Supplementary information

InCHlib – Interactive Cluster Heatmap for web applications

Škuta C.[1,2], Bartůněk, P.[2] and Svozil, D.[1,2*]

[1] Laboratory of Informatics and Chemistry, Faculty of Chemical Technology, Institute of Chemical Technology Prague, Technická 5, CZ-166 28 Prague, Czech Republic

[2] CZ-OPENSCREEN, Institute of Molecular Genetics of the ASCR, v. v. i., Vídeňská 1083, CZ-142 20 Prague, Czech Republic

Contact: svozild@vscht.cz

1. *inchlib_clust* clustering options

inchlib_clust is a Python script that performs data clustering and prepares input data for *InCHlib*. *inchlib_clust* can be used both from the command line or *Python* code. Its documentation and use examples are available from the *InCHlib* website [1]. *inchlib_clust* performs a hierarchical clustering using *fastcluster* [2] library. The following linkages and distance measures are available.

Linkages:

1. single - the single/min/nearest algorithm
2. complete - the complete/max/farthest algorithm
3. average - the average/UPGMA algorithm
4. weighted - the weighted/WPGMA algorithm
5. centroid - the centroid/UPGMC algorithm
6. median - the median/WPGMC algorithm
7. ward - the Ward/incremental algorithm

Distance measures:

1. braycurtis - the Bray-Curtis distance
2. canberra - the Canberra distance
3. chebyshev - the Chebyshev distance
4. cityblock - the Manhattan distance
5. correlation - the Correlation distance
6. cosine - the Cosine distance
7. dice - the Dice dissimilarity (boolean)
8. euclidean - the Euclidean distance
9. hamming - the Hamming distance (boolean)
10. jaccard - the Jaccard distance (boolean)
11. kulsinski - the Kulsinski distance (boolean)
12. mahalanobis - the Mahalanobis distance
13. matching - the matching dissimilarity (boolean)
14. minkowski - the Minkowski distance
15. rogerstanimoto - the Rogers-Tanimoto dissimilarity (boolean)
16. russellrao - the Russell-Rao dissimilarity (boolean)
17. seuclidean - the normalized Euclidean distance
18. sokalmichener - the Sokal-Michener dissimilarity (boolean)
19. sokalsneath - the Sokal-Sneath dissimilarity (boolean)
20. sqeuclidean - the squared Euclidean distance
21. yule - the Yule dissimilarity (boolean)

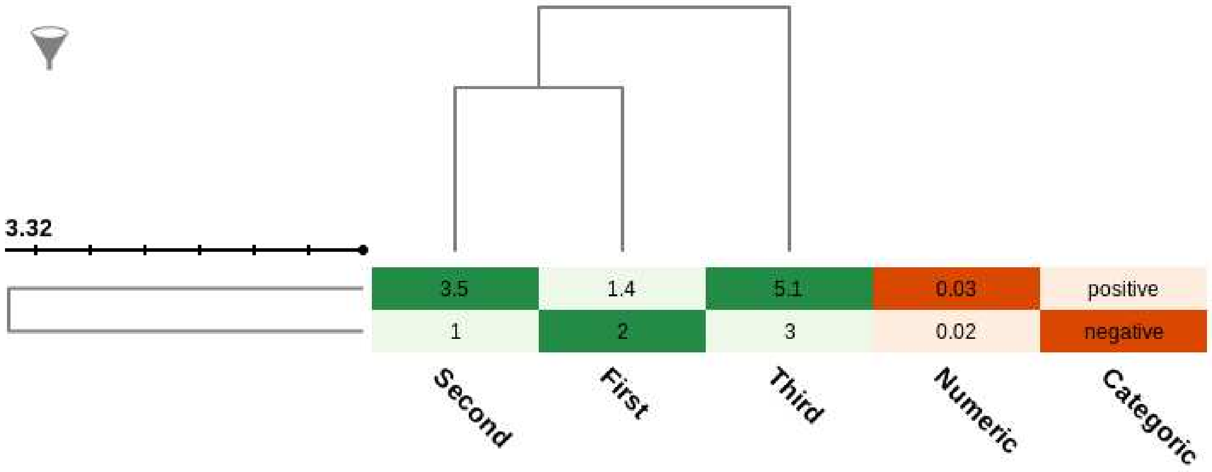
2. *InCHlib* deployment

In this section, a short example demonstrates how to integrate *InCHlib* into a web page. The commented *HTML/JavaScript* code is followed by the rendered cluster heatmap. The input data file example.json is given in Additional file 1.

```
<html>
    <head>
        <script src="/path/to/jquery-2.0.3.min.js"></script> //import jQuery
        <script src="/path/to/kinetic-v5.0.0.min.js"></script> //import KineticJS
        <script src="/path/to/inchlib-1.0.0.js"></script> //import InCHlib
        <script>
        $(document).ready(function() { //execute when the web page is ready
            var inchlib = new InCHlib({ //InCHlib instantiation
                target: "inchlib", //define target element
                metadata: true, //input file contains metadata
                column_dendrogram: true, //input file contains column dendrogram
            });

            inchlib.read_data_from_file("/path/to/example.json"); //read data from file
            inchlib.draw(); //draw cluster heatmap
        });
        </script>
    </head>

    <body>
        <div id="inchlib"></div> //target element
    </body>
 </html>
```

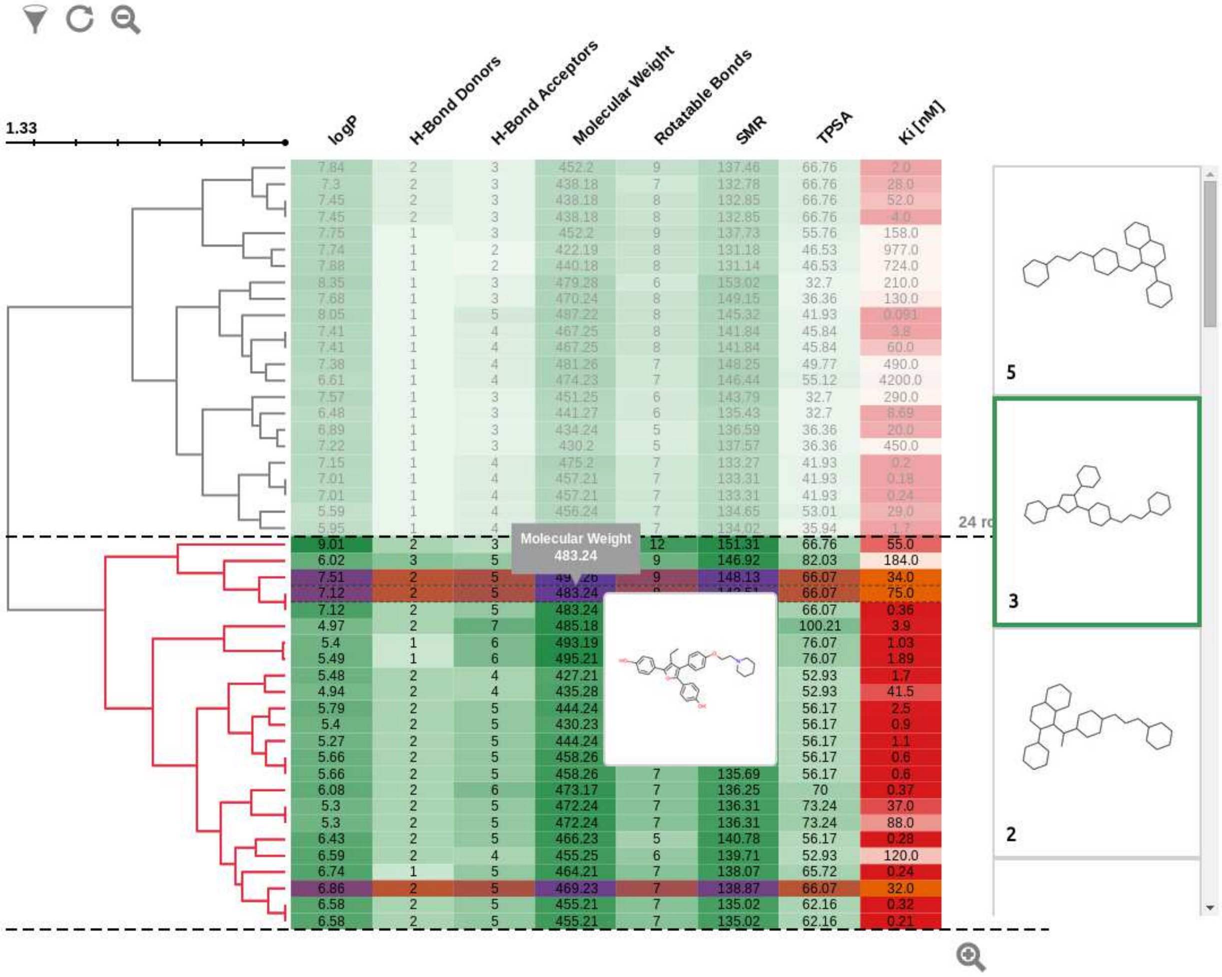


3. Use examples

In this section, various use cases demonstrating *InCHlib's* versatility are presented.

3.1. Chemical biology data

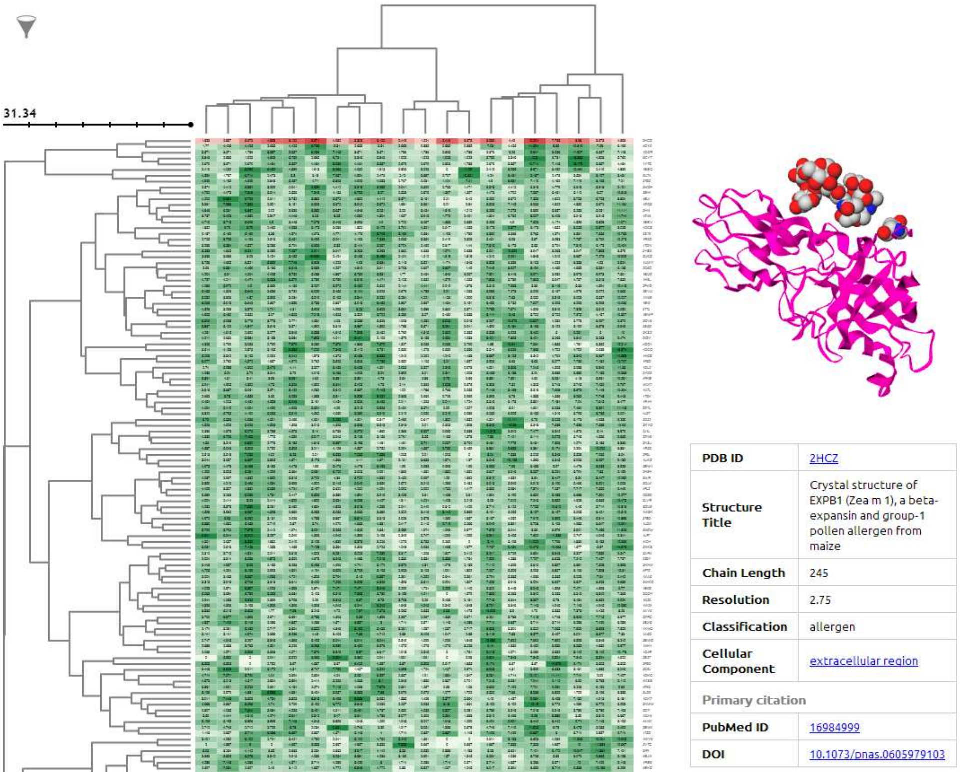
InChlib use in chemical biology is demonstrated by this example. The data set consists of 8 physico-chemical and structural properties of 195 *estrogen receptor alpha (ERα)* ligands obtained from the *ChEMBL* database [3]. The ligand properties are: the logarithm of the octanol-water partition coefficient (*logP*), molar refractivity (*SMR*), topological polar surface area (*TPSA*), molecular weight, and number of rotatable bonds, hydrogen-bond donors, hydrogen-bond acceptors and aromatic rings. Each ligand is also characterized by its K_i value (equilibrium dissociation constant determined in inhibition studies) that is considered as the medata.



logP	H-Bond Donors	H-Bond Acceptors	Molecular Weight	Rotatable Bonds	SMR	TPSA	Ki [nM]
7.84	2	3	452.2	9	137.46	66.76	2.0
7.3	2	3	438.18	7	132.78	66.76	28.0
7.45	2	3	438.18	8	132.85	66.76	52.0
7.45	2	3	438.18	8	132.85	66.76	4.0
7.75	1	3	452.2	9	137.73	55.76	158.0
7.74	1	2	422.19	8	131.18	46.53	977.0
7.88	1	2	440.18	8	131.14	46.53	724.0
8.35	1	3	479.28	6	153.02	32.7	210.0
7.68	1	3	470.24	8	149.15	36.36	130.0
8.05	1	5	497.22	8	145.32	41.93	0.091
7.41	1	4	467.25	8	141.84	45.84	3.8
7.41	1	4	467.25	8	141.84	45.84	60.0
7.38	1	4	481.26	7	148.25	49.77	490.0
6.61	1	4	474.23	7	146.44	55.12	4200.0
7.57	1	3	451.25	6	143.79	32.7	290.0
6.48	1	3	441.27	6	135.43	32.7	8.69
6.89	1	3	434.24	5	136.59	36.36	20.0
7.22	1	3	430.2	5	137.57	36.36	450.0
7.15	1	4	475.2	7	133.27	41.93	0.2
7.01	1	4	457.21	7	133.31	41.93	0.18
7.01	1	4	457.21	7	133.31	41.93	0.24
5.59	1	4	456.24	7	134.65	53.01	29.0
5.95	1	4	[illegible]	7	134.02	35.94	1.7
9.01	2	3	[illegible]	12	151.31	66.76	55.0
6.02	3	5	[illegible]	9	146.92	82.03	184.0
7.51	2	5	[illegible]	9	148.13	66.07	34.0
7.12	2	5	483.24	[illegible]	[illegible]	66.07	75.0
7.12	2	5	483.24	[illegible]	[illegible]	66.07	0.36
4.97	2	7	485.18	[illegible]	[illegible]	100.21	3.9
5.4	1	6	493.19	[illegible]	[illegible]	76.07	1.03
5.49	1	6	495.21	[illegible]	[illegible]	76.07	1.89
5.48	2	4	427.21	[illegible]	[illegible]	52.93	1.7
4.94	2	4	435.28	[illegible]	[illegible]	52.93	41.5
5.79	2	5	444.24	[illegible]	[illegible]	56.17	2.5
5.4	2	5	430.23	[illegible]	[illegible]	56.17	0.9
5.27	2	5	444.24	[illegible]	[illegible]	56.17	1.1
5.66	2	5	458.26	[illegible]	[illegible]	56.17	0.6
5.66	2	5	458.26	7	135.69	56.17	0.6
6.08	2	6	473.17	7	136.25	70	0.37
5.3	2	5	472.24	7	136.31	73.24	37.0
5.3	2	5	472.24	7	136.31	73.24	88.0
6.43	2	5	466.23	5	140.78	56.17	0.28
6.59	2	4	455.25	6	139.71	52.93	120.0
6.74	1	5	464.21	7	138.07	65.72	0.24
6.86	2	5	469.23	7	138.87	66.07	32.0
6.58	2	5	455.21	7	135.02	62.16	0.32
6.58	2	5	455.21	7	135.02	62.16	0.21

3.2. 3D protein structure

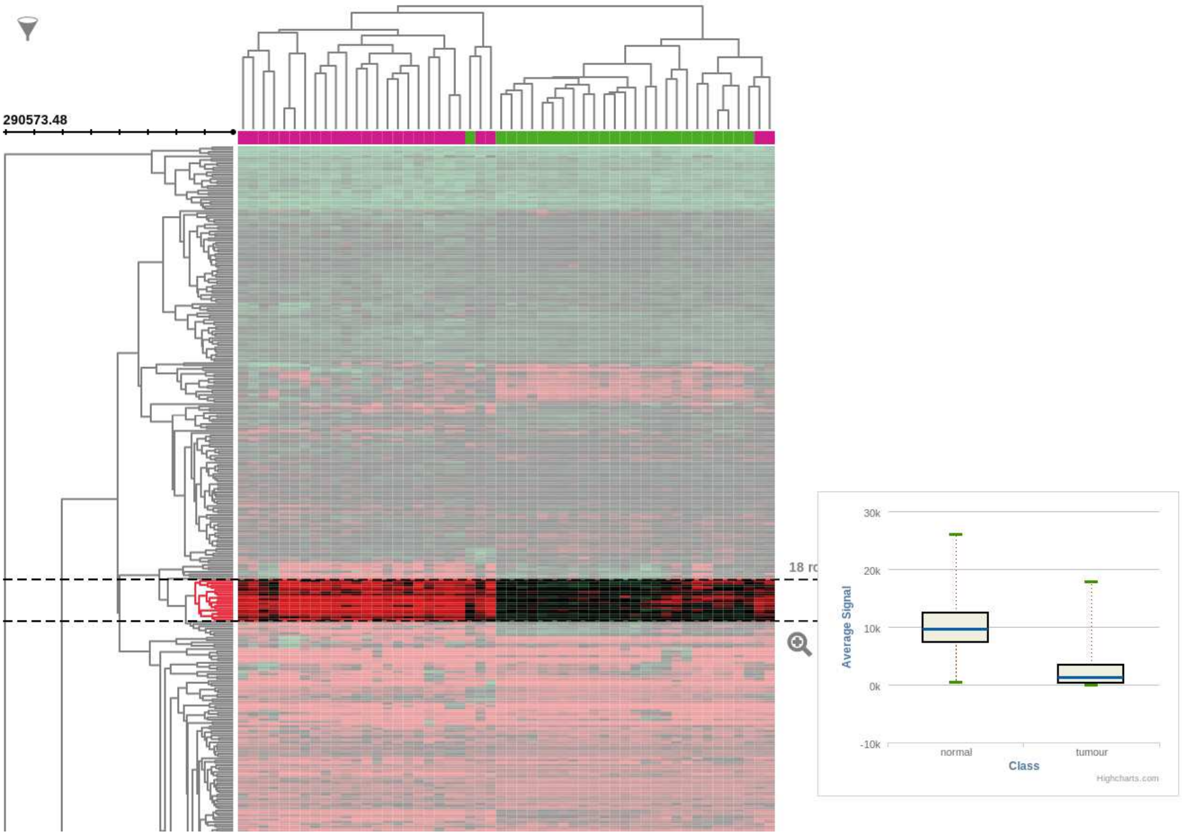
This example demonstrates how to combine *InCHlib* with 3D protein structure visualization. Each protein is characterized by the percentage of its 20 amino acid (green data matrix). There are no metadata in this example. 3D structure visualization is delivered by the *GLmol 3D viewer* [4]. Various information about the structure with links to external databases are summarized in the table below the 3D visualization.



PDB ID	2HCZ
Structure Title	Crystal structure of EXPB1 (Zea m 1), a beta-expansin and group-1 pollen allergen from maize
Chain Length	245
Resolution	2.75
Classification	allergen
Cellular Component	extracellular region
Primary citation	
PubMed ID	16984999
DOI	10.1073/pnas.0605979103

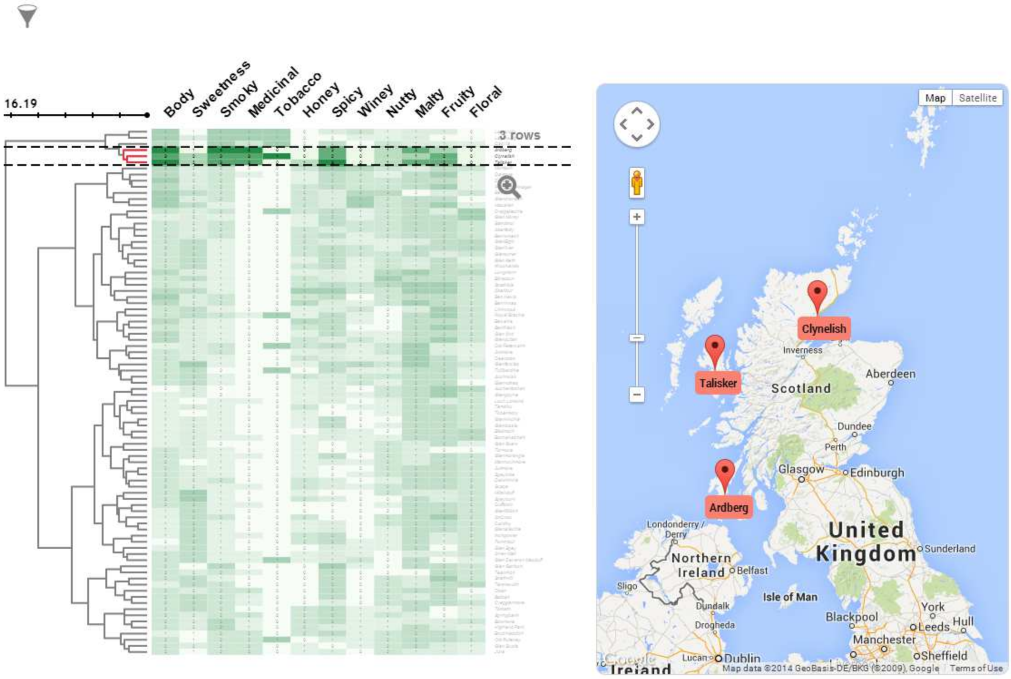
3.3. Microarray data

In this example, use of *InCHlib* for microarray data analysis is demonstrated. The expression levels of 48,803 genes (rows) were measured for 52 tumour and normal samples (columns) [5]. Original 48,803 genes were compressed to 500 rows. Column metadata distinguish between normal (green) and tumour (magenta) samples. The boxplot on the right, provided by the *Highcharts* library [6], shows expression level distribution in tumour and normal classes.



3.4. Scotch whisky distilleries

This example demonstrates the *InChlib* versatility. Data do not originate from a biomedical domain, instead clustering of Scotch whiskies by their twelve taste characteristics was performed. In this example, *InCHlib* is integrated with the *Google map* service.



4. References

1. **inchlib_clust Documentation** [http://openscreen.cz/software/inchlib/inchlib_clust]
2. Müllner D: **fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python**. *Journal of Statistical Software* 2013, **53**(9):1-18.
3. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B *et al*: **ChEMBL: a large-scale bioactivity database for drug discovery**. *Nucleic acids research* 2012, **40**(Database issue):D1100-1107.
4. **GLmol Molecular Viewer** [http://webglmol.sourceforge.jp/index-en.html]
5. Kolar M, Szabo P, Dvorankova B, Lacina L, Gabius HJ, Strnad H, Sachova J, Vlcek C, Plzak J, Chovanec M *et al*: **Upregulation of IL-6, IL-8 and CXCL-1 production in dermal fibroblasts by normal/malignant epithelial cells in vitro: Immunohistochemical and transcriptomic analyses**. *Biology of the cell / under the auspices of the European Cell Biology Organization* 2012, **104**(12):738-751.
6. **Highcharts JS** [http://www.highcharts.com/]