Good Practice in Traditional Chinese Medicine Research in the Post-genomic Era

GP-TCM

223154

D4.13

Handbook of guidelines for use of *in-silico* tools in CHM research

Document description	
Name of document	Handbook of guidelines for use of *in-silico* tools & databases in CHM research
Abstract	Handbook of guidelines for use of *in-silico* tools and databases in CHM research.
Document identifier	D4.13
Document class	Deliverable
Version	1
Author(s)	David Barlow, Tom Ehrman, Alessandro Buriani
Date of creation	21st April 2011
Date of last modification	30th April 2011
Status	Final
Destination	European Commission
WP number	WP4

TABLE OF CONTENTS

1 HANDBOOK OF GUIDELINES FOR USE OF IN-SILICO TOOLS IN CHM RESEARCH

1.1 Criteria to guide the design and choice of CHM database for *in-silico* studies

Criteria that might be used (a) in defining the content to be acquired in creation of a new CHM database, and (b) to score the utility of an existing CHM database, were established in WP4 D4.12.

The database features deemed to be of greatest importance were identified as:

1. those relating to its content – particularly in terms of the chemical structure data held (criteria C1-C11);

2. the versatility of Boolean searches that are catered for – the ideal being that searches involving multiple search fields are allowed, one of which is preferably a molecular structure based search field (criteria C12-C14),

3. the nature of the allowed chemical structure output – the ideal being that multiple chemical structures can be output, each in 3D format (criteria C15-C16).

The maximum score achievable is 50.

Table 1 – Quality evaluation criteria for CHM databases

Evaluation Criteria	*Description of criteria*	*Score*
C1	Contains information on CHM usage?	**2**
C2	Contains botanical information of plants used in CHM?	**2**
C3	Contains toxicological data on CHM phytochemicals?	**2**
C4	Contains (known) biological activity data on CHM phytochemicals?	**2**
C5	Contains (predicted) biological activity data on CHM phytochemicals?	**2**
C6	Contains chemical structure data on CHM phytochemicals?	**2**
C7	Contains 2D structures of CHM phytochemicals?	**2**
C8	Contains 3D structures of CHM phytochemicals?	**2**
C9	Contains data on the molecular mechanism(s) of CHM phytochemicals?	**2**
C10	Data on >1,000 unique CHM phytochemicals	**2**
C11	Database is live, curated and updated?	**2**
C12	Database can be queried by chemical structure?	**2**
C13	Searches with ≥ 2 search fields possible?	**2**
C14	Searches can be performed using combinations of search fields?	**2**
C15	Chemical structure(s) can be output in 1D (as SMILES strings)?	**1**
	Chemical structure(s) can be output in 2D?	**2**
	Chemical structure(s) can be output in 3D (as single conformers)?	**4**
	Chemical structure(s) can be output in 3D (as conformer libraries)?	**8**
C16	Facility for (batch) output of multiple chemical structures	**5**
C17	Database is reported in an article in a peer reviewed journal	**2**

1.2 Criteria to guide the choice and use of *in-silico* analytical tools in CHM research

1. *In-silico* tools should be applied in CHM research through close collaboration between computational specialists and CHM experimentalists.

2. The usefulness and choice of *in-silico* screening tools is dependent on the reason for their use and/or the nature of the output to be generated. The knowledge of the computational specialist is essential in guiding decisions made here.

3. *In silico* tools are best applied in CHM research as a means: to seek out potential mechanisms of action of their constituents; to identify putative new leads for drugs; and to summarise and/or visualize the complex patterns embedded within the output generated through associated 'omics studies. Given the complexity of CHM, and the ensuing difficulties in performance of experimental studies, *in-silico* studies clearly offer an economical and efficient way of exploring the problem landscape and thus helping to define suitable hypotheses for subsequent testing in (*in-vitro* and/or *in-vivo*) laboratory studies.

4. Several categories of software are of relevance in CHM research. Those of use for virtual screening and/or the identification of potential mechanisms of action of CHM constituenst include:

1. **Ligand based screening programs** Pre-requisite(s) for use: knowledge of compounds with known activity; use: to identify putatively active compounds; tools available: classification/regression trees (including Random Forest), linear discriminants analysis, artificial neural networks, support vector machines.

2. **Pharmacophore programs** Can be either ligand-based (LB), or target-based (TB) (the latter being superior/preferable); pre-requisite(s) for use: 3D structures of known ligands to chosen targets (LB), or known 3D structures of target protein(s), and ideally known 3D structure(s) of known complex(es)) (TB); use: to identify putative active compounds; programs available: LiganScout[1]

3. **Docking programs** Pre-requisites for use: known 3D structure(s) of target proteins; use: to 'dock' potential small molecule ligands into protein active sites, optimising their topographical and chemical complementarity, and scoring their interaction. Programs available: FlexX[2], Gold[3], Dock[4], Glide[5], MolDock[6], AutoDock[7], and LigandFit[8].

Other types of software tool include:

4. **Pattern recognition software** Use: post-screening analyses (involving dimensionality reduction); algorithms employed: principle components analysis, multi-dimensional scaling, self-organising maps, and various forms of cluster analysis.

5. **Proteomics and/or genomics data visualisation and analysis tools** Use: Application specific programs for statistical processing and visualisation of data output from DNA micro-array experiments, MS proteomics experiments etc.

1.3 References

[1] Langer, T Wolber, G (2004) Virtual combinatorial chemistry and in silico screening: efficient tools for lead discovery? Pure Appl Chem, 76: 991-996.

[2] Rarey, M, Wefing, S, Lengauer, T (1996) Placement of medium-sized molecular fragments in active sites of proteins. J. Comp.-aided Mol. Design 10: 41-54.

[3] Jones, G, Willett, P, Glen, RC, Leach, AR, Taylor, R (1997) Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267: 727-748.

[4] Shoichet, BK, Kuntz, ID (1993) Matching chemistry and shape in molecular docking. Prot. Eng. 6: 723-732.

[5] Halgren, TA, Murphy, RB, Friesner, RA, Beard, HS, Frye, LL, Pollard, WT, Banks, JL (2004) Glide: a new approach for rapid, accurate docking and scoring: 1. Method assessment of docking accuracy. J. Med. Chem. 47: 1739-1749.#

[6] Thomsen, R, Christensen, MH (2006) MolDock: a new technique for high-accuracy molecular docking. J. Med. Chem. 49: 3315-3321.

[7] Goodsell, DS, Morris, GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. J. Mol. Recogn. 9: 1-5.

[8] Venkatachalam, CM, Jiang, X, Oldfield, T, Waldman, M (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J. Mol Graphics Model. 21: 289-307.