May, **2008**. (The abstract is downloadable from the site http://www.massbank.jp/en/document.html).
[29] C. Hopley, T. Bristow, A. Lubben, A. Simpson, E. Bull, K. Klagkou, J. Herniman, J. Langley. Towards a universal product ion mass spectral library – reproducibility of product ion spectra across eleven different mass spectrometers. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1779.
[30] K. Volná, M. Holcapek, L. Kolárová, K. Lemr, J. Cáslavsky, P. Kacer, J. Poustka, M. Hubálek. Comparison of negative ion electrospray mass spectra measured by seven tandem mass analyzers towards library formation. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 101.
[31] Geeknet, Inc. SourceForge.net. http://sourceforge.net/projects/massbank/. [Last accessed: March 2010].
[32] Institute for Advanced Biosciences, Keio University. Mass++ Manual. http://www.massbank.jp/en/manual.html. [Last accessed: March 2010].
[33] RIKEN, Plant Science Center. Platform for RIKEN Metabolomics MS/MS spectral tag (MS2T) viewer. http://prime.psc.riken.jp/lcms/ms2tview/ms2tview.html. [Last accessed: March 2010].
[34] Creative Commons. Creative Commons Attribution Licenses. http://creativecommons.org/. [Last accessed March 2010].
[35] R. G. Cote, P. Jones, L. Martens, R. Apweiler, H. Hermjakob. The ontology lookup service: more data and better tools for controlled vocabulary queries. *Nucleic Acids Res.* **2008**, *36*, W372.

Copyright © 2010 John Wiley & Sons, Ltd.

NPC Natural Product Communications

2011
Vol. 6
No. 1
7 - 10

Multi-stage Mass Spectrometric Analysis of Saponins in Glycyrrhiza radix

Ken Tanaka[a,*], Kosuke Hayashi[a], Abrar Fahad[b] and Masanori Arita[b]

[a]*Department of Medicinal Resources, Institute of Natural Medicine, University of Toyama, 2630 Sugitani, Toyama 930-0194, Japan*

[b]*RIKEN Plant Science Center, Department of Biophysics and Biochemistry, The University of Tokyo, Bunkyo-ku Hongo 7-3-1, 113-0033 Tokyo, Japan*

ktanaka@inm.u-toyama.ac.jp

Received: October 16th, 2010; Accepted: December 17th, 2010

The fragmentation pathways of six triterpenoid saponins from Glycyrrhiza radix were investigated using LC-MS/MS. Depending on the structure and the substitution pattern, different molecular adduct ions, $[M+NH_4]^+$ or $[M+H]^+$, were observed in the positive ESI spectra. In the positive MS^n spectra from the molecular adduct ions, characteristic product ions corresponding to the loss of dehydrated glucuronic acid or glucuronic acid were detected and they indicated the type of substitution and structural modification. Fragment ions originating from the sapogenin moiety in the positive mass spectra were predominantly provided by saponins having an 11-oxo-12-ene structure. On the other hand, the saponins gave fragment ions corresponding to the sugar moiety in the negative mass spectra. These results indicate the specific property of saponins that have the 11-oxo-12-ene structure to localize positive or negative charge in the mass spectrometric ionization and fragmentation process. Information obtained from the present study can be utilized for structural elucidation of triterpenoid saponins in the Glycyrrhiza radix by LC-MS.

Keywords: Glycyrrhiza radix, LC-MS/MS, electrospray ionization, fragmentation.

Glycyrrhiza Radix (licorice), which is derived from the roots and stolons of *Glycyrrhiza* species, has been used as a crude drug for over 2000 years in many Asian and European countries [1a]. Glycyrrhiza Radix exhibits a variety of pharmacological properties, including antiulcer, anticancer, antioxidative, antiviral, hepatoprotective, antiinflammatory, expectorant and endocrinological activities [1b]. These can be attributed to the chemical constituents, such as the triterpenoid saponins, flavonoids and chalcones [1c]. Triterpenoid saponins are the major chemical constituent of licorice (4–20%), and more than 100 saponins have been identified from *Glycyrrhiza* sp., so far [2].

Recently, LC-MS has been used for the qualitative and quantitative analysis of the chemical constituents of Glycyrrhiza Radix, as well as for the characterization of licorice constituents in prepared Chinese medicines [3a,3b]. However, these studies focused on the LC-MS analysis of flavonoids and glycyrrhizin, the latter a major saponin of licorice, and relatively little has been reported on the systematic interpretation of the MS^n fragmentation of triterpenoid saponins in licorice other than the molecular ions of the new compounds. Mass spectrometry is the most sensitive method for simultaneous detection of hundreds of compounds in a single extract and has been recognized as an important tool for the quantitative evaluation of crude drugs. However, undoubtedly, mass spectrometry is not a perfect methodology for identifying chemically diverse natural products in plant extracts when authentic standards are not available. Therefore, accumulation of mass spectral information of standard compounds is necessary in order to be able to mine the natural products in extracts and for the practical elucidation of the structure of natural products. In the present study, the fragmentation of six saponins was investigated using multi-stage mass spectrometric analysis to provide information that could be utilized in the structural elucidation of triterpenoid saponins in licorice.

The structures of the six saponins are shown in Figure 1. Most saponins in licorice exist as glucuronides. The sapogenins are oleanane type triterpenes with 11-oxo-12-ene (**1**, **2** and **6**), 12-ene (**3**), 11,13(18)-diene (**4**) and 11-oxo-12-ene with a lactone ring between 30-COOH and 22-OH (**5**) as functional groups.

In Figure 2, the positive ESI-MS and MS^2 spectra of **1** are shown. Compound **1** provides $[M+H]^+$ and $[M+NH_4]^+$ ions at *m/z* 823.4073 and 840.4457, respectively. Furthermore, the $[(\text{dehydrated aglycone})+H]^+$ ion is also observed at *m/z* 453.3350. In the MS^2 spectrum, from the $[M+H]^+$ ion, product ions at *m/z* 647.3763 and *m/z* 453.3332 were observed. The loss of 176 amu (from *m/z* 823.4073 to *m/z* 647.3763) and the loss of 370 amu (from *m/z* 823.4073 to *m/z* 453.3332) corresponded to cleavage of the dehydrated

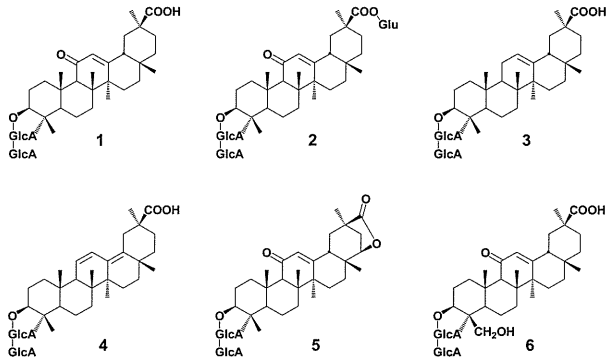


Figure 1: Structures of saponins analyzed in this study.
1: glycyrrhizin, **2**: licoricesaponin A3, **3**: licoricesaponin B2, **4**: licoricesaponin C2, **5**: licoricesaponin E2, **6**: licoricesaponin G2, GlcA: glucuronic acid, Glu: glucose.

glucuronic acid moiety and di-glucuronic acid moiety, respectively. All the product ions detected in the spectrum are derived from the sapogenin moiety, while ions originating from the sugar moiety are not observed. The results indicate that the positive charge of the $[M+H]^+$ ion of **1** was localized at the sapogenin moiety in the process of fragmentation.

In the MS^2 spectrum from the $[(\text{dehydrated aglycone})+H]^+$ ion at *m/z* 453.3350, two significant product ions were detected: at *m/z* 435.3236, attributed to the loss of H_2O, and at *m/z* 407.3304, attributed to the loss of HCOOH. In addition, fragment ions having an odd number of electrons at *m/z* 436.3336 and 408.3408 can also be observed with relatively high abundance. It is well known that ions having an even number of electrons predominantly provide the fragment ions with an even number of electrons by cleavage of the neutral molecule, except for compounds having the specific structure to stabilize the radical cation (even electron rule) [4]. The results indicate that the dehydrated aglycone of **1** has the characteristic property to stabilize radical cations. The negative ESI-MS spectrum and the MS^2 spectra of **1** are shown in Figure 3. The $[M-H]^-$ ion was detected at *m/z* 821.3941. In the MS^2 spectrum from the $[M-H]^-$ ion, the product ion at *m/z* 351.0569 predominates and is attributable to dehydrated di-glucuronic acid. Further fragment ions from the dehydrated di-glucuronic acid ion are annotated, as shown in the MS^3 spectrum (Figure 3). In contrast with the positive charge on the glycyrrhizin molecule, these mass spectra indicate that the negative charge of the $[M-H]^-$ ion of **1** is mainly retained by the sugar moiety in the process of fragmentation.

The positive ESI-MS, MS^2 and MS^3 spectra of **2** and **6** are shown in Figure 4. The compounds provide $[M+H]^+$ ions at *m/z* 985.4692 and 839.4069, respectively, and very weak ammonium adduct and $[(\text{dehydrated sapogenin})+H]^+$ ions are also observed. The respective $[M+H]^+$ ions showed similar fragmentation patterns in the MS^2 spectra and the observed product ions are annotated as shown in Figure 4. Compared with the fragmentation of **1**, **2** provides the ion corresponding to the loss of dehydrated glucose in the first step of fragmentation, and **6** gives relatively intense dehydrated fragments. As with the MS^3 fragmentation of **1**, **2** predominantly shows ions resulting from the cleavage of water and formic acid from major product ions corresponding to the dehydrated sapogenin in the MS^3 spectra. On the other hand, **6** gives several product ions attributable to the loss of water, formic acid and formaldehyde in the MS^3 spectrum. In the MS^3 spectra of both compounds, intense fragment ions with odd numbers of electrons are observed and they correlate with ions observed with an even number of electrons produced by the loss of water and either formic acid or formaldehyde. Compound **5** shows $[M+NH_4]^+$ ions in the ESI-MS spectrum and the fragmentation of the MS^2 spectra from the $[M+NH_4]^+$ ions and $[(\text{dehydrated sapogenin})+H]^+$ ions are similar to those of **1** (Figure 5).

The fragmentation in the negative ESI-MS spectra and the MS^2 spectra of **2**, **5** and **6** are similar to those of **1** and the MS^2 spectra from the respective $[M-H]^-$ ions show the ion due to dehydrated di-glucuronic acid. The positive ESI-MS spectra, and the MS^2 and MS^3 spectra of **3** and **4**, licoricesaponin without the oxygen at C-11, are shown in Figure 6. Unlike **1**, **2**, **5** and **6**, for licoricesaponin having

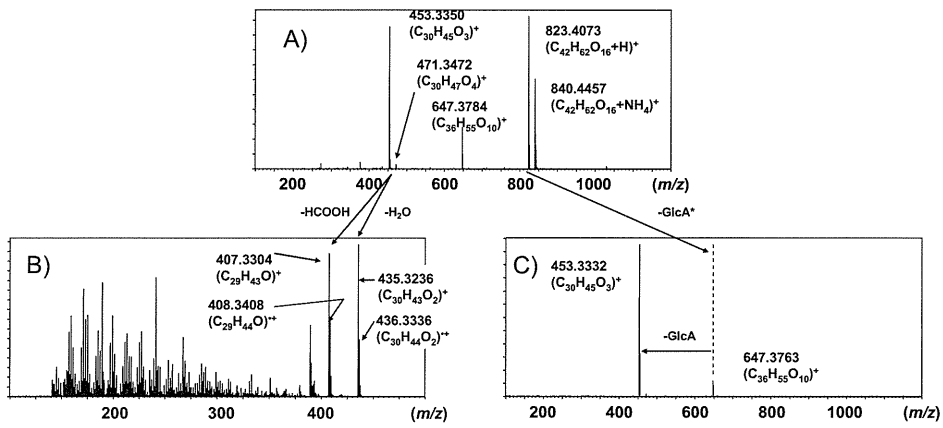


Figure 2: Positive ESI-MS spectrum and MS^2 spectra of **1**.
A)ESI-MS spectrum of **1**, B) MS^2 spectrum from the ion at *m/z* 453.3350 in A), C) MS^2 spectrum from the $[M+H]^+$ ion at *m/z* 823.4073 in A).

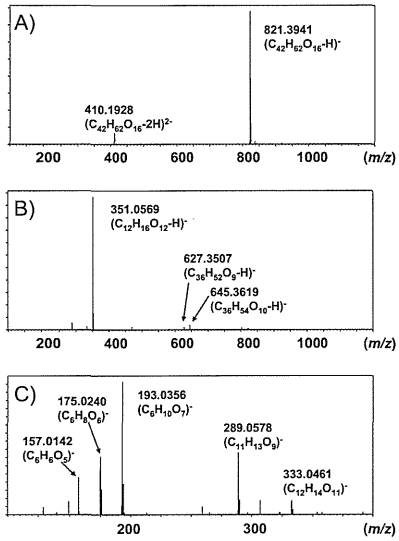


Figure 3: Negative ESI-MS spectrum and MS^2 spectra of **1**.
A) ESI-MS spectrum of **1**, B) MS^2 spectrum from the $[M-H]^-$ ion at *m/z* 821.3941 in A), C) MS^3 spectrum from the ion at *m/z* 351.0569 in B).

the 11-oxo-12-ene structure, fragment ions attributable to the sugar moiety are observed in the MS^2 spectra. The presence of fragment ions corresponding to the sapogenin moiety and the sugar moiety indicate that the positive charges of the compounds were retained by both parts of the saponin structure. These findings show that the 11-oxo-12-ene structure plays a specific role in retaining the positive charge in the processes of ionization and fragmentation of saponins derived from licorice.

In the present study, we investigated the fragmentation pathways of triterpenoid saponins in Glycyrrhiza radix using a LC-MS/MS method. Depending on the structure and the substitution pattern, each saponin gave the prominent molecular adduct ion distinctively. The differences in the fragmentation pathways from the molecular adduct ions should indicate the type of structural modifications, which can be utilized for elucidating the structures of triterpenoid saponins in Glycyrrhiza radix by LC-MS.

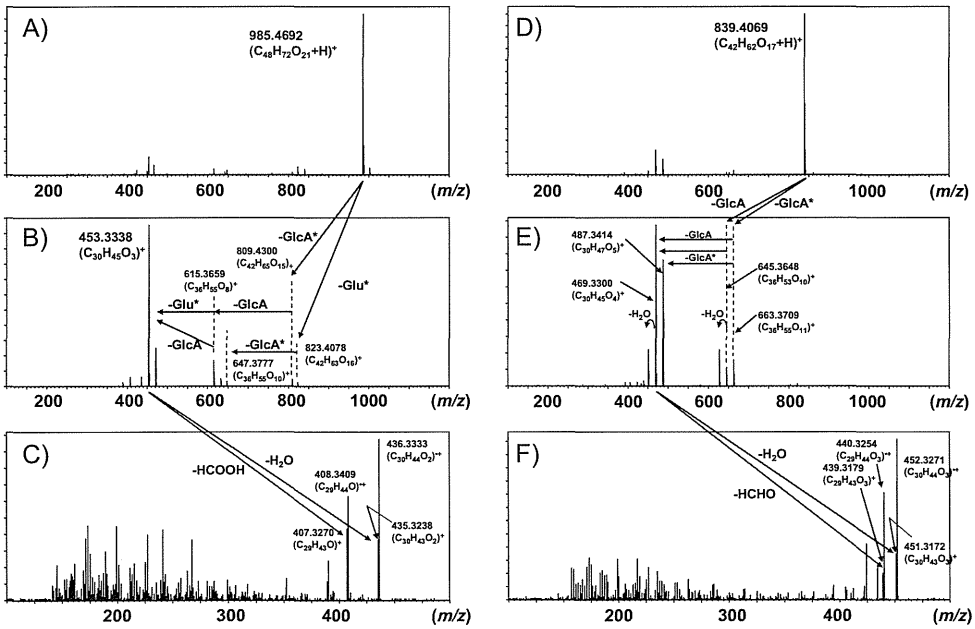


Figure 4: Positive ESI-MS, MS^2 and MS^3 spectra of **2** and **6**.
A) ESI-MS spectrum of **2**, B) MS^2 spectrum from the $[M+H]^+$ ion at *m/z* 985.4692 in A), C) MS^3 spectrum from the ion at *m/z* 453.3338 in B), D) ESI-MS spectrum of **6**, E) MS^2 spectrum from the $[M+H]^+$ ion at *m/z* 839.4069 in D), F) MS^3 spectrum from the ion at *m/z* 469.3300 in E).

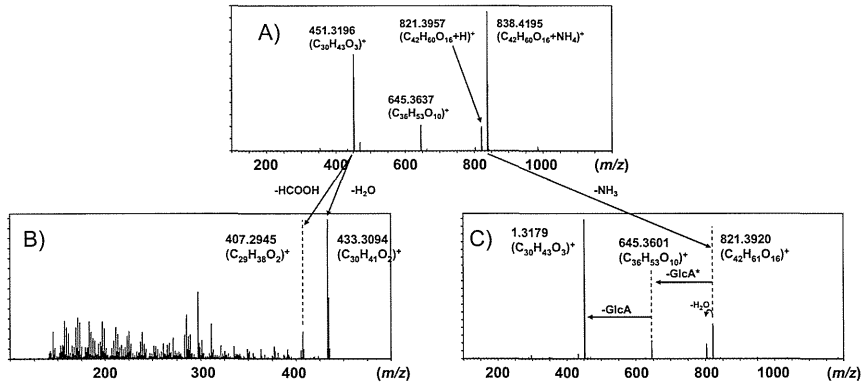


Figure 5: Positive ESI-MS and MS^2 spectra of **5**.
A) ESI-MS spectrum of 1, B) MS^2 spectrum from the ion at *m/z* 451.3196 in A), C) MS^2 spectrum from the $[M+NH_4]^+$ ion at *m/z* 838.4195 in A).

Experimental

Analytical sample preparation and reagents: Analytical grade glycyrrhizin (**1**) was purchased from Wako Chemical Co. Ltd (Osaka, Japan). Licoricesaponins A3 (**2**), B2 (**3**), C2 (**4**), E2 (**5**) and G2 (**6**) (Figure 1) were isolated from a drug sample (TMPW No. 24238) purchased from Uchida Wakanyaku Co. Ltd. (Tokyo, Japan). The isolated compounds were identified by comparing their ^{1}H- and ^{13}C-NMR spectra with those reported in the literature [5-8]. Other analytical grade chemicals and HPLC grade chromatographic solvents were also purchased from either Wako Chemical Co. Ltd (Osaka, Japan) or Nacalai Tesque, Inc. (Kyoto, Japan). The samples were dissolved in methanol at a concentration of 0.1 mg/mL. Each solution was filtered through a 0.2 μm Millipore filter unit (Advantec, Japan), and 1 μL of the filtrate was injected into an LC-MS system for analysis.

Analytical instruments: LC-MS analyses were performed with a Shimadzu LC-IT-TOF mass spectrometer equipped with an ESI interface. The ESI parameters were as follows: source voltage +4.5 kV (positive mode) and –3.5 kV (negative mode), capillary temperature 200°C, nebulizer gas 1.5 L/min. The mass spectrometer was operated in both positive and negative ion modes scanning from *m/z* 100 to 1500. A Waters Atrantis T_3 column (2.1 mm i.d. x 150 mm), maintained at 40°C, was used. The mobile phase was a binary eluent of (A) 5 mM ammonium acetate solution, (B) CH_3CN under the following gradient conditions: 0–30 min linear gradient from 10% to 100% B, 30–40 min isocratic at 100% B. The flow rate was 0.2 mL/min. In MS/MS analysis, the mass spectrometer was set to survey ions in the range *m/z* 100–1500 and to subject the most abundant ions in the scan to automatic MS/MS analysis.

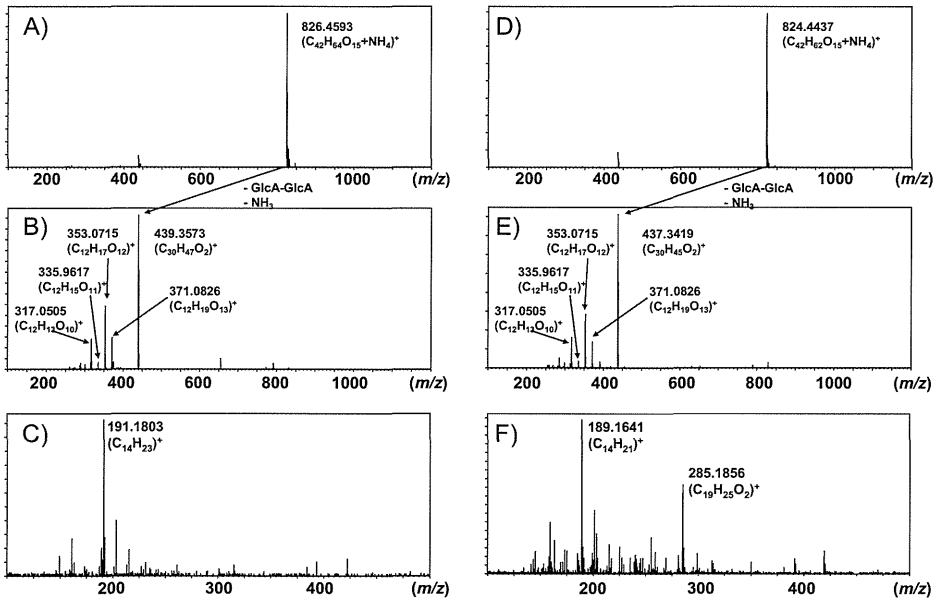


Figure 6: Positive ESI-MS, MS^2 and MS^3 spectra of **3** and **4**.
A) ESI-MS spectrum of **3**, B) MS^2 spectrum from the $[M+NH_4]^+$ ion at *m/z* 826.4693 in A), C) MS^2 spectrum from the ion at *m/z* 439.3573 in B), D) ESI-MS spectrum of **4**, E) MS^2 spectrum from the $[M+NH_4]^+$ ion at *m/z* 824.4437 in D), F) MS^2 spectrum from the ion at *m/z* 437.3419 in E).

References

[1] (a) Blumenthal M, Goldberg A, Brinckmann J. (**2000**) *Herbal Medicine: Expanded Commission E Monographs*. American Botanical Council, Newton. 233–236; (b) Society of Japanese Pharmacopoeia (**2007**) *The Japanese Pharmacopoeia*, 15th ed., The Ministry of Health, Labour and Welfare (English edition), Tokyo, p. 1026; (c) Asl MN, Hosseinzadeh H. (**2008**) Review of pharmacological effects of *Glycyrrhiza* sp. and its bioactive compounds. *Phytotherapy Research*, ***22***, 709–724.

[2] Buckingham J. (**2007**) *The Dictionary of Natural Products*. Chapman & Hall/CRC, Florida.

[3] Zhang Q, Ye M. (**2009**) Chemical analysis of the Chinese herbal medicine Gan-Cao (licorice). *Journal of Chromatography A*, ***1216***, 1954–1969; (b) Meng X, Li H, Song F, Liu C, Liu Z, Liu S. (**2009**) Studies on triterpenoids and flavones in *Glycyrrhiza uralensis* Fisch. by HPLC-ESI-MSn and FT-ICR-MSn. *Chinese Journal of Chemistry*, ***27***, 299–305.

[4] Gross JH. (2004) Mass Spectrometry, Springer, Heidelberg, p. 226.

[5] Yahara S, Nishioka I. (**1984**) Flavonoid glucosides from licorice. *Phytochemistry*, ***23***, 2108–2109.

[6] Kitagawa I, Hori K, Sakagami M, Zhou J, Yoshikawa M. (**1993**) Saponin and sapogenol. XLVIII. On the constituents of the roots of *Glycyrrhiza uralensis* Fischer from northeastern China. (2). Licorice-saponins D3, E2, F3, G2, H2, J2, and K2. *Chemical & Pharmaceutical Bulletin*, ***41***, 1337–1345.

[7] Kitagawa I, Hori K, Sakagami M, Hashiuchi F, Yoshikawa M, Ren J. (**1993**) Saponin and sapogenol. XLIX. On the constituents of the roots of *Glycyrrhiza inflata* Batalin from Xinjiang, China. Characterization of two sweet oleanane-type triterpene oligoglycosides, apioglycyrrhizin and araboglycyrrhizin. *Chemical & Pharmaceutical Bulletin*, ***41***, 1350–1357.

[8] Hatano T, Takagi M, Ito H, Yoshida T. (**1998**) Acylated flavonoid glycosides and accompanying phenolics from licorice. *Phytochemistry*, ***47***, 287–293.

Variation of Glycyrrhizin and Liquiritin Contents within a Population of 5-Year-Old Licorice (*Glycyrrhiza uralensis*) Plants Cultivated under the Same Conditions

Mareshige KOJOMA,*[,a] Shigeki HAYASHI,[b,#] Toshiro SHIBATA,[b] Yutaka YAMAMOTO,[c] and Haruo SEKIZAKI[a]

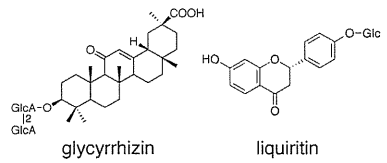
[a] *Faculty of Pharmaceutical Sciences, Health Sciences University of Hokkaido; Tobetsu, Hokkaido 061–0293, Japan:* [b] *Hokkaido Division, Research Center for Medicinal Plant Resources, National Institute of Biomedical Innovation; Nayoro, Hokkaido 096–0065, Japan: and* [c] *Tochimoto Tenkaido Co., Ltd.; Oniya, Kaibara, Tamba, Hyogo 669–3315, Japan.*
Received November 23, 2010; accepted March 28, 2011; published online May 24, 2011

Cultivated licorice plants (*Glycyrrhiza uralensis* FISCH.) contain smaller amounts of the triterpene saponin glycyrrhizin than wild licorice plants. To resolve this problem and to breed strains with high-glycyrrhizin content we determined the glycyrrhizin content of 100 samples of *G. uralensis* that were propagated from seed and grown under the same conditions in the field for 5 years. There was a 10.2-fold variation in glycyrrhizin content among these plants, ranging from 0.46 to 4.67% (average 2.11±0.90%). There was also a wide variation in liquiritin content, ranging from 0.11 to 2.65% (average 1.00±0.49%). The glycyrrhizin content was positively correlated with that of liquiritin in the taproots (r^2=0.5525). Our results indicate that there are various genetic strains for glycyrrhizin and liquiritin synthesis within a population of plants propagated from seed. The selected high-glycyrrhizin and liquiritin strains will be useful for licorice production and studies on biosynthetic analysis of glycyrrhizin and liquiritin.

Key words *Glycyrrhiza uralensis*; glycyrrhizin; liquiritin; Glycyrrhizae Radix; Leguminosae; licorice

Glycyrrhizae Radix, the underground material derived from licorice, *Glycyrrhiza uralensis* FISCH. (Leguminosae) and some other *Glycyrrhiza* species, is extensively used as an herbal medicine in worldwide.[1)] Biological activities of licorice have been attributed to glycyrrhizin, liquiritin (Fig. 1), and other compounds. *G. uralensis* is a perennial herb that grows primarily in the semi-arid zones of Asia. Licorice products are derived from wild or cultivated *G. uralensis* from China or other Asian countries. In recent decades, however, over-harvest has gradually exhausted the natural resources in these regions, and several habitats are undergoing desertification. Thus, it is important to optimized cultivation conditions of *G. uralensis* to adequately substitute for the collection of wild resources. In particular, high quality and high glycyrrhizin content are desirable attributes of *G. uralensis* for medicinal use. According to the Japanese Pharmacopoeia, the glycyrrhizin content of Glycyrrhizae Radix for medicinal use must be at least 2.5%.[2)] World Health Organization guidelines specify that glycyrrhizin content should be at least 4% for Glycyrrhizae Radix.[3)] However, the glycyrrhizin content is lower (<2.5%) in cultivated plants of *G. uralensis* than that in wild ones.[4)] Therefore, there is a demand for improved cultivars with high glycyrrhizin contents. Generally, there is wide genetic variation among plants propagated from seed, but there is no clear information on the variation in glycyrrhizin and liquiritin contents among seed-generated *G. uralensis*.

In this study, we hypothesized that there are various genetic strains within a population of *G. uralensis* grown from seed, and that some individuals would show higher glycyrrhizin and liquiritin contents. Accordingly, these could be selected to obtain a high-glycyrrhizin and high-liquiritin strain.



glycyrrhizin liquiritin

Fig. 1. Chemical Structure of Glycyrrhizin (Glycyrrhizic Acid) and Liquiritin

MATERIALS AND METHODS

Plant Materials and Cultivation Conditions *Glycyrrhiza uralensis* FISCH. plants were cultivated in a field for 5 years. A voucher specimen (Accession No. HK 15739-09) has been deposited in the Herbarium of the Division of Hokkaido, Research Center for Medicinal Plant Resources, National Institute of Biomedical Innovation, Japan (Hokkaido Division, NIBIO). Approximately 7000 seeds were sown on 7 June 2004 in the agricultural research field of the Hokkaido Division, NIBIO (Hokkaido, Japan, 44°21′N, 142°27′E). The average annual temperature at the field site is 6.1 °C (range, −28.6 to 33.3 °C; Japan Meteorological Agency) and the rainfall is 891 mm (2004—2008). The plants were grown at a density of 12.5 plants/m^2 (row spacing, 80 cm; plant spacing, 10 cm). Prior to planting, manure (20000 kg/ha), calcium carbonate (1000 kg/ha), and fertilizer (NPK 56 : 56 : 56 kg/ha) were applied to the field. Additional fertilizers were applied as follows: NPK 12 : 88 : 60 kg/ha in August 2004; calcium carbonate 1000 kg/ha and NPK 84 : 84 : 84 kg/ha in May 2005; NPK 18 : 132 : 90 kg/ha in August 2005; calcium carbonate 1000 kg/ha and NPK 84 : 84 : 84 kg/ha in May 2006; NPK 18 : 132 : 90 kg/ha in August 2006; and calcium carbonate 1000 kg/ha and NPK 126 : 126 : 126 kg/ha in May 2007. The underground parts of 600 plants were unintentionally dug up on 27 August 2008

* To whom correspondence should be addressed. e-mail: kojoma@hoku-iryo-u.ac.jp
# Equal contribution with first author.
© 2011 Pharmaceutical Society of Japan

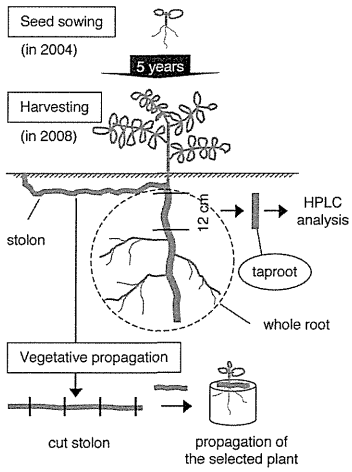


Fig. 2. Cultivation Process, Diagram of Harvested Parts, and Vegetative Propagation from *Glycyrrhiza uralensis* Stolons

Taproots (12-cm segments) were used for glycyrrhizin analyses.

(Fig. 2), taking particular care not to damage the roots (approximately 50—100 cm in length) and stolons (50—150 cm). Then, we visually selected 100 plants that had large root system and no damage by insects or disease from the 600 plants, washed in running tap water and then dried (50 °C, 10 d).

Glycyrrhizin and Liquiritin Analyses We analyzed the glycyrrhizin and liquiritin contents of the taproots cut from the whole root mass of the 100 selected plants (Fig. 2). It is possible that the glycyrrhizin and liquiritin contents are different according to the part of the root; therefore, we selected a 12-cm piece of taproot taken from the basal part of the root system for analysis. The taproot pieces were dried (50 °C) and ground to a fine powder (<150 μm). Quantitative analysis by HPLC of glycyrrhizin was performed according to the procedure in the Japanese Pharmacopoeia[2] with minor modifications. Dried samples (250 mg) were extracted in 50% EtOH (12.5 ml) with reciprocal shaking (150 rpm) for 15 min, and then ultrasonicated for 10 min. After centrifugation at 2500 ***g*** for 10 min, the supernatant was collected. This was repeated twice and extracts were made up to a final volume of 25 ml. A volume of 20 μl was used for HPLC analysis. The HPLC system consisted of an LC-2000 Plus system (JASCO, Tokyo, Japan), a TSKgel ODS-80T$_S$ QA column (5 μm, 150×4.6 mm, TOSOH, Tokyo, Japan), and a Mightysil guard column RP-18 GP 5-4.6 (Kanto Kagaku, Tokyo, Japan) with column temperature of 30 °C. A solvent system consisting of 2% (v/v) acetic acid : CH_3CN (60 : 40) was used at a flow rate of 0.6 ml/min. Elution of compounds was monitored by measuring absorbance at 254 nm. A glycyrrhizin standard was purchased from Tokyo Chemical Industry (Tokyo, Japan). Liquiritin analysis was performed as described by Kitagawa *et al.*[5] with minor modifications. Samples (250 mg) were extracted in 90% MeOH (12.5 ml), and then extraction was carried out in the same way as that used for glycyrrhizin analysis, as described above. The HPLC system was the same as that used for glycyrrhizin analysis, except for the column (Mightysil RP-18 GP 250-4.6 column; 5 μm, 250×4.6 mm, Kanto Kagaku, Tokyo, Japan; column temperature, 40 °C). A solvent system consisting of 2% (v/v) acetic acid : CH_3CN (80 : 20) was used at a flow rate of 0.7 ml/min. Elution of compounds was monitored by measuring absorbance at 257 nm. A liquiritin standard was purchased from Kishida Chemical Co. (Osaka, Japan).

RESULTS

Glycyrrhizin Content Glycyrrhizin contents of the 100 samples of *G. uralensis* that were grown under the same conditions for 5 years are shown in Fig. 3A. There was a wide variation in glycyrrhizin content; the highest was 4.67% (plant No. 79), 10.2-fold higher than the lowest (0.46% in No. 21). The average glycyrrhizin content was 2.11±0.90%. Of the 100 samples of *G. uralensis*, five had glycyrrhizin contents of 0—1%, 47 had 1—2%, 32 had 2—3%, 9 had 3—4%, and 7 plants had 4—5%.

Root Weight The fresh root weight (whole root, Fig. 2) of each of the 100 samples of *G. uralensis* was recorded not to dry within 24 h after digging up (data not shown). There were considerable differences in the whole root weights of the plants. The average whole root fresh weight was 163.6±61.0 g, with a 5.9-fold difference between the highest (431.6 g, No. 15) and the lowest (73.0 g, No. 53). There was no clear relationship between glycyrrhizin content and whole root weight (r^2=0.0082, Pearson's correlation coefficient). The taproot diameter of each of the 100 samples of *G. uralensis* was also recorded (data not shown). Taproot diameter ranged from 18.1 mm (No. 53) to 36.0 mm (No. 15), with an average of 26.5±3.9 mm. There was a positive correlation between whole root weight and taproot diameter r^2=0.6532). However, there was no clear correlation between glycyrrhizin content and taproot diameter (r^2=0.0205).

Liquiritin Content The liquiritin contents of the taproots are shown in Fig. 3B. There was a wide variation in liquiritin content among the 100 samples of *G. uralensis*. The highest liquiritin content was 2.65% (No. 15), 24.1-fold higher than the lowest (0.11% in No. 55). The average liquiritin content was 1.00±0.49%. Of the 100 samples of *G. uralensis*, 56 had liquiritin contents of 0—1%, 41 had 1—2%, and three had 2—3%. Liquiritin content was not correlated with whole root fresh weight (r^2=0.0042) or taproot diameter (r^2=0.0303).

Relationship between Glycyrrhizin Content and Liquiritin Content The glycyrrhizin content was positively correlated with that of liquiritin in the taproots (Fig. 4, r^2=0.5525).

DISCUSSION

In this study, we observed wide variations in glycyrrhizin and liquiritin contents among 100 samples of 5-year-old cultivated *Glycyrrhiza uralensis*. It is thought that the environmental conditions for cultivation, genetic characteristics of each plant influence the difference of plant secondary metabolites production. We think that the difference of the each content of glycyrrhizin and liquiritin in this study is chiefly caused by a difference of the biosynthesis ability at each gene level. Glycyrrhizin is generated with the triterpenoid biosynthetic pathway. In *Glycyrrhiza* plants, the biosynthesis of glycyrrhizin involves the initial cyclization of 2,3-oxidosqualene to a triterpene, β-amyrin, followed by a series of oxidative reactions and glucuronylation. Genes for

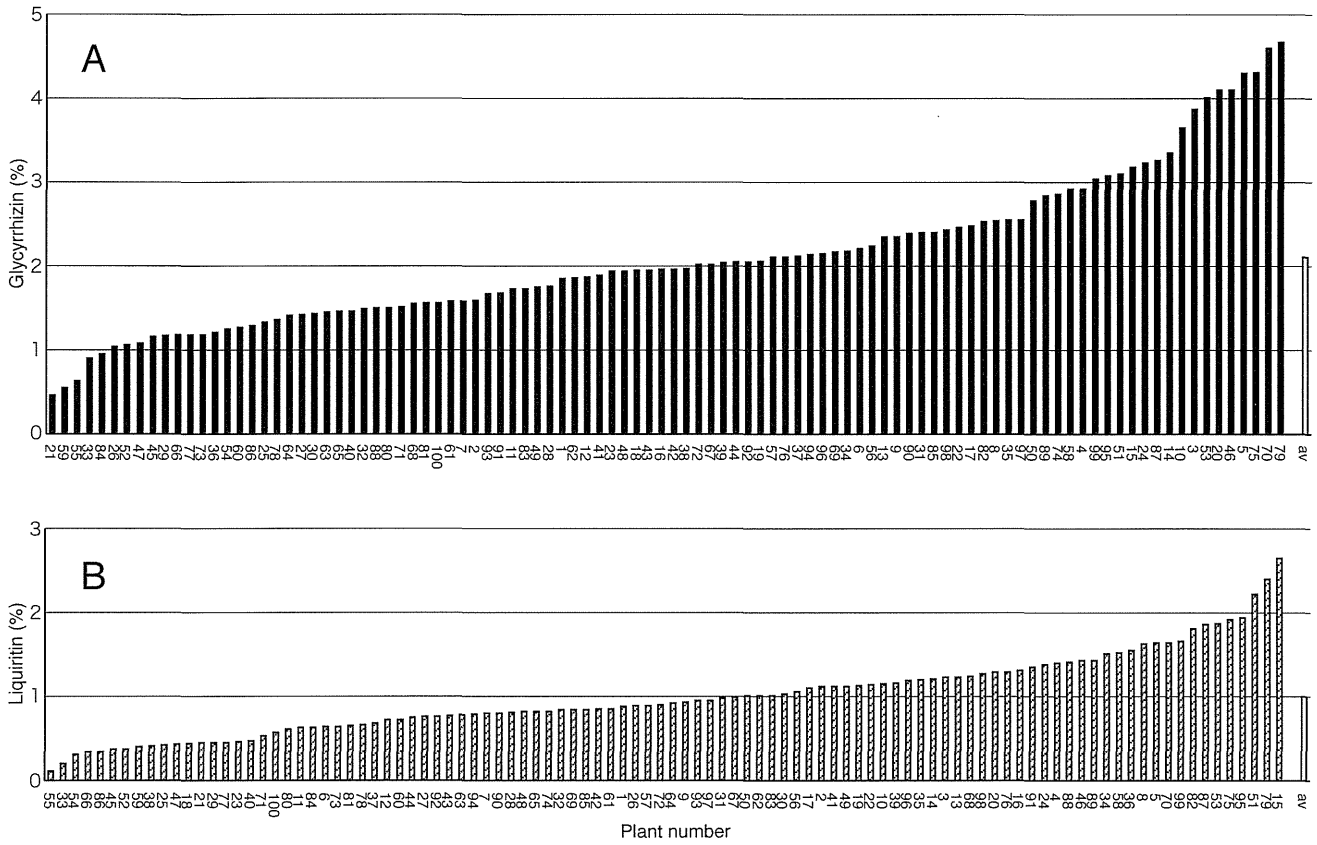


Fig. 3. Glycyrrhizin (A) and Liquiritin (B) Contents of Taproots from 100 Plants of 5-Year-Old *Glycyrrhiza uralensis* (av, Average of 100 Plants)

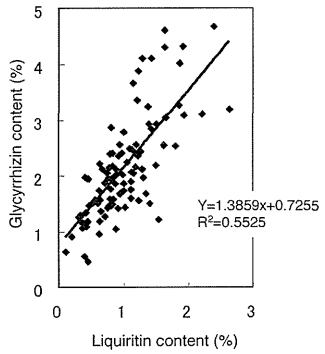


Fig. 4. Correlation between Glycyrrhizin Content and Liquiritin Content of Taproots from 100 Plants of 5-Year-Old *Glycyrrhiza uralensis*

two of the key enzymes in this pathway were discovered in *G. glabra*: *GgSQS1* encodes squalene synthase, which catalyzes the synthesis of 2,3-oxidosqualene from squalene, and *GgbAS1* encodes β-amyrin synthase, which produces β-amyrin from 2,3-oxidosqualene.[6,7] The *CYP88D6* gene, which was isolated from *G. uralensis*, encodes a cytochrome P450 monooxygenase that catalyzes the two-step oxidation of β-amyrin at C-11 to produce 11-oxo-β-amyrin.[8] Recently, Sudo *et al.* analyzed 56857 expressed sequence tags of *G. uralensis*.[9] On the other hand, liquiritin is generated with the flavonoid synthetic pathway. Studies on the roles of cytochrome P450s in the production of flavonoids (to which class liquiritin belongs) are also progressed in *Glycyrrhiza* plants.[10] The glycyrrhizin content was positively correlated with that of liquiritin in the present study, even though glycyrrhizin and liquiritin are synthesized *via* different biosynthetic pathways. It is assumed that the each biosynthetic pathway of the triterpenoid and flavonoid takes part in parallel or the upstream part in these two biosynthetic pathways takes part in the biosynthesis regulations of glycyrrhizin and liquiritin. However, the detail reason for the positive correlation between glycyrrhizin and liquiritin content is unclear now. These phenomena in this study will be further clarified by genetic analysis of secondary metabolites synthesis including glycyrrhizin and liquiritin. We expect that the present results will contribute to such biosynthetic studies.

We are propagating the selected high contents of glycyrrhizin and liquiritin plants by stolon cuttings (Fig. 2) and *in vitro* micro-propagation using tissue culture.[11] Thus, it may be possible to produce high quality strains of *G. uralensis* with high glycyrrhizin and high liquiritin contents. The positive correlation between glycyrrhizin and liquiritin content is important information for selection or breeding of high glycyrrhizin and high liquiritin strains. There has been no report on the cultivation of *Glycyrrhiza* plants with only high content of glycyrrhizin. Yamamoto and Tani[4] reported glycyrrhizin contents of 1.52±0.57% in taproots of cultivated 4-year-old *G. uralensis*. This level is lower than that required for medicinal use. The result may be due to the contamination with low glycyrrhizin individuals. By using cloned plants of the selected strains in our study, it will be possible to clarify the effects of environmental factors such as cultivation period, fertilizer, temperature, *etc.* on glycyrrhizin and liquiritin production.

Acknowledgments This work was partially supported by Grants-in-Aid for Scientific Research from the Ministry

of Education, Culture, Sports, Science and Technology of Japan, a Health Sciences Research Grant from the Ministry of Health, Labour and Welfare of Japan, and a Grant for Research for Promoting Technological Seeds from the Japan Science and Technology Agency.

REFERENCES

1) Hayashi H., Sudo H., *Plant Biotechnol.*, **26**, 101—104 (2009).
2) The Society of Japanese Pharmacopoeia, "Glycyrrhizae Radix," The Japanese Pharmacopoeia (Pharmacopoeia of Japan), 15th ed., Ministry of Health, Labour and Welfare of Japan, Tokyo, Japan, 2006, p. 1295.
3) World Health Organization, "WHO Monographs on Selected Medicinal Plants," Vol. 1, World Health Organization, Geneva, Switzerland, 1999, p. 187.
4) Yamamoto Y., Tani T., *J. Trad. Med.*, **19**, 87—92 (2002).
5) Kitagawa I., Chen W. Z., Taniyama T., Harada E., Hori K., Kobayashi M., Ren J., *Yakugaku Zasshi*, **118**, 519—528 (1998).
6) Hayashi H., Hirota A., Hiraoka N., Ikeshiro Y., *Biol. Pharm. Bull.*, **22**, 947—950 (1999).
7) Hayashi H., Huang P., Kirakosyan A., Inoue K., Hiraoka N., Ikeshiro Y., Kushiro T., Shibuya M., Ebizuka Y., *Biol. Pharm. Bull.*, **24**, 912—916 (2001).
8) Seki H., Ohyama K., Sawai S., Mizutani M., Ohnishi T., Sudo H., Akashi T., Aoki T., Saito K., Muranaka T., *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14204—14209 (2008).
9) Sudo H., Seki H., Sakurai N., Suzuki H., Shibata D., Toyoda A., Totoki Y., Sasaki Y., Iida O., Shibata T., Kojoma M., Muranaka T., Saito K., *Plant Biotechnol.*, **26**, 105—107 (2009).
10) Ayabe S., Akashi T., *Phytochem. Rev.*, **5**, 271—282 (2006).
11) Kojoma M., Ohyama K., Seki H., Hiraoka Y., Asazu N. S., Sawa S., Sekizaki H., Yoshida S., Muranaka T., *Plant Biotechnol.*, **27**, 59—66 (2010).

DATABASE FOR CRUDE DRUGS AND KAMPO MEDICINE

MASANORI ARITA[1,2,*] arita@bi.s.u-tokyo.ac.jp
MIWA YOSHIMOTO[3] yoshimotom@jaist.ac.jp
KAZUHIRO SUWA[4] suwa@liorect.jp
AKI HIRAI[5] SHIGEHIKO KANAYA[2,5] {hirai,skanaya}@gtc.naist.jp
NAOTOSHI SHIBAHARA[6] KEN TANAKA[7] {shiba1,ktanaka}@inm.u-toyama.ac.jp

[1] *Department of Biophysics and Biochemistry, Graduate School of Science The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[2] *Metabolome Informatics Unit, RIKEN Plant Science Center 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan*
[3] *School of Information Science, Japan Advanced Institute of Science and Technology 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*
[4] *The Liorect, Co., Ltd. 5-10-5 Shimoochiai, Chuo-ku, Saitama 338-0002, Japan*
[5] *Department of Bioinformatics and Genomics, Graduate School of Information Science Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan*
[6] *Research Center for Ethnomedicine, Institute of Natural Medicine, University of Toyama 2630 Sugitani, Toyama 930-0194, Japan*
[7] *Department of Medicinal Resources, Institute of Natural Medicine, University of Toyama 2630 Sugitani, Toyama 930-0194, Japan*

A wiki-based repository for crude drugs and Kampo medicine is introduced. It provides taxonomic and chemical information for 158 crude drugs and 348 prescriptions of the traditional Kampo medicine in Japan, which is a variation of ancient Chinese medicine. The system is built on MediaWiki with extensions for inline page search and for sending user-input elements to the server. These functions together realize implementation of word checks and data integration at the user-level. In this scheme, any user can participate in creating an integrated database with controlled vocabularies on the wiki system. Our implementation and data are accessible at http://metabolomics.jp/wiki/.

Keywords: Kampo, Chinese medicine, database, wiki

1. Introduction

1.1. *What is Kampo medicine?*

Kampo medicine (KM) is different from the traditional Chinese medicine (TCM). Japanese Pharmacopoeia (JP or "Nihon Yakkyoku Ho") [1] contains around 150 crude drugs, and they are prescribed *in combination with* Western drugs. The combined

---

* Part of this work first appeared as "Ontology checking and integration of crude-drug and kampo information" in the Proceedings of the 2nd International Conference of Biomedical Engineering and Informatics (BMEI2009, Tianjin China, 3:1304-7) and as "Database for Crude Drugs and Kampo Medicines" in the Proceedings of the 2010 Annual Conference of the Japanese Society of Bioinformatics (JSBi2010, Fukuoka Japan, T03).

prescription is unparalleled in any traditional medicine, including traditional Korean medicine and TCM; Japan is a unique country where crude drugs have been administered in Western standards with records for around 40 years. In this sense, KM is neither complementary nor alternative medicine. This distinction from Chinese or other traditional medicines is sometimes emphasized in Japan. (In China, on the other hand, all Asian medicines are integrated into TCM.) Since traditional medicines, contrary to patented Western medicine, cannot escape from the debate of intellectual property rights, accurate understanding of their origins and history is important. In the World Health Organization (WHO) terminology, KM is defined as "the medicine traditionally practiced in Japan, based on ancient Chinese medicine" [2].

1.2. *Kampo formula*

In KM, prescribed formula is a weighted combination of multiple crude drugs, and is specified by a list of drug names and their amounts. As an example, "Kakkonto" (pueraria decoction) formula is shown in Table 1.

Table 1. Standard Formula of Kakkonto (pueraria decoction). These ingredients are simmered and the decocted extract is administered orally.

English Name	Latin Name of Original Plants	Amount (g/day)
Pueraria Root	*Pueraria lobata*	4
Jujube	*Zizyphus jujuba*	3
Ephedra Herb	*Ephedra sinica*	3
Glycyrrhiza	*Glycyrrhiza uralensis*	2
Cinnamon Bark	*Cinnamomum cassia*	2
Peony Root	*Paeonia lactiflora*	2
Ginger	*Zingiber officinale*	2

In Japan, 236 official Kampo formulas are approved by the Ministry of Health, Labour and Welfare (MHLW) and available in an official literature source [3]. The internationally official name for each Kampo prescription is written in Romanized Japanese kana script [4]. (Therefore, Kakkonto in English is "Kakkonto", not "Pueraria decoction".) Nevertheless, standardization of KM is not straightforward mainly for two reasons. First, quality of herbal products may vary depending on their source materials, place of cultivation, and their production process. Second, each formula may be personalized according to the health condition of patients. To internationalize KM and to utilize accumulated data for knowledge extraction, medical records need to be digitized and organized systematically.

1.3. *Necessity for a public information repository*

In the genome-sequencing era, the wealth of clinical information hitherto accumulated will inevitably face the matching amount of genome information such as the recently launched 1000 plants (1KP) or other genome projects [5,6]. An immediate task will be the resolution of discrepancy between biochemical knowledge such as molecular interaction/function and medical knowledge such as hypo/hypertensive activity. In

addition, linking information on active compounds with their biosynthetic genes remains to be another important task [7]. To achieve this goal, the following three steps are necessary [8]:

1. Data accumulation in which quantity is more emphasized than quality;
2. Data curation/organization, i.e., creating ontology by a community effort; and
3. Knowledge extraction from the data for unbiased analysis.

We consider that a wiki-based repository can achieve these steps. The advantage of using wiki is the user-friendliness and flexibility to accommodate various data types such as pictures and chromatograms. It also reduces the cost for data revision and updates. Here we introduce our initial implementation, which also includes information from various collaborators including the members of Kanaya Laboratory at Nara Institute of Science and Technology; Tochimoto Tenkaido Co Ltd; Alps Pharmaceutical lnd. Co Ltd; and NTS Publishing Inc.

2. Methods

2.1. Wiki-based vocabulary control: HTML Forms

The notable difference between the relational database system and wiki systems is the notion of data integrity. Standard wiki is merely a collection of web-pages, in which users can edit contents arbitrarily. To standardize and control user inputs, we need a mechanism to guide users to follow certain ontology. The easiest solution is to support a choice list, i.e., the function to import certified terms from other information sources in order to control vocabulary. In more practical examples, a wiki user should use interactive contents such as drop-down lists and check boxes. Still, the mechanism for such interactive choices should not be predefined, and the word choices be only suggested, not enforced.

Many dynamic web pages utilize the 'HTML Forms' to create interactive contents. The W3C recommendation of the 'Forms' describes components such as drop-down box, check box, or (radio) button, and they are supported by all web browsers. Obviously, the above mentioned solution can be achieved by integrating the Forms into wiki, and it is also the simpler way to implement interactive components than to use, say, JavaScript.

We implemented the core set of the Forms into our MediaWiki system: specifically, `<form>` and `<input>` HTML tags for achieving single/multiple choices were allowed inside wiki. However, our goal is not just writing such tags inside the wiki. In their original usage, selection items and tag commands are written within the same HTML page. In our wiki implementation, on the other hand, items to be chosen are imported from another page on demand so that possible selection are externally given and can be shared among multiple pages.

To realize word selection written in external pages, a mechanism to duplicate information between pages is required. The mechanism cannot be achieved by describing a reference function (e.g. by using programs) inside wiki pages, because selected words must be instantiated as a page source instead of being just *displayed* as a

result of computation. For example, let us imagine implementation of a click button to edit certain data field in a wiki page (Fig. 1).

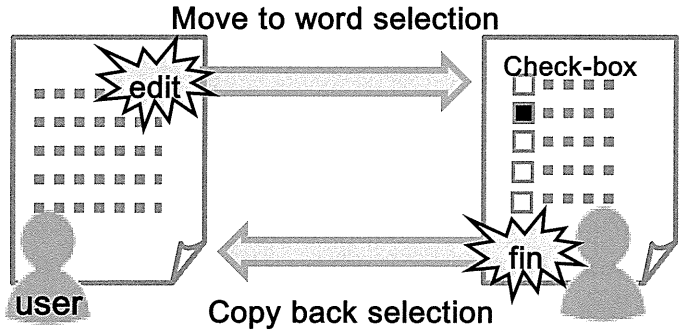


Figure 1. Flow of the word choosing process. The user invokes a transition by pressing the edit/finish button. Executed programs invoked by HTML Forms must be editable on the user side. Resources on the wiki server are inaccessible and data are copied only between client pages.

When a user clicks this button, another page (or window) for word selection appears, and a user chooses any set of words by using check boxes. After selection, the user returns to the original page by pressing a 'finish' button, and the selected words are *copied* into the original page. After the choice, the selected words can be edited arbitrarily. There are three required parameters in this workflow: (1) location of the page under edit (i.e. where the choice button is); (2) location of the selection list and the finish button; and (3) the list of words currently chosen. In the implementation, two MediaWiki commands were introduced:

`{{#formtag:` *type* `|` *parameter*`}}` and `{{#get:` *parameter*`}}`,

where *type* is either '`form`' or '`input`', and *parameter* is a string directly passed to the HTML tag. Specifically, the `#formtag:form` command is translated into:
`<form action="cgi-bin/foo.cgi" method="post">`*parameter*`</form>`,
and the `#formtag:input` command is translated into `<input` *parameter*`>` by the wiki engine.

The `#get` command is used to obtain parameter values which are posted to the server. Through these commands, the cgi script (indicated as foo.cgi in the above sample code) receives the three parameters: source page, selection page, and the current word choice. When a 'submit' option is invoked on the source page, the user will be forwarded to the selection page. The selection page is preprocessed to display interactive components such as check boxes. This preprocessing will be explained in the next subsection. When the user finishes the selection of words, another 'submit' option is invoked. This time, the user will be forwarded back to the source page, and the page must be updated to reflect the new selection. To designate the page contents to be replaced, the source page has a special marker '`++variable name++`' into which the new selection is inserted. With these settings, it is not hard to implement the required workflow in the cgi program.

2.2. Preprocessing a selection list using the Lua programming language

If the preprocessing part were fixed, i.e. how and where to place check boxes and drop down lists were PHP-coded as in standard 'special pages' of MediaWiki, then not much flexibility would be left for end users. Since the selection page is better to be freely designed like normal wiki pages, the layout of `form` elements such as check boxes and buttons is controlled by the Lua programming language [9]. Lua is a lightweight functional language and is designed as an embeddable scripting engine [10]. In our implementation, a Lua process is called by a special command
{{#lua: source *code* | *standard input*}}.
The standard output of the Lua process will be embedded where the command is placed. For security reasons, file-accessing libraries of Lua are nullified by overriding its primitives, and the running time is limited by an external daemon process. Since Lua is a full-fledged programming language, it adds universal computation power to all end users. The only limitations are the running time limit and closed I/O libraries. (Note that the I/O is closed only for the Lua language. The I/O of MediaWiki is left untouched.)
In a typical implementation style, a selection mechanism is implemented as a Lua program in one page, and the selection data are located in other pages. This strategy separates edit buttons and related functions from user-defined data, and improves reusability of Lua programs.

2.3. Data collection

Information of crude drugs, their key identifiers (Latin or scientific names), and indications were manually collected from the JP 15th Edition [1] (The most recent version is the 16th Edition, released in April 2011). Kampo formulas, especially the variation in their compositions, were taken from other literature sources [3,11,12]. For the database, we have accumulated information of 158 crude drugs and about 348 Kampo formulas. The full lists of crude drugs and formulas are available on the page "http://metabolomics.jp/wiki/Persist:CrudeDrugList" and "Persist:KampoList" [†], respectively. The total number of literature sources exceeded 60, and 7677 academic references were also integrated. Each reference item occupies one wiki page, and is associated with Kampo formulas to which it describes. From every page of Kampo formulas, associated references are linked automatically.

3. Results

3.1. Overview of the Kampo data repository

In Fig. 2, the overview of the Kampo data repository is depicted. Users are expected to visit the top page at "Category:CD", which is linked to many index pages. Index pages

[†] Hereafter, we indicate only page titles of our wiki site, and the common server address (http://metabolomics.jp/wiki/) will be omitted.

are automatically generated pages by searching keywords in each data group such as names of crude drugs or Kampo formulas. The primary keys in our repository are Latin names for crude drugs and Romanized names for Kampo formulas. Such keys are used to represent explicit links between data pages. When no primary key is shared between data pages, related links are generated by on-demand searches. For example, each Kampo formula is a combination of crude drugs; each formula page can have explicit links to crude drugs. On the other hand, each crude-drug page does not keep track of all information on Kampo formulas, whose contents may change in the repository. Therefore, the crude-drug name in Latin is searched against all formula pages at the time of page rendering; hit pages are listed as links which look like just the same as explicit links for end users. Since the system is wiki-based, each page is freely editable through a browser.

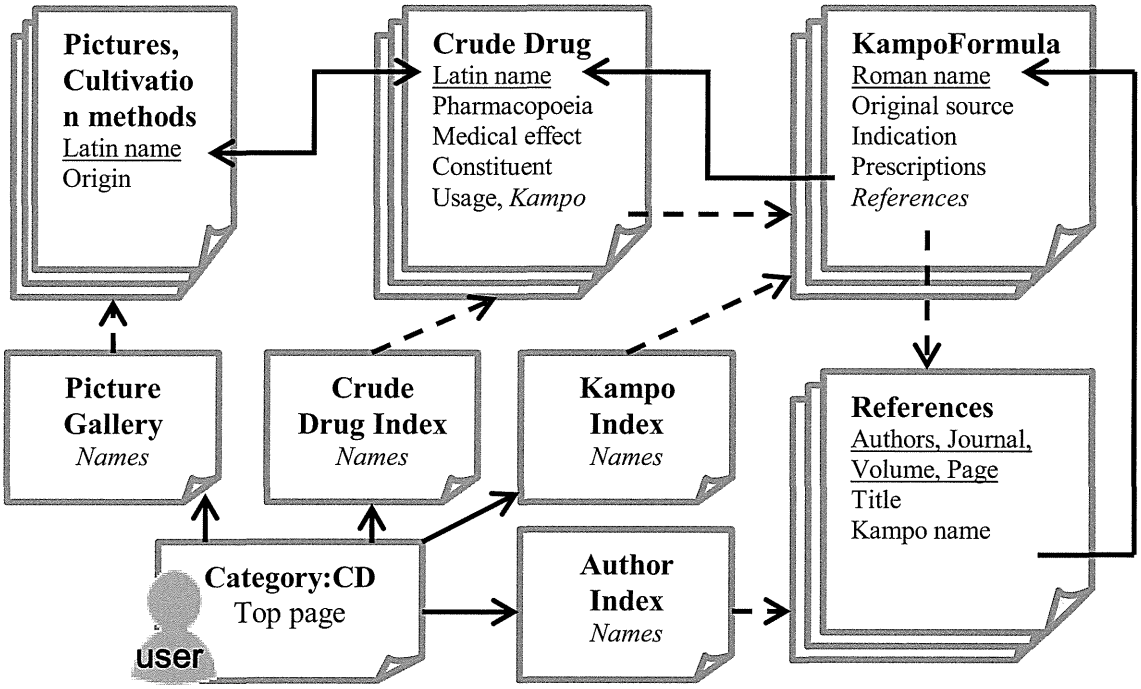


Figure 2. Overview of the Kampo repository. Each square box indicates one wiki page, and overlaid boxes indicate page groups. Solid lines show explicit links. Dashed lines show automatic links generated by page searches. Page titles are in bold face, page contents in standard face, and automatically generated contents in italic face.

3.2. Large variation in Kampo formula

Not a few differences in ingredients and their names were found in Kampo formulas from various literature sources. Of total 348 Kampo formulas from over 60 literatures, 30% (108 formulas) were described in a single literature only. Among the rest, only 27% (65) had identical drug ingredients in multiple literature sources. When drug amounts were also considered, only 2% (8 formulas: Dokkatsuto, Jashoshito, Jingyobofuto, Jingyokyokatsuto, Jogan'ippo, Oshosan, Satotsuko, and Shobaito) had exactly the same doses in multiple references. Most varied was Shakanzoto, which had

15 dose variations. In our site, each formula can be accessed by name, for example, as "Kampo:Shakanzoto" on the wiki platform.

This result does not mean, however, that KM is unorganized. First, Kampo formula is originally designed to be prescribed for each patient by considering symptoms and conditions. Second, some crude drugs can be considered almost identical. For example, (Kan)Shokyo (dried ginger) and Kankyo (steamed and dried ginger) are made from the same ginger root and only the production process is different (steaming or not). The same is true for Kippi and Chimpi (both from sour orange peel) or Keihi and Keishi (both from cinnamon bark). When these names were considered identical, the number of inconsistent formulas was reduced to less than 50. Typical variation was addition of optional ingredients.

Even after this merging process, we could find several unique formulas in the book titled "Natural Medical Resource 2nd Edition" (in Japanese). To give some examples, Choijokito prescribed for constipation contains Uncaria hook, which is used as vasodilator. Choreito for kidney disorder contains Angelica root which is used for feminine ailment. It is noteworthy that we could locate these putative errors through integration and comparison with other literature sources.

3.3. Extraction of modern Kampo usage

For each Kampo formula, related keywords were extracted from associated reference titles. Frequent co-occurrence of Japanese keywords, for example more than 4 in reference titles, already produced observations such that "Anchusan" is used as parasiticide against anisakis, "Kakkonto" against influenza, "Tokishakuyakusan" for dementia, or "Chotosan" for hypertension. Currently each formula page in our database displays automatically extracted keywords from reference titles. Such words are not described in official indications but reflect modern trends of Kampo usage. There is a severe limitation for on-demand computation inside wiki pages, but if such relationships are once output (e.g. downloaded from the wiki site), they can be further used for network analysis.

3.4. Linking with plant taxa and English information

Each crude drug is linked with information of original plants (currently no information for animals or minerals). Also available is statistical information at the plant family level. Although details of crude drugs have been available only in Japanese by tradition, the database contains English information for 148 formulas taken from the special edition of Kampo, Acupuncture and Integrated Medicine "Current Kampo Medicine" (2006) [13] and 119 English package inserts for ethical use compiled by Japan Pharmaceutical Reference [‡]. Page titles are in Latin names and additional herbal information is also available in English. Bioactivities and reference information are

[‡] Its full list is shown on the "Index:JPR" page.

organized through the mechanisms in Section 2, and registered users can edit both data processing programs in Lua as well as data themselves.
It is worth mentioning that the drug- and Kampo-list pages on our server are not always created and curated manually: data redundancy can be minimized as in the relational database system (RDB). Except for key identifiers corresponding to primary keys in RDB, the data are only *referenced* inside multiple pages through the extensive use of key identifiers and inline page search commands [9]. For example, pages for crude drugs contain information of related Kampo formulas, and these links are automatically generated by page searches. Similarly, composition of Kampo formulas is automatically linked to each crude-drug page. Therefore, wiki editors do not need to care about manual addition of error-prone hyperlinks. Since all functions are implemented at the user level, a user can verify details of the computation flow by browsing page sources and can modify or copy them if necessary.

4. Discussion

4.1. Related databases in Asian countries

After much discussion on biodiversity and environment protection, national institutions in major East Asian countries commenced collection of information on traditional medicines. In Japan, Research Center for Medicinal Plant Resources (National Institute of Biomedical Innovation) launched a database project on chemical constituents, biosynthetic genes, biological activities of 75 medicinal plants in collaboration with several Japanese universities in 2010 [14]. The center also released a pilot database for medicinal plants early in 2010 [15]. Korea Institute of Oriental Medicine designed a portal site OASIS (Oriental Medicine Advanced Searching Integrated System) for establishing an alliance among Korean research institutions on medicinal plants [16]. China Academy of Traditional Chinese Medicine provides Traditional Chinese Medicine Database System, which includes literature information (more than 120,000 references from 800 journals) , materia medica (herbal medicine), medical formula, clinical medicine, and so on [17]. The often mentioned drawback of these websites is their exclusive nature. Data access is usually restricted to collaborators. It is very hard for foreign users to navigate these sites. We have contacted Research Center for Medicinal Plant Resources in Japan and agreed to collaborate for creating a more open, user-friendly database system. Our wiki design will therefore contribute to the Japan-wide consortium of the medicinal plant resources. Indeed, the crude drug information will become a part of new database system at Institute of Natural Medicine, University of Toyama under the authorization of the Medical and Pharmaceutical Society for WAKAN-YAKU in Japan.

4.2. Relationship with semantic webs and XML

Semantic web is the next-generation web standard and its quintessence are hyperlinks with predicates: each link is assigned a predicate so that the relationship of links is machine-interpretable. Inline page searches are also supported in several implementations (for examples, see the "SemanticWiki" page in English Wikipedia). Although there is no doubt about its superiority over current monotone hyperlinks, ontology of predicates themselves becomes an important issue for discussion. Since each data provider can add custom-made predicates, streamlining and understanding predicates is even more difficult than to standardize original data. Interested readers are recommended to visit websites for predicate ontology such as the Open Biological and Biomedical Ontologies (OBO) [18] or Vocabulary of Interlinked Datasets (VOID) [19]. The disadvantage of XML is similar. It bloats up the data amount and the custom set of tags or predicates is sometimes complicated. An easy-to-understand mechanism to manage predicates will be needed and our proposal fits into this niche. Our idea does not conflict with Semantic Wiki or XML. It can be utilized not only for standard wiki-based sites but also for such semantic-wiki sites.

Not a few biological databases have been converted to the wiki style [20], but an approach to manage ontology through inline page search or dynamic page copying is unheard of. By extending data contents to metabolome and genome information, we plan to develop interdisciplinary information portal using this approach.

4.3. Limitations of the wiki-based approach

The proposed approach uses full-text search extensively. The search is performed on demand to collect necessary information (e.g. statistics) and to create links to other pages. The strategy is therefore not scalable. (In MediaWiki, page contents are recommended to be less than 30 Kbytes in each page by default.) Users do not suffer from navigation delay only when the number of pages remains less than tens of thousands; pre-computation such as using a suffix array will be necessary for a larger scale. The pre-computation is not straightforward, however, because the update needs to be on-demand and dynamic. The development of dynamic data structures is a major research topic to improve scalability. The wiki-based approach also does not fit with large-scale data processing using application program interface (API) due to its flexible format. Although users can download full page contents using the original interface of MediaWiki (see its user manual for details), much efforts will be needed to integrate the system into conventional style of data federation using fully defined formats and protocols. Since MediaWiki uses SQL database in its backend, using the wiki-based interface only for human-computer interaction is the likely solution.

5. Conclusion

We created a database for crude drugs and Kampo medicine using a wiki system. In biology, there are numerous wiki-based database projects. In a standard setting, each

page is edited independently and detection of data duplicates or installation of ontology is difficult. In such situations, our approach would be an easy choice than to introduce a whole new system and to force all users to reformat their data accordingly.

Acknowledgments

This work is partially supported by JST Institute for Bioinformatics Research and Development and Grant-in-Aid for Publication of Scientific Research Results & Scientific Research on Priority Areas "Systems Genomics" from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] The Japanese Pharmacopeia, JP 15th Edition
http://www.mhlw.go.jp/topics/bukyoku/iyaku/yakkyoku/english.html
[2] WHO Regional Office for the Western Pacific. *WHO International Standard Terminologies in Traditional Medicine in the Western Pacific Region*, World Health Organization, Western Pacific Region, 2007
[3] Pharmaceutical Affairs Bureau of MHLW. *Handbook for Formulatinq Kampo Medicines for General Uses* ("*Ippan-yo Kampo Shoho No Tebiki*" in Japanese) Yakugyo Jihosha, Tokyo, 1975
Latest revision: http://www.japal.org/contents/pdf/notice/20100401_0401-2.pdf
[4] Tsutani, K. *et al.* Standard Kampo Formula Nomenclature (in Japanese) *Japanese Journal of Oriental Medicine* 56(4):611-622, 2005
Available online at http://www.jsom.or.jp/html/magazine.html
[5] The oneKP Project by the University of Alberta. http://www.onekp.com/
[6] PlantGDB by the National Science Foundation, USA. http://www.plantgdb.org/
[7] Walsh, C.T. and Fischbach, M.A. Natural Products Version 2.0: Connecting Genes to Molecules, *Journal of the American Chemical Society*, 132:2469–2493, 2010
[8] Arita, M. What can metabolomics learn from genomics and proteomics?, *Current Opinion in Biotechnolgy*, 20:610-615, 2009
[9] Arita, M., Suwa, K. Search extension transforms Wiki into a relational system: A case for flavonoid metabolite database, *BioData Mining*, 1:7, 2008
[10] Ierusalimschy, R. *Programming in Lua* (2nd Edition). Lua.org, ISBN 85-903798-2-5, 2006
[11] Harashima, H. (Ito, M. and Kitayama, T. eds.) *Shoyakutan* (in Japanese) *NTS Inc.*, Tokyo, 2007
[12] Satake, M., Ida, Y., Nemoto, Y. (eds.) *Practical Guide to Crude Drugs* ("*Shoyaku Kaisetsu*" in Japanese) Yakugyo Jihosha, Tokyo, 2001
[13] KAIM Special Edition - Current Kampo Medicine.
http://www.kaim.us/special_TOC.html
[14] Yakuji Nippo Headline News. Aug 4, 2010 (in Japanese)
http://www.yakuji.co.jp/entry20132.html (last accessed Jan 21, 2011)

[15] Medicinal Plant Database (in Japanese) http://wwwts9.nibio.go.jp/mpdb.html
[16] OASIS Portal Site (in Korean) http://oasis.kiom.re.kr/portal/index.jsp
[17] Traditional Chinese Medicine Database System (mainly in Chinese) http://cowork.cintcm.com/engine/wdbintro1.jsp
[18] The Open Biological and Biomedical Ontology: http://www.obofoundry.org/
[19] Vocabulary of Interlinked Datasets: http://vocab.deri.ie/void
[20] Giles, J. Key biology databases go wiki, *Nature news* 445:691, 2007