Standardize 1.0

NANOBRIDGES
-ACollaborative Project

UNIWERSYTET GDAŃSKI

Marie Curie Actions
EUROPEAN UNION

THE AUTHORS ARE GRATEFUL FOR THE FINANCIAL SUPPORT FROM THE EUROPEAN COMMISSION THROUGH THE MARIE CURIE **IRSES** PROGRAM, NANOBRIDGES PROJECT (FP7-PEOPLE-2011-IRSES, GRANT AGREEMENT NUMBER 295128).

Last updated: 22.05.2015

Standardize 1.0

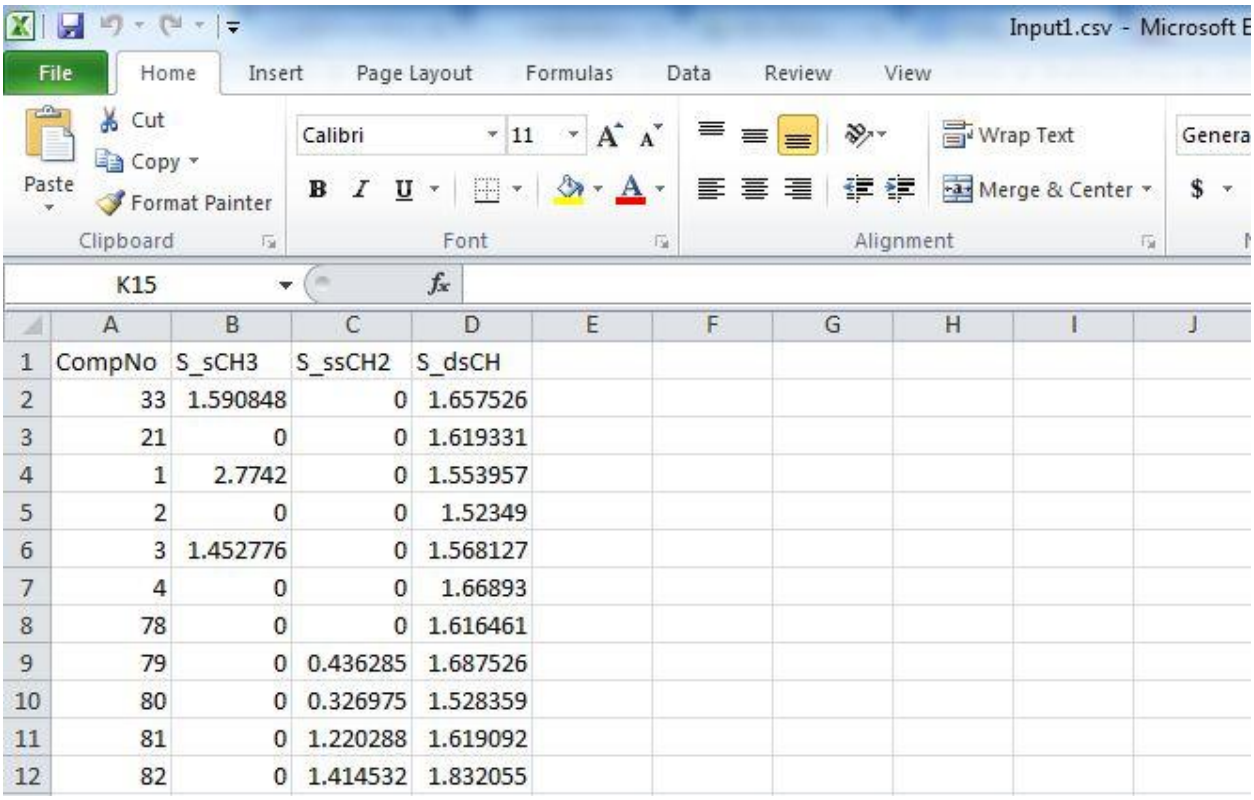
Background: The standardization of dataset means adjusting the values measured on different scale/units to a uniform scale.The data preprocessing plays a crucial role while using the dataset for different chemometric studies such QSAR, data mining etc**.** This step is very important when dealing with parameters of different units and scales.

Formula/Equation: The standardized value of descriptor is calculated by the following equation:

$$D_{std} = \sum_{i=0}^{n} \frac{D_i - D_{mean}}{SD} \quad (1)$$

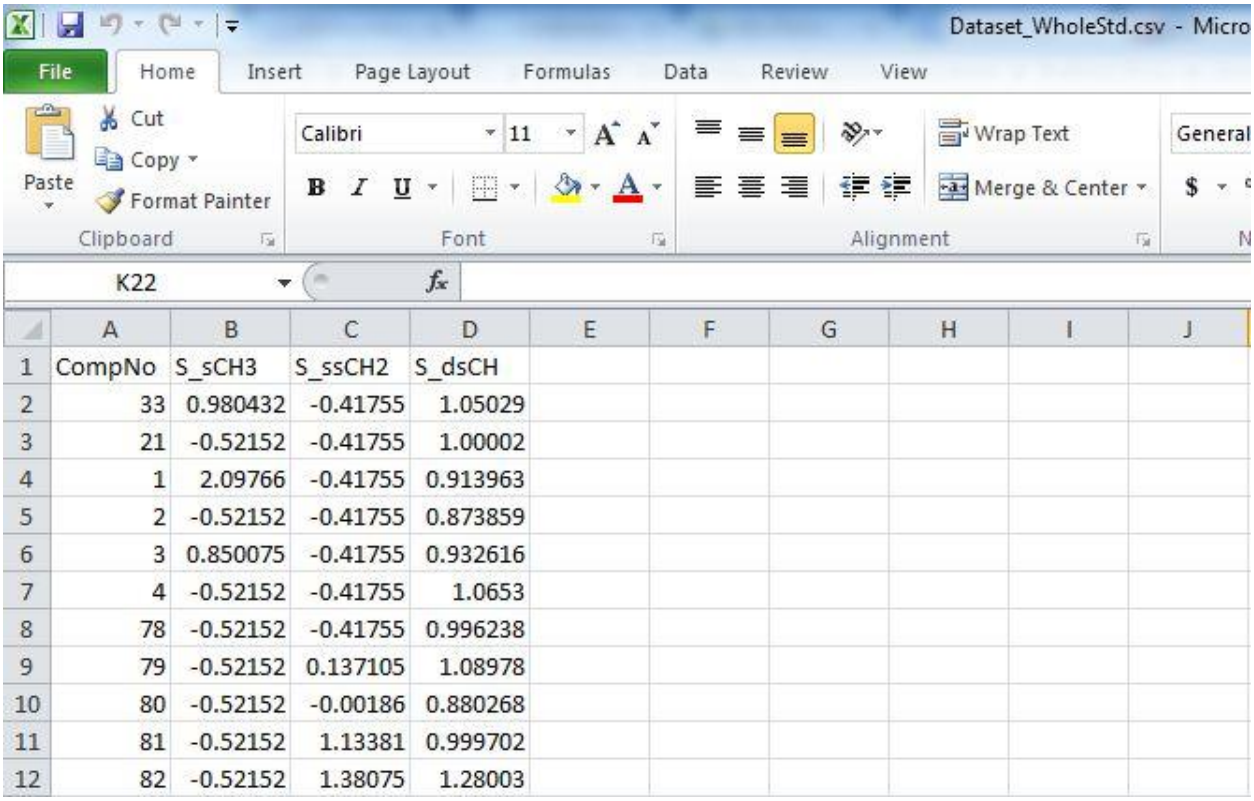
In Eq (1), $D_{std}$is the scaled/standardized descriptor, D_i is the unscaled descriptor, and D_{mean} and SD are the descriptor average and standard deviation.

Input file format: This program takes a input file in CSV format.



	A	B	C	D	E	F	G	H	I	J
1	CompNo	S_sCH3	S_ssCH2	S_dsCH						
2	33	1.590848	0	1.657526						
3	21	0	0	1.619331						
4	1	2.7742	0	1.553957						
5	2	0	0	1.52349						
6	3	1.452776	0	1.568127						
7	4	0	0	1.66893						
8	78	0	0	1.616461						
9	79	0	0.436285	1.687526						
10	80	0	0.326975	1.528359						
11	81	0	1.220288	1.619092						
12	82	0	1.414532	1.832055						

Output file format:



Dataset_WholeStd.csv - Micro

File | Home | Insert | Page Layout | Formulas | Data | Review | View

Cut, Copy, Format Painter, Paste — Clipboard

Calibri, 11 — Font

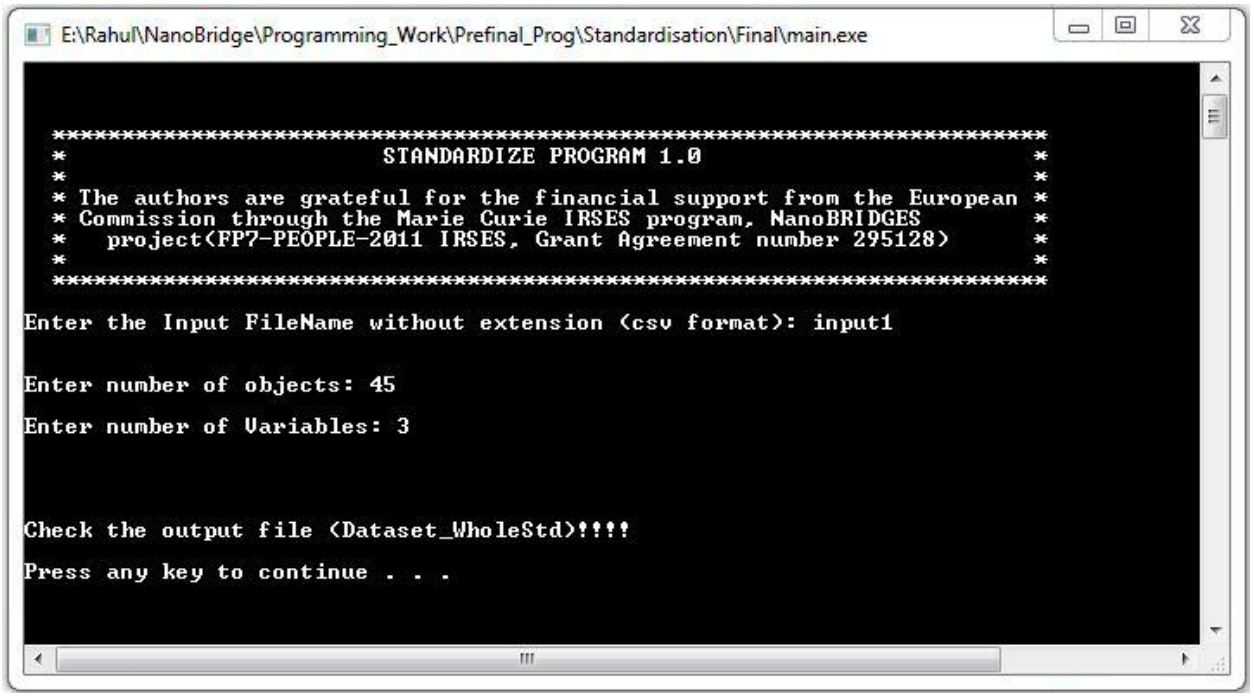
Wrap Text, Merge & Center — Alignment

General

K22

	A	B	C	D	E	F	G	H	I	J
1	CompNo	S_sCH3	S_ssCH2	S_dsCH						
2	33	0.980432	-0.41755	1.05029						
3	21	-0.52152	-0.41755	1.00002						
4	1	2.09766	-0.41755	0.913963						
5	2	-0.52152	-0.41755	0.873859						
6	3	0.850075	-0.41755	0.932616						
7	4	-0.52152	-0.41755	1.0653						
8	78	-0.52152	-0.41755	0.996238						
9	79	-0.52152	0.137105	1.08978						
10	80	-0.52152	-0.00186	0.880268						
11	81	-0.52152	1.13381	0.999702						
12	82	-0.52152	1.38075	1.28003						

How to run a program:



```
E:\Rahul\NanoBridge\Programming_Work\Prefinal_Prog\Standardisation\Final\main.exe

 ********************************************************************************
 *                        STANDARDIZE PROGRAM 1.0                               *
 *                                                                              *
 * The authors are grateful for the financial support from the European         *
 * Commission through the Marie Curie IRSES program, NanoBRIDGES                *
 *   project(FP7-PEOPLE-2011 IRSES, Grant Agreement number 295128)              *
 *                                                                              *
 ********************************************************************************

Enter the Input FileName without extension (csv format): input1

Enter number of objects: 45

Enter number of Variables: 3



Check the output file (Dataset_WholeStd)!!!!

Press any key to continue . . .
```

Disclaimer

For academic purpose only.

The program Standardize has been developed in C++ language and is validated on the known data sets. This program is compatible with both 32- and 64-bit Windows operating system. Please report for discrepancy of result for any other dataset. Contact us at any of the following addresses:

Dr. Tomasz Puzyn,
NanaBRIDGESProject
Coordinator,
Faculty of Chemistry,
University of Gdansk,
Gdansk,
Poland 80-952
Email Id: puzi@qsar.eu.org

Dr. Kunal Roy,
Drug Theoretics and Cheminformatics Lab.,
Dept. of Pharmaceutical Technology,
Jadavpur University,
Kolkata, West Bengal,
INDIA-700032
Email Id: kunalroy_in@yahoo.com

Rahul BalasahebAher,
PhdResearch Scholar,
Drug Theoretics and Cheminformatics Lab.,
Dept. of Pharmaceutical Technology,
Jadavpur University,
Kolkata, West Bengal,
INDIA-700032
E-mail Id: rahulba26@gmail.com(*for any queries regarding the program)

References:

1. Aher RB, Roy K. Understanding the structural requirements in diverse scaffolds for the inhibition of P. falciparum Dihydroorotate Dehydrogenase (*Pf*DHODH) using 2D-QSAR, 3D-Pharmacophore and Structure-Based energy-optimized pharmacophore models. Combinatorial Chemistry & High Throughput Screening. http://benthamscience.com/journal/abstracts.php?journalID=cchts&articleID=127211