and a gene but also various other types of co-occurrences data, such as gene–gene and compound–compound co-occurrence data.

We propose a probabilistic model, which we call a mixture aspect model (MAM), coupled with an efficient algorithm for estimating its parameters. MAM is an extension of a probabilistic model, called the aspect model (AM) developed in natural language processing (Hofmann, 2001), with one significant difference. MAM can incorporate different types of co-occurrence data efficiently. More formally, the probabilistic structure of MAM is a weighted mixture of (normalized) AMs, and each component (i.e. AM) handles one type of co-occurrence data. For example, we can have three different components corresponding to three different co-occurrences of compound–gene, gene–gene and compound–compound. These three datasets might be handled by AM by regarding all three as only one type of co-occurrence data. However MAM has roughly two significant advantages, compared with this way of using AM. First, it has a weight for each component, so that the users can control the weight for each co-occurrence dataset. Obviously, AM cannot do this. The second advantage is both time and space efficiency. When we have T datasets all having co-occurrences of N events, MAM considers $T \cdot N^2$ combinations at most, whereas AM must consider a maximum of $T^2 \cdot N^2$. In practice, N (i.e. the number of compounds or genes) reaches at least a few thousands, so that even if T is a relatively small number, this difference would be pronounced.

Our algorithm for estimating the probability parameters of MAM is based on the EM (Expectation–Maximization) algorithm (Dempster *et al.*, 1977) that locally maximizes the likelihoods of given data. Once the probability parameters of MAM are estimated, MAM can predict the likelihood for any pair of events, such as a pair of a chemical compound and a gene. MAM can find new biological-related compound–gene pairs that have not yet been found in current biology and medical literatures.

In our experiments, we generated three types of co-occurrence datasets: gene–gene, compound–compound and compound–gene from the Medline records (Wheeler *et al.*, 2005). We evaluated our method by not only these datasets, but also an independent (human-curated) dataset of chemical compound and gene relationships in the ChEBI database. We first checked the performance of MAM to predict the co-occurrences of compounds and genes by using cross-validation, starting with compound–gene pairs and then adding compound–compound pairs, followed by gene–gene pairs. Experimental results have shown that adding gene–gene (or compound–compound) pairs improved the performance of using compound–gene pairs only, with the difference being statistically significant. In particular, we found that adding compound–compound pairs is the most effective in improving the performance of predicting compound–gene pairs. We then performed the experiment on predicting the biological-related compounds and genes in the ChEBI database, and found that the performance improvement was obtained in almost the same way. These results indicate that combining all these datasets is effective in our problem setting, and that MAM and its learning algorithm are extremely useful for obtaining the results. Finally, we computed the likelihood of each of all unknown compound–gene (more precisely, drug–gene) pairs and selected the top 20 of them according to the likelihoods. We thus showed a list of them that have the highest likelihoods given by MAM trained by all given datasets and examined the validity of these pairs from biological, medical and pharmaceutical viewpoints.

2 RELATED WORK

Mining the Medline text for biomedical knowledge discovery has become a very active field in bioinformatics recently. One of the important applications is to discover the relationship among genes, proteins, disease phenotype and chemical compounds. Co-occurrence in Medline is a simple, effective and popular technique to identify biological relationships among different entities. This technique is based on the hypothesis that entities appearing in the same Medline record are more likely to be biologically related. This hypothesis has been verified by many researchers. Jenssen *et al.* (2001) presented a gene-to-gene co-occurrence network called PubGene using over 10 million Medline records. They randomly selected 500 pairs of genes that co-occurred once and 500 pairs of genes that co-occurred more than five times in the Medline, then manually analyzed the biological relationship of these pairs by expertise. They found that the accuracy of biological relationship identification is ~60% for the first group, and 72% for the second. In further analysis, they found that almost all errors were owing to the failures in gene name recognition. Chang *et al.* (2004) also identified related genes and drugs based on their co-occurrence in the titles and abstracts of publications in Medline. They manually examined the biological relationship of 100 gene–drug pairs. They found that out of the 100 pairs (50 of them with largest number of co-occurrence, and another 50 of them randomly selected), 70 shared some biological relationships. From these studies, we can see that co-occurrence methods can successfully find biological relationships, and most of the failures are because of the difficulty of biological entity name identification in extracting Medline texts. We emphasize that in our experiment we generated our co-occurrence data not directly from Medline texts, but from human curated datasets (for further details, see Section 4), consequently avoiding errors that may occur in gene name or chemical compound name identification.

Some studies have combined co-occurrence methods with natural language processing techniques, such as shallow parsing, full parsing and constructing templates (Yandell *et al.*, 2002; Blasckel *et al.*, 2002). Their goal is to extract and clarify the detailed relationships among biological entities, such as protein–protein interaction (Blasckel *et al.*, 1999), protein–drug interaction (Rindflesch *et al.*, 2000) and gene–mutation pairs (Rebholz-Schuhmann *et al.*, 2004).

Some researchers have, however, attempted to find implicit relationships between biological entities of not having direct co-occurrences in the literature (Wiren *et al.*, 2004). For example, Perez-Iratxeta *et al.* (2002) used the fuzzy set theory to analyze the relationships between the co-occurrence of MeSH terms in different categories and the co-occurrence of a MeSH term and a GO (Gene Ontology) term in Medline records and scored the implicit associations between symptoms of diseases and GO terms by fuzzy relations.

In contrast to these existing approaches, our focus is placed on implicit 'compound–gene' relations in the literature, and our approach is based on statistical learning using a probabilistic model that is an extension of the so-called AM (Hofmann, 2001). This AM has already proved effective in a lot of applications for analyzing co-occurrence data, such as informational retrieval, computational linguistics and collaborative filtering (Hofmann, 2004; Si *et al.*, 2003). We emphasize that our statistical learning based approach develops a noise-robust probabilistic model and a systematic and

efficient algorithm for estimating the parameters of our model from different types of multiple co-occurrence data.

3 METHODS

3.1 Notations

We define the notations that are used throughout this paper. We denote a variable by a capitalized letter, e.g. U, and its value as the same letter in lower case, e.g. u. To explain a particular model for the co-occurrence of a gene and a compound, we define the following symbols in particular. Let G be an observable random variable taking on values $g_1, \ldots, g_S$, each of which corresponds to a gene. Similarly, let C be an observable random variable taking on $c_1, \ldots, c_T$, each of which corresponds to a chemical compound. Let Z be a discrete-valued latent variable taking on values $z_1, \ldots, z_H$, each of which corresponds to a latent cluster, where H is the number of clusters. Let θ be a set of parameters for the model to be optimized in the learning process, and let π be a mixture parameter (i.e. weight) of a component of our model that the users can specify. Let D be a set of all examples.

3.2 Mixture aspect model (MAM)

We begin by describing the AM for two-mode and co-occurrence data (Hofmann, 2001). With latent clusters z_h $(h = 1, \ldots, H)$, AM gives the log-likelihood for a co-occurrence of (u, v) in the following form:

$$\log p(u, v; \theta) = \log \sum_h p(u|z_h;\theta)p(v|z_h;\theta)p(z_h;\theta).$$

So the log-likelihood for D by this model is given as follows:

$$\log p(D;\theta) = \sum_{i,j} N_{i,j} \log p(u_i, v_j;\theta),$$

where $N_{i,j}$ is the number of co-occurrences of (u_i, v_j).

The purpose of this paper is to handle multiple different types of co-occurrence data with overlapping variable. More concretely, we can assume that we have two datasets, in which one has two random variables U and V, and the other has V and W. For these two datasets, we now define a new probabilistic model that is a mixture of two AMs, which we call two-component mixture aspect model (2MAM). The log-likelihood for D with two datasets for this model is given as follows:

$$\log p(D;\theta) = \pi_{UV} \sum_{i,j} \frac{N_{i,j}}{N_{UV}} \log \sum_h p(u_i|z_h;\theta)p(v_j|z_h;\theta)p(z_h;\theta) + \pi_{VW} \sum_{j,k} \frac{M_{j,k}}{N_{VW}} \log \sum_h p(v_j|z_h;\theta)p(w_k|z_h;\theta)p(z_h;\theta),$$

where $\pi_{UV} + \pi_{VW} = 1$ for U and V, $N_{i,j}$ and $M_{j,k}$ are the number of co-occurrences of (u_i, v_j) and (v_j, w_k), respectively, $N_{UV} = \sum_{i,j} N_{i,j}$ for U and V, and $N_{VW} = \sum_{j,k} M_{j,k}$ for V and W.

We note that both the first and second terms in this equation use the same probability parameter $p(v|z;\theta)$. Therefore, the parameter must be controlled by both datasets. We can easily see that this mixture model for two datasets can be extended to a mixture model for an arbitrary number of datasets. We note that if each of these datasets has a random variable that appears in more than one dataset, this model is different from AM. This is particularly true when estimating its parameters, each of which corresponds to a variable appearing more than once. These parameters must be trained (controlled) by more than one dataset. The detailed algorithm for estimating this type of parameters is described for a particular case of the co-occurrence of a chemical compound and a gene in Section 3.4.

3.3 Mixture aspect model for predicting co-occurrences of compound–gene

When there is only one type of co-occurrence data (i.e. compound–gene pairs), this dataset can be handled by AM. If another dataset like gene–gene pairs is added to this dataset, these two datasets can be handled by 2MAM. For example, if we have two types of co-occurrence data, such as compound–gene and gene–gene pairs, the log-likelihood for all the data D by 2MAM is written as follows:

$$\log p(D;\theta) = \pi_{CG} \sum_{i,j} \frac{N_{i,j}}{N_{CG}} \log \sum_h p(c_i|z_h;\theta)p(g_j|z_h;\theta)p(z_h;\theta) + \pi_{GG} \sum_{j,j'} \frac{M_{j,j'}}{N_{GG}} \log \sum_h p(g_j|z_h;\theta)p(g_{j'}|z_h;\theta)p(z_h;\theta),$$

where $N_{CG} = \sum_{i,j} N_{i,j}$ and $N_{i,j}$ is the number of co-occurrences of (c_i, g_j), and $N_{GG} = \sum_{j,j'} M_{j,j'}$ and $M_{j,j'}$ is the number of co-occurrences of $(g_j, g_{j'})$.

In this paper, we consider three types of co-occurrence data: compound–gene, gene–gene and compound–compound pairs. We also present a probabilistic model for this data, which we call three-component mixture aspect model (3MAM). The log-likelihood for all data D can be given by 3MAM as follows:

$$\log p(D;\theta) = \pi_{CG} \sum_{i,j} \frac{N_{i,j}}{N_{CG}} \log \sum_h p(c_i|z_h;\theta)p(g_j|z_h;\theta)p(z_h;\theta) + \pi_{GG} \sum_{j,j'} \frac{M_{j,j'}}{N_{GG}} \log \sum_h p(g_j|z_h;\theta)p(g_{j'}|z_h;\theta)p(z_h;\theta) + \pi_{CC} \sum_{i,i'} \frac{L_{i,i'}}{N_{CC}} \log \sum_h p(c_i|z_h;\theta)p(c_{i'}|z_h;\theta)p(z_h;\theta).$$

In the above equation, $\pi_{CG} + \pi_{GG} + \pi_{CC} = 1$, $N_{CC} = \sum_{i,i'} L_{i,i'}$ and $L_{i,i'}$ is the number of $(c_i, c_{i'})$ pairs.

In this paper, even though we used three types of data, it is evident that 3MAM can incorporate another type of co-occurrence data if it can improve the predictive performance of 3MAM.

3.4 Estimating probability parameters

Given training data D and the number of clusters H, a popular criterion for estimating the probabilities of a probabilistic model is the maximum likelihood (ML). Parameters are estimated to maximize the log-likelihood of data D:

$$\theta^{\mathrm{ML}} = \arg\max_{\theta} \log p(D;\theta).$$

The most popular approach for obtaining an ML estimator of a probabilistic model is a time-efficient general scheme called the EM (Expectation–Maximization) algorithm (Dempster *et al.*, 1977) that provides a local maximum. In general, the EM algorithm starts with a random set of initial parameter values and iterates both the expectation step (E-step) and the maximization step (M-step) alternately until a certain convergence criterion is satisfied.

3.4.1 AM We begin to explain the EM algorithm for AM for only one type of co-occurrence data, i.e. compound–gene pairs. The log-likelihood for D is given in Section 3.2, and the E- and M-steps can be given as follows:

E-step:

$$p(z_h|c_i, g_j;\theta) = \frac{p(c_i|z_h;\theta)p(g_j|z_h;\theta)p(z_h;\theta)}{\sum_{h'} p(c_i|z_{h'};\theta)p(g_j|z_{h'};\theta)p(z_{h'};\theta)}.$$

M-step:

$$\theta_{c_i|z_h} \propto \sum_j N_{i,j}\, p(z_h|c_i, g_j;\theta_{\mathrm{old}}),$$

$$\theta_{g_j|z_h} \propto \sum_i N_{i,j}\, p(z_h|c_i, g_j;\theta_{\mathrm{old}}),$$

$$\theta_{z_h} \propto \sum_{i,j} N_{i,j}\, p(z_h|c_i, g_j;\theta_{\mathrm{old}}).$$

3.4.2 2MAM Next, we show the EM algorithm for the case in which we add another type of co-occurrence data, such as gene–gene pairs to compound–gene pairs. In this case we use 2MAM, and so the log-likelihood for these datasets is given in Section 3.3. The E- and M-steps for 2MAM can be given as follows:

E-step:

$$p(z_h|c_i,g_j;\theta)=\frac{p(c_i|z_h;\theta)p(g_j|z_h;\theta)p(z_h;\theta)}{\sum_{h'}p(c_i|z_{h'};\theta)p(g_j|z_{h'};\theta)p(z_{h'};\theta)},$$

$$p(z_h|g_j,g_{j'};\theta)=\frac{p(g_j|z_h;\theta)p(g_{j'}|z_h;\theta)p(z_h;\theta)}{\sum_{h'}p(g_j|z_{h'};\theta)p(g_{j'}|z_{h'};\theta)p(z_{h'};\theta)}.$$

M-step:

$$\theta_{c_i|z_h}\propto\sum_j N_{i,j}\,p(z_h|c_i,g_j;\theta_{\text{old}}),$$

$$\theta_{g_j|z_h}\propto\pi_{\text{CG}}\sum_i\frac{N_{i,j}}{N_{\text{CG}}}p(z_h|c_i,g_j;\theta_{\text{old}})+\pi_{\text{GG}}\sum_{j'}\frac{M_{j,j'}}{N_{\text{GG}}}p(z_h|g_j,g_{j'};\theta_{\text{old}}),$$

$$\theta_{z_h}\propto\pi_{\text{CG}}\sum_{i,j}\frac{N_{i,j}}{N_{\text{CG}}}p(z_h|c_i,g_j;\theta_{\text{old}})+\pi_{\text{GG}}\sum_{j',j''}\frac{M_{j',j''}}{N_{\text{GG}}}p(z_h|g_{j'},g_{j''};\theta_{\text{old}}).$$

3.4.3 3MAM Finally, we show the case in which we use all the three types of co-occurrence data such as compound–gene, gene–gene and compound–compound pairs. We use 3MAM, and so the log-likelihood for these datasets is also given in Section 3.3. The E- and M-steps for 3MAM can be given as follows:

E-step:

$$p(z_h|c_i,g_j;\theta)=\frac{p(c_i|z_h;\theta)p(g_j|z_h;\theta)p(z_h;\theta)}{\sum_{h'}p(c_i|z_{h'};\theta)p(g_j|z_{h'};\theta)p(z_{h'};\theta)},$$

$$p(z_h|g_j,g_{j'};\theta)=\frac{p(g_j|z_h;\theta)p(g_{j'}|z_h;\theta)p(z_h;\theta)}{\sum_{h'}p(g_j|z_{h'};\theta)p(g_{j'}|z_{h'};\theta)p(z_{h'};\theta)},$$

$$p(z_h|c_i,c_{i'};\theta)=\frac{p(c_i|z_h;\theta)p(c_{i'}|z_h;\theta)p(z_h;\theta)}{\sum_{h'}p(c_i|z_{h'};\theta)p(c_{i'}|z_{h'};\theta)p(z_{h'};\theta)}.$$

M-step:

$$\theta_{c_i|z_h}\propto\pi_{\text{CG}}\sum_j\frac{N_{i,j}}{N_{\text{CG}}}p(z_h|c_i,g_j;\theta_{\text{old}})+\pi_{\text{CC}}\sum_{i'}\frac{L_{i,i'}}{N_{\text{CC}}}p(z_h|c_i,c_{i'};\theta_{\text{old}}),$$

$$\theta_{g_j|z_h}\propto\pi_{\text{CG}}\sum_i\frac{N_{i,j}}{N_{\text{CG}}}p(z_h|c_i,g_j;\theta_{\text{old}})+\pi_{\text{GG}}\sum_{j'}\frac{M_{j,j'}}{N_{\text{GG}}}p(z_h|g_j,g_{j'};\theta_{\text{old}}),$$

$$\theta_{z_c}\propto\pi_{\text{CG}}\sum_{i,j}\frac{N_{i,j}}{N_{\text{CG}}}p(z_h|c_i,g_j;\theta_{\text{old}})+\pi_{\text{GG}}\sum_{j',j''}\frac{M_{j',j''}}{N_{\text{GG}}}p(z_h|g_{j'},g_{j''};\theta_{\text{old}})+\pi_{\text{CC}}\sum_{i',i''}\frac{L_{i',i''}}{N_{\text{CC}}}p(z_h|c_{i'},c_{i''};\theta_{\text{old}}).$$

Table 1. The size of co-occurrence datasets

Item	Size
Medline records	63 940
Gene type	22 292
Gene–gene	174 077
Chemical compound type	3454
Compound–compound	20 443
Compound–gene	47 217

4 EXPERIMENTAL RESULTS

4.1 Data

MAM can incorporate any type of co-occurrence data, but as mentioned in Section 1, in this paper we focus on the existing literature only. We derived our datasets from all records that have been stored in Medline and were published from 1960 to 2004.

We first used the 'Locus ID' (Pruitt and Maglott, 2001) to check if a gene is in an abstract, by using a list of links which is available at ftp://ftp.ncbi.nih.gov/refseq/LocusLink in December 2004. Each link connects a Locus ID with a PubMed ID. We focused on 'human' genes only and selected Medline records containing one or more human genes using this list. We then generated co-occurrence data on genes from the selected Medline records. In order to produce meaningful gene–gene co-occurrence pairs, we skipped the Medline records, each of which has more than 103 genes. This is because some Medline records report all genes in the microarray experiment[1].

We then used the CAS Registry numbers as defined in the records of Medline to find a chemical compound in a document. Using the selected Medline records, we generated co-occurrence data on compound pairs. (For the details of the CAS Registry numbers, see www.cas.org/EO/regsys.html.) We finally generated co-occurrence data on compound–gene pairs from the selected Medline records using both the CAS registry numbers and the above link list of genes. We used this list on Locus IDs because it is a curated list and is probably the most reliable data source, to the best of our knowledge. Table 1 shows the sizes of the three co-occurrence datasets. We note that 'Medline records' in the table is the number of Medline records that we used to derive our three types of datasets.

4.2 Performance evaluation by cross-validation

4.2.1 Evaluation procedure We evaluated the performance of MAM using cross-validation on predicting compound–gene pairs.

We tested four types of models to predict compound–gene pairs. That is, we first tested AM using the co-occurrence data of compound–gene pairs only, and then tested two different 2MAM by adding compound–compound [2MAM (CG + CC)] and gene–gene [2MAM (CG + GG)] pairs. Finally, we made use of all three types of co-occurrence data to train 3MAM.

To examine the effect of the size of the training dataset to the performance of the probabilistic model, we set five different ratios of the size of training to test data, 3:1, 2:1, 1:1, 1:2 and 1:3, in the cross-validation experiment. For example, in the 3:1 case, we randomly divide the original compound–gene data into four subsets

[1]For example, Medline ID 12477932 has more than 9000 human genes.

Table 2. Percentage of the AUCs and the t-values (in parentheses) obtained by 50 rounds of cross-validation on compound–gene pairs

Model	Ratio of training to test data				
	3:1	2:1	1:1	1:2	1:3
3MAM (CG + CC + GG)	**96.0**	**95.5**	**94.5**	**92.8**	**91.5**
2MAM (CG + CC)	95.0 (**81.4**)	94.5 (**73.9**)	93.2 (**60.3**)	91.1 (**88.6**)	89.6 (**94.9**)
2MAM (CG + GG)	92.3 (**193.8**)	91.6 (**168.0**)	89.8 (**158.6**)	87.7 (**209.2**)	86.4 (**197.4**)
AM (CG)	89.0 (**232.2**)	88.0 (**202.4**)	86.0 (**190.5**)	83.6 (**285.5**)	82.0 (**357.4**)

of roughly equal size, and then alternatively select one subset as the test data and the other three subsets as training data. We carried out 50 rounds of this cross-validation to reduce possible biases occurring in only a few rounds and averaged the results obtained. When we add another type of training data, keeping the same training compound–gene pairs for each round of cross-validation, we added one or more other types of co-occurrence data to train 2MAM or 3MAM. Then, the prediction was performed on the same test dataset.

We note that AM cannot make any predictions on a compound–gene pair in the test data if one component of this pair does not appear in the training data. Thus, we removed all such co-occurrence pairs in the test data, and the remaining pairs were used as positive test examples[2]. We then randomly generated the same number of compound–gene pairs which are not found in both training and test as negative test examples. We checked the performance of each of the models tested by the ability to discriminate positive from negative test examples.

4.2.2 Evaluation measures Once we estimated the probability parameters of a probabilistic model from training data, we computed the likelihood of each compound–gene pair in test data and ranked all pairs according to their likelihoods. We evaluated these ranked pairs by AUC (area under the ROC curve).

An ROC (receiver operator characteristic) curve is drawn by plotting 'sensitivity' against 'false positive rate', using the ranked compound–gene pairs. The sensitivity (or true positive rate) is the proportion of the number of correctly predicted positive examples to the total number of positive examples. The false positive rate is the proportion of the number of false positive examples to the total number of negative examples. The AUC, a popular metric for measuring the performance of different models (Bradley, 1997), can be computed as the area under this ROC curve. We can see that the larger the AUC, the better the performance of the model. We further used the paired sample two-tailed t-test to statistically evaluate the performance difference of the two models. Since we run cross-validation 50 times, we have at least 100 values in each of the five different ratios, and so if the t-value is >3.50 then the difference is >99.9% statistically significant in any ratio.

4.2.3 Parameter settings The stopping condition we adopted for our EM estimation was when the improvement of the observed log-likelihood between two successive EM iteration is <0.001. We used a uniform distribution for the weights (i.e. π) of both 2MAM and 3MAM in all cases. As mentioned in Section 1, MAM is a space-efficient model, but we note that our datasets require a huge memory space. When we set $H = 128$, there were altogether ~3 250 000 [$\fallingdotseq (22\,292 + 3454) \times 128$ for $p(g|z)$ and $p(c|z)$] parameters to be estimated for 3MAM. It costed around 800 MB of memory and took ~30 min to execute on a Linux workstation with dual Intel Xeon 3.0 GHz processors and 8 GB of main memory.

[2]We emphasize that this experimental setting is advantageous to AM and not to MAM.

Table 3. Percentage of the AUCs and the t-values (in parentheses) obtained by predicting biological-related compound–gene pairs in ChEBI

Model	3MAM (CG + CC + GG)	2MAM (CG + CC)	2MAM (CG + GG)	AM (CG)
AUC (t-value)	**86.3**	86.1 (1.42)	82.3 (**21.6**)	80.1 (**30.9**)

4.2.4 Results Table 2 shows the AUC for each model at different data settings and the t-values between the AUC of 3MAM and that of another model. This table clearly showed that 3MAM outperformed the other three models and its performance was followed by 2MAM (CG + CC), 2MAM (CG + GG) and AM. We note that the difference between 3MAM and AM reached ~7–10%. This performance improvement is significant, because the AUC of AM reached 82–89% already, and so it is usually hard to improve these values. Furthermore, the t-values showed that 3MAM outperformed all other models by a statistically significant factor in all cases. These results indicate that incorporating compound–compound and gene–gene pairs improved the predictive performance obtained by compound–gene pairs only. Another empirical finding from these results is that incorporating compound–compound pairs was more effective in improving the predictive performance than incorporating gene–gene pairs, even though the size of the compound–compound pairs is smaller than (in fact, less than one-eighth of) that of gene–gene pairs.

4.3 Performance evaluation by ChEBI

4.3.1 Test data description We further examined the performance of our probabilistic models by using an independent dataset, i.e. we trained our models using our datasets generated from the Medline records, and the trained models were applied to another dataset.

We used a list of the relationships between chemical compounds and proteins in the ChEBI database to generate a test dataset. We first extracted 26 091 pairs of chemical compounds and human genes from this list by replacing a protein with one or more corresponding genes according to the UniProt database (Bairoch *et al.*, 2005). We then removed a compound–gene pair from them if it appears in our training dataset or if one component of this pair does not appear in the

Table 4. Top 20 pairs of drugs and genes

CAS registry number	Drug name	Locus ID	Gene name	Log-likelihood
19545-26-7	Wortmannin	5594	*MAPK1*: Mitogen-activated protein kinase 1	−2.615
16561-29-8	Tetradecanoylphorbol acetate	5590	*PRKCZ*: Protein kinase C, zeta	−2.764
23214-92-8	Doxorubicin	1029	*CDKN2A*: Cyclin-dependent kinase inhibitor 2A	−2.992
73-22-3	Tryptophan	5705	*PSMC5*: Proteasome 26S subunit	−3.000
10102-43-9	Nitric Oxide	959	*TNFSF5*: Tumor necrosis factor, member 5	−3.027
66-81-9	Cycloheximide	5970	*RELA*: V-rel reticuloendotheliosis viral oncogene homolog A	−3.030
33419-42-0	Etoposide	4193	*MDM2*: Transformed 3T3 cell double minute 2	−3.033
50-02-2	Dexamethasone	3458	*IFNG*: Interferon, gamma	−3.037
15663-27-1	Cisplatin	581	*BAX*: BCL2-associated X protein	−3.060
521-18-6	Dihydrotestosterone	2099	*ESR1*: Estrogen receptor 1	−3.061
53-85-0	Dichlororibofuranosylbenzimidazole	2963	*GTF2F2*: General transcription factor IIF, polypeptide 2	−3.103
50-07-7	Mitomycin	7157	*TP53*: Tumor protein p53	−3.104
320-67-2	Azacitidine	6622	*SNCA*: Synuclein, alpha	−3.111
33069-62-4	Paclitaxel	581	*BAX*: BCL2-associated X protein	−3.148
133407-82-6	Leucine aldehyde	7124	*TNF*: Tumor necrosis factor, member 2	−3.203
10540-29-1	Tamoxifen	5241	*PGR*: Progesterone receptor	−3.208
7722-84-1	Hydrogen peroxide	596	*BCL2*: B-cell CLL/lymphoma 2	−3.213
67526-95-8	Thapsigargin	5580	*PRKCD*: Protein kinase C, delta	−3.215
59-14-3	Bromodeoxyuridine	1027	*CDKN1B*: Cyclin-dependent kinase inhibitor 1B	−3.221

compound–gene pairs in the training dataset of AM for the benefit of AM. The number of final compound–gene pairs was 11 743, and we used them as positive examples in the test dataset. We generated the same number of negative examples, which are not in the training dataset and the positive test dataset.

4.3.2 Evaluation procedure We trained four different models, AM (CG), 2MAM (CG + GG), 2MAM (CG + CC) and 3MAM (CG + CC + GG) using the datasets generated from the Medline records and then predictions were performed on the test dataset generated in the above manner. We randomly generated the negative test dataset 50 times and the results were averaged over the 50 runs to reduce a possible bias. The evaluation measures and parameter settings were the same as those of the cross-validation experiment.

4.3.3 Results Table 3 shows the average AUC for each model and the *t*-values between the AUC of 3MAM and that of another model. This table clearly showed that 3MAM outperformed three other models, and was followed by 2MAM (CG + CC), 2MAM (CG + GG) and AM. This result indicates that our model is extremely effective to improve the predictive performance for an independent dataset as well. The improvement over 2MAM (CG + GG) and AM were especially significant, but the advantage over 2MAM (CG + CC) was slight (statistically insignificant). This result confirms the empirical finding that incorporating compound–compound pairs was more effective in improving the predictive performance than incorporating gene–gene pairs.

4.4 Mining and analyzing unknown drug-gene relationships

We trained 3MAM by using all datasets, i.e. all gene–gene, compound–compound and compound–gene pairs, and computed the likelihood of each of all possible compound–gene pairs that are not in the training dataset. We repeated this run 50 times and computed the sum of all 50 likelihoods for each compound–gene pair. We sorted all the pairs according to their likelihoods and removed the pairs whose chemical compounds do not have pharmacological activity. Thus each chemical compound of the remaining pairs is a drug. From this list of pairs, we finally selected the top 20 pairs so that a drug did not appear again when we scanned the sorted pairs from the top to the bottom. Table 4 shows the list of these 20 pairs with their log-likelihoods.

These pairs are unknown pairs in the literature, but our method suggested that each of them has a strong relationship between a drug and a gene. In fact, we can see a biological relationship for each pair of a drug and a gene. Below, we briefly describe the biological, medical and pharmaceutical relationship on each pair of the list, for only the top five pairs owing to the space limitations.

Wortmannin, of the first pair, is an inhibitor for the phosphatidylinositide kinase (PI3-kinase) pathway. Substrates of PI kinases are important signaling molecules that affect a wide range of biological processes (Zewail *et al.*, 2003). In particular, recent studies have revealed that wortmannin affects proteasome-mediated degradation and chromatin remodeling. MAPK1 is also a popular kinase, and so wortmannin is expected to be an inhibitor of the MAPK pathway too.

The second pair of tetradecanoylphorbol acetate and protein kinase C seems to have more direct relation than that of the first pair. A pharmacological action of tetradecanolyphorbol acetate is a carcinogen. However, a number of processes important in certain diseases, such as solid tumors, are facilitated by the action of protein kinase C, and so inhibitors of protein kinase C have the potential to be anticancer drugs (Shih *et al.*, 1999). Tetradecanolyphorbol acetate might become one of them.

Doxorubicin of the third pair is known as an antineoplastic agent, i.e. a drug intended to inhibit or prevent the maturation and proliferation of neoplasms that may become malignant, by targeting

the DNA. However, CDKN2A is an inhibitor of cyclic-dependent kinase (CDK) which is involved in replicative senescence, cell immortalization and tumor generation. Thus, these facts imply that the two molecules of this pair are strongly related.

A biological relation on the fourth pair of tryptophan and proteasome 26s subunit would be easily expected, because 26S proteasomes are mainly involved in the degradation of tryptophan hydroxylase (Kojima *et al.*, 2002). The fifth pair would also have a clear biological relation. A member of the tumor necrosis factors can be a ligand for CD40, and CD40 ligation can stimulate nitric oxide production (Bingaman *et al.*, 2000).

These facts indicate that unknown pairs that are predicted by our method to be related with each other have, in fact, strong biological relations. This analysis would be possible for most of the pairs in this list. Above all, the analysis on these pairs show that our probabilistic model and its learning algorithm can capture significant relationships between chemical compounds (more specifically, drugs) and genes.

5 CONCLUDING REMARKS

We have proposed a probabilistic model, composed of a mixture of aspect models, each of which is for one type of co-occurrence data, coupled with its learning algorithm. Our model can combine a number of different types of co-occurrence data efficiently, and in fact, our experimental results have shown that incorporating different types of datasets improved the predictive performance drastically.

In our experiments, we used a uniform distribution for the component weights (π) of our mixture model to allow users additional control. Interesting future work would adjust the weights to achieve the maximum predictive performance. It would also be interesting to investigate the possibility of incorporating totally different types of co-occurrence data (e.g. microarray co-expressions and protein–protein interactions) to the current literature data. Another appealing extension is to extract large scale co-occurrence data directly from Medline by incorporating latest natural language processing (NLP) techniques.

ACKNOWLEDGEMENTS

This work is supported in part by Bioinformatics Education Program 'Education and Research Organization for Genome Information Science' and Kyoto University 21st Century COE Program 'Knowledge Information Infrastructure for Genome Science' with support from MEXT (Ministry of Education, Culture, Sports, Science and Technology), Japan.

Conflict of Interest: none declared.

REFERENCES

Bairoch,A. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.

Bingaman,A. *et al.* (2000) The role of CD40L in T cell-dependent nitric oxide production by murine macrophages. *Transpl. Immunol.*, **8**, 195–202.

Blaschke,C. *et al.* (1999) Automatic extraction of biological information from scientific text: protein–protein interactions. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **7**, 60–67.

Blaschke,C. *et al.* (2002) Information extraction in molecular biology. *Brief Bioinform.*, **3**, 154–165.

Bradley,A. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30**, 1145–1159.

Brooksbank,C. *et al.* (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.*, **33**, D46–D53.

Chang,J.T. and Altman,R.B. (2004) Extracting and characterizing gene-drug relationships from the literature. *Pharmacogenetics*, **14**, 577–586.

Dempster,A. *et al.* (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, **39**, 1–38.

Hofmann,T. (2001) Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, **42**, 177–196.

Hofmann,T. (2004) Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, **22**, 89–115.

Jenssen,T. *et al.* (2001) A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.*, **28**, 21–28.

Kanehisa,M. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.

Kojima,M. *et al.* (2000) Rapid turnover of tryptophan hydroxylase is driven by proteasomes in RBL2H3 cells, a serotonin producing mast cell line. *J. Biochem. (Tokyo)*, **127**, 121–127.

Perez-Iratxeta,C. *et al.* (2002) Association of genes to genetically inherited diseases using data mining. *Nat. Genet.*, **31**, 316–319.

Pruitt,K. and Maglott,D. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137–140.

Rebholz-Schuhmann,D. *et al.* (2004) Automatic extraction of mutations from Medline and cross-validation with OMIM. *Nucleic Acids Res.*, **32**, 135–142.

Rindflesch,T. *et al.* (2000) EDGAR: extraction of drugs, genes and relations from biomedical literature. *Pac. Symp. Biocomput.*, **5**, 517–528.

Shih,S.C. *et al.* (1999) Role of protein kinase C isoforms in phorbol ester-induced vascular endothelial growth factor expression in human glioblastoma cells. *J. Biol. Chem.*, **274**, 15407–15414.

Si,L. and Jin,R. (2003) Flexible mixture model for collaborative filtering. In *Proceedings of the 12th International Conference on Machine Learning*, August 21–24, 2003, Washington, DC, USA. AAAI Press, pp. 704–711.

Stapley,B. and Benoit,G. (2000) Biobibliometrics: information retrieval and visualization from co-occurrence of gene names in MEDLINE abstracts. *Pac. Symp. Biocomput.*, **5**, 529–540.

Wheeler,D. *et al.* (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **33**, D39–D45.

Wren,J.D. *et al.* (2004) Knowledge discovery by automated identification and ranking of implicit relationships. *Bioinformatics*, **20**, 389–398.

Yandell,M.D. and Majoros,W.H. (2002) Genomics and natural language processing. *Nat. Rev. Genet.*, **3**, 601–610.

Zewail,A. *et al.* (2003) Novel functions of the phosphatidylinositol metabolic pathway discovered by a chemical genomics screen with wortmannin. *Proc. Natl Acad. Sci. USA*, **100**, 3345–3350.

Prostaglandin E_2 Stimulates Granulocyte Colony-Stimulating Factor Production via the Prostanoid EP2 Receptor in Mouse Peritoneal Neutrophils[1]

Yukihiko Sugimoto,[2]* Yoko Fukada,[2]* Daisuke Mori,[2]* Satoshi Tanaka,* Hana Yamane,* Yasushi Okuno,† Katsuya Deai,* Soken Tsuchiya,* Gozoh Tsujimoto,† and Atsushi Ichikawa[3]*‡

G-CSF is a hemopoietic growth factor involved in granulocytic differentiation of progenitor cells. In this study, we investigated the effects of PGE_2 on G-CSF production in murine peritoneal neutrophils in vitro and in vivo. PGE_2 augmented LPS-primed G-CSF release from peritoneal neutrophils. This augmentation was mimicked by a type E prostanoid receptor (EP)2-selective agonist but not by other EP-specific agonists. Indeed, the effect of PGE_2 on G-CSF release was abolished in neutrophils isolated from *EP2*-deficient mice. PGE_2 and an EP2 agonist have the ability to stimulate *G-CSF* gene expression even in the absence of LPS. In the casein-induced peritonitis model, the appearance of G-CSF in the casein-injected peritoneal cavity associated well with the timing of neutrophil infiltration as well as PGE_2 levels in exudates, with a peak value at 6 h postinjection. Inhibition of endogenous PG synthesis by indomethacin resulted in a marked decrease in G-CSF content and neutrophil number in the peritoneal cavity. Moreover, *EP2*-deficient mice exhibited a strikingly reduced G-CSF content in peritoneal exudates with comparable responses in neutrophil migration and local PGE_2 production at 6 h postinjection. These results suggest that the PGE_2-EP2 system contributes to the local production of G-CSF during acute inflammation. *The Journal of Immunology,* **2005, 175: 2606–2612.**

Prostaglandin E_2 is one of the most important arachidonate metabolites synthesized by the action of cyclooxygenase (1). This lipid mediator is involved in a wide range of diseases, including inflammation, by exerting pleiotropic actions (2). Administration of PGE_2 alone does not cause any significant responses, demonstrating that PGE_2 on its own has little inflammatory capacity. In contrast, in the presence of other mediators, PGE_2 can synergistically amplify the local inflammatory response. For instance, PGE_2 has been shown to enhance zymosan-stimulated IL-10 production in macrophages (3). In contrast to these proinflammatory activities, PGs are also known to inhibit the production of proinflammatory cytokines by macrophages activated with LPS; PGE_2 has been shown to inhibit LPS-induced IL-12 production and LPS-stimulated TNF-α production in resident macrophages (4, 5). Thus, PGE_2 works as a modulator of cellular immune responses initiated by other stimulants.

G-CSF is a member of a family of hemopoietic growth factors that are required for proliferation and differentiation of hemopoietic progenitor cells (6, 7). Administration of G-CSF increases peripheral blood neutrophil counts in many species, including humans (8, 9). G-CSF is released by monocytes and endothelial cells in response to proinflammatory inputs such as LPS and TNF-α (7, 10). Such inflammation-induced G-CSF has been considered to contribute to the maintenance of a number of neutrophil lineage cells in the bone marrow (7, 11). Indeed, in the casein-induced peritonitis model, G-CSF is produced in the peritoneal cavity at the time of neutrophil migration, suggesting that neutrophils can produce G-CSF for this purpose (11). PGE_2 has been shown to enhance LPS-induced G-CSF formation in human monocyte or macrophage-like THP-1 cells (12, 13). However, it remains unknown whether PGE_2 modulates G-CSF formation from neutrophils, and there are no reports evaluating the contribution of PGE_2 to local G-CSF production within an inflammatory area. The biological actions of PGE_2, including its effects on cellular immune responses, are mediated by the G protein-coupled receptor, type E prostanoid receptor (EP; reviewed in Refs. 14 and 15). EP can be divided into four distinct pharmacological classes: EP1, EP2, EP3, and EP4. Among them, both the EP2 and EP4, which are coupled to the stimulation of adenylyl cyclase, have been shown to exist throughout a wide range of immune cells, including dendritic cells (5, 16, 17). However, the precise EP subtypes that mediate the immunoregulatory actions of PGE_2 remain to be established.

In this study, we have investigated the effects of PGE_2 on G-CSF production from peritoneal polymorphonuclear leukocytes (PMN[4]) (3) in vitro and in vivo. We show that PGE_2 in the absence of LPS stimulation is capable of inducing the production of G-CSF by peritoneal PMNs. Although both EP2 and EP4 are expressed in peritoneal PMNs, this effect of PGE_2 is mediated by EP2 but not by EP4. Moreover, EP2 signaling was found to be a physiological stimulus of the local production of G-CSF within an inflammatory site.

*Department of Physiological Chemistry and †Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan; and ‡School of Pharmaceutical Sciences, Mukogawa Women's University, Hyogo, Japan

Received for publication January 28, 2004. Accepted for publication June 3, 2005.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported in part by a grant from the Sankyo Foundation of Life Science, Grant-in-Aid for Scientific Research on Priority Areas (14013036, 14021053, 15019050, 15012234, and 16012238) from the Ministry of Education, Culture, Sports Science, and Technology of Japan, and from the Ministry of Health and Labor of Japan.

[2] Y.S., Y.F., and D.M. contributed equally to this work.

[3] Address correspondence and reprint requests to Dr. Atsushi Ichikawa, School of Pharmaceutical Sciences, Mukogawa Women's University, 11-68, Kyuban-chou, Koshien, Nishinomiya, Hyogo 663-8179, Japan. E-mail address: aichikaw@mwu.mukogawa-u.ac.jp

[4] Abbreviations used in this paper: PMN, polymorphonuclear leukocyte; CHO, Chinese hamster ovary; dbcAMP, dibutyryl cAMP.

0022-1767/05/$02.00

Materials and Methods

Animals

Female C57BL/6 mice (12 wk of age) purchased from Japan SLC were used as wild-type mice. The generation of EP2-deficient (*ptger2*$^{-/-}$) mice has been described previously (18). EP2-deficient mice, backcrossed for 10 generations to C57BL/6 mice, were maintained on a 12-h light/dark cycle under specific pathogen-free conditions. All experimental procedures were approved by the Committee on Animal Research of Kyoto University Faculty of Pharmaceutical Sciences.

Materials

LPS from *Escherichia coli* O55:B5, casein, dibutyryl cAMP, and indomethacin were obtained from Sigma-Aldrich. The ^{125}I-labeled cAMP assay system was purchased from Amersham Bioscience, and the ELISA kits for G-CSF and MIP-2 quantification were purchased from R&D Systems. PGE_2 in the peritoneal fluid was quantified using an enzyme immunoassay kit (Cayman Chemical). PGE_2 was purchased from Funakoshi. RPMI 1640 medium and FBS were from Invitrogen Life Technologies (LPS < 30 pg/ml). The EP-specific agonists, DI-004, AE1–259, AE-248, and AE1–329 were generous gifts from Ono Pharmaceutical. The specificities of the agonists were analyzed by measuring the binding affinity to the respective EP subtype expressed in Chinese hamster ovary (CHO) cells, as reported previously (5, 19). PGE_2, EP agonists, dibutyryl cAMP (dbcAMP), other reagents, and culture media were confirmed to be endotoxin-free (<0.1 EU/ml endotoxin) using the *Limulus* Amoebocyte Lysate assay (Endospecy; Seikagaku Kogyo). All other chemicals were commercial products of reagent grade.

Preparation of peritoneal PMNs and immunostaining

Female mice were injected i.p. with 2 ml of 5% (w/v) casein in sterile saline and were sacrificed by cervical dislocation 6 h after injection. The lavage fluid (4 ml) was collected in a syringe, and exudated peritoneal cells were precipitated by centrifugation. PMNs were further purified from the peritoneal cells by Percoll stepwise density gradient (1.090 and 1.070 g/ml) centrifugation (600 × *g* for 20 min at 4°C). The purity of PMNs was >95% as determined by staining with May-Grünwald-Giemsa. PMNs were suspended in RPMI 1640 medium containing 10% heat-inactivated FBS, 150 μM 2-ME, and 100 μM sodium pyruvate. Immunofluorescence studies were performed as described previously (20). LPS-stimulated PMNs were centrifuged onto round coverglasses (ϕ = 18.0 mm), which were then placed in 12-well culture plates. For G-CSF staining, a goat anti-G-CSF Ab (1:50) (Santa Cruz Biotechnology) was used as the primary Ab, and a rhodamine-conjugated donkey anti-goat IgG Ab (1:200) (Chemicon International) was used as the secondary antibody. A FITC-conjugated rat anti-Gr-1 Ab (1:200) (BD Pharmingen) was used for Gr-1 staining. The stained cells were observed using a confocal laser scanning microscope, LSM5Pascal (Carl Zeiss). No significant staining was observed when cells were incubated without primary Ab (G-CSF) or incubated with FITC-conjugated isotype control (Gr-1) (Fig. 1*A*).

Measurement of G-CSF production and cAMP accumulation

PMNs (5 × 10^5 cells/well) were incubated with medium containing FBS (10%) or BSA (1%) with or without 100 ng/ml LPS for the indicated times at 37°C in 5% CO_2. After incubation, each culture was centrifuged at 300 × *g* for 5 min at 4°C to remove the cells. The amounts of G-CSF in the supernatant were assayed using ELISA kits, according to the manufacturer's instructions (R&D Systems). The values obtained from the medium only were used as the value at 0 h. The cAMP levels in PMNs were determined as described previously (5). Cells (5 × 10^5 cells/well) were washed with Krebs-HEPES buffer (pH 7.4) containing 100 μM Ro-20-1724 and preincubated for 10 min. Reactions were started by the addition of test reagents along with 100 μM Ro-20-1724. After incubation for the indicated time at 37°C, reactions were terminated by the addition of 10% trichloracetic acid. The cAMP content of the cells was then measured by the cAMP radioimmunoassay kit (Amersham Biosciences).

RT-PCR and microarray analysis

For RNA extraction, PMNs (3.5–5 × 10^7 cells/ml) were collected and total RNA prepared by the acid guanidinium thiocyanate-phenol chloroform method (21). For RT-PCR analysis, cDNA was amplified using primers specific for G-CSF and GAPDH. GAPDH primers were purchased from Invitrogen Life Technologies. The primer sequences for G-CSF were 5′-CTGTGGCAAAGTGCACTATGGTCAGGACG-3′ (sense primer) and 5′-GGATGTTGCCAACTTTGCCACCACCAT-CTG-3′ (anti-sense primer). For microarray analysis, total RNA was isolated by a combination of the guanidinium thiocyanate-phenol chloroform method and RNeasy column chromatography (Qiagen). The obtained RNA was amplified, labeled, and prepared for hybridization to GeneChip Murine Expression 430 oligonucleotide arrays (Affymetrix) using standard methods as described previously (22).

Casein-induced peritonitis

Mice were injected i.p. with 2 ml of a 5% (w/v) solution of casein in PBS. Indomethacin (10 mg/kg) was s.c. injected 2 h before the casein treatment. In additional experiments, peritoneal inflammatory responses were induced by an i.p. injection of 2 ml of 0.3% sodium thioglycollate (w/v in distilled water). At the indicated times after injection, mice were sacrificed by cervical dislocation. The abdominal cavity was then injected with 4 ml of PBS and massaged to ensure adequate mixing of the cell population with the harvesting fluid. Cells in the peritoneal cavity were then collected using a syringe. Cytocentrifuge preparations of peritoneal cells were made and then stained with May-Grünwald-Giemsa. The G-CSF and MIP-2 contents in harvested peritoneal washout fluids were assayed with ELISA kits, and PGE_2 contents were assayed with an enzyme immunoassay kit (Cayman Chemical).

Statistical analysis

Data are shown as means ± SEM. Comparison of two groups was performed by the Student's *t* test. For comparison of more than two groups with comparable variances, one-way ANOVA was performed first, and then the Dunnett's test was used to evaluate the pairwise group difference. The presented results are representative of at least three independent experiments.

Results

Effects of PGE_2 on LPS-stimulated G-CSF release from mouse peritoneal PMNs

To uncover the potential roles of PGE_2 in neutrophils, we examined the effects of PGE_2 on LPS-stimulated production of G-CSF in mouse peritoneal PMNs. In PMNs purified from mouse peritoneal cells, G-CSF was immediately produced upon LPS treatment. More than 95% of the PMN preparation were Gr-1^+ cells, all of which were also stained with the anti-G-CSF Ab (Fig. 1*A*). No significant staining was observed in Gr-1^- cell or when cells were incubated with FITC-conjugated isotype control in the absence of anti-G-CSF Ab. PGE_2 at 1 μM augmented the effects of LPS on G-CSF production by ~2- to 3-fold (Fig. 1*B*). Such an enhancing effect of PGE_2 on G-CSF production was observed even at 10 nM, but carbacyclin, a stable prostacyclin analog, failed to enhance G-CSF production even at 1 μM (Fig. 1*C*). We further tested the possibility that the PGE_2 synthesized by neutrophils themselves may stimulate LPS-induced G-CSF production. However, pretreatment of the cells with indomethacin failed to alter the LPS-induced G-CSF release from neutrophils.

PGE_2 augmented LPS-induced G-CSF production via the EP2 by stimulating cAMP accumulation in mouse peritoneal PMNs

There are four subtypes of PGE_2Rs (EPs), which differ in their signal transduction pathways. We previously demonstrated that peritoneal PMNs express mRNAs for the Gs-coupled EPs, EP2 and EP4, but not for the Gq-coupled EP1 nor the Gi-coupled EP3 (23). Indeed, when we performed the cAMP formation assay on peritoneal PMNs, PGE_2 dose dependently increased cAMP content. In addition, AE1-329, an EP4 agonist, as well as AE1-259, an EP2 agonist, elicited cAMP accumulation, indicating that PMNs express functional EP2 and EP4 (Fig. 2*A*). We next examined which EP agonist can mimic the PGE_2 stimulation of G-CSF production in PMNs. However, among the four EP-selective agonists, only an EP2 agonist augmented LPS-induced G-CSF production while an EP4 agonist could not (Fig. 2*B*). In the PMNs isolated from *EP2*-deficient mice, PGE_2, as well as an EP2 agonist, failed to augment LPS-induced G-CSF production, although a membrane-permeable cAMP analog, dbcAMP, augmented the LPS-induced G-CSF response as observed in PMNs from wild-type mice

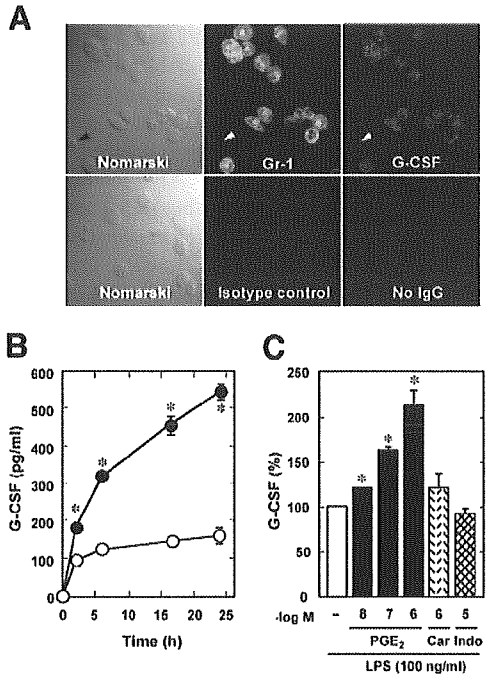

FIGURE 1. Exogenous PGE$_2$ augments LPS-induced G-CSF release from peritoneal PMNs. *A*, Expression of G-CSF in Gr-1$^+$ cells. LPS-treated cells were double-stained with an anti-G-CSF Ab (G-CSF) and an anti-Gr-1 Ab (Gr-1) (*upper panels*). No staining was observed in Gr-1$^-$ cell (arrowheads) or when cells were incubated with FITC-conjugated isotype control (isotype control) in the absence of anti-G-CSF Ab (no IgG). *B*, Time course of LPS-stimulated G-CSF release. PMNs (1×10^6 cells/well) were stimulated with 100 ng/ml LPS in the presence (●) or in the absence of 1 μM PGE$_2$ (○), and G-CSF levels in the medium were determined. *, $p < 0.05$ vs vehicle. *C*, Effects of PGE$_2$, carbacyclin, and indomethacin on LPS-induced G-CSF release. PMNs were preincubated with (Indo) or without indomethacin for 1 h, followed by incubation in medium containing LPS supplemented with vehicle (−), carbacyclin (Car), or PGE$_2$ for 6 h. The G-CSF levels are represented as percentages of vehicle. *, $p < 0.05$ vs vehicle.

(Fig. 2, *C* and *D*). These results suggested that PGE$_2$ stimulates G-CSF production via the EP2 through an increase in cAMP levels.

PGE$_2$ itself stimulates G-CSF gene expression in peritoneal PMNs

We next examined whether PGE$_2$ has the ability to stimulate G-CSF release from PMNs in the absence of LPS. When the PMNs were cultured in medium supplemented with 1% BSA in the absence of LPS and FBS, G-CSF was undetectable in the culture medium. Even under such conditions, PGE$_2$, as well as an EP2 agonist, stimulated G-CSF production of PMNs (Fig. 3*A*). dbcAMP also stimulated G-CSF release in these cells (data not shown). To test the possibility that PGE$_2$ via cAMP accumulation stimulates G-CSF gene expression, we examined the effects of EP agonists and dbcAMP on G-CSF mRNA levels in PMNs by using semiquantitative PCR. Although we did not detect G-CSF in the serum-free medium, the vehicle-treated cells showed a low level of basal expression. The cells exposed to 1 μM PGE$_2$ or 1 μM of an EP2 agonist for 6 h showed an ~5-fold increase in expression of G-CSF mRNA (Fig. 3*B*). The cells treated with dbcAMP showed an 8.5-fold increase, whereas treatment with an EP4 agonist failed to significantly increase G-CSF mRNA levels. These results suggested that PGE$_2$ stimulates G-CSF production via the EP2 at least in part by increasing G-CSF mRNA levels in PMNs.

To estimate the contribution of EP2 or cAMP signaling on the effect of PGE$_2$ on PMNs, we examined the changes in gene expression profiles of PMNs treated with PGE$_2$, an EP2 agonist, an EP4 agonist or dbcAMP. Of the ~24,000 genes represented on the oligonucleotide array, genes with altered expression values upon PGE$_2$ treatment (log ratio of PGE$_2$ vs control was >1.3 or less than −1.3) were selected and regarded as PGE$_2$-regulated genes (193 genes) (Fig. 3*C*). Approximately 80% of the PGE$_2$-regulated genes (151 genes) were regulated by dbcAMP in the same manner. An EP2 agonist also mimicked the effect of PGE$_2$ in 68% of the PGE$_2$-regulated genes (131 genes), whereas an EP4 agonist did so only in 12% of the PGE$_2$-regulated genes (24 genes). Most of the EP2-regulated genes (122 genes, 93%) were not altered by an EP4 agonist, and 84% of such "EP2 only" genes (102 genes) showed PGE$_2$-like responses also upon dbcAMP treatment. This group of genes, with their expression levels commonly altered by PGE$_2$, an EP2 agonist and dbcAMP but not by an EP4 agonist, includes

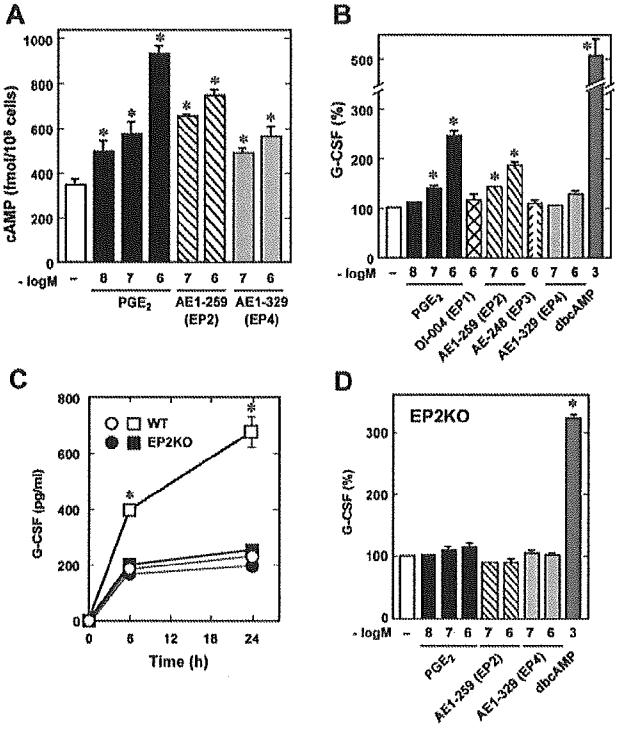

FIGURE 2. PGE$_2$ stimulates LPS-induced G-CSF release from PMNs via the EP2. *A*, Effects of PGE$_2$ and EP agonists on cAMP formation. PMNs were stimulated with the indicated concentrations of PGE$_2$, AE1-259 (EP2 agonist), or AE1-329 (EP4 agonist) for 10 min at 37°C. *, $p < 0.05$ vs vehicle. *B*, Effects of PGE$_2$, EP agonists, and dbcAMP on LPS-induced G-CSF release. PMNs were incubated in medium containing 100 ng/ml LPS with or without PGE$_2$, DI-004 (EP1 agonist), AE1-259 (EP2 agonist), AE-248 (EP3 agonist), AE1-329 (EP4 agonist), and dbcAMP for 6 h. G-CSF levels represent percentages of the vehicle. *, $p < 0.05$ vs vehicle. *C*, Time course of LPS-stimulated G-CSF release from PMNs lacking the EP2. PMNs isolated from C57BL/6 (open symbols, WT) or *EP2*-deficient mice (closed symbols, EP2KO) were incubated in medium containing 100 ng/ml LPS with (squares) or without (circles) 1 μM PGE$_2$ for the indicated time periods. *, $p < 0.05$ vs the corresponding value for vehicle-treated cells. *D*, PGE$_2$ but not dbcAMP fails to stimulate LPS-induced G-CSF release from PMNs of *EP2*-deficient mice. PMNs were incubated in medium containing 100 ng/ml LPS with or without dbcAMP or each EP agonist for 6 h. *, $p < 0.05$ vs vehicle.

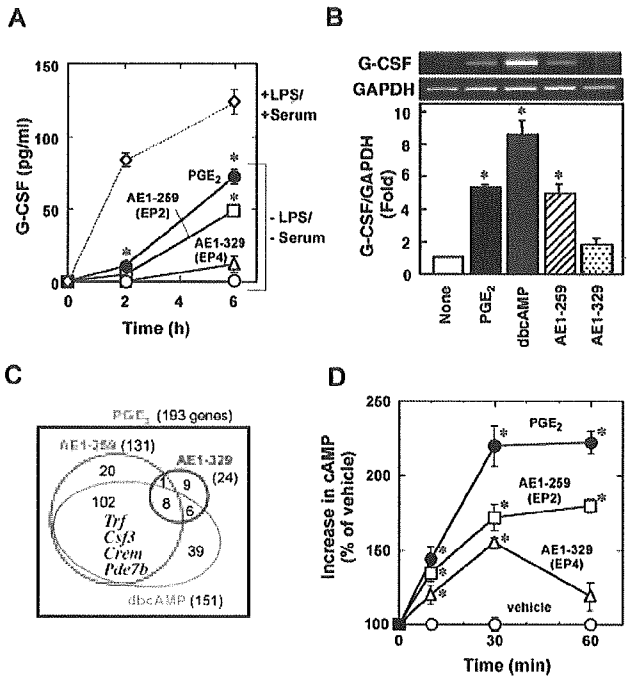

FIGURE 3. G-CSF release is induced by EP2-elicited persistent cAMP accumulation. *A*, PGE_2 or an EP2 agonist alone stimulates G-CSF production in PMNs. PMNs (1×10^6 cells/ml) were incubated in medium supplemented with 10% FBS and 100 ng/ml LPS (◇) or serum-free medium supplemented with vehicle (○), 1 μM PGE_2 (●), 1 μM AE1-259 (□), or 1 μM AE1-329 (△) for the indicated times. *, $p < 0.05$ vs corresponding value for vehicle. *B*, Effect of EP agonists and dbcAMP on G-CSF (*Csf3*) gene expression. PMNs were incubated at 37°C for 6 h in serum-free medium supplemented with vehicle (none), PGE_2, AE1-259, AE1-329 (1 μM each), or dbcAMP (1 mM). Total RNA isolated from each sample was subjected to RT-PCR analysis. The G-CSF/GAPDH values are represented as the fold of the value for the vehicle. *, $p < 0.05$ vs vehicle. *C*, An EP2 agonist and dbcAMP but not an EP4 agonist showed PGE_2-like changes in gene expression of PMNs. Each total RNA used in *B* was subjected to GeneChip analysis. Genes with altered expression values upon PGE_2 treatment (log ratio of PGE_2 vs vehicle >1.3 or less than −1.3) were selected and subdivided into groups corresponding to pairwise overlaps shown in the Venn diagrams. *D*, Time course of cAMP accumulation by EP2 and EP4. PMNs were stimulated with vehicle (○), PGE_2 (●), AE1-259 (□), or AE1-329 (△) (1 μM each) for the indicated times at 37°C. Values represent the percentage increase in cAMP over vehicle-treated cells. *, $p < 0.05$ vs vehicle.

G-CSF (*Csf3*) and other genes known to be up-regulated by cAMP signals such as transferrin (*Trf*) and phosphodiesterase 7 (*Pde7b*) (Table I) (24, 25). Thus, expression changes upon treatment with PGE_2, an EP2 agonist, and dbcAMP are closely associated, and the molecular events induced by PGE_2 are likely to be caused by the activation of cAMP signaling. To further explore the possible difference in cAMP signaling between EP2 and EP4, we examined the time course of EP agonist-induced cAMP increase (Fig. 3*D*). PGE_2 increased cAMP levels over the basal until 30 min and sustained this level until 60 min. An EP2 agonist mimicked this profile even though its maximal activity was lower than PGE_2. On the other hand, an EP4 agonist increased cAMP levels in a manner similar to an EP2 agonist until 30 min but failed to maintain this level until 60 min. Such chronic activation in cAMP signaling may be required for G-CSF gene expression. Based on these results, we conclude that PGE_2 stimulation of G-CSF production is mediated by the persistent activation of cAMP signaling.

Inhibition of PGE_2 synthesis and EP2 gene deficiency lead to decreased G-CSF levels in peritoneal exudates in casein-induced peritonitis

To explore the possibility that PGE_2 synthesized locally within an inflammation site enhances G-CSF production in vivo, we used the casein-induced peritonitis model, in which neutrophils accumulate within the peritoneal cavity, and examined the relationship between the number of infiltrated neutrophils and G-CSF and PGE_2 content in the peritoneal fluid (Fig. 4). In this model, casein injection into the peritoneal cavity was followed by immediate PMN infiltration preceding macrophage accumulation; infiltrated PMNs appeared at 2 h postinjection and by 6 h were increased by >10-fold. PMN number was maintained at high levels until 16 h postinjection. Macrophages were constant in number until 6 h postinjection and then transiently increased up to 5-fold at 16 h (data not shown). The PMN ratio in total peritoneal cells was 0% before the injection and increased after the injection to ~50 and 90% for 2 and 6 h, respectively. In this peritonitis model, the casein injection initially activates peritoneal macrophages; activated macrophages are then thought to produce PGE_2. Indeed, when we measured the PGE_2 contents in the peritoneal exudates, rapid PGE_2 production was observed; a significant amount of PGE_2 could already be detected at 2 h postinjection and increased by 2-fold reaching a peak level at 6 h and was decreased by 16 h. On the other hand, G-CSF was undetectable at 2 h postinjection, but a large amount was detected at 6 h and decreased by 16 h. Thus, PGE_2 production precedes the appearance of G-CSF in the peritoneal cavity.

To assess whether PGE_2 triggers G-CSF production, we tested the effects of indomethacin pretreatment on the appearance of G-CSF release, as well as PMN infiltration. Subcutaneous administration of indomethacin 2 h before casein injection decreased PGE_2 levels both at 2 and 6 h after the injection (Fig. 5). Under such conditions, the increase in PMN infiltration was markedly diminished at 6 h postinjection while the number of macrophages was not affected significantly. In parallel with the decrease in the number of infiltrated PMNs, the levels of G-CSF in the peritoneal exudates also decreased. No significant changes were observed in the levels of MIP-2, one of the critical chemokines that attracts

Table I. *Representative genes whose changes in expression levels were up-regulated by PGE_2, an EP2 agonist and dbcAMP, but not by an EP4 agonist*[a]

Gene Symbol	Gene Name	Log_2 (Fold) (vs Vehicle)				Unigene ID
		PGE_2	AE1–259	AE1–329	dbcAMP	
Trf	Transferrin	6.4	3.8	NC	8.5	37214
Csf3	G-CSF	3.8	2.1	NC	3.4	5244
Crem	cAMP-responsive element modulator	3.6	2.2	NC	5.9	1238
Pde7b	Phosphodiesterase 7B	2.7	2.0	NC	2.6	100580

[a] The effect of PGE_2, AE1–259 (EP2 agonist), AE1–329 (EP4 agonist), and dbcAMP is indicated as a logarithm of the fold change vs the expression level of the vehicle. NC, not changed.

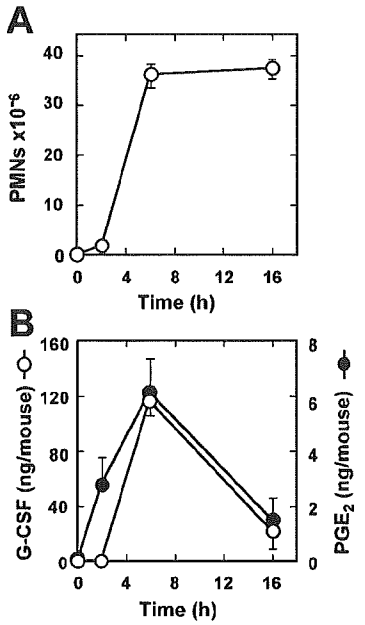

FIGURE 4. Association of PGE$_2$ production and G-CSF content in casein-induced peritonitis. *A*, PMN accumulation in casein-induced peritonitis. Mice were injected i.p. with 2 ml of a 5% (w/v) solution of casein, and peritoneal cells were collected at the indicated times after the injection. The numbers of PMNs were determined by staining with May-Grunwald-Giemsa. *B*, PGE$_2$ production precedes the appearance of peritoneal G-CSF. The contents of PGE$_2$ and G-CSF in the supernatants of peritoneal lavage fluids were determined by EIA and ELISA, respectively.

neutrophils to the site of infection. Thus, the inhibition of endogenous PG synthesis was followed by a decrease in PMN infiltration, resulting in a reduction of G-CSF levels.

To evaluate the positive correlation between EP2 signaling and local G-CSF production, we examined the effects of EP2 gene deficiency on G-CSF contents in peritoneal exudates. In EP2-deficient mice, the casein injection was followed by an increase in PGE$_2$ levels as seen in wild-type animals (Fig. 6). Although the injection-primed PMN infiltration and the numbers of peritoneal macrophages in EP2-deficient mice were indistinguishable from those of wild-type mice, the G-CSF content at 6 h postinjection was diminished to 30% in EP2-deficient mice compared with wild-type mice. Thus, EP2 deficiency results in low levels of G-CSF in the peritoneal exudates. Because the infiltrated PMNs would be exposed to PGE$_2$, a decrease in local G-CSF production may at least be partly due to the lack of EP2 in PMNs.

Discussion

One of the remarkable findings in the current study is that PGE$_2$ appears to induce G-CSF release from peritoneal PMNs. PGE$_2$ has been shown to positively or negatively regulate the production of cytokines and chemokines such as TNF-α, IL-1β, IL-8, IL-12, MCP-1, and MIP-1 from monocytes and neutrophils (2, 3, 23–29). PGE$_2$ in most cases promotes or inhibits the production of cytokines and chemokines induced by stimuli such as LPS or TNF-α and thus has been considered to be a modulator of immune responses (30). Indeed, it has been shown that endogenous PGE$_2$ contributes to LPS-induced *G-CSF* gene expression in monocytes (12, 13). Hareng et al. (13) recently demonstrated that dbcAMP enhances LPS-primed promoter activity via a cAMP-responsive element located at −240 bp of the *G-CSF* gene in THP-1 cells. We also found that indomethacin inhibits LPS-induced G-CSF production and that PGE$_2$ reverses this inhibition in the murine resident peritoneal macrophages (D. Mori, S. Tsuchiya, and Y. Sugimoto, unpublished observation). The enhancing effects of PGE$_2$ on G-CSF production in monocytes/macrophages require an input of the LPS signal; PGE$_2$ itself does not induce basal G-CSF production. In contrast, in peritoneal PMNs, PGE$_2$, or dbcAMP alone appears to induce G-CSF production as shown in the current study. However, although G-CSF levels were undetectable in the serum-free culture medium, we detected a faint but significant amount of mRNA expression for the *G-CSF* gene in the PMNs in the absence of PGE$_2$ (Fig. 3, *A* and *B*). Because the peritoneal PMNs have already received activation signals from multiple inputs such as adhesion molecules and chemoattractants during migration, the infiltrated cells may still have active signals in particular intracellular signaling cascades such as the MAPK- or ρ-cascades (31–33). We tested whether PGE$_2$ alone can really induce G-CSF production in bone marrow-derived PMNs (20), but marrow neutrophils failed to stimulate production and gene expression of G-CSF upon PGE$_2$ or

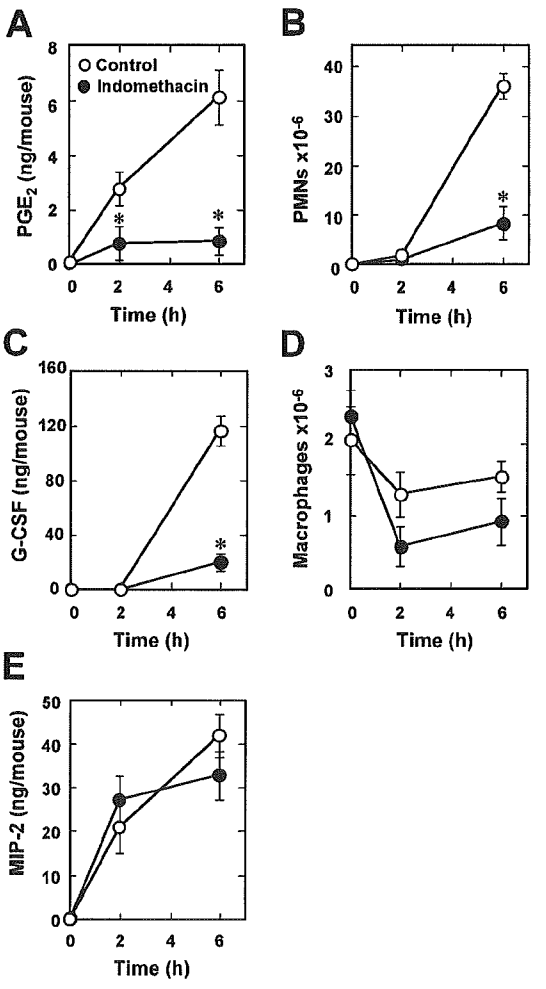

FIGURE 5. Effect of indomethacin on the numbers of PMNs and macrophages and the levels of PGE$_2$, G-CSF, and MIP-2 in the peritoneal cavity. Mice were s.c. administered with (●) or without (○) indomethacin (10 mg/kg) 1 h before casein injection. Peritoneal cells were collected at the indicated times after the casein injection. The numbers of PMNs (*B*) and macrophages (*D*) were determined by staining with May-Grunwald-Giemsa. The levels of PGE$_2$ in the lavage fluid (*A*) was measured by enzyme immunoassay, and G-CSF (*C*) and MIP-2 (*E*) were determined by ELISA. *, $p < 0.05$ vs control.

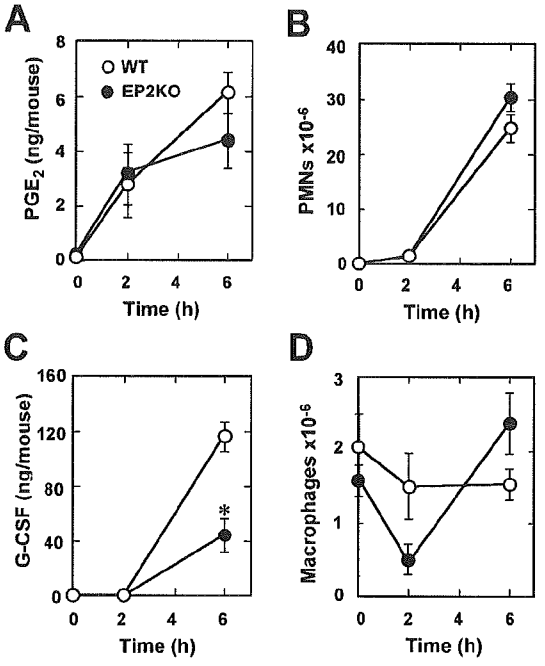

FIGURE 6. Effect of *EP2* gene deficiency on G-CSF levels in peritoneal exudates. Wild-type C57BL/6 (○) and *EP2*-deficient mice (●) were treated with casein by i.p. injection. Peritoneal cells were collected, the numbers of PMNs (*B*) and macrophages (*D*) were determined by staining with May-Grunwald-Giemsa, and the contents of PGE_2 (*A*) and G-CSF (*C*) in the lavage fluid were determined by EIA and ELISA, respectively. *, $p < 0.05$ vs wild type.

dbcAMP treatment (data not shown). The inability of marrow neutrophils to respond to PGE_2 suggested that a preactivated signal associated with migration is necessary for the basal expression of the *G-CSF* gene. Thus, PGE_2 is likely to augment such basal promoter activity of the *G-CSF* gene. However, this PGE_2 potentiation is considered to play a key role in local G-CSF production in the peritoneal cavity as discussed below.

The peritoneal PMNs used in the current study express functional prostanoid receptors EP2 and EP4, but only EP2 can sustain intracellular cAMP increase for >30 min (Fig. 3*D*). We previously found in the peritoneal macrophages that both gene expression and cAMP-producing activity of EP4 is down-regulated by LPS treatment (5). In the peritoneal PMNs, *EP4* gene expression was constant during PGE_2 and/or LPS treatment (data not shown), indicating that the inability of EP4 to increase cAMP for a longer period is not due to down-regulation of *EP4* gene expression. In our initial analysis, the two Gs-coupled receptors EP2 and EP4 were different in their sensitivity to agonist-induced desensitization when expressed in CHO cells (34, 35); EP4 undergoes agonist-induced desensitization, but no desensitization occurs in EP2. Such agonist-induced desensitization of the EP4 may cause its inability to sustain increases in cAMP levels for longer periods. When the PMNs were preincubated with an EP2 agonist for 30 min followed by incubation in the absence of this agonist for 6 h, the cells were no longer able to produce a significant amount of G-CSF (data not shown). Therefore, persistent increase in cAMP levels is likely to be required for G-CSF production in the peritoneal PMNs, although its precise mechanism is unknown. Moreover, PGE_2-elicted changes in gene expression profiles of PMNs were mimicked by an EP2 agonist and dbcAMP but not by an EP4 agonist (Fig. 3*C*). Because membrane permeable dbcAMP is considered to constantly activate cAMP-dependent signaling, these results support the notion that PGE_2 alters gene expression profiles through chronic activation of cAMP signaling by the EP2. It is interesting in this respect that genes functionally antagonizing cAMP signaling such as cAMP-specific phosphodiesterase *Pde7b* and cAMP-responsive element modulator *Crem* were up-regulated by an EP2 agonist and dbcAMP (Table I).

In the in vivo studies, we found a positive association between PGE_2 levels and G-CSF contents in peritoneal exudates in the casein-induced peritonitis model; both G-CSF and PGE_2 contents increased at 6 h and decreased at 16 h postinjection (Fig. 4), and attenuation of PGE_2 production by indomethacin decreased G-CSF levels at 6 h postinjection (Fig. 5). The absence of an effect of indomethacin in the in vitro system suggested that PMNs themselves produce only a small amount of PGE_2. Therefore, resident macrophages are likely to be a source of PGE_2 critical for G-CSF production in the peritoneal cavity. In a previous report, Metcalf et al. found in the casein-induced peritonitis that the appearance of G-CSF in the peritoneal cavity correlates with the appearance of PMNs and that the concentration of G-CSF in the peritoneal fluid is higher than in the serum (11). Based on these findings, they concluded that G-CSF locally produced in the peritoneal cavity should be responsible for the increased G-CSF levels in the circulation. Then, which cell type is responsible for the G-CSF production, resident macrophages, or the newly elicited neutrophils? We found that in the murine resident peritoneal macrophages, PGE_2 augmented LPS-induced G-CSF production as Hareng et al. (13) reported for THP-1 cells. However, in the case of the macrophages, EP4 is responsible for the PGE_2 potentiation of G-CSF release because an EP4 agonist mimicked the effect of PGE_2 more effectively than an EP2 agonist (D. Mori, S. Tsuchiya, and Y. Sugimoto, unpublished observation). Therefore, EP2 is unlikely to be involved in G-CSF release from macrophages, and a striking reduction in G-CSF levels in the *EP2*-deficient mice can be ascribed to loss of G-CSF production from the PMNs (Fig. 6). Thus, we conclude that EP2 signaling is essential for local G-CSF production and also that the PMNs are the physiologically critical source of G-CSF production at least in this peritonitis model.

Casein-induced PMN infiltration was markedly attenuated by the inhibition of endogenous PG synthesis but not by *EP2* gene disruption. Such a difference in cell profiling suggests that some prostanoid receptor other than EP2 participates in this process. We previously demonstrated that PGI_2 is critical for an inflammation-associated increase in vascular permeability; PGI_2R-deficient mice showed a striking reduction in exudate volume in carrageenan-induced pleurisy (36), suggesting that PGI_2 is the major prostanoid affecting vascular permeability in this model. Alternatively, other prostanoid receptor systems may be suppressing some chemotactic factors essential for neutrophil migration such as MIP-2 (37). Indeed, some prostanoids, including PGE_2, have been suspected to suppress MIP-2 production (38). However, we failed to detect a significant decrease in peritoneal MIP-2 levels in the indomethacin-treated mice (Fig. 5). Recently, mice null for the IL-17R were reported to be impaired in neutrophil recruitment and in the local production of G-CSF in a *Klebsiella pneumoniae* lung infection model (39). Involvement of the endogenous PG system in IL-17 signaling is a likely possibility that should be examined.

In conclusion, we found that PGE_2-EP2 signaling plays a key role in inflammation-related G-CSF production. Recombinant G-CSF has been clinically used as a therapeutic drug for neutropenia (9). Interestingly, it has been reported that the occurrence of neutropenia is associated with the use of aspirin-like drugs (40) and that chronic aspirin treatment in vivo but not in vitro results in

functionally immature blood neutrophils (41). Whether PGE_2 contributes to physiological G-CSF production in marrow cells or granulocytes is an interesting issue to be addressed.

Acknowledgments

We thank Dr. Sachiko Oh-ishi (Kitasato Institute, Tokyo, Japan) for her invaluable discussions and helpful advice on this work. We are grateful to Helena A. Popiel and Sachiko Terai-Yamaguchi for careful reading and secretarial assistance.

Disclosures

The authors have no financial conflict of interest.

References

1. Funk, C. D. 2001. Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* 294: 1871–1875.
2. Harris, S. G., J. Padilla, L. Koumas, D. Ray, and R. P. Phipps. 2002. Prostaglandins as modulators of immunity. *Trends Immunol.* 23: 144–150.
3. Shinomiya, S, H. Naraba, A. Ueno, I. Utsunomiya, T. Maruyama, S. Ohuchida, F. Ushikubi, K. Yuki, S. Narumiya, Y. Sugimoto, A. Ichikawa, and S. Oh-ishi. 2001. Regulation of TNF-α and interleukin-10 production by prostaglandins I2 and E2: studies with prostaglandin receptor-deficient mice and prostaglandin E receptor subtype-selective synthetic agonists. *Biochem. Pharmacol.* 61: 1153–1160.
4. van der Pouw Kraan, T. C., L. C. Boeije, R. J. Smeenk, J. Wijdenes, and L. A. Aarden. 1995. Prostaglandin-E2 is a potent inhibitor of human interleukin 12 production. *J. Exp. Med.* 181: 775–779.
5. Ikegami, R., Y. Sugimoto, E. Segi, M. Katsuyama, H. Karahashi, F. Amano, T. Maruyama, H. Yamane, S. Tsuchiya, and A. Ichikawa. 2001. The expression of prostaglandin E receptors EP2 and EP4 and their different regulation by lipopolysaccharide in C3H/HeN peritoneal macrophages. *J. Immunol.* 166: 4689–4696.
6. Demetri, G. D., and J. D. Griffin. 1991. Granulocyte colony-stimulating factor and its receptor. *Blood* 78: 2791–2808.
7. Boneberg, E. M., and T. Hartung. 2002. Molecular aspects of anti-inflammatory action of G-CSF. *Inflamm. Res.* 51: 119–128.
8. Welte, K., J. Gabrilove, M. H. Bronchud, E. Platzer, and G. Morstyn. 1996. Filgrastim (r-metHuG-CSF): the first 10 years. *Blood* 88: 1907–1929.
9. Hubel, K., and A. Engert. 2003. Clinical applications of granulocyte colony-stimulating factor: an update and summary. *Ann. Hematol.* 82: 207–213.
10. Clark, S. C., and R. Kamen. 1987. The human hematopoietic colony-stimulate factors. *Science* 236: 1229–1237.
11. Metcalf, D., L. Robb, A. R. Dunn, S. Mifsud, and L. Di Rago. 1996. Role of granulocyte-macrophage colony-stimulating factor and granulocyte colony-stimulating factor in the development of an acute neutrophil inflammatory response in mice. *Blood* 88: 3755–3764.
12. Lee, M. T., K. Kaushansky, P. Ralph, and M. B. Ladner. 1990. Differential expression of M-CSF, G-CSF, and GM-CSF by human monocytes. *J. Leukocyte Biol.* 47: 275–282.
13. Hareng, L., T. Meergans, S. von Aulock, H. D. Volk, and T. Hartung. 2003. Cyclic AMP increases endogenous granulocyte colony-stimulating factor formation in monocytes and THP-1 macrophages despite attenuated TNF-α formation. *Eur. J. Immunol.* 33: 2287–2296.
14. Coleman, R. A., W. L. Smith, and S. Narumiya. 1994. International Union of Pharmacology classification of prostanoid receptors: properties, distribution, and structure of the receptors and their subtypes. *Pharmacol. Rev.* 46: 205–229.
15. Sugimoto, Y., S. Narumiya, and A. Ichikawa. 2000. Distribution and function of prostanoid receptors: studies from knockout mice. *Prog. Lipid Res.* 39: 289–314.
16. Fedyk, E. R., J. M. Ripper, D. M. Brown, and R. P. Phipps. 1996. A molecular analysis of PGE receptor (EP) expression on normal and transformed B lymphocytes: coexpression of EP1, EP2, EP3β and EP4. *Mol. Immunol.* 33: 33–45.
17. Kabashima, K., D. Sakata, M. Nagamachi, Y. Miyachi, K. Inaba, and S. Narumiya. 2003. Prostaglandin E2-EP4 signaling initiates skin immune responses by promoting migration and maturation of Langerhans cells. *Nat. Med.* 9: 744–749.
18. Hizaki, H., E. Segi, Y. Sugimoto, M. Hirose, T. Saji, F. Ushikubi, T. Matsuoka, Y. Noda, T. Tanaka, N. Yoshida, S. Narumiya, and A. Ichikawa. 1999. Abortive expansion of the cumulus and impaired fertility in mice lacking the prostaglandin E receptor subtype EP_2. *Proc. Natl. Acad. Sci. USA* 96: 10501–10506.
19. Suzawa, T., C. Miyaura, M. Inada, T. Maruyama, Y. Sugimoto, F. Ushikubi, A. Ichikawa, S. Narumiya, and T. Suda. 2000. The role of prostaglandin E receptor subtypes (EP1, EP2, EP3, and EP4) in bone resorption: an analysis using specific agonists for the respective EPs. *Endocrinology* 141: 1554–1559.
20. Tanaka, S., K. Deai, A. Konomi, K. Takahashi, H. Yamane, Y. Sugimoto, and A. Ichikawa. 2004. Expression of L-histidine decarboxylase in granules of elicited mouse polymorphonuclear leukocytes. *Eur. J. Immunol.* 34: 1472–1482.
21. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162: 156–159.
22. Sugimoto, Y., H. Tsuboi, Y. Okuno, S. Tamba, S. Tsuchiya, G. Tsujimoto, and A. Ichikawa. 2004. Microarray evaluation of EP4 receptor mediated prostaglandin E2 suppression of 3T3-L1 adipocyte differentiation. *Biochem. Biophys. Res. Commun.* 322: 911–917.
23. Yamane, H., Y. Sugimoto, S. Tanaka, and A. Ichikawa. 2000. Prostaglandin E_2 receptors, EP2 and EP4, differentially modulate TNF-α and IL-6 production induced by lipopolysaccharide in mouse peritoneal neutrophils. *Biochem. Biophys. Res. Commun.* 278: 224–228.
24. Suire, S., I. Fontaine, and F. Guillou. 1995. Follicle stimulating hormone (FSH) stimulates transferrin gene transcription in rat Sertoli cells: *cis*- and *trans*-acting elements involved in FSH action via cyclic adenosine 3′,5′-monophosphate on the transferrin gene. *Mol. Endocrinol.* 9: 756–766.
25. Lee, R., S. Wolda, E. Moon, J. Esselstyn, C. Hertel, and A. Lerner. 2002. PDE7A is expressed in human B-lymphocytes and is up-regulated by elevation of intracellular cAMP. *Cell. Signal.* 14: 277–284.
26. Standiford, T. J., S. L. Kunkel, M. W. Rolfe, H. L. Evanoff, R. M. Allen, and R. M. Strieter. 1992. Regulation of human alveolar macrophage- and blood monocyte-derived interleukin-8 by prostaglandin E2 and dexamethasone. *Am. J. Respir. Cell Mol. Biol.* 6: 75–81.
27. Zhong, W. W., P. A. Burke, M. E. Drotar, S. R. Chavali, and R. A. Forse. 1995. Effects of prostaglandin E2, cholera toxin and 8-bromo-cyclic AMP on lipopolysaccharide-induced gene expression of cytokines in human macrophages. *Immunology* 84: 446–452.
28. Nataraj, C., D. W. Thomas, S. L. Tilley, M. T. Nguyen, R. Mannon, B. H. Koller, and T. M. Coffman. 2001. Receptors for prostaglandin E_2 that regulate cellular immune responses in the mouse. *J. Clin. Invest.* 108: 1229–1235.
29. Takayama, K., G. Garcia-Cardena, G. K. Sukhova, J. Comander, M. A. Gimbrone, Jr., and P. Libby. 2002. Prostaglandin E_2 suppresses chemokine production in human macrophages through the EP4 receptor. *J. Biol. Chem.* 277: 44147–44154.
30. Coleman, R. A., I. Kennedy, P. P. A. Humphrey, K. Bunce, and P. Lumley. 1990. Prostanoids and their receptors. In *Comprehensive Medicinal Chemistry*, Vol. 3. C. Hansch, P. G. Sammes, J. B. Taylor, and J. C. Emmett, eds. Pergamon, Oxford, p. 643–714.
31. Smolen, J. E., T. K. Petersen, C. Koch, S. J. O'Keefe, W. A. Hanlon, S. Seo, D. Pearson, M. C. Fossett, and S. I. Simon. 2000. L-selectin signaling of neutrophil adhesion and degranulation involves p38 mitogen-activated protein kinase. *J. Biol. Chem.* 275: 15876–15884.
32. Dib, K., F. Melander, and T. Andersson. 2001. Role of p190RhoGAP in β_2 integrin regulation of RhoA in human neutrophils. *J. Immunol.* 166: 6311–6322.
33. Ley, K. 2002. Integration of inflammatory signals by rolling neutrophils. *Immunol. Rev.* 186: 8–18.
34. Nishigaki, N., M. Negishi, and A. Ichikawa. 1996. Two Gs-coupled prostaglandin E receptor subtypes, EP2 and EP4, differ in desensitization and sensitivity to the metabolic inactivation of the agonist. *Mol. Pharmacol.* 50: 1031–1037.
35. Fujino, H., K. A. West, and J. W. Regan. 2002. Phosphorylation of glycogen synthase kinase-3 and stimulation of T cell factor signaling following activation of EP2 and EP4 prostanoid receptors by prostaglandin E_2. *J. Biol. Chem.* 277: 2614–2619.
36. Murata, T., F. Ushikubi, T. Matsuoka, M. Hirata, A. Yamasaki, Y. Sugimoto, A. Ichikawa, Y. Aze, T. Tanaka, N. Yoshida, et al. 1997. Altered pain perception and inflammatory response in mice lacking prostacyclin receptor. *Nature* 388: 678–682.
37. Greenberger, M. J., R. M. Strieter, S. L. Kunkel, J. M. Danforth, L. L. Laichalk, D. C. McGillicuddy, and T. J. Standiford. 1996. Neutralization of macrophage inflammatory protein-2 attenuates neutrophil recruitment and bacterial clearance in murine *Klebsiella pneumonia*. *J. Infect. Dis.* 173: 159–165.
38. Kyrkanides, S., A. H. Moore, J. A. Olschowka, J. C. Daeschner, J. P. Williams, J. T. Hansen, and M. Kerry O'Banion. 2002. Cyclooxygenase-2 modulates brain inflammation-related gene expression in central nervous system radiation injury. *Brain Res. Mol. Brain Res.* 104: 159–169
39. Ye, P., F. H. Rodriguez, S. Kanaly, K. L. Stocking, J. Schurr, P. Schwarzenberger, P. Oliver, W. Huang, P. Zhang, J. Zhang, et al. 2001. Requirement of interleukin 17 receptor signaling for lung CXC chemokine and granulocyte colony-stimulating factor expression, neutrophil recruitment, and host defense. *J. Exp. Med.* 194: 519–527.
40. Strom, B. L., J. L. Carson, R. Schinnar, E. S. Snyder, M. Shaw, and F. E. Lundin Jr. 1993. Nonsteroidal anti-inflammatory drugs and neutropenia. *Arch. Intern. Med.* 153: 2119–2124.
41. Barton, A. E., D. L. Bayley, M. Mikami, C. G. Llewellyn-Jones, and R. A. Stockley. 2000. Phenotypic changes in neutrophils related to anti-inflammatory therapy. *Biochim. Biophys. Acta* 1500: 108–118.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Food and Chemical Toxicology 43 (2005) 529–535
Food and Chemical Toxicology
www.elsevier.com/locate/foodchemtox

Comprehensive analysis of the effect of phytoestrogen, daidzein, on a testicular cell line, using mRNA and protein expression profile

Tetsuya Adachi [a,*], Yasushi Okuno [a], Shigeo Takenaka [b], Kazuyuki Matsuda [c], Naoki Ohta [b], Kyoka Takashima [d], Koji Yamazaki [d], Daisuke Nishimura [d], Kazutaka Miyatake [c], Chisato Mori [d], Gozoh Tsujimoto [a]

[a] *Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, Yoshida-Shimoadachicho, Sakyo-ku, Kyoto 606-8501, Japan*
[b] *Department of Veterinary Sciences, Osaka Prefecture University, Osaka 591-8501, Japan*
[c] *Department of Applied Biological Chemistry, Osaka Prefecture University, Osaka 591-8501, Japan*
[d] *Department of Bioenvironmental Medicine, Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan*

Received 18 October 2004; accepted 12 December 2004

Abstract

In this study, we examined the effects of exposure to phytoestrogen (daidzein), 17β-estradiol (E_2), diethylstilbestrol (DES) and staurosporin on the TM4 testicular cell line, using comprehensive analysis, such as cDNA microarray and two-dimension polyacrylamide gel electropholesis (2D-PAGE) analysis, and we demonstrated if these toxicogenomic analyses could classify the chemical compounds. First, RNA was extracted from TM4 cells that had been treated with daidzein (80 μM), DES, E_2 (40 μM) and staurosporin (100 nM) for 30 min. We performed cDNA microarray analysis, and the expression ratio data thus obtained were then analyzed using hierarchical clustering. This hierarchical clustering showed that daidzein exposure induced a different effect on gene expression change from that of E_2, DES and staurosporin. Next, protein extracted from TM4 cells also underwent cDNA microarray analysis for 3 h. We performed 2D-PAGE analysis, and the spot intensity ratio data thus obtained were analyzed using hierarchical clustering. As with cDNA microarray, the hierarchical clustering of protein spot ratios showed that daidzein exposure induced a different effect on gene expression change from that of the other substances. In conclusion, we have demonstrated for the first time that classification of these chemicals can be performed by clustering analysis, using data from cDNA microarray and 2D-PAGE analyses, and that exposure to daidzein induces effects different from those of E_2, DES and staurosporin.

Keywords: Phytoestrogen; cDNA microarray; Two-dimension polyacrylamide gel electrophoresis; Testicular cell line; TM4

1. Introduction

Phytoestrogens, such as daidzein, genistein and coumestrol, are defined as estrogenic compounds. They are found in plants (Kurzer and Xu, 1997), and are known to bind to estrogen receptors α (ERα) and β (ERβ) (Kuiper et al., 1998). Phytoestrogens reportedly have beneficial effects on cancer (including reproductive cancer), cardiovascular disease, brain function, alcohol abuse, osteoporosis and menopausal symptoms (Hirayama, 1984; Tajima and Tominaga, 1985; Koo, 1988; Seveson, 1989; Lee et al., 1991; Goodman et al., 1997). However, it has been reported that neonatal exposure to genistein in animal experiments also causes uterine

Abbreviations: 2D-PAGE, two-dimension polyacrylamide gel electrophoresis; DES, diethylstilbestrol; DMSO, dimethyl sulfoxide; E_2, 17β-estradiol; ER, estrogen receptor; MS, mass spectrometry

* Corresponding author. Tel./fax: +81 75 753 6688.
E-mail address: adachi@tom.life.h.kyoto-u.ac.jp (T. Adachi).

0278-6915/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.fct.2004.12.006

adenocarcinomas (Newbold et al., 2001). Moreover, as reported in our previous study, mRNA expression changes were detected in neonatal genistein-treated mouse testes using a cDNA microarray (Adachi et al., 2004). These compounds also have adverse effects on reproductive systems (Kurzer and Xu, 1997; Bingham et al., 1998).

As discussed above, it is not been clear by what mechanism phytoestrogen exerts these effects on reproductive organs. Therefore, it is necessary to develop methods of estimating the effects of phytoestrogen on reproductive organs. The most applicable and efficient approaches to developing such methods are comprehensive analysis using DNA microarray and two-dimension polyacrylamide gel electrophoresis (2D-PAGE). DNA microarray is known to be a powerful and high-throughput method for monitoring the expression of thousands of genes simultaneously (Duggan et al., 1999; Rockett and Dix, 1999; Hossain et al., 2000; Adachi et al., 2002; Komiyama et al., 2003). Moreover, 2D-PAGE is the only method currently available that is capable of simultaneously separating and quantifying hundreds or thousands of proteins (Garrels, 1989). Many reports of studies using 2D-PAGE have recently been published.

In this study, we examined the effects of TM4 Sertoli cells on exposure to phytoestrogen, daidzein, using DNA microarray and 2D-PAGE. Diethylstilbestrol (DES) and 17β-estradiol (E_2), which are known to be estrogenic compounds, and staurosporin, which is known to be a tyrosine kinase inhibitor, were used as controls. Daidzein is known to be an estrogenic compound, and an inactive form of tyrosine kinase inhibitor (Magee and Rowland, 2004). We also conducted a clustering analysis, using data from the DNA microarray and 2D-PAGE, in order to determine the degree of similarity of the mRNA and protein expression of phytoestrogen to that following treatment with DES, E_2 and staurosporin. Moreover, we demonstrated if these toxicogenomic analyses could classify the chemical compounds.

2. Materials and methods

2.1. Chemicals

Daidzein (minimum 98%), DES (minimum 99%), E_2 (minimum 98%) and staurosporin (minimum 95%) were purchased from Sigma–Aldrich (St Louis, MO). Dimethyl sulfoxide (DMSO), which was used as solvent, was also purchased from Sigma–Aldrich.

2.2. Cell cultures

TM4 cells, a mouse Sertoli cell line, were obtained from American Type Culture Collection (Manassas, VA). Cells were cultured in 1:1 mixture of Ham's F12 medium and Dulbecco's modified Eagle's medium (Invitrogen Corp., Carlsbad, CA) with 1.2 g/l sodium bicarbonate and 15 mM HEPES, 92.5%; horse serum, 5%; fetal bovine serum, 2.5%; penicillin, 100 unit/ml; and streptomycin, 100 μg/ml at 37 °C in a 5% CO_2 atmosphere, with fluid renewal thereafter every 2 days.

2.3. Cell viability

Cells (1×10^4 cells/well) were plated on 96-well microtiter plates and incubated overnight at 37 °C. On the following day, TM4 cells were treated with daidzein (0.4, 4, 40, 80 μM), DES (0.04, 0.4, 4, 40 μM), E_2 (0.04, 0.4, 4, 40 μM) and stauroporin (0.1, 1, 10, 100 nM) dissolved in DMSO for 3, 6, 24 and 48 h. Control cells were treated only with DMSO. The final concentration of DMSO never exceeded 0.1%. The viability of TM4 cells treated with these chemicals was determined by WST-1 assay, which involved obtaining absorbance of a test wavelength of 450 nm in a colorimetric assay using (4-[3-(4-iodophenyl)-2-(4-nitrophenyl)-2H-5-tetrazolio]-1, 3,-benzene disulfonate) (WST-1, Dojindo, Kumamoto, Japan) (Ishiyama et al., 1996).

2.4. RNA extraction and cDNA microarray analysis

Cells (1×10^6 cells/well) were plated on ϕ 6 cm plates and incubated overnight at 37 °C. The next day, TM4 cells were treated for 30 min with daidzein (80 μM), DES (40 μM), E_2 (40 μM) and stauroporin (100 nM) dissolved in dimethyl sulfoxide (DMSO). The dose of chemicals was determined from the change of cell viability by chemical exposure. Moreover, since the decrease of cell counts on chemical exposure was induced by incubation for more than 24 h, and transcriptional changes were induced before the induction of the change of cell viability, we also determined the incubation time. Total RNA was purified from treated cells using SEPAZOL reagent (Nakalai Tesque, Kyoto, Japan). Poly(A)$^+$ RNA was purified using Oligotex-dT30 (TakaraBio, Kyoto, Japan). An cDNA microarray, containing 2942 cDNA probes purchased from Incyte Genomics Inc (Palo Alto, CA), was constructed in-house as previously described (Schena et al., 1995) by Asahi Techno Glass Corporation. Microarray assay was performed by the method described in our previous reports (Adachi et al., 2002, 2004; Yoshikawa et al., 2000), except for the use of single labeling (Cy5) of cDNA in control and test-substance treatments. The array was washed, and then scanned, using a fluorescent laser-scanning device (ScanArray4000, GSI Lumonics, Tokyo, Japan). Numerical microarray data obtained were normalized

using classical linear regression techniques. Changes in genetic expression was changed by microarray, were confirmed using real-time PCR, as previously described (Adachi et al., 2004).

2.5. Protein extraction and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) analysis

Cells (1×10^6 cells/well) were plated on ϕ 6 cm plates and incubated overnight at 37 °C. The next day, TM4 cells were treated for 3 h with daidzein (80 μM), DES (40 μM), E_2 (40 μM) and stauroporin (100 nM) dissolved in DMSO. The dose of chemicals was determined in a similar manner as above. Moreover, since the translational changes were induced before the induction of the change of cell viability and after the transcriptional changes, we also determined the incubation time. Cells were scraped, using 5 ml of phosphate buffer saline; they were then sonicated, and centrifuged to remove debris. Next, we obtained two-dimensional polyacrylamide gel electropholesis (2D-PAGE) maps with an immobilized pH gradient strip. The sample protein (10 μg) was resolved by its charge on a 7 cm non-linear pH 3–10 immobilized gradient strip (Bio-Rad, Tokyo, Japan) in the presence of 7 M urea, 2 M thiourea, 5% SB 3–10 and 5% carrier ampholyte. The identified proteins were then separated according to size, using SDS-PAGE on a 5–20% polyacrylamide gel (Wako Pure Chemicals, Osaka, Japan). Each sample was tested once. In a typical experiment, the samples were processed in parallel. The 25 μl sample was mixed with the 100 μl lysis solution described above, and was adsorbed onto a strip over night at room temperature. The proteins were focused at 200 V for 15 min, then 400 V for 10 min, 1000 V for an hour, 1500 V for 6 h, total V h, at 25 °C in a NR-1410R an iso-electric focusing apparatus (Nihon Eido, Tokyo, Japan). Then an immobilized strip transferred made to migrate onto a ready-made gel for 2D-PAGE at 20 mA/gel with a NA-1013 mini slab gel electrophoresis apparatus (Nihon Eido). The proteins were visualized by staining with Sypro Ruby (Invitrogen Corp., Carlsbad, CA), and detected by means of an FLA3000 (Fuji Film, Tokyo, Japan). The acquired images were analyzed by means of PDQUEST software ver. 7.1 (Bio-Rad). The spots were detected after background correction in the gel images, followed by manual editing of all gels examined. A match-set was created that included all of the gels to be compared. A standard gel (master) was generated from the image with the highest spot number by including additional spots from the other gels.

2.6. Clustering analysis

Clustering analysis was performed in order to examine the relationship between mRNA and protein expression patterns following treatment with daidzein, DES, E_2 and staurosporin. The ratio (chemical treatment/DMSO treatment) obtained by microarray analysis was used for clustering analysis of mRNA expression. The Pearson coefficient algorithm and single linkage method in the computer software 'R' were used to convert the expression ratio obtained by microarray data to a logarithmic expression ($\log_{1.5}X$). The spot intensity ratio (chemical treatment/DMSO treatment) obtained by 2D-PAGE was also converted to logarithmic representation ($\log_{1.5}X$) by means of the Pearson coefficient algorithm and the single linkage method, using computer software 'R'.

3. Results

3.1. Cell viability following the treatment with daidzein, E_2, DES and staurosporin

We first examined the effects of treatment with daidzein, E_2, DES and staurosporin on cell viability. Daidzein treatment for 24 and 48 h at a concentration of 80 μM reduced cell viability; it also reduced cell viability after a 24-h treatment at a concentration of 40 μM (Fig. 1A). Treatment with E_2 and DES for 24 and 48 h at a concentration of 40 μM reduced cell viability, as did E_2 treatment for 6 h (Fig. 1B and C). Staurosporin treatment for 24 and 48 h at concentrations of 10 and 100 μM also reduced cell viability (Fig. 1D).

3.2. cDNA microarray analysis for detection of gene expression changes due to chemical exposure

In order to elucidate changes in mRNA expression in TM4 cells resulting from exposure to daidzein, DES, E_2 and staurosporin, mRNA was extracted from daidzein-treated, DES-treated, E_2-treated, staurosporin-treated and DMSO-treated cells for 30 min. The majority of genes examined showed only small differences in expression between treatment and non-treatment. Their expression ratios (chemical treatment/DMSO treatment) were between 1.5 and 0.67. However, the numbers of the genes with expression change greater than 1.5-fold or less than 0.67-fold following chemical treatment are shown in Table 1. Change of these gene expressions was not detected in low dose chemical exposure, and was not detected in 3, 6, 24, 48-h incubation (data not shown). Table 2 shows genes that shared in common the fact that their expression was changed as a result of exposure to daidzein, DES, E_2 and staurosporin against DMSO treatment. Confirmation of gene expressions was obtained using real-time RT-PCR (data not shown). Expression ratio data were also analyzed using hierarchical clustering, and yielded temporal relationships between the expression profiles of 2942 genes (Fig. 2).

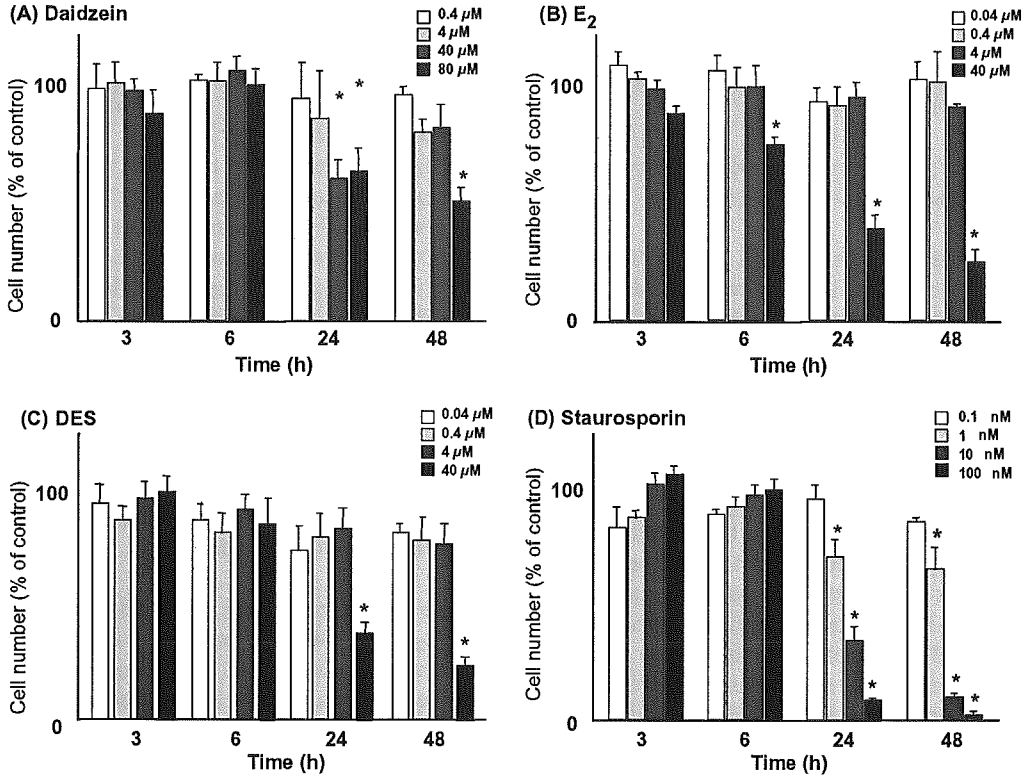

Fig. 1. TM4 Cell viability following treatment for 3, 6, 24 and 48 h with daidzein (0.4, 4, 40, 80 μM) (A), DES (0.04, 0.4, 4, 40 μM) (B), E_2 (0.04, 0.4, 4, 40 μM) (C) and stauroporin (0.1, 1, 10, 100 nM) (D), dissolved in DMSO. Relative cell viability was calculated after viability in DMSO treatment was normalized as 100%. $^{**}P < 0.01$, $^{*}P < 0.05$ vs DMSO treatment.

Table 1
Number of the genes whose expression was changed by exposure to chemicals, as determined using DNA microarray

Chemical	Number of genes changed by chemical exposure[a]	
	1.5>	0.67<
Daidzein	14	36
E_2	11	34
DES	24	37
Staurosporin	25	57

[a] Fold expression in microarray data, based on the ratio of fluorescence in chemical-treated cells to that of vehicle (DMSO)-treated cells.

Table 2
Genes that shared in common the fact that their expression was altered by chemical exposure, identified using cDNA microarray analysis

Accession no.	Gene name[a]
AF012251	Casein kinase II, alpha prime subunit
AF107296	Mitotic checkpoint protein kinase BUB1B
BC034168	Pleckstrin homology domain
BC054445	Death effector domain
U40796	Endonuclease

[a] The top hits among genes on the GenBank database are shown.

Fig. 2. Hierarchical clustering analysis using cDNA microarray data. Expression ratio obtained on the basis of microarray data was converted to logarithmic representation ($\log_{1.5}X$), and the Pearson coefficient algorithm and single linkage method in computer software 'R' were used.

3.3. 2D-PAGE analysis for detection of protein expression resulting from chemical exposure

Next, to elucidate changes in protein expression resulting from chemical exposure, protein was extracted

from daidzein-treated, DES-treated, E_2-treated, staurosporin-treated and DMSO-treated cells for 3 h. 2D-gel images of daidzein and E_2 treatment are shown in Fig. 3A and B, respectively. The number of protein spots that appeared in 2D-gel as a result of test substance or DMSO treatment are shown in Table 3. Table 3 also shows the number of spots whose intensities were changed as a result of treatment with the test substance as compared to DMSO (spot intensity ratios, greater than 1.5-fold and less that 0.67-fold). Spot intensity ratio data were analyzed using hierarchical clustering (Fig. 4).

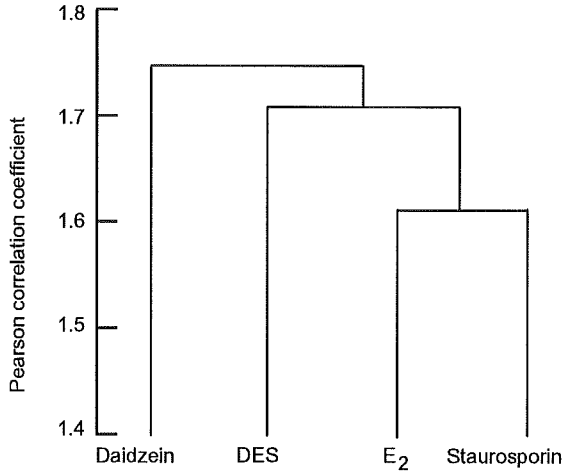

Fig. 4. Hierarchical clustering analysis using 2D-PAGE data. Expression ratio obtained by 2D-PAGE data was converted to logarithmic representation ($\log_{1.5}X$), and the Pearson coefficient algorithm and single linkage method in the computer software 'R' were used.

4. Discussion

It has been reported that phytoestrogens have beneficial effects on cancer (including reproductive cancer), cardiovascular disease, brain function, alcohol abuse, osteoporosis and menopausal symptoms (Hirayama, 1984; Tajima and Tominaga, 1985; Koo, 1988; Seveson, 1989; Lee et al., 1991; Goodman et al., 1997). However, these compounds also have adverse effects on reproductive systems (Kurzer and Xu, 1997; Bingham et al., 1998). Phytoestrogen displays estrogenic and tyrosine kinase inhibition activity. The effects of phytoestrogen on reproductive organs are still not clearly understood. It is therefore necessary to develop methods of

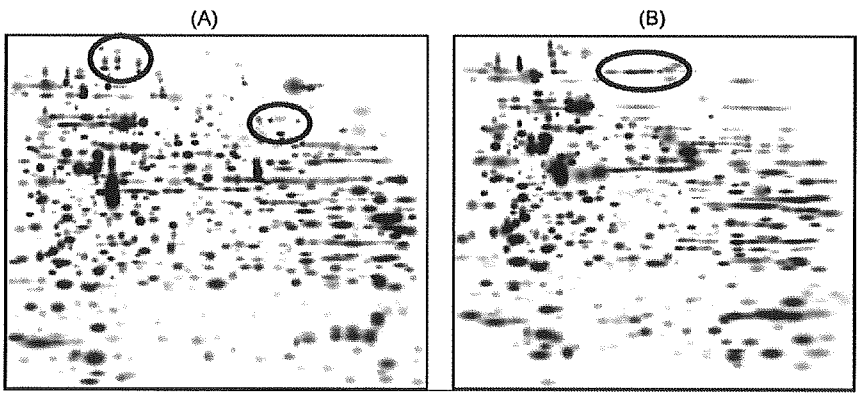

Fig. 3. A representative result of 2D-PAGE. The proteins were visualized by staining with Sypro Ruby (Invitrogen Corp.) and detected by a FLA3000 (Fuji Film). The acquired images were analyzed by means of PDQUEST software ver. 7.1 (Bio-Rad). (A) daidzein treatment, (B) E_2 treatment. Circle area, for example, shows the expression difference between daidzein and E_2 treatment.

Table 3
Number of the spots whose protein expression was changed by exposure to chemicals, as determined using 2D-PAGE

Chemical	Number of total spots[a]	Number of non-changed spots, compared to vehicle[b]	Number of spots changed by chemical exposure[b]	
			1.5> (Visible spot[c])	0.67< (Invisible spot[c])
Daidzein	500	133	256 (121)	278 (167)
E_2	460	134	256 (99)	256 (186)
DES	465	105	284 (108)	266 (190)
Staurosporin	443	79	299 (119)	288 (223)
Vehicle	546			

[a] Number of spots, which were visualized by staining with Sypro Ruby (Invitrogen Corp., Carlsbad, CA).
[b] Fold expression for 2D-PAGE data, based on the ratio of fluorescence for chemical-treated cells to that of vehicle (DMSO)-treated cells.
[c] "Visible spot" and "invisible spot" are spots that, respectively, were or were not detected as a result of chemical exposure, as compared to vehicle (DMSO)-treated cells.

evaluating the effects of phytoestrogen on reproductive organs. In this study, we estimated the effects of phytoestrogen using TM4 Sertoli cells, compared with two estrogenic compounds (DES and E_2) and one kinase inhibition compound (staurosporin).

First, we examined cell viability following exposure to test and control substances. Exposure to daidzein inhibited cell viability in a time- and dose-dependent manner (Fig. 1A). The results showed that 80 μM daidzein inhibits TM4 cell viability. Indeed, it has been reported that in cases of colon cancer, high doses (50, 100 μM) of daidzein induce apoptotic cell death (Guo et al., 2004). On the other hand, exposure to DES, E_2 and staurosporin also inhibited cell viability. However, the daidzein concentration at which cell growth was inhibited was higher than with the other substances (Fig. 1B–D). Our study suggests that exposure to daidzein had less hazardous effects on TM4 cell viability than the other substances, such as estrogenic compounds and tyrosine kinase inhibitors. Next, we used a DNA microarray to analyze gene expression, in order to determine the mRNA expression profile when the viability of TM4 cells was reduced as a result of exposure to daidzein, E_2, DES or staurosporin. Five genes were detected that shared in common the fact that their expression changed as a result of exposure to daidzein, DES, E_2 and staurosporin, as compared to DMSO treatment (Table 2). Since the number of genes whose expression was changed by exposure to each chemical, was very low (5 genes), we surmise that the various substances used brought about different changes in gene expression, which affects cell viability. However, since 5 genes are related with cell signaling, cell proliferation and apoptosis, these genes may trigger the change of protein synthesis, resulting the change of cell viability. In addition, these expression ratio data were analyzed using hierarchical clustering, and yielded temporal relationships between expression profiles (Fig. 2). This hierarchical clustering showed that daidzein exposure affected TM4 cell viability by causing changes in gene expression that differed from that caused by E_2, DES and staurosporin.

Moreover, to elucidate changes in protein expression resulting from chemical exposure, protein was extracted from daidzein-treated, DES-treated, E_2-treated, staurosporin-treated and DMSO-treated cells for 3 h, using 2D-PAGE. Protein spot intensity ratio data were analyzed using hierarchical clustering (Fig. 4). As with cDNA microarray, this hierarchical clustering of protein spot ratio showed that daidzein exposure affected TM4 cell viability as a result of a different gene expression changes than those resulting from treatment with E_2, DES and staurosporin.

Two clustering analyses of mRNA expression change data and protein expression change data revealed that hierarchical clustering of cDNA microarray data was consistent with that of 2D-PAGE data, and that daidzein exposure affected TM4 cell viability as a result of changes in both mRNA and protein expression that differed from those associated with exposure to E_2, DES, or staurosporin (Figs. 2 and 4). In this study, hierarchical clustering of cDNA microarray data was consistent with that of 2D-PAGE data, in spite of the fact that mRNA expression changes did not always coincide with protein expression changes. We demonstrated for the first time that using the data on both mRNA expression and protein expression profiles makes it possible to distinguish these chemicals. Recently, DNA microarray analysis has been used in the study of signal transduction (Fambrough et al., 1999), in drug discovery (Marton et al., 1999), in diagnosis of cancer (Alizadeh et al., 2000), and in analysis of single nucleotide polymorphism (Cargill et al., 1999; Halushka et al., 1999). Moreover, 2D-PAGE analysis and spot sequence analysis have also been used in the study of signal transduction (Sheffield and Gavinski, 2003) and cancer (Rai and Chan, 2004), among other fields. Based on the results of this study, we surmise that a combination of microarray data analysis with 2D-PAGE data analysis makes it possible to classify endocrine disrupter chemicals. However, we will conduct a future study in order to identify, through mass spectrometry (MS) analysis, which chemicals cause changes in protein spots. In conclusion, using cDNA microarray and 2D-PAGE analysis, we demonstrated that exposure to daidzein, E_2, DES and staurosporin induces changes in mRNA and protein expression in TM4 cells. We also demonstrated for the first time that these chemicals can be classified by clustering analysis, using data from cDNA microarray and 2D-PAGE analysis, and that exposure to daidzein induced different effects from those induced by E_2, DES and staurosporin. We surmise that both mRNA expression and protein expression profiles can be used as tools for classifying these chemicals, and may be useful in predicting their effects. In this study, we used only daidzein, DES, E_2 and staurosporin, however, in the future study, mRNA expression and protein expression profile of testicular cell exposed to many various chemical compounds, including phytoestrogen, supports to determine not only the traditional single function but also the multi-functional events. Moreover, in the future study, the classification of many endocrine disrupter chemicals, whose action is unknown, using the data of cDNA microarray and 2D-PAGE helps to predict the similar or different effect of the testis on the exposure to endocrine disrupter chemicals. We propose that, from a toxicogenomic standpoint, combination of cDNA microarray analysis with 2D-PAGE analysis is a useful tool for classification of chemicals, and that using MS analysis, the genes and proteins thus detected can be useful as biological markers of the effects of exposure to relevant chemicals.

Acknowledgments

We are grateful to the Dr. Hitoshi Tainaka (Asahi Techno Glass Co. Funabashi, Japan) for the donation cDNA microarrays. This study was supported by grants-in-aid from the Ministry of Education, Science, Sports, Culture and Technology of Japan, as part of the 21st Century COE program "Knowledge Information Infrastructure for Genome Science". It was supported by the Fund for Endocrine Disrupters provided by the Ministry of the Environment, Japan.

References

Adachi, T., Komiyama, M., Ono, Y., Koh, K.B., Sakurai, K., Shibayama, T., Kato, M., Yoshikawa, T., Seki, N., Iguchi, T., Mori, C., 2002. Toxicogenomic effects of neonatal exposure to diethylstilbestrol on mouse testicular gene expression in the long term: a study using cDNA microarray analysis. Molecular Reproduction and Development 63, 17–23.

Adachi, T., Ono, Y., Koh, K.B., Takashima, K., Tainaka, H., Matsuno, Y., Nakagawa, S., Todaka, E., Sakurai, K., Fukata, H., Iguchi, T., Komiyama, M., Mori, C., 2004. Long-term alteration of gene expression without morphological change in testis after neonatal exposure to genistein in mice: toxicogenomic analysis using cDNA microarray. Food and Chemical Toxicology 42, 445–452.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson Jr., J., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., Staudt, L.M., 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403, 503–511.

Bingham, S.A., Atkinson, C., Liggins, J., Bluck, L., Coward, A., 1998. Phyto-oestrogens: Where are we now? British Journal of Nutrition 79, 393–406.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S., 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature Genetics 22, 231–238.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P., Trent, J.M., 1999. Expression profiling using cDNA microarrays. Nature Genetics 21, 10–14.

Fambrough, D., McClure, K., Kazlauskas, A., Lander, E.S., 1999. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of gene. Cell 97, 727–741.

Garrels, J.I., 1989. The QUEST system for quantitative analysis of two-dimensional gels. Journal of Biological Chemistry 264, 5269–5282.

Goodman, M.T., Hankin, J.H., Wilkens, L.R., Lyu, L.C., McDuffie, K., Liu, L.Q., Kolonel, L.N., 1997. Diet, body size, physical activity, and the risk of endometrial cancer. Cancer Research 57, 5077–5085.

Guo, J.M., Xiao, B.X., Liu, D.H., Grant, M., Zhang, S., Lai, Y.F., Guo, Y.B., Liu, Q., 2004. Biphasic effect of daidzein on cell growth of human colon cancer cells. Food and Chemical Toxicology 42, 1641–1646.

Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., Chakravarti, A., 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nature Genetics 22, 239–247.

Hirayama, T., 1984. Epidemiology of stomach cancer in Japan. Journal of Clinical Oncology 14, 159–168.

Hossain, M.A., Bouton, C.M., Pevsner, J., Laterra, J., 2000. Induction of vascular endothelial growth factor in human astrocytes by lead. Involvement of a protein kinase C/activator protein-1 complex-dependent and hypoxia-inducible factor 1-independent signaling pathway. Journal of Biological Chemistry 275, 27874–27882.

Ishiyama, M., Tominaga, H., Shiga, M., Sasamoto, K., Ohkura, Y., Ueno, K., 1996. A combined assay of cell viability and in vitro cytotoxicity with a highly water-soluble tetrazolium salt, neutral red and crystal violet. Biological and Pharmaceutical Bulletin. 19, 1518–1520.

Komiyama, M., Adachi, T., Mori, C., 2003. Analysis of toxicogenomic response to endocrine disruptors in the mouse testis. In: Inoue, T., Pennie, W.D. (Eds.), Toxicogenomics. Springer-Verlag Tokyo, Tokyo, pp. 156–162.

Koo, L.C., 1988. Dietary habits and lung cancer risk among Chinese females in Hong Kong who never smoked. Nutrition Cancer 11, 155–172.

Kuiper, G.G., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., van der Saag, P.T., van der Burg, B., Gustafsson, J.A., 1998. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinology 139, 4252–4263.

Kurzer, M.S., Xu, X., 1997. Dietary phytoestrogens. Annual Review of Nutrition 17, 353–381.

Lee, H.P., Gourley, L., Duffy, S.W., Esteve, J., Lee, J., Day, N.E., 1991. Dietary effects on breast-cancer risk in Singapore. Lancet 337, 1197–1200.

Magee, P.J., Rowland, I.R., 2004. Phyto-oestrogens, their mechanism of action: current evidence for a role in breast and prostate cancer. British Journal of Nutrition 91, 513–531.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett Jr., D.E., Hartwell, L.H., Brown, P.O., Friend, S.H., 1999. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nature Medicine 4, 1293–1301.

Newbold, R.R., Banks, E.P., Bullock, B., Jefferson, W.N., 2001. Uterine adenocarcinoma in mice treated neonatally with genistein. Cancer Research 61, 4325–4328.

Rai, A.J., Chan, D.W., 2004. Cancer proteomics: serum diagnostics for tumor marker discovery. Annals of the New York Academy of Sciences 1022, 286–294.

Rockett, J.C., Dix, D.J., 1999. Application of DNA arrays to toxicology. Environmental Heath Perspective 107, 681–685.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Seveson, R.K., 1989. A prospective study of demographics, diet, and prostate cancer among men of Japanese ancestry in Hawaii. Cancer Research 49, 1857–1860.

Sheffield, L.G., Gavinski, J.J., 2003. Proteomics methods for probing molecular mechanisms in signal transduction. Journal of Animal Sciences 81, 48–57.

Tajima, K., Tominaga, S., 1985. Dietary habits and gastro-intestinal cancers: a comparative case-control study of stomach and large intestine cancers in Nagoya, Japan. Japan Journal of Cancer Research 76, 705–716.

Yoshikawa, T., Nagasugi, Y., Azuma, T., Kato, M., Sugano, S., Hashimoto, K., Masuho, Y., Muramatsu, M., Seki, N., 2000. Isolation of novel mouse genes differentially expressed in brain using cDNA microarray. Biochemical and Biophysical Research Communications 275, 532–537.