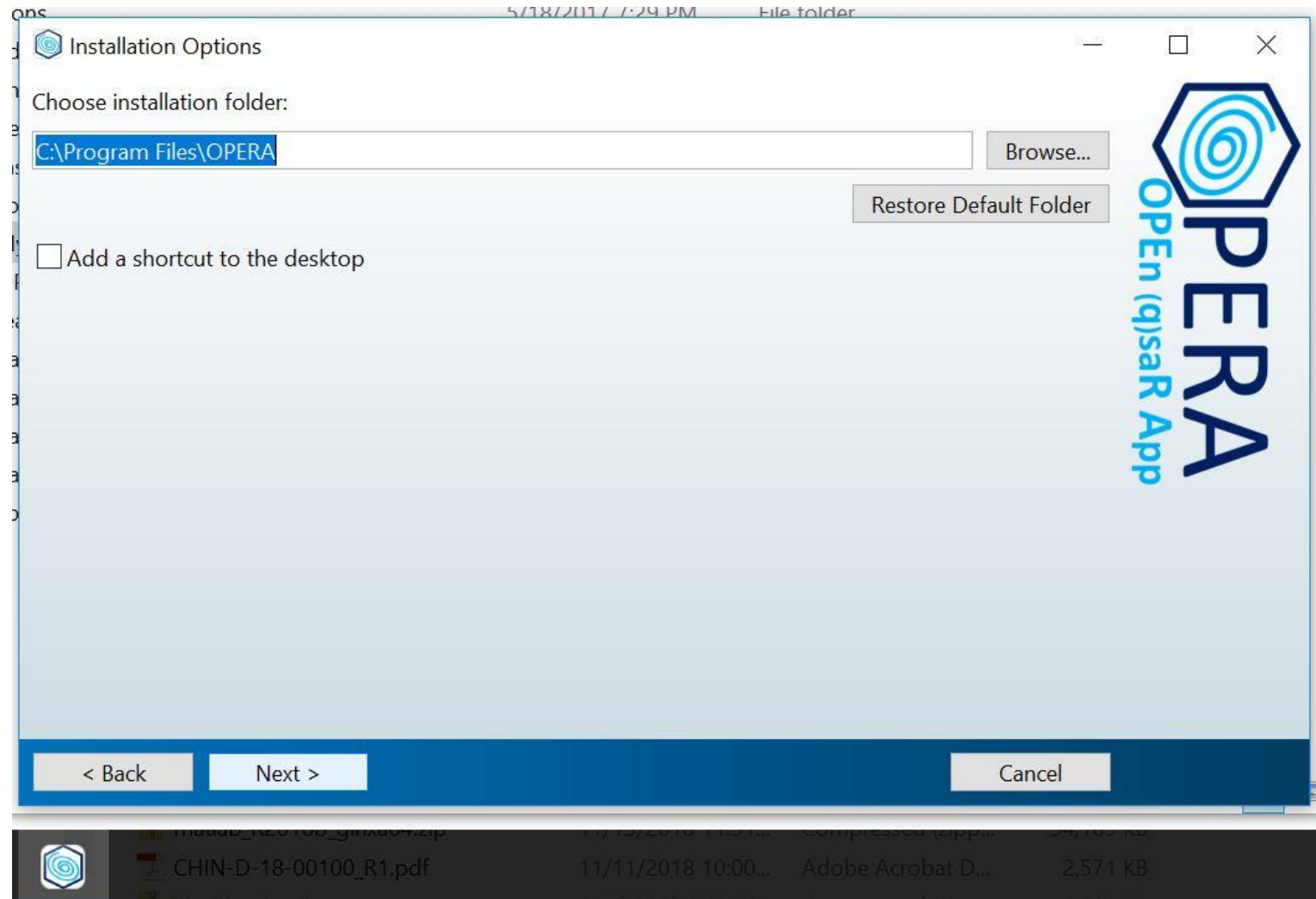
The installer will unpack and initialize quickly



First intro to OPERA, click next when ready



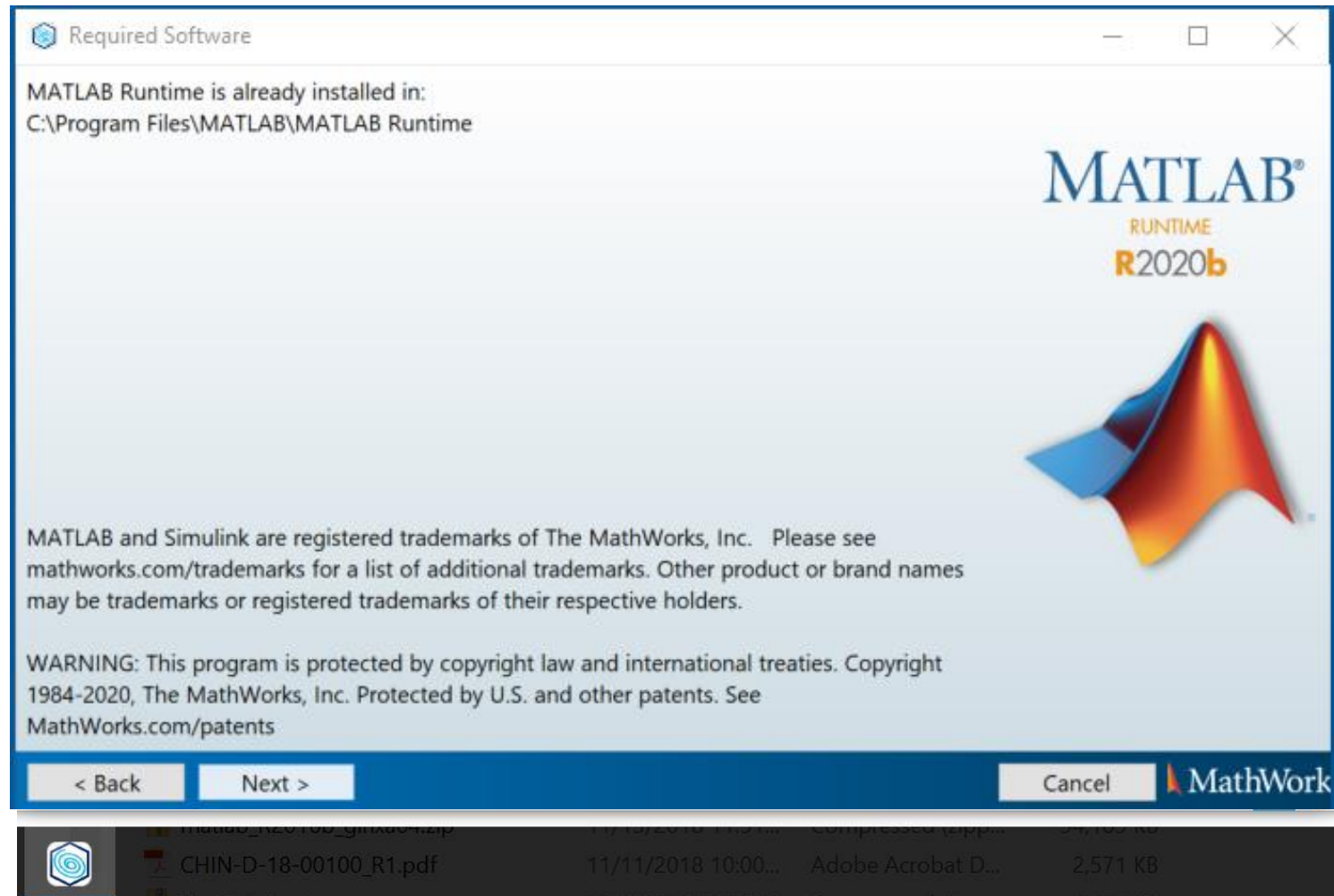
Recommended install options.



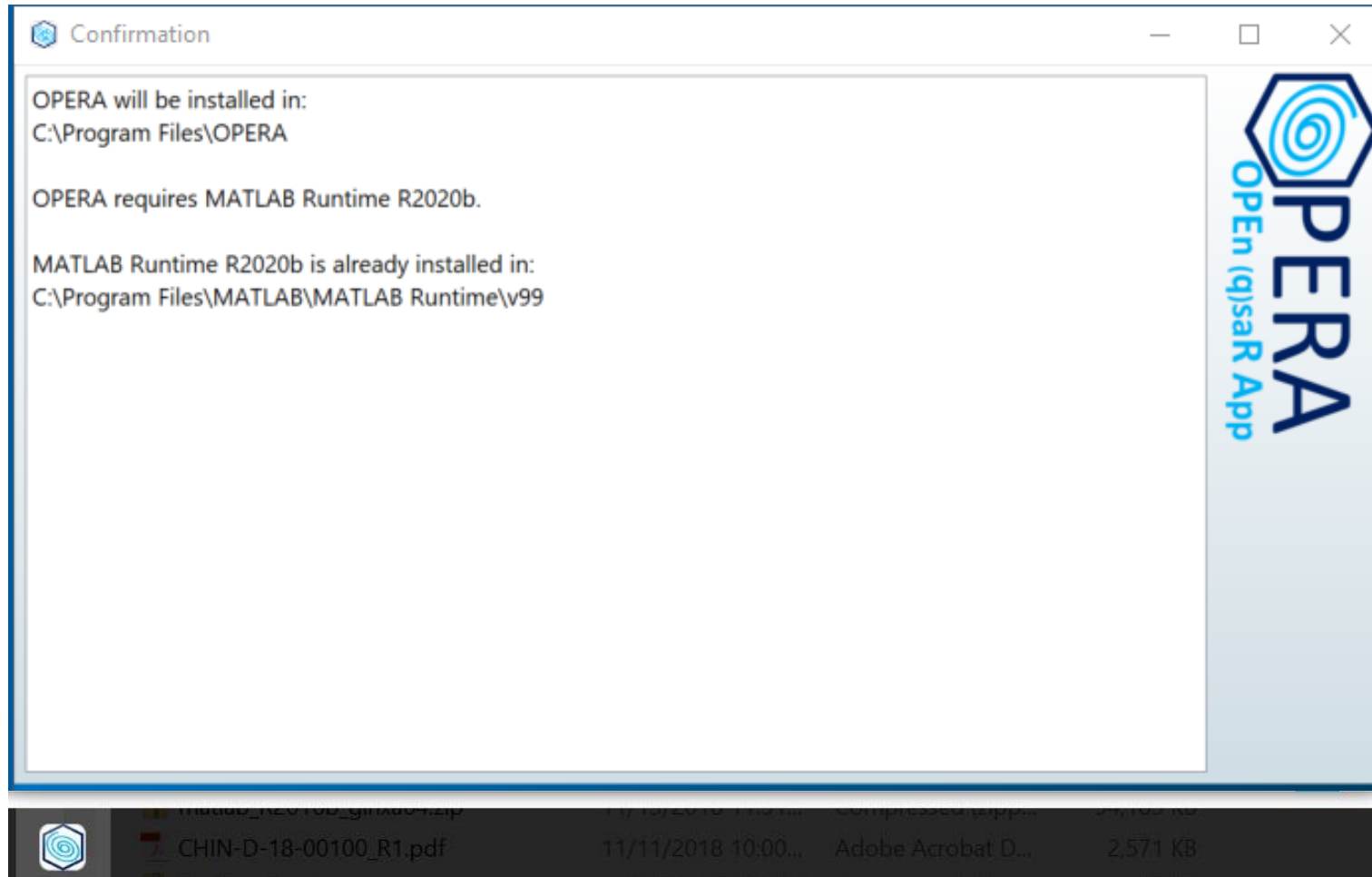
It is important to

- For GUI, select add shortcut
- For command line, the shortcut is provided in the unzipped file. Just copy paste it to the desktop.
- Note that if the default installation folder is modified:
- The GUI will adapt automatically
- The command line shortcut location needs to be updated and further instructions will be provided during the first run.

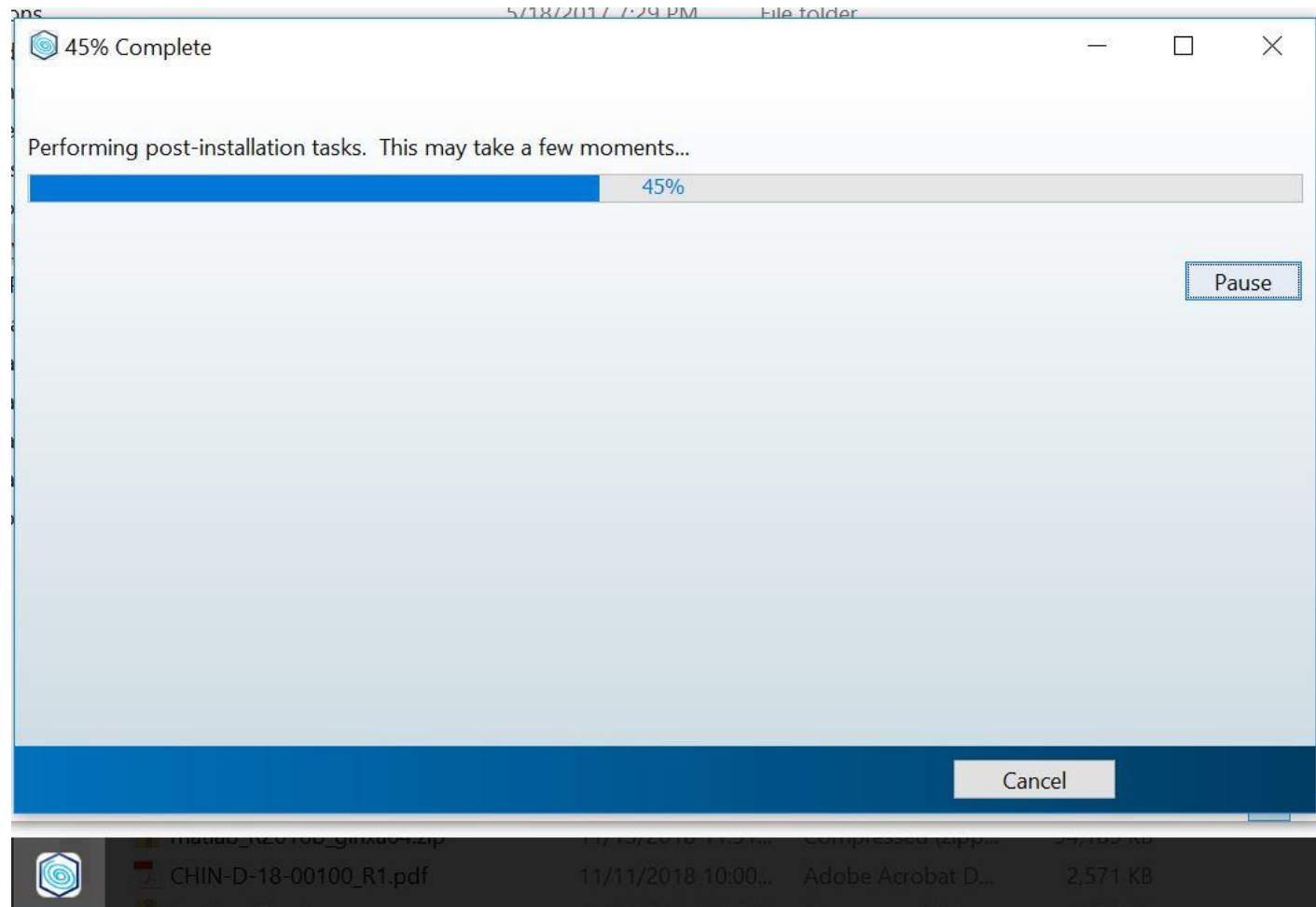
Just click next



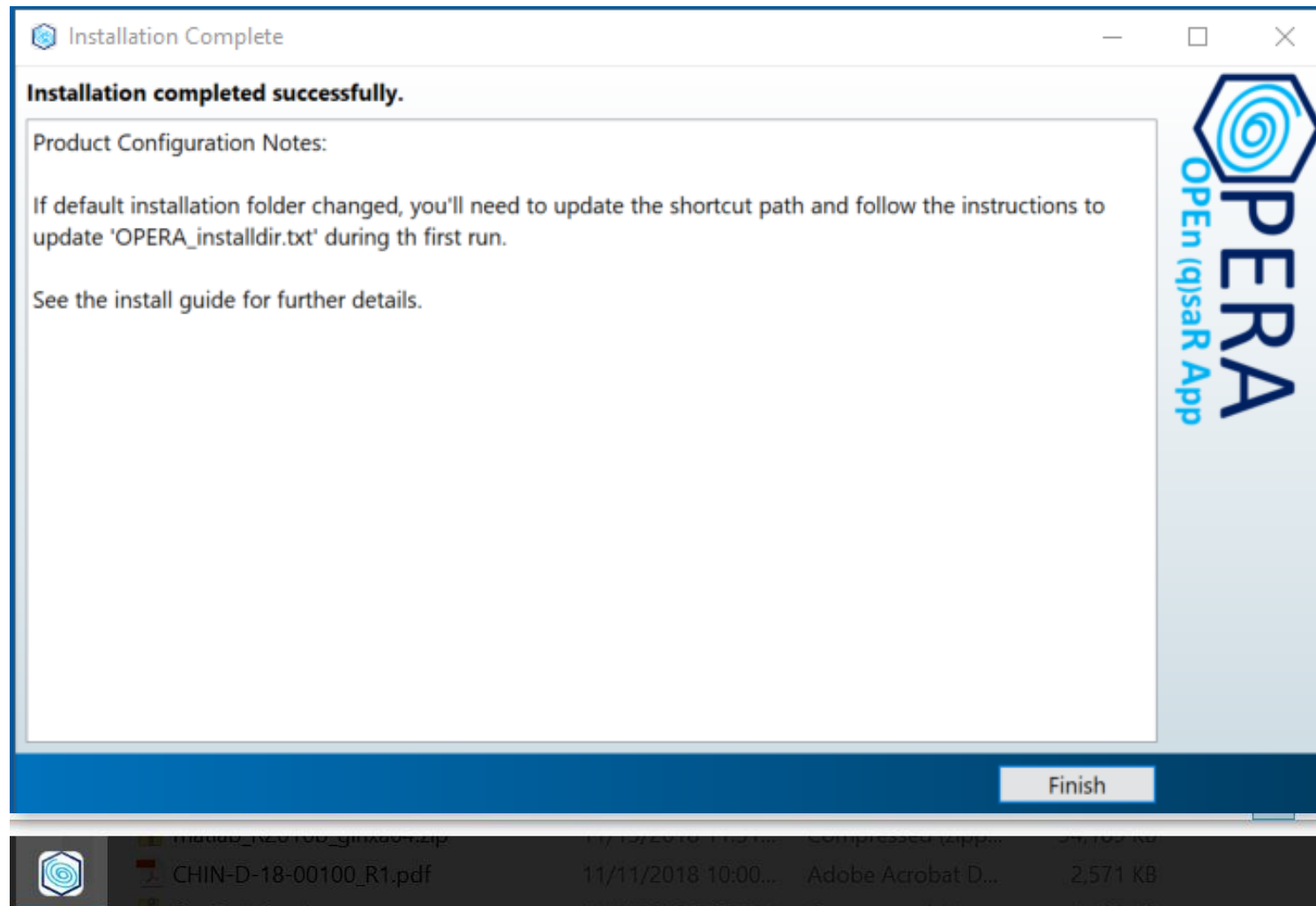
Confirm and install.



Wait while downloading and/or installing the runtime. (next time will go faster)









That's it! IT's OPERAtional.



Installed files

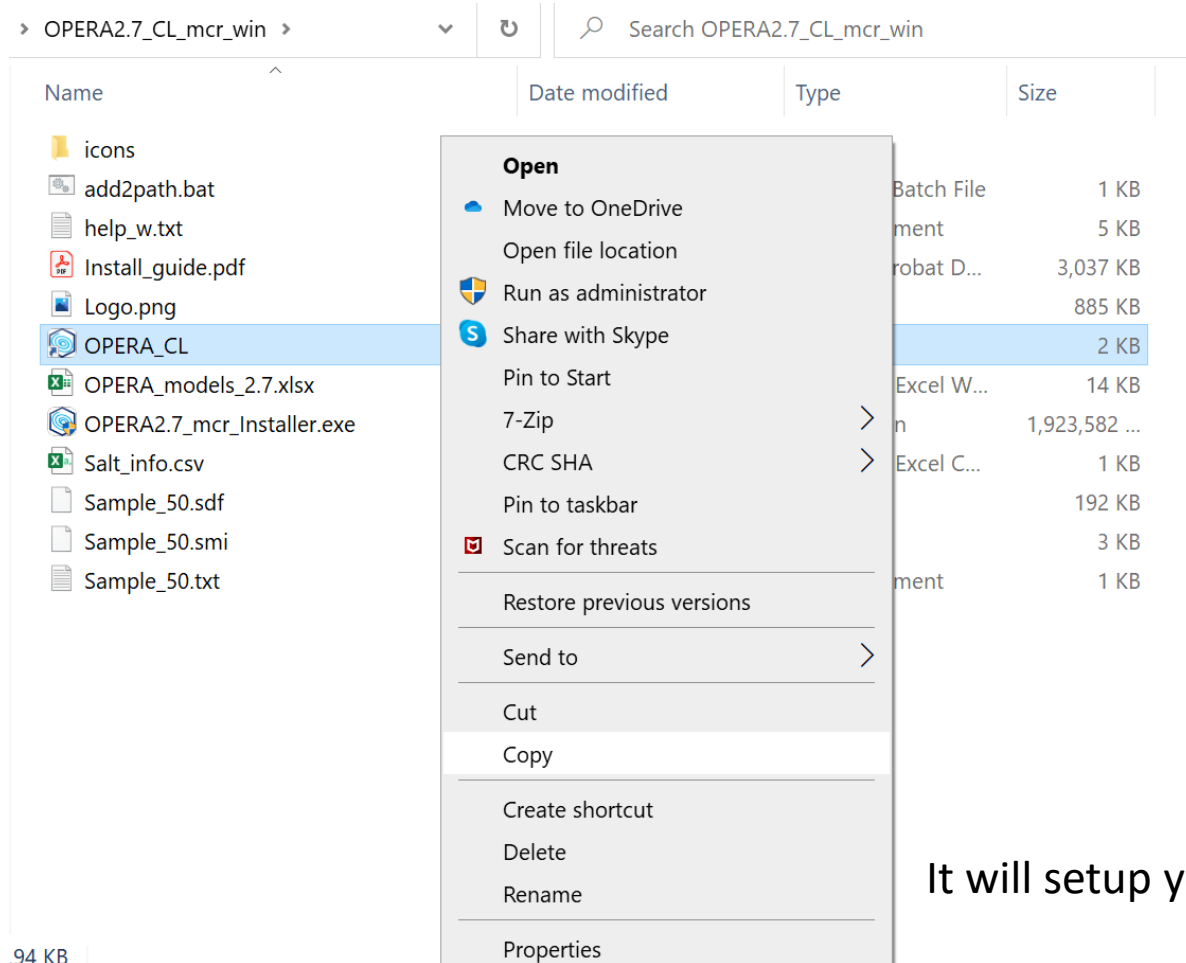
OPERA files

→ ▾ ↑ 📁 > This PC > Local Disk (C:) > Program Files > OPERA > application			
▼ 📁 OPERA	^		
> 📁 appdata			
📁 application			
> 📁 sys			
> 📁 uninstall			
> 📁 Realtek			
> 📁 STMicroelectron			
	Name	Type	Size
	 CDKDescUI-2.0.jar	Executable Jar File	11,672 KB
	 desc_fp.xml	XML File	1 KB
	 icon.ico	Icon	7 KB
	 OPERA.exe	Application	19,279 KB
	 padel-full-1.00.jar	Executable Jar File	30,943 KB
	 splash.png	PNG File	72 KB

Runtime files

C > Local Disk (C:) > Program Files > MATLAB > MATLAB Runtime > v94		
	^	
Name	Date modified	Type
📁 appdata	12/21/2018 10:37...	File folder
📁 bin	12/21/2018 11:06...	File folder
📁 etc	12/21/2018 10:37...	File folder
📁 extern	12/21/2018 10:37...	File folder
📁 help	12/21/2018 11:06...	File folder
📁 java	12/21/2018 10:36...	File folder
📁 lib	12/21/2018 11:06...	File folder
📁 mcr	12/21/2018 10:36...	File folder
📁 polyspace	12/21/2018 11:03...	File folder
📁 remote	12/21/2018 11:05...	File folder
📁 resources	12/21/2018 11:06...	File folder
📁 runtime	12/21/2018 10:37...	File folder
📁 settings	12/21/2018 10:36...	File folder
📁 src	12/21/2018 11:05...	File folder
📁 sys	12/21/2018 11:06...	File folder
📁 toolbox	12/21/2018 11:06...	File folder
📁 ui	12/21/2018 11:05...	File folder
📁 uninstall	12/21/2018 10:37...	File folder
📄 MCR_license.txt	2/3/2015 2:40 PM	Text Document
📄 patents.txt	1/16/2018 3:26 PM	Text Document
📄 trademarks.txt	12/28/2013 2:08 ...	Text Document
📄 VersionInfo.xml	2/23/2018 1:39 PM	XML Document

For command line in Windows only: copy the provided shortcut to your desktop.



It will setup your environment variable and run the application in command line.

Double click the shortcut to run the app

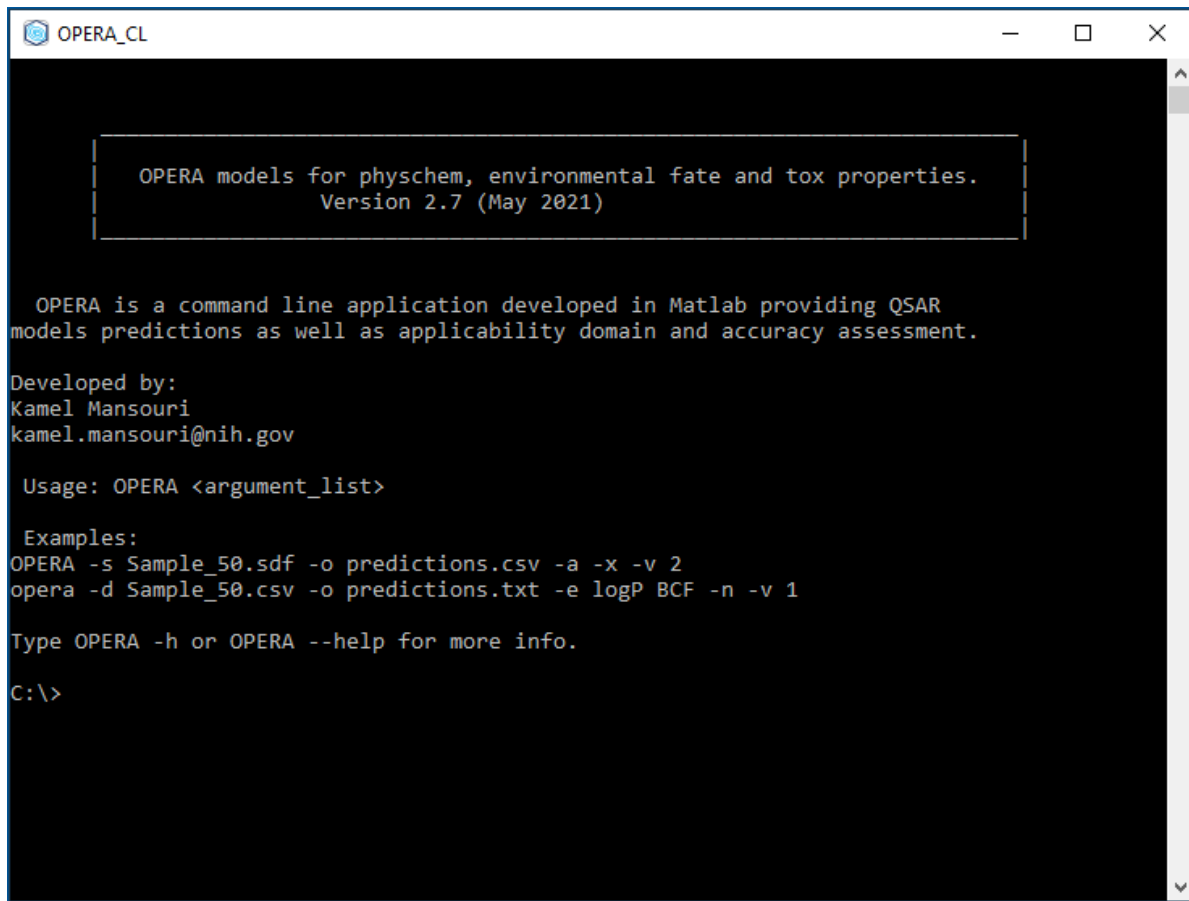


Command line



GUI

Ready to run the models



OPERA_CL

```
OPERA models for physchem, environmental fate and tox properties.  
Version 2.7 (May 2021)
```

OPERA is a command line application developed in Matlab providing QSAR models predictions as well as applicability domain and accuracy assessment.

Developed by:
Kamel Mansouri
kamel.mansouri@nih.gov

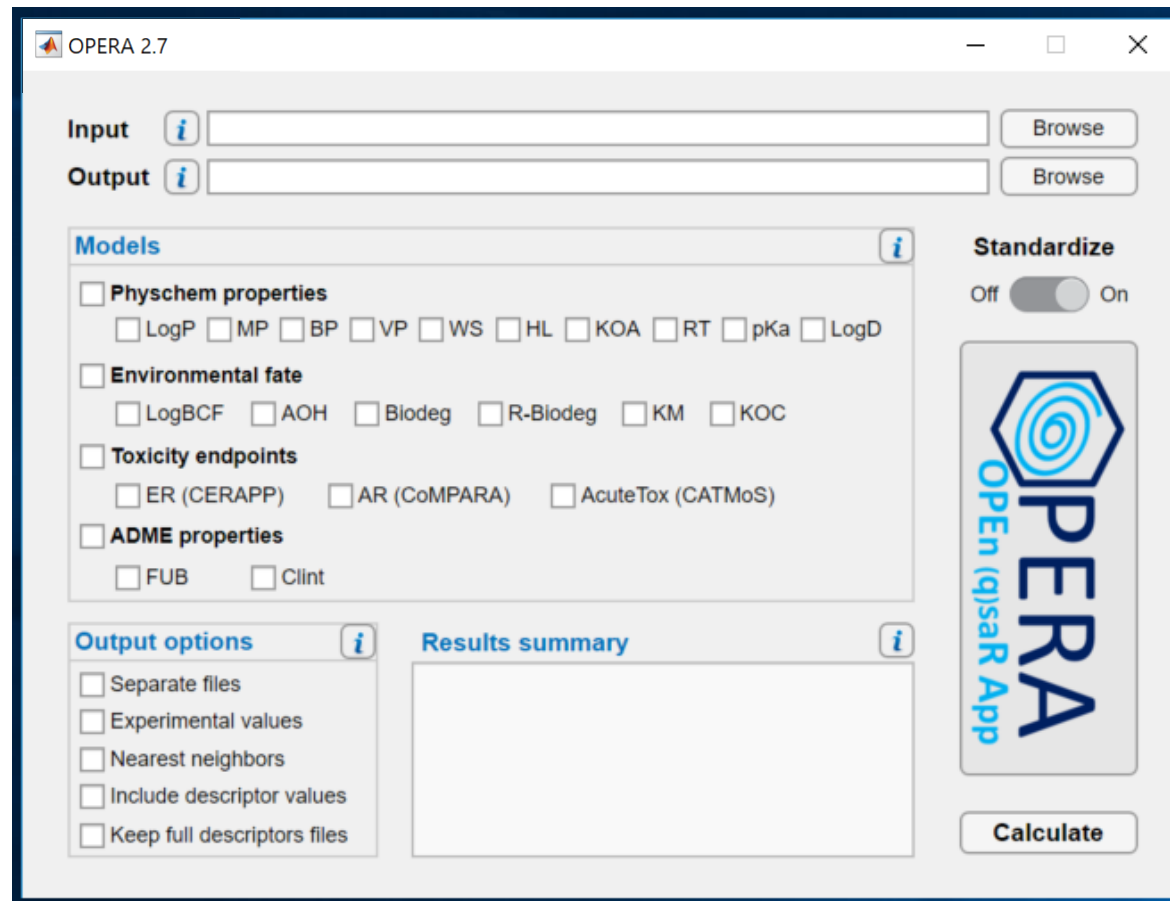
Usage: OPERA <argument_list>

Examples:
OPERA -s Sample_50.sdf -o predictions.csv -a -x -v 2
opera -d Sample_50.csv -o predictions.txt -e logP BCF -n -v 1


Type OPERA -h or OPERA --help for more info.


C:\>


Command line



OPERA 2.7


Input 

Output 


Models 


- ☐ **Physchem properties**
☐ LogP ☐ MP ☐ BP ☐ VP ☐ WS ☐ HL ☐ KOA ☐ RT ☐ pKa ☐ LogD
- ☐ **Environmental fate**
☐ LogBCF ☐ AOH ☐ Biodeg ☐ R-Biodeg ☐ KM ☐ KOC
- ☐ **Toxicity endpoints**
☐ ER (CERAPP) ☐ AR (CoMPARA) ☐ AcuteTox (CATMoS)
- ☐ **ADME properties**
☐ FUB ☐ Clint

Standardize
Off ☐ On ☐

Output options 

- ☐ Separate files
- ☐ Experimental values
- ☐ Nearest neighbors
- ☐ Include descriptor values
- ☐ Keep full descriptors files

Results summary 

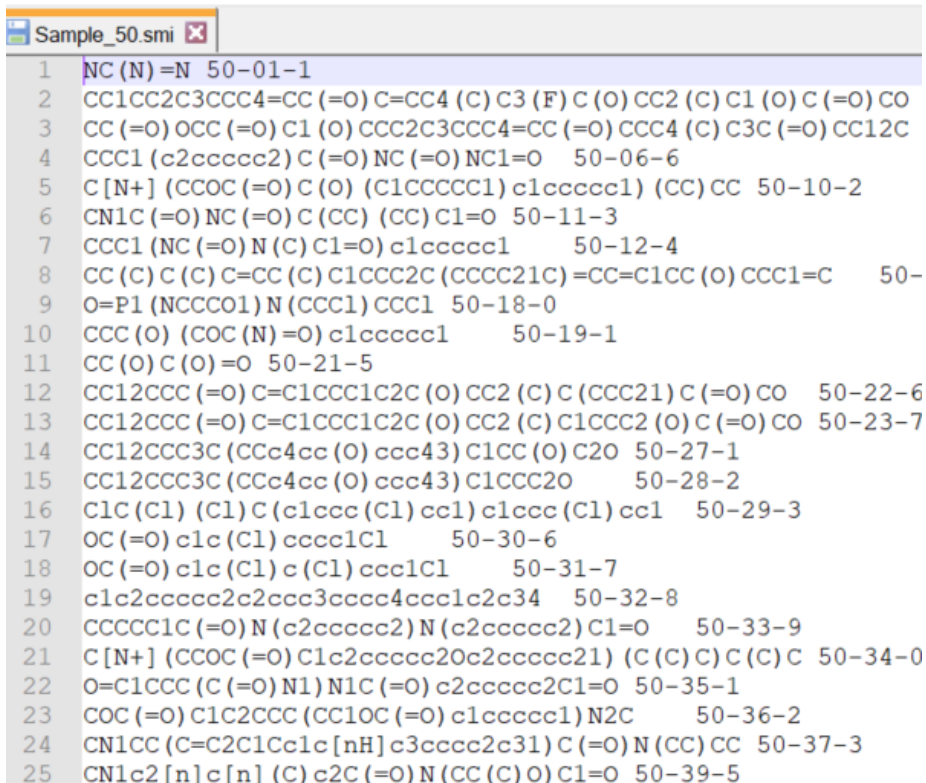


GUI

Accepted Input files are QSAR-ready structures in:

SMILES file:

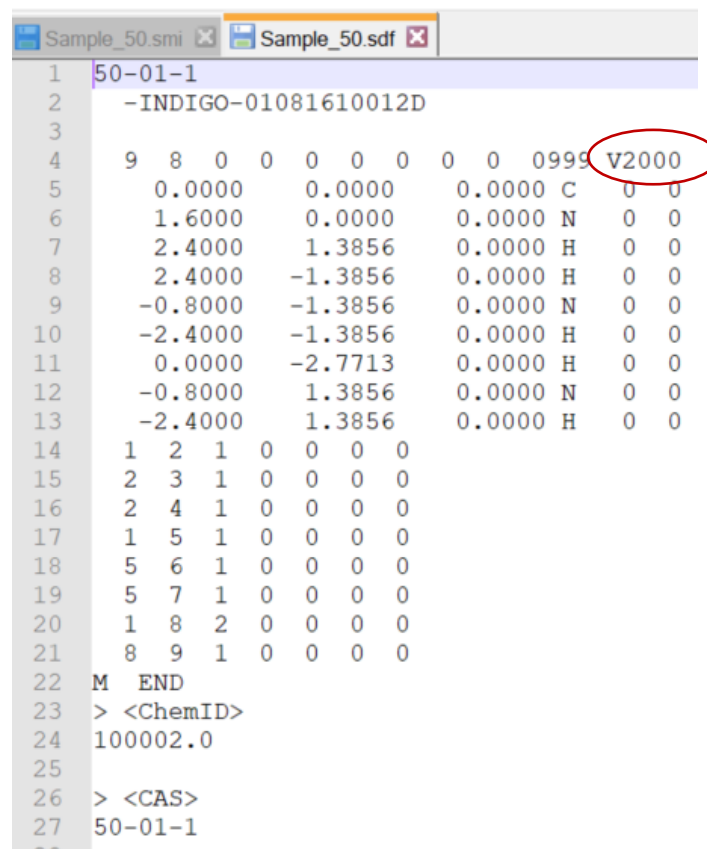
- Extension .smi
- Tab delimited text file (structure \t ID)
- No headers & no empty lines



```
1 NC(N)=N 50-01-1
2 CC1CC2C3CCC4=CC(=O)C=CC4(C)C3(F)C(O)CC2(C)C1(O)C(=O)CO 50-22-6
3 CC(=O)OCC(=O)C1(O)CCC2C3CCC4=CC(=O)CCC4(C)C3C(=O)CC12C 50-23-7
4 CCC1(c2ccccc2)C(=O)NC(=O)NC1=O 50-06-6
5 C[N+](CCOC(=O)C(O)(C1CCCC1)c1ccccc1)(CC)CC 50-10-2
6 CN1C(=O)NC(=O)C(CC)(CC)C1=O 50-11-3
7 CCC1(NC(=O)N(C)C1=O)c1ccccc1 50-12-4
8 CC(C)C(C)C=CC(C)C1CCC2C(CCCC21C)=CC=C1CC(O)CCC1=C 50-
9 O=P1(NCCC1)N(CCC1)CCC1 50-18-0
10 CCC(O)(COC(N)=O)c1ccccc1 50-19-1
11 CC(O)C(O)=O 50-21-5
12 CC12CCC(=O)C=C1CCC1C2C(O)CC2(C)C(CCC21)C(=O)CO 50-22-6
13 CC12CCC(=O)C=C1CCC1C2C(O)CC2(C)C1CCC2(O)C(=O)CO 50-23-7
14 CC12CCC3C(CCC4cc(O)ccc43)C1CC(O)C2O 50-27-1
15 CC12CCC3C(CCC4cc(O)ccc43)C1CCC2O 50-28-2
16 ClC(Cl)(Cl)C(c1ccc(Cl)cc1)c1ccc(Cl)cc1 50-29-3
17 OC(=O)c1c(Cl)cccc1Cl 50-30-6
18 OC(=O)c1c(Cl)c(Cl)ccc1Cl 50-31-7
19 c1c2ccccc2c2ccc3cccc4ccc1c2c34 50-32-8
20 CCCCC1C(=O)N(c2ccccc2)N(c2ccccc2)C1=O 50-33-9
21 C[N+](CCOC(=O)C1c2ccccc2Oc2ccccc21)(C(C)C)C(C)C 50-34-0
22 O=C1CCC(C(=O)N1)N1C(=O)c2ccccc2C1=O 50-35-1
23 COC(=O)C1C2CCC(CC1OC(=O)c1ccccc1)N2C 50-36-2
24 CN1CC(C=C2C1Cc1c[nH]c3ccccc2c31)C(=O)N(CC)CC 50-37-3
25 CN1c2[n]c[n](C)c2C(=O)N(CC(C)O)C1=O 50-39-5
```

SDF/Mol file (v2000):

- Extension .sdf/.mol
- Different blocks
- Single or multiple molecules



```
1 50-01-1
2 -INDIGO-01081610012D
3
4 9 8 0 0 0 0 0 0 0 0 0999 V2000
5 0.0000 0.0000 0.0000 C 0 0
6 1.6000 0.0000 0.0000 N 0 0
7 2.4000 1.3856 0.0000 H 0 0
8 2.4000 -1.3856 0.0000 H 0 0
9 -0.8000 -1.3856 0.0000 N 0 0
10 -2.4000 -1.3856 0.0000 H 0 0
11 0.0000 -2.7713 0.0000 H 0 0
12 -0.8000 1.3856 0.0000 N 0 0
13 -2.4000 1.3856 0.0000 H 0 0
14 1 2 1 0 0 0 0
15 2 3 1 0 0 0 0
16 2 4 1 0 0 0 0
17 1 5 1 0 0 0 0
18 5 6 1 0 0 0 0
19 5 7 1 0 0 0 0
20 1 8 2 0 0 0 0
21 8 9 1 0 0 0 0
22 M  END
23 > <ChemID>
24 100002.0
25
26 > <CAS>
27 50-01-1
```

Descriptors files (CL only):

- Extension .csv
- Calculated by PaDEL and/or CDK
- Continuous and/or fingerprints

	A	B	C	D	E	F	G
	Name	nAcid	ALogP	ALogp2	AMR	apol	naAromAt n.
	100001	0	0	0	0	43.20152	20
	100002	0	0	0	0	48.0551	22
	100003	0	0	0	0	39.68152	18
	100004	0	0	0	0	14.56076	6
	100005	0	1.8631	3.471142	29.4773	19.8951	0
	100006	0	0.1248	0.015575	9.7308	27.78793	10
	100007	0	0	0	0	31.30793	14
	100008	0	0.4562	0.208118	4.601	29.54793	12
0	100009	0	0.642	0.412164	5.5021	26.02793	10
1	100010	0	0	0	0	22.93434	10
2	100011	0	0.642	0.412164	5.5021	26.02793	10
3	100012	0	0	0	0	27.78793	12
4	100013	0	0.789	0.622521	15.0732	21.17434	6
5	100014	0	0.5906	0.348808	7.7935	20.98076	9
5	100015	0	1.284	1.648656	11.0042	20.74793	6
7	100016	0	1.926	3.709476	16.5063	23.84152	6
3	100017	0	2.568	6.594624	22.0084	26.9351	6
9	100018	0	0.0686	0.004706	23.2621	18.56152	0
0	100019	0	0.5729	0.328214	14.2678	23.84152	6

New input option since v2.3

Text file (.txt) with chemical ID if known:

- CASRN,
- DTXSID,
- InChiKey

1	50-01-1
2	DTXSID3020384
3	50-04-4
4	50-06-6
5	50-10-2
6	FWJKNZONDWOGMI-UHFFFAOYSA-N
7	DTXSID9023257
8	50-14-6
9	50-18-0
10	50-19-1
11	DTXSID7023192
12	50-22-6
13	DTXSID7020714
14	50-27-1
15	50-28-2
51	8052-31-1
52	12125-02-9



IDs are used to identify and use structures from the knowledge base of OPERA containing all ~900k QSAR ready structures available on the EPA CompTox chemicals dashboard.

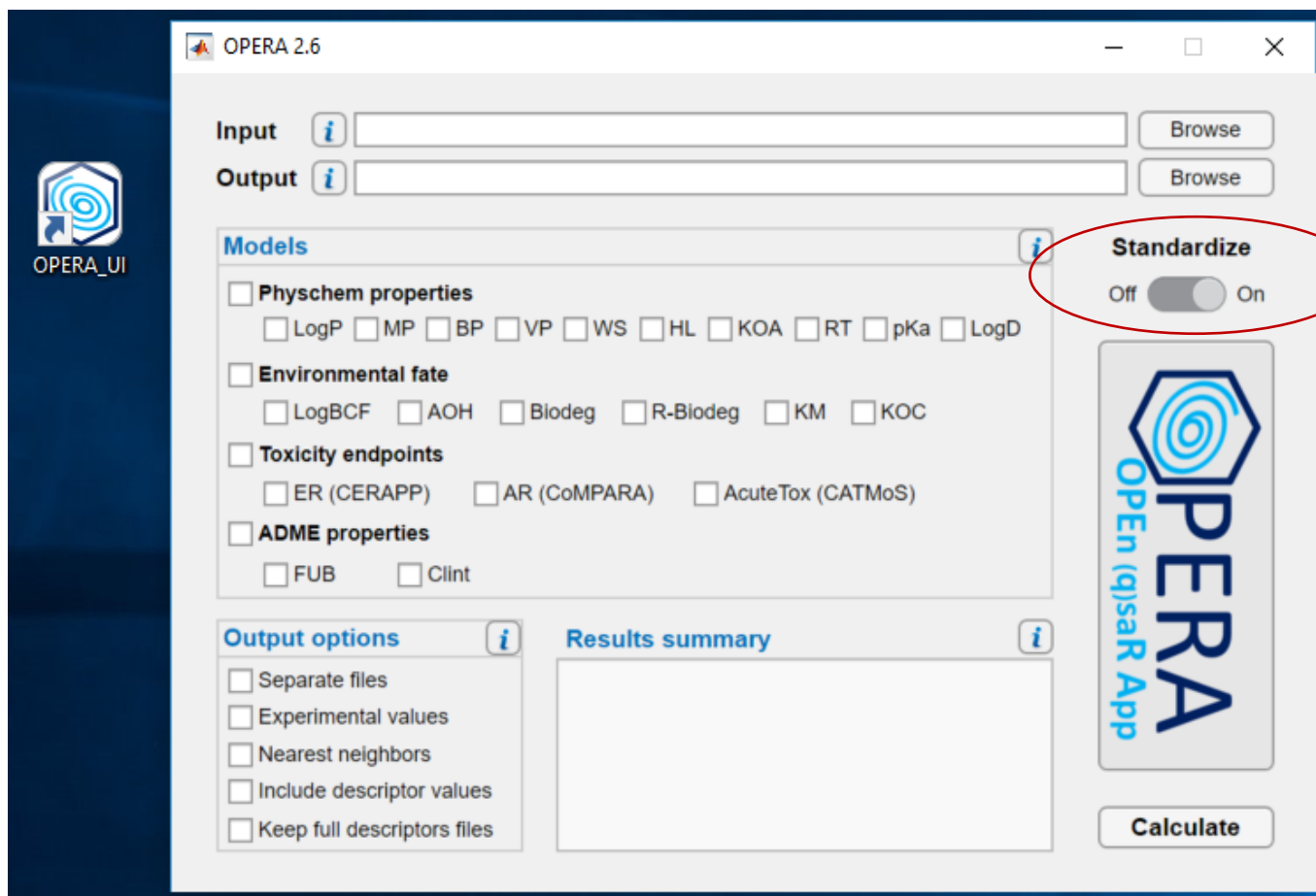
1	MoleculeID	FoundBy	LogP_pred
2	50-01-1	CASRN	-1.4006686
3	DTXSID3020384	DTXSID	1.920045958
4	50-04-4	CASRN	2.102684892
5	50-06-6	CASRN	1.469012545
6	50-10-2	CASRN	-1.070394867
7	FWJKNZONDWOGMI-UHFFFAOYSA-N	InChiKey	1.151133944
8	DTXSID9023257	DTXSID	1.689380242
9	50-14-6	CASRN	8.375363568
10	50-18-0	CASRN	0.635404951
11	50-19-1	CASRN	-1.019179944
12	DTXSID7023192	DTXSID	-0.720712697
13	50-22-6	CASRN	1.94212741
14	DTXSID7020714	DTXSID	1.61077559
15	50-27-1	CASRN	2.451817819
16	50-28-2	CASRN	2.888888888
52	8052-31-1	NotFound	NaN
53	12125-02-9	Error: Inorganic	NaN

Unique column, 1 ID/chemical

If ID is not found, the user needs to enter a structure instead.

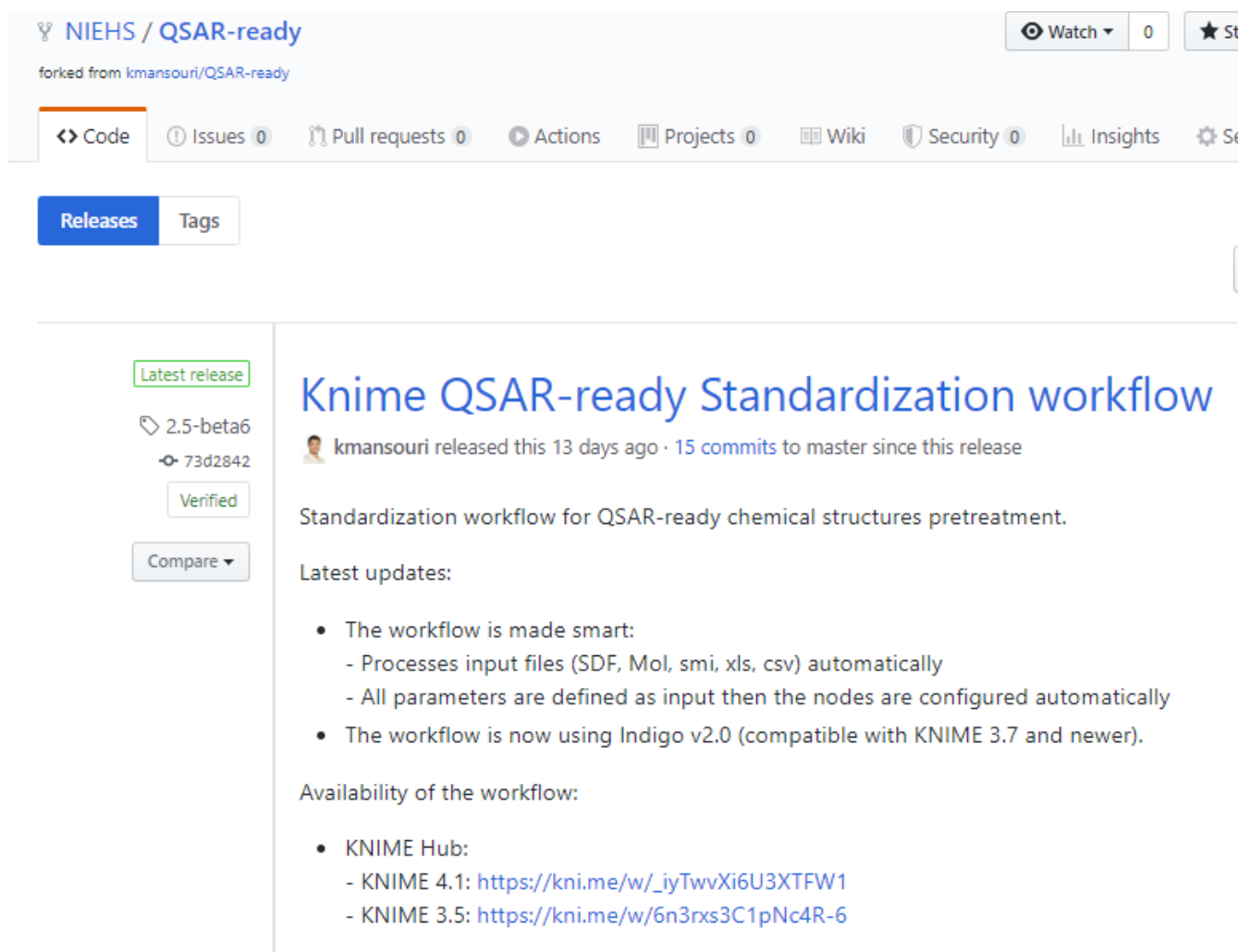
QSAR-ready structures:

- Option 1 (since v2.6): Use the standardize option in OPERA



QSAR-ready structures:

- Option 2: run the KNIME QSAR-ready standardization workflow:
<https://github.com/NIEHS/QSAR-ready>

The image shows the GitHub repository page for 'NIEHS / QSAR-ready'. At the top, it says 'forked from kmansouri/QSAR-ready'. Below this are tabs for 'Code', 'Issues', 'Pull requests', 'Actions', 'Projects', 'Wiki', 'Security', and 'Insights'. There are also buttons for 'Releases' and 'Tags'. The 'Releases' tab is selected, showing the 'Latest release' as '2.5-beta6' with a commit hash of '73d2842' and a 'Verified' badge. Below the release information is the title 'Knime QSAR-ready Standardization workflow' by 'kmansouri', released 13 days ago with 15 commits. The description states it is a 'Standardization workflow for QSAR-ready chemical structures pretreatment.' Under 'Latest updates:', it lists two points: the workflow is smart (processes input files automatically and configures nodes automatically) and it now uses Indigo v2.0. Under 'Availability of the workflow:', it lists KNIME Hub links for KNIME 4.1 and KNIME 3.5.

NIEHS / QSAR-ready
forked from kmansouri/QSAR-ready

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

Releases Tags

Latest release
2.5-beta6
73d2842
Verified
Compare

Knime QSAR-ready Standardization workflow

kmansouri released this 13 days ago · 15 commits to master since this release

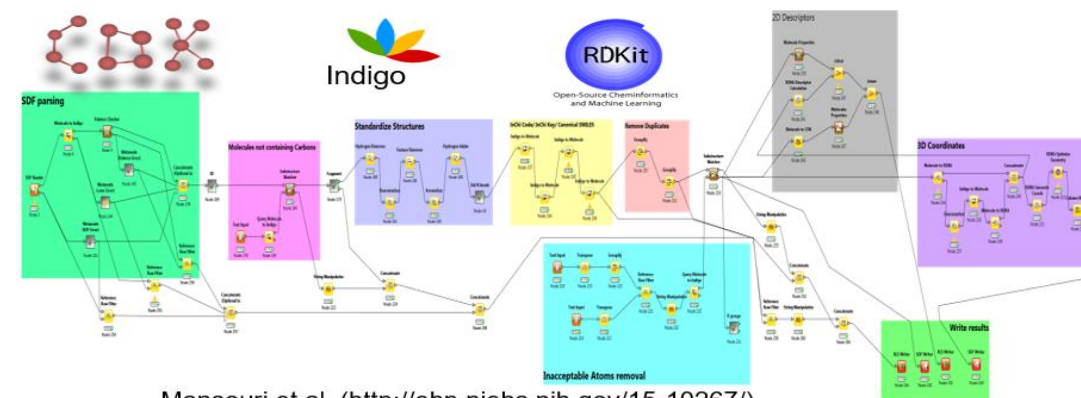
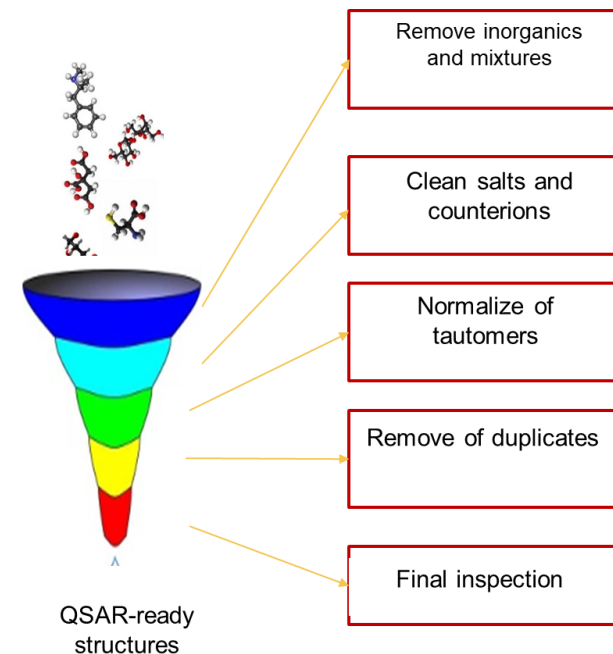
Standardization workflow for QSAR-ready chemical structures pretreatment.

Latest updates:

- The workflow is made smart:
 - Processes input files (SDF, Mol, smi, xls, csv) automatically
 - All parameters are defined as input then the nodes are configured automatically
- The workflow is now using Indigo v2.0 (compatible with KNIME 3.7 and newer).

Availability of the workflow:

- KNIME Hub:
 - KNIME 4.1: https://kni.me/w/_iyTwwXi6U3XTFW1
 - KNIME 3.5: <https://kni.me/w/6n3rxs3C1pNc4R-6>



Mansouri et al. (<http://ehp.niehs.nih.gov/15-10267/>)

QSAR-ready structures:

- Option 3: download the QSAR-ready SMILES from the EPA CompTox Dashboard:
https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

The screenshot shows the EPA CompTox Dashboard interface. At the top, the EPA logo and navigation links (Home, Advanced Search, Batch Search, Lists, Predictions, Downloads) are visible. A progress bar at the top indicates six steps, with Step 6 labeled "Step Six: Click 'Download'". Below the progress bar, the "Select Output Format:" dropdown is set to "Excel". A large blue "Download" button is prominently displayed. On the left, the "Customize Results" section includes checkboxes for "Chemical Identifiers" (DTXSID, Chemical Name, DTXCID, CAS-RN, InChIKey, IUPAC Name) and "Structures" (Mol File, SMILES, InChI String, MS-Ready SMILES, and QSAR-Ready SMILES, which is circled in red). A tooltip explains that QSAR-Ready SMILES are desalted, de-isotoped, and in neutral forms. On the right, the "Presence in Lists:" section lists various chemical inventories and databases, each with a checkbox and a link.

Select Output Format:
Excel

Download

Customize Results

- ☐ Select All
- ☐ Select All in Lists

Chemical Identifiers

- ☒ DTXSID
- ☒ Chemical Name
- ☐ DTXCID
- ☒ CAS-RN
- ☐ InChIKey
- ☐ IUPAC Name

Structures

- ☐ Mol File
- ☐ SMILES
- ☐ InChI String
- ☐ MS-Ready SMILES
- ☒ QSAR-Ready SMILES

Intrinsic And Predicted Properties

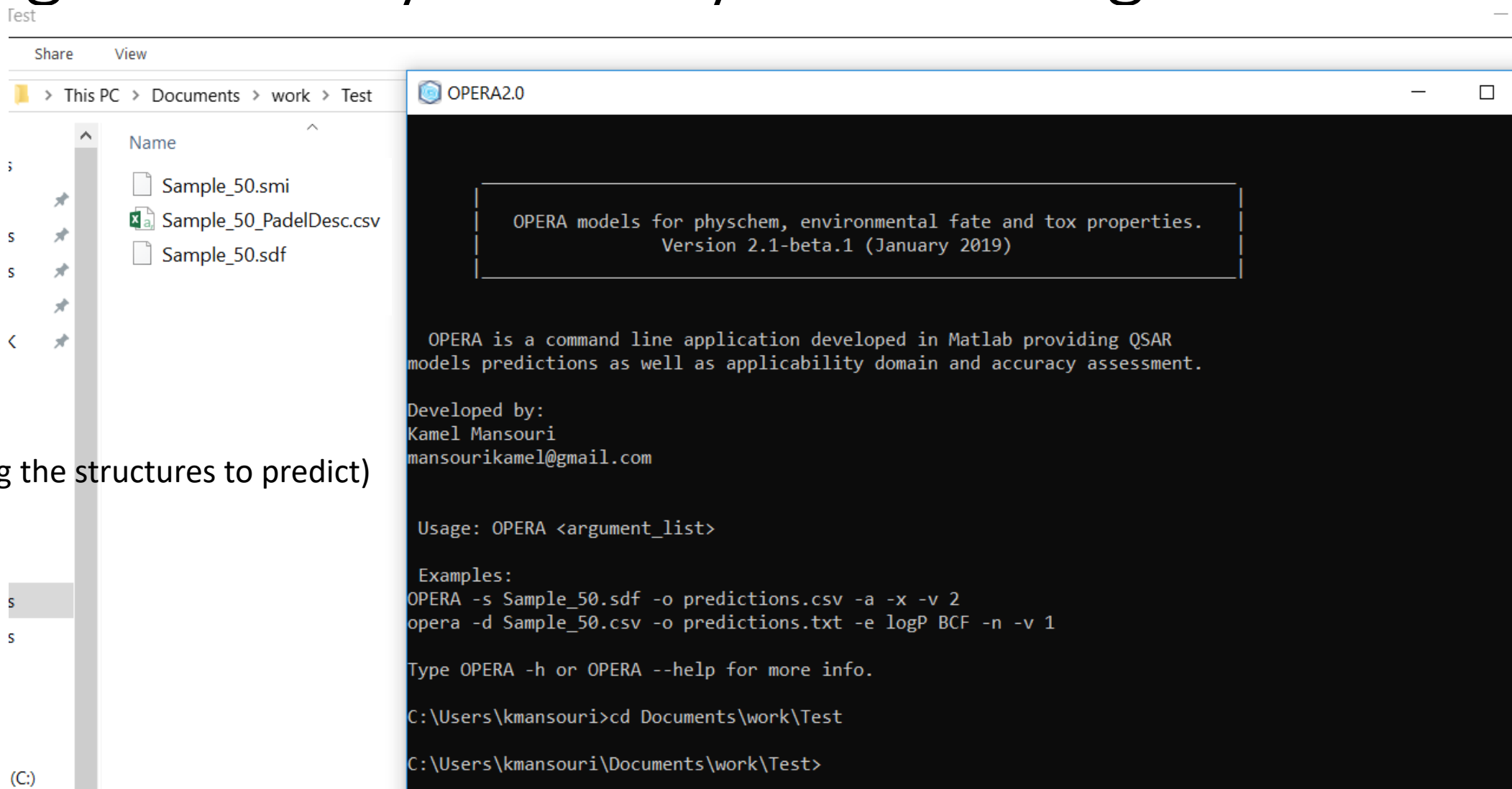
- ☐ Molecular Formula
- ☐ Average Molecular Weight

Presence in Lists:

- ☐ 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities
- ☐ AEGLs: Acute Exposure Guideline Levels
- ☐ ANDROGEN: Androgen Receptor Chemicals
- ☐ ARTICLE: Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014)
- ☐ ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances
- ☐ ATSDR: Toxic Substances Portal Chemical List
- ☐ California Office of Environmental Health Hazard Assessment
- ☐ CERAPP: Collaborative Estrogen Receptor Activity Prediction Project
- ☐ CHEMINV: ToxCast/Tox21 Chemical inventory available as DMSO solutions (20181123)
- ☐ CHEMINV: EPA Chemical Inventory for ToxCast
- ☐ CHEMINV: EPA ToxCast CHEMINV list of volatiles
- ☐ CHEMINV: EPA ToxCast Cheminventory chemicals with stability problems
- ☐ EPA ToxCast Cheminventory DMSO Insolubles
- ☐ EPA ToxCast Cheminventory List of Reactives
- ☐ PubChem database from the University of Alberta
- ☐ ANTIBIOTIC list of antibiotics
- ☐ DRUGS: Pharmaceutical List with EU, Swiss, US Consumption Data

QSAR-Ready SMILES are the SMILES representations of the desalted, de-isotoped, stereo-neutral forms of chemical structures associated with particular chemical substances. One QSAR-Ready SMILES can map to multiple individual chemical substances. They differ from MS-Ready SMILES based on a more restricted set of salt counterions primarily.

Command line: Change directory “cd” to your working folder



(Folder containing the structures to predict)

Use the commands in the help file or type: “opera -h” for help

```
C:\Users\kmansouri\Documents\work\Test>opera -h
```

OPERA2.0

```
OPERA models for physchem, environmental fate and tox properties.  
Version 2.1-beta.1 (January 2019)
```

OPERA is a command line application developed in Matlab providing QSAR models predictions as well as applicability domain and accuracy assessment. All models are built on curated data from public domain. Molecular descriptors are calculated using PaDEL and CDK software.

Usage: OPERA <input> <output> [Options]

By default, all endpoints will be calculated.

Examples:

```
OPERA -s Sample_50.sdf -o predictions.csv -a -x -n -v 2
```

```
opera -d Sample_50.csv -o predictions.txt -e logP BCF -v 1
```

Input:

```
-s, --SDF, --MOL, --SMI Structure file containing the molecule(s) to be  
                           predicted. IDs will be assigned if the file does not contain molecule names.  
                           Molecular descriptors will be calculated using PaDEL software.  
-d, --Descriptors      pre-calculated PaDEL descriptors in csv file. If the first column is not  
                           "Name" as the standard PaDEL output, molecule IDs will be assigned.  
-m, --Mat, --ascii     Matlab matrix or ascii file containing PaDEL descriptors.  
-i, --MolID            Molecule names in csv file.  
-t, --SaltInfo         Salt IDs to improve melting point predictions. List provided in Salts.xls  
-l, --Labels           Descriptor labels. Necessary if the descriptor file does not contain labels  
                           or contains more than the 1444 PaDEL 2D descriptors.
```

Output:

```
-o, --Output            Output file containing the predictions, applicability domain and accuracy  
                           information. File extension could be csv or txt. The output will contain by default:  
                           Molecule ID, predicted value (pred), Applicability domain (AD), Similarity index  
                           (Sim_index) and accuracy estimate (Conf_index).  
-n, --Neighbors        Add 5 nearest neighbors from training set (CAS, InChIKeys, Observed and predicted values)  
-O, --FullOutput       Output file containing all prediction details and used descriptors in csv format.  
-x, --Seperate         Separate output file for each endpoint.
```

Miscellaneous:

```
-v, --Verbose           Verbose level: 0=silent (default), 1=minimum details, 2=full details.  
-a, --All              All endpoints to be calculated (default).  
-c, --Clean            Clean temporary files (generated during descriptor calculation).  
-logP, -BCF...         List endpoints to be calculated (case insensitive). 'BCF'/'logBCF', 'BP', 'logP', 'MP',  
                           'VP'/'logVP', 'WS', 'AOH', 'BioDeg', 'RB'/'ReadyBiodeg', 'HL'/'logHL', 'KM'/'logKM',  
                           'KOA', 'Koc'/'logKoc', 'RT', 'pKa', 'logD', 'CERAPP'/'ER', 'CoMPARA'/'AR', 'CATMoS/AcuteTox'.  
                           Groups of Endpoints: 'StrP' (Structural properties), 'PC'/'Physchem', 'EnvFate'/'EF', 'Tox' (ER, AR, AcuteTox).  
-e, --Endpoint         List endpoints to be calculated.  
-h, --Help             Display this help file and exit.  
-V, --Version          Version of the application
```

OPERA is a command line application developed in Matlab providing QSAR models predictions as well as applicability domain and accuracy assessment. All models are built on curated data from public domain. Molecular descriptors are calculated using PaDEL and CDK software.

Usage: OPERA <input> <output> [Options]
By default, all endpoints will be calculated.
Examples:
OPERA -s Sample_50.sdf -o predictions.csv -a -x -n -v 2
opera -d Sample_50.csv -o predictions.txt -e logP BCF -v 1

Input:

```
-s, --SDF, --MOL, --SMI Structure file containing the molecule(s) to be  
                           predicted. IDs will be assigned if the file does not contain molecule names.  
                           Molecular descriptors will be calculated using PaDEL software.  
-d, --Descriptors      pre-calculated PaDEL descriptors in csv file. If the first column is not  
                           "Name" as the standard PaDEL output, molecule IDs will be assigned.  
-m, --Mat, --ascii     Matlab matrix or ascii file containing PaDEL descriptors.  
-i, --MolID            Molecule names in csv file.  
-t, --SaltInfo         Salt IDs to improve melting point predictions. List provided in Salts.xls  
-l, --Labels           Descriptor labels. Necessary if the descriptor file does not contain labels  
                           or contains more than the 1444 PaDEL 2D descriptors.
```

Output:

```
-o, --Output            Output file containing the predictions, applicability domain and accuracy  
                           information. File extension could be csv or txt. The output will contain by def  
                           Molecule ID, predicted value (pred), Applicability domain (AD), Similarity inde  
                           (Sim_index) and accuracy estimate (Conf_index).  
-n, --Neighbors        Add 5 nearest neighbors from training set (CAS, InChIKeys, Observed and predict  
-O, --FullOutput       Output file containing all prediction details and used descriptors in csv forma  
-x, --Seperate         Separate output file for each endpoint.
```

Miscellaneous:

```
-v, --Verbose           Verbose level: 0=silent (default), 1=minimum details, 2=full details.  
-a, --All              All endpoints to be calculated (default).  
-c, --Clean            Clean temporary files (generated during descriptor calculation).  
-logP, -BCF...         List endpoints to be calculated (case insensitive). 'BCF'/'logBCF', 'BP', 'logP',  
                           'VP'/'logVP', 'WS', 'AOH', 'BioDeg', 'RB'/'ReadyBiodeg', 'HL'/'logHL', 'KM'/'logKM',  
                           'KOA', 'Koc'/'logKoc', 'RT', 'pKa', 'logD', 'CERAPP'/'ER', 'CoMPARA'/'AR', 'CATM  
                           Groups of Endpoints: 'StrP' (Structural properties), 'PC'/'Physchem', 'EnvFate'  
-e, --Endpoint         List endpoints to be calculated.  
-h, --Help             Display this help file and exit.  
-V, --Version          Version of the application
```


Examples:

Usage: OPERA <argument_list>

General command: "OPERA <input> <output> <options> " no specific order for the arguments.

```
>opera -s Sample_50.sdf -o predictions.csv -logP -v 1
```

Using the sample sdf file (50 chemicals) to calculate logP only in minimum verbose mode (v=1). Endpoints names (only) are case insensitive.

```
>opera -s Sample_50.sdf -o predictions.csv
```

Simplest command specifying only the input and output. By default, OPERA will calculate all endpoints and output them to the same file in silent mode (V=0 by default).

```
>opera -d PadelDesc.csv -o predictions.csv -LOGP -BCF -v 2
```

Running one of the previous commands will generate the descriptor file "PadelDesc.csv" for this list of chemicals that you can use to run other models.

```
>opera -d PadelDesc.csv -o predictions.csv -physchem -v 2 -s Sample_50.sdf
```

This runs a list of models at the same time without running all of them. "physchem" or "pc" will run 10 models.

```
>opera -s Sample_50.smi -o predictions.csv -CATMoS -v 1
```

This will run CATMoS models only using "CATMoS" or "AcuteTox" (case insensitive).

```
>opera -s Sample_50.sdf -o predictions.csv -all -n -x -v 1 -c
```

This will run all models. "-n" to get the nearest neighbors and their experimental and predicted data. "-x" will output each one of them in a separate file. "-c" or "-clean" to delete generated descriptors afterwards.

```
>opera -s Sample_50.sdf -d PadelDesc.csv -o predictions.csv -Tox -n -x -v 1
```

This will run all toxicity endpoints including CERAPP (ER), CoMPARA (AR) and CATMoS (AcuteTox).

Running the provided sample structures

```
OPERA_CL

C:\Users\kmansouri\Downloads>opera -s Sample_50.sdf -st -o pred.csv -logp -pKa -logD -EF -catmos -cl -v 1

Endpoints to be calculated:
LOGP, PKA, LOGD, BCF, AOH, BIODEG, RBIODEG, KM, KOC, CATMOS and CL

Initializing and loading models...

===== Structures standardization =====
Input structures: 50.
Generating QSAR-ready structures...
Standardized structures: 50.

===== Molecular Descriptors =====
Loaded structures: 50.
PaDEL calculating 2D descriptors...
PaDEL descriptors calculated for: 50 molecules.
Loading of PaDEL descriptors file...
Checking loaded variables.
Loaded 1444 PaDEL descriptors for 50 molecules.
CDK 2.0 calculating 2D descriptors...
CDK descriptors calculated for: 50 molecules.
Loading of CDK descriptors file...
Checking loaded variables.
Loaded 286 CDK descriptors for 50 molecules.
PaDEL generating fingerprints...
PaDEL fingerprints generated for: 50 molecules.
Reading file with salt information.
The number of molecules with salt information:0

===== Running The Models =====
Predicting LogP values (Log10)...
The provided salt information will be considered in the predictions
Predicting pKa values (unitless)...
Loading of fingerprints file...
Loaded 10145 PaDEL fingerprints for 50 molecules.
Predicting LogD values (Log10)...
----- Env. Fate Endpoints -----
Predicting LogOH values (Log10 cm3/molecule-sec)...
Predicting LogBCF values (Log10)...
Predicting Biodeg. half-life values (Log10 days)...
Predicting Ready-Biodegradability (Binary 0/1)...
Predicting LogKm half-life values (Log10 days)...
Predicting LogKoc values (Log10 L/Kg)...
Predicting Clint values...
Predicting Acute Oral Tox. endpoints (CATMoS)...

===== End Of Calculation =====
50 molecules predicted. Total process time: 00:01:47.
```

Command line

The screenshot shows the OPERA 2.6 graphical user interface. It features a top section for 'Input' and 'Output' with file paths and 'Browse' buttons. Below this is a 'Models' section with checkboxes for various prediction categories: Physchem properties (LogP, MP, BP, VP, WS, HL, KOA, RT, pKa, LogD), Environmental fate (LogBCF, AOH, Biodeg, R-Biodeg, KM, KOC), Toxicity endpoints (ER, AR, AcuteTox), and ADME properties (FUB, Clint). To the right of the models is a 'Standardize' toggle switch set to 'Off'. Below the models is an 'Output options' section with checkboxes for 'Separate files', 'Experimental values', 'Nearest neighbors', 'Include descriptor values', and 'Keep full descriptors files'. On the right side, there is a large vertical logo for 'OPERA Open (q)sar App'. At the bottom right is a 'Calculate' button. A 'Results summary' box on the bottom right provides a detailed log of the process: 'Loaded structures from SDF file: 50', 'Standardized structures: 50', 'Calculated PaDEL descriptors: 1444 (11 sec)', 'Calculated CDK descriptors: 286 (10 sec)', 'Generated PaDEL fingerprints: 10146 (27 sec)', 'Predicted structures: 50 (10 sec)', and 'Total processing time: 00:02:01.'

GUI

Sample outputs:

“predictions.csv”

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	LogP_prec	AD_LogP	AD_index	Conf_index	MP_pred	AD_MP	AD_index	Conf_index	BP_pred	AD_BP	AD_index	Conf_index
2	-1.16675	1	0.480419	0.514066	56.96995	0	0.845536	0.591406	257.4774	0	0.148019	0.459055
3	1.865338	1	0.900462	0.860512	240.0609	1	0.886896	0.77831	391.2263	0	0.156367	0.289241
4	2.220271	1	0.908599	0.854532	225.4009	1	0.868502	0.769243	418.5021	0	0.198701	0.29612
5	1.418476	1	0.888307	0.832005	178.3617	1	0.907997	0.792323	255.0721	0	0.102487	0.429512
6	3.263077	1	0.475522	0.511194	113.4288	1	0.423574	0.396192	303.5064	1	0.218644	0.377263
7	1.250174	1	0.88672	0.826925	140.6504	1	0.870612	0.678357	244.3635	0	0.199114	0.408801
8	1.661618	1	0.884391	0.880548	140.7473	1	0.869315	0.704217	313.0569	0	0.227134	0.451173
9	8.521042	0	0.423516	0.451172	109.0241	1	0.918952	0.808039	388.1919	0	0.288908	0.478521
10	0.794911	1	0.916139	0.831594	53.26655	1	0.889102	0.805115	300.1238	0	0.215153	0.397557
11	-1.34435	1	0.542563	0.511082	70.8003	1	0.862021	0.603289	300.3755	0	0.275189	0.513219
12	-0.76893	1	0.880398	0.785205	45.61057	1	0.87657	0.805093	208.6963	1	0.419094	0.402072
13	2.075537	1	0.885388	0.774448	190.2265	1	0.890244	0.776764	400.3502	0	0.227297	0.322959
14	1.653221	1	0.887992	0.897769	221.5952	1	0.91094	0.869202	389.2867	0	0.161782	0.29391
15	2.635143	1	0.863338	0.699103	270.1702	1	0.856718	0.608162	382.0755	0	0.189731	0.315207
16	3.907586	1	0.87718	0.772113	196.6646	1	0.881577	0.794906	372.7094	1	0.273881	0.367613
17	6.91244	1	0.871424	0.810684	104.6093	1	0.899273	0.844186	350.3782	1	0.549401	0.606396
18	2.64288	1	0.919871	0.754499	157.4414	1	0.94492	0.84299	293.8553	1	0.729356	0.715913
19	3.536705	1	0.611748	0.663621	147.8905	1	0.94981	0.796142	300.6033	1	0.552324	0.69184
20	6.128521	1	0.947304	0.88721	223.6441	1	0.978058	0.708651	485.9157	0	0.937575	0.871229

“predictions.txt”

```
Predicted LogP values...
=====
Molecule 50-01-1:
LogP predicted= -1.167
AD: inside
AD_index= 0.48
Conf_index= 0.51
Molecule 50-02-2:
LogP predicted= 1.865
AD: inside
AD_index= 0.90
Conf_index= 0.86
Molecule 50-04-4:
LogP predicted= 2.220
AD: inside
AD_index= 0.91
Conf_index= 0.85
Molecule 50-06-6:
LogP predicted= 1.418
AD: inside
AD_index= 0.89
Conf_index= 0.83
Molecule 50-10-2:
LogP predicted= 3.263
AD: inside
AD_index= 0.48
Conf_index= 0.51
Molecule 50-11-3:
LogP predicted= 1.250
AD: inside
AD_index= 0.89
Conf_index= 0.83
```

The output format depends
on the specified extension:
csv or txt

For each model, the predicted value is associated with applicability domain and accuracy estimates.

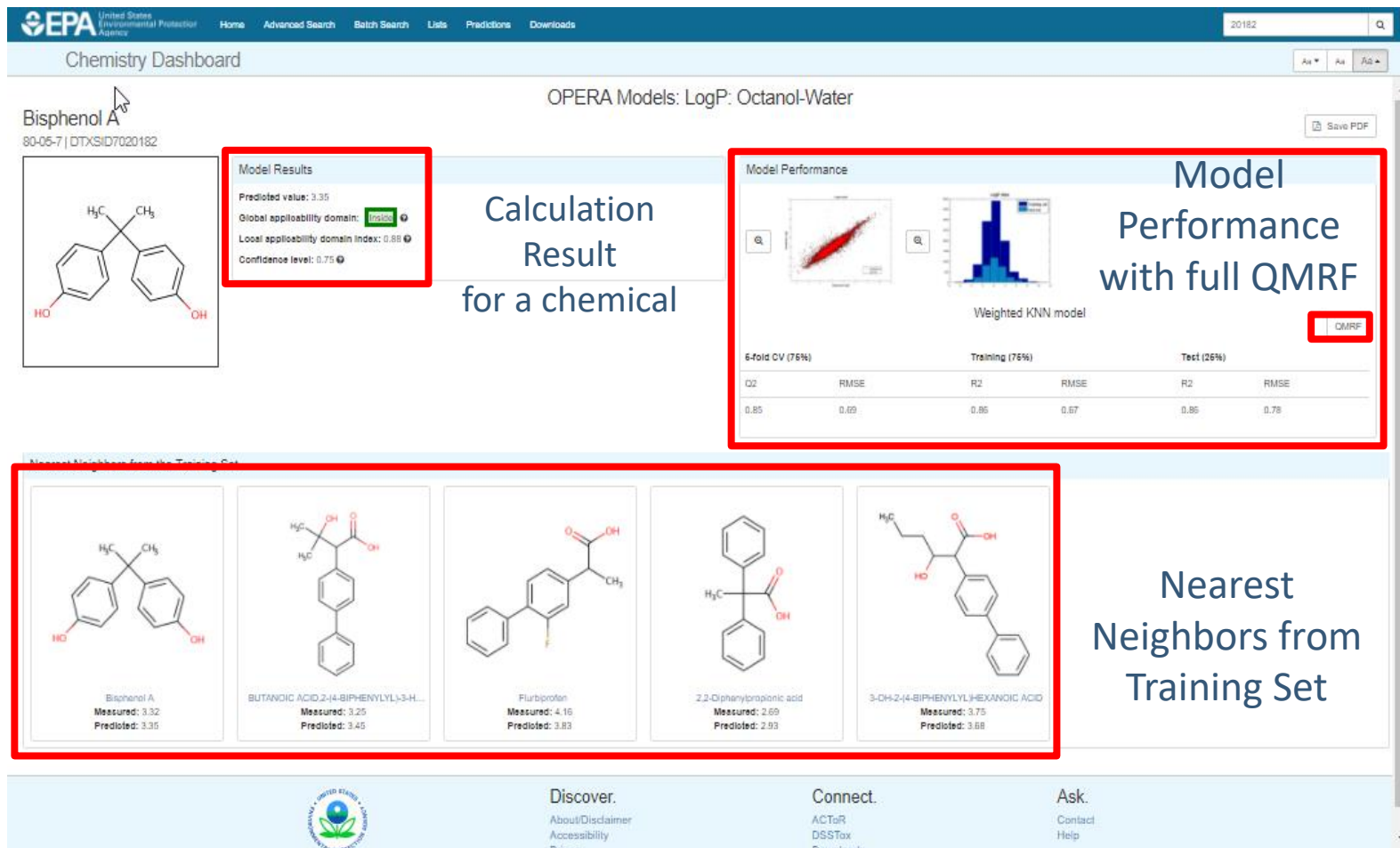
AD_logP (0/1): Global applicability domain considering the whole chemical space of the model.

AD_index [0-1]: Local applicability domain based on the similarity to the 5 nearest neighbors

Conf_index [0-1]: Accuracy estimate based on the predictions of the 5 nearest neighbors.

OPERA prediction report on the EPA Dashboard

<https://comptox.epa.gov/dashboard>



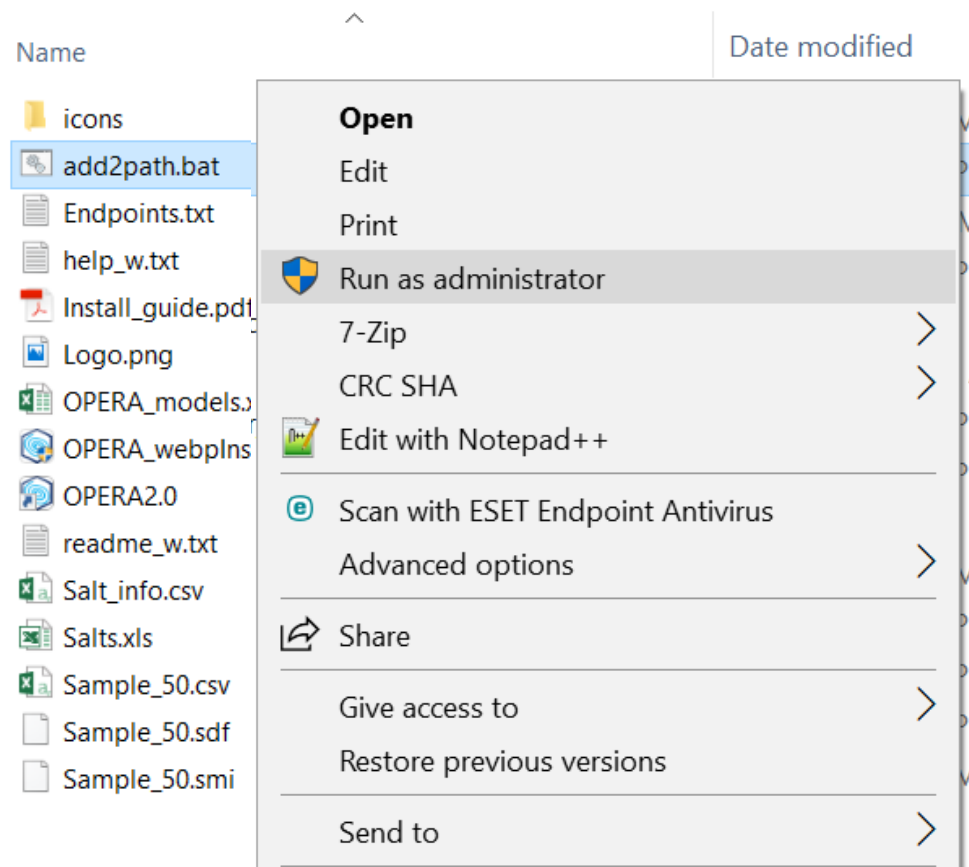
More details in the QMRF reports available on the EPA dashboard, the Github repo and the JRC QSAR database

Optional and only for expert users of command line:

Run add2path.bat as admin to permanently edit system %PATH%.

Only needed to install for all users or to use in a pipeline/ workflow.

Run only ONCE to avoid truncated path!



Thank you!

Register on the GitHub repo for the updates <https://github.com/NIEHS/OPERA>

- [1] Mansouri, K. et al. SAR and QSAR in Env. Res. (2016). <https://doi.org/10.1080/1062936X.2016.1253611>
- [2] Mansouri K. et al. J Cheminform (2018) <https://doi.org/10.1186/s13321-018-0263-1>.
- [3] The CompTox Chemistry Dashboard (<https://comptox.epa.gov/dashboard>)
- [4] Williams A. J. et al. J Cheminform (2017) <https://doi.org/10.1186/s13321-017-0247-6>
- [5] JRC QSAR Model Database <https://qsardb.jrc.ec.europa.eu/qmrf/endpoint>
- [6] Mansouri, K. et al. EHP (2016) <https://doi.org/10.1289/ehp.1510267>
- [7] Mansouri, K. et al. J Cheminform (2019) <https://doi.org/10.1186/s13321-019-0384-1>
- [8] Mansouri, K. et al. EHP (2020) <https://doi.org/10.1289/EHP5580>
- [9] Mansouri, K. et al. EHP (2021) <https://doi.org/10.1289/EHP8495>

Your feedback will be appreciated!
Kamel.mansouri@nih.gov