

EuroFIR-BASIS — a combined composition and biological activity database for bioactive compounds in plant-based foods

Jørn Gry^a, Lucinda Black^b,
Folmer D. Eriksen^a,
Kirsten Pilegaard^a,
Jenny Plumb^c, Mike Rhodes^c,
Darina Sheehan^b,
Mairéad Kiely^b and
Paul A. Kroon^{c,*}

^aThe National Food Institute, Technical University of Denmark, Mørkøj Bygade 19, 2860 Soborg, Denmark

^bDepartment of Food and Nutritional Sciences, University College Cork, Ireland

^cThe Institute of Food Research, Colney Lane, Norwich, Norfolk NR4 7UA, UK
(Tel.: +44 (0)1603 255236; fax: +44 (0)1603 507723; e-mail: paul.kroon@bbsrc.ac.uk)

Mounting evidence suggests that certain non-nutrient bioactive compounds promote optimal human health and reduce the risk of chronic disease. An Internet-deployed database, EuroFIR-BASIS, which uniquely combines food composition and biological effects data for plant-based bioactive

compounds, is being developed. The database covers multiple compound classes and 330 major food plants and their edible parts with data sourced from quality-assessed, peer-reviewed literature. The database will be a valuable resource for food regulatory and advisory bodies, risk authorities, epidemiologists and researchers interested in diet and health relationships, and product developers within the food industry.

Introduction

There has been much interest in recent years in the effects of non-nutrient components of the diet in influencing health and well-being. Most of these “bioactive compounds” occur in plant-based foods and for the purposes of this report are defined as “inherent non-nutrient constituents of food plants with anticipated health promoting/beneficial and/or toxic effects when ingested”. Foremost among these non-nutrient bioactive components are the secondary metabolites of plants. It is proposed that the presence of these components can largely explain the epidemiological evidence for the health benefits derived from a diet rich in fruit and vegetables (Walter, 2003). Such diets are correlated with reduced risk of the development of chronic disorders, such as cardiovascular diseases and cancer. However, individual bioactive components can have potential toxicological as well as beneficial effects, and compounds with potential beneficial effects may have negative effects at higher doses. Thus, the intake levels of such compounds, as well as the nature and dose-activity relationships of their biological effects, are of considerable interest.

Plants produce a vast array of secondary metabolites (i.e. those that are not the product of primary metabolism), covering many different compound classes (see Fig. 1 and Table 1). These metabolites may have important functions in plants, mainly in their interaction with the environment and for defence against invading micro-organisms. Within each class, plants tend to produce large numbers of closely related compounds. For example, the number of described flavonoid structures present in nature runs into several thousands (Harborne, 1999).

Plant secondary metabolites have been classified according to their structure and biosynthetic relationships. In most cases, the main groups of secondary metabolites are divided into sub-groups. For example, flavonoids are

* Corresponding author.

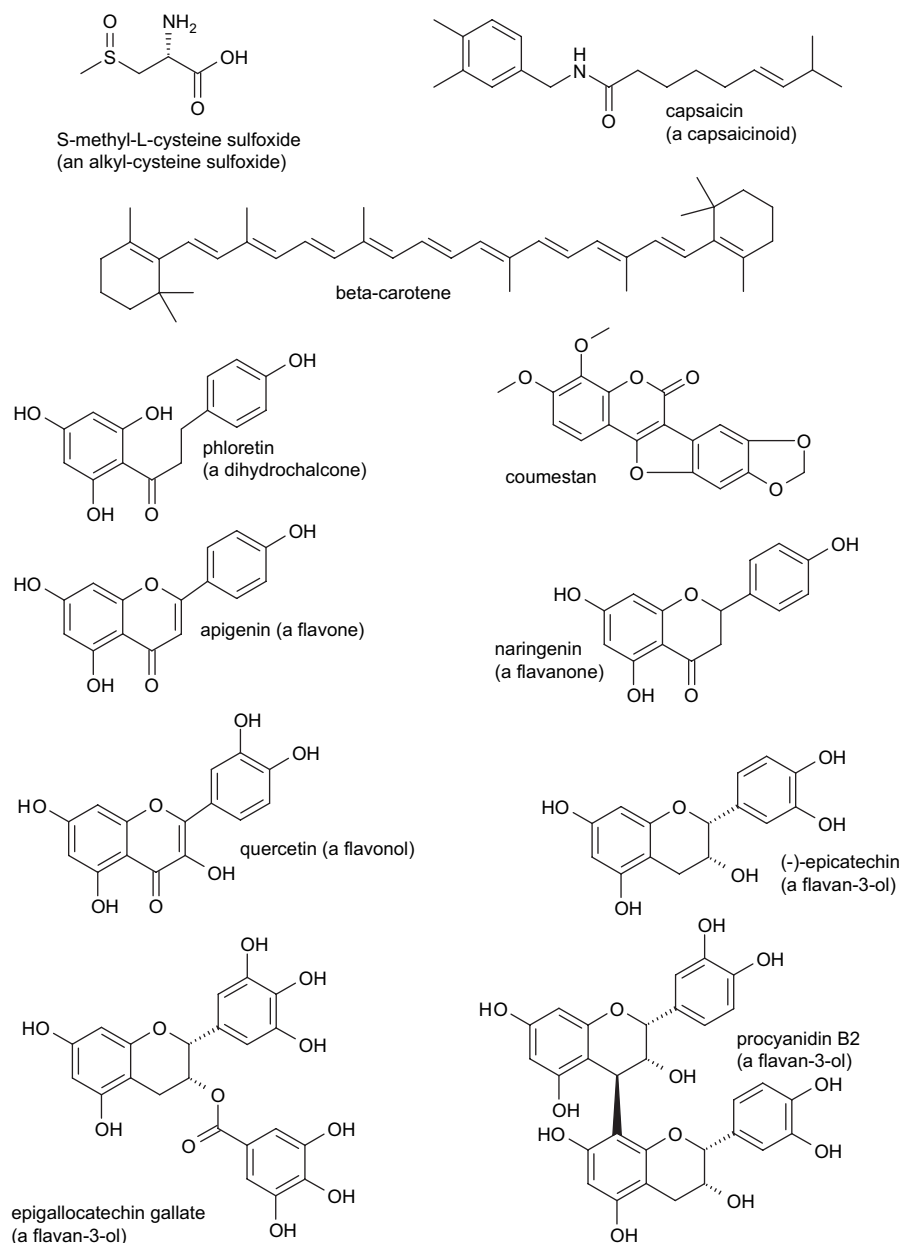


Fig. 1. Example structures for the different classes of bioactives currently included in the EuroFIR-BASIS database.

divided into flavones, flavanones, flavonols, flavan-3-ols, anthocyanidins, isoflavones and other sub-groups. This vast number of related compounds poses problems in evaluating the dietary role of plant bioactives. In the past, some compounds have been classified as either ‘beneficial’ or ‘detrimental’ in terms of impact on human health, but this is rather simplistic since the nature of the biological responses to many bioactive compounds is dependent on the dose. A further complication is that some largely inert plant secondary metabolites are converted to highly active compounds during food processing or after consumption of food. Examples include the essentially inert glucosinolates (found in cruciferous plant species) and

alk(en)yl-cysteine sulfoxides (found in *Allium* species) that are, upon disruption of the plant tissues, converted to highly active isothiocyanates and thiosulfinates/disulfides, respectively.

The EuroFIR-BASIS database, that is the topic of this article, contains compositional and biological effects data for bioactive compounds of plant origin, for which some evidence of a potential beneficial effect has been reported. Other databases have been developed concerned solely with the composition of certain groups of bioactives (e.g. USDA database of critically evaluated flavonoids data; Holden, *et al.*, 2005), and also databases concerned solely with limited biological activities (e.g. antioxidant/radical

Compound class	Examples	Typical plant food sources	Examples of reported biological activities
Alk(en)yl-cysteine sulfoxides	S-methyl-L-cysteine sulfoxide, S-propenyl-L-cysteine sulfoxide	Onions, garlic, leeks	Induction of phase-2 enzymes and apoptosis, arrest of cell cycle
Capsaicinoids	Capsaicin, dihydrocapsaicin	Chilli peppers Sweet peppers	Induction of cellular antioxidant responses, inhibition of LDL-oxidation
Carotenoids	Beta-carotene, lycopene, cryptaxanthine	Tomatoes, carrots, bell peppers	Antioxidant, anti-inflammatory, anti-carcinogenic activities
Coumestans	Coumestan, coumestrol	Soya bean, mung bean	Hepatoprotective, inhibition of Na(+)/K(+)-ATPase
(Dihydro)chalcones	Phloretin	Apples	Inhibition of intestinal glucose uptake
Flavonoids			
Flavones	Apigenin, luteolin	Celery, parsley	Antioxidant, antiproliferative, anti-hypertensive, anti-carcinogenic,
Flavanones	Naringenin, hesperetin	Citrus fruits	anti-thrombotic, cell cycle arrest, induction of phase-2 enzymes,
Flavonols	Quercetin, kaempferol	Onions, tea, green beans, tomatoes	inhibition of phase-1 enzymes, inhibition of LDL-oxidation, improvement of vascular tone
Flavan-3-ols	(+)-Catechin, (-)epicatechin, procyanidin B1, procyanidin B2	Tea, cocoa, apples, berries, certain beans	
Anthocyanidins	Cyanidin, delphinidin, pelargonidin	Blackcurrants, blueberries, strawberries	
Isoflavones	Daidzein, genistein	Soy beans	
Glucosinolates/ isothiocyanates	Glucoraphanin/sulphoraphane	Broccoli, cabbage, Brussel's sprouts	Antiproliferative, cell cycle arrest, induction of phase-2 enzymes, inhibition of carcinogen-induced cancer formation/progression
Lignans	Secoisolariciresinol, matairesinol	Linseed, fruits and vegetables	Estrogenic
Phenolic acids	Ferulic acid, salicylic acid	Coffee, cereal bran, fruits	Anti-inflammatory
Phytosterols	Campestanol, sitosterol	Wheat	Cholesterol lowering
Polyacetylenes	Falcarinol, falcarindiol	Carrots, celery, parsley	Anti-carcinogenic
Stilbenes	Resveratrol, trans-piceid	Grapes, peanuts	Antioxidant, cardio-protective, lifespan extension

critically assessed compositional and 75 toxicological records.

The EuroFIR-BASIS project

Given the mounting interest in the potential for food-based health-beneficial bioactives to provide protection against diseases, the need to gather information on the intake and activity of such compounds, and make it available in a convenient and widely accessible form, was recognised. The information was to be gathered in a new system, EuroFIR-BASIS, availing of the expertise obtained in the development in the EU-NETTOX and -BASIS projects. Compositional and biological information on potential protective bioactive compounds from BASIS, as well as the critical assessment systems and the electronic input forms developed in BASIS, were the starting point for the development of the EuroFIR-BASIS, including the transfer of over 2000 compositional records.

In contrast to previous databases, EuroFIR-BASIS was designed to focus on compounds that showed evidence of beneficial health effects. In addition, EuroFIR-BASIS will include data for processed plant products, such as wine, bread, chocolate and soy products. A further objective was to move from a CD-medium to an Internet-based system, to produce an accessible database that would be convenient and of value to a broad spectrum of users. EuroFIR-BASIS was designed to be compatible with the structure and organisation of the national nutrient composition databases within the overall EuroFIR project. This included using the same database structure, and incorporating and using LanguaL, the international framework for food description (www.LanguaL.org), to ensure that similar search terms could be used to search all systems.

The following classes of plant secondary metabolites were identified for initial inclusion in EuroFIR-BASIS: alk(en)yl-cysteine sulfoxides, capsaicinoids, carotenoids, coumestans, flavonoids, glucosinolates, lignans, phenolic acids, phytosterols, polyacetylenes and stilbenes. These classes of putative health-beneficial compounds were derived from the list of bioactive compound classes of food plants as reported from the BASIS project (Gry *et al.*, 2002), but with some classes removed: classes were removed when the expert evaluators agreed that the literature evidence generally indicated toxic effects rather than beneficial effects. Fig. 1 and Table 1 show examples of typical compounds in each class, with their plant origin and examples of reported biological activities.

Currently, EuroFIR-BASIS allows inputs for approximately 330 major European food plants including their edible parts and a selected group of processed plant products. Following the initial transfer of ~2000 compositional records from BASIS, over 6000 compositional records have been added to EuroFIR-BASIS (i.e. ~8000 records to date). As emerging data for other

classes of bioactive compounds become available, the compound class list will be updated. For each compound class, information on the chemistry, occurrence, distribution in the plant, and factors affecting their levels in food, will be provided. In addition, an objective system for the assessment of the data quality is being designed, which is compatible with the national food composition databases within the EuroFIR project. In subsequent phases of EuroFIR-BASIS, it is hoped to broaden the scope of the database by including a wider range of compounds, plants and processed plant foods. The EuroFIR-BASIS database will provide critically assessed compositional and biological information on bioactive food plant constituents. The database will be very comprehensive in that it will cover a large number of compounds from the major classes of bioactives, and there will be extensive coverage among major European food plants. It is not envisaged that the database will be comprehensive in terms of containing data from every available published report; there are two major reasons for this: (1) systematic searches with keywords cannot identify all the reports containing relevant information and (2) the data inputting process is relatively time-consuming because of the amount of data extracted.

At present a series of report formats are being evaluated to establish useful outputs for users of the database. This completely unique database will be easily accessible *via* the Internet and is compatible with the overall EuroFIR databank system. The database is aimed primarily at national health authorities and international regulatory and advisory bodies (e.g. DG SANCO, European Food Safety Authority (EFSA) and CODEX/JECFA), academics, and to scientists in the food industry.

The database will serve as a useful tool:

- in the evaluation of genetically modified food plants (bioactive food plant constituents are considered to be key substances)
- in relation to evaluation of novel foods
- in relation to evaluation of natural plant compounds used as ingredients/additives in foods
- in relation to consideration of health claims
- in relation to diet and health consideration of food plants.

Development of EuroFIR-BASIS

The EuroFIR bioactives work package, responsible for the development of EuroFIR-BASIS, involves the input of 28 participants from 17 countries, corresponding to approximately 6 person years; 3 person years for software development, database management and administration, and 3 person years for data extraction and critical assessment. Participants are organised into three task groups, dealing with the plant food sources (Plant List Group), the compositional data (Composition

Evaluators Group) and the biological effects data (Biological Effects Group) (see Fig. 2). The work package leader, the task group leaders and the database managers oversee the development of the database. The organisation of the various functions that are working to develop and implement the EuroFIR-BASIS database, their relationships with other EuroFIR functions, are illustrated in Fig. 2.

Development of the food plant lists

The EuroFIR plant list task group (PLG) comprises four compilers and two database consultants. A list of the major food plants consumed in Europe is currently being compiled and will be available online by the end of 2007. Major food plants have been selected, and the correct scientific names, including author names (Brummitt & Powell, 1992), have been identified, alongside high-level taxonomy (Judd, Campbell, Kellogg, Stevens, & Donoghue, 2002; Stevens, 2005) and the relevant plant part used for food (Spjut, 1994). Beyond a major list of food plants, a list of exotic food plants that are used infrequently or by ethnic minority groups in Europe is being prepared. The following

resources have been utilised in developing the food plant lists:

- World Economic Plants (Wiersema & Leon, 1999),
- PROSEA (Plant Resources of South-East Asia) (PROSEA, 1992–2002),
- PROTA (Plant Resources of Tropical Africa) (PROTA, 2002–2006),
- Cornucopia II (Facciola, 1998),
- Mansfeld's World Database of Agricultural and Horticultural Crops (Hanelt & Institute of Plant Genetics and Crop Plant Research, 2001),
- EuroFIR-NETTOX list of food plants (Pilegaard, Eriksen, Soerensen, & Gry, 2006) and
- The New Worlds Fruit database (IPGRI, 2000–2007).

The present EuroFIR-BASIS list contains the same food plants as the NETTOX list produced in 1997 but has (1) been updated and (2) had plant parts added (Pilegaard *et al.*, 2006). In 2008 a new European food plant list will replace the current food plant list from 1997. Lists of exotic food and health food plants are under preparation. These lists, when complete, will be transferred into the EuroFIR-BASIS database, along with colour pictures of the edible part of major food plants and corresponding descriptions.

Development of the compositional database

The EuroFIR composition evaluators task group (CEG) currently comprises 12 specialist evaluators who extract and critically assess compositional data on the selected classes of bioactive compounds (see Table 1). EuroFIR-BASIS compositional information is currently compiled using data extracted from literature published in peer-reviewed journals only. For each compound class, keywords are used to conduct literature searches using three online portals/databases: CabWeb[®] (CAB Partnership, 2003–2005, www.cabweb.net), ISI Web of Science[™] (<http://portal.isiknowledge.com/portal.cgi>) and CAS SciFinder (<http://www.cas.org/SCIFINDER/scicover2.html>). Citations are retrieved from journals covering a broad range of disciplines including agriculture, biology, environmental sciences, medicine, life sciences, nutrition and chemistry. In addition, general taxonomic names for food plants are searched. The abstracts are reviewed by CEG evaluators who have expertise relating to the compound classes to which they are assigned, and articles containing potentially valuable quantitative data on bioactive constituents in human foods are retrieved. All searches are fully documented, together with the total number of references that may contain data suitable for entry into the database.

From retrieved references, the data for each food/compound combination are inputted to the database *via* an online electronic form based on the original BASIS

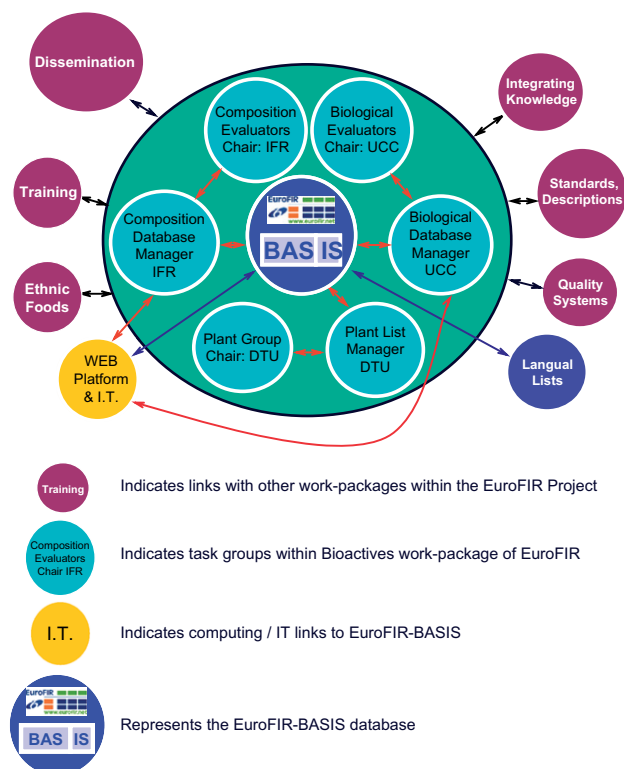


Fig. 2. Organisational structure showing the integration of various components that makes up EuroFIR-BASIS and their interaction with other work packages in the EuroFIR project. The EuroFIR-BASIS work package is led by Dr Jørn Gry of the Danish Technical University, Søborg, Denmark with assistance from Dr Máiréad Kiely (University College Cork, Ireland) and Dr Paul Kroon (Institute of Food Research, Norwich, UK).

input form but further developed and tested within the EuroFIR project. The input form is designed to promote uniformity in the way in which individual evaluators present and evaluate the data. Primary data from the selected publications are inputted to the database and evaluated for quality in six key areas: food description; component description; representativeness of sample (sample plan); sample handling; analytical methodology; and analytical performance. Evaluators enter information extracted from the publication, including information concerning the following:

- (a) bibliographic reference (including authors, title, citation details, brief description of report aims/contents),
- (b) food information (including plant species, plant part, cultivar, maturity, country of origin, growing conditions),
- (c) processing of the plant/plant part/food,
- (d) sampling information (including number and size of collected samples, sample year, sample handling, sample plan, analytical sample size and number of replicates),
- (e) compositional information (including compound and class, source of analytical standard, analytical method applied, reported levels (average level, minimum, maximum, standard deviation or error, and units)).

Where possible all input fields are in the form of pick-lists, using the LanguaL food description system where appropriate, to facilitate uniformity in the way plant foods and compounds are described, both within the database and with the other EuroFIR databases. All compositional data used within the database are linked with full citation information and web links to the original document.

Following the inputting of the primary data, the evaluators complete a section concerned with data quality (the critical evaluation) which is organised as a number of sub-sections. For each subsection, the evaluator is required to select from a choice of 'yes/no', or a set of integer scores ranging from 1 to 5 as described in Table 2. A numerical score (The Critical Evaluation score, CE score) is automatically calculated according to the answers/scores inputted by the evaluator. The possible scores range from 20 to 100, with a score of 20 obtained if all questions are answered 'no' and all the numerical scores are rated '1', whereas a score of 100 is obtained only if all questions

are answered 'yes' and all numerical scores are rated '5' (=highest). Currently, primary reports concerned with bio-active composition in plants/foods that score very highly in all the EuroFIR-BASIS scoring sections are rare — typical limitations include lack of sample plan, incomplete or lack of validation of method of analysis, key details missing for plant description, and insufficient evidence for an unequivocal compound identification. The distribution of CE scores for the data that were extracted from EuroFIR-BASIS following the first 4 months of assessments using this system were as follows:

All inputs are submitted to a EuroFIR-BASIS database manager who inspects all fields in the form and (i) looks for inconsistencies (e.g. inedible plant parts selected, data expressed on a dry weight basis but no mention of drying in analytical methodology, discrepancy between plant common and botanical names) and (ii) ensures that textural fields provide clear information. Subsequently, the data are accepted into the database. All evaluators are fully trained in the use of the form, and regular evaluator assessments are conducted to check evaluator performance and ensure uniformity between evaluators. The evaluator assessment procedure is conducted as follows:

- (i) all compositional evaluators complete input forms for the same three published reports selected by the CEG Database Manager,
- (ii) the inputted data (extracted data and quality scores) are combined in a summary report,
- (iii) the CEG Chair and Database Manager prepare ideal input forms for the three primary references based on their joint assessment of the reference and on the information in the summary report, and this is distributed to all evaluators with additional comments that provide the rationale for each field input of the data extraction and quality scoring sections,
- (iv) each individual evaluation is checked against the 'ideal' evaluation by the CEG Database Manager and Chair, and
- (v) where necessary, individual evaluators are given further training by the Database Manager and are subsequently required to complete further evaluations that are checked by the CEG Chair and Database Manager.

With respect to the data extraction sections, the evaluation process is used to ensure firstly that there is fidelity in data extraction and secondly that evaluators transfer the data to the correct parts of the form. With regard to the second point, all modifiable fields in the input form are linked to automatic pop-up help screens that serve to describe precisely the nature of the information that is required.

The database includes not only data extracted directly from peer-reviewed documents, but also general textural

Quality score range	Frequency (%)
0–39	0.0
40–49	7.9
50–59	16.5
60–69	36.2
70–79	22.0
80–89	10.2
>90	7.1

Table 2. An overview of the EuroFIR-BASIS Critical Evaluation Scoring System

Data quality criterion	Evaluator response	Supporting information for evaluator
Plant/food description (including processing)	Rate 1-2-3-4-5	Are you certain that the plant name or manufactured (pre-packed food) is correct? Are the plant cultivar, season and year of growth and geographic origin included? The varietal type to which the plant belongs (e.g. white or red onion) should have been included in the paper if it is likely to influence the level of component. Was the processing of the plant/food properly described?
Representativeness of sample (sample plan)	Rate 1-2-3-4-5	Is there a sampling plan? Have appropriate sampling procedures been applied? Is this a representative sample for this food plant? Were there a suitable number of samples taken?
Sample handling	Rate 1-2-3-4-5	The evaluator must decide whether handling BEFORE analysis (but after sampling) was appropriate and that it would not affect the level of the component (was sample thawed, cut, peeled, pressed, homogenised, dried, exposed to light or oxygen, etc? water content?)
Component description	Rate 1-2-3-4-5	Has the chemical identity of the phytochemical been correctly established? 5 – There are no doubts about component identity e.g. NMR, FTIR confirmation of structure, evidence from multiple complementary methods (e.g. retention time comparison plus UV/visible spectra plus single ion/disintegration MS plus chemical/enzymatic degradation) 3 – There is some uncertainty in the component identification, e.g. MS used when isomers may be present, retention time with diode array detection only, or the use of a well established method but in a novel setting 1 – The chemical identity is poorly established e.g. by retention time comparison only.
Analytical methodology	Rate 1-2-3-4-5	Are appropriate extraction procedures applied for the components in question? Has the recovery of components been estimated, e.g. through spiking experiments (with the spike added at a level similar to that of the analyte in the food or plant under consideration)? Has an internal standard been used and is it appropriate and added at a suitable level, or has the component been quantified using the standard addition method? Has there been an assessment of the purity of the separated component purity (e.g. comparison of UV and MS quantifications)? Has the linearity of the detector response been demonstrated? Were standard curves of the components used for quantification? Is sample work-up/sample quantification sufficiently reproducible? Purity of standards
Analytical performance	Rate 1-2-3-4-5	Yes: An in-house reference sample was analysed for quality control purposes No: An in-house reference sample was NOT analysed for quality control purposes OR the analysis was not documented

information on the compound classes, including information on most common plant genera/species where the compounds are found, the factors that affect their levels and the breakdown products that may be produced during processing. Structures and synonyms for all compounds in the database are included.

Development of the biological effects database

The aim of the biological effects task group (BEG) is to extract, and critically assess, data from published reports concerned with the biological effects of plant-based bioactive compounds. Eight specialist evaluators, chosen for their knowledge of the biological properties of particular compound classes, extract data from *in vitro*, animal model and human studies. Data on the biological effects of bioactive compounds are deployed in the EuroFIR-BASIS database alongside the composition data.

Reference selection begins with online searching of peer-reviewed publications using PubMed (www.pubmed.gov), ISI Web of Knowledge™ (<http://portal.isi-knowledge.com/portal.cgi>) and CAS SciFinder (<http://www.cas.org/SCIFINDER/scicover2.html>). Search terms include the compound name, compound class or food plant, with all searches fully documented, including search term, search limits and the number of references found. The database manager identifies references that examine the biological effects of compounds and food plants included in EuroFIR-BASIS and abstracts are sent to the appropriate evaluator, with consideration of their area of expertise. The evaluator then prioritises references for inclusion in EuroFIR-BASIS. Papers are selected in reverse chronological order and, except in the case of carcinogenic effects where *in vitro* and animal studies are more prevalent, human studies are prioritised.

Target biological systems and pathologies include cardiovascular health, obesity, metabolic health, type 2 diabetes, cancer and bone health. References are selected to ensure adequate compound coverage, with emphasis on the prioritised food plants in EuroFIR-BASIS. To ensure coverage of key papers, search results are periodically crosschecked with the review literature and the science citation index.

Critical assessments of references are conducted online, using input forms originally designed in BASIS but further developed, tested and optimised within the EuroFIR project. The input form comprises pick-lists, using the LanguaL food description system where appropriate, and text boxes with a defined number of characters. The pick-lists and LanguaL descriptors are a key component in maintaining the relational aspect of the database and ensuring harmony between the task groups, as well as with the EuroFIR nutrient databank. Evaluators enter information from the literature concerning the following:

- (a) bibliographic reference (including authors, title, citation details, web link to the original document and a brief description of the study)
- (b) plant information (including plant species, plant part and country of origin)
- (c) processing of the plant (including heat treatment, cooking method, treatment applied and preservation method)
- (d) test material (including compound, compound class, source, purity and measured quantity)
- (e) *in vitro* or *in vivo* (animal or human) study information (*in vitro* information includes cell type/line, positive and negative control, treatment concentration and duration, standard assay; *in vivo* information includes species, strain/race, gender, route of administration, experimental design, dose, treatment duration and major parameters studied)
- (f) results (including experimental outcome, effective and non-effective levels and adverse effects)
- (g) biomarkers (including a description of the biomarker studied (e.g. HDL cholesterol) and whether a significant effect was observed or not).

Evaluators also complete a section concerned with data quality which is organised into six components:

1. Study design (including study type, blinding, number of study groups, gender, inclusion and exclusion criteria, placebo/control and duration of treatment)
2. Subject (including description of the subjects, animals or cells)
3. Test material (including identity, quantity, carrier, processing and sourcing)

4. Conduct of study (including compliance, treatment of dropouts and documentation of adverse effects)
5. Methodology (including methods, analytical quality control and statistical treatment of data)
6. Results (including appropriateness, description and conclusions).

The evaluator assigns each component a score from 1 to 5 based on the quality of each component. In an equivalent system to that used for composition data, a numerical score (the Critical Evaluation score, CE score) will be automatically calculated according to the component scores assigned by the evaluator. Once the system is finalised, the CE score will present the user with a clear and objective indication of the strengths and weaknesses of a particular study.

All completed input forms are submitted to a EuroFIR-BASIS database manager who checks for any inconsistencies (e.g. discrepancy between plant common and botanical names, discrepancy between pick list and text box information) and ensures that text boxes provide clear and sufficient information, including explanation of any abbreviations. If necessary, the input form is returned to the evaluator for revision before acceptance into the database.

All evaluators are fully trained in the prioritisation of references, use of the input form and quality assessment of references. Evaluator assessments are conducted on a regular basis to monitor performance and ensure uniformity between evaluators. The evaluator assessment procedure is conducted as follows:

- (i) all BEG evaluators complete input forms for the same three references, selected by the database manager
- (ii) the BEG chair and database manager prepare ideal input forms for the references based on their joint assessment of the paper and the extracted data from the submitted input forms
- (iii) the ideal input form is distributed to all evaluators with additional comments for each data field and quality scoring section
- (iv) each individual evaluation is checked against the 'ideal' evaluation by the BEG database manager and chair
- (v) where necessary, individual evaluators are given further training by the database manager and are subsequently required to complete further evaluations that are checked by the BEG chair and database manager.

All data fields in the input form are linked to a help text that describes the information required in that section and provides an example input. In addition, four example input forms are available for viewing – two *in vitro*, one animal model and one human study – along with the

original article from which data was extracted. Regular meetings, in addition to ongoing email and telephone communication, ensure that evaluators have constant support from the database manager in the development of the database.

EuroFIR-BASIS outputs

The EuroFIR-BASIS database will provide users with a large data set comprising compositional and biological effects data that can be mined in a number of ways, and which can be used to generate reports containing selected categories of data in various formats.

Firstly, the database containing the raw extracted data will be searchable by the user *via* a simple yet highly flexible user-driven process. The user will be able to extract data based on their selections from drop-down lists and/or category 'buttons'. For example, a user interested in the content of quercetin in onions would choose 'quercetin' from the Compound drop-down list, 'onions' or '*Allium cepa* L.' from the Plant Food drop-down list and subsequently extract all the records from the database. The user will also be able to perform a more specific extraction and limit the search to, for example, 'quercetin' in 'onions' grown in the 'UK' (Country of Origin) with a CE score >75. The results of such an extraction would be a table of extracted data generated using the default report format.

The second part of the process concerns the report format, which is also user-driven. Users will be able to select from a wide variety of options concerning the fields they would like to be included in the report. The aforementioned user interested in 'quercetin' in 'onions' would be able to select 'Variety', 'Growing Conditions', 'GMO (yes/no)?', 'Sample Year', 'Method of Analysis', etc... and tailor the output to include their desired additional information. Once the output format has been defined by the user, the report will be generated. At this point, the user will be able to select any individual value from the report and open up the complete associated input form, i.e. providing all the data inputted by the original evaluator including the citation details, hyper-links to the original paper, textural inputs and the field-by-field breakdown of the CE score. The data in the report can be easily exported for display using various spreadsheet, database and word processing software packages.

Users interested in biological effects data will be able to search for and export information in a similar way to composition data. For example, users interested in the biological effects of quercetin would select 'quercetin' from the Compound drop-down list and subsequently extract all records regarding the biological effects of quercetin. Similarly, users interested in studies regarding a specific plant would use the Plant Food drop-down list, and users interested in a specific biological effect would use the Biological Effect drop-down list to obtain all records pertaining to their query. In a more specific

search, a user might be interested in the biological effects of quercetin in relation to HDL cholesterol *in vivo* and would select 'quercetin' from the Compound drop-down list, 'HDL cholesterol' from the Biological Effects drop-down list and check the '*in vivo*' box.

As with composition data, reports from the biological effects data are user-driven. A user would select a variable of interest, such as compound or biomarker, and would then select the additional variables they wish to export. Any information field from the input form can be exported including citation information, compound information, plant information, experimental design, dose, treatment duration, results, biomarkers and CE score. For example, a user may wish to export the biological effects data for quercetin including the following variables: compound, plant source, experimental design, dose, biomarker, significant effect, CE score and EuroFIR-BASIS reference code. All chosen data can be exported as tables into spreadsheet, database or word processing software packages and for further information on any particular study, a user can access the individual input form by using a hyper-link from the exported reference code.

Associated activities

Training of evaluators, both existing and new, will be continuous throughout the project to ensure consistency between evaluators' assessments and compatibility with the quality framework developed within EuroFIR-BASIS. One of the core aims is to enable future compilers to continue using the same methods and procedures that have been established during the development of EuroFIR-BASIS. For this reason, there is documentation outlining all Standard Operating Procedures used in the search and capture of literature on the compositional and biological effects of bioactives, including interaction between the database managers and evaluators. Documentation for the basis of selection of major European food plants, along with preferred scientific names, will also be included. In addition, all quality assurance procedures are being documented, including guidance used by expert evaluators during data evaluations, data entry and quality assessment. All of the procedures and processes will be compiled to form a Compilers' and Users' Manual for the EuroFIR-BASIS database.

Uses and users of EuroFIR-BASIS

The EuroFIR project includes a Users Advisory Group (UAG) comprising experts in bioactive constituents. This group provides valuable feedback to the EuroFIR-BASIS team regarding the requirements of potential users of the database. It is envisaged that one of the primary users of EuroFIR-BASIS will be the regulatory affairs sector. For example, uses may include the assessment of genetically modified plants and novel food plants, the investigation of health claims, and risk assessment. The database will also serve the food industry, for example, during the evaluation and development of novel foods. For scientists and

epidemiologists, the database will be a valuable resource for investigating food and health relationships, and for assessing exposure to bioactive constituents in human populations.

Conclusions

The role of bioactive compounds in health is of increasing interest to both the scientific community and the food industry. EuroFIR-BASIS is being developed as a valuable resource for advancement of research and to facilitate regulatory and survey type activities in the area of bioactive compounds, and will provide data to a wide range of users. The online-deployment of the database ensures that it is an accessible resource, and the inclusion of a broad range of compound classes, food plants and processed plant foods ensures that it can be applied widely. As research in bioactive compounds progresses, future development of EuroFIR-BASIS may see the expansion of the database to include additional compound classes and compounds from non-plant foods.

Acknowledgements

The authors thank Paul Needs (IFR) for providing the structures of the bioactive compounds. This work was completed on behalf of the EuroFIR Consortium and funded under the EU FP6 'Food Quality and Safety Programme' (Contract n° FP6-513944).

References

- Brummitt, R. K., & Powell, C. E. (1992). *Authors of plant names*. Kew: Royal Botanic Gardens.
- Facciola, S. (1998). *Cornucopia II: A source book of edible plants*. Vista, CA, USA: Kampong Publications.
- Gry, J., Holm, S., Morgan, M., Møller, A., Preece, R., Speijers, G., et al. (2002). Bioactive Substances in Food Plant Information System (BASIS 1999–2001) Final Report (1999–2001). Executive Summary, Annexes – BASIS Database, Reports, 3.07 – Compound Classes. <<http://www.foodcomp.dk/basis/>>
- Gry, J., Jørgen, W., Kovatsis, A., Møller, A., Rhodes, M., Rosa, E., et al. (1998). *Guide to resources generated by the EU-AIR NETTOX Project (1995–1997)*. Søborg: Danish Veterinary and Food Administration.
- Hanelt, P., & Institute of Plant Genetics and Crop Plant Research. (Eds.). (2001). *Mansfeld's encyclopaedia of agricultural and horticultural crops*. Springer Verlag. © of the electronic version: IPK Gatersleben. <<http://mansfeld.ipk-gatersleben.de/Mansfeld/>>
- Harborne, J. B. (1999, January). In: J. B. Harborne (Ed.), *The handbook of natural flavonoids*. John Wiley & Sons Inc.
- Holden, J. M., Bhagwat, S. A., Haytowitz, D. B., Gebhardt, S. E., Dwyer, J. T., Peterson, J., et al. (2005). Development of a database of critically evaluated flavonoids data: application of USDA's data quality evaluation system. *Journal of Food Composition Analysis*, 18, 829–844.
- IPGRI (International Plant Genetic Resources Institute) (2000–2007). New world fruit database. <<http://www.ipgri.cgiar.org/Regions/Americas/programmes/TropicalFruits/default.asp>>
- Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., & Donoghue, M. J. (2002). *Plant systematics: A phylogenetic approach*, (2nd ed.). Sunderland, MA, USA: Sinauer Associates.
- Pellegrini, N., Serafini, M., Salvatore, S., Del Rio, D., Bianchi, M., & Brighenti, F. (2006). Total antioxidant capacity of spices, dried fruits, nuts, pulses, cereals and sweets consumed in Italy assessed by three different in vitro assays. *Molecular Nutrition and Food Research*, 50, 1030–1038.
- Pilegaard, K., Eriksen, F. D., Soerensen, H., & Gry, J. (2006). EuroFIR-NETTOX list of food plants. Online edition. <www.EuroFIR.net>
- Spjut, R. W. (1994). A systematic treatment of fruit types. *Memoirs of the New York Botanical Garden*, 70, 1–182.
- Stevens, P. F. (2005, May). Angiosperm phylogeny website. Version 6. <<http://www.mobot.org/MOBOT/research/APweb/>>
- Walter, P. (2003). Non-nutritive bioactive food constituents of plants importance for nutrition and health. *International Journal for Vitamin and Nutrition Research*, 73(2).
- Wiersema, J. H., & Leon, B. (1999). *World economic plants: A standard reference*. Boca Raton, Florida, USA: CRC Press.