Supporting Information:

In silico prediction and automatic LC-MS^n annotation of green tea metabolites in urine

*Lars Ridder**[1,2], Justin J. J. van der Hooft,[1,3,5,†] Stefan Verhoeven,[2] Ric C.H. de Vos,[3,4,5] Jacques Vervoort,[1,3] Raoul J. Bino[1]*

[1] Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands

[2] Netherlands eScience Center, Science Park 140, 1098 XG Amsterdam, The Netherlands

[3] Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands

[4] Plant Research International, Wageningen University and Research Centre, P.O. Box 16, 6700 AA, Wageningen

[5] Centre for BioSystems and Genomics, P.O. box 98, 6700 AB, Wageningen, The Netherlands.

Table S-1 Reaction rules for intestinal digestion and human phase II biotransformations

Rule	Smirks
Conversions in the gut	
glycosidase	[CH1$(C-&@O):1]-&!@[O:2][#6:3]>>[C:1]O.[O:2][#6:3]
esterase	[C$(C=O):1]-&!@[O:2][#6:3]>>[C:1]O.[O:2][#6:3]
C-ring_fiss_1	[C$([CH1]c2[cH1]ccc[cH1]2):1]1[CH1]([OH1])[CH2]c3c([OH1])[cH1]c([OH1])[cH1]c3O1>>[C:1]C(O1)CCC1=O
lactone_hydrol1	[O:1]=[C:2]1[O:3][CH1$(C[CH2]c2[cH1]ccc[cH1]2):4][CH2:5][CH2:6]1>>[O:1]-[C:2]1=[O:3].[C:4][C:5][C:6]1
lactone_hydrol2	O=C1O[CH1$(C[CH2]c2[cH1]ccc[cH1]2):4][CH2:5][CH2:6]1>>O=C1O.O[C:4][C:5][C:6]1
C-ring_fiss_2	[C$([CH0]c2[cH1]ccc[cH1]2):1]1=C([OH1])C(=O)c3c([OH1])[cH1]c([OH1])[cH1]c3O1>>[C:1]C(=O)O
alpha-ox	[c:1][CH2:2]C(=O)[OH1]>>[c:1][C:2](=O)O
beta-ox	[CH2:4][CH2]C(=O)[OH1]>>[C:4](=O)O
reduction	[C;$(C(=O)[OH1]):1][CH1:2]=[CH1:3][c:4]>>[C:1][C:2]-[C:3][c:4]
p_dehydrox	[c$(c1c[cH1]c([CH2]C)[cH1]c1):1][OH1]>>[c:1]
decarbox	[c$(c1ccccc1):1]C(=O)[OH1]>>[C:1]
Phase II conjugations	**(only rules that applied to the set of tea compounds and their gut metabolites are shown)**
O-gluc_1	[C;!$(C1CCOCC1);!$(C1COCCC1);!$(C(O)=O):1][O:2][H]>>[C:1][O:2]C1OC(C(O)=O)C(O)C(O)C1O
O-gluc_2	[c:1][O:2][H]>>[c:1][O:2]C1OC(C(O)=O)C(O)C(O)C1O
O-gluc_4	[C:1][C;!$(C(O)(=O)C1OCCCC1):2](=[O:3])[O:4][H]>>[C:1][C:2](=[O:3])[O:4]C1OC(C(O)=O)C(O)C(O)C1O
O-gluc_5	[c:1][C:2](=[O:3])[O:4][H]>>[c:1][C:2](=[O:3])[O:4]C1OC(C(O)=O)C(O)C(O)C1O
O-sulf_1	[c:1][O:2][H]>>[c:1][O:2]S(=O)(=O)O
O-sulf_2*	[C;!$(C=O);!$(CC[OH1]):1][O:2][H]>>[C:1][O:2]S(=O)(=O)O
O-methyl	[c:1][O:2][H]>>[c:1][O:2]C
glycination_1	[c:1][C:2](=[O:3])[O][H]>>[c:1][C:2](=[O:3])NCC(=O)O
glycination_2*	[C!$(CN):1][C:2](=[O:3])[O][H]>>[C:1][C:2](=[O:3])NCC(=O)O

* These rules were excluded in the "selected" set described in Table S-2

Table S-2 Effects of the number of reaction steps and selecting a subset of most common phase 2 biotransformation rules on the number of generated compounds and computation times.

The selection of most important rules was based on the criterion that more than 5% of the metabolites they predict for a large and diverse set of clinically studied compounds[43] agreed with the clinical observations. The selected rules included glucuronidation of aromatic as well as aliphatic hydroxyl and carboxyl groups, methylation and sulfation of aromatic hydroxyl groups only, and glycination of aromatic carboxyl groups only.

phase II rules	# reaction steps	# total compounds	computation time (min.) *	#annotations presented in Fig S-1
selected	2	4905	6	
all	2	6086	7	
selected	3	18952	30	
all	3	27245	41	

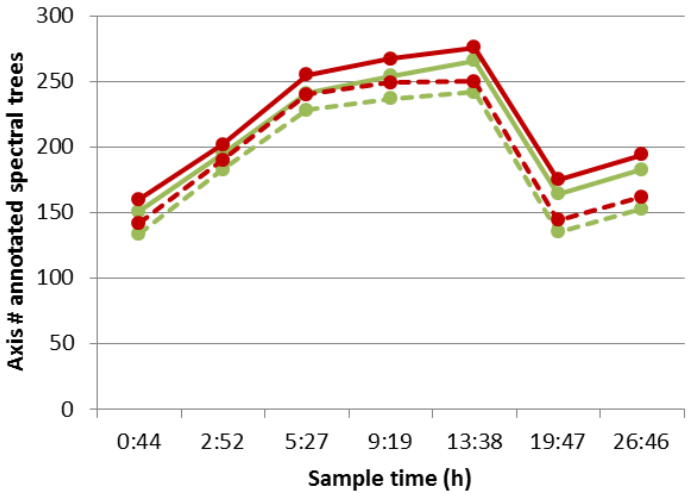
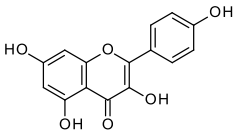
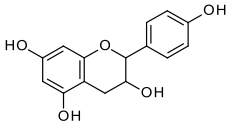
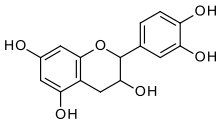
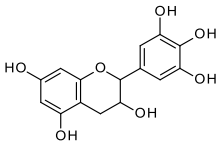
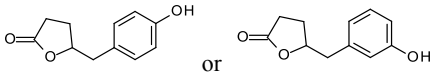
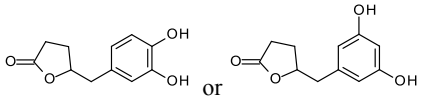
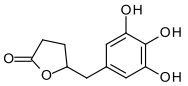
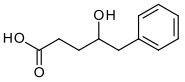
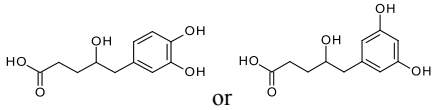
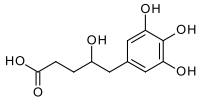
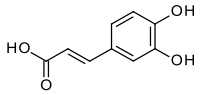
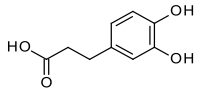
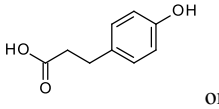
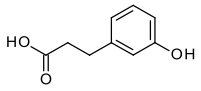
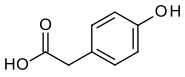
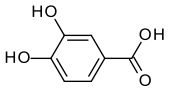
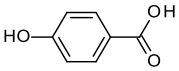
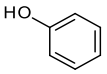
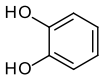
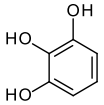


Figure S-1 The number of automatically annotated parent ions in seven consecutive urine sample data sets, after consumption of tea, using four different virtual metabolite libraries. Legend as indicated in Table S-2.

Table S-3 Overview of the metabolites identified in urine. Consecutive columns provide details of the main original components in tea, the intermediate species that may be absorbed from the intestinal tract and the metabolites that were assigned to urine components on the basis of our automatic annotation. Compounds newly identified in the present urine samples are underlined. Asterisks indicate structures that were not present in PubChem (on Jan. 28, 2013)

Tea components	Intermediate after hydrolytic and microbiotic conversion rules	Phase 2 conjugates generated and annotated in urine (# positional isomers)
Kaempferol Kaempferol-3-*O*-glucoside Kaempferol-3-*O*-polyglycoside	 Kaempferol	*O*-glucuronide
(Epi-)afzelechin-3-*O*-gallate		<u>(Epi-)afzelechin-*O*-glucuronide</u> <u>(Epi-)afzelechin-*O*-sulfate</u>
(Epi-)catechin (Epi-)catechin-3-*O*-gallate	 (Epi-)catechin	*O*-sulfate-*O*-methyl (4*)
(Epi-)gallocatechin (Epi-)gallocatechin-3-*O*-gallate	 (Epi-)gallocatechin	*O*-sulfate-*O*-methyl (2*) <u>*O*-glucuronide-*O*-methyl (2*)</u> *O*-sulfate-di-*O*-methyl (2*)
(Epi-)afzelechin-3-*O*-gallate (Epi-)catechin (Epi-)catechin-3-*O*-gallate	 5-(4'-hydroxyphenyl)-γ-valerolactone	*O*-glucuronide(2*) *O*-sulfate*
(Epi-)catechin (Epi-)catechin-3-O-gallate	 5-(dihydroxyphenyl)-γ-valerolactone	<u>non-conjugated</u> *O*-glucuronide (3*) *O*-sulfate (2*) *O*-sulfate-*O*-methyl (2*) *O*-glucuronide-*O*-sulfate (1*) *O*-glucuronide-*O*-methyl (4*)
(Epi-)gallocatechin (Epi-)gallocatechin-3-*O*-gallate	 5-(3',4',5'-trihydroxyphenyl)-γ-valerolactone	*O*-glucuronide (2*) *O*-glucuronide-*O*-methyl (3*) <u>*O*-glucuronide-di-*O*-methyl*</u> *O*-sulfate (2*) *O*-sulfate-*O*-methyl (2*) <u>*O*-sulfate-di-*O*-methyl*</u> *O*-glucuronide-*O*-sulfate*
(Epi-)afzelechin-3-*O*-gallate	 4-hydroxy-5-phenyl-valeric acid	*O*-glucuronide* *O*-sulfate*
(Epi-)catechin (Epi-)catechin-3-*O*-gallate		*O*-glucuronide (2*) *O*-sulfate (2*) *O*-glucuronide-O-methyl (4*) *O*-sulfate-O-methyl (2*)

	4-hydroxy-5-dihydroxyphenyl-valeric acid	
(Epi-)gallocatechin (Epi-)gallocatechin-3-*O*-gallate	 4-hydroxy-5-trihydroxyphenyl-valeric acid	*O*-glucuronide-*O*-methyl (2*)
Chlorogenic acid (Epi-)gallocatechin-3-*O*-caffeate (Epi-)gallocatechin-3-*O*-ferulate	 Caffeic acid	<u>Caffeic acid-*O*-sulfate*</u> Ferulic acid-*O*-glucuronide (2) Ferulic acid-*O*-sulfate (2)
	 3-(3,4-dihydroxyphenyl)propionic acid	*O*-sulfate <u>*O*-methyl-*O*-sulfate</u> <u>(dihydroferulic acid-*O*-sulfate)</u>
(Epi-)afzelechin-3-*O*-gallate (Epi-)catechin (Epi-)catechin-3-*O*-gallate (Epi)gallocatechin-3-*O*-ferulate (Epi-)gallocatechin-3-*O*-p-coumarate	 or  3-(hydroxyphenyl)propionic acid	<u>*O*-methyl-*O*-glucuronide (2*)</u>
Quercetin, kaempferol	 Hydroxyphenylacetic acid	<u>*O*-sulfate</u>
(Epi-)catechin (Epi-)catechin-3-*O*-gallate	 Protocatechuic acid	<u>Non-conjugated</u> Vanillic acid glucuronide (2*) Vanillic acid sulfate
(Epi-)catechin-*O*-p-benzoate (Epi-)gallocatechin-*O*-p-benzoate	 Hydroxybenzoic acid	<u>Non-conjugated</u> Hippuric acid 2-hydroxyhippuric acid 3-hydroxyhippuric acid 4-hydroxyhippuric acid <u>Methoxyhippuric acid</u>
(Epi-)catechin-*O*-p-benzoate (Epi-)gallocatechin-*O*-p-benzoate	 Phenol	*O*-glucuronide *O*-sulfate
	 Benzenediol (e.g. catechol)	*O*-glucuronide *O*-sulfate <u>*O*-methyl-O-sulfate (2x)</u>
Gallic acid Catechin-3-*O*-gallate Gallocatechin-3-*O*-gallate Prodelphinidin-*O*-gallate (2) Gallocatechin-catechin-gallate (3) Procyanidin-*O*-gallate (2) Epiafzelechin-3-*O*-gallate (1) Epicatechin 3,5-di-*O*-gallate (1)	 Benzenetriol (e.g. pyrogallol)	*O*-glucuronide (1*+<u>1</u>) *O*-sulfate (2*) *O*-sulfate-*O*-methyl (5*) <u>*O*-sulfate-di-*O*-methyl*</u> <u>*O*-glucuronide-*O*-methyl (3)</u> <u>*O*-glucuronide-di-*O*-methyl*</u> <u>*O*-glucuronide-*O*-sulfate*</u> <u>*O*-glucuronide-*O*-sulfate-*O*-methyl*</u> <u>di-*O*-glucuronide*</u>

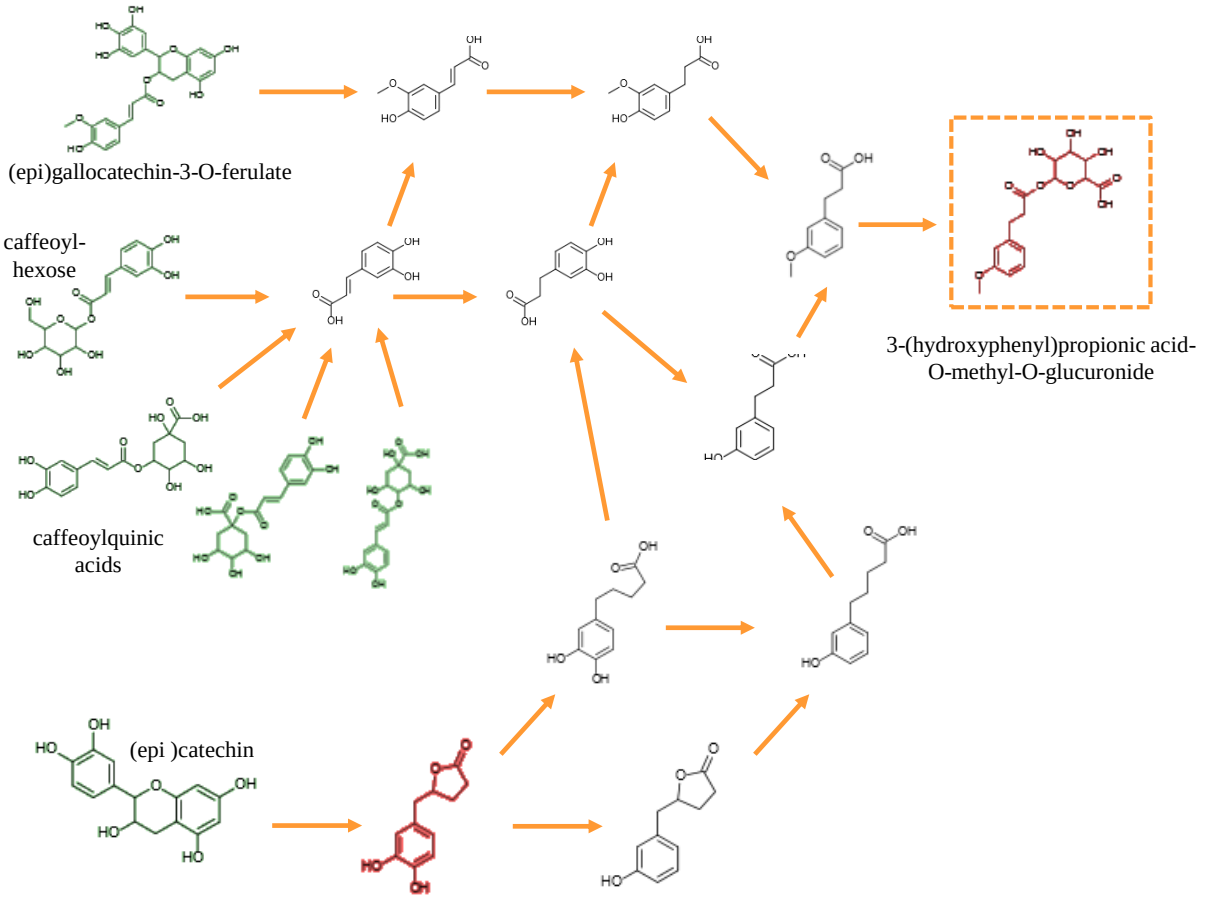


Figure S-2. The *in silico* pathways via which 3-(hydroxyphenyl)propionic acid-*O*-methyl-*O*-glucuronide is generated from (epi)gallocatechin-3-*O*-ferulate, caffeoyl-hexose, caffeoylquinic acids (e.g. chlorogenic acid) as well as (epi-)catechin, corresponding to the highlighted pathways in Figure 5. Green structures represent green tea components and red structures represent metabolites confirmed in the urine LC-MS^n profile at t=9:19 h.