User Manual

AD-Scan: A Tool for Post-QSAR Applicability Domain Compliance Scanning of Small Molecules

AD-Scan is an open-source application designed for analysing the prediction made by QSAR model. The software is capable of estimating experimental activity range for predicted activity, scanning against applicability domains of QSAR model. It is also capable of scanning of test compound against applicability domain defined by Lipinski's rule of five by set of compounds. The tool is useful for all QSAR model builders, for evaluation of query prediction against applicability domain of model & Lipinski's rule.

Availability: The AD-Scan tools, example data and online documentation are freely accessible at http://csircimap.in/ads/.

Contact: f.khan@cimap.res.in

Workflow: Figure 1 provides details about inputs and outputs concerned with AD-Scan tool. **Tool inputs are:** Number of Observations used in deciding the applicability domain, Dataset for applicability domain, predicted value for query compound through model, number of query variations in query set, Hydrogen bond acceptors, Hydrogen bond donors, Molecular weight, and LogP. **Tool outputs are:** Overall decision for query compound for being in applicability domain, Applicability domain check for various combinations as HBD vs. HBA, MW vs. HBA, LOGP vs. HBA, MW vs. HBD, LOGP vs. HBD, and LOGP vs. MW. Output data was provided in the form of applicability domain score, decision and relative percentage approach towards applicability domain.

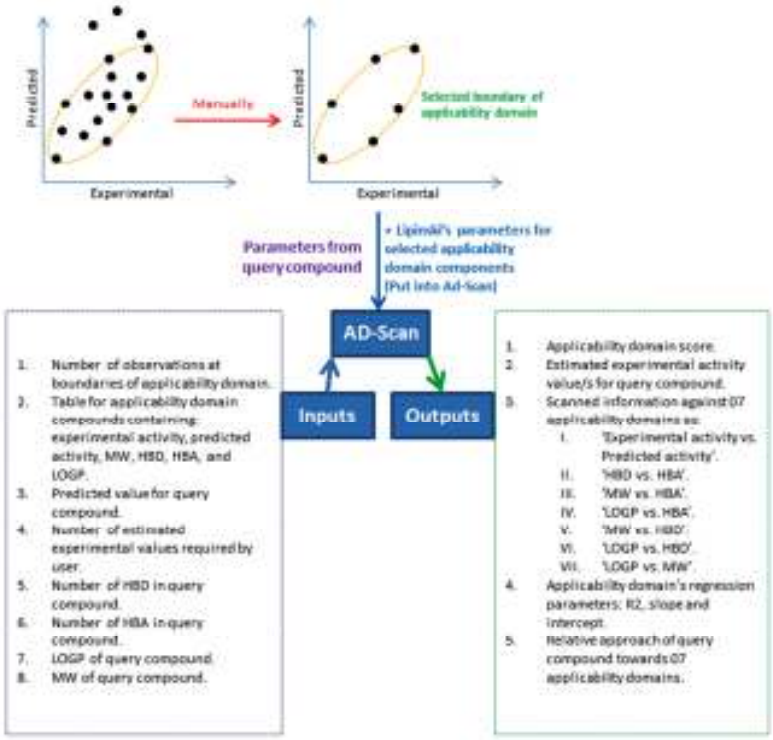

Figure 1. Pictorial representation of workability of AD-Scan

How to run the application: The application is available in two formats: (i) online and (ii) command-line (Figure 2, 3 & 4). Both web & command-line applications need same data entries. These entries are: (i) number of observations at boundaries of applicability domain, (ii) table for applicability domain compounds containing: experimental activity, predicted activity, MW, HBD, HBA, and LOGP, (iii) predicted value for query compound, (iv) number of experimental estimated values required by user, (v) Number of HBD in query compound, (vi) Number of HBA in query compound, (vii) LOGP of query compound and (viii) MW of query compound.

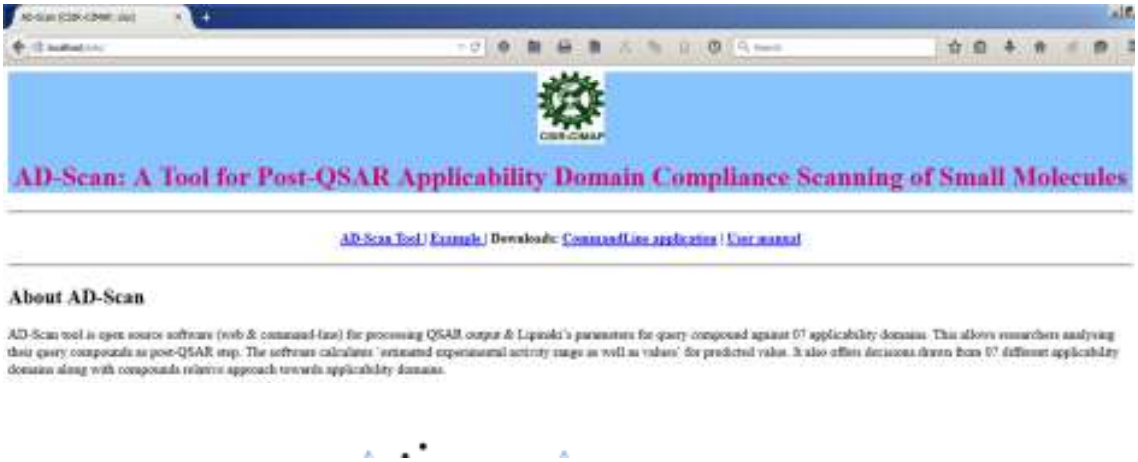

Figure 2. Partial view of website

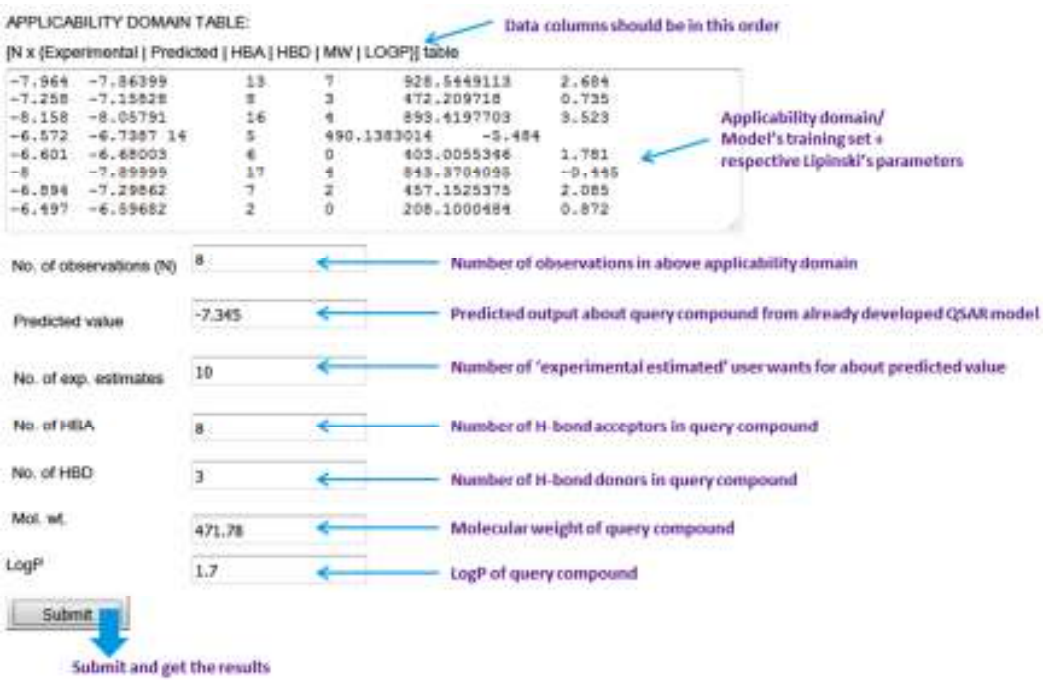

Figure 3. Inputs to AD-Scan web application

To run this program from command line, use following command: >"java AppDomCheck4 <noOfObservations> <ExpPredTxt> <testPredValue> <number of experimental values to be calculated against predicted activity of query molecule> <HBA> <HBD> <MW> <LogP>".

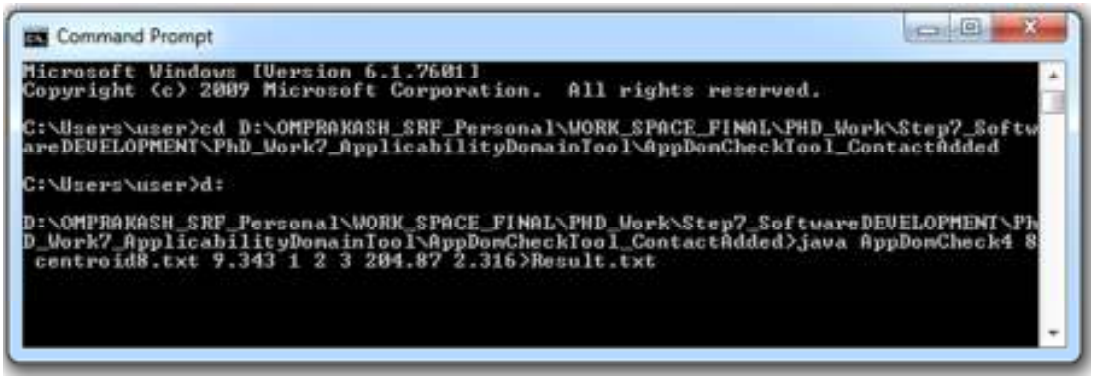

Figure 3. Inputs to AD-Scan Command-line application

Understanding the outputs:

Training set score is the control for screening of query compound. Estimated experimental range is shown below score of training set. Decision about query compound being within applicability domain or not is shown along with relative approach of query compound towards applicability domain. Statistics about applicability domain is also shown. Estimation of query compound against other six applicability domain criteria is followed by these experimental activity results. Details about outputs are shown in Figure 5.

```
=====================================
APPLICABILITY DOMAIN CHECK FOR TESTED COMPOUND
Contact Dr. Feroz Khan (f.khan@cimap.res.in, ferozkhan306@gmail.com) OR Om Prakash (mailboxtop@gmail.com) for any query
Address: Central Institute of Medicinal and Aromatic Plants, P.O.-CIMAP, Near Kukrail Picnic Spot, Lucknow - 226015, India
=====================================

Applicability Domain score With Training set: 9.220149909999995
Test compound's predicted activity range from: 9.505515500054504
Test compound's predicted activity range to: 10.494484494345496

-----------------------------------------------------------------------------------------------
expEstimated,   TestPredEstimation, ADscoreWithTestComp,    ADscoreWithTrainingset, AppDomDecision, Relative (%) approach towards AD,
-----------------------------------------------------------------------------------------------

13.190017770712812, 9.505515500054504,  2.597452896195916, 9.220149909999995,  1.0,    71.82651773589445,
13.315385773634102, 9.604412404063604,  2.559685672512061 4, 9.220149909999995,  1.0,    72.23013389696976,
13.435753776555394, 9.703309303032704,  2.521918448404131,  9.220149909999995,  1.0,    72.64775005804509,
13.556121779476685, 9.802206202021804,  2.484151224294153, 9.220149909999995,  1.0,    73.05736421912021,
13.676489782397976, 9.901103101010904,  2.446384000192685,  9.220149909999995,  1.0,    73.46696238019552,
13.796857785319267, 10.000000000000004, 2.408616776085995,  9.220149909999995,  1.0,    73.87659854127084,
13.917225788240557, 10.098896898989103, 2.370849551979239,  9.220149909999995,  1.0,    74.28621470234597,
14.037593791161848, 10.197793797978203, 2.333082327872509,  9.220149909999995,  1.0,    74.69583086342128,
14.157961794083139, 10.296690696967303, 2.295315103765778 6, 9.220149909999995,  1.0,    75.10544702449666,
14.278329797004443,  10.395587595954403, 2.257547879659048 5, 9.220149909999995,  1.0,    75.51506318557189,
-----------------------------------------------------------------------------------------------
Note: AppDomDecision: '1' suggests existance of test compound's activity within 'Applicability domain'
Note: AppDomDecision: '0' suggests test compound's activity as outlier

Information about Test molecule's predicted activity and regression model plot are as:
Number of observations used for deciding applicability domain: 8
ExpPredTst file name: centroid8.tst
QSAR model's R2: 0.9505515500054504
Test molecule's predicted activity through QSAR model: 10.0
Regression line slope: 0.8216211666631134
Regression line intercept on Y-axis: -1.3357903898590706


=====================================
Applicability domain check for HBD vs HBA
=====================================
Applicability Domain score With Training set: 24.0
```

Control for screening of query compound

Estimated experimental range for query compound

Estimated experimental views
(Here number of estimates depends on user)

Statistics about applicability domain

Above table is the description about analysis of query compound with one applicability domain.

(Similarly proceeded for other six applicability domains)

Figure 5. File outputs delivered from the AD-Scan