Chemico-Biological Interactions 223 (2014) 38–44

Contents lists available at ScienceDirect

Chemico-Biological Interactions

journal homepage: www.elsevier.com/locate/chembioint

An *in silico* toxicogenomics approach for inferring potential diseases associated with maleic acid

Ying-Chi Lin [a,b,1], Chia-Chi Wang [a,b,C,d,1], Chun-Wei Tung [a,b,d,*]

[a] School of Pharmacy, College of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan
[b] Ph.D. Program in Toxicology, College of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan
[c] Institute of Environmental Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan
[d] National Environmental Health Research Center, National Health Research Institutes, Miaoli County, Taiwan

ARTICLE INFO

Article history:
Received 7 May 2014
Received in revised form 25 July 2014
Accepted 5 September 2014
Available online 17 September 2014

Keywords:
Maleic acid
Toxicogenomics
Diseases
In silico analysis

ABSTRACT

Maleic acid is a multi-functional chemical widely applied in the manufacturing of polymer products including food packaging. However, the contamination of maleic acid in modified starch has raised the concerns about the effects of chronic exposure to maleic acid on human health. This study proposed a novel toxicogenomics approach for inferring functions, pathways and diseases potentially affected by maleic acid on humans by using known interactions between maleic acid and proteins. Neuronal signal transmission and cell metabolism were identified to be most influenced by maleic acid in this study. The top disease categories inferred to be associated with maleic acid were mental disorder, nervous system diseases, cardiovascular diseases, and cancers. The results from the *in silico* analysis showed that maleic acid could penetrate the blood–brain barrier to affect the nervous system. Several functions and pathways were further analyzed and identified to give insights into the mechanisms of maleic acid-associated diseases. The toxicogenomics approach may offer both a better understanding of the potential risks of maleic-acid exposure to humans and a direction for future toxicological investigation.

1. Introduction

Maleic acid is cis-isomer of butenedioic acid used as a fragrance ingredient and pH adjuster in beauty products or cosmetics [1]. This chemical is also widely used in the manufacturing of polymer products including food packaging and is listed as a legal indirect component in foods in both the United States and the European Union countries. The oral LD_{50} of the maleic acid are 708 and 2400 mg/kg in rat and mouse, respectively [2]. Generally, the health risk associated with oral exposure to maleic acid is considered to be low. Maleic anhydride, which is rapidly converted to maleic acid when encountering water, had been illegally added to modified starch to enhance favorable properties, such as elasticity. The Food and Drug Administration in Taiwan uncovered the illegal high levels of maleic acid in food products such as tapioca starch, noodles, and hotpot ingredients and suspended the sale and distribution of these contaminated food products [3]. The adulteration of maleic anhydride in modified starch gives rise to the concern about the long-term human oral exposure to maleic acid, especially in Taiwan.

Maleic acid has been reported to induce the nephrotoxicity in rabbits, rats and dogs [1,4–6]. Maleic acid induced the renal tubular injury and cell necrosis in the proximal tubules of treated rats, which resulted in phenotypes resembling Fanconi's syndrome, a disease characterized by increased urinary excretion of glucose, amino acids, phosphate and proteins [5,6]. Maleic acid interfered renal proximal Na^+ and H^+ transport and inhibited the activity of proximal tubule Na-K-ATPase and H-ATPase [7,8]. However, the toxicological effects of maleic acid on human health are still largely unknown. Despite the lack of epidemiological studies, the integration and bioinformatics analysis of the available data in chemical–gene/protein interaction will generate possible mechanisms for evaluating the effects of maleic acid on human health.

The Comparative Toxicogenomics Database (CTD) consisting of high-confident chemical–gene interactions curated from selected literatures is a useful resource for understanding chemical induced diseases [9]. The database has been successfully applied to analyze diseases associated with phthalates [10]. Currently, there are only a few curated interacting genes associated with maleic acid in this database, which was insufficient for bioinformatics analyses.

* Corresponding author at: 100 Shih-Chuan 1st Road, Kaohsiung 80708, Taiwan. Tel.: +886 7 3121101x2316; fax: +886 7 3135612.
E-mail addresses: yclin@kmu.edu.tw (Y.-C. Lin), chiachiwang@kmu.edu.tw (C.-C. Wang), cwtung@kmu.edu.tw (C.-W. Tung).
[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.cbi.2014.09.004
0009-2797/

Table 1
Top ranking proteins interacted with maleic acid.

Gene symbol	Name	Score	Species*
GOT2	Glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)	0.846	H,D,R
HMOX1	Heme oxygenase (decycling) 1	0.836	H,D,R
PTS	6-Pyruvoyltetrahydropterin synthase	0.809	H,D,R
TEK	TEK tyrosine kinase, endothelial	0.800	H
KDR	Kinase insert domain receptor (a type III receptor tyrosine kinase)	0.800	H
HGF	Hepatocyte growth factor (hepapoietin A; scatter factor)	0.800	H,R
GOT1	Glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1)	0.764	H,D,R
GRIK1	Glutamate receptor, ionotropic, kainate 1	0.728	H,R
HTR1A	5-Hydroxytryptamine (serotonin) receptor 1A	0.722	H
GLS2	Glutaminase 2 (liver, mitochondrial)	0.699	H,R
HTR1B	5-Hydroxytryptamine (serotonin) receptor 1B	0.696	H
GRIN2B	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B	0.694	H,D,R
GRIN2A	Glutamate receptor, ionotropic, N-methyl D-aspartate 2A	0.652	H,D,R
ELANE	Elastase, neutrophil expressed	0.613	H
DRD5	Dopamine receptor D5	0.612	H
DRD2	Dopamine receptor D2	0.611	H,D,R
FOLH1	Folate hydrolase (prostate-specific membrane antigen) 1	0.557	H
HTR2B	5-Hydroxytryptamine (serotonin) receptor 2B	0.528	H,D
ALPL	Alkaline phosphatase, liver/bone/kidney	0.526	H,D
ALPI	Alkaline phosphatase, intestinal	0.526	H,D
KLRB1	Killer cell lectin-like receptor subfamily B, member 1	0.526	H
HTR2A	5-Hydroxytryptamine (serotonin) receptor 2A	0.482	H
PTMA	Prothymosin, alpha	0.476	H
DRD4	Dopamine receptor D4	0.471	H,R
SIGMAR1	Sigma non-opioid intracellular receptor 1	0.438	H,D
GGT2	Gamma-glutamyltransferase 8 pseudogene	0.431	H
GGT5	Gamma-glutamyltransferase 5	0.431	H,D
GGT7	Gamma-glutamyltransferase 7	0.431	H,D,R
GGT1	Gamma-glutamyltransferase 1	0.431	H,D
PDE2A	Phosphodiesterase 2A, cGMP-stimulated	0.415	H,D,R
TREH	Trehalase (brush-border membrane glycoprotein)	0.414	H,D,R

* H: Human; D: Dog; R: Rat.

In an effort to further evaluate the effects of maleic acid on human health, we designed a novel toxicogenomics approach incorporating the largest chemical–protein interaction database STITCH [11], Gene Ontology (GO) enrichment analysis and disease inference. First, the proteins interacting with maleic acid were identified from STITCH database. Subsequently, the interacting proteins were utilized to infer diseases, functions and pathways associated with maleic acid. As a result, diseases including mental disorders, cancers and diseases of nervous and cardiovascular systems were inferred through the toxicogenomics analysis approach with affected functions and pathways providing a better insight to the influence of maleic acid on human health. The toxicogenomics analysis approach is also expected to be useful for other environmental and industry chemicals.

2. Material and methods

2.1. Chemical–protein interactions

Chemical–protein interaction data was retrieved from STITCH database of version 3.1 [11], an aggregated database of interactions connecting over 300,000 chemicals and 2.6 million proteins from 1133 organisms. The interaction data was curated as three interaction scores from different data sources of experiments, databases and text-mining representing the likelihood or relevance of interactions [12]. The experiment part consists of direct chemical–protein binding data with experimental evidence derived from PDSP K_i [13] and PDB [14]. The database part contains interaction data mainly from pathway databases of KEGG [15] and Reactome [16]. The text-mining data was constructed by extracting information of interactions from literatures from MEDLINE and OMIM using text-mining techniques of both a simple co-occurrence scheme and a more complex natural language processing approach [12]. In addition to the three scores, STITCH database provides a combined score S for a given chemical–protein interaction generated by combining the three scores of corresponding evidence types using a Bayesian scoring scheme defined as: $S = 1 - \prod_i (1 - S_i)$, where S_i denotes the three scores of corresponding evidence types [17]. This study utilized the combined score to represent a chemical–protein interaction whose score is an integer value ranging from 0 (no interaction) to 1 (high-confident interaction). The default cutoff for confident interactions is 0.4 [11,17]. Chemical–protein interactions were transferred among species based on the sequence similarity of the proteins [17]. The chemical–protein interactions have also been successfully utilized to predict non-genotoxic hepatocarcinogenicity [18,19].

2.2. Enrichment analysis of Gene Ontology, pathway and disease

Gene Ontology (GO) is the controlled vocabulary of describing gene products which covers three domains, namely cellular component (CC), molecular function (MF), and biological process (BP) [20]. GO offers a useful resource for studying gene functions [21]. The identification of enriched GO terms from a given gene list could give insights into the overrepresented functions of the genes. GO terms have also been utilized as features for machine learning [22,23]. In addition to GO terms, the enrichment analyses of pathways and diseases are also helpful tools for better understanding of the influenced pathways and diseases relevant to maleic acid. The pathway-gene relationships for enrichment analysis are based on KEGG [24] and REACTOME [25] pathway databases. The analysis of enriched diseases utilize MEDIC disease vocabulary [26], a combination of Medical Subject Headings (MeSH) [27] and Online Mendelian Inheritance in Man (OMIM) [28] databases. Several web