Joint INFOODS/EuroFIR workshop

Component Identification

Discussion slides prepared by
Ian Unwin (IDUFIC/EuroFIR AISBL)

9th International Food Data Conference

NBI Conference Centre, Norwich, UK

17th September 2011

Previous meeting

Previous decisions:

1. INFOODS will continue to use a dash – for unknown or unspecified or mixed method of determination while EuroFIR will use the three notations U, UN or UNK.
2. In INFOODS, conversion factors are regarded as components and always start with X.
3. EuroFIR uses A Method Indicator (e.g. "Niacin equivalents calculated from niacin and tryptophan" [MI0421]) indicating how to calculate the niacin equivalent. INFOODS treats them as component identifiers with metric units and where all vitamin equivalents are the sums of the contributing components which are adjusted for their relative vitamin activity by specific conversion factors (e.g. NIAEQ = NIA + TRP/60).
4. agreement to add mineral with valence states, e.g. AS3+
5. INFOODS agreed to add version numbers to their future releases of component identifiers and to add a field with the corresponding EuroFIR component names.
6. Agreement to add only those components which are used in food composition and/or are determined separately for food composition purposes. No firm agreement was reached on additives (e.g. cyclamate)
7. It was suggested to add '_LABEL' for all conversion factors and components in the database for labelling purposes. However, this requires further discussions.
8. Different component names were discussed and harmonization reached in most cases

INFOODS has worked on a set of new tagnames but has not released them yet (to be done until the end of 2011)

Background

Data levels

Based on Greenfield & Southgate, Eurofoods recommendations and EuroFIR modifications relating to databases

Level 1

- Data sources
  - documents reporting food composition data

Level 2

- Reported data
  - compositional data extracted from data sources

Level 3

- Compiled data
  - representative data, scrutinised and possibly aggregated

Level 4

- Derived data
  - calculated (e.g. from recipe) or otherwise adapted data

General issues

Scope – component types

- Analysed nutrients
  - Well-defined structures ↔ Rational methods
  - Ill-defined structures and mixtures ↔ Empirical methods
- Bioactives ↔ Analysed
- Ingredients, such as: ↔ Recipe calculation
  - Additives and fortification
  - Nutrient substitutes e.g. salatrims
  - Mixtures defined by specification e.g. maltodextrins
- Properties
  - Food properties
  - Nutritional measures

- Extra types for INFOODS tagnames
  - Method variants
  - Calculation factors

General issues

Requirements – 1

- Food component identification
  - in data interchange
  - in database management systems
  - ? to data users
  - in defining calculation formulae
- Canonical
  - Unambiguous ↔ refers to only one possible substance
  - Unique ↔ substance has only one possible identifier
- Compatible with interchange specification
  - INFOODS tagnames ↔ SGML data element tagnames
  - EuroFIR identifiers ↔ XML data element content
- Straightforward interpretation
  - when defined
  - when assigned
  - when viewed

General issues

Requirements – 2

- Alphabetical listing
  - by code (i.e. tagname or identifier)
  - by component descriptor
- Other listings
  - by component group
  - importance/priority
  - ?
- Documentation (to be discussed later)
- Other requirements?

General issues

Coverage policies

- As follow from Scope and Requirements
- Additives and ingredients
- Intrinsic and synthetic forms
- Bioactives
- Any others?

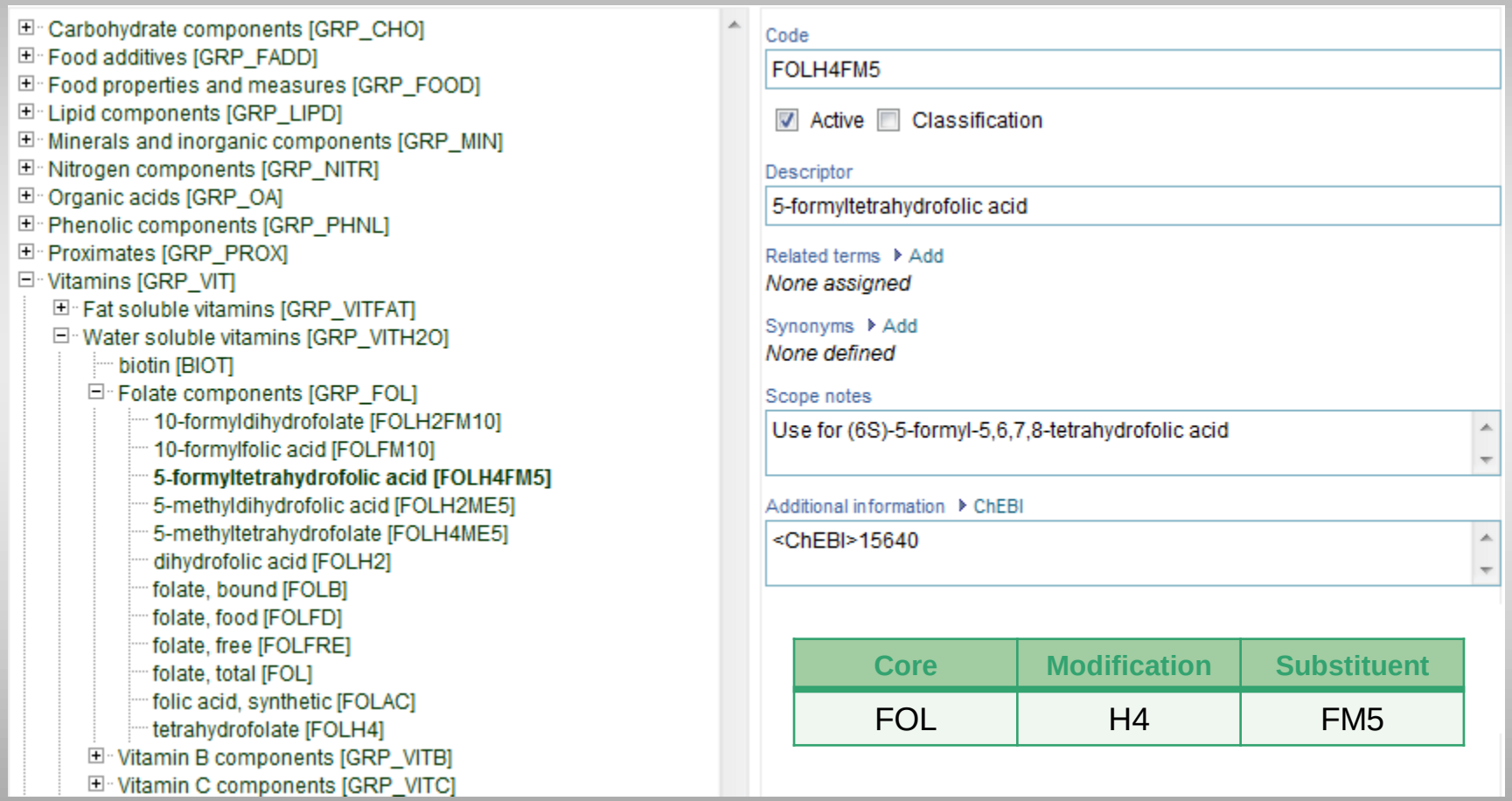
General issues

Code definition principles

- Character set
  - Uppercase alphabetics plus numbers
  - Hyphen (-) and underscore (_)
  - Colon (:) – available in EuroFIR identifiers but not tagnames
  - ? plus sign (+); open and close parentheses (); ....
- Structured codes
- Inverted format
  - Core structure
  - Modifications to core structure
  - Substitutions and isomers
- Stereoparents
- Consistent forms
  - xxxS for sugars, e.g. SUCS for sucrose

Code definition principles

Inverted format



Core	Modification	Substituent
FOL	H4	FM5

Code definition principles

Stereoparents

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY
and
INTERNATIONAL UNION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN)

Nomenclature of Retinoids

Recommendations 1981

http://www.chem.qmul.ac.uk/iupac/misc/ret.html

World Wide Web version prepared by G. P. Moss

Department of Chemistry, Queen Mary University of London,
Mile End Road, London, E1 4NS, UK
e-mail g.p.moss@qmul.ac.uk

These recommendations are as close as possible to the printed version prepared for publication by G.P. Moss [see *Arch. Biochem. Biophys.*, 1983, **224**, 728-731; *Eur. J. Biochem.*, 1982, **129**, 1-5; *J. Biol. Chem.*, 1983, **258**, 5329-5333; *Pure Appl. Chem.*, 1983, **55**, 721-726; *Biochemical Nomenclature and Related Documents*, 2nd edition, Portland Press, 1992, pages 247-251; copyright IUPAC and IUBMB; reproduced with the permission of IUPAC and IUBMB]. If you need to cite these rules please quote these references as their source. In the web version footnotes have been converted into notes following the paragraph to which they apply. In setting up the World Wide Web version an error was detected and the appropriate correction has been made. The change has been marked by which is a link to details of the change and where it applies. A PDF (128 kB) of the printed version is available.

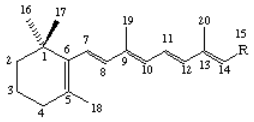
Code definition principles

Stereoparent – retinoids

Ret-2. Stereoparents

A stereoparent is a parent compound whose name implies stereochemistry, which will not need to be stated explicitly.

It is convenient to omit the explicit representation of C and H atoms in the skeletal formulae of retinoids as follows:



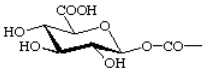
(**1**) R = CH_2OH (**6**) R = CH_2NH_2

(**2**) R = CHO (**7**) R = CH=NOH

(**3**) R = CO_2H (**8**) R = $CH=N[CH_2]_4CHNH_2CO_2H$

(**4**) R = CH_3 (**9**) R = $CO_2C_2H_5$

(**5**) R = CH_2OCOCH_3 (**10**) R- =



Ret-2.1. Retinol

The compound (**1**) (2*E*,4*E*,6*E*,8*E*)-3,7-dimethyl-9-(2,6,6-trimethylcyclohex-1-en-1-yl)nona-2,4,6,8-tetraen-1-ol (see note 1), also known as vitamin A, vitamin A alcohol, vitamin A_1, vitamin A_1 alcohol, axerophthol or axerol, should be designated retinol (see note 2).

Note 1 The numbering system of the systematic name is different from above. See Ret-3.

Note 2 WHO-approved nonproprietary names.

Ret-2.2. Retinal

The compound (**2**) also known as vitamin A aldehyde, vitamin A_1 aldehyde, retinene or retinene$_1$ should be designated retinal (see note 1) or, if liable to be confused with the adjective retinal (pertaining to the retina), retinaldehyde (see note 2).

Note 1 Recommended for chemical usage.

Note 2 Recommended for nutritional usage [ref 16].

Ret-2.3. Retinoic Acid

The compound (**3**) also known as tretinoin (see note), vitamin A acid or vitamin A_1 acid should be designated retinoic acid.

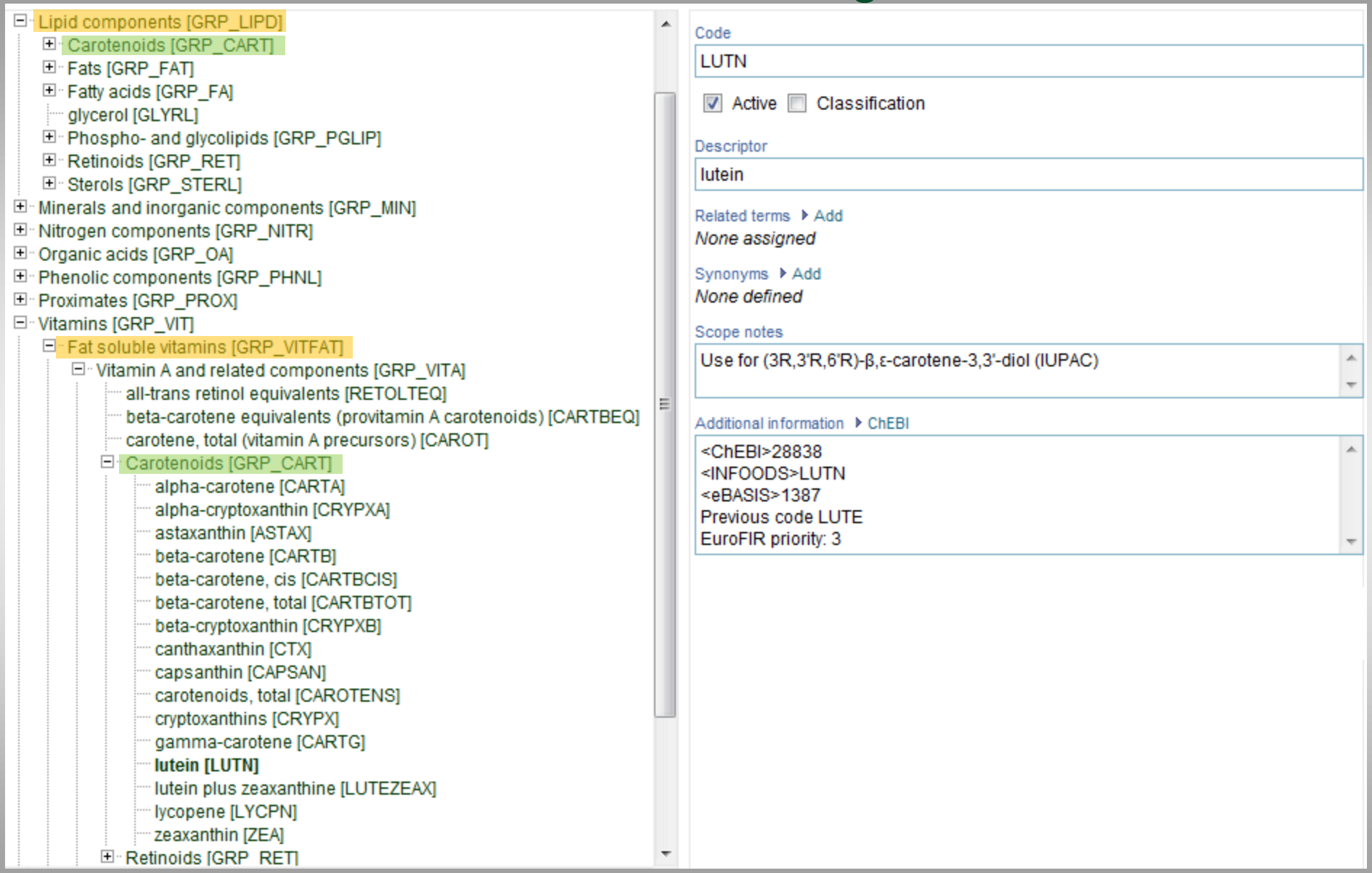
General issues

Component grouping

- Purposes
  - bring together related components
  - simplify entry into list of components (expand or collapse group headings)
  - group term (broad term) helps identify the specific component

- Types
  - Classification ↔ single hierarchy
  - Polyhierarchical ↔ component term can have more than one parent

Thesaurus Manager



INFOODS component grouping proposal – 1

Annex 1

Draft INFOODS proposition of components

group	Description
01	General description of food
02	Macronutrients including energy
021	Energy
022	Protein
0221	Amino acids: individual and aggregations
0222	Nitrogen and protein expressions, conversion factors
0223	Protein components
0224	Other nitrogen-containing components
0225	Protein quality
023	Fat, fatty acids, fatty acid conversion factor
0231	Fatty acids
0232	Fat components
0233	Phospholipids
0234	Fat quality
024	Carbohydrates, carbohydrate fractions
0241	Sugars
02411	Monosaccharides
02412	Disaccharides
0242	Oligosaccharides
0243	Polysaccharides
02431	Dietary fibre, fibre, dietary fibre fractions
024311	Dietary fibre fractions
025	Water
026	Alcohols
027	Ash and other solids
028	Polyols
029	Organic acids
03	Minerals
033	Heavy metals/ contaminants
034	Other minerals

INFOODS component grouping proposal – 2

04	Vitamins
041	Fat soluble vitamins
0411	Vitamin A, retinol
0412	Carotenoids
0413	Vitamin D
0414	Vitamin E
0415	Vitamin K
042	Water soluble vitamins
0421	Thiamin (vitamin B1)
0422	Riboflavin (vitamin B2)
0423	Folate
0424	Niacin
0425	Pantothenic acid
0426	Vitamin B6
0427	Vitamin B12
0428	Biotin
0429	Vitamin C
05	Sterols
051	Plant sterols
052	Cholesterol
06	Bioactive compounds
061	Flavonoids
0611	Flavanols/flavans
0612	Flavonones
0613	Flavones
0614	Flavonols
0615	Isoflavonoids/isoflavones
062	Tannins
063	Phenolic acids
064	Other bioactive compounds
065	Antioxidant activities
07	Biogenic amines
08	Purines
09	Additives
10	Choline and derivates
11	?? Not knowing where to classify
12	Miscellaneous

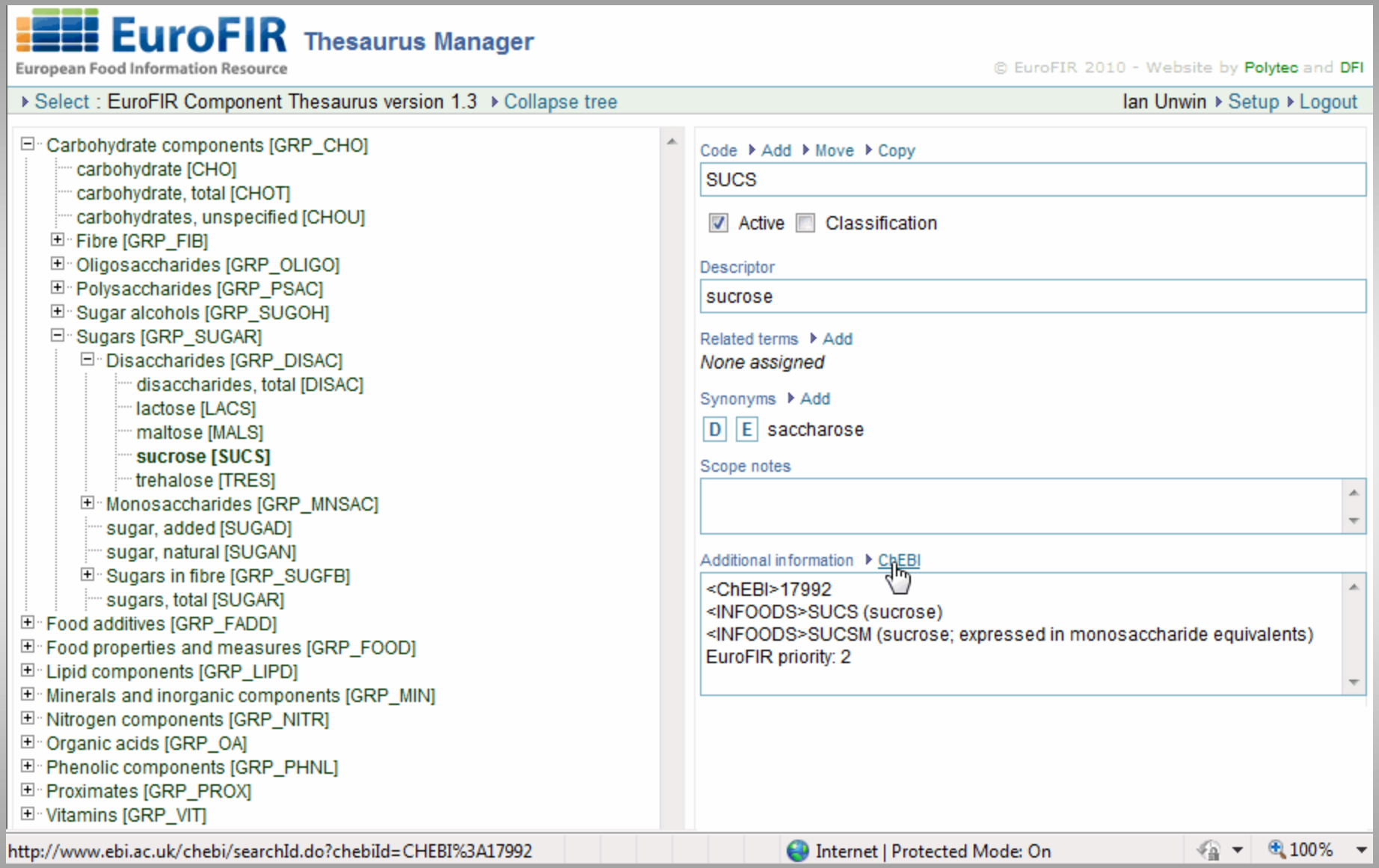
General issues

Documentation

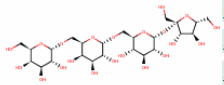
Based on EuroFIR component identifiers managed in Thesaurus Manager

- Overview webpage
  - http://www.eurofir.eu/eurofir_aisbl/products/eurofir_thesauri/component_thesaurus
  - links to information on viewing Component Thesaurus
  - links to discussion forum and topics
- Documentation in Thesaurus Manager
  - Code ↔ Descriptor
  - scope notes
  - synonyms
  - INFOODS tagname ↔ EuroFIR identifier mapping
  - links to external information, e.g. ChEBI (www.ebi.ac.uk/chebi/)
    - structure diagram
    - IUPAC and further nomenclature
    - chemical notation
    - registry numbers (Chemical Abstracts Service and others)
    - onward links to other resources
    - citations
    - ontology (hierarchical classification + chemical relationships)

Thesaurus Manager



Chemical Entities of Biological Interest (ChEBI)



ChEBI Name	**stachyose**
ChEBI ID	**CHEBI:17164**
Definition	An *O*-glycosylglycoside consisting of sucrose having an α-D-galactosyl-(1→6)-α-D-galactosyl moiety attached at the 6-position of the glucose.
Stars	★★★ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:26749, CHEBI:9248, CHEBI:15105, CHEBI:1148754

See structure as: ☑ Image ☐ Applet

Download Molfile

- Find compounds which contain this structure
- Find compounds which resemble this structure

InChI	InChI=1S/C24H42O21/c25-1-6-10(28)14(32)17(35)21(41-6)39-3-8-11(29)15(33)18(36)22(42-8)40-4-9-12(30)16(34)19(37)23(43-9)45-24(5-27)20(38)13(31)7(2-26)44-24/h6-23,25-38H,1-5H2/t6-,7-,8-,9-,10+,11+,12-,13-,14+,15+,16+,17-,18-,19-,20+,21+,22+,23-,24+/m1/s1
InChIKey	InChIKey=UQZIYBXSHAGNOE-XNSRJBNMSA-N
SMILES	OC[C@H]1O[C@H](OC[C@H]2O[C@H](OC[C@H]3O[C@H](O[C@]4(CO)O[C@H](CO)[C@@H](O)[C@@H]4O)[C@H](O)[C@@H](O)[C@@H]3O)[C@H](O)[C@@H](O)[C@H]2O)[C@H](O)[C@@H](O)[C@H]1O

Formula	Source
C24H42O21	KEGG COMPOUND

Net Charge: 0

Average Mass: 666.57768

ChEBI Ontology

Tree view

Outgoing
- stachyose (CHEBI:17164) **is a** *O*-glycosylglycoside (CHEBI:35319)
- stachyose (CHEBI:17164) **has functional parent** sucrose (CHEBI:17992)
- stachyose (CHEBI:17164) **is a** tetrasaccharide (CHEBI:50126)

IUPAC Name

β-D-fructofuranosyl α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)-α-D-glucopyranoside

Synonyms	Sources
α-D-Gal*p*-(1→6)-α-D-Gal*p*-(1→6)-α-D-Glc*p*-(1↔2)-β-D-Fru*f*	JCBN
O-alpha-D-galactopyranosyl-(1->6)o-alpha-D-galactopyranosyl-(1->6)O-alpha-D-galactopyranosyl-beta-D-fructofuranoside	ChEBI
Stachyose	KEGG COMPOUND

Database Links	Databases
C01613	KEGG COMPOUND
G00278	KEGG GLYCAN
1148754	ChEMBL

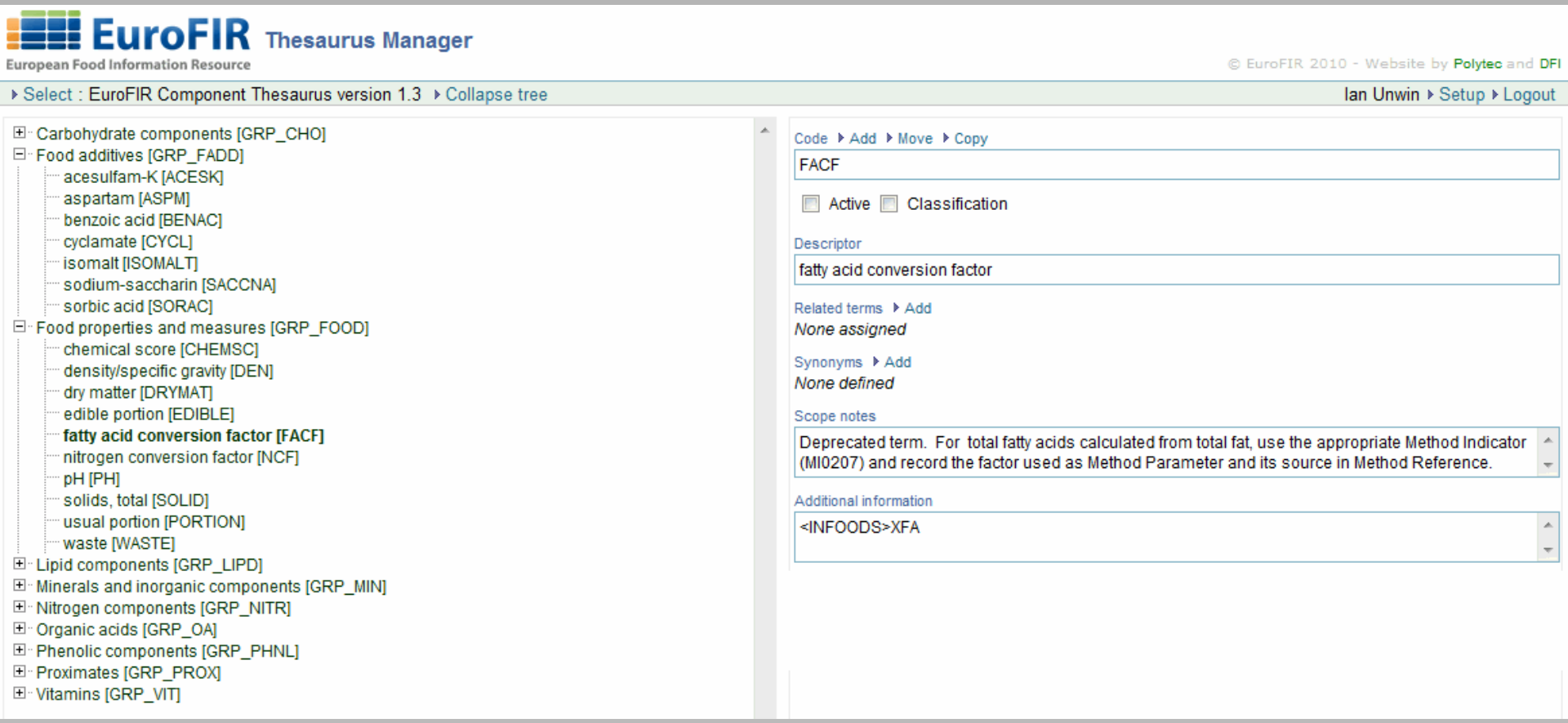
View more database links

Registry Numbers	Types	Sources
10094-58-3	CAS Registry Number	KEGG COMPOUND
1445746	Reaxys Registry Number	Reaxys
1696669	Reaxys Registry Number	Reaxys
470-55-3	CAS Registry Number	KEGG COMPOUND
470-55-3	CAS Registry Number	ChemIDplus
74246	Reaxys Registry Number	Reaxys

Citations

Wagner JR,Becker R,Gumbmann MR,Olson AC (1976)
Hydrogen production in the rat following ingestion of raffinose, stachyose

Thesaurus Manager



General issues

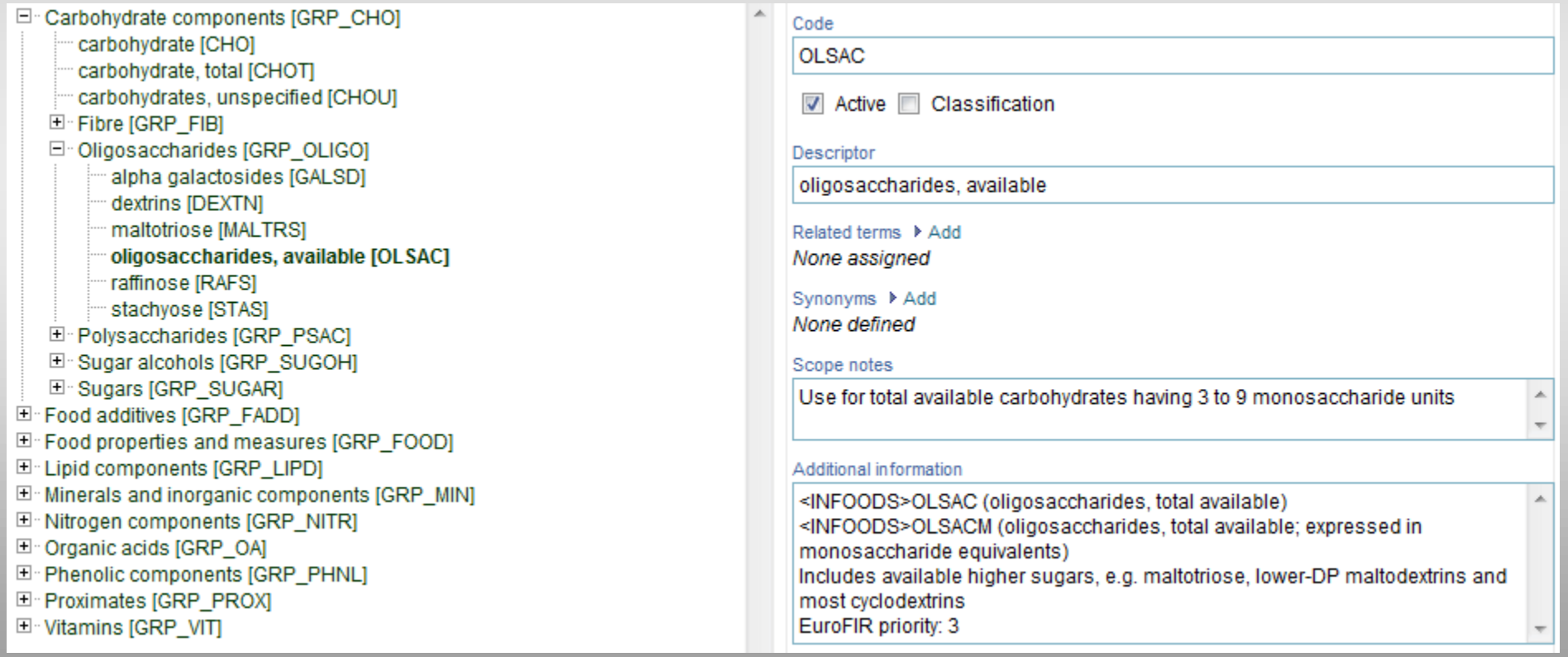
Decision making process

- Timeliness
  - Data management requires the component identifier
    - EuroFIR Data Repository
  - Provisional codes?
- Level of discussion
  - within well-defined area ↔ little discussion needed
  - difficult area ↔ careful discussion needed
- Discussion of proposals
  - INFOODS and EuroFIR jointly or separately
  - Separate proposals or collected proposals
  - Discussion forum
- Decision on proposals
- Implementation of proposals
  - INFOODS tagname list
  - EuroFIR Component Thesaurus
  - harmonised terms and documentation

Carbohydrates

Oligo- and polysaccharide definitions

- oligosaccharides, available [OLSAC]
  - ? oligosaccharide, total
- as ingredients
  - oligo- and polysaccharide mixtures
    - e.g. maltodextrins



Fatty acids

Individual fatty acids – previous A

Tentative agreements reached:

- Both organizations will use only uppercase letters
- For the cis/trans isomers, both organizations will:
  - use C and T (not Z and E);
  - put C for cis or T for trans in front of double bond position counting from the COOH end, if it is known where they occur;
  - put at the end of the component identifier either a single T if at least one double bond is trans or a single C if it is known that all double bonds are cis isomers;
  - identify unknown positions of known cis/trans isomers by putting at the end of component identifier as many C and T as double bonds while putting C first followed by T, e.g. CCT if two cis isomers and one trans exist without knowing their position;
  - for individual fatty acids with unknown cis/trans isomers, both INFOODS and EuroFIR will not add any specific sign, e.g. F16D2 or F16:2, respectively;.
  - for fatty acids with unknown number of double bounds, INFOODS will not add any sign and EuroFIR UN for unidentified, e.g. F16 vs. F16:UN
  - for the sum of polyunsaturated fatty acids (FAPU), it is recommended to create new component identifiers in both systems to indicate cis/trans fatty acids (e.g. FAPUC and FAPUT) and promote their use.
- iso/anteiso: I and AI
- for sums of fatty acids, additional info on which fatty acids are included should be added

Fatty acids

Individual fatty acids – previous B

Agreement still needed:

- radical naming
- natural vs. introduced trans isomers through hydrogenation

Outstanding questions

- why is iso sometimes I and sometimes ISO, e.g. F17:0I and F18:2ISO
- why is trans sometimes TRS and sometimes T, and cis sometimes CIS and sometimes C (new question)

Fatty acids

Individual fatty acids – 1

n-Notation and monounsaturated fatty acids

Initial discussion points:

- Does the use in component codes of the *n*-notation specification (N*x*) for *trans* or substituted fatty acids cause any problems in defining the structure represented?
- Are there any exceptions to the rule that the *n*-notation specification (N*x*) should always be the last part of the component code?
- Are there any reasons for not changing the code for ricinoleic acid from F18:1N9O to F18:1O12CN9 (or the tagname from F18D1N9O to F18D1O12CN9)?

Fatty acids

Individual fatty acids – 2

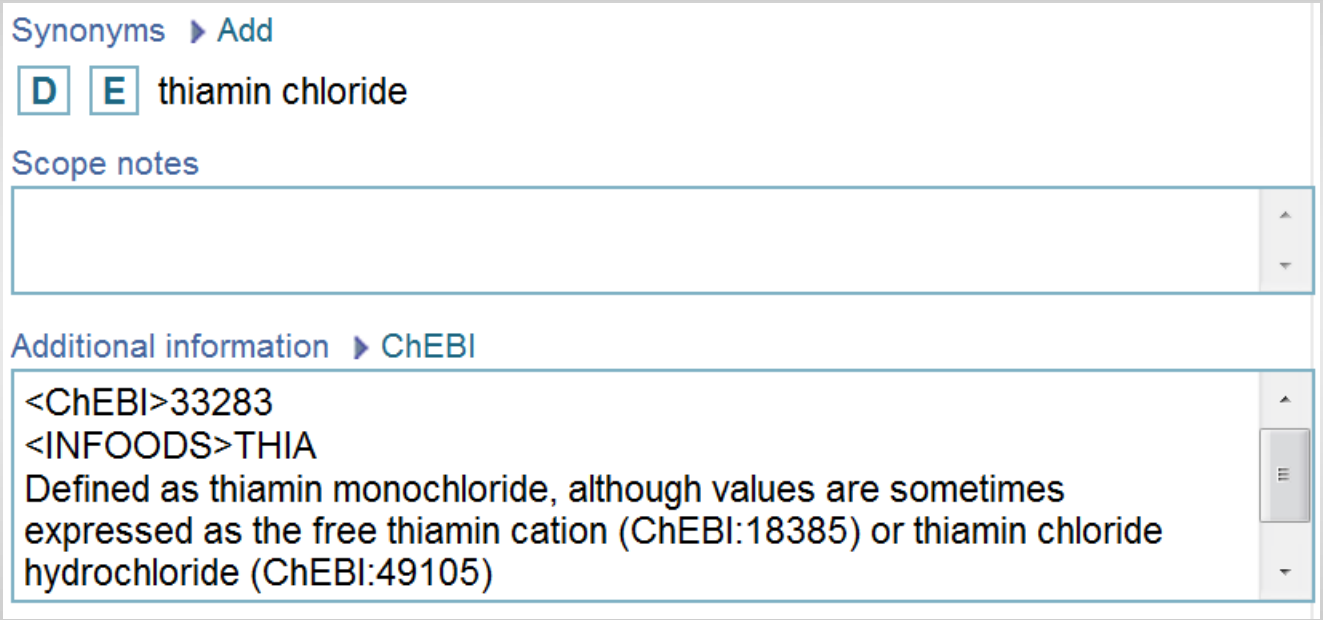
Polyunsaturated fatty acids

Initial discussion points:

- The use of *n*-notation (N*x*) for trans polyunsaturated fatty acids must be limited to the methylene separated isomers, for example with F18:2TTN3 representing 12-*trans*,15-*trans*-octadecadienoic acid. The code for 9-*trans*,15-*trans*-octadecadienoic acid would be F18:2T9T12. Is this policy logical?
- If all positions cited before the final *n*-notation (N*x*) are counted from the carboxylate end, F18:3TCCN3 (for example) would identify 9-*trans*,12-*cis*,15-*cis*-octadecatrienoic acid. Is this an acceptable approach?
- Is the use of short forms (e.g. F18:3CN3) for the all-cis isomers, explicit forms (e.g. F18:3TTTN3) for the all-trans isomers and forms such as F18:3TN3 for total trans isomers the best approach to take?

Vitamins

- Thiamin chloride [THIA]
  - not thiamin(1+)
  - not thiamin chloride hydrochloride

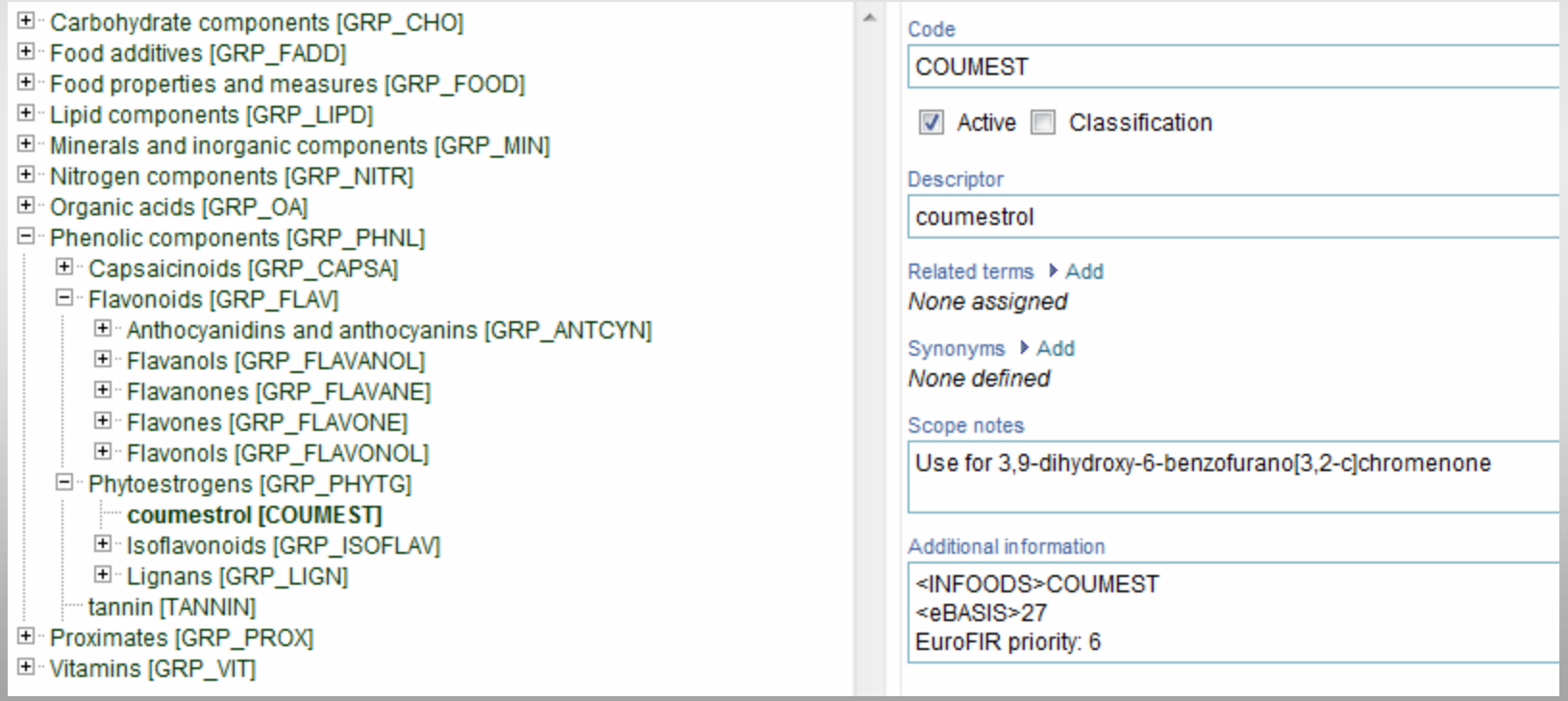


- Folates
  - further aspects to discuss?

Bioactives

Code definition principles – 1

- Codes similar to others, not meaningless identifiers
- Core structure as XX_
  - e.g. Glucosinolates as GS_XXX
- Codes similar to others, not meaningless identifiers
- Interface to existing codes



Bioactives

Code definition principles – 2

- Carbohydrate components [GRP_CHO]
- Food additives [GRP_FADD]
- Food properties and measures [GRP_FOOD]
- Lipid components [GRP_LIPD]
- Minerals and inorganic components [GRP_MIN]
- Nitrogen components [GRP_NITR]
- Organic acids [GRP_OA]
- Phenolic components [GRP_PHNL]
  - Capsaicinoids [GRP_CAPSA]
  - Flavonoids [GRP_FLAV]
    - Anthocyanidins and anthocyanins [GRP_ANTCYN]
    - Flavanols [GRP_FLAVANOL]
    - Flavanones [GRP_FLAVANE]
    - Flavones [GRP_FLAVONE]
    - Flavonols [GRP_FLAVONOL]
      - kaempferol [KAEMF]
      - myricetin [MYRIC]
      - **quercetin [QUERCE]**
      - rutin [RUTIN]
  - Phytoestrogens [GRP_PHYTG]
    - coumestrol [COUMEST]
    - Isoflavonoids [GRP_ISOFLAV]
    - Lignans [GRP_LIGN]
  - tannin [TANNIN]
- Proximates [GRP_PROX]
- Vitamins [GRP_VIT]

Code

QUERCE

☑ Active ☐ Classification

Descriptor

quercetin

Related terms ▸ Add

None assigned

Synonyms ▸ Add

None defined

Scope notes

Use for 3,3',4',5,7-pentahydroxy-2-phenylchromen-4-one

Additional information ▸ ChEBI

<ChEBI>16243
<INFOODS>QUERCE
<eBASIS>1509
EuroFIR priority: 4

1509	Quercetin
1883	Quercetin 3,4'-diglucoside
1886	Quercetin 3-rhamnodiglucoside
1882	Quercetin 4'-glucoside
1511	Quercetin glycosides
1512	Quercetin-3-arabinoside
1517	Quercetin-3-galactoside
1518	Quercetin-3-gluco-7-rhamnoside
1519	Quercetin-3-glucoside
1520	Quercetin-3-glucuronide
1524	Quercetin-3-rhamnodiglucoside
1525	Quercetin-3-rhamnoside
1526	Quercetin-3-rutinoside
1528	Quercimeritrin

Bioactives

Code definition principles – 3

- Locants
  - 2, 3, 4 not *o*, *m*, *p*
  - coding of 3,4'-diglucoside
- Coding as condensation products
  - glucosides

Further issues

- Speciation
  - oxidation state
  - ionic species
  - radicals
- Combination coding
  - mixtures
  - condensation products
- ?