JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Efficient Approach for Simultaneous Estimation of Multiple Health-Promoting Effects of Foods

Kiyoko Nagahama,[†] Nozomu Eto,[*,†,§] Kunihito Yamamori,[#] Kazuo Nishiyama,[§] Yoichi Sakakibara,[†,§] Takako Iwata,[†,⊗] Asuka Uchida,[†,⊥] Ikuo Yoshihara,[#] and Masahito Suiko[†,§]

[†]Miyazaki Prefectural Industrial Support Foundation, 16500-2 Higashi-Kaminaka, Sadowara-cho, Miyazaki 880-0303, Japan

[§]Department of Biochemistry and Applied Biosciences, Faculty of Agriculture, University of Miyazaki, 1-1 Gakuen Kibanadai-nishi, Miyazaki 889-2192, Japan

[#]Department of Computer Science and Systems Engineering, Faculty of Engineering, University of Miyazaki, 1-1 Gakuen Kibanadai-nishi, Miyazaki 889-2192, Japan

ABSTRACT: The investigation of new food constituents for purposes of disease prevention or health promotion is an area of increasing interest in food science. This paper proposes a new system that allows for simultaneous estimation of the multiple health-promoting effects of food constituents using informatics. The model utilizes expression data of intracellular marker proteins as descriptors that reply to stimulation of a constituent. To estimate three health-promoting effects, namely, cancer cell growth suppression activity, antiviral activity, and antioxidant stress activity, each model was constructed using expression data of marker proteins as input data and health-promoting effects as the output value. When prediction performances of three types of mathematical models constructed by simple, multiple regressions, or artificial neural network (ANN), were compared, the most adequate model was the one constructed using an ANN. There were no statistically significant differences between the actual data and estimated values calculated by the ANN models. This system was able to simultaneously estimate health-promoting effects with reasonable precision from the same expression data of marker proteins. This novel system should prove to be an interesting platform for evaluation of the health-promoting effects of food.

KEYWORDS: health-promoting effects, food constituents, artificial neural network, linear regression, marker protein

INTRODUCTION

Health-promoting effects of foods have recently attracted the attention of researchers because foods have the potential to prevent disease and promote health in ways not anticipated by traditional nutrition science. The potential of foods for disease prevention is supported by literature papers including a prospective cohort study in Japan in which the consumption of fruits was associated with a lower risk of cardiovascular disease[1] and a higher intake of total fruits and vegetables was associated with a dose-dependent decrease in the risk of esophageal squamous cell carcinoma.[2] In another epidemiologic study on fruit and vegetable intake and cancer risk, a meta-analytic approach to evaluate evidence from a case-control and prospective study suggests a protective effect of both fruits and vegetables toward reducing the risk of cancer.[3] Moreover, fruits and vegetables are rich in polyphenols, and their effects in the prevention of various diseases are widely supported by observations of epidemiological studies, experimental studies on animals or cultured human cell lines, and clinical studies.[4]

There are many health-promoting effects of food that should be evaluated, for example, antioxidative activity,[5,6] antimutagenic activity,[7−9] and cancer cell growth suppression activity.[10−12] In addition, some natural products and constituents have multiple health-promoting effects. For instance, catechins, contained in tea, are shown to possess multiple health-promoting effects including prevention of cancer[13] and cardiovascular disease,[14] antioxidant or pro-oxidant activities,[15] and antiobesity, antidiabetic,[16] and anti-inflammatory effects.[17] Thus, when food constituents are investigated for their potential effects, optimal approaches will include the capability to screen for multiple effects at once, thus saving both time and labor.

In the field of drug discovery, identification of new drug candidates employs the use of not only in vitro assays such as cell-based assays,[18] but also in silico assays such as computational methods.[19−22] In silico assays are useful for the assessment of pharmacokinetics, toxicity, and pharmacological effects. Also, they are beneficial for narrowing the number of compounds that need to be subjected to chemical assays, thus saving time and money. Computational approaches such as analysis of quantitative structure−activity relationships (QSAR)[23,24] and artificial neural networks (ANN)[25,26] have been applied to search for a food with a single health-promoting effect on a specific target molecule. To achieve this rapidly and cost effectively, a new screening system using an in silico assay is proposed to evaluate multiple health-promoting effects of food simultaneously.

MATERIALS AND METHODS

Food Constituents, Drugs, and Food Extracts. The tested food constituents were as follows: lipoic acid, curcumin, resveratrol, epigallocatechin-3-gallate (EGCG), arachidonic acid, epigallocatechin (EGC), kaempferol, chlorogenic acid, galangin, and linoleic acid (Sigma-Aldrich

Received: May 10, 2011
Revised: July 8, 2011
Accepted: July 11, 2011
Published: July 11, 2011

Table 1. Antibodies Used for ELISA

marker protein	capture antibody	detection antibody	secondary antibody
thioredoxin	mouse mAb[a] (Serotec, Oxford, U.K.)	goat pAb[b] (R&D Systems Inc., Minneapolis, MN)	mouse pAb, peroxidase conjugated (Pierce, Rockford, IL)
survivin	mouse mAb (established)	goat pAb (R&D Systems Inc.)	mouse pAb, peroxidase conjugated (Pierce)
FADD	rabbit pAb (USBiological, Swampscott, MA)	mouse mAb (BD Transduction Lab, San Diego, CA)	goat pAb, peroxidase conjugated (ICN Pharmaceuticals Inc., Aurora, OH)
TXNRD1	rabbit pAb (LabFrontier, Seoul, South Korea)	mouse mAb (Abcam Ltd., Cambridge, U.K.)	goat pAb, peroxidase conjugated (ICN Pharmaceuticals Inc.)
Hsp90	mouse mAb (BD Transduction Lab.)	goat pAb (Santa Cruz Biotechnology, Santa Cruz, CA)	mouse pAb, peroxidase conjugated (Pierce)
MxA	mouse mAb (KYOWA MEDEX Co.,Ltd., Tokyo, Japan)	biotinylated mouse mAb (established)	streptavidin, peroxidase conjugated (GE Healthcare Biosciences, Little Chalfont, U.K.)
tNOX	mouse mAb (established)	biotinylated mouse mAb (established)	streptavidin, peroxidase conjugated (GE Healthcare Biosciences)
NQO1	mouse mAb (Abnova, Taipei, Taiwan)	goat pAb (IMGENEX Co., San. Diego, CA)	mouse pAb, peroxidase conjugated (Pierce)
Hsp70	Duoset IC kit (R&D Systems Inc.)		
XIAP	Duoset IC kit (R&D Systems Inc.)		
ERK2	Duoset IC kit (R&D Systems Inc.)		
p53	Duoset IC kit (R&D Systems Inc.)		
Bcl-2	Duoset IC kit (R&D Systems Inc.)		
GAPDH	whole-cell normalization kit (Active Motif, Carlsbad, CA)		

[a] mAb, monoclonal antibody. [b] pAb, polyclonal antibody.

Co., St. Louis, MO); glycitein, quercetin, γ-aminobutyric acid (GABA), and capsaicin (Wako Pure Chemical Industries, Ltd., Osaka, Japan); *cis*-9,*trans*-11 conjugated linoleic acid (CLA), and *trans*-10,*cis*-12 CLA (Cayman Chemical Co., Ann Arbor, MI); cyanidin, pelargonidin, and delphinidin (Extrasynthèse, Genay, France); genistein (Wako Chemicals, USA, Inc., Dallas, TX); daidzein (Fujicco Co., Ltd., Kobe, Japan); rosmarinic acid (MP Biomedicals, Inc., Irvine, CA); and benzyl isothiocyanate (BITC) (Tokyo Chemical Industry Co., Ltd., Tokyo, Japan). The tested drugs were as follows: simvastatin, lovastatin, and pravastatin (Wako Pure Chemical Industries, Ltd.); fluvastatin and atorvastatin (Toronto Research Chemicals, Inc., North York, Canada); ribavirin (MP Biomedicals, Inc.); and interferon-α 2b (Prospec-Tany Technogene Ltd., Rehovot, Israel). Edible plants collected in Miyazaki, Japan, were as follows: leaves and roots of Japanese radish (*Raphanus sativus*); roots of burdock (*Arctium lappa*); leaves of carrot (*Daucus carota*), spearmint (*Mentha spicata*), rosemary (*Rosmarinus officinalis*), lemon balm (*Melissa officinalis*), stevia (*Stevia rebaudiana*), sweet basil (*Ocimum basilicum*), green tea (*Camellia sinensis*), and blueberry (*Vaccinium virgatum*). Each sample of lyophilized powder (1 g) was extracted by vortexing for 30 s with 80% ethanol (30 mL), whereas blueberry leaf powder was extracted with 80% ethanol or hot water. The extracts were filtered through filter paper (filter paper no. 2, Toyo, Tokyo, Japan), concentrated with a vacuum evaporator, and completely dried with a freeze-dryer. The lyophilized extracts were redissolved in dimethyl sulfoxide.

Cells. The human hepatocellular carcinoma cells HepG2 were maintained in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich Co.) supplemented with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin (Sigma-Aldrich Co.). HCV replicon cells, human hepatoma HuH-7 cells carrying a HCV subgenomic replicon derived from the genotype 1b isolate Con1,[27] were maintained in DMEM supplemented with Glutamax (Invitrogen, Carlsbad, CA), 10% FCS, 1% penicillin/streptomycin (Invitrogen), and 500 μg/mL of G418 (Invitrogen). HepG2/ARE cells, in which a luciferase-based ARE reporter construct derived from human NAD(P)H dehydrogenase [quinone] 1 (NQO1) was stably transfected, were established by modifications of the method of Boerboom et al.[28] The cells were maintained in DMEM (Sigma-Aldrich Co.) supplemented with 10% FCS, 1% penicillin/streptomycin (Sigma-Aldrich Co.), and 500 μg/mL of G418 (Sigma-Aldrich Co.). All cells were maintained in a humidified atmosphere containing 5% CO_2 at 37 °C.

Measurement of the Expression Levels of Marker Proteins. HepG2 cells were seeded into 100 mm tissue culture dishes at 3×10^5 cells/mL and incubated for 24 h. Subsequently, the medium was replaced with one containing a food constituent or extract. After exposure for 24 h, cells were washed with cold phosphate-buffered saline (PBS), solubilized with ice-cold lysis buffer (1 mM EDTA, 0.005% Tween 20, and 0.5% Triton-X-100 in PBS) containing protease inhibitor cocktail (Roche, Basel, Switzerland). The protein concentrations of the cell lysates were measured by DC protein assay (Bio-Rad Laboratories, Hercules, CA), and then the sample was set to 1 mg/mL. Then, the 14 kinds of marker proteins were measured by sandwich enzyme-linked immunosorbent assay (ELISA) to detect the intracellular response generated by the food constituent or extract. Measured proteins were as follows: thioredoxin, survivin, heat shock protein 70 (Hsp70), X-linked inhibitor of apoptosis protein (XIAP), Fas-associated death domain protein (FADD), thioredoxin reductase 1 (TXNRD1), heat shock protein 90 (Hsp90), IFN-inducible antiviral protein Mx (MxA), tumor-associated hydroquinone oxidase (tNOX), NQO1, tumor suppressor p53 (p53), extracellular signal-regulated kinase 2 (ERK2), B-cell lymphoma 2 (Bcl-2), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Table 1 shows the kits or combinations of antibodies used in sandwich ELISA to detect each marker protein. Microtiter plates (96 well) were coated with 100 μL of capture antibody in 50 mM carbonate buffer (pH 9.6) and incubated overnight at 4 °C. The plates were blocked with 300 μL of 5% skim milk in PBS and incubated overnight at 4 °C. Subsequently, the cell lysates (100 μL) were added into each well and incubated for 2 h at 37 °C. All of the cell lysates were analyzed in duplicate. The plates were incubated with 100 μL of detection antibody

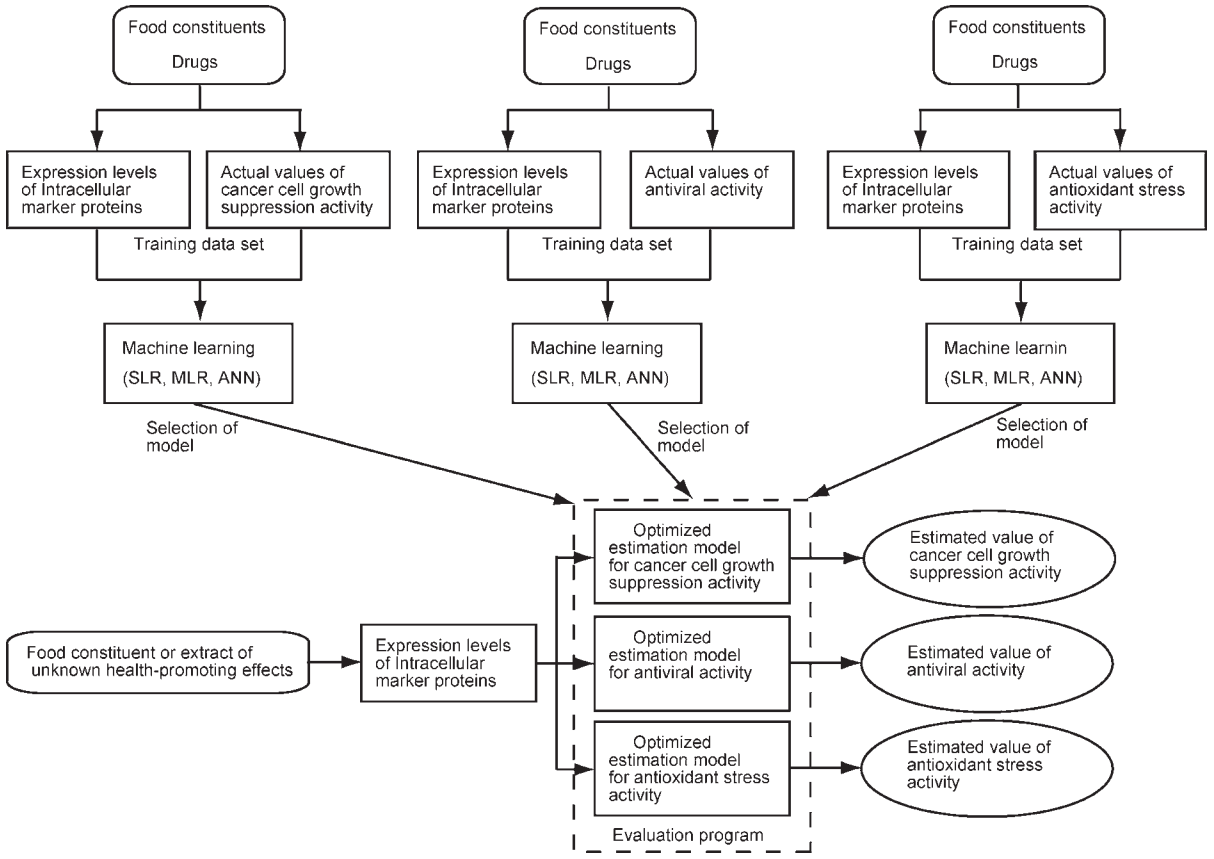

Figure 1. Conceptual diagram of the evaluation system to estimate plural health-promoting effects of food constituents from expression levels of intracellular marker proteins. Training data sets were collected on activity values of health-promoting effect and expression level of intracellular proteins by using the same food constituents. Then estimation models were constructed using health-promoting effect as the dependent variable or output value and relative expression of 13 kinds of marker proteins as independent variables or input values. Subsequently, the optimized model was selected in all of the built models. Finally, activity values of plural health-promoting effects can be estimated simultaneously from the expression data of marker proteins by the evaluation system.

for 1 h at 37 °C, followed by incubation with a secondary antibody. The peroxidase reaction was initiated by the addition of 100 μL of substrate solution [0.1 M citrate buffer (pH 4.0) containing 0.003% H_2O_2, and 0.3 mg/mL 2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonic acid) (Wako)]. After 10 min, the absorbance was measured at 405 and 490 nm with a multichannel microtiter plate reader (Vmax; Molecular Devices, Redwood City, CA). ELISA data were normalized by comparison with the absorbance of GAPDH. The relative marker protein expression level was represented by a relative value in comparison with the control.

Determination of Cancer Cell Growth Suppression Activity. Cancer cell growth suppression activities for food constituents and extracts were determined by a cell proliferation assay. HepG2 cells were inoculated into a 96-well microtiter plate (Corning) at 1×10^4 cells/well and cultured for 24 h. The cells were incubated in the presence or absence of food constituents or extracts. After 48 h of incubation, the cell survival rate was measured by the WST-8 [2-(2-methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2*H*-tetrazolium, monosodium salt] cell-counting kit (Dojindo, Kumamoto, Japan) according to the manufacturer's protocol. The WST-8 assay is based on a water-soluble formazan reaction occurring only in viable cells.[29] Cancer cell growth suppression activity was calculated by using the following equation: cancer cell proliferation rate = (absorbance of test)/(absorbance of control). The absorbance in the microplate wells was measured at 450 nm with a multichannel microtiter plate reader (Vmax; Molecular Devices).

Measurement of Antiviral Activity. Antiviral activity was evaluated by measuring the impact on suppression of HCV replication using a HCV subgenomic replicon system.[27] The HCV replicon cells in DMEM supplemented with Glutamax and 5% FCS (5×10^3 cells/well) were plated in 96-well plates and were cultured for 24 h. Then, the cells were treated with each food constituent or extract for 72 h. HCV replicon levels were determined by the luciferase reporter assay, and the cell proliferation rate was determined by the WST-8 assay. The luciferase activity was quantified using the Steady-Glo luciferase assay system (Promega, Madison, WI) according to the manufacturer's protocol, and the luminescence was measured by the DTX 800 multimode detector (Beckman Coulter, Fullerton, CA). The relative luciferase activity was represented by the relative value in comparison with the control. Antiviral activity was defined as follows: viral replication rate = (relative luciferase activity)/(cell proliferation rate).

Determination of Antioxidant Stress Activity. Antioxidant stress activity was evaluated by induction of ARE-mediated gene expression using a reporter assay.[28] The HepG2/ARE cells were inoculated into 96-well plates at 4×10^4 cells/well. After 24 h of incubation, the cells were treated with each food constituent or extract for 24 h. The luciferase activity was quantified using the Bright-Glo luciferase assay system (Promega) according to the manufacturer's protocol, and the luminescence was measured by using a multiplate reader (GENios, Tecan Japan Co. Ltd., Japan). Luciferase activity and cell proliferation rate were calculated by using the same method as for antiviral activity. Antioxidant stress activity was calculated as follows: ARE-luciferase activity rate = (relative luciferase activity)/(cell proliferation rate).

Table 2. Training Data Set 1: Marker Protein Expression Rate and Experimental Values of Health-Promoting Effects Used To Build the Models

		marker protein expression rate													health-promoting effects		
compd	concn (μM)	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	12[l]	13[m]	A[n]	B[o]	C[p]
lipoic acid	100	0.613	0.664	0.720	0.592	0.496	0.696	0.580	0.667	0.428	0.565	0.679	0.628	0.626	0.990	0.8957	0.834
	300	0.707	0.630	0.772	0.675	0.447	0.896	0.572	0.836	0.500	0.763	0.736	0.625	0.696	1.041	0.4827	1.251
	1000	0.966	0.627	0.773	0.999	0.594	1.164	0.699	1.300	1.310	1.460	0.850	0.691	1.009	0.998	0.1178	2.711
EGCG	7	0.878	0.881	0.723	0.768	0.872	0.781	0.714	0.771	0.569	0.967	0.925	0.743	0.762	0.944	1.3515	0.630
	20	0.882	0.734	0.589	0.656	0.671	0.694	0.399	0.622	0.446	0.902	0.868	0.609	0.625	0.905	0.9307	0.740
	50	1.605	1.119	0.731	1.207	1.008	0.968	0.614	1.070	1.047	1.475	1.300	0.974	0.990	0.870	0.0354	0.672
daidzein	25	0.632	0.627	0.712	0.654	1.121	0.973	0.481	0.672	0.519	0.689	0.735	0.674	0.665	0.958	1.1245	2.891
	50	0.528	0.552	0.627	0.577	0.879	1.217	0.514	0.707	0.594	0.848	0.803	0.755	0.702	0.906	1.0149	2.519
	150	0.737	0.531	0.717	0.830	0.968	1.153	0.571	0.748	0.717	0.945	0.858	0.942	0.868	0.832	0.5375	1.806
glycitein	10	0.882	0.932	0.821	0.871	1.016	0.935	0.788	0.856	0.555	1.004	0.968	0.820	0.962	1.086	0.8283	0.995
	30	0.888	0.739	0.684	0.935	1.256	0.834	0.796	0.788	0.433	0.929	0.954	0.720	0.939	1.044	1.1602	1.476
	100	1.652	0.604	0.802	1.343	1.850	0.773	0.679	1.082	0.759	1.129	1.242	1.029	1.379	0.966	0.9859	1.801
quercetin	5	0.844	0.956	0.834	0.780	1.136	0.908	1.109	0.824	0.686	0.922	0.884	0.853	0.838	1.109	0.9532	0.860
	15	0.917	0.871	0.741	0.725	0.870	0.810	1.127	0.783	0.757	0.920	0.878	0.843	0.823	1.063	NT[q]	0.733
	60	1.417	1.158	0.674	0.808	0.989	0.604	0.629	0.761	0.873	1.019	0.935	0.980	0.686	0.904	0.4172	2.871
cyanidin	40	0.633	0.749	0.847	0.726	1.134	0.891	0.600	0.742	0.640	0.744	0.798	0.833	0.784	0.983	1.2056	0.538
	150	0.536	0.738	0.673	0.658	0.903	0.683	0.459	0.678	0.632	0.854	0.879	0.856	0.739	0.822	0.0391	0.450
	400	0.598	0.776	0.784	0.669	0.884	0.797	0.633	0.820	0.631	0.720	0.836	1.094	0.854	0.451	NT	0.538
pelargonidin	100	0.784	0.853	0.983	0.848	1.085	0.916	0.909	0.820	0.529	1.069	0.984	0.756	0.913	1.000	0.7858	0.402
	250	1.112	0.907	1.249	0.849	1.150	1.080	1.027	0.879	0.633	1.440	1.310	1.022	1.016	1.025	0.4207	NT
	800	1.367	0.617	1.508	0.862	1.071	0.875	0.892	0.571	0.861	1.845	1.411	1.149	1.067	0.902	NT	2.330
delphinidin	15	1.032	0.956	0.995	1.050	1.083	1.020	0.816	1.017	0.521	1.066	1.125	0.924	0.953	0.966	1.0568	0.827
	70	1.216	0.871	0.902	1.169	1.256	0.931	0.882	0.842	0.487	0.970	1.163	0.976	0.847	0.847	0.5343	0.717
	200	1.251	0.847	0.865	1.238	1.078	0.863	0.753	0.519	0.513	1.132	1.445	1.206	0.614	0.528	NT	1.930
curcumin	4	0.952	0.952	1.003	0.824	1.554	1.400	0.672	0.989	0.776	0.950	0.944	0.977	0.894	1.009	1.1212	1.741
	15	1.017	0.685	0.981	0.908	1.288	1.091	0.768	0.894	0.871	1.072	0.999	0.864	0.819	0.994	NT	NT
	40	2.934	0.692	2.250	1.122	1.963	1.373	0.800	0.898	0.900	1.377	1.155	1.015	0.839	1.021	NT	2.094
GABA	100	0.995	1.255	0.911	1.021	1.390	1.190	1.209	1.064	0.743	0.951	1.177	0.964	1.035	1.002	1.0369	1.247
	300	1.143	1.222	0.967	1.085	1.337	1.107	1.058	1.143	0.697	0.778	1.322	1.011	1.177	1.006	1.1237	0.941
	1000	1.214	1.185	0.981	1.038	1.326	1.009	1.223	1.391	0.978	0.803	1.305	1.059	1.305	1.032	1.0846	0.788
resveratrol	10	0.724	0.790	0.797	0.851	0.785	0.979	0.622	0.731	0.406	1.049	1.017	0.893	0.914	0.960	NT	2.778
	30	0.598	0.778	0.913	0.841	0.705	0.931	0.460	0.647	0.305	1.010	1.009	1.041	0.880	0.858	NT	2.310
	80	0.694	0.559	0.946	1.093	0.784	0.727	0.573	0.605	0.673	1.067	1.104	2.406	0.982	0.563	NT	1.172
arachidonic acid	15	1.016	1.041	0.982	0.874	1.139	0.940	1.086	0.891	0.572	0.903	1.128	1.013	0.912	1.030	0.7954	1.277
	45	1.294	1.077	1.060	1.007	0.777	1.255	1.022	0.860	0.596	0.940	1.412	1.151	1.019	0.932	0.4984	1.073
	100	1.718	0.895	1.032	1.121	1.091	1.111	0.981	0.699	0.823	1.015	1.566	1.199	1.005	0.806	0.0378	0.727
CLA12C	1	0.900	1.016	1.076	0.887	1.153	0.935	1.058	0.823	0.629	0.855	0.891	0.761	0.865	0.959	NT	1.203
	3	0.929	1.090	1.132	0.949	1.799	1.009	1.012	0.854	0.987	0.875	0.961	0.931	0.951	0.962	0.7944	1.111

dx.doi.org/10.1021/jf201836g |J. Agric. Food Chem. 2011, 59, 8575–8588

Table 2. Continued

		marker protein expression rate													health-promoting effects		
compd	concn (μM)	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	12[l]	13[m]	A[n]	B[o]	C[p]
	10	0.953	1.002	1.021	0.951	1.318	0.986	0.923	1.061	1.048	0.899	0.949	0.926	1.011	0.965	0.1139	1.015
CLA9C	10	1.032	0.887	0.995	1.054	1.183	0.871	1.198	0.883	0.727	0.909	0.974	0.955	0.837	0.898	0.5918	0.552
	30	1.054	0.940	0.945	1.026	1.462	0.879	1.151	0.822	0.888	0.860	1.003	0.866	0.917	0.891	0.1460	0.638
	100	0.965	0.886	0.842	1.008	0.878	0.865	1.033	0.900	1.058	0.919	0.986	0.826	0.867	0.827	0.0358	0.935
kaempferol	6	0.967	0.987	0.827	0.939	1.349	1.162	0.900	1.003	0.943	1.328	1.376	0.977	0.917	0.981	NT[q]	0.954
	20	1.047	0.998	0.838	0.949	1.466	1.294	0.710	0.983	1.040	1.312	1.468	1.024	0.963	0.928	NT	2.411
	60	1.161	0.920	0.744	1.018	1.439	1.187	0.722	1.026	1.110	0.980	1.420	1.203	1.004	0.686	NT	2.924
IFN	100 IU/mL	0.921	0.870	0.843	0.996	1.424	1.240	0.863	2.603	0.665	1.158	1.145	0.898	1.006	1.014	0.0032	0.990
	300 IU/mL	1.141	1.176	0.921	1.128	2.165	1.564	1.049	5.268	0.736	1.216	1.269	1.062	1.106	1.005	0.0003	1.433
	1000 IU/mL	1.184	1.104	0.946	1.220	1.660	1.504	0.914	6.962	0.973	1.543	1.136	1.054	1.256	1.000	0.0001	1.030
ribavirin	2 μg/mL	1.066	0.903	0.597	0.896	0.966	0.727	0.891	0.730	0.874	1.061	1.025	0.944	1.066	1.043	NT	1.583
	10 μg/mL	1.096	0.626	0.549	0.776	1.045	0.655	0.759	0.627	1.009	1.173	1.137	1.475	0.935	0.971	0.4728	2.217
	30 μg/mL	1.186	0.479	0.523	0.759	1.180	0.602	0.536	0.800	1.109	1.037	0.939	1.259	0.884	0.926	0.0999	2.282
fluvastatin	7.5	1.006	0.796	0.878	1.019	1.220	0.823	0.709	0.820	0.407	0.969	0.844	0.929	0.868	0.933	0.1470	0.973
	15	1.014	0.731	0.741	0.914	1.352	0.679	0.497	0.706	0.274	0.876	0.713	0.735	0.725	0.921	0.0641	0.845
	50	1.250	0.715	0.825	1.030	1.262	0.683	0.556	0.844	0.512	0.916	0.809	0.895	0.902	0.788	NT	0.728
atorvastatin	3.5	0.775	0.612	0.859	0.811	0.970	0.834	0.720	0.802	0.511	0.982	1.004	0.733	0.742	1.075	0.5317	1.675
	10	0.910	0.707	0.947	1.061	0.926	0.894	0.655	0.861	0.636	1.229	1.240	0.855	0.812	1.027	0.0386	1.396
	35	0.942	0.641	0.982	1.095	0.559	0.892	0.715	0.987	0.860	1.192	1.247	1.063	0.943	0.762	NT	1.076
simvastatin	3.5	1.087	0.916	0.869	0.839	1.334	0.853	0.912	0.799	0.885	1.169	1.014	0.765	0.799	1.028	0.8126	1.144
	10	1.099	0.886	0.848	0.833	0.969	0.682	0.695	0.826	0.635	1.235	1.096	0.907	0.835	0.949	0.0516	1.003
	35	1.590	0.902	1.214	0.938	1.830	1.076	0.673	1.180	1.059	1.236	1.279	1.029	1.019	0.586	NT	0.822
pravastatin	100	1.122	0.820	0.957	0.935	1.126	0.876	0.883	0.915	0.688	1.260	1.208	0.900	0.822	0.926	NT	1.586
	300	1.132	0.719	0.903	0.875	0.848	0.781	0.721	0.822	0.620	1.240	1.249	0.856	0.786	0.963	0.8990	1.117
	1000	1.212	0.690	0.852	0.981	0.586	0.757	0.666	0.807	0.841	1.230	1.263	0.954	0.787	0.938	0.0402	0.856
chlorogenic acid	20	0.786	0.924	0.716	0.808	1.244	0.969	1.191	0.851	0.472	0.855	0.961	0.810	0.806	0.971	1.0075	0.950
	70	0.889	0.981	0.670	0.826	1.513	1.066	1.326	0.985	0.497	0.877	1.076	0.871	0.888	1.001	NT	0.810
	200	1.053	1.025	0.794	0.917	1.840	0.998	1.378	1.176	0.785	0.937	1.107	1.054	1.049	0.924	0.5526	0.733
rosmarinic acid	5	0.638	0.673	0.688	0.759	0.604	0.827	0.864	0.822	0.554	0.739	0.881	0.673	0.859	0.949	1.2471	0.771
	15	0.550	0.672	0.653	0.885	0.659	0.900	0.795	0.877	0.489	0.757	0.958	0.705	0.855	0.982	1.0712	0.604
	50	0.509	0.733	0.587	0.699	0.684	0.780	0.817	0.819	0.729	0.771	0.886	0.774	0.830	0.993	1.3326	0.550
galangin	8	0.856	0.911	0.956	0.998	1.010	1.143	0.626	1.126	0.624	1.023	1.237	0.932	1.105	0.967	1.6717	1.885
	15	0.670	0.932	0.840	0.961	1.075	1.084	0.629	1.077	0.662	1.052	1.344	0.954	1.096	0.958	1.6554	1.686
	50	0.823	0.703	0.743	1.006	0.818	0.876	0.550	1.065	1.067	1.555	1.374	0.940	1.174	0.665	NT	1.252
capsaicin	10	0.833	1.229	1.009	0.909	3.029	1.147	1.023	1.109	0.793	1.461	1.180	0.937	1.081	0.933	0.6250	0.922
	60	1.113	1.246	1.066	0.957	2.759	1.144	1.181	1.131	0.847	1.530	1.272	1.018	1.131	0.860	0.6405	0.947
	150	1.518	0.837	0.987	0.952	2.602	1.133	1.252	1.024	1.157	1.665	1.266	1.177	1.357	0.633	NT	0.567
BITC	1.5	1.104	1.089	1.204	0.836	4.098	1.310	1.352	1.113	0.646	1.683	1.346	1.054	1.127	0.925	0.9488	1.271
	5	1.459	1.265	2.104	0.915	6.846	1.702	1.639	1.442	0.786	2.615	1.747	1.404	1.484	0.912	0.3782	2.713
	15	2.194	1.175	5.002	0.757	4.421	1.440	1.730	0.961	1.051	2.208	1.473	1.427	1.351	0.742	NT	NT

dx.doi.org/10.1021/jf201836g |J. Agric. Food Chem. 2011, 59, 8575–8588

Table 2. Continued

compd	concn (μM)	marker protein expression rate													health-promoting effects		
		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	12[l]	13[m]	A[n]	B[o]	C[p]
linoleic acid	20	0.942	1.028	1.102	1.036	1.249	0.941	1.458	0.863	0.682	0.824	0.928	0.714	0.805	0.935	NT	1.199
	50	0.970	0.920	1.010	0.987	1.441	1.027	1.476	0.851	0.834	0.912	1.019	0.664	1.051	0.945	0.7219	1.281
	150	1.080	0.943	0.942	1.011	1.403	0.995	1.524	0.958	1.004	1.013	1.028	0.586	1.198	0.829	NT	2.901

[a] 1, thioredoxin. [b] 2, survivin. [c] 3, Hsp70. [d] 4, XIAP. [e] 5, FADD. [f] 6, TXNRD1. [g] 7, Hsp90. [h] 8, MxA. [i] 9, tNOX. [j] 10, NQO1. [k] 11, ERK2. [l] 12, p53. [m] 13, Bcl-2. [n] A, cancer cell growth suppression activity. [o] B, antiviral activity. [p] C, antioxidant stress activity. [q] NT, not tested.

Design of the Evaluation System. A conceptual diagram of the evaluation system is shown in Figure 1. Three effects were chosen to estimate the plural health-promoting effects of the food constituents, namely, cancer cell growth suppression activity, antiviral activity, and antioxidant stress activity. Mathematically, these estimations were carried out by single linear regression (SLR) or multiple linear regression (MLR) or by using an ANN as the nonlinear model.

Data Sets. The data sets used in this study consisted of 23 food constituents, 7 drugs, and 12 food extracts (Tables 2 and 3). These data were divided into two groups, that is, training and test data sets. To construct the mathematical models, the training and test data sets were used to build the model and to verify the performance of the models, respectively. In addition, two kinds of training data sets were prepared.

Training Data Set 1. Training data set 1 consisted of 21 food constituents and 6 drugs (Table 2). The SLR and MLR were constructed by training data set 1, and ANN1 was learned by the same data set.

Training Data Set 2. ANN2 was constructed to improve the prediction performance of ANN1 and learned by the data generated from training data set 1 in Table 2. To avoid overfitting to training data set 1, the pseudorandom numbers were generated by the random function of the GNU Compiler Collection and then were added to training data set 1. A different size of noise was randomly and individually added to the relative expression of marker proteins and actual value of the health-promoting effect. Therefore, the additive noise was assumed to mimic the experimental error. Briefly, when the cells were treated with a certain concentration of food constituent, i, let m_j^i be the relative expression of marker protein j and let a_k^i be the actual value of the health-promoting effect k. Then, a matrix illustrated by eq 1 was created. Here, N is the number of all combinations of food constituents and concentration.

$$\begin{bmatrix} m_1^1 & \cdots & m_j^1 & \cdots & m_{13}^1 & a_k^1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^i & \cdots & m_j^i & \cdots & m_{13}^i & a_k^i \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^N & \cdots & m_j^N & \cdots & m_{13}^N & a_k^N \end{bmatrix} \quad (1)$$

Next, the average of each column of eq 1 was calculated using eqs 2 and 3, and a matrix illustrated by eq 4 was created.

$$\bar{m}_j = \frac{1}{N}\sum_{i=1}^{N} m_j^i \quad (2)$$

$$\bar{a}_k = \frac{1}{N}\sum_{i=1}^{N} a_k^i \quad (3)$$

$$\begin{bmatrix} \bar{m}_1 & \cdots & \bar{m}_j & \cdots & \bar{m}_{13} & \bar{a}_k \end{bmatrix} \quad (4)$$

Then, a noise to be added to $\bar{m}_j$ and $\bar{a}_k$ was generated. The noise, which generated against relative expression of marker proteins, ε_j, is uniform distribution random numbers that satisfy $-\alpha\bar{m}_j \le \varepsilon_j \le +\alpha\bar{m}_j$. Similarly, the noise for the actual value of the health-promoting effects, ε_k', is uniform distribution random numbers that satisfy $-\alpha\bar{a}_k \le \varepsilon_k' \le +\alpha\bar{a}_k$. Here, α, which satisfies $0 \le \alpha \le 0.1$ as cancer cell growth suppression and antioxidant stress activity or $0 \le \alpha \le 0.05$ as antiviral activity, is a constant. After some preliminary test runs, α was determined to optimize performance. When the data of P for each certain concentration of food constituent were synthesized, ε_j^p or ε'^p_k was the pth noise. The training data sets including the noise illustrated by eq 5 were created by repetition processing by the addition of the random numbers to m_j^i or a_k^i.

$$\begin{bmatrix} m_1^1 + \varepsilon_1^1 & \cdots & m_j^1 + \varepsilon_j^1 & \cdots & m_{13}^1 + \varepsilon_{13}^1 & a_k^1 + \varepsilon'^1_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^1 + \varepsilon_1^p & \cdots & m_j^1 + \varepsilon_j^p & \cdots & m_{13}^1 + \varepsilon_{13}^p & a_k^1 + \varepsilon'^p_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^1 + \varepsilon_1^P & \cdots & m_j^1 + \varepsilon_j^P & \cdots & m_{13}^1 + \varepsilon_{13}^P & a_k^1 + \varepsilon'^P_k \\ m_1^2 + \varepsilon_1^1 & \cdots & m_j^2 + \varepsilon_j^1 & \cdots & m_{13}^2 + \varepsilon_{13}^1 & a_k^2 + \varepsilon'^1_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^2 + \varepsilon_1^P & \cdots & m_j^2 + \varepsilon_j^P & \cdots & m_{13}^2 + \varepsilon_{13}^P & a_k^2 + \varepsilon'^P_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^N + \varepsilon_1^1 & \cdots & m_j^N + \varepsilon_j^1 & \cdots & m_{13}^N + \varepsilon_{13}^1 & a_k^N + \varepsilon'^1_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ m_1^N + \varepsilon_1^P & \cdots & m_j^N + \varepsilon_j^P & \cdots & m_{13}^N + \varepsilon_{13}^P & a_k^N + \varepsilon'^P_k \end{bmatrix} \quad (5)$$

Test Data Sets. The unused data sets for training, that is, 2 food constituents, a drug, and 12 food extracts, were used for testing the performance of the SLR, MLR, ANN1, and ANN2 (Table 3). Unlike the generated training data set 2, the test data sets did not add noise. To test whether a wide range of activities in compounds can be estimated by the mathematical model, two food constituents (genistein and EGC) and a drug (lovastatin) were used. Genistein has a cancer cell growth suppression activity[12,30−34] and a defense effect against oxidative stress.[35−43] Lovastatin[44−47] has suppression activity against HCV replication. EGC was chosen as compound, which exhibited weak activity against the three health-promoting effects. Food extracts were used to test for crude samples.

Linear Regression Model. Linear regression models were used to build models to estimate the three health-promoting effects from intracellular protein expression in response to stimulation with food constituents. Simple and multiple linear regressions are the most widely known modeling methods. The linear regressions relate one dependent variable y to one (SLR) or several (MLR) independent variables x_i by eq 6

$$y = \alpha + \sum_{i=1}^{n} \beta_i x_i \quad (6)$$

Table 3. Test Data Sets: Marker Protein Expression Rate and Experimental Values of Health-Promoting Effects Used To Verify the Models and Estimated Values Calculated from the Optimized ANN2 Models

		marker protein expression rate													health-promoting effects					
															actual value			estimated value from ANN2		
food extract	concn (μg/mL)	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	12[l]	13[m]	A[n]	B[o]	C[p]	A[n]	B[o]	C[p]
blueberry leaves (HW[q])	5	0.913	1.128	0.966	0.811	0.967	0.997	0.898	0.809	0.784	1.133	0.919	0.979	0.934	0.948 ± 0.032	0.155 ± 0.031	0.662 ± 0.090	1.048	0.164	0.937
blueberry leaves (EtOH[r])	5	0.722	1.004	0.820	0.790	1.007	0.993	0.990	0.924	0.878	1.245	0.857	0.913	0.931	0.926 ± 0.058	0.023 ± 0.025	0.766 ± 0.219	1.046	0.179	1.148
Japanese radish roots	1000	0.827	0.949	0.896	0.875	1.113	1.034	0.846	0.738	0.722	0.929	0.929	0.917	0.888	0.905 ± 0.018	0.459 ± 0.148	1.173 ± 0.077	1.045	0.329	0.950
Japanese radish leaves	1000	1.718	1.149	1.171	1.061	1.209	0.966	1.084	1.116	1.039	1.266	1.323	1.096	0.847	0.559 ± 0.007	0.336 ± 0.184	2.233 ± 0.236	0.639	0.072	1.605
carrot leaves	400	3.240	1.288	1.604	0.977	1.363	1.182	1.276	0.620	0.666	1.703	2.373	1.653	0.665	0.873 ± 0.041	0.746 ± 0.295	1.401 ± 0.189	0.890	0.702	2.221
green tea leaves	40	0.932	1.091	1.132	0.952	1.740	1.125	1.133	0.918	0.750	1.186	1.307	1.186	0.944	0.902 ± 0.009	0.019 ± 0.012	0.914 ± 0.247	0.953	0.392	0.936
	120	1.282	1.198	1.237	1.098	1.627	1.211	0.952	1.036	1.085	1.457	1.566	1.412	1.082	0.738 ± 0.005	0.058 ± 0.067	0.921 ± 0.562	0.853	0.112	0.955
burdock roots	300	0.867	0.940	0.868	0.890	0.832	1.126	1.052	0.834	0.828	1.109	0.835	0.902	0.911	0.990 ± 0.004	0.054 ± 0.023	0.882 ± 0.152	1.047	0.207	0.962
	1000	1.006	0.967	0.844	0.873	0.874	1.126	1.245	0.812	0.837	1.252	0.784	0.920	0.838	0.941 ± 0.014	0.010 ± 0.011	1.877 ± 0.173	1.042	0.084	2.201
spearmint leaves	100	1.016	1.237	1.077	0.884	1.185	0.999	1.092	0.827	0.865	1.085	1.032	1.261	0.896	0.961 ± 0.016	0.318 ± 0.094	0.617 ± 0.178	1.009	0.159	0.935
rosemary leaves	50	1.567	0.598	1.653	1.363	1.352	1.345	1.303	0.869	1.267	1.206	1.150	0.905	1.115	0.768 ± 0.034	0.668 ± 0.481	2.666 ± 0.547	0.698	0.527	2.457
lemon balm leaves	300	1.166	1.392	1.055	1.162	1.480	1.118	1.623	0.789	0.927	1.181	1.412	1.491	0.953	0.867 ± 0.058	0.374 ± 0.170	1.144 ± 0.100	0.886	0.073	0.936
stevia leaves	60	0.939	0.978	0.928	0.969	1.015	1.075	1.013	1.120	0.974	0.917	1.061	1.026	1.063	1.001 ± 0.026	0.477 ± 0.143	0.631 ± 0.084	0.947	0.457	0.933
	200	1.119	1.185	1.067	1.126	1.101	1.205	1.032	1.045	1.076	1.054	1.232	1.256	0.850	0.951 ± 0.012	0.158 ± 0.055	0.743 ± 0.053	0.877	0.081	0.938
sweet basil leaves	120	0.953	1.164	0.952	1.060	1.351	1.026	1.159	0.876	0.893	0.909	1.087	1.038	0.907	0.963 ± 0.020	0.068 ± 0.035	0.935 ± 0.084	0.920	0.128	0.933
	400	1.514	2.120	1.561	1.124	2.338	1.395	1.948	0.952	1.150	1.469	1.632	1.821	1.189	0.703 ± 0.032	0.086 ± 0.084	2.281 ± 0.202	0.857	0.072	1.340
lovastatin	5 μM	1.018	0.836	1.056	0.917	0.593	0.960	0.835	0.908	0.701	1.192	1.135	0.897	0.877	1.022 ± 0.044	0.799 ± 0.228	0.964 ± 0.234	1.038	0.616	1.133
	25 μM	1.046	0.642	1.069	0.933	0.754	0.847	0.734	0.850	0.675	1.332	1.287	1.000	0.876	0.964 ± 0.032	0.032 ± 0.011	0.899 ± 0.051	0.977	0.113	1.146
genistein	20 μM	0.832	0.702	0.711	0.859	0.655	1.009	0.824	1.041	0.655	1.096	1.006	1.131	1.033	0.832 ± 0.023	1.527 ± 0.095	1.772 ± 0.401	0.955	1.478	1.152
	60 μM	1.412	1.533	0.853	1.357	0.774	1.364	0.912	1.520	1.919	1.866	1.354	3.824	1.767	0.582 ± 0.015	1.021 ± 0.187	2.392 ± 0.517	0.531	1.137	2.261
EGC[s]	30 μM	1.066	1.012	0.798	1.086	1.207	1.068	1.026	1.053	0.695	1.154	1.649	1.106	1.045	0.972 ± 0.020	1.001 ± 0.153	0.575 ± 0.085	0.884	0.657	0.865
	60 μM	0.878	1.016	0.752	1.183	0.881	0.929	0.810	1.147	1.107	0.906	1.282	1.232	1.104	1.007 ± 0.027	0.489 ± 0.098	0.674 ± 0.308	0.996	1.088	0.803

[a] 1, thioredoxin. [b] 2, survivin. [c] 3, Hsp70. [d] 4, XIAP. [e] 5, FADD. [f] 6, TXNRD1. [g] 7, Hsp90. [h] 8, MxA. [i] 9, tNOX. [j] 10, NQO1. [k] 11, ERK2. [l] 12, p53. [m] 13, Bcl-2. [n] A, cancer cell growth suppression activity. [o] B, antiviral activity. [p] C, antioxidant stress activity. [q] HW, hot water extract. [r] EtOH, 80% ethanol extract. [s] EGC, epigallocatechin.

dx.doi.org/10.1021/jf201836g |J. Agric. Food Chem. 2011, 59, 8575–8588

Table 4. Regression Coefficients and Statistics of the Fits Obtained from SLR Models

	cancer cell growth suppression activity				antiviral activity				antioxidant stress activity			
variable	intercept	regression coeff	R^2 [a]	RMSE[b]	intercept	regression coeff	R^2	RMSE	intercept	regression coeff	R^2	RMSE
TXN	0.965	−0.045	0.016	0.125	1.353	−0.710	0.171**[c]	0.419	1.032	0.273	0.019	0.684
survivin	0.856	0.073	0.012	0.126	0.667	−0.015	0.000	0.460	1.761	−0.520	0.021	0.684
Hsp70	0.951	−0.033	0.019	0.125	0.834	−0.205	0.010	0.458	0.985	0.362	0.019	0.684
XIAP	1.043	−0.135	0.026	0.125	1.411	−0.833	0.083*	0.441	1.378	−0.074	0.000	0.691
FADD	0.936	−0.012	0.008	0.126	0.699	−0.034	0.005	0.459	1.164	0.112	0.019	0.684
TXNRD1	0.885	0.035	0.004	0.126	0.674	−0.020	0.000	0.460	0.617	0.714	0.052*	0.673
Hsp90	0.882	0.043	0.010	0.126	0.674	−0.023	0.000	0.460	1.588	−0.322	0.017	0.685
MxA	0.899	0.019	0.017	0.125	0.797	−0.129	0.080*	0.442	1.341	−0.029	0.001	0.690
tNOX	1.005	−0.118	0.041	0.124	1.219	−0.800	0.126**	0.430	0.858	0.621	0.038	0.677
NQO1	0.988	−0.064	0.027	0.125	1.146	−0.468	0.101*	0.436	0.666	0.601	0.071*	0.666
ERK2	1.068	−0.136	0.055*	0.123	0.901	−0.232	0.012	0.458	0.772	0.496	0.024	0.682
p53	1.152	−0.246	0.224***	0.111	1.289	−0.699	0.073*	0.443	0.971	0.360	0.016	0.685
Bcl2	0.987	−0.073	0.010	0.126	0.658	−0.004	0.000	0.460	0.964	0.371	0.009	0.688

[a] R^2, coefficient of determination. [b] RMSE, root-mean-square error. [c] *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

where y, α, β_i, and x_i represent the dependent variable, intercept, regression coefficient, and independent variable, respectively. The models were constructed using health-promoting effect as the dependent variable and relative expression of marker protein as independent variable (Table 2). The goodness-of-fit was evaluated by a coefficient of determination (R^2) and root-mean-square error (RMSE). The RMSE was calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_{pred} - y_{obs})^2}{n}} \quad (7)$$

where y_{pred}, y_{obs}, and n represent estimated value, actual value, and number of samples, respectively. The F test was used to test the significance of the established regression equations, and $p < 0.05$ was accepted as significant. SLR and MLR analyses were performed by the regression analysis tool of Microsoft Excel 2003.

ANN Model. ANN is a computerized mathematical model designed to emulate the architecture of the brain, and it is a powerful nonlinear modeling technique offering solutions to problems that have not been clearly formulated. A multilayer feedforward network with a backpropagation learning algorithm was used.[48] Every neuron in each layer is connected to every neuron of the adjacent layer by weighted links. To estimate with the least possible error, these weights must be adjusted. The activation function, eq 8, means that the input, y_j, into a neuron is multiplied by its corresponding connection weights, w_{jk}, and summed. Then, the sum, x_k, is transformed using the sigmoid function, eq 9. The sigmoid function is one of the most commonly used transfer functions. The calculated value, y_k, is the output of the considered neuron. All neurons in hidden and output layers are calculated similarly. Finally, the result was sent to the output neuron, and then the estimated value, y_k, was calculated. In eqs 8 and 9, j and k represent hidden and output layers, respectively.

$$x_k = \sum_{j=1} W_{jk} y_j \quad (8)$$

$$y_k = f(x_k) = \left(\frac{1}{1 + e^{-x_k}}\right) \quad (9)$$

Structure of ANN. ANNs used in this paper were three-layer networks with 13 neurons in the input layer, 5 or 6 neurons in the hidden layer, and 1 neuron in the output layer. After some preliminary test run, the number of neurons in the hidden layer was determined to optimize performance. ANN1 consisted of 5 neurons in the hidden layer, whereas ANN2 consisted of 6 neurons in the hidden layer. ANNs were constructed using the health-promoting effect as the output value and relative expression of 13 kinds of marker proteins as input data, namely, TXN, survivin, Hsp70, XIAP, FADD, TXNRD1, Hsp90, MxA, tNOX, NQO1, ERK2, p53, and Bcl-2 (Table 2). The output variable data (also called teacher signal) should be normalized into the range from 0 to 1 for the transfer function using the sigmoid function. Therefore, the values were normalized by scaling linearly between 0.1 and 0.9 using minima and maxima.

Verification and Selection of the Optimized ANN. To select the optimized ANN, the main parameters of the ANNs were optimized. The range of initial weight was varied from −2 to 2. The learning rate of ANN1 was varied from 0.2 to 0.6 in steps of 0.1, and the momentum factor was varied from 0.2 to 0.8 in steps of 0.05. The learning rate of ANN2 was varied from 0.6 to 0.7 in steps of 0.05, and the momentum factor was 0.4. An ANN was trained repeatedly until the allowable error became <0.1 or until the learning epoch reached 40000. Then, the trained ANN was retested by the training data sets. Additionally, the ANN was tested against the test data sets that had not been included in the ANN learning (Table 3). The optimized ANN, which gave the low RMSE and the high R^2, was selected in all of the built ANN.

Comparison of the Prediction Performance of the Models. To verify the performance of each estimation model, SLR, MLR, ANN1, and ANN2 were examined using test data (Table 3). The goodness-of-fit was evaluated by R^2 and RMSE. The similarity of the actual and estimated values was evaluated by analyzing their medians by the Mann−Whitney test. The variances of the actual and estimated values were also compared by Levene's test. If the p value is >0.05, there is no significant evidence to conclude that the actual data and the data from the models differ. All statistical analysis was performed with Systat 13 (Systat Software, Inc., Point Richmond, CA).

RESULTS

Estimation of Health-Promoting Effects by SLR. In this study, we have attempted to build models to estimate three health-promoting effects, namely, cancer cell growth suppression activity, antiviral activity, and antioxidant stress activity, from intracellular protein expression in response to stimulation with food constituents. A SLR model was first used to examine

Table 5. Correlation Matrix between Marker Protein Expressions as Variables

	TXN	survivin	Hsp70	XIAP	FADD	TXNRD1	Hsp90	MxA	tNOX	NQO1	ERK2	p53	Bcl-2
TXN	1												
survivin	0.244	1											
Hsp70	0.598	0.299	1										
XIAP	0.468	0.264	0.053	1									
FADD	0.428	0.523	0.609	0.099	1								
TXNRD1	0.323	0.511	0.464	0.252	0.617	1							
Hsp90	0.294	0.642	0.475	0.176	0.596	0.442	1						
MxA	0.097	0.316	0.031	0.341	0.194	0.508	0.136	1					
tNOX	0.425	0.193	0.215	0.266	0.224	0.269	0.252	0.180	1				
NQO1	0.529	0.272	0.596	0.224	0.720	0.505	0.331	0.228	0.417	1			
ERK2	0.520	0.430	0.397	0.515	0.487	0.506	0.350	0.164	0.369	0.681	1		
p53	0.321	0.085	0.312	0.295	0.316	0.172	0.083	0.071	0.269	0.412	0.519	1	
Bcl-2	0.383	0.466	0.407	0.446	0.614	0.550	0.529	0.363	0.482	0.599	0.622	0.399	1

Table 6. Regression Coefficients and Statistics of the Fits Obtained from MLR Models

	cancer cell growth suppression activity		antiviral activity		antioxidant stress activity	
variable	intercept	regression coeff	intercept	regression coeff	intercept	regression coeff
	1.060		1.291		0.753	
TXN		0.098		−0.310		0.200
survivin		0.086		0.476		−0.789
Hsp70		−0.065		−0.595		−0.388
XIAP		−0.205		−1.214		−0.253
FADD		−0.049		0.019		0.024
TXNRD1		0.110		0.288		1.552
Hsp90		0.095		−0.235		−0.414
MxA		0.013		−0.159		−0.199
tNOX		−0.161		−0.551		0.414
NQO1		0.083		−0.525		0.465
ERK2		−0.112		0.617		−0.189
p53		−0.181		−0.755		0.120
Bcl-2		0.124		2.065		−0.071
R^2	0.345**[a]		0.567***		0.248	
adjusted R^2	0.218		0.437		0.095	
RMSE	0.102		0.303		0.599	
N	81		57		78	

[a] **, $P < 0.01$; ***, $P < 0.001$.

whether three health-promoting effects can be estimated by relative expression of a single marker protein compared with the control. The models were constructed using the health-promoting effects as the dependent variable and relative expression of a marker protein as the independent variable (Table 2). The regression coefficients, intercepts, and statistics of fits obtained from the SLR models are shown in Table 4. An F test was used to test the significance of the established regression equations. The models that had significant explanatory power were the regression equations using ERK2 or p53 for cancer cell growth suppression activity; TXN, XIAP, MxA, tNOX, NQO1, or p53 for antiviral activity; TXNRD1 or NQO1 for antioxidant stress activity (Table 4). Although the model estimated for cancer cell growth suppression activity by p53 had the highest coefficient of determination, R^2, in all SLR models, the model could explain only 22.4% of the total variance. Any SLR models by a single marker protein had low R^2. Therefore, a SLR model would not be adequate to estimate health-promoting effects by marker protein expression.

Estimation of Health-Promoting Effects by MLR. SLR models that estimate health-promoting effects by single marker protein expression did not have sufficient explanatory power; thus, we attempted to build a MLR model using 13 marker protein expressions. Table 5 summarizes the correlation coefficients between any two descriptors, namely, marker protein expressions. The mean was 0.372, and the highest correlation coefficient was 0.720. Thus, significant multicollinearity did not exist among marker proteins expressions. Then MLR models

Table 7. Parameters of the Optimized ANN1Models

	cancer cell growth suppression activity	antiviral activity	antioxidant stress activity
number of input neurons	13	13	13
number of hidden neurons	5	5	5
number of output neurons	1	1	1
learning rate	0.80	0.70	0.75
momentum factor	0.3	0.5	0.3
final prediction error (training data)			
R^2	0.799	0.954	0.828
RMSE	0.057	0.099	0.294

were constructed using health-promoting effect as the dependent variable and relative expression of 13 kinds of marker proteins as independent variables, namely, TXN, survivin, Hsp70, XIAP, FADD, TXNRD1, Hsp90, MxA, tNOX, NQO1, ERK2, p53, and Bcl-2 (Table 2). The regression coefficients, intercepts, and statistics of fits obtained from the MLR models are shown in Table 6. The results of *F* testing showed that the MLR models for cancer cell growth suppression activity and antiviral activity estimation had significant explanatory power, whereas the MLR model for antioxidant stress activity was not useful to estimate activity. The R^2 values of the MLR models that had significant explanatory power were higher than those of the SLR models. In addition, the model for antioxidant stress activity had also high R^2 in comparison with the SLR models. Furthermore, the RMSE values from the MLR models were smaller than those of the SLR models, so that prediction error from any MLR models improved relative to the SLR models. These results suggest that the use of multimarker proteins may be appropriate for estimation of health-promoting effect in comparison with a single marker protein.

Estimation of Health-Promoting Effects by ANN. Because the R^2 of the MLR models, types of linear models, were low, the models were not able to adequately explain the total variance. Therefore, we attempted to build ANN as a nonlinear model. The ANN models were trained with an error back-propagation algorithm by using health-promoting effect as the dependent variable and relative expressions of 13 kinds of marker proteins as independent variables (Table 2). Optimized parameters and statistics of fits obtained from the ANN1 models are shown in Table 7. Compared with the SLR and the MLR models, the R^2 of the optimized ANN1 models for three health-promoting effects were improved. The ANN1 models were also improved in RSME compared with the SLR and the MLR models. These results suggest that the ANN1 models, types of nonlinear models, will be adequate to calculate health-promoting effects from intracellular protein expressions.

Comparison of Prediction Performance of the Models by the External Test Data. Two food constituents, a drug, and 12 food extracts, which had not been used as the training data, were used as the test data for further external validation of the models (Table 3). Predictive abilities of the models were confirmed by the RMSE and the R^2 between actual and estimated values. Statistical differences between actual and estimated values were evaluated by Mann−Whitney and Levene's tests. As can be seen in Table 8, the most adequate values of the RMSE and R^2 of actual and estimated values are reached using the ANN1 models in each of three health-promoting effects. All of the *p* values from Table 8 indicate that there is no significant evidence to conclude that the actual data and the data from the ANN1 models differ. Taken together, these results indicate that the most adequate models to estimate for three health-promoting effects are ANN1 models.

Improvement in Prediction Performance of ANN Models. ANN models based on nonlinear systems were most adequate to estimate for three health-promoting effects from intracellular protein expressions. However, prediction performances of the ANN1 models were not enough to estimate because the training data might only include a relatively small number of data points against the number of variables. To improve the prediction performance of the ANN models, ANN2 models were built using the synthesized training data. Optimized parameters of the ANN2 models are shown in Table 9. Prediction performances of the models were verified by the test data, and the estimated values are shown in Table 3. These results indicated that the ANN2 models gave basically correct estimated values, although a few exceptional errors between the actual and the estimated values were observed in the models for antiviral and antioxidant stress activities. Statistics of fits obtained from the optimized ANN2 models are shown in Table 9 and Figure 2. The RMSE and R^2 from the ANN2 models were better than those from the ANN1 models (Table 8) in each of the three health-promoting effects, so that prediction performances of the ANN2 models were more improved over those of the ANN1 models. In addition, all of the *p* values in Table 9 were >0.05, demonstrating that there were no statistically significant differences between the actual data and estimated values calculated by the ANN2 models. Therefore, the ANN2 models have a statistically satisfactory goodness of fit from the modeling point of view. These results showed that the most adequate models to estimate for three health-promoting effects were the ANN2 models that were built by the synthesized training data. Taken together, the ANN2 models could estimate three health-promoting effects simultaneously with reasonable accuracy.

DISCUSSION

A mathematical model with descriptors of compound structure and/or physical and chemical characteristics, a QSAR, is well-known as a means of estimating the physiological and chemical activity of a compound in the field of pharmacy.[49−52] QSAR is an effective prediction method for a single compound of already known structure; however, it is not suitable for a functional prediction of food constituents. Often, a test sample is not an isolated or purified compound; instead, a crude extract of food is evaluated. In this case, the molecular structure information of a test compound cannot be used as a descriptor. Hence, the expression data of marker proteins in the cell that replied to stimulation by a compound were utilized as the descriptors in this study. A change in expression level of an intracellular protein reflects the various phenomena that occur in a cell. It is known that some marker proteins are related to pathophysiological functions, for example, caspase family enzymes[53] and p53[54] contribute to apoptosis and tumor necrosis

Table 8. Comparison of the Statistical Parameters in the Test Data Using Every Applied Model

		cancer cell growth suppression activity					antiviral activity					antioxidant stress activity				
					hypothesis testing (p value)					hypothesis testing (p value)					hypothesis testing	
test type		R	R^2	RMSE	Mann−Whitney	Levene's test	R	R^2	RMSE	Mann−Whitney	Levene's test	R	R^2	RMSE	Mann−Whitney	Levene's test
SLR	variable															
	TXN	0.412	0.169	0.126	0.489	0.001	−0.188	0.035	0.604	0.062	0.226	0.408	0.166	0.612	0.027	0.003
	survivin	−0.434	0.188	0.151	0.734	0.001	0.194	0.038	0.475	0.010	0.000	−0.208	0.044	0.697	0.130	0.001
	Hsp70	0.382	0.146	0.131	0.534	0.000	0.148	0.022	0.456	0.008	0.000	0.440	0.194	0.621	0.030	0.000
	XIAP	0.516	0.266	0.122	0.319	0.002	−0.225	0.051	0.481	0.025	0.001	−0.487	0.237	0.652	0.040	0.000
	FADD	0.350	0.122	0.134	0.681	0.000	0.362	0.131	0.477	0.006	0.000	0.195	0.038	0.639	0.038	0.001
	TXNRD1	−0.533	0.284	0.139	0.716	0.000	−0.061	0.004	0.476	0.010	0.000	0.586	0.343	0.612	0.022	0.001
	Hsp90	−0.296	0.088	0.141	0.935	0.000	0.235	0.055	0.474	0.010	0.000	−0.410	0.168	0.684	0.056	0.000
	MxA	−0.450	0.203	0.136	0.630	0.000	−0.341	0.116	0.495	0.004	0.000	−0.224	0.050	0.649	0.040	0.004
	tNOX	0.635	0.403	0.112	0.265	0.003	−0.091	0.008	0.485	0.130	0.020	0.522	0.272	0.607	0.018	0.002
	NQO1	0.635	0.403	0.124	0.519	0.001	−0.147	0.022	0.470	0.021	0.000	0.515	0.266	0.604	0.024	0.005
	ERK2	0.307	0.094	0.125	0.392	0.001	−0.220	0.048	0.479	0.010	0.000	0.134	0.018	0.661	0.021	0.004
	p53	0.598	0.357	0.133	0.076	0.631	−0.305	0.093	0.672	0.275	0.335	0.417	0.174	0.618	0.020	0.001
	Bcl-2	0.509	0.259	0.127	0.550	0.001	−0.341	0.116	0.478	0.010	0.000	0.404	0.163	0.624	0.036	0.061
MLR		0.520	0.271	0.121	0.474	0.489	−0.104	0.011	0.725	0.647	0.192	0.401	0.161	0.605	0.065	0.999
ANN1		0.750	0.562	0.094	0.519	0.753	0.807	0.651	0.246	0.897	0.470	0.687	0.472	0.552	0.769	0.003

Table 9. Parameters and Statistics of the Fits Obtained from the Optimized ANN2 Models

	cancer cell growth suppression activity	antiviral activity	antioxidant stress activity
number of input neurons	13	13	13
number of hidden neurons	6	6	6
number of output neurons	1	1	1
learning rate	0.70	0.65	0.70
momentum factor	0.4	0.4	0.4
final prediction error (training data)			
R^2	0.938	0.959	0.871
RMSE	0.054	0.109	0.284
final prediction error (test data)			
R^2	0.714	0.741	0.646
RMSE	0.082	0.209	0.384
hypothesis testing (p value)			
Mann−Whitney	0.286	0.614	0.236
Levene's test	0.878	0.783	0.382

factor-α[55,56] and interleukin-6[55,57] contribute to inflammation. In this study, an empirical modeling method was adopted because intracellular proteins including ones that have an unknown relationship to health-promoting effects were assumed marker proteins. An empirical modeling method is adaptable using experimental results. The model based on biochemical descriptors was able to estimate health-promoting effects of a test sample within an allowable error even if the sample was an extract including a complex mixture of components. Furthermore, if the expression data of marker proteins are used as common data, it is expected that activities of plural health-promoting effects can be estimated at the same time. We evaluated three health-promoting effects, cancer cell growth suppression activity, antiviral activity, and antioxidant stress activity, to check whether plural health-promoting effects could be estimated at the same time. The model was able to estimate three health-promoting effects from the data common to plural effects, namely, expression data of marker proteins that was not used for learning of the model (Table 3). The estimated values obtained from the model were similar to the actual values for a wide range of activities. Physiological activities of the previously reported compounds and extracts were confirmed by the model. For example, genistein and the rosemary extract have a cancer cell growth suppression activity,[12,30−34] and a defense effect against oxidative stress.[35−43] Lovastatin[44−47] and blueberry leaf extract[58] have suppression activities against HCV replication. Our results suggest that plural health-promoting effects of a compound or extract can be estimated from the expression data of marker proteins efficiently. Using only a single wet experiment, namely, measurement of the expression level of a cellular protein in response to stimulation by a test compound, made it possible to estimate multiple health-promoting effects using a calculation with high accuracy.

Generally, a clinical test involves a single tumor marker, but there are many cases when a combination of several plural markers allows a more precise diagnosis. Prostate-specific antigen (PSA) is the most widely used serum biomarker for early detection of prostate cancer.[59] However, the utility of PSA has

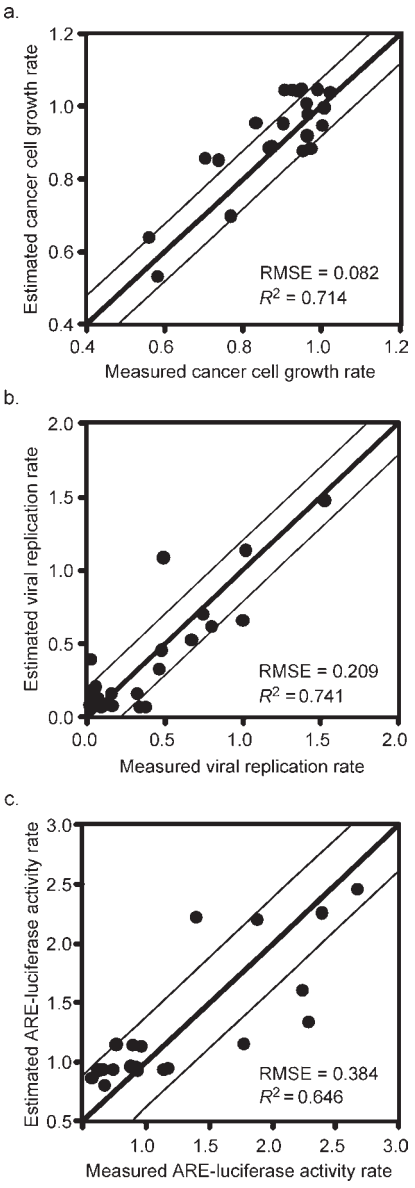

Figure 2. Correlation of actual versus estimated values with test data using the optimized ANN2 models: (a) cancer cell growth suppression activity; (b) antiviral activity; (c) antioxidant stress activity. The area between narrow lines expresses the range of the RMSE between actual and estimated values. The RMSE and the R^2 between actual and estimated values are given in each graph.

been limited by a lack of specificity within the 4−10 ng/mL range, such that a second biochemical marker, namely, free PSA, must be measured and the free-to-total serum PSA ratio can be used to increase specificity.[60] Furthermore, according to a recent paper, a mass spectral serum profiling method that is based on a multimarker was higher in accuracy, sensitivity, specificity, a positive predictive value, and a negative predictive value than diagnosis by only PSA.[61] Our results indicated that the MLR analysis model using the multimarker had higher prediction accuracy than the SLR analysis models using the single marker (Tables 4 and 5). If a relationship between a marker protein and a health-promoting effect does not depend on a protein expression level, it may be difficult to predict a health-promoting effect only using a single marker. In addition, it will be difficult to predict a health-promoting effect only by a single marker when plural mechanisms are present for one health-promoting effect. For example, the antiproliferative effect accompanying apoptosis and/or cell cycle arrest in cancer cells is regulated by a complex signal transduction pathway. p53, one of the tumor suppressor genes, participates in the regulation of cell cycle progression at G_1/S and G_2/M phases and induction of apoptosis.[54] Genistein enhances p53 expression in HepG2 cells, and an antiproliferative effect is shown.[32] In this study, genistein enhanced expression of p53 and had an effect on the antiproliferation of HepG2 cells (Table 3). However, it is not always true that an antiproliferative effect in cancer cells is regulated via the p53-dependent pathway. Carnosic acid, a component of rosemary, shows an antiproliferative effect through G_2 cell cycle arrest in p53-deficient human prostatic cancer PC3 cells.[62] Additionally, to our knowledge there have been no previous reports of an antiproliferative effect through a p53-dependent pathway by rosemary extracts and/or components such as carnosic acid and carnosol. As shown in Table 3, an antiproliferative effect was shown by the rosemary and the Japanese radish leaves, and the expression level of p53 remained unchanged. In this way it would be difficult to evaluate a health-promoting effect only by a single marker, when there was a large difference in the expression level of the intracellular protein that was influenced by the different constituents with a similar health-promoting effect.

ANN was the best among the tested models. Therefore, it seems that a relationship between health-promoting effects and expression patterns of marker proteins was nonlinear. Because expression patterns of intracellular proteins are very complicated, linear separation may have been difficult. An ANN trained by error back-propagation[48] can resolve a problem that could not be divided into a linear solution. Therefore, for solution of various complicated problems, ANN is applied in many fields, for example, toxicology,[63] pharmacy,[64] food safety, and quality analysis.[65] ANN, which can solve a complicated problem, would be suitable for our model that has 13 descriptors of plural marker protein expression levels.

Although the plural health-promoting effects can be estimated from the expression data of marker protein, there are still some points that should be improved. Food constituents used for training data were biased toward polyphenols. Besides polyphenols of plant origin, food products contain many constituents, for example, proteins and peptides of animal and marine products origin, polysaccharides of mushrooms, and fatty acids of animal, marine, and plant origin. There is a possibility that an estimation method based on an empirical modeling approach cannot be estimated precisely in a constituent of the kind that is not learned. Therefore, the model must be trained using data of various kinds of food constituents. From the point of view of modeling, if the number of marker proteins, or descriptors, is increased, the accuracy of the estimate will rise. However, the number of necessary data points increases if the number of descriptors increases, because overfitting occurs when the number of data is insufficient compared with the number of descriptors. Because it is difficult to collect enormous amounts of data to prevent overfitting, adequate marker proteins should be chosen and unnecessary markers should be excluded from the model. Therefore, in our laboratory we are currently studying ways to choose better marker proteins.

In conclusion, plural health-promoting effects can be estimated simultaneously from the expression data of marker proteins. This system will be effective as a prediction model that

can estimate plural health-promoting effects simultaneously. Being able to presume plural health-promoting effects by measuring only the expression data of the marker proteins means they are more promptly and handily assessable because they need not be measured by using plural methods. Furthermore, if more health-promoting effects can be presumed from the expression data of the marker proteins at the same time, it will be useful as a first screening method of food constituents for various beneficial purposes.

AUTHOR INFORMATION

Corresponding Author

*Postal address: Department of Biochemistry and Applied Biosciences, Faculty of Agriculture, University of Miyazaki, 1-1 Gakuen Kibanadai-nishi, Miyazaki 889-2192, Japan. Phone: +81-985-58-7210. Fax: +81-985-58-7210. E-mail: neto@cc.miyazaki-u.ac.jp.

Present Addresses

⊗Department of Veterinary Physiology, Faculty of Agriculture, University of Miyazaki.

⊥Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki.

Funding Sources

This work was supported by a Grant-in-Aid from the Collaboration of Regional Entities for the Advancement of Technological Excellence (CREATE) from the Japan Science and Technology Agency.

ACKNOWLEDGMENT

We thank Dr. Michinori Kohara (The Tokyo Metropolitan Institute of Medical Science) for providing HCV replicon cells.

REFERENCES

(1) Takachi, R.; Inoue, M.; Ishihara, J.; Kurahashi, N.; Iwasaki, M.; Sasazuki, S.; Iso, H.; Tsubono, Y.; Tsugane, S. Fruit and vegetable intake and risk of total cancer and cardiovascular disease: Japan Public Health Center-Based Prospective Study. *Am. J. Epidemiol.* **2008**, *167*, 59–70.

(2) Yamaji, T.; Inoue, M.; Sasazuki, S.; Iwasaki, M.; Kurahashi, N.; Shimazu, T.; Tsugane, S. Fruit and vegetable consumption and squamous cell carcinoma of the esophagus in Japan: the JPHC study. *Int. J. Cancer* **2008**, *123*, 1935–1940.

(3) Riboli, E.; Norat, T. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *Am. J. Clin. Nutr.* **2003**, *78*, 559S–569S.

(4) Scalbert, A.; Manach, C.; Morand, C.; Rémésy, C.; Jiménez, L. Dietary polyphenols and the prevention of diseases. *Crit. Rev. Food Sci. Nutr.* **2005**, *45*, 287–306.

(5) Cao, G.; Alessio, H.; Cutler, R. Oxygen-radical absorbance capacity assay for antioxidants. *Free Radical Biol. Med.* **1993**, *14*, 303–311.

(6) Schlesier, K.; Harwat, M.; Böhm, V.; Bitsch, R. Assessment of antioxidant activity by using different in vitro methods. *Free Radical Res.* **2002**, *36*, 177–187.

(7) Shishu; Kaur, I. P. Antimutagenicity of curcumin and related compounds against genotoxic heterocyclic amines from cooked food: the structural requirement. *Food Chem.* **2008**, *111*, 573–579.

(8) Spada, P.; de Souza, G.; Bortolini, G.; Henriques, J.; Salvador, M. Antioxidant, mutagenic, and antimutagenic activity of frozen fruits. *J. Med. Food* **2008**, *11*, 144–151.

(9) King, A.; Shaughnessy, D.; Mure, K.; Leszczynska, J.; Ward, W.; Umbach, D.; Xu, Z.; Ducharme, D.; Taylor, J.; Demarini, D.; Klein, C. Antimutagenicity of cinnamaldehyde and vanillin in human cells: Global gene expression and possible role of DNA damage and repair. *Mutat. Res.* **2007**, *616*, 60–69.

(10) Sun, J.; Chu, Y.; Wu, X.; Liu, R. Antioxidant and antiproliferative activities of common fruits. *J. Agric. Food Chem.* **2002**, *50*, 7449–7454.

(11) Chu, Y.; Sun, J.; Wu, X.; Liu, R. Antioxidant and antiproliferative activities of common vegetables. *J. Agric. Food Chem.* **2002**, *50*, 6910–6916.

(12) Cheung, S.; Tai, J. Anti-proliferative and antioxidant properties of rosemary *Rosmarinus officinalis*. *Oncol. Rep.* **2007**, *17*, 1525–1531.

(13) Chen, D.; Milacic, V.; Chen, M. S.; Wan, S. B.; Lam, W. H.; Huo, C.; Landis-Piwowar, K. R.; Cui, Q. C.; Wali, A.; Chan, T. H.; Dou, Q. P. Tea polyphenols, their biological effects and potential molecular targets. *Histol. Histopathol.* **2008**, *23*, 487–496.

(14) Hodgson, J. M.; Croft, K. D. Tea flavonoids and cardiovascular health. *Mol. Aspects Med.* **2010**, *31*, 495–502.

(15) Lambert, J. D.; Elias, R. J. The antioxidant and pro-oxidant activities of green tea polyphenols: a role in cancer prevention. *Arch. Biochem. Biophys.* **2010**, *501*, 65–72.

(16) Thielecke, F.; Boschmann, M. The potential role of green tea catechins in the prevention of the metabolic syndrome — a review. *Phytochemistry* **2009**, *70*, 11–24.

(17) Singh, R.; Akhtar, N.; Haqqi, T. M. Green tea polyphenol epigallocatechin-3-gallate: inflammation and arthritis [corrected]. *Life Sci.* **2010**, *86*, 907–918.

(18) Michelini, E.; Cevenini, L.; Mezzanotte, L.; Coppa, A.; Roda, A. Cell-based assays: fuelling drug discovery. *Anal. Bioanal. Chem.* **2010**, *398*, 227–238.

(19) Bajorath, J. Computational analysis of ligand relationships within target families. *Curr. Opin. Chem. Biol.* **2008**, *12*, 352–358.

(20) Ekins, S.; Mestres, J.; Testa, B. In silico pharmacology for drug discovery: applications to targets and beyond. *Br. J. Pharmacol.* **2007**, *152*, 21–37.

(21) Ekins, S.; Mestres, J.; Testa, B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* **2007**, *152*, 9–20.

(22) Valerio, L. G. In silico toxicology for the pharmaceutical sciences. *Toxicol. Appl. Pharmacol.* **2009**, *241*, 356–370.

(23) Martinez-Mayorga, K.; Medina-Franco, J. L. Chemoinformatics — applications in food chemistry. *Adv. Food Nutr. Res.* **2009**, *58*, 33–56.

(24) Valerio, L. G.; Arvidson, K. B.; Chanderbhan, R. F.; Contrera, J. F. Prediction of rodent carcinogenic potential of naturally occurring chemicals in the human diet using high-throughput QSAR predictive modeling. *Toxicol. Appl. Pharmacol.* **2007**, *222*, 1–16.

(25) Cabrera, A. C.; Prieto, J. M. Application of artificial neural networks to the prediction of the antioxidant activity of essential oils in two experimental in vitro models. *Food Chem.* **2010**, *118*, 141–146.

(26) Bucinski, A.; Zielinski, H.; Kozlowska, H. Artificial neural networks for prediction of antioxidant capacity of cruciferous sprouts. *Trends Food Sci. Technol.* **2004**, *15*, 161–169.

(27) Sakamoto, H.; Okamoto, K.; Aoki, M.; Kato, H.; Katsume, A.; Ohta, A.; Tsukuda, T.; Shimma, N.; Aoki, Y.; Arisawa, M.; Kohara, M.; Sudoh, M. Host sphingolipid biosynthesis as a target for hepatitis C virus therapy. *Nat. Chem. Biol.* **2005**, *1*, 333–337.

(28) Boerboom, A.; Vermeulen, M.; van der Woude, H.; Bremer, B.; Lee-Hilz, Y.; Kampman, E.; van Bladeren, P.; Rietjens, I.; Aarts, J. Newly constructed stable reporter cell lines for mechanistic studies on electrophile-responsive element-mediated gene expression reveal a role for flavonoid planarity. *Biochem. Pharmacol.* **2006**, *72*, 217–226.

(29) Tominaga, H.; Ishiyama, M.; Ohseto, F.; Sasamoto, K.; Hamamoto, T.; Suzuki, K.; Watanabe, M. A water-soluble tetrazolium salt useful for colorimetric cell viability assay. *Anal. Commun.* **1999**, *36*, 47–50.

(30) Chodon, D.; Ramamurty, N.; Sakthisekaran, D. Preliminary studies on induction of apoptosis by genistein on HepG2 cell line. *Toxicol In Vitro* **2007**, *21*, 887–891.

(31) Frey, R.; Li, J.; Singletary, K. Effects of genistein on cell proliferation and cell cycle arrest in nonneoplastic human mammary epithelial cells: involvement of Cdc2, p21(waf/cip1), p27(kip1), and Cdc25C expression. *Biochem. Pharmacol.* **2001**, *61*, 979–989.

(32) Chang, K.; Kung, M.; Chow, N.; Su, S. Genistein arrests hepatoma cells at G2/M phase: involvement of ATM activation and upregulation of p21waf1/cip1 and Wee1. *Biochem. Pharmacol.* **2004**, *67*, 717–726.

(33) Su, S.; Chow, N.; Kung, M.; Hung, T.; Chang, K. Effects of soy isoflavones on apoptosis induction and G2-M arrest in human hepatoma cells involvement of caspase-3 activation, Bcl-2 and Bcl-XL downregulation, and Cdc2 kinase activity. *Nutr. Cancer* **2003**, *45*, 113–123.

(34) Yamasaki, M.; Fujita, S.; Ishiyama, E.; Mukai, A.; Madhyastha, H.; Sakakibara, Y.; Suiko, M.; Hatakeyama, K.; Nemoto, T.; Morishita, K.; Kataoka, H.; Tsubouchi, H.; Nishiyama, K. Soy-derived isoflavones inhibit the growth of adult T-cell leukemia cells in vitro and in vivo. *Cancer Sci.* **2007**, *98*, 1740–1746.

(35) Raschke, M.; Rowland, I.; Magee, P.; Pool-Zobel, B. Genistein protects prostate cells against hydrogen peroxide-induced DNA damage and induces expression of genes involved in the defence against oxidative stress. *Carcinogenesis* **2006**, *27*, 2322–2330.

(36) Sierens, J.; Hartley, J.; Campbell, M.; Leathem, A.; Woodside, J. Effect of phytoestrogen and antioxidant supplementation on oxidative DNA damage assessed using the comet assay. *Mutat. Res.* **2001**, *485*, 169–176.

(37) Wu, H.; Chan, W. Genistein protects methylglyoxal-induced oxidative DNA damage and cell injury in human mononuclear cells. *Toxicol In Vitro* **2007**, *21*, 335–342.

(38) Chan, W.; Yu, J. Inhibition of UV irradiation-induced oxidative stress and apoptotic biochemical changes in human epidermal carcinoma A431 cells by genistein. *J. Cell Biochem.* **2000**, *78*, 73–84.

(39) Foti, P.; Erba, D.; Riso, P.; Spadafranca, A.; Criscuoli, F.; Testolin, G. Comparison between daidzein and genistein antioxidant activity in primary and cancer lymphocytes. *Arch. Biochem. Biophys.* **2005**, *433*, 421–427.

(40) Liang, H.; Qiu, S.; Shen, J.; Sun, L.; Wang, J.; Bruce, I.; Xia, Q. Genistein attenuates oxidative stress and neuronal damage following transient global cerebral ischemia in rat hippocampus. *Neurosci. Lett.* **2008**, *438*, 116–120.

(41) Wei, H.; Zhang, X.; Wang, Y.; Lebwohl, M. Inhibition of ultraviolet light-induced oxidative events in the skin and internal organs of hairless mice by isoflavone genistein. *Cancer Lett.* **2002**, *185*, 21–29.

(42) Satoh, T.; Izumi, M.; Inukai, Y.; Tsutsumi, Y.; Nakayama, N.; Kosaka, K.; Shimojo, Y.; Kitajima, C.; Itoh, K.; Yokoi, T.; Shirasawa, T. Carnosic acid protects neuronal HT22 Cells through activation of the antioxidant-responsive element in free carboxylic acid- and catechol hydroxyl moieties-dependent manners. *Neurosci. Lett.* **2008**, *434*, 260–265.

(43) Satoh, T.; Kosaka, K.; Itoh, K.; Kobayashi, A.; Yamamoto, M.; Shimojo, Y.; Kitajima, C.; Cui, J.; Kamins, J.; Okamoto, S.; Izumi, M.; Shirasawa, T.; Lipton, S. Carnosic acid, a catechol-type electrophilic compound, protects neurons both in vitro and in vivo through activation of the Keap1/Nrf2 pathway via S-alkylation of targeted cysteines on Keap1. *J. Neurochem.* **2008**, *104*, 1116–1131.

(44) Ikeda, M.; Kato, N. Life style-related diseases of the digestive system: cell culture system for the screening of anti-hepatitis C virus (HCV) reagents: suppression of HCV replication by statins and synergistic action with interferon. *J. Pharmacol. Sci.* **2007**, *105*, 145–150.

(45) Kim, S.; Peng, L.; Lin, W.; Choe, W.; Sakamoto, N.; Kato, N.; Ikeda, M.; Schreiber, S.; Chung, R. A cell-based, high-throughput screen for small molecule regulators of hepatitis C virus replication. *Gastroenterology* **2007**, *132*, 311–320.

(46) Kapadia, S.; Chisari, F. Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2561–2566.

(47) Ye, J.; Wang, C.; Sumpter, R. J.; Brown, M.; Goldstein, J.; Gale, M. J. Disruption of hepatitis C virus RNA replication through inhibition of host protein geranylgeranylation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15865–15870.

(48) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536.

(49) Arab Chamjangali, M.; Beglari, M.; Bagherian, G. Prediction of cytotoxicity data (CC(50)) of anti-HIV 5-phenyl-1-phenylamino-1*H*-imidazole derivatives by artificial neural network trained with Levenberg−Marquardt algorithm. *J. Mol. Graph. Model* **2007**, *26*, 360–367.

(50) Lim, C.; Fujiwara, S.; Yamashita, F.; Hashida, M. Prediction of human skin permeability using a combination of molecular orbital calculations and artificial neural network. *Biol. Pharm. Bull.* **2002**, *25*, 361–366.

(51) Jaiswal, K.; Naik, P. Distinguishing compounds with anticancer activity by ANN using inductive QSAR descriptors. *Bioinformation* **2008**, *2*, 441–451.

(52) Lara, J.; Wohlhueter, R.; Dimitrova, Z.; Khudyakov, Y. Artificial neural network for prediction of antigenic activity for a major conformational epitope in the hepatitis C virus NS3 protein. *Bioinformatics* **2008**, *24*, 1858–1864.

(53) Cohen, G. Caspases: the executioners of apoptosis. *Biochem. J.* **1997**, *326* (Part 1), 1−16.

(54) Haupt, S.; Berger, M.; Goldberg, Z.; Haupt, Y. Apoptosis − the p53 network. *J. Cell Sci.* **2003**, *116*, 4077–4085.

(55) Balkwill, F.; Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **2001**, *357*, 539–545.

(56) Sethi, G.; Sung, B.; Aggarwal, B. TNF: a master switch for inflammation to cancer. *Front. Biosci.* **2008**, *13*, 5094–5107.

(57) Rose-John, S.; Scheller, J.; Elson, G.; Jones, S. Interleukin-6 biology is coordinated by membrane-bound and soluble receptors: role in inflammation and cancer. *J Leukoc Biol* **2006**, *80*, 227–236.

(58) Takeshita, M.; Ishida, Y.; Akamatsu, E.; Ohmori, Y.; Sudoh, M.; Uto, H.; Tsubouchi, H.; Kataoka, H. Proanthocyanidin from blueberry leaves suppresses expression of subgenomic hepatitis C virus RNA. *J. Biol. Chem.* **2009**, *284*, 21165–21176.

(59) Catalona, W.; Smith, D.; Ratliff, T.; Basler, J. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. *JAMA, J. Am. Med. Assoc.* **1993**, *270*, 948–954.

(60) Catalona, W.; Smith, D.; Wolfert, R.; Wang, T.; Rittenhouse, H.; Ratliff, T.; Nadler, R. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. *JAMA, J. Am. Med. Assoc.* **1995**, *274*, 1214–1220.

(61) Oh, J.; Lotan, Y.; Gurnani, P.; Rosenblatt, K.; Gao, J. Prostate cancer biomarker discovery using high performance mass spectral serum profiling. *Comput. Methods Programs Biomed.* **2009**, *96*, 33–41.

(62) Johnson, J.; Syed, D.; Heren, C.; Suh, Y.; Adhami, V.; Mukhtar, H. Carnosol, a dietary diterpene, displays growth inhibitory effects in human prostate cancer PC3 cells leading to G2-phase cell cycle arrest and targets the 5′-AMP-activated protein kinase (AMPK) pathway. *Pharm. Res.* **2008**, *25*, 2125–2134.

(63) Dohnal, V.; Kuca, K.; Jun, D. What are artificial neural networks and what they can do? *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.* **2005**, *149*, 221–224.

(64) Winkler, D. Neural networks as robust tools in drug lead discovery and development. *Mol. Biotechnol.* **2004**, *27*, 139–168.

(65) Huang, Y.; Kangas, L.; Rasco, B. Applications of artificial neural networks (ANNs) in food science. *Crit. Rev. Food Sci. Nutr.* **2007**, *47*, 113–126.