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

GP-TCM

223154

D4.4, D4.5, D4.6 and D4.7

D4.4 Report of the discussion group on biological target oriented database

D4.5 Biological target oriented database for in vitro research on CHM

D4.6 Report on quality criteria and scoring of the CHM database

D4.7 Update of CHM target oriented database with quality scores

Document description	
Name of document	Because of the high volume of work it has been decided to subsume the four deliverable reports into one document. Thus this current report covers the achievements of work package 4 in all the following areas: D4.4 Report of the discussion group on biological target oriented database D4.5 Biological target oriented database for *in vitro* research on CHM D4.6 Report on quality criteria and scoring of the CHM database D4.7 Update of CHM target oriented database with quality scores
Abstract	The report details the activities of the group in the 18 month period.
Document identifier	D4.4;D4.5;D4.6;D4.7
Document class	Deliverable
Version	1
Author(s)	A. Buriani, P. J. Hylands and J. Zhou
Date of creation	13 April 2010
Date of last modification	4 November 2010
Status	Final
Destination	European Commission
WP number	WP4

Contents

1 D4.4: REPORT OF THE DISCUSSION GROUP ON BIOLOGICAL TARGET ORIENTED DATABASE

1.1 Background

Following the kick-off meeting held at the host institution, King's College London, of the Coordinator of WP4 in October 2009, it had been planned to hold the follow-up meeting in February or March 2010. However, due to administrative problems and work pressures it has not been possible to hold a face-to-face meeting. Nonetheless the Coordinator had coordinated a collective discussion with other members and collected their opinions and feedbacks. The discussion was then developed further during the follow-up meeting and the general GP-TCM meeting in July 2010.

1.2 Results of the discussion group

Given that it proved impossible to arrange a second WP4 meeting convenient for all participants early in 2010, the discussion developed among the work package members via e-mails and personal contacts, and the discussion was finalised in July 2010, during the follow-up meeting in London and the general GP-TCM meeting which took place in the following days at Henley.

As a first decision, it was agreed that rather than a real database, a data list, even in a simple form such as MS Office Access, having a simple search engine sufficient to the scope of the task, and that it should be implemented in the GP-TCM website.

The online repository cannot be orientated solely on molecular targets, since the choice of the papers will be based on disease areas and/or plants, so the first level of the entries should be disease and plant, while molecular targets could be one of the subheadings or could be listed among the characteristics for each paper.

Unlike the scoring procedure, for which the application of inclusion/exclusion criteria should be used (see D4.6), the only restriction for the addition of a scientific paper in the repository is language and that it is listed in major databases like PubMed. It is suggested that English should be the only language admitted for the repository, at least initially. Later a working group could be organised to select papers in different languages and provide a proper translation. Initially the following suggestion was given on how to focus the selection of the papers:

1st step: The diseases should be agreed on with the other two WPs dealing with pharmacology (WP5 and WP6; as far as it could be anticipated at the time the two most likely disease areas should be cancer and diabetes).

2nd step: the choice of phytocomplexes (from plant or traditional formulas) to be taken into consideration should be guided by their application on the chosen disease areas.

3rd step: The amount of literature and reviews should be then taken into account, but the final choice of the herbal preparations should be agreed with WP1 and WP2.

4th step: WP3 should be given indication of the herbal preparations chosen for this initial phase, so that they can concentrate their efforts on those. It was recognised that a uniform database will otherwise be very difficult to organize and also may cause problems with synergies between the work packages. The most important criterion is that the final choice of the herbal preparations should be agreed with WP1, WP2 and WP3, so that all four WPs can concentrate their efforts on those.

5th step: before making accessible to the public a disease area, a minimum of one hundred papers should be listed, including clinical and *in-vivo* studies (at this stage it seemed really important that all the WPs be coordinated and collaborate regarding the scope). It is recognised that this is an arbitrarily chosen number of articles but it is felt that this is the

minimum that might provide a representative result. This number will of course grow as time passes.

Papers will be added as pdfs, which cannot be searched with common search engines. Thus each paper, before being added to the repository, should also be indexed so that it can usefully be traced using simple key words. These should list: TCM formula names, TCM plant names, Latin binomial and common names, single molecules, disease name, syndrome, molecular target, authors, institutions, etc. An alternative might be to publish the PubMed link – this could be the only solution if we want to let the public to consult it. This also addresses the issue of copyright.

Monographs and/or herbal constituents, fingerprinting and other botanical and chemical data, as well as toxicology data should be linked for reference. As previously suggested, the existing KCL/SIMM Chinese Herbal Constituents database (CHCD) and Bio-active Plant Compounds database (BPCD), in combination with data from WP1, 2 and 3 could be considered for use as a basis for this endeavour.

These initial concepts were further developed during the July 2010 meeting in London with respect to the strategy to be followed by members in order to choose the research area to concentrate the analytical work. In consideration of the importance of the issue for the entire GP-TCM, it was decided to extend the discussion with members of all the other work packages during the 1st GP-TCM Annual General Meeting, held in Henley following the WP4 follow-up meeting in London. At this stage of the project, in fact, one of the most urgent issues was the rationale to be used in the choice of the papers for evaluation, as well as to build up the online database. The importance of this question goes well beyond WP4, and has implications for the project as a whole, given that the choice of research areas can cast its influence over each work package, if a common thread is aimed at.

During the general meeting at Henley, the discussion involved members from most WPs and finally the strategy shifted from being disease-oriented, to plant-oriented. It was thus agreed to concentrate on plants used in selected TCM formulas. As a suggestion to start with, the TCM formula “Liu Wei Di Huang Wan” was indicated for its meaningful experimental work, given also that the 6 species used in the formula were also among those already classified in a plant list set up by WP1. Together with other 6 species suggested from another TCM formula, it was then agreed that these plants, all of which are frequently used in TCM, would be the initial focus of WP4 work, as well as the focus for other work packages.

Subsequently the following list of herbs (Table 1) was provided by WP1 lead Prof. Simmonds to all WPs for their guidance:

The list shows comprehensive naming information on the constituents on the formulae under a few categories:
- Latin binomial name
- Latin binomial name used in Chinese Pharmacopoeia
- Accepted Latin binomial name
- Latin binomial synonyms
- Comments
- Plant part used
- Pharmaceutical name from China Pharmacopoeia 2005
- Two transliteration versions using pin yin.

TABLE 1

Chinese Pharmacopoeia species (CP05 English)	Accepted Latin scientific name (Flora of China) or other comment	Part used in TCM	Pharmaceutical name (from Chinese Pharmacopoeia 2005 (NOTE 2)	Chinese name (pin yin)	Latin scientific synonyms 1: from Flora of China (NOTE 1)
Alisma orientalis (Sam.) Juzep.	*Alisma orientale* (Sam.) Juz. (note slightly different ending)	tuber (called a 'rhizome' in Chinese. Pharm.)	**Rhizoma Alismatis**	泽泻 'Ze Xie' or 'Zexie'	1: Alisma plantago-aquatica L. var. orientale Sam.; Alisma plantago-aquatica subsp. orientale (Sam.) Sam.
Cornus officinalis Sieb. et Zucc.	*Cornus officinalis* Sieb. & Zucc.	ripe fruit	**Fructus Corni**	山茱萸 'Shan Zhu Yu' or 'Shanzhuyu'	1: Macrocarpium officinale (Sieb. & Zucc.) Nak.; 2: Cornus officinalis var. koreana Kitam.,
Dioscorea opposita Thunb.	*Dioscorea polystachya* Turcz. (NOTE 3)	tuber (called a 'rhizome' in Chinese. Pharm.)	**Rhizoma Dioscoreae**	山药 'Shan Yao' or 'Shanyao'	1. Dioscorea batatas Dec. ; D. decaisneana Carrière; D. doryphora Hance; D. potaninii Prain & Burk.; D. rosthornii Diels; D. swinhoei Rolfe 2. D. batatas forma clavata Makino; D. batatas forma daikok Makino; D. batatas forma flabellata Makino; D. batatas forma rakuda Makino; D. batatas forma tsukune Makino; D. cayennensis var. pseudobatatas Hauman; D. pseudobatatas (Hauman) Herter
Paeonia suffruticosa Andr.	*Paeonia ostii* T. Hong & J. X. Zhang (NOTE 4)	rootbark	**Cortex Moutan**	牡丹皮 'Mu Dan Pi' or 'Mudanpi'	1. Paeonia ostii var. lishizhenii B. A. Shen; P. suffruticosa Andrews subsp. ostii (T. Hong & J. X. Zhang) Halda.
Poria cocos (Schw.) Wolf.	*Wolfiporia extensa* (Peck) Ginns	sclerotium (mass of mycelium)	**Poria**	茯苓 'Fu Ling' or 'Fuling'	Poria cocos (Schw.) Wolff; Wolfiporia cocoa (F.A. Wolf) Ryvarden & Gilb
Rehmannia glutinosa Libosch	*Rehmannia glutinosa* (Gaertner) Liboschitz ex Fischer & CA Meyer;	processed rhizome (called a 'root tuber' in Chinese. Pham.); processing involves stewing or steaming in yellow rice wine.	**Radix Rehmanniae Praeparata**	熟地黄 'Shu Di Huang' or 'Shudihuang'; (NOTE 5)	1. Digitalis glutinosa Gaertner; Rehmannia chinensis Liboschitz ex Fischer & C. A. Meyer; R. glutinosa var. hemsleyana Diels; R. glutinosa var. huechingensis Chao & Shih; R. glutinosa forma huechingensis (Chao & Shih) P. G. Hsiao; R. glutinosa forma purpurea Matsuda
2 species: *Ganoderma ludicum* (Leyss. Ex Fr.) Karst. and *Ganoderma sinensis* Zhao, Xu et Zhang	*Ganoderma lucidum* (Curtis) P. Karst but this is not a Chinese species, although the name given to many samples	sporophore (spore-bearing part)	**Ganoderma**	灵芝 (Ling Zhi)	
Angelica pubescens Maxim. *f . biserrata* Shan et Yuan		root	**Pubescent Angelica Root**	独活'Du Huo'	
Astragalus membranaceus Bge.		root	**Astragali Radix**	黄芪 'Huang Qi'	

Atractylodes macrocephala Koide			**Atractylodis Ovatae Rhizoma**	白术 'Bai Zhu'	
Scutellaria baicalensis George		root	**Scutellariae Radix**	黄芩 'Huang Qin'	
Tripterygium wilfordii Hook.*f.*		root	**Radix Tripterygii Wilfordii**	Lei Gong Teng'	

NOTE 1 - (http://flora.huh.harvard.edu/china); 2: additional synonyms from World Checklist of Selected Plant Families (http://apps.kew.org/wcsp)
NOTE 2 - (N.B. these names are reversed in the CP2010 e.g. **'Rhizoma Alismatis'** becomes **'Alismatis Rhizoma'**)
NOTE 3 - (the CP05 name *'D. opposita* Thunb.' is a nomenclatural synonym which has been misapplied; the name *D. opposita* is therefore superfluous and illegitimate and scientifically should not be used (following advice from M. Gilbert (Kew/Missouri) - email to C. Leon 25/7/09 and P. Wilkin (Kew) - (pers.comm 9/09).
NOTE 4 - Following a major Chinese taxonomic revision (De-yuan Hong et al., 1999. A revision of the *Paeonia suffruticosa* complex (Paeoniaceae) Nordic Journal of Botany, vol. 19 (3): 289 – 300) the name *'P. suffruticosa* Andr.' is no longer considered botanically to be the source of the TCM herb **'Cortex Moutan'**. The Flora of China follows this view. The correct scientific name for the source of **Cortex Moutan** is *Paeonia ostii* T. Hong & J. X. Zhang (the name *P. suffruticosa* has been assigned to another species). But, for purposes of literature searching however it is necessary to use the *P. suffruticosa* since this is the name coined in TCM.
NOTE 5 - Do not confuse with the separate drug **'Radix Rehmanniae'** otherwise known as 'Di Huang', 'Dihuang', 'Xiandihuang' or 'Shengdihuang' which is not processed.

This list of species has taken over a year to compile. At the start of the project it was agreed that we would target popular herbs used in TCM, but soon realised we needed to define what we meant by "popular". Members of the WP1 obtained lists from companies trading species used in TCM in main land China, Hong Kong and Europe as well as lists from colleagues in China and Hong Kong. Once the data were collated, we had a list of over 100 species commonly used in traditional Chinese medicines in China and Europe. However, this list was too long. The use of the 100+ species in the treatment of conditions including cancer, diabetes, irritable bowel and skin diseases was checked and the list of plants was reduced from over 100 to 53. At Henley, it was suggested we look at some popular TCM formula and it was agreed to select the six species used in the CHM formula "Liu Wei Di Huang Wan", plus six more species used in many other TCM formula.

Although this list provides a starting point for the initial analysis, later the work could be usefully expanded by the inclusion of likely (and rare) adulterants, and closely related species, etc.

In September 2010, the WP4 Coordinator assigned each herb to a WP4 member to begin the search and selection of the articles to be used for the on line repository (see also D4.6).

In conclusion, the initial approach changed and Diseases and molecular targets will not guide the choice of the research fields, but will still be highlighted for each paper in the online repository so that user will be able to search the data-list with those criteria anyway.

2 D4.5: BIOLOGICAL TARGET ORIENTED DATABASE FOR *IN VITRO* RESEARCH ON CHM

2.1 Suggestions for repository I

In the first page of the online repository, a brief explanation of the project and its scope should be given. The growing nature of the list of papers should be emphasised and an explanation of inclusion/exclusion criteria and scoring methodology provided.

A list of plants from CHM have been provided and each plant has been assigned to each WP4 member (see D4.4). It is foreseeable that subgroups for choosing and evaluating scientific papers will be formed (see D4.6). The Coordination Office (CO) will assist WP4 members to review the work and help them to add the papers to the repository. After the end of the project, the repository will operate, as the rest of the website, under the guidance of the European Society of Chinese Medicine Research, to be constituted.

A minimum of one hundred papers should be listed, including clinical and *in-vivo* studies, in each disease heading.

For various reasons (see below and D4.6) the repository will host papers with and without scoring. Each paper should thus be flagged as “quality scored” or “not evaluated”.

To help users in their navigation into the data list contents, a suggestion on how to establish the logical structure of the repository could be:

1st level: Disease (Diabetes, Cancer, etc)/Plant

2nd level: claimed molecular target 1, 2, 3,.... n, unknown

3rd level: type of research (only when other WPs will have provided articles from their areas): 1 - clinical studies (including case report) (papers provided by WP6); 2 *in-vivo* studies (papers provided by WP5), *ex-vivo* studies (WP4-WP5), *in-vitro* studies (WP4), *in-silico* studies (WP4), epidemiological studies on traditional use (WP6), reports from traditional use (none of these can be evaluated for quality, but might be useful for the researcher in the field).
The use of a functional genomics approach should be flagged on each paper (later it could be considered to add a specific subheading)

4th level: TCM formulas; TCM single plants (this could be limited to the list of about one hundred plants currently most traded, suggested from work packages 1 and 2); TCM single molecules (these can't be quality scored)

5th level: each plant or formula should be linked to its references for composition, fingerprinting, extraction, etc. This part of the repository should be provided by WP1 and WP2), as well as for toxicology data (from WP3).

When there is uncertainty on which category of each level should the paper fall into, then the paper should be listed in both categories.

Alternatively to having different levels, specific flags could be added to each paper or a list of characteristics tabbed.
Each paper will be indexed using simple key words. These should list: TCM formula names, TCM plant names, Latin binomial and common names, single molecules, disease name, syndrome, molecular target, authors, institutions, etc.

A search engine should be provided to search such key words.

Monographs and/or herbal constituents, fingerprinting and other botanical and chemical data, as well as toxicology data should be linked for reference. The existing KCL/SIMM Chinese Herbal Constituents database (CHCD) and Bio-active Plant Compounds database (BPCD), in combination with data from WP1, WP2 and WP3 could be used. These databases which were mentioned in the WP4 kick-off meeting in 2009 have been developed in King's College London and are currently the subject of negotiation with a US company who are looking to license them and make them available on the web. An important feature of the negotiation will be academic access for WP members – an update on these discussions will be made available as soon as possible.

In the web pages, there should also hyperlinks such as:

Phytochemical Informatics of Traditional Chinese Medicine & Therapeutic Relevance
http://dx.doi.org/10.1021/ci700155t

Virtual Screening of Chinese Herbs with Random Forest
http://dx.doi.org/10.1021/ci600289v

Phytochemical Databases of Chinese Herbal Constituents & Bioactive Plant Compounds with Known Target Specificities
http://dx.doi.org/10.1021/ci600288m

TCMGeneDIT (Taiwan)
http://tcm.lifescience.ntu.edu.tw/

TCM-ID (Singapore)
http://tcm.cz3.nus.edu.sg/group/tcm-id/tcmid.asp

3 D4.6: REPORT ON QUALITY CRITERIA AND SCORING OF THE CHM DATABASE

3.1 Background and general principles

Background: During the WP4 kick-off meeting, the WP4 discussion group agreed and developed the criteria. The implementation of these criteria resulted in unsatisfactory validation results, so they were changed during the follow-up meeting. A final modification of the scoring procedure was then optimised following further testing.

The passages which led to the final scoring procedure are presented in their temporal sequence up to the final version developed in October 2010 following the trial and validation of the scoring procedure agreed and dedicated to the issue during the WP4 follow-up meeting:

Following the team work during the WP4 kick-off meeting (see D4.1 and D4.2) and further discussions among work package members, it was decided that all the sections of a paper should be evaluated, though with a different level of importance in the global scoring. Each evaluation criterion was established with a collective work where all the WP members participated to reach a final agreement on the criteria, their relative weight and guidelines for their correct application. Criteria were then validated by their experimental application to the same papers by all the WP members and results compared.

General principles were maintained throughout the different versions, while the scoring procedure was progressively optimised.

It is also noticeable that the criteria, for the complexity of the field to be evaluated, will require continuing discussion and optimization, which will develop during the entire course of the project and will especially benefit from the contributions of other WPs and user's feedback. The scores on the database will be accordingly updated.

3.2 Some inclusion/exclusion issues

A method for the initial selection of papers should be established which should have inclusion/exclusion criteria, before a paper is evaluated, though a wider choice should be used for selecting papers to be included in the database, which should include as many papers as possible for each field so as to become a general reference for researchers in the field. There is no suggestion to exclude at this stage those articles just to be poor (by having a low score) – the **main aim of the current work is not to judge scientific contents, but to identify the requirements needed for a paper to be considered of good quality standard**. The lack of some aspects just might make it impossible to compare a paper to a given standard.

First of all, there will be a specific rationale to choose the papers to be evaluated, and to this scope it was agreed with all the other WPs in July 2010 that the choice should be guided by a list of selected plants from CHM formulas (see D4.4), so that at the end of the project we will have more chances of nearing a “state of the art” repository for the specific subjects. Also, with the contribution of other WPs, the repository could be used as a valuable reference by research scientists ranging from *in-vitro* to clinical studies.

The formulas or plants should have known clinical activity shown in clinical studies, or even just have evidence from traditional practice (but then the papers should be listed in a separate subheading of the repository, see D4.5), otherwise it will not be possible to establish a parallel between clinical, *in-vitro* and *in-vivo* WPs.

Papers on herbal preparations that have been studied only *in vitro* might be included just to have a comprehensive data list of the papers on the subject.

3.3 Some scoring issues

- When a paper deals with purified compounds scores are high for all the criteria related to the identification of the material used. Nevertheless the findings cannot be used to explain the biological activity of the original phytocomplex, which is one of the issues for TCM research. Moreover such papers simply follow the general criteria for scientific papers and should then be excluded from evaluation, even though they can be included in the appropriate section of the database (single bioactive principles, see D4.5).
- Papers focusing on the extraction and identification of single bioactive components, rather than dealing primarily on biological activity might be included in the database but, like those dealing with single molecules, cannot be scored.
- When *in-vitro* papers include meaningful experiments carried out *in-vivo*, they should be scored higher. Since the focus of WP4 is *in-vitro* work, this suggestion is a guide for reviewers. It was agreed during the WP4 follow-up meeting in July that some bonus points should be added to the score for articles which included meaningful *in-vitro* work.
- Papers dealing with formulas should be considered like those using single herbs, given that the identification of the components or fingerprint data are present.
- If components are not described then the paper should refer to previous literature where the components have been described. Evidence should be provided that the phytocomplex is identical in all respects.
- In addition, when a preparation used for *in-vitro* studies is identical with that used for therapy and/or in *in-vivo* studies, a higher score should be given. As in the third bullet point above, the precise operation of this suggestion was agreed during the follow-up meeting in July 2010.

3.4 Quality evaluation procedures

The discussion among WP members has allowed a consensus to be reached on several basic principles that need to be applied when choosing the papers for quality evaluation, some of which have been highlighted in D4.6 in the paragraph dedicated to inclusion/exclusion criteria. In general all papers dealing with *in-vitro* pharmacology of phytocomplexes related to TCM will be considered, regardless of the technique used. In the beginning it was decided that the focus of this work will be on those herbs suggested from WP2 on the basis of their actual use, importance and amount of scientific work. Later a list of plants was provided as a guide, but the research areas will be extended in the future following the same structure and criteria to include other plants.
After a paper is chosen, a group of at least three experts (triplet) should proceed to evaluate it according to the following scheme:

<u>First version of the scoring procedure</u>

Paper Sections considered:

1. Title and Abstract
2. Introduction and background
3. Methods (plant description)
4. Methods (experimental)
5. Objectives
6. Results
7. Conclusion and Discussion

Each section of a paper is to be scored according to the following scale:

5 very good
4 good
3 satisfactory
2 inadequate
1 wholly inadequate
0 not applicable.

Where the most important criteria is
number 4,
the next number 3,
the next number 6,
the next number 7,
the next number 2
the next number 5, and finally
the number 1.

The weighting is achieved by multiplying the score for criterion 4 by 7 (the number of criteria), that for the next most important (criterion 3) by 6 (the number of criteria less 1), etc. Thus, an article deemed to be very good (and so scoring 5) in all categories will have a cumulative score of 140. An article scoring only 1 for each of the 7 criteria would obtain a score of only 28.

First set of criteria

Number	***Section in article to be reviewed***	***Description of criterion***	***Relative importance of criterion***
1	Title and abstract	Does either the or abstract or both provide the herbal medicinal product's • Latin binomial name • the part of the plant used and • the type of preparation tested?	**7**
2	Introduction and background	• Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented? • Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	**5**
3	Methods – plant description	Do the methods include or indicate the • herbal product name • the Latin binomial name (including the botanical authority) and • family name for each herbal ingredient • the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number	**2**

		• the part of the plant used to make the product or the extract • the processing used to make the test material, • type and concentration of the extraction solvent used and the ratio of the herbal drug to extract • the method of authentication of the herbal raw material, including details of any voucher specimen • whether the test material had been subjected to fingerprinting and by what methods and by whom and • whether any special testing/purity testing had been carried out and by whom • whether the material had been standardised, and by what process and by whom?	
4	Methods – experimental	Does the description provide a • detailed unequivocal description of the test system • detailed unequivocal description of the experimental protocol • description of and justification for the statistical methodology used?	**1**
5	Objectives	Are the specific objectives and hypothesis and rationale for the selection of the test models used included?	**6**
6	Results	Do the results include the • sample size; clearly defined primary and secondary outcome measures • statistics • data showing the selectivity of the biological effect and its specificity	**3**
7	Conclusion and Discussion	• appropriate strategy and robustness of the conclusions • interpretation of the result in light of the product tested and dosage regimen used? • Has the conclusion been linked to clinical use? • Do the conclusions *support* or *contradict* existing findings?	**4**

It is important to note that some papers may not reach an overall score considered worthy of detailed consideration and inclusion but they may contain some interesting aspects such as a relevant *in-vitro* model or methodology. They will be taken into account for the database just for this detail.
For each paper there will be three or more total scores from which the mean value will be evaluated and the range considered.

Using the same quality levels used for evaluating each section, ranges of mean total scores were initially proposed as follows:

1. Wholly inadequate: papers scoring from 28 to 60
2. inadequate: papers scoring from 61 to 80
3. sufficient: papers scoring from 81 to 100
4. good: papers scoring from 101 to 120

5. very good: papers scoring from 121 to 140

When the mean of the total scores and the values of the single total scores fall into a certain range, then the evaluation of the paper is final. If the single totals fall in a different range with respect to the mean of the total score, then the reviewers need to further discuss until a consensus is reached. In this case a larger number of reviewers might be needed. The details of this process were later refined at the follow-up meeting and during the subsequent validation of the method. In the end this last classification was totally eliminated, it was considered to be too judgmental and to not give any extra value to the scoring procedure.

3.5 Trial of the evaluation procedure and quality criteria

The proposed procedure was tested for its suitability by distributing 11 papers kindly provided by Professor Simmonds, to the WP members for their evaluation. Results were correlated by the Coordinator and a final qualification was proposed for each paper. Given the experimental nature of the work, the following results were further discussed during the follow-up meeting in July to optimize the procedure and make it suitable for the final guidelines.

A discussion of these results formed an important agenda item for the meeting which resulted in the adjustment of the guidelines for their use or more radical changes. All participants felt that this trial was a very important aspect of the group's initial work because it would lead to a protocol to be applied more widely and comprehensively throughout the project and contribute to the deliverables of the other work packages as well.

Titles of the papers selected for the first pilot study are given in Table 1a.

Table 1a papers selected for the first evaluation

Number	***Title***
1	Shang P, Qian AR, Yang TH, Jia M, Mei QB, Cho CH, Zhao WM, Chen ZN Experimental study of anti-tumor effects of polysaccharides from *Angelica sinensis*. *World J Gastroenterol* 2003, **9**(9): 1963-1967
2	Tze-chen Hsieh, Xiaohua Lu, Jennifer Chea and Joseph M. Wu Prevention and Management of Prostate Cancer Using PC-SPES: A Scientific Perspective. *J. Nutr.* 2002, **132**: 3513S–3517S
3	Dong-Chan Kim, Se-Young Choi, Sun-Hee Kim, Bong-Sik Yun, Ick-Dong Yoo Nanga. Ravi Prakash Reddy, Ho Sup Yoon, and Kyong-Tai Kim Isoliquiritigenin Selectively Inhibits H2 Histamine Receptor Signaling. *Mol Pharmacol* 2006, **70**:493–500
4	Minsook Ryua, Eun Hye Kimc, Mison Chund, Seunghee Kangd, Bumsang Shime, Young-Beob Yu, Gajin Jeong Jong-Soo Lee Astragali Radix elicits anti-inflammation *via* activation of MKP-1, concomitant with attenuation of p38 and Erk. *Journal of Ethnopharmacology* 2008, **115**:184–193
5	Tetsuyuki Takahashi, Nobuo Takasuka, Masaaki Iigo, Masaki Baba, Hoyoku Nishino, Hiroyuki Tsuda and Toru Okuyama Isoliquiritigenin, a flavonoid from licorice, reduces prostaglandin E2 and nitric oxide, causes apoptosis, and suppresses aberrant crypt foci development. Cancer Sci 2004, **95**(5): 448-53
6	Jing-Tian XIE, Chong-Zhi WANG, Bin ZHANG, Sangeeta Ram MEHENDALE, Xiao-Li LI, Shi SUN, Aung Htun HAN, Wei DU, Tong-Chuan HE, and Chun-Su YUAN *In Vitro* and *in vivo* Anticancer Effects of American Ginseng Berry: Exploring Representative Compounds *Biol. Pharm. Bull.* 2009, **32**(9):1552—1558
7	Winnie Lai Ting Kan, Chi Hin Cho, John A. Rudd, Ge Lin Study of the anti-proliferative effects and synergy of phthalides from *Angelica sinensis* on colon cancer cells *Journal of Ethnopharmacology* 2008, **120**:36–43
8	Nu-Man Tsai, Yi-Lin Chen, Chau-Chin Lee, Po-Chen Lin, Yeung-Leung Cheng, Wen-Liang Chang, Shinn-Zong Lin and Horng-Jyh Harn

	The natural compound n-butylidenephthalide derived from *Angelica sinensis* inhibits malignant brain tumor growth *in vitro* and *in vivo Journal of Neurochemistry* 2006, **99**:1251–1262
9	Ying-Kun Qiu, De-Qiang Dou, Li-Ping Cai, Hai-Ping Jiang, Ting-Guo Kang, Bing-You Yang, Hai-Xue Kuang, Michael ZC Li Dammarane-type saponins from *Panax quinquefolium* and their inhibition activity on human breast cancer MCF-7 cells *Fitoterapia* 2009, **80**: 219–222
10	Kazuto WASHIDA, Yoshiyuki ITOH, Takashi IWASHITA, and Kyosuke NOMOTO Androgen Modulators from the Roots of *Paeonia lactiflora* (Paeoniae Radix) Grown and Processed in Nara Prefecture, Japan *Chem. Pharm. Bull.* 2009, **57**(9): 971—974
11	M Noda, RL Vogel, AM Craig, J Prahl, HF DeLuca, DT Denhardt. Enhancement of 1,25-dihydroxyvitamin D3 –and All *trans* retinoic acid-induced HL-60 leukemia cell differentiation by *Panax ginseng*. *PNAS* December 1, 1990 vol. 87 no. 24 9995-9999

The papers were thus evaluated by the WP members as shown in Table 2 and Table 3.

Table 2 and Table 3 show the scores of the article given by the individual participants and their relative individual ranking, respectively. This is shown most clearly in a graphical representation in Figure 1. Detailed comments from some WP4 members using the quality criteria and scoring are attached as Appendix II.

The results show that there was a wide variability in the scores given by each member and, most of all, in no case was it possible to assign a final quality score according to the criteria and the procedures, in fact the single total values of each paper fall outside the range where the mean value falls, according to the suggestion list in section 3.4 above. According to the procedure, then the reviewers should discuss the article until a consensus is reached.

It is thus proposed that the processing criteria be refined following extensive discussion at the follow-up meeting. This should lead to the reduction in variation by trying to make the criteria more objective rather than subjective. This discussion can best be achieved face to face – this was in fact one of the drivers for the follow-up meeting.

Table 2 Total Scores of each article by individual reviewers

	reviewer								
article	***A***	***B***	***C***	***D***	***E***	***F***	***G***	*mean*	*SD*
1	90	73	61	54	99	68	92	**77**	17
2	73	90	52	68	106	74	73	**77**	17
3	113	105		140	91	94	98	**107**	18
4	80	94	76	82	108	98	138	**97**	21
5	68	56		129	87	87	74	**84**	25
6	102	111	93	120	116	127	93	**109**	13
7	134	79	67	99	112	100	94	**98**	24
8	109	113		140	130	86	93	**112**	21
9	123	78	58	67	80	60	46	**73**	25
10	96	107	71	71	93	73	67	**83**	16
11			63	79	103	84		**82**	16

Table 3 – Ranking of each article by individual reviewers

article number	*reviewer*						
	A	***B***	***C***	***D***	***E***	***F***	***G***
1	7	9	6	11	1	10	6
2	9	6	8	9	2	8	8
3	3	4		1	9	4	2
4	8	5	2	6	6	3	1
5	10	10	11	3	10	5	7
6	5	2	1	4	4	1	4
7	1	7	4	5	5	2	3
8	4	1		2	3	6	5
9	2	8	7	10	11	11	10
10	6	3	3	8	8	9	9
11				7	7	7	

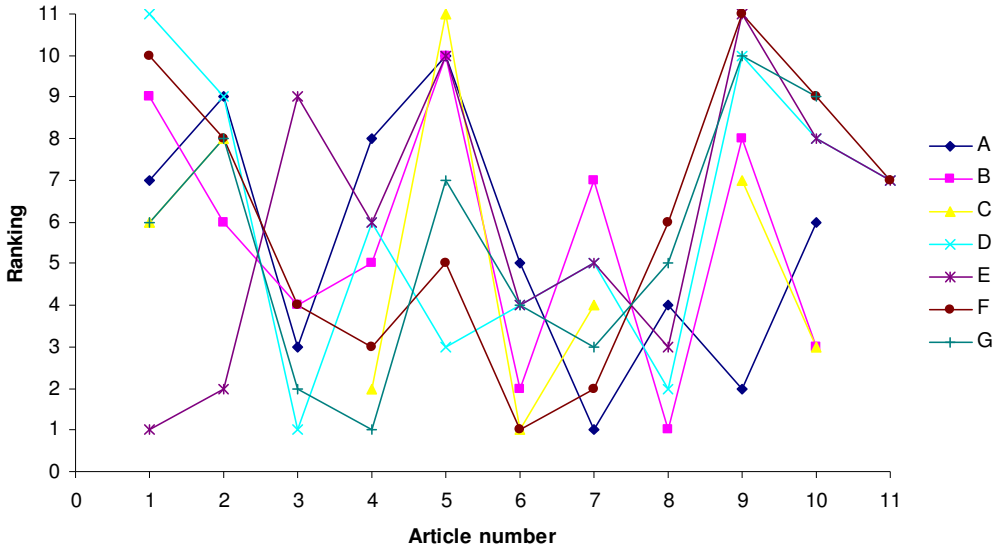

Figure 1 A chart showing the variation in article evaluation; A-G show individual reviewers; rankings range from 1-11, where 1 represents the article judged to be of highest quality

3.6 The results of the original criteria

WP4 Kick-off meeting was held at the Waterloo Campus of the King's College London, on 16th – 18th October 2009. This meeting gathered work package members to discuss tasks and deliverables of the WP. The meeting resulted in proposed criteria for literature evaluation, which were trialed before the follow-up meeting. The agenda for the latter meeting included an evaluation and revision of the critical assessment of the selected articles with the goal of adopting **definitive criteria** to be used throughout the next phases of the project, as well as addressing the month 18 deliverables.

As the original criteria showed above, WP4 members used these criteria to evaluate 11 selected new papers, which were kindly provided by Prof. Monique Simmonds. Tables 2 and 3 show the total scores and ranks of different reviewers, indicating large variations.

Table 4 - Total Scores of each article by individual reviewers

	reviewer								
article	***A***	***B***	***C***	***D***	***E***	***F***	***G***	*mean*	*SD*
1	90	73	**61**	**54**	99	**68**	92	**77**	17
2	73	90	**52**	68	106	**74**	**73**	**77**	17
3	113	105		**140**	91	94	98	**107**	18
4	80	94	**76**	82	**108**	98	**138**	**97**	21
5	68	**56**		129	87	87	74	**84**	25
6	102	**111**	93	**120**	**116**	**127**	93	**109**	13
7	134	**79**	**67**	99	112	100	94	**98**	24
8	109	113		**140**	**130**	86	93	**112**	21
9	**123**	78	**58**	**67**	80	**60**	**46**	**73**	25
10	96	107	**71**	**71**	93	**73**	**67**	**83**	16
11			**63**	**79**	103	84		**82**	16

Notes: Blue colour highlights lower scores. Red colour highlights higher scores. A large variation among the reviewer's scores is highlighted.

Table 5 – Ranking of each article by individual reviewers

	reviewer								
article number	***A***	***B***	***C***	***D***	***E***	***F***	***G***	*article number*	***A***
1	7	9	6	11	**1**	10	6	*1*	7
2	9	6	8	**9**	2	8	8	*2*	9
3	3	4		**1**	9	4	2	*3*	3
4	8	5	2	6	6	3	**1**	*4*	8
5	10	10	11	3	10	5	7	*5*	10
6	5	2	**1**	4	4	**1**	4	*6*	5
7	**1**	7	4	5	5	2	3	*7*	1
8	4	**1**		2	3	6	5	*8*	4
9	2	8	7	10	11	11	10	*9*	2
10	6	3	3	8	8	9	9	*10*	6
11				7	7	7		*11*	

Notes: Rank 1 is the best article. All the individual reviewers selections are different.

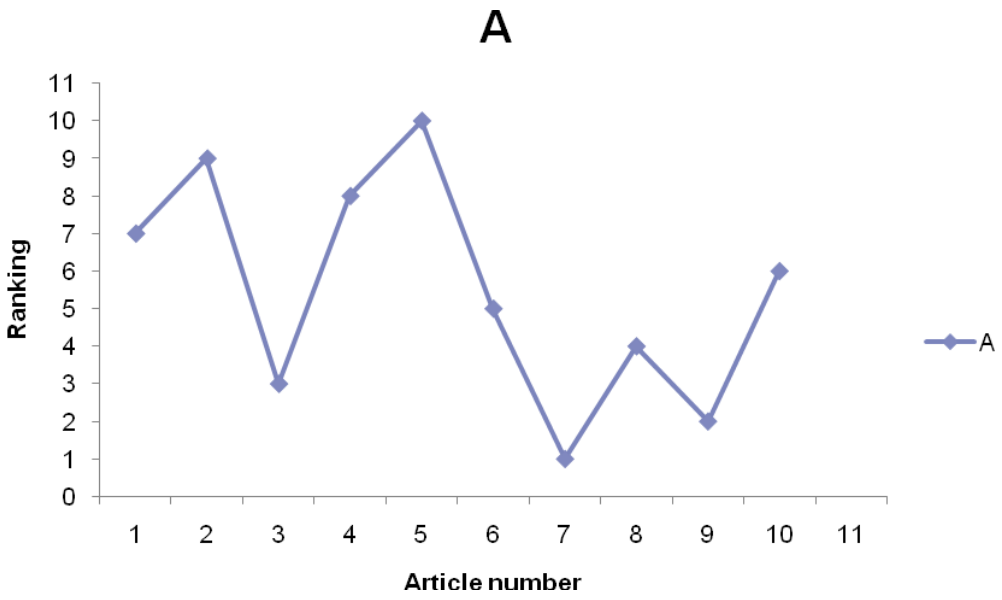

Figure 2 Ranking of Reviewer A

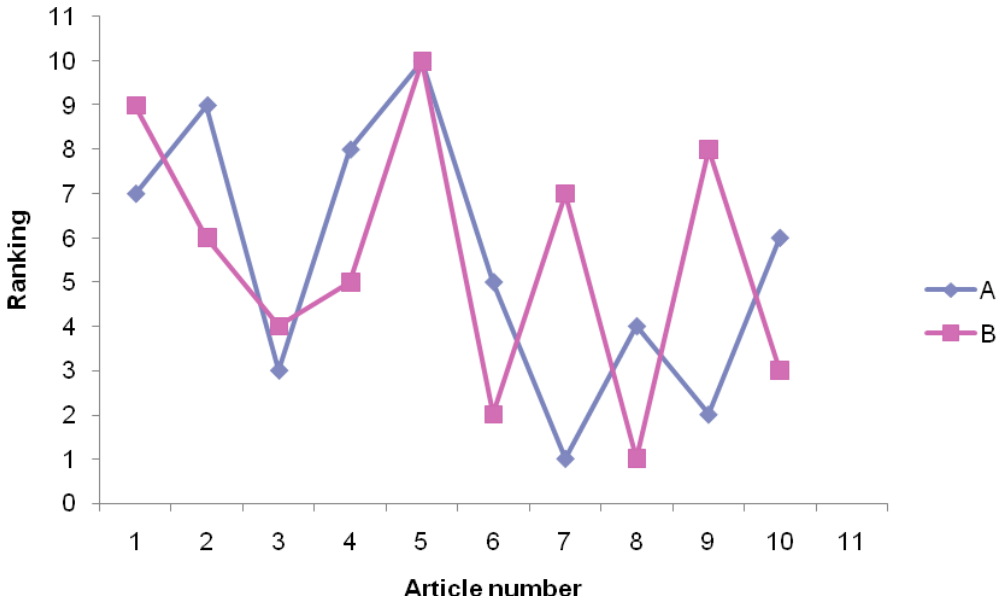

Figure 3 Ranking of Reviewers A and B

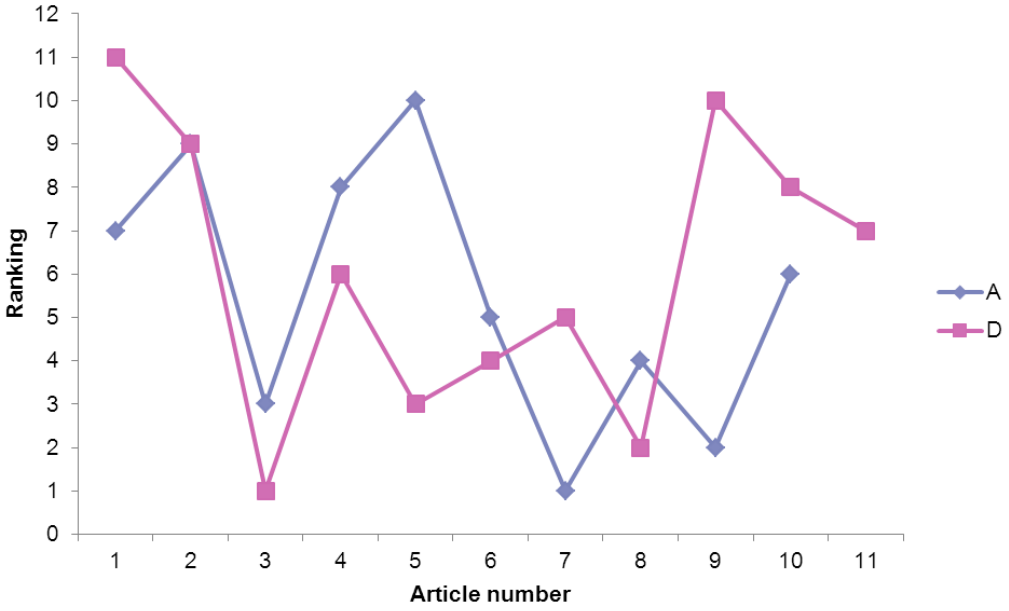

Figure 4 Ranking of Reviewers A and D

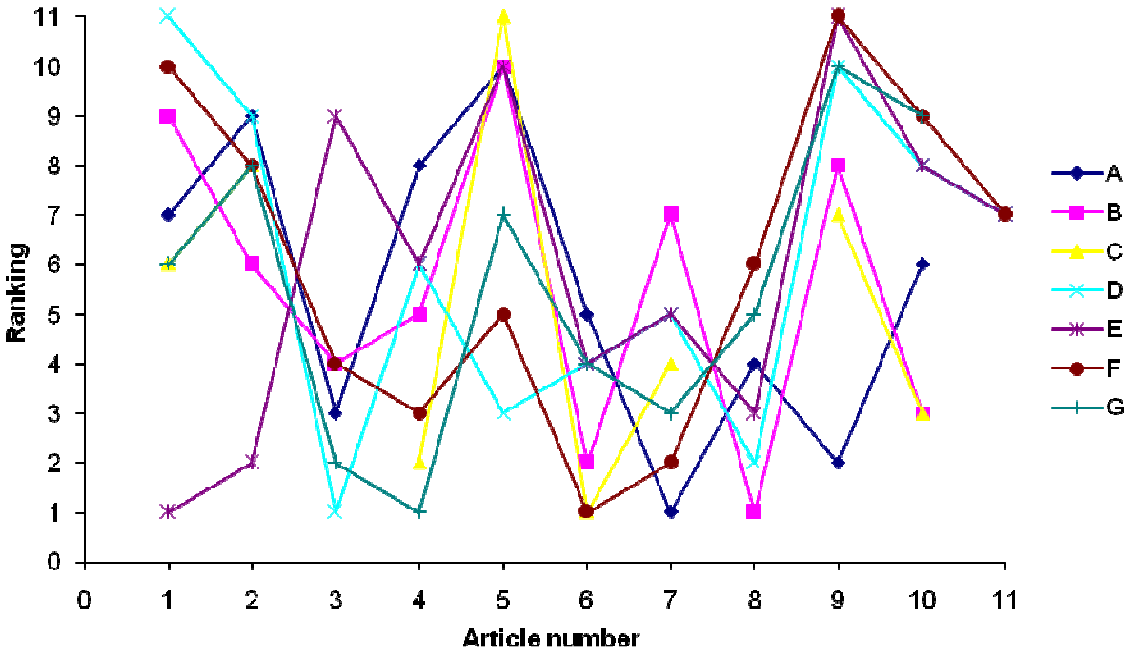

Figure 5 A chart showing the variation in article evaluation; A-G show individual reviewers; rankings range from 1-11, where 1 represents the article judged to be of highest quality

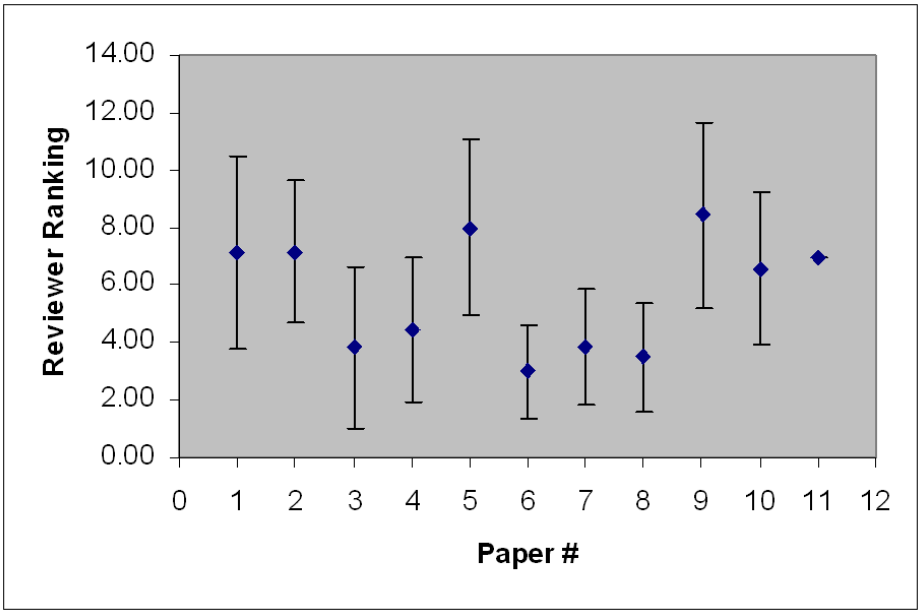

Figure 6 The standard deviation of rank for all the selected papers

3.7 The modified criteria generated in the WP4 follow-up meeting

In the follow-up meeting, WP4 members revised the original criteria. Based on the results of the trial evaluation, especially on the comments from Dr. Qihe Xu, WP4 members generated modified criteria. The modified criteria are showed in table 6.

Table 6 the modified criteria in the WP4 follow-up meeting

section	Modified criteria in the follow-up meeting
Title and abstract	
	Name of plant or product used and use
	Do these accurately describe contents of paper?
Introduction and background	Does rationale relate to traditional medicinal use
	Does it include adequate background
	Does it clearly state if study is new, or
	extend existing knowledge
Methods - plant description	Include or indicate the
	Herbal product name
	Solvent, ratio, time and temperature of of extraction,
	Method of authentication,
	Voucher specimen details
	Proprietary name or extract name with manufacturer and batch no
	Plant part
	Processing or other treatment
	Type, concentration and solvent
	Any fingerprinting including technique and by whom
	Standardisation, by what process and by whom?

The modified criteria were then tested. We used these modified criteria to evaluate one article. The WP4 members were randomly divided into three groups, each group having three members.

The Test evaluation was performed on the paper:

" *Astragali radix elicits anti-inflammation via activation of MKP-1, concomitant with attenuation of p38 and Erk.* Journal of Ethnopharmacology 2008, 115:184-193 – M.Ryu, E.H. Kim, M. Chun, S. Kang, B. Shim, Y-B Yu, G. Jeong, J-S Lee."

Using the modified criteria, three groups gave the different marks. The marks are presented below:

52/105; 47/105 and 62/105.

These assessments were closer to each other than before (using the first criteria) and so then the modified criteria were refined again. The application of each single criteria was then discussed, as well as the scoring system as a whole, reviewing both the methodology and the inclusion criteria.

New criteria

Briefly, in the new evaluation scheme, the scores are given according to the total number of the criteria present in the section, multiplied by the "relevance factor" attributed to each section. In section 3b it is possible to give more scores as bonuses, in virtue of the presence of methodological strategies which can benefit both a systems biology approach (*i.e.,* assays using multisystem approaches are valued more than single biological paradigms), and that can be more directly related to clinical application (*i.e.,* the use of human samples). These experimental approaches are considered to be more representative in CHM research. The value from extra bonuses is added to the final score as it is, without multiplying it by the relevance factor.

Sections were reduced to 5, since the "objectives section" is not always present in all journals and they can be found elsewhere in the articles. In some cases requirements are fulfilled even when they are reported in a section different from the one indicated in the criteria.

The final score is given by the sum of the sections' scores plus the extra bonuses.

Table 7 - New criteria

No	***Section in article to be reviewed***	***Relative importance of criterion***	***Relevance factor***	***Description of criteria***
1	Title and abstract	**5**	1	4 criteria: Does either the title or abstract or both provide the herbal medicinal product's 1 - Latin binomial name of the plant (or clear name of the product). This point will be scored in this section even when the Latin binomial name might be present elsewhere in the paper? 2 - part of the plant used? 3 – experimental use of the plant/product? 4 – are the article's contents accurately described?
2	Introduction and background	**4**	2	4 criteria: 1 – is the scientific background presented? 2 – is an explanation of the rationale presented, including a brief statement of the reasons for the study with reference to the specific herbal medicinal product and its traditional use? 3 - is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated? (novelty) 4 – are the objectives of the study clear? (the objectives will be scored in this section even when, depending on the journal, might be found elsewhere in the paper)
3a	Methods – (plant description)	2	4	Number of criteria are variable depending on the type of material used: GENERAL (apply to all) 8 criteria: do the methods describe: 1 - Solvent used and ratio for the extraction 2 – time of extraction 3 - temperature of extraction 4 - yield of extraction 5 – is the method of authentication of the herbal raw material indicated? 6 – are details of any voucher specimen included? 7 – has the test material been subjected to simple chemical constituent profiling and/or complex fingerprinting (by what methods and by whom) 8 – has the material been standardised (by what process and by whom) IN CASE OF UNPROCESSED PLANT AND/OR MIXTURES OF PLANTS 2 criteria:

				1 – is the herbal product name clearly indicated? 2 – is the part of the plant used to make the product specified? IN CASE OF A PROCESSED PLANT 5 criteria: 1 - are the processed product name or the extract name and the name of the manufacturer of the product indicated? 2 - is the batch number of the herbal product name indicated? 3 – is the part of the plant used to make the product or the extract specified? Where applicable: 4 – is the type of preparation to make the test material described? 5 – is the yield of the extraction to make the test material indicated? IN THE CASE OF A PROPRIETARY PRODUCT 5 criteria: 1 - are the proprietary product name or the extract name and the name of the manufacturer indicated? 2 – is the batch number of the product indicated? 3 – is the part of the plant used to make the product or the extract indicated? Where applicable: 4 – is the type preparation to make the test material described? 5 – is the yield of the extraction to make the test material indicated?
3b	Methods (experimental	2	4	5 criteria: 1 – are the details of administration/application of test material(s) described? 2 – is the test system of relevance for TCM studies? (human enzyme, human cells etc...) 3 – are proper controls used? 4 – are there quality controls and/or characterization of model(s)? 5 – is there a description of and justification for the statistical methodology used? according to the type of test used a bonus is added to the final score for this section (i.e. after multiplication by the relevance factor) as follows: • <u>protocols</u>: single protocols: none • multiple protocols (using different strategies to address the same issue from different perspectives): 3 points • microarray studies: 5 points • <u>test systems</u>: purified target molecules: none cell components: 1 point cell based tests: single type of cell cultures: 2 points co-cultures: 3 points tissue culture: 3 points whole blood 4 isolated organs 5 computer models 6 more bonus points are added if the paper specifies: dose justification 3 supplementation of the data by in-vivo tests 4
	Objectives			See "introduction and background"
4	Results	1	5	7 criteria

				1 - is it clear that n >= 3 (independent repeats)? 2 – is the sample size appropriate? 3 – are controls used? 4 – are the outcome measures clearly defined? 5 - are the test results appropriately used in the statistical analyses? 6 - is the significance clearly established in figures and tables? 7 - are there data showing selectivity of biological effect and its specificity?
5	Conclusion and Discussion	3	3	4 criteria 1 – can conclusions be considered robust in light of the results? 2 - are the results appropriately interpreted in light of the product test dosage regimen used? 3 - are conclusions linked to clinical use? 4 - do conclusions/results support or contradict existing findings?

Testing of the new scoring criteria

Following the decision to focus on herbs used in TCM preparations, a list of herbs used in two different TCM formulations was prepared (provided by Prof. Simmonds, see Table 1) and assigned by WP4 coordinator to each WP4 member (Table 8). The list shows comprehensive naming information on the constituents on the formulae under a few categories:

The assignment of the herbs is as follows and most participants gave their consensus:

Table 8 – Herb assignment

parent plant name	**member**
Alisma orientale (Sam.) Juzep.	AlessandroBuriani
Cornus officinalis Sieb. et Zucc.	Enrica Bosisio
Dioscorea opposita Thunb.	Atanas Atanasov
Paeonia suffruticosa Andr.	Monique Simmonds
Poria cocos (Schw.) Wolff	Verena Dirsch
Rehmannia glutinosa Libosch.	Maria Carrara
Ganoderma	LauraMaria Laura Garcia Bermejo
Angelica pubescens Maxim. *f.biserrata* Shan et Yuan	Angelika Vollmar
Astragalus membranaceus Bge.	Stefan Zahler
Atractylodes macrocephala Koide	Qihe Xu
Scutellaria baicalensis George	Tai-ping Fan
Tripterygium wilfordii Hook.*f.*	Jue Zhou

The WP4 Coordinator then gave the following instructions for assignees:
step 1: paper collection on the assigned herb.
step 2: selection of the articles to be included in the datalist based on the subject (in vitro pharmacology, functional genomics etc....)
step 3: application of inclusion/exclusion criteria:

a - only papers dealing with *in vitro* pharmacology of phytocomplexes related to CHM are to be considered;
b - articles on purified compounds are excluded, given that they follow the roles of any other scientific paper;
c - the paper has to be listed in the Pub Med search engine;
d - articles must be in English language;
e - as a rule the articles will have to be published in the last 5 years, even though exceptions will be taken into consideration if properly justified).
Note that these criteria apply to the evaluation process, while the inclusion of papers in the appropriate section of the data list has no limitations, since they can be included even if they are not evaluated
step 4: selection of one paper by each member for the evaluation, collection by the coordinator, and preparation of the final list of the papers to be scored.
step 5: mailing of the papers to the members and scoring of the articles according to the new criteria.
step 6: scores are sent to the Coordinator
step 7: comparison of the result and validation of the scoring procedure

Eight members concluded the assigned work on article search and selection (Table 9), and sent the paper to the coordinator. Articles were thus checked for eligibility and a final list of 5 articles was sent to all WP4 members for evaluation.

Table 9 – Article eligibility

plant	plant assignee	article	notes
Atractylodes macrocephala Koide	Qihe Xu	Reactive oxygen species mediation of Baizhu-induced apoptosis in human leukemia cells – Huang H-L et al. - Journal of Ethnopharmacology 97 (2005) 21–29	Eligible for evaluation
Dioscorea opposita	Atanas Atanasov	Neuroprotective effects of *Dioscorea opposita* on scopolamine-induced memory impairment in *in vivo* behavioral tests and *in vitro* assays. Yang MH et al. - Journal of Ethnopharmacology 121 (2009) 130–134	Eligible for evaluation
Cornus officinalis	Enrica Bosisio	Fructus Corni suppresses hepatic gluconeogenesis related gene transcription,enhances glucose responsiveness of pancreatic beta-cells, and prevents toxin induced beta-cell death – Chen CC et al. - Journal of Ethnopharmacology 117 (2008) 483–490	Eligible for evaluation
Alisma orientale	Alessandro Buriani	*In vitro* Antidiabetic Activities of Five Medicinal Herbs used in Chinese Medicinal Formulae – Lau CH et al. - *Phytother. Res.* 22, 1384 –1388 (2008)	Eligible for evaluation
Rehmannia glutinosa	Maria Carrara	Hot water-extracted Lycium barbarum and Rehmannia glutinosa inhibit proliferation and induce apoptosis of hepatocellular carcinoma cells – Chao J C-J et al. - *World J Gastroenterol* 2006 July 28; 12(28): 4478-4484	Eligible for evaluation
Poria cocos	Verena Dirsch	Inhibition of Tumor-Promoting Effects by Poricoic Acids G and H and Other Lanostane-Type Triterpenes and	The paper is actually dealing with purified compounds and is

		Cytotoxic Activity of Poricoic Acids A and G from *Poria cocos* – Ukiya M - *J. Nat. Prod.* 2002, *65,* 462-465	not considered for evaluation
Angelica pubescens	Angelika Vollmar	Angelmarin, a novel anti-cancer agent able to eliminate the tolerance of cancer cells to nutrient starvation – Awale S. Et al. - Bioorganic & Medicinal Chemistry Letters 16 (2006) 581–583	The paper is actually dealing with purified compounds and is not considered for evaluation
Ganoderma	Maria Laura Garcia Bermejo	Ganoderic acid T inhibits tumor invasion *in vitro* and *in vivo* through inhibition of MMP expression – Chen N-H. - Pharmacological Reports 2010 – 62 - 150-163	The paper is actually dealing with purified compounds and is not considered for evaluation

In order to reduce individual mistakes, miscalculations and misunderstandings between members, an electronic Excel spreadsheet pre-set to perform automatically all the calculations, was prepared and sent to all members, asking to use it for scoring the papers:

Article #	Section 1 (x1)	Section 2 (x2)	Section 3a (x4	Section 3b (x4	Section 4 (x5)	Section 5 (x3)	Bonus from Section 3b	total score
1	0	0	0	0	0	0		
	0	0	0	0	0	0		0
2	0	0	0	0	0	0		
	0	0	0	0	0	0		0
3	0	0	0	0	0	0		
	0	0	0	0	0	0		0
4	0	0	0	0	0	0		
	0	0	0	0	0	0		0
5	0	0	0	0	0	0		
	0	0	0	0	0	0		0
6	0	0	0	0	0	0		
	0	0	0	0	0	0		0
7	0	0	0	0	0	0		
	0	0	0	0	0	0		0
8	0	0	0	0	0	0		
	0	0	0	0	0	0		0
9	0	0	0	0	0	0		
	0	0	0	0	0	0		0
10	0	0	0	0	0	0		
	0	0	0	0	0	0		0
11	0	0	0	0	0	0		
	0	0	0	0	0	0		0

Figure 7 – Evaluation spreadsheet

Second collective scoring test using the new criteria

Papers chosen by WP4 participants for evaluation (see Table 9):
1 - *Reactive oxygen species mediation of Baizhu-induced apoptosisin human leukemia cells*
2 - *Neuroprotective effects of Dioscorea opposita on scopolamine-induced memory impairment in in vivo behavioural tests and in vitro assays*
3 - *Fructus Corni suppresses hepatic gluconeogenesis related gene transcription, enhances glucose responsiveness, of pancreatic beta-cells, and prevents toxin induced beta-cell death*
4 - *In vitro antidiabetic activities of five medicinal herbs used in Chinese Medicinal Formulae*
5 - *Hot water-extracted Lycium barbarum and Rehmannia glutinosa inhibit proliferation and induce apoptosis of hepatocellular carcinoma cells*

To date, six evaluators sent back their scorings as follows:

Evaluator N.1

#	Section 1 (x1)	Section 2 (x2)	Section 3a (x4)	Section 3b (x4)	Section 4 (x5)	Section 5 (x3)	Bonus	total score
1	4	3	8	3	4	2		
	4	6	32	12	20	6	5	**85**
2	3	4	4	3	5	3		
	3	8	16	12	25	9	10	**83**
3	4	4	7	3	6	3		
	4	8	28	12	30	9	5	**96**
4	4	4	8	5	5	3		
	4	8	32	20	25	9	6	**104**
5	3	4	4	4	5	4		
	3	8	16	16	25	12	5	**85**

Evaluator N.2

#	Section 1 (x1)	Section 2 (x2)	Section 3a (x4)	Section 3b (x4)	Section 4 (x5)	Section 5 (x3)	Bonus	total score
1	4	3	5	2	2	2		
	4	6	20	8	10	6	5	**59**
2	4	4	3	2	6	2		
	4	8	12	8	30	6	9	**77**
3	4	4	7	4	7	2		
	4	8	28	16	35	6	5	**102**
4	4	4	6	3	6	2		
	4	8	24	12	30	6	5	**89**
5	4	4	6	3	6	0		
	4	8	24	12	30	0	5	**83**

Evaluator N.3

#	Section 1 (x1)	Section 2 (x2)	Section 3a (x4)	Section 3b (x4)	Section 4 (x5)	Section 5 (x3)	Bonus	total score
1	3	4	7	3	4	2	0	
	3	8	28	12	20	6	3	**80**
2	2	4	4,5	3	1	1		
	2	8	18	12	5	3	10	**58**
3	4	4	7	2	5	4		
	4	8	28	8	25	12	6	**91**
4	2	4	8	2	5	2,5		
	2	8	32	8	25	7,5	5	**87,5**
5	2	4	3	1	1	1		
	2	8	12	4	5	3	2	**36**

Evaluator N.4

#	Section 1 (x1)	Section 2 (x2)	Section 3a (x4)	Section 3b (x4)	Section 4 (x5)	Section 5 (x3)	Bonus	total score
1	4	4	7	3	4	2		
	4	8	28	12	20	6	5	**83**
2	3	4	2	3	5	3		
	3	8	8	12	25	9	10	**75**
3	4	4	8	3	6	4		
	4	8	32	12	30	12	5	**103**
4	4	4	8	5	5	2		
	4	8	32	20	25	6	6	**101**
5	4	4	5	4	5	4		
	4	8	20	16	25	12	5	**90**

Evaluator N.5

#	Section 1 (x1)	Section 2 (x2)	Section 3a (x4)	Section 3b (x4)	Section 4 (x5)	Section 5 (x3)	Bonus	total score
1	4	3	4	3	2	2		
	4	6	16	12	10	6	5	**59**
2	4	5	3	4	3	4		
	4	10	12	16	15	12	10	**79**
3	5	5	3	2	2	2		
	5	10	12	8	10	6	2	**53**
4	3	5	5	4	3	3		
	3	10	20	16	15	9	5	**78**
5	4	2	2	3	2	2		
	4	4	8	12	10	6	2	**46**

Evaluator N.6

#	1 Section 1 (x1)	2 Section 2 (x2)	3 Section 3a (x4)	4 Section 3b (x4)	5 Section 4 (x5)	6 Section 5 (x3)	7 Bonus	total score
1	1	8	28	12	10	6		65
2	3	8	28	68	30	36		168
3	4	8	44	48	35	9		148
4	1	8	28	44	35	3		130
5	6	25	40	28	8	3		110

Six WP4 members participated to the evaluation. 4 members (1, 2, 3, 5) were among those participating to the follow-up meeting where the new criteria were agreed on. Two members (4 and 6) did not participate, but one (4) had a face to face meeting before the evaluation with one of the WP4 kick off meeting participants.

It is apparent that the variability among different evaluators is improved with the scoring procedure. Evaluator number 6 usually gives a much higher score with respect to the other evaluators. This first observation is probably linked to the fact that the evaluator just received the new scoring procedure without having a chance of a face to face explanation.

Having found a reasonable motif why evaluator number 6 has been giving higher scores respect to the others, we can consider it an outlier and as such exclude it from further collective analysis. Nevertheless this finding gives us a further guideline for the application of the scoring criteria, *i.e.* the clear need to explain in person the criteria to a member before he or she proceeds to the evaluation.

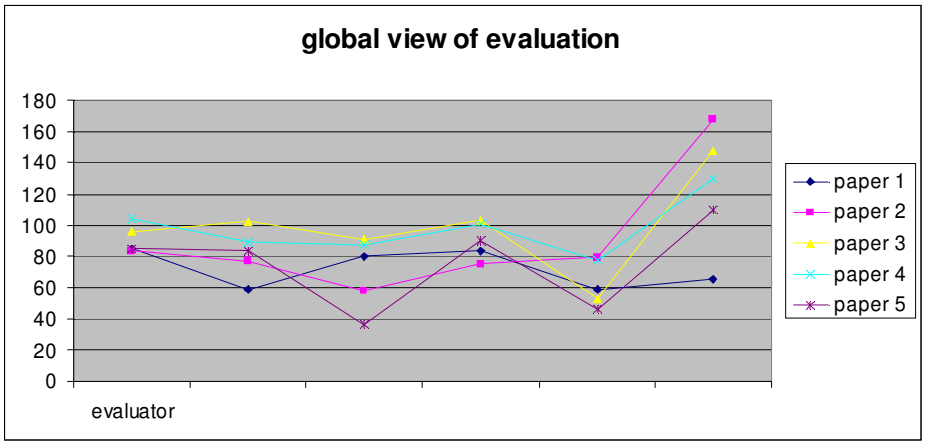

Figure 8 – Global view of evaluation

The analysis of the results at this point proceeds without taking into account evaluator number 6. The graph shows several differences in the evaluation of the papers, but with no clear

underestimation of superevaluation trends related directly to the approach of the single evaluators. See figure 9 and table 10

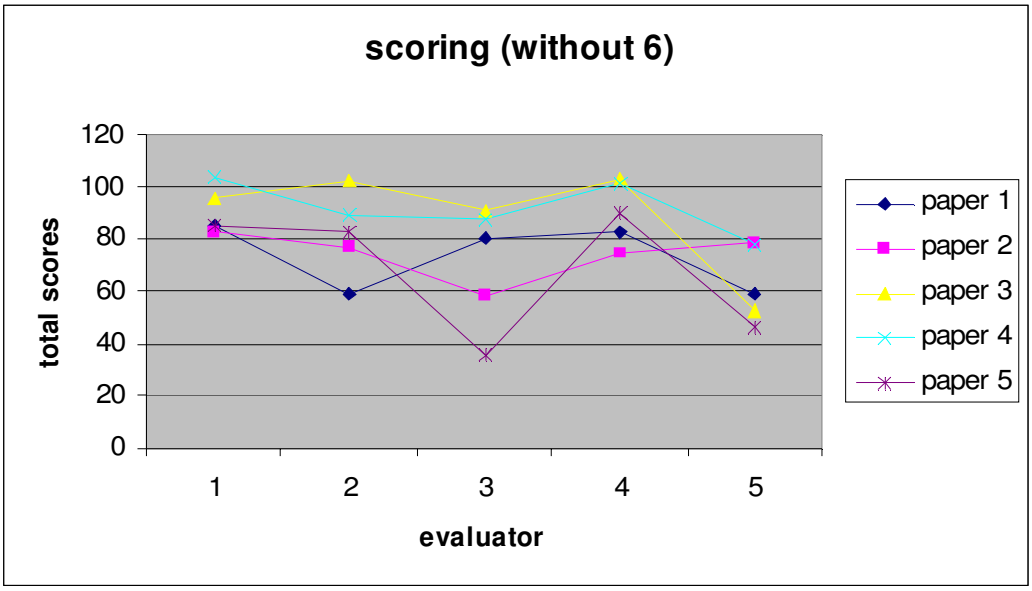

Figure 9 – Scoring results after omission of 6

TABLE 10 Scores of 5 papers, 5 reviewers

evaluator	paper 1	paper 2	paper 3	paper 4	paper 5
1	85	83	96	104	85
2	59	77	102	89	83
3	80	58	91	87,5	36
4	83	75	103	101	90
5	59	79	53	78	46
mean	**73,2**	**74,4**	**89**	**91,9**	**68**
S.D.	**13**	**9,6**	**20,7**	**10,6**	**25**

Most disagreement is found for papers n.3 and n.5 (SD higher than 20). Looking into those two papers it is possible to try to track back the specific article sections where most differences are found (Table 11 and 12 - raw scores represent the original scores for each criterion before they are multiplied by the relevance factor):

TABLE 11 Raw scores for paper 3								
evaluator	raw score section 1	raw score section 2	raw score section 3a	raw score section 3b	row score section 5	row score section 6	bonus	total score
1	4	4	7	3	6	3	5	96
2	4	4	7	4	7	2	5	102
3	4	4	7	2	5	4	6	91
4	4	4	8	3	6	4	5	103
5	5	5	3	2	2	2	2	53
mean	**4.2**	**4.2**	**6.4**	**2.8**	**5.2**	**3.0**	**4.6**	**89.0**
SD	**0.4**	**0.4**	**1.9**	**0.8**	**1.9**	**1.0**	**1.5**	**20.7**
% var.	**10.6**	**10.6**	**30.5**	**29.9**	**37.0**	**33.3**	**33.0**	**23.3**

TABLE 12 Raw scores by section for paper 5								
evaluator	raw score section 1	raw score section 2	raw score section 3a	raw score section 3b	raw score section 5	raw score section 6	bonus	total score
1	3	4	4	4	5	4	5	85
2	4	4	6	3	6	0	5	83
3	2	4	3	1	1	1	2	36
4	4	4	5	4	5	4	5	90
5	4	2	2	3	2	2	2	46
mean	**3.4**	**3.6**	**4.0**	**3.0**	**3.8**	**2.2**	**3.8**	**68.0**
SD	**0.9**	**0.9**	**1.6**	**1.2**	**2.2**	**1.8**	**1.6**	**25.0**
% var	**26.3**	**24.8**	**39.5**	**40.8**	**57.1**	**81.3**	**43.2**	**36.8**

The overall largest percentage of variation for total scores is found for paper n. 5 (36.8%) and most of it lies in the scoring of sections 6 (81.3 %) and 5 (57.1 %). This pattern is similar, although attenuated, to paper 3 with sections 5 and 6 again being the most variable (33.3% and 37%, respectively). In the new scoring procedure section n 5 (results) represent the criterion weighing heaviest on the total score since the value is amplified by a relevance factor of 5. It can also be noticed that the single scores that contribute most to the variability, are low (0 and 1), but they are not given consistently by the same evaluator.

Considering the nature of the sections, they can be listed according to the type of the criteria used in terms of objectivity:

Section 1(title and abstract): formal/objective
Section 2 (introduction): subjective
Section 3a (methods - plant description): formal/objective
Section 3b (methods – experimental): formal/objective
Section 4 (results): some formal/objective, some subjective

Section 5 (conclusions/discussion): subjective

The two sections with the highest variability are also among the less objective ones, which in part can explain the scoring differences among members.

Differences found in the scoring of formal criteria are most likely due to misunderstanding or misinterpretation of the written criteria and as such could be corrected with a more precise definition of the criteria themselves, or with a more careful explanation to each evaluator. On the contrary, differences found for subjective criteria are rather related to each one's personal views, none of which, among peers, can be considered more appropriate than another. Differences in personal judgements are in fact a value per se, especially when putting together quality guidelines. Still the weight that has been given to such criteria should be questioned, given that quality is supposed to be based upon objective values. Basically, when dealing with quality evaluation of the formal presentation of a scientific paper, objective criteria should weight more than the subjective ones. Moreover, when considering subjective criteria in a scientific paper, one might actually indirectly and inappropriately judge the paper in terms of scientific value, which is something that is supposed to have been done already by the journal's editors and reviewers.

Maintaining the indicated criteria even though they can be considered subjective, could and should be done in the evaluation scheme, since they represent a guarantee that existing cultural differences within the scientific community are taken into account, but cannot be weighted more than the objective ones, otherwise it would not be possible, to lower down the intrinsic variability of the system and reach a consensus on a consistent quality evaluation procedure.
With these considerations in mind the relevance factors have been reconsidered so that variability due to personal opinions and views will not be eliminated, but simply deflated down to a more appropriate dimension.

The relevance factors were then adjusted according to both their objectivity as well as the importance of the section:

Section 1 (title/abstract): 1,5
Section 2 (introduction): 1
Section 3a (plant description): 3
Section 3b (experimental): 3
Section 4 (results): 2
Section 5 (discussion): 1

Some thought must also be given to the bonuses issue. A bonus is given if a paper contains some aspect that makes it more relevant in terms of functional genomics, and more in general, of systems biology applied to TCM research. Some experimental models, even when not using high throughput assays, might be using experimental designs that allow a more multisystem view of the results. Still if a paper is using correctly an experimental model, without using a systemic approach, its quality should not be judged less. Quality is indeed a different issue with respect to the scientific approach, which is instead a relevant aspect for the specific aim of GP-TCM. The suggestion is to enucleate the score for bonuses from the quality evaluation and keep it separate to mention in the presentation of the paper in terms of degree of relevance for functional genomics in TCM research. It is worth to consider that the criteria used to define the bonuses should also be part of the guidelines for researchers in TCM who want to use a functional genomics and/or systems biology approach.

Using the new scoring procedure the differences, though maintained, are indeed attenuated (Fig 10 - 11 and Table 13):

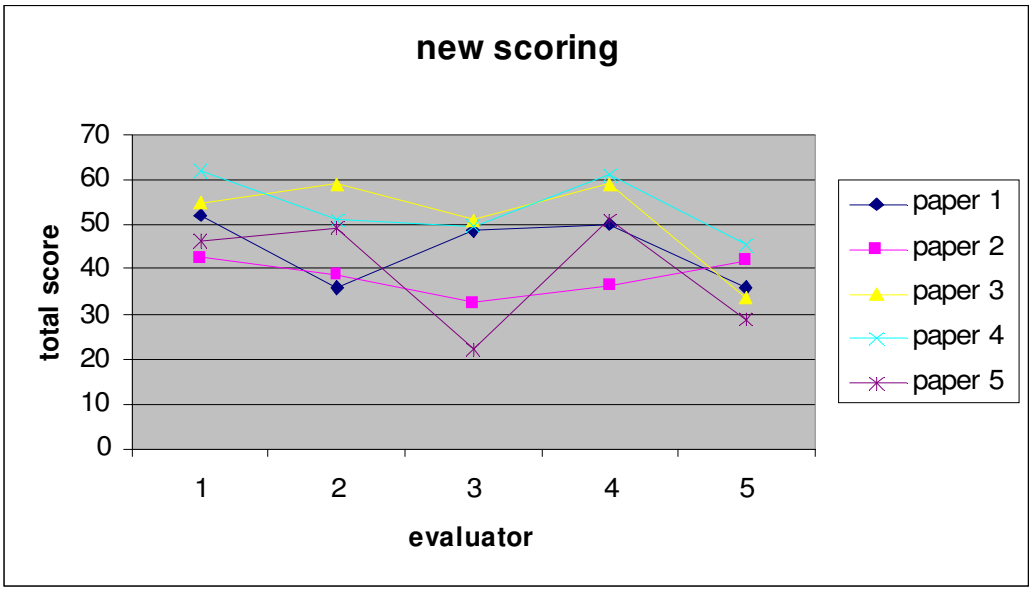

Figure 10 – Scores by evaluator using the new procedure

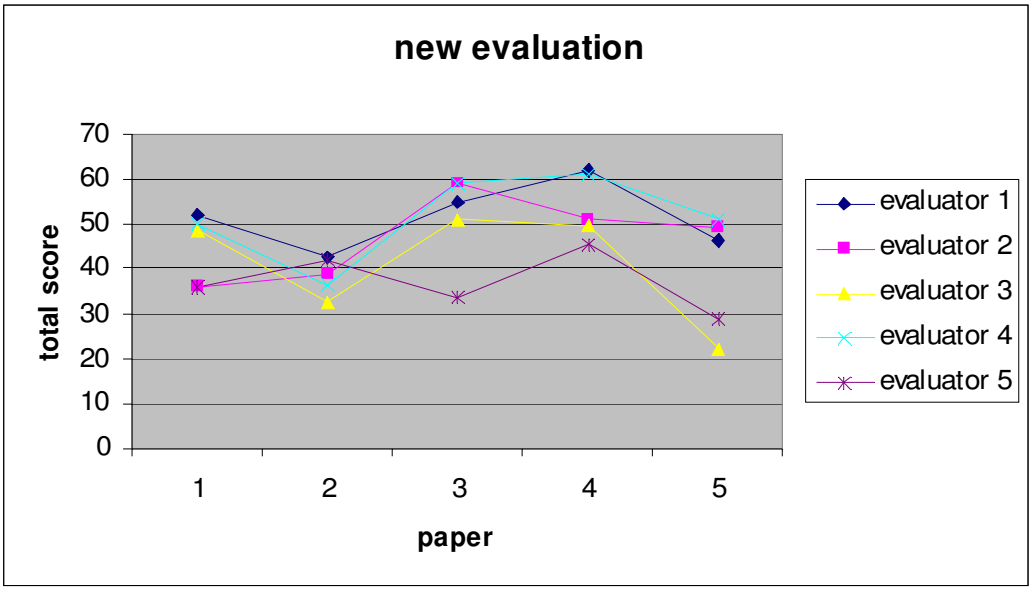

Figure 11 – paper's evaluations using the new procedure

TABLE 13 New criteria applied by each evaluator					
evaluator	**paper 1**	**paper 2**	**paper 3**	**paper 4**	**paper 5**
1	**52**	**42,5**	**55**	**62**	**46,5**
2	**36**	**39**	**59**	**51**	**49**
3	**48,5**	**32,5**	**51**	**49,5**	**22**
4	**50**	**36,5**	**59**	**61**	**51**
5	**36**	**42**	**33,5**	**45,5**	**29**
mean	**44,5**	**38,5**	**51,5**	**53,8**	**39,5**
S.D.	**7,8**	**4,1**	**10,5**	**7,3**	**13,1**
% VAR.	**17,7**	**10,7**	**20,6**	**13,6**	**33,2**

A drastic improvement is seen in the standard deviations among scores (see Tables 10 and 13 for comparison) even though the trend is maintained and thus each evaluator's judgement is not artificially modified.

Looking at the total scorings according to the different evaluators, 4 out of 5 evaluators have a rather similar trend while the other one has a different evaluation from all the others for paper n. 3 (a score of 22, almost half of the mean value of the scores).

Three evaluators agree on the best paper (n 4), while for the remaining two it is the second best. Two evaluators agree on the worst paper (n 2), while two others score it as the second worst and one second best.

Considering that the maximum score possible is 82, the papers scores range from 48% (38,5) to 66% (53,8) of the best possible score.

Considering the bonuses separately:

TABLE 14 (bonuses)					
evaluator	paper 1	paper 2	paper 3	paper 4	paper 5
1	5	10	5	6	5
2	5	9	5	5	5
3	3	10	6	5	2
4	5	10	5	6	5
5	5	10	2	5	2
mean	**4.6**	**9.8**	**4.6**	**5.4**	**3.8**
S.D.	**0.9**	**0.4**	**1.5**	**0.5**	**1.6**
% VAR.	**19.4**	**4.6**	**33.0**	**10.1**	**43.2**

The bonuses for papers 1, 2 and 4 are rather consistent among evaluators, while the differences increase for n. 3 and 5. Interestingly those are the same papers with the most variable quality scores as well.

Conclusions from the trial:

- Evaluators should be given explanations directly on the scoring procedures
- There is an improvement with respect to the previous procedure and most scores show the same trends and acceptable variabilities among different evaluators
- It is possible to further improve the variability of the scorings, maintaining the different opinions of the single evaluator, by lowering the value of relevance factors and adjusting them taking into account both the objectivity as well as the scientific importance of the criteria (see the final table with the list of criteria)
- The variability of the scoring reflects different scientific viewpoints, which should be maintained in the scoring process. The evaluation process for each paper should then be carried out at least by three different evaluators. In case of variability higher that 25% of the total score, the paper should be further evaluated by at least by two more scientists until a variability below 25% is achieved.
- The quality score should be considered separately for each paper from the relevance bonuses for TCM research. Bonuses will be maintained in the evaluation procedure, but will lead to a new different parameter: Relevance for TCM Research.
- The present scoring procedure is an achievement, but it is also a starting point, a common basis which is intended to be further improved during the ongoing work of the evaluators. It is foreseeable that at least once a year the criteria might be updated. Accordingly, the scores previously attributed will be updated in the on line repository and the date of the scoring registered and shown. For this reason each evaluator will have to store each paper's working sheet to be eventually updated.
- In order to organize the scoring procedure, the WP4 Coordinator will appoint the triplets of evaluators (four groups, based on the number of WP4 members)

Following these considerations a new scoring sheet was produced to be used for future evaluations

Procedure for Article Evaluation

WP4 members have been assigned a specific herb related to CHM (see Table 8). Most of them have collected the papers using a Pub Med search, which will have to be updated at least every six months.

Every two months the Coordinator will ask each assignee to send at least ten papers for the evaluation. The papers should be chosen among those collected according to the following procedure:

step 1: formation of triplets of WP4 members for articles evaluation, by the coordinator. The composition of each triplet will benefit the presence of at least one member who had previously participated to the selection of the criteria during the follow-up meeting. This member, supported by the CO, will have to contact the others in order to guarantee that a good explanation for each criterion is given. The triplets should represent operative units helping the organization and the coordination of the project and a spokesperson will be named for each triplet.

step 2: the Coordinator sends each assignee the latest version of the criteria, bonus, inclusion/exclusion criteria, scoring procedures and scoring spreadsheet.

step 3: each assignee will perform the search and collection of the articles on the assigned herb (with a bi-monthly update).

step 4: each assignee will perform the selection of the articles to be included in the datalist based on the subject (in vitro pharmacology, functional genomics etc....) based also on inclusion/exclusion criteria:

a - only papers dealing with *in-vitro* pharmacology of phytocomplexes related to CHM are to be considered;

b - articles on purified compounds are excluded, given that they follow the roles of any other scientific paper;

c - the paper has to be listed in the Pub Med search engine;

d - articles must be in English language;

e - as a role the articles will have to be published in the last 5 years, even though exceptions will be taken into consideration if properly justified).

Note that these criteria apply to the evaluation process, while the inclusion of papers in the appropriate section of the data list has no limitations, since they can be included even if they are not evaluated

step 5: at least 10 papers will be selected each two months by each member for the evaluation and the spokesperson of the triplet will send them to the Coordinator

step 6: collection of the articles by the coordinator, eligibility check and preparation of the final lists (one for each triplet) of the papers to be scored.

step 7: mailing of the papers to the triplet spokesperson and their distribution

step 8: scoring of the articles according to the new criteria, and classification of each paper according to the following characteristics, which, besides the scoring and the bonus, correspond to the rest of the fields present in the on-line repository

- **Article** (title, authors, journal, year and pages)
- **Plant** (scientific name of the main plant)
- **WP4 reference member**
- **Plant/s used** (if a phytocomplex from one or more plants are used)
- **Formula used** (if a phytocomplex from a formula is used)
- **Purified molecules used** (if single compounds are used)
- **Disease/s**
- **molecular target/s and/or mechanisms investigated**

Table 15 – Final Criteria

No	*Section in article to be reviewed*	*Relative importance of criterion*	*Relevance factor*	*Description of criteria*	*Post evaluation comments*
1	Title and abstract	5	1,5	4 criteria: Does either the title or abstract or both provide the herbal medicinal product's 1 - Latin binomial name of the plant (or clear name of the product). This point will be scored in this section even when the latin binomial name might be present elsewhere in the paper) 2 - part of the plant used? 3 – experimental use of the plant/product? 4 – are the article's contents accurately described?	The scoring in this section have been the least variable. The relevance factor has been increased fro 1 to 1,5 in consideration of the objectivity of its criteria
2	Introduction and background	**4**	1	4 criteria: 1 – is the scientific background presented? 2 – is an explanation of the rationale presented, including a brief statement of the reasons for the study with reference to the specific herbal medicinal product and its traditional use? 3 - is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated? (novelty) 4 – are the objectives of the study clear? (the objectives will be scored in this section even when, depending on the journal, might be found elsewhere in the paper)	The relevance factor has been decreased given the high degree of subjectivity of the criteria of this section
3a	Methods – (plant description)	2	3	Number of criteria are variable depending on the type of material used: GENERAL (apply to all) 8 criteria: do the methods describe: 1 - Solvent used and ratio for the extraction 2 – time of extraction 3 - temperature of extraction 4 - yield of extraction 5 – is the method of authentication of the herbal raw material indicated? 6 – are details of any voucher specimen included? 7 – has the test material been subjected to simple chemical constituent profiling and/or complex fingerprinting (by what methods and by whom) 8 – has the material been standardised (by what process and by whom) IN CASE OF UNPROCESSED PLANT AND/OR MIXTURES OF PLANTS 2 criteria: 1 – is the herbal product name clearly indicated? 2 – is the part of the plant used to make the product specified?	The relevance factor in this section has been decreased to 3, following a general reconsideration of the size of such factors. Nevertheless it is the higer relevance factor, in consideration of both the importance of the section in terms of scientific quality as well as objectivity of criteria

				IN CASE OF A PROCESSED PLANT 5 criteria: 1 - are the processed product name or the extract name and the name of the manufacturer of the product indicated? 2 - is the batch number of the herbal product name indicated? 3 – is the part of the plant used to make the product or the extract specified? Where applicable: 4 – is the type of preparation to make the test material described? 5 – is the yield of the extraction to make the test material indicated? IN THE CASE OF A PROPRIETARY PRODUCT 5 criteria: 1 - are the proprietary product name or the extract name and the name of the manufacturer indicated? 2 – is the batch number of the product indicated? 3 – is the part of the plant used to make the product or the extract indicated? Where applicable: 4 – is the type preparation to make the test material described? 5 – is the yield of the extraction to make the test material indicated?	
3b	Methods (experimental	2	3	5 criteria: 1 – are the details of administration/application of test material(s) described? 2 – is the test system of relevance for TCM studies? (human enzyme, human cells etc...) 3 – are proper controls used? 4 – are there quality controls and/or characterization of model(s)? 5 – is there a description of and justification for the statistical methodology used?	The relevance factor in this section has been decreased to 3, following a general reconsideration of the size of such factors. Nevertheless it is the higer relevance factor, in consideration of both the importance of the section in terms of scientific quality as well as objectivity of criteria
	Objectives			See "introduction and background"	
4	Results	1	2	7 criteria 1 - is it clear that n >= 3? 2 – is the sample size appropriate? 3 – are controls used? 4 – are the outcome measures clearly defined? 5 - are the test results appropriately used in the statistical analyses? 6 - is the significance clearly established in figures and tables? 7 - are there data showing selectivity of biological effect and its specificity?	The relevance factor in this section has been decreased to 2, following a general reconsideration of the size of such factors and the relative subjectivity of some of the criteria.
5	Conclusion and Discussion	3	1	4 criteria 1 – can conclusions be considered robust in light of the results?	The relevance factor has been decreased given the high degree of subjectivity

				2 - are the results appropriately interpreted in light of the product test dosage regimen used? 3 - are conclusions linked to clinical use? 4 - do conclusions/results support or contradict existing findings?	of the criteria of this section

The bonuses are taken out of the quality scoring system and constitute a separate score of relevance for TCM in vitro pharmacology, where any aspect allowing a more systems biology approach, or a close relation to clinical findings is counted:

BONUS

protocols:

- single protocols: none
- multiple protocols: 3 points
- high throughput analysis/microarray studies: 5 points

test systems:

- purified target molecules: none
- cell components: 1 point (2 if human)

cell based tests:

- single type of cell cultures: 2 points (3 if human)
- co-cultures: 3 points (4 if human)
- tissue culture: 3 points (4 if human)
- whole blood: 4 points (5 if human)
- isolated organs: 5 points (6 in human)

other:

- computer models: 6 points
- supplementation of the data by in-vivo tests: 4 points
- supplementation of the data by clinical findings: 6 points

Article #	Section 1 (x1.5)	Section 2 (x1)	Section 3a (x3)	Section 3b (x3)	Section 4 (x2)	Section 5 (x1)	total score	Bonus from section 3b
1	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
2	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
3	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
4	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
5	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
6	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
7	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
8	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
9	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
10	0	0	0	0	0	0		
	0	0	0	0	0	0	0	
11	0	0	0	0	0	0		
	0	0	0	0	0	0	0	

Figure 12 - New scoring spreadsheet

step 9: scores and descriptions are collected by the triplet's representative and sent to the Coordinator

step 10: validation of the scoring procedure by the Coordinator's office

step 11: addition of the papers to the repository by the CO, or by the triplets representatives

4 D4.7: UPDATE OF CHM TARGET ORIENTED DATABASE WITH QUALITY SCORES

4.1 Background

The database represents one of the pivotal achievements of WP4, it aims at being one of the most exploitable tools for researchers in TCM and in the end it should represent an important piece of reference for the scientific community. Considering the limited amount of resources that can be dedicated to this particular deliverable of the project, WP4 members have agreed on several points:

- More than a real database, the tool should be an on line repository of in vitro CHM research articles
- The repository will not be target guided; rather the main headings will be plant and disease. The molecular target/mechanism will be highlighted as one of the key characteristics in each paper
- The repository should be hosted in the WP4 dedicated pages of the GP-TCM web site
- The articles will be listed according to specific headings (plant, disease), while others will be noted in the description of the papers, using searchable keywords
- The fields initially covered will be limited to those indicated in D4.4 (table 1) and more plants will be covered in the future

The discussion on the subject, especially the technical parts, though it started from the WP4 kick-off meeting in October 2009, reached a final agreement only during the first GP-TCM general meeting in July 2010. This aspect, together with the delayed final draft of the scoring criteria (D4.6), did not allow the creation of the sample of the dedicated web pages till October 2010.

4.2 Some issues to be considered for the on line repository

An important issue is related to copyright, at least for those journals that do not allow free downloads of papers for the general public. This issue was on the agenda during the July 2010 WP4 meeting. It was then agreed that the institutions participating to the project could make an agreement with the consortium, for permission to use their libraries and databases. In the meantime the repository will be made available only to members of the participating institutions, so the access to the web pages would not be open to the general public. It was also decided that a version of the database where papers would be linked via their PubMed citation will be considered, so that the repository will be made available to the public.
It is foreseen that once the web pages will be accessible and papers available to the public or members, the repository will be constantly updated by the WP4 website administrator with the material coming from reviewers of each section and will continue to be managed even after the end of the project, just like the rest of the web pages, by the European Society of Chinese Medicine Research, to be constituted.

4.3 Building up the online WP4 repository

Detailed criteria for the on line repository are indicated in D4.4 and D4.5, including the minimum number of papers that can be considered adequate before a specific voice can be made available to the public. This number has been established in 100 for each disease area and will include quality scored and unscored papers (see D4.5).

During the 1st GP-TCM Annual Meeting in Henley (28 -30 July 2010), WP4 members had requested assistance from the GP-TCM project manager regarding the use of the project website to display some of their deliverable material, especially the online repository. In order to allow the WP4 members to proceed with their planned work, a practical solution had been agreed to be implemented to assist them with the use of website.

The members of WP4 would upload all the paper information on to the website as instructed by the project manager. The group should also produce the detailed categories for these papers (i.e., main disease names, sub-disease names etc) to indicate clearly how the papers will be arranged/listed under the website. The project manager would then liaise with the project website developer to find out ways to best present the WP4 information. WP4 members should work closely with the project manager, who would provide them general assistance, to update the prepared pages to get it as close as possible to their original requirements. This requirement was initially as follows:

The viewer should first see a page containing some kind of search engine and a list of links with Main Categories. The search engine could be the standard search engine that is built in our website or Google set to do searching just on our website. Main Categories links could be for example DISEASE, PLANT etc. When the viewer clicks on some of the Main Categories – a new page will open that will contain the hyperlinks with respective Subcategories of the Main Category (e.g., Subcategories of the Main Category DISEASE will be CANCER, DIABETES etc.). When the viewer clicks on some of the hyperlinks representing a Subcategory – a new page will open containing list with all papers that are related to it (for example when CANCER is clicked – a new page will open containing a list with all papers in our depository that are related to cancer).
The project manager indicated that, due to limited technical capability of the project website (and budget), the WP4 requirement might not be fully designed as specified above, nevertheless, the CO would look for the best possible solution that is available with the resources in hand.

Following the contacts between the Project Manager and the web site developer, some adjustments were done, which were considered satisfactory by the WP4 Coordination Team for the online repository. WP4 Coordination, in collaboration with the project manager, have thus been working towards establishing the final version of the WP4 repository at the GP-TCM website. Currently, the database has been implemented by the WP4 Coordination team members. The database comes with an introductory section, where the aim and focus of this study are defined. Furthermore, detailed information regarding the main sections (i.e. diseases, plants) is also provided for the interested readers. Having checked the information, the user is then directed to the bottom of the page, where the list of diseases or plants is provided. The user can view total number of papers under each main section, expand the main section to view the sub-sections, and choose a plant, a disease or sub-disease to directly see the full paper details associated with these titles. The papers will be ordered according to their scores. The WP4 members are providing the paper evaluations and the characteristics of the papers that have to be highlighted in the notes. Hence, the paper details and their evaluation score details will be uploaded to the website once WP members' evaluations are finalised. Since the database is an ongoing project and presently has insufficient number of uploaded papers, at the moment it can only be accessed by the consortium members. However, WP4 envisages reaching the sufficient number of articles in the next few months, and have the database open to public by late 2010 - early 2011.

Articles will be constantly reviewed by triplets of WP4 members, scored, characterised and uploaded as described (see step 8 of final scoring procedure - D4.6).

A scheme of how the scores and characteristics of a paper will be presented for its uploading is shown in figure 13, where a first set of scored papers is reported.

article	plant	WP4 reference scientist	plant/s used	formula used	purified molecule/s used	disease/s	molecular target/s and/or mechanism/s investigated	general quality score (mean ± SD)	relevance score for TCM
Reactive oxygen species mediation of Baizhu-induced apoptosis in human leukemia cells – Huang H-L et al. - Journal of Ethnopharmacology 97 (2005) 21–29	*Atractylodes macrocephala Koide*	Dr. Qihe Xu	Atractylodes macrocephala Koide (Baizhu)			cancer	apoptosis, ROS generation.	45 ±8	4,6 ±0,9
Neuroprotective effects of Dioscorea opposita on scopolamine-induced memory impairment in in vivo behavioral tests and in vitro assays. Yang MH et al. - Journal of Ethnopharmacology 121 (2009) 130–134	*Dioscorea opposita*	Dr. Atanas Atanasov	Dioscorea opposita			neurodegenarative diseases	neuroprotection from oxidative stress	39 ±4	9,8 ±0,4
Fructus Corni suppresses hepatic gluconeogenesis related gene transcription, enhances glucose responsiveness of pancreatic beta-cells, and prevents tozin induced beta-cell death – Chen CC et al. - Journal of Ethnopharmacology 117 (2008) 483–490	*Cornus officinalis*	Prof. Enrica Bosisio	Cornus officinalis		loganin, ursolic acid	diabetes	reduced PEPCK gene expression for hepatic gluconeogenesis, β-cell protection against toxic challenge, insulin secretion enhancement in hyperglycemic conditions	52 ±11	4,6 ±1,5
In vitro Antidiabetic Activities of Five Medicinal Herbs used in Chinese Medicinal Formulae – Lau CH et al. - Phytother. Res. 22, 1384 –1388 (2008)	*Alisma orientale*	Dr. Alessandro Buriani	Alisma orientale, Cornus officinalis, Schisandra chinensis, Poria cocos, Dioscorea opposita			diabetes	intestinal glucose uptake (absorption), inhibition of hepatic gluconeogenesis, increase in peripheral glucose uptake (insulin-dependent and independent	54 ±7	5,4 ±0,5
Hot water-extracted Lycium barbarum and Rehmannia glutinosa inhibit proliferation and induce apoptosis of hepatocellular carcinoma cells – Chao J C-J et al. - World J Gastroenterol 2006 July 28; 12(28): 4478-4484	*Rehmannia glutinosa*	Prof. Maria Carrara	Rehmannia glutinosa, Lycium barbarum			cancer	inhibition of proliferation, stimulation of p53-induced apoptosis	40 ±13	3,8 ±1,6
Inhibition of Tumor-Promoting Effects by Poricoic Acids G and H and Other Lanostane-Type Triterpenes and Cytotoxic Activity of Poricoic Acids A and G from Poria cocos – Ukiya M - J. Nat. Prod. 2002, 65, 462-465	*Poria cocos*	Prof. Verena Dirsch			poricoic acid G, poricoic acid H, poricoic acid B, poricoic acid A, tumulosic acid, dehydro-tumulosic acid, 3-epidehydro-tumulosic acid, polyporenic acid C, 25-hydroxy-3-epidehydro-tumulosic acid, dehydroabietic acid methyl ester	cancer	cytotoxicity	article not meeting inclusion criteria for evaluation	article not meeting inclusion criteria for evaluation
Angelmarin, a novel anti-cancer agent able to eliminate the tolerance of cancer cells to nutrient starvation – Awale S. Et al. - Bioorganic & Medicinal Chemistry Letters 16 (2006) 581–583	*Angelica pubescens*	Prof. Angelika Vollmar			Angelmarin	cancer	cytotoxicity under nutrient-deprivation conditions	article not meeting inclusion criteria for evaluation	article not meeting inclusion criteria for evaluation
Ganoderic acid T inhibits tumor invasion in vitro and in vivo through inhibition of MMP expression – Chen N-H.- Pharmacological Reports 2010 – 62 - 150-163	*Ganoderma*	Prof. Laura Maria Garcia Bermejo			ganoderic acid T	cancer	inhibition of proliferation, inhibition of adhesion to ECM, induction of homotypic aggregation, inhibition of cell migration, down-regulation of the expression of matrix metalloproteinase-9 (MMP-9), inducible nitric oxide synthase (iNOS), and urokinase-type plasminogen activator (uPA).	article not meeting inclusion criteria for evaluation	article not meeting inclusion criteria for evaluation

Fig.13 – characteristics and scores for uploading articles

Briefly each triplet of evaluators will have to score the assigned papers and to note their following characteristics:

- Article (title, authors, journal, year and pages)
- Plant (scientific name of the main plant)
- WP4 reference member
- Plant/s used (if a phytocomplex from one or more plants are used)
- Formula used (if a phytocomplex from a formula is used)
- Purified molecules used (if single compounds are used)
- Disease/s
- molecular target/s and/or mechanisms investigated

The notes will be given to the person in charge of uploading the papers (the CO or a representative of the triplet of evaluators themselves) in the format shown in Fig. 13

Each field will be appropriately shown in the web pages (see Fig. 14). A link to the PubMed citation will also be provided for each paper, and key words to be implemented in the search engine are taken from those listed in the paper's notes.

Figure 14 - WP4 repository at the GP-TCM website

APPENDIX I: ROLES OF BENEFICIARIES AND EXPERTS AS AT 14TH APRIL 2010:

Prof Peter Hylands (UK), **Coordinator**
Dr Jue Zhou (UK), **WP4 Assistant**
Dr Fan Qu (UK), **Research Associate**
Prof Verena Dirsch (Austria), **Deputy Coordinator**
Dr Elke Heiss (Austria), **Postdoctoral assistant**
Dr Atanas Atanasov (Austria), **Postdoctoral assistant**
Dr Alessandro Buriani (University of Padova, Italy), **Assistant Coordinator**
Dr David Barlow (UK), **Beneficiary**
Dr Maria Laura Garcia Bermejo (Spain), **Beneficiary**
Prof Enrica Bosisio (Italy), **Beneficiary**
Dr Mario Dell'Agli (Italy), **research assistant**
Prof Maria Carrara (Italy), **Beneficiary**
Dr Hani El-Nezami (China), **Beneficiary**
Dr Tai-Ping Fan (UK), **Beneficiary**
Prof Monique Simmonds (UK), **Beneficiary**
Prof Angelika Vollmar (Germany), **Beneficiary**
Dr Stefan Zahler (Germany), **Beneficiary**
Dr Qihe Xu (UK), **Beneficiary**
Dr Halil Uzuner, **GP-TCM Project Manager**

Table 1 – WP4 membership contact details

Name
Prof Peter Hylands (UK), **Coordinator**
Dr Jue Zhou (UK), **WP4 assistant**
Dr Fan Qu (UK), **Research associate**
Prof Verena Dirsch (Austria), **Deputy Coordinator**
Dr Elke Heiss, **Postdoctoral assistant**
Dr Atanas Atanasov, **Postdoctoral assistant**
Dr Alessandro Buriani (Italy), **Assistant Coordinator**
Dr David Barlow (UK), **Beneficiary**
Dr Maria Laura Garcia Bermejo (Spain), **Beneficiary**
Prof Enrica Bosisio (Italy), **Beneficiary**
Dr Mario Dell'Agli, **Research assistant**
Prof Maria Carrara (Italy), **Beneficiary**
Dr Hani El-Nezami (China), **Beneficiary**
Dr Tai-Ping Fan (UK), **Beneficiary**
Prof Monique Simmonds (UK), **Beneficiary**
Dr Marcia Tolfts, **Postdoctoral assistant**
Prof Angelika Vollmar (Germany), **Beneficiary**
Dr Stefan Zahler, **Postdoctoral assistant**
Dr Qihe Xu (UK), **Beneficiary**
Dr Halil Uzuner, **GP-TCM Project Manager**

APPENDIX II: TABLE OF SELECTED REVIEWER'S COMMENTS

Table 1 –Comments and scores from some reviewers

Note: *RIC= Relative importance of criterion*

article	*Section in article*	*Reviewer A*			*Reviewer F*			
		RIC	*Score*	*Note*	*RIC*	*Score*		*Note*
1	Title and abstract	7	1x5	Latin binomial name: Yes; part of the plant used: Yes; type of preparation tested: Yes	1	4	x1	4
	Introduction and background	5	3x5	Yes: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented?	2	5	x3	15
	Methods – plant description	2	6x3	Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	3	2	x6	12
	Methods – experimental	1	7x3	(1) What's the difference between Angelica sinensis (Oliv.) Diels, Angelica sinensis (Oliv.) and Angelica sinensis? The first and its synonyms should better be used throughout the paper; (ii) no description of authentication of the herbal raw material, no voucher specimen; no fingerprinting; (iii) results of AP composition anaysis were given but the methodology used and who did the analysis not told; (iv) the material had not been standardized.	4	3	x7	21
	Objectives	6	4x2	(i) Detailed unequivocal description of the test system, experimental protocol and the statistical methodology are satisfactory; (ii) No reference and catalogue No. for the 3 murine tumours, an HHCC cell line and the human embryo dermal fibroblast cell line; authentication of cell lines? (iii) Cannot comment on the *in vitro* assays as I am not familiar to these models.	5	1	x2	2
	Results	3	5x3		6	2	x5	10
	Conclusion and Discussion	4		(i) Sample sizes defined; primary and secondary outcome measures clearly defined. However, high doses of AP0 reduces	7	1	x4	4

				thymus weight but there was increased mortalityin S180-transplanted mice, which was not interpreted; (ii) statistics seems to be OK; (iii) data seemed to be OK but body weight and Thymus weight are not specific parameters to refect ascites volume and any beneficial effects.				
	Other	sum	90	Confidence score: 60-80%	Sum			68
2	Title and abstract	7	1x3	Title should include or even focus on baicalein.	1	4	x1	4
	Introduction and background	5	3x3	The introduction on PC-SPEC is not clear at the beginning. How many reports were based on the herbal formula only? How many are complicated due to addition of Western medicines? Are mechanisms of action of its pure herbal formala as stated in the 2nd sentence of para 1? Any known baicalein's anti-cancer property should be introduced in more details	2	5	x3	15
	Methods – plant description	2	6x2	(i) Original producer of PC-SPEC? Is the PC-SPEC purely herbal preparation? Does it contain warfarin, diethylstilbestrol, indomethacin and additive estrogen of non-herbalk origin? Purity of baicalein?	3	2	x6	12
	Methods – experimental	1	7x3	Negative control for PC-SPEC and Baicalein? If this paper, as suggested by its title, meant to compare PC-SPEC and baicalein, in all studies, PC-SPEC should be used as a positive control (This paper only used PC-SPEC in fig4 but not fig1-3).	4	2	x7	14
	Objectives	6	2x3	Should be more focused on baicalein?	5	4	x2	8
	Results	3	5x2	Loading control for fig.3a? Poorly quantitative? How individual repeats for each exp? Reproducibility? Varibility? No statistics provided at all.	6	2	x5	10
	Conclusion and Discussion	4	4x3	Jumping from PC-SPEC to and from baicalein should be more careful; it is another jump from baicalein only to all dietary flavonoids as this paper does not show that all dietary flavonoids share the anti-cancer properties as bacalein.	7	3	x4	12
	Other		73	Conflict of interest not stated; the work was reported in a meeting sponsored by industry.	Sum:			74
3	Title and abstract	7	1x5	Latin binomial name: Yes; part of the plant used: Yes; type of preparation tested:Yes	1	5	x1	5

	Introduction and background	5	3x5	Yes but how many independent repeats and how is the reproducibility of Figures 2-4 not stated: sample size; clearly defined primary and secondary outcomemeasures	2	5	x3	15
	Methods – plant description	1	7x4	Yes: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented? Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	3	2	x6	12
	Methods – experimental	6	2x4	No. statistics (Fig. 8 – t test was used although there are multiple groups; Fig. 2-4 has no statistics as there was one experiment only).Yes data showing the selectivity of the biological effect and its specificity	4	3	x7	21
	Objectives	3	5x3	Yes: appropriate strategy and robustness of the conclusions; Yes: interpretation of the result in light of the product tested and dosage regimen used? Yes Has the conclusion been linked to clinical use? NA Do the conclusions support or contradict existing findings?	5	5	x2	10
	Results	4	4x4	Yes detailed unequivocal description of the test system. Yes detailed unequivocal description of the experimental protocol.Yes description of and justification for the tatistical methodology used?	6	3	x5	15
	Conclusion and Discussion	?		Yes Are the specific objectives and hypothesis and rationale for the selection of the test models used included?	7	4	x4	16
	Other		113	Compounds-only studies get high scores in 3. confidence score: 60-80%	Sum:			94
4	Title and abstract	7	1x4	Latin binomial name: Yes; the part of the plant used: Yes. the type of preparation tested: No	1	3	x1	3
	Introduction and background	5	3x5	Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented?: Yes. Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated? Yes	2	3	x3	9
	Methods – plant description	2	6x4	herbal product name: Yes. the Latin binomial name (including the botanical authority) and family name for each herbal ingredient: Yes. the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number: NA.	3	5	x6	30

				the part of the plant used to make the product or the extract:Yes. the processing used to make the test material, type and concentration of the extraction: Yes. solvent used and the ratio of the herbal drug to extract: Yes. the method of authentication of the herbal raw material, including details of any voucher specimen:HPLC-DAD/Yes. whether the test material had been subjected to fingerprinting and by what methods and by whom: No. whether any special testing/purity testing had been carried out and by whom: HPLC-DAD. whether the material had been standardised, and by what process and by whom? No				
	Methods – experimental	1	7x2	(A)detailed unequivocal description of the test system: Yes but- RT-PCR was not quantitative and results were likely not reliable;(B)detailed unequivocal description of the experimental protocol: Yes but Whether negative controls of drug extracts were used in *in vitro* studies were not described; in *in vivo* studies, if cells from all 10 or so animals were pooled, how can we trust them as a group? It says that there was no significant variation among group members – if so, sounding that counting cells of individual aniaml is possible, why the samples were pooled? (C) description of and justification for the statistical methodology used? Yes but Duncan's NMR test has a bad reputation as it leads to high false positives; also test test is not appropriate for one-to-one comparison among multiple groups.	4	3	x7	21
	Objectives	6	2x3	Fig. 1-3: Group-pooled samples could not lead to conclusions on differences among groups as the standard variation is unknown. Fig. 4: The dose- and time-course data without vehicle control are not trustable and none of the observation was impressive given the poor loading controls and the non-quantitative nature of the methods. How many independent experiments? Fig.5: bad loading control; Fig.6: single consituent in HPLC-DAD for QC?	5	4	x2	8
	Results	3	5x2	Data do not support conclusions	6	3	x5	15
	Conclusion and Discussion	4	4x1		7	3	x4	12
	Other				Sum:			98

5	Title and abstract	7	1x5	Latin binomial name: Yes; part of the plant used: NA; type of preparation tested: NA	1	4	x1	4
	Introduction and background	5	3x3	Yes/?: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented? – Not clear why RAW264.7 mouse macrophage cell line but not colon cancer cells were used for experiments on COXs and iNOS. Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	2	3	x3	9
	Methods – plant description	2	6x3	NA the Latin binomial name (including the botanical authority) and Yes: family name for each herbal ingredient- flavonoids. No: the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number. NA the part of the plant used to make the product or the extract.NA the processing used to make the test material, NA the method of authentication of the herbal raw material, ...NA whether the test material had been subjected to fingerprinting No whether any special testing/purity testing had been carried out. No whether the material had been standardised.	3	2	x6	12
	Methods – experimental	1	7x2	Yes but (i) RT-PCR is semi-quantitative; (ii) MTT is not a specific method for cell proliferation; it can be affected by cell proliferation, cell death as well as cell metabolic status; (iii) PI staining and flow cytometry is not specific for apoptosis: detailed unequivocal description of the test system.Yes but only t test was used, even for comparison of multiple groups!!!: description of and justification for the statistical methodology used?	4	4	x7	28
	Objectives	6	2x2	No: Are the specific objectives (not focused) and hypothesis (no specific hypothesis) and rationale for the selection of the test models (use of macrophage cell line as an *in vitro* model not justified) used included?	5	3	x2	6
	Results	3	5x2	No: sample size (single dishes per group for NO and PGE2 studies and flow cytometric analysis of apoptosis and caspase 3 were alll for one experiment only and statistics was based on 3 measurements of the same samples; no statement on how many times each experiment has been done and the reproducibility of	6	4	x5	20

				experiments); clearly defined primary and secondary outcome measures. No data showing the selectivity of the biological effect and its specificity (apoptosis and proliferation assays not specific enough).				
	Conclusion and Discussion	4	4x2	No appropriate strategy and robustness of the conclusions.No interpretation of the result in light of the product tested and dosage regimen used?Yes/No Has the conclusion been linked to clinical use?Yes/No Do the conclusions support or contradict existing findings?	7	2	x4	8
	Other				Sum:			87
6	Title and abstract	7	1x2	Latin binomial name: No; part of the plant used: Yes; type of preparation tested: Yes but not clear	1	3	x1	3
	Introduction and background	5	3x5	Yes: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented? Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	2	5	x3	15
	Methods – plant description	2	6x3	Yes herbal product name.Yes the Latin binomial name (including the botanical authority) .Yes family name for each herbal ingredient.NA the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number.Yes the part of the plant used to make the product or the extract.Yes the processing used to make the test material, Yes type and concentration of the extraction solvent used and the ratio of the herbal drug to extract.No the method of authentication of the herbal raw material, including details of any voucher specimen.No whether the test material had been subjected to fingerprinting and by what methods and by whom and .Yes whether any special testing/purity testing had been carried out and by whom (HPLC).No whether the material had been standardised, and by what process and by whom?	3	5	x6	30
	Methods – experimental	1	7x3	Yes but in 'MTS assay' for MTS does not stand for 'modified trichrome stain': detailed unequivocal description of the test.Yes detailed unequivocal description of the experimental protocol.Yes	4	5	x7	35

				but for comparison of multiple groups ANOVA and t test were used: description of and justification for the statistical methodology used?				
	Objectives	6	2x5	Yes Are the specific objectives and hypothesis and rationale for the selection of the test models used included?	5	4	x2	8
	Results	3	5x4	Yes sample size; clearly defined primary and secondary outcome measures.Yes statistics (as in 4).Yes data showing the selectivity of the biological effect and its specificity	6	4	x5	20
	Conclusion and Discussion	4	4x4	Yes appropriate strategy and robustness of the conclusions.Yes interpretation of the result in light of the product tested and dosage regimen used?.Yes Has the conclusion been linked to clinical use?Yes Do the conclusions *support* or *contradict* existing findings?	7	4	x4	16
	Other	sum	102	Confidence score: 60-80%	Sum			127
7	Title and abstract	7	1x5	Latin binomial name: Yes; part of the plant used: Yes; type of preparation tested:Yes	1	2	x1	2
	Introduction and background	5	3x5	Yes: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented?Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	2	5	x3	15
	Methods – plant description	2	6x4	Yes herbal product name.Yes the Latin binomial name (including the botanical authority).Yes family name for each herbal ingredient.Yes but no batch number the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number.Yes the part of the plant used to make the product or the extract.Yes the processing used to make the test material, Yes type and concentration of the extraction solvent used and the ratio of the herbal drug to extract.Yes but no voucher number the method of authentication of the herbal raw material, including details of any voucher specimen.No whether the test material had been subjected to fingerprinting and by what methods and by whom.Yes whether any special testing/purity testing had been carried out and by whom.No whether the material had been standardised, and by what process and by whom?	3	4	x6	24

	Methods – experimental	1	7x5	Yes detailed unequivocal description of the test system.Yes detailed unequivocal description of the experimental protocol.Yes description of and justification for the statistical methodology used?	4	2	x7	14
	Objectives	6	2x5	Yes Are the specific objectives and hypothesis and rationale for the selection of the test models used included?	5	3	x2	6
	Results	3	5x5	Yes sample size; clearly defined primary and secondary outcome measures.Yes statistics.Yes data showing the selectivity of the biological effect and its specificity	6	3	x5	15
	Conclusion and Discussion	4	4x5	Yes appropriate strategy and robustness of the conclusions.Yes interpretation of the result in light of the product tested and dosage regimen used?Yes Has the conclusion been linked to clinical use?Yes Do the conclusions support or contradict existing findings?	7	2	x4	8
	Other		134	Confidence score: 60-80%	Sum			84
8	Title and abstract	7	1x4	Latin binomial name: yes but no authority; the part of the plant used: NA; the type of preparation tested: yes	1	5	x1	5
	Introduction and background	5	3x4	Previous reports on A. sinensis (Oliv.) and cancer were not reviewed (but slightly mentioned in dscussoion –literature 44) – a sentence “few studies have been made of possible antitumor effects of A. sinensis“ is not enough	2	4	x3	12
	Methods – plant description	2	6x5	Yes: herbal product name, the Latin binomial name (including the botanical authority), the part of the plant used to make the product or the extract, the processing used to make the test material, type and concentration of the extraction solvent used and the ratio of the herbal drug to extract. NA: the method of authentication of the herbal raw material, including details of any voucher specimen; whether the test material had been subjected to fingerprinting and by what methods and by whom; whether any special testing/purity testing had been carried out and by whom; whether the material had been standardised, and by what process and by whom.	3	3	x6	18

	Methods – experimental	1	7x3	(i) Many cell lines, esp. those gifts and those from less authorative sources (such as ATCC): were they authenticated? Some have even no reference; (ii) fixation before PI staining for cell cycle analysis was not described. (iii) measures: Western blot: expression of apoptosis-related genes was illustrated using one sample per group. How representative is this? How much was the variation within each group? (iv) Statistics: Student's t test should not be used in comparison of >2 groups! (v) How many indepedent exps were done for each *in vitro* study not stated; n not stated for all its *in vitro* studies.	4	5	x7	35
	Objectives	6	2x5	Not clearly described.	5	2	x2	10
	Results	3	5x4	(i) sample size: N=6 per group is small; (ii) clearly defined primary and secondary outcome measures: Yes, but not quantitative; (iii) Yes: Data showing the selectivity of the biological effect and its specificity.	6	3	x5	15
	Conclusion and Discussion	4	4x3	(i)" No evidence of AS-C–induced cytotoxic effects was found in liver or kidney after a single dose of 500 mg/kg (either i.p. or s.c.)" means very little as you normally cannot treat cancer by one injection only; (ii) although p53-independent effect is conclusive, it is less so about the p53-dependent pathway; (iii) The authors said "it is possible that AS-C might initially cause DNA damage followed by phosphorylation of p53 and induction of p16 expression" – will the DNA damage be a side effect or is it only specific to caner cells?	7	2	x4	8
	Other	Sum	109	Confidence score: 60-80%	Sum			103
9	Title and abstract	7	1x5	Latin binomial name: Yes; part of the plant used: Yes; type of preparation tested:Yes	1	4	x1	4
	Introduction and background	5	3x5	Yes: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented? Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	2	2	x3	6
	Methods –	2	6x4	Yes herbal product name.Yes the Latin binomial name (including	3	3	x6	18

	plant description			the botanical authority).Yes family name for each herbal ingredient.NA the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number.Yes the part of the plant used to make the product or the extract.Yes the processing used to make the test material, Yes type and concentration of the extraction solvent used and the ratio of the herbal drug to extract.Yes the method of authentication of the herbal raw material, including details of any voucher specimen.No whether the test material had been subjected to fingerprinting and by what methods and by whom.No whether any special testing/purity testing had been carried out and by whom.No whether the material had been standardised, and by what process and by whom?				
	Methods – experimental	1	7x4	Yes detailed unequivocal description of the test system.Yes detailed unequivocal description of the experimental protocol.NA description of and justification for the statistical methodology used?	4	3	x7	21
	Objectives	6	2x5	Yes Are the specific objectives and hypothesis and rationale for the selection of the test models used included?	5	1	x2	2
	Results	3	5x5	Yes/No sample size; clearly defined primary and secondary outcome measures. NA statistics.Yes data showing the selectivity of the biological effect and its specificity	6	1	x5	5
	Conclusion and Discussion	4	4x4	Yes appropriate strategy and robustness of the conclusions.Yes interpretation of the result in light of the product tested and dosage regimen used?No Has the conclusion been linked to clinical use?Yes Do the conclusions *support* or *contradict* existing findings?	7	1	x4	4
	Other	Sum	123	It's a biophysics and biochemistry paper with little biology data; confidence score: <60%	Sum			60
10	Title and abstract	7	1x5	Latin binomial name: Yes; part of the plant used: Yes; type of preparation tested:Yes	1	5	x1	5
	Introduction and background	5	3x5	Yes: Are the scientific background and explanation of the rationale including a brief statement of the reasons for the study with reference to the specific herbal medicinal product being tested presented? Yes: Is there a statement reflecting whether the study is new, building on existing knowledge or that traditional indications are being investigated?	2	5	x3	15

	Methods – plant description	2	6x3	Yes herbal product name.Yes but no authority: the Latin binomial name (including the botanical authority) .Yes family name for each herbal ingredient.NA the proprietary product name or the extract name and the name of the manufacturer of the product and its batch number.Yes the part of the plant used to make the product or the extract.Yes the processing used to make the test material, Yes type and concentration of the extraction solvent used and the ratio of the herbal drug to extract.No/yes: no authentication procedure described; there is voucher deposit but voucher number not provided: the method of authentication of the herbal raw material, including details of any voucher specimen.No whether the test material had been subjected to fingerprinting and by what methods and by whom .No whether any special testing/purity testing had been carried out and by whom; No whether the material had been standardised, and by what process and by whom?	3	2	x6	12
	Methods – experimental	1	7x3	Yes detailed unequivocal description of the test system.Yes detailed unequivocal description of the experimental protocol.No description of and justification for the statistical methodology used?	4	3	x7	21
	Objectives	6	2x5	Yes Are the specific objectives and hypothesis and rationale for the selection of the test models used included?	5	1	x2	2
	Results	3	5x3	Yes but experiments only did twice (n=2): sample size; clearly defined primary and secondary outcome measures.No statistics.Yes & No (AR binding is OK but suppression of cell growth may be through AR-independent manner; antagonism to AR needs more evidence) data showing the selectivity of the biological effect and its specificity	6	2	x5	10
	Conclusion and Discussion	4	4x3	Yes but not robust enough about AR antagonism: appropriate strategy and robustness of the conclusions.Yes but not robust enough about AR antagonism: interpretation of the result in light of the product tested and dosage regimen used?Yes Has the conclusion been linked to clinical use?Yes Do the conclusions *support* or *contradict* existing findings?	7	2	x4	8
	Other			It's mainly a biophysics and biochemistry paper with some biology data; confidence score: <60%	Sum			73

Other comments from the application of the second set of criteria

1. In 1, "This point will be scored in this section even when the latin binomial name might be present elsewhere in the paper?", the question mark should be a full stop, if I am right?

2. In 2, I found all the papers got the same maximum score of 4. Do you think that our criteria could indeed differentiate good and the bad in the "Introduction"?

3. In 3a, a unprocessed plant study can earn a score of 8+2= 10, but any processed or proprietary product could earn a maximum of 8+5 =13. Given the relative relevance factor of 4, a paper on unprocessed plant medicine study could earn 3x4 =12 - this is a lot. Do you think that this is fair?

4. In 3b, "2 - is the test system of relevance for TCM studies? (human enzyme, human cells etc...) sounds a bit vague and could be potentially misleading. Do you mean that only human systems get a score? I give a score if only it is a valid model including rat cells, etc.

5. In 3b, "multiple protocols" get a bonus of 3, what does this mean? Does it mean "using different strategies to address the same issue from different perspectives"?

6. In 3b, I do not know if a paper using multiple cell lines to support a phenomenon is applicable to multiple disease areas such as in paper 1 should be given any bonus.

7. In 3b, I do not know why computer models get the highest bonus of 6. Are computerised models more relevant to patients than isolated cells, tissues and organs (bonus of 2-4) and/or supplementation by in vivo tests (bonus of 4)?

8. In 4, "1 - is it clear that n >= 3?" is very misleading. Does this mean number of independent repeats? Many paper write N=x, only to show sample numbers in a single study or in pooled studies.

9. In 4, "2 - is the sample size appropriate?" is difficult to decide. If the variation is low, some times 3-4 per group is sufficient to show the different groups; if the variation is high, even 6-10 is not enough.

10. In some criteria, such as in 5, I feel difficult to give a full score, is it possible to give a half score?