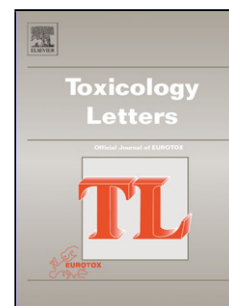


Journal Pre-proof

Prediction of the allergic mechanism of haptens *via* a reaction-substructure-compound-target-pathway network system

Peiwen Di, Zengrui Wu, Hongbin Yang, Weihua Li, Yun Tang, Guixia Liu



PII: S0378-4274(19)30291-7
DOI: <https://doi.org/10.1016/j.toxlet.2019.09.019>
Reference: TOXLET 10580

To appear in: *Toxicology Letters*

Received Date: 8 May 2019
Revised Date: 30 July 2019
Accepted Date: 25 September 2019

Please cite this article as: Di P, Wu Z, Yang H, Li W, Tang Y, Liu G, Prediction of the allergic mechanism of haptens *via* a reaction-substructure-compound-target-pathway network system, *Toxicology Letters* (2019), doi: <https://doi.org/10.1016/j.toxlet.2019.09.019>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.

Prediction of the allergic mechanism of haptens via a reaction-substructure-compound-target-pathway network system

Peiwen Di, Zengrui Wu, Hongbin Yang, Weihua Li, Yun Tang*, Guixia Liu*

Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China

* Corresponding authors. Tel: +86-21-64250811; Fax: +86-21-64251033.

E-mail addresses: gxliu@ecust.edu.cn (G. Liu), ytang234@ecust.edu.cn (Y. Tang).

Highlights:

- A network system was developed to predict reactions, targets and pathways of haptens.
- It is the first time to explore the relationships between reactions and skin targets.
- 12 key substructures were identified to determine the protein binding reactions.
- 33 skin targets and 12 pathways were acquired to understand the allergic mechanism.

ABSTRACT

Skin sensitization, frequently leading to allergic contact dermatitis (ACD), is authenticated to be a significant endpoint in the field of drug discovery and cosmetics. The initiation of ACD, also known as the skin sensitization mechanism, has been documented as an adverse outcome pathway (AOP), which can be studied experimentally and computationally. In this study, we collected 154 haptens and applied systems toxicology methods to develop a reaction-substructure-compound-target-pathway network system. For the collected haptens, their key substructures were identified and associated with their protein binding reactions. The targets of haptens, including the known targets collected from four databases and the potential targets predicted *via* our balanced substructure-drug-target network-based inference (bSDTNBI) method, were matched to skin proteins to obtain skin targets. The dermatitis-related pathways were enriched and were subject to literature verification. The network system we developed can be applied to predict the reactions, targets and pathways of new haptens, which contributed to evaluating chemical safety and optimizing chemical structures. The study of skin sensitization mechanism is helpful for understanding the skin immunity and resisting ACD.

Keywords: Skin sensitization; Allergic contact dermatitis; Adverse outcome pathway; Systems toxicology; Pathway enrichment; Hazard identification

Abbreviations

Ac	acylation reactions
ACC	accuracy
ACD	allergic contact dermatitis
AOP	adverse outcome pathway
bSDTNBI	balanced substructure-drug-target network-based inference
CAS	chemical abstracts service
CD	cluster of differentiation
DC	dendritic cell
FP	false positive
HIF-1	hypoxia-inducible factor 1
KC	keratinocyte
MA	Michael addition reactions
MoSS	Molecular Substructure Miner
NF	negative frequency
NF-kappa B	nuclear factor-kappa B
OECD	the Organization for Economic Cooperation and Development
PF	positive frequency
PI3K-Akt	phosphatidylinositol 3'-kinase-Akt
SB	Schiff base reactions
SMILES	simplified molecular input line entry system
S _N 1	unimolecular nucleophilic substitution reactions
S _N 2	bimolecular nucleophilic substitution reactions
S _N Ar	nucleophilic aromatic substitution reactions
TCR	T cell receptor
TLR	Toll-like receptor
TNF	tumor necrosis factor
TP	true positive

1. Introduction

Skin sensitization is a complicated dermato-immunologic process caused by the local contact of haptens with skin, and often leads to the clinical symptoms known as allergic contact dermatitis (ACD). An estimated 15-20% of the general population suffers from ACD, making skin sensitization a significant human health issue that should be considered in chemical hazard and risk assessment (Di et al., 2019; Mehling et al., 2012; Peiser et al., 2012).

Systems toxicology, which replaces “one compound-one target” paradigm with “one compound-multiple targets” strategy, is the incorporation of traditional toxicology with network models and quantitative analysis of molecular and functional changes (Sturla et al., 2014). The systems toxicology methods have been testified in our previous studies (Peng et al., 2019; Wang et al., 2017; Wang et al., 2018) to be considerate and promising methods for studying toxicological mechanisms, and were applied in this article to predict the allergic mechanism of haptens and study the mechanism of skin sensitization.

The initiation of ACD, also known as the skin sensitization mechanism, has been documented as an adverse outcome pathway (AOP) by the Organization for Economic Cooperation and Development (OECD) (OECD, 2013). The AOP for skin sensitization includes 4 key mechanistic events: (1) The haptens are covalently bound to endogenous proteins in the skin. (2) The epidermal keratinocytes (KCs) are activated. (3) The dendritic cells (DCs) are activated. (4) The haptenated protein

fragments (antigens) are presented to T cells, causing the activation and proliferation of T cells (Ezendam et al., 2016; Goebel et al., 2012; Kleinstreuer et al., 2018). The activated T cells are repeatedly exposed to the same allergen. Ultimately, the adverse outcome occurs, namely ACD.

The molecular initiating event of the skin sensitization AOP is the covalent binding of haptens to endogenous proteins in the skin. The protein binding reactions mainly include the following 5 categories: Michael addition reactions (MA); Schiff base reactions (SB); unimolecular/bimolecular nucleophilic substitution reactions (S_N1/S_N2); acylation reactions (Ac); nucleophilic aromatic substitution reactions (S_NAr) (Aptula et al., 2005; Di et al., 2019; Goebel et al., 2012; Urbisch et al., 2016). There have been some computational tools with simulators to predict skin sensitization. The “protein binding alerts for skin sensitization by OASIS v1.4” wizard of OECD QSAR Toolbox (<https://www.qsartoolbox.org/>) (Diderich, 2010) has been introduced to be the endpoint-specific profiler to identify the protein binding events of skin sensitization. The protein binding reactions of haptens can be inferred from their key substructures. However, the key substructures associated with the protein binding reactions have not been sorted out yet.

In addition, another issue worth exploring is how to identify the skin targets of haptens and thereby obtain the dermatitis-related pathways. With regard to new haptens, part of their targets can be collected from databases, but other targets cannot be directly searched. Since it is time-consuming and almost impossible to determine

all the targets of haptens by experimental methods, the potential targets in this study were predicted *via* our balanced substructure-drug-target network-based inference (bSDTNBI) method, an integrated network and chemoinformatics tool for the systematic prediction of drug-target interactions and drug repositioning (Wu et al., 2016; Wu et al., 2017). The targets of haptens, including the known targets and the potential targets, were matched to the skin proteins to obtain the skin targets of haptens. The dermatitis-related pathways can be enriched by the skin targets of haptens.

The workflow diagram illustrated in Fig. 1 contributed to comprehending our research procedures. In this study, we collected 154 haptens and intended to achieve the following 4 objectives: (1) identifying the key substructures and associating them with the protein binding reactions; (2) acquiring the skin targets and enriching the dermatitis-related pathways; (3) exploring the relationship between protein binding reactions and skin targets; (4) constructing the network system and verifying the rationality of prediction results. The reaction-substructure-compound-target-pathway network system can be applied not only to predict the reactions, targets and pathways of new haptens, but also to help comprehend the mechanism of skin sensitization.

2. Materials and methods

2.1. Collection of haptens and their protein binding reactions

In this study, 154 haptens along with their chemical abstracts service (CAS) registry numbers, theoretical protein binding reactions and potency classes were

collected from Dearden et al. (2015). It was noteworthy that the objects of this study were haptens rather than prohaptens or prehaptens. Though numerous skin sensitizers were available in the databases, such as the OECD's eChemPortal database and the NICEATM database (Clippinger et al., 2018), they cannot be used in this study because we cannot pick out haptens from these skin sensitizers.

154 haptens included 62 haptens with weak potency, 64 haptens with moderate potency, 16 haptens with strong potency and 12 haptens with extreme potency. Their simplified molecular input line entry system (SMILES) codes were collected from NCI/CADD Chemical Identifier Resolver (<https://cactus.nci.nih.gov/chemical/structure>) and ChemIDplus Advanced (<https://chem.nlm.nih.gov/chemidplus/>) using CAS registry numbers. Their PubChem CID codes were collected from the pubchem database (<https://pubchem.ncbi.nlm.nih.gov/>).

The protein binding reactions of haptens were identified from two aspects. On one hand, Dearden et al. (2015) has listed the theoretical protein binding reactions of 154 haptens, which were collected from Gerberick et al. (2005) and Kern et al. (2010). On the other hand, the “protein binding alerts for skin sensitization by OASIS v1.4” wizard of OECD QSAR Toolbox was employed to identify the protein binding reactions of haptens considering the inhibitory effect of electron and stereo factors on the protein binding in the skin. Therefore, the protein binding reactions of 154 haptens were identified by combining two results.

2.2. Identification of the key substructures

The key substructures were defined as molecular functional groups which were of high frequency in localized compounds and of low frequency in the other compounds. The key substructures were separately analyzed for each type of protein binding reaction, so as to comprehend the reaction mechanisms of known haptens and predict the reaction mechanisms of new haptens. When seeking for the key substructures associated with the specific reaction, we labeled the haptens with this reaction as active and labeled the haptens with other reactions as inactive. In this study, MoSS (Molecular Substructure Miner) and SARpy were jointly applied to analyze the key substructures of haptens. It was noteworthy that we intended to identify the key substructures associated with MA, SB, S_N2 and Ac reactions because of the inadequate number of the haptens with S_NAr reactions or S_N1 reactions.

MoSS on the KNIME (Konstanz Information Miner) platform (<https://www.knime.com/>) (Berthold et al., 2007) was a program applied to search frequent molecular substructures and discriminative fragments in a database of molecule descriptions. The KNIME workflow for identifying key substructures was sequentially linked by 3 nodes: the File Reader node, the MoSS node, the XLS Writer node (as shown in Fig. 1). The File Reader node was applied to import the SMILES codes and the protein binding reactions of 154 haptens. In the MoSS node, the atomic numbers of fragments were controlled in 2-10 and the pure carbon fragments were ignored. After running the workflow, we acquired the key substructures and their frequencies from the XLS Writer node.

SARpy (Ferrari et al., 2013) was a free standalone and user-friendly tool which applied “string mining” to directly identify substructures from the SMILES codes. The parameter used in SARpy was “optimal” in the acquisition of substructures. The substructure frequency analysis (Jensen et al., 2007) was applied to associate the key substructures with each compound. In order to evaluate the frequency of each substructure, the positive frequency (*PF*), negative frequency (*NF*) and accuracy (*ACC*) were calculated with the following formulas:

$$PF = \frac{TP}{P} \quad (1)$$

$$NF = \frac{FP}{N} \quad (2)$$

$$ACC = \frac{TP}{TP + FP} \quad (3)$$

In the formulas, *P* represents the number of localized compounds; *N* represents the number of outside compounds; *TP* (true positive) represents the number of localized compounds which contain the substructure; *FP* (false positive) represents the number of outside compounds which contain the substructure.

The substructures were regarded as key substructures if they simultaneously satisfied the following two criteria: positive frequency (*PF*) > 0.1 and negative frequency (*NF*) < 0.05. After the key substructures of haptens were identified by MoSS and SARpy, the compound-substructure pairs and substructure-reaction pairs were ascertained and could be applied to build a compound-substructure-reaction subnetwork.

2.3. Acquisition of the targets

First of all, it was clear that the criteria for the targets in this study were the following 5 points: (1) K_i , K_d , IC_{50} or $EC_{50} \leq 10 \mu M$; (2) the target protein could be described by a sole UniProt accession number; (3) the target protein was tagged as “reviewed” in the UniProt database (<https://www.uniprot.org/>) (Apweiler et al., 2004); (4) the target protein was derived from Homo sapiens; (5) the molecular weight of compound ranged from 100-600 Daltons (Wu et al., 2016).

The known targets of 154 haptens were discovered from 4 databases including ChEMBL Database (<https://www.ebi.ac.uk/chembl/>) (Gaulton et al., 2012), Binding Database (<http://www.bindingdb.org/bind/index.jsp>) (Liu et al., 2007), IUPHAR/BPS Guide to PHARMACOLOGY (<http://www.guidetopharmacology.org/>) (Pawson et al., 2014) and PDSP Ki Database (<https://pdsp.unc.edu/databases/kidb.php>) (Roth et al., 2000). These 4 databases provided the compound-target pairs and could be applied to collect the experimentally-determined targets of haptens.

The potential targets of 154 haptens were speculated with our bSDTNBI method, which is an integrated network and chemoinformatics tool for large-scale compound-target interactions prediction and drug repositioning. The schematic diagram of bSDTNBI method was illustrated in Fig. 2. The principle and application of bSDTNBI method were described in the Section 1 of the Supporting Information. In this study, the parameters were set as $\alpha = \beta = 0.1$, $\gamma = -0.5$, and $k = 2$, as they were adopted in our previous study (Wu et al., 2016). We regarded the top 30 predicted targets of each hapten as potential targets, so we obtained 4620 potential

compound-target pairs of 154 haptens. If the potential compound-target pairs overlapped with the known compound-target pairs, we removed the potential pairs and retained the known pairs. Therefore, the targets of 154 haptens were acquired as a result of collecting the known targets and predicting the potential targets.

The targets of 154 haptens were matched to the skin proteins to obtain the skin targets of haptens. The skin proteins were collected from 3 databases, including HuGE Navigator (<https://phgkb.cdc.gov/PHGKB/hNHome.action>) (Yu et al., 2008), the Comparative Toxicogenomics Database (<http://ctdbase.org/>) (Davis et al., 2018) and the Human Protein Atlas Database (<http://www.proteinatlas.org/>) (Uhlén et al., 2015). The intersections of the targets of haptens and the skin proteins were regarded as the skin targets of haptens, which were applied to enrich pathways.

2.4. Enrichment of the dermatitis-related pathways

The implement of pathway enrichment used in this study was the functional annotation tool in the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (<https://david.ncifcrf.gov/>) (Huang et al., 2009). We uploaded the official gene symbols, limited annotations by Homo sapiens and set current background to be Homo sapiens. We acquired 5 kinds of pathways from different sources and primarily thought about frequently-used KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. The pathways were regarded as key pathways when satisfying the condition of $p\text{-value} < 0.05$. The details about KEGG pathways were acquired from the website <https://www.kegg.jp/kegg/pathway.html> to

comprehend the pathways.

In this study, the dermatitis-related pathways were enriched from the skin targets of haptens, and were subject to literature verification. The pathways associated with other diseases, such as “inflammatory bowel disease” (hsa05321), could also be obtained, but they were neglected. Various studies have been conducted on the targets and pathways of ACD (Cottrez et al., 2015; Lambrechts et al., 2010), which could be used to verify the targets and pathways obtained in this paper. If there were some enriched pathways that were not mentioned in the literature, we analyzed them to judge whether they were indeed related to ACD.

2.5. Exploration of the relationship between protein binding reactions and skin targets

With regard to each hapten, we acquired its protein binding reactions and its skin targets, which provided the condition for us to investigate the relationship between protein binding reactions and skin targets. We employed Fisher’s exact test (Fisher, 1922) and utilized the *Score* value to quantify the correlation degree of the relationship with the following formula:

$$Score = \frac{\binom{X}{x} \binom{Y}{y}}{\binom{X+Y}{x+y}} = \frac{X!Y!(x+y)!(X+Y-x-y)!}{x!y!(X-x)!(Y-y)!(X+Y)!} \quad (4)$$

where x represents the number of compounds which undergo the reaction and act on the target; y represents the number of compounds which undergo the reaction but do not act on the target; X represents the number of compounds which act on the target; Y

represents the number of compounds which do not act on the target.

The sum of x and y is equal to the number of compounds which undergo the reaction, while the sum of X and Y is equal to the number of compounds. Considering that the haptens with S_NAr reactions or S_N1 reactions were insufficient to be studied, we separately calculated the *Score* values for MA, SB, S_N2 and Ac reactions with regard to each skin target. There were high correlations if the *Score* values < 0.05 . By evaluating numerous *Score* values, we could quantify the correlation degree of the reaction-target relationship, in order to single out the high-correlated reaction-target pairs.

2.6. Construction of the network system

The compound-substructure pairs were obtained via MoSS and SARpy. The compound-reaction pairs were obtained via data collection and OECD QSAR Toolbox. The substructure-reaction pairs were associated. The compound-target pairs were obtained via data collection and bSDTNBI method. The target-pathway pairs were obtained via pathway enrichment. The reaction-target relationship was explored via Fisher's exact test. Therefore, we acquired six sets of pairs, and accordingly constructed the reaction-substructure-compound-target-pathway network system. The network system could be applied to predict the reactions, targets and pathways of haptens, and comprehend the mechanism of skin sensitization. The schematic diagram was charted with Cytoscape (<https://cytoscape.org/>) to receive a global perspective of the associations.

3. Results

3.1. The haptens and their protein binding reactions

The chemical names, CAS registry numbers, PubChem CID codes, SMILES codes, EC3 values and potency classes of 154 haptens were summarized in Table S1. In this article, the “protein binding reactions obtained *via* data collection” were abbreviated as “theoretical reactions”, while the “protein binding reactions obtained *via* OECD QSAR Toolbox” were abbreviated as “OECD reactions”. The “theoretical reactions” and “OECD reactions” of 154 haptens were also listed in Table S1.

Comparing the “theoretical reactions” and “OECD reactions”, we discovered that the “theoretical reactions” and “OECD reactions” of 154 haptens were basically the same. According to the literature (Dearden et al., 2015; Gerberick et al., 2005; Kern et al., 2010), 154 haptens included 45 haptens with MA reactions, 35 haptens with SB reactions, 45 haptens with S_N2 reactions, 23 haptens with Ac reactions, 4 haptens with S_NAr reactions, and 2 haptens with S_N1 reactions. So, all the haptens possessed “theoretical reactions”. Of 154 haptens, 144 haptens had “OECD reactions” but the other 10 haptens had no “OECD reactions”. Of the 144 haptens with “OECD reactions”, only one hapten had different “theoretical reaction” and “OECD reaction”, which was iodopropynyl butylcarbamate (CAS: 55406-53-6). According to the literature (Kern et al., 2010), it was identified as a hapten with S_N2 reaction, because it possessed the halogen and could act as a S_N2 electrophile. The OECD QSAR Toolbox identified it as a hapten with Ac reaction, because it may undergo ester

aminolysis. In our opinion, we agreed with the view of literature and regarded it as a hapten with S_N2 reaction.

As a conclusion, the OECD QSAR Toolbox could accurately identify the protein binding reactions of most haptens, and had the advantage of quickly determining the reaction types of numerous haptens. For new haptens, we could obtain the protein binding reactions not only through the OECD QSAR toolbox, but also through their key substructures.

3.2. The key substructures of haptens

With MoSS and SARpy, we identified 12 key substructures and evaluated their frequencies, which were listed in Table 1. The 121 compound-substructure pairs obtained with 154 haptens and 12 key substructures were summarized in detail in Table S2. For ease of understanding the quantitative descriptions and visualizing the results of key substructures, the diagram of compound-substructure associations was illustrated in Fig. 3. As shown in Table 1 and Fig. 3, the number of key substructures associated with MA, SB, S_N2 , and Ac reactions was 1, 3, 3, and 5, respectively.

Substructure I, which possesses a double bond with electron-withdrawing substituent carbonyl group, can act as a Michael acceptor. Substructures II, III, IV, which are reactive carbonyl compounds, can act as Schiff base electrophiles. Substructures V, VI, which possess the halogens, can act as S_N2 electrophiles. Substructure VII, which possesses the sulfonyl, can act as a S_N2 electrophile. Substructures VIII, IX, X, XI, XII, which possess the carbonyl and leaving groups,

can act as acylating agents.

Therefore, 12 substructure-reaction pairs were obtained and their protein binding reactions were illustrated in Table 1. The 12 key substructures were associated with 4 kinds of protein binding reactions (MA, SB, S_N2, Ac). The haptens having S_NAr reactions or S_N1 reactions were relatively few and not enough to be studied. Similar problems have arisen in previous studies (Aptula et al., 2005). However, after comparing the only 4 haptens having S_NAr reactions, we could speculate that the haptens with S_NAr reactions usually possessed aromatic rings with leaving substituents, which was consistent with the condition for the occurrence of S_NAr reactions.

After acquiring 121 compound-substructure pairs, 12 substructure-reaction pairs, and 154 compound-reaction pairs, we obtained the compound-substructure-reaction groups (Table S2) and accordingly built the compound-substructure-reaction subnetwork.

3.3. The skin targets of haptens

As for 154 haptens, 243 known compound-target pairs were collected from 4 databases. We regarded the top 30 predicted targets of each compound as potential targets, so we obtained 4620 potential compound-target pairs of 154 haptens. Among 4620 potential compound-target pairs, there were 59 potential pairs that overlapped with the known pairs and should be removed. Therefore, we collected 4804 non-redundant compound-target pairs including 243 known pairs and 4561 potential

pairs (Table S3). 435 non-redundant targets of haptens were obtained. We matched 435 targets of haptens to skin targets, and obtained 33 dermatitis-related targets of haptens (Table 2). As for 154 haptens and 33 dermatitis-related targets of haptens, we obtained 523 compound-target pairs (Table S2, Table S3, Table S4), which were applied to build the network system.

3.4. The dermatitis-related pathways

After carrying out the pathway enrichment with 33 dermatitis-related targets of haptens, we obtained 12 dermatitis-related pathways. Their p-values, related targets and main functions were listed in Table 3. After acquiring 33 dermatitis-related targets of haptens and 12 dermatitis-related pathways, we obtained 52 target-pathway pairs (Table 3), which were applied to build the network system.

12 dermatitis-related pathways referred to a wide variety of biological processes. These 12 pathways could be classified into 4 categories: 2 pathways about metabolism, 5 pathways in the environmental information processes, 2 pathways in the cellular processes and 3 pathways in organismal systems.

The 2 pathways about metabolism were “metabolism of xenobiotics by cytochrome P450” and “steroid hormone biosynthesis”. “Metabolism of xenobiotics by cytochrome P450” refers to the biodegradation and metabolism of xenobiotics, while “steroid hormone biosynthesis” refers to the lipid metabolism.

The 5 pathways in the environmental information processes were “NF-kappa B signaling pathway”, “TNF signaling pathway”, “HIF-1 signaling pathway”,

“sphingolipid signaling pathway” and “PI3K-Akt signaling pathway”. “NF-kappa B signaling pathway”, induced by TNF-alpha, IL-1 or byproducts of bacterial and viral infections, affects nuclear factor-kappa B (NF-kappa B) which is a family of transcription factors that regulate genes related to immunity, inflammation and cell survival. “TNF signaling pathway” regulates tumor necrosis factor (TNF), which is a vital cytokine that can induce a variety of intracellular signaling pathways. “HIF-1 signaling pathway” regulates hypoxia-inducible factor 1 (HIF-1), which is a transcription factor and a major regulator of oxygen homeostasis. “Sphingolipid signaling pathway” regulates sphingomyelin and its metabolic products, which have the second messenger function in many cell signaling pathways. “PI3K-Akt signaling pathway”, activated by various types of cell stimulation or toxic damage, regulates basic cellular functions such as transcription, translation, proliferation, growth and survival.

The 2 pathways in the cellular processes were “cell adhesion molecules” and “apoptosis”. “Cell adhesion molecules” plays a key role in a series of biological processes including hemostasis, immune response, inflammation, embryogenesis and nerve tissue development. “Apoptosis” is a genetically programmed process which is controlled by many interrelating processes and eliminates damaged or redundant cells by activating caspases.

The 3 pathways in organismal systems were “T cell receptor signaling pathway”, “Toll-like receptor signaling pathway” and “bile secretion”. “T cell receptor signaling

pathway”, the crux to the effective responses of the immune system, requires the involvement of T cell receptors (TCRs) and costimulatory molecules. “Toll-like receptor signaling pathway” regulates Toll-like receptors (TLRs) which recognize structurally conserved molecules derived from microbes and activate immune cell responses. “Bile secretion” refers to the secretion of bile, which can not only digest and absorb fats and fat-soluble vitamins in the small intestine, but also eliminate excess cholesterol and many wastes, bilirubin, drugs and toxic compounds.

According to the p-values of 12 dermatitis-related pathways, we discovered that “NF-kappa B signaling pathway”, “T cell receptor signaling pathway”, “Toll-like receptor signaling pathway” and “TNF signaling pathway” may be the primary pathways in the process of skin sensitization and dermatitis induction, which was consistent with previous studies (Martin, 2012; Peiser et al., 2012; Ryan et al., 2007).

3.5. Evaluation results of the reaction-target relationship

In allusion to each of the 33 dermatitis-related targets of haptens, we calculated 4 *Score* values for MA, SB, S_N2 and Ac reactions, and subsequently obtained 132 *Score* values. The calculation processes and calculation results of 132 *Score* values were enumerated in Table S5. After that, we sought out 15 targets which were highly correlated with reactions, and obtained 19 high-correlated reaction-target pairs (*Score* values < 0.05) (Table 4, Table S2, Table S5).

The 19 high-correlated reaction-target pairs indicated that there was a certain relationship between protein binding reactions and targets of skin sensitization. As for

haptens with MA reactions, the ABCB1, ABCG2, CYP1A1, DPP4, F3 and NFE2L2 may be primary targets. As for haptens with SB reactions, the DPP4, PTPRC, TLR9 and TNF may be primary targets. As for haptens with S_N2 reactions, the ABCG2, EPHX1, NFE2L2, SELP and TLR2 may be primary targets. As for haptens with Ac reactions, the ACE, HSD11B1, NFE2L2 and ST6GAL1 may be primary targets. Although the relationship between protein binding reactions and skin targets has been confirmed to be existing, there has been no literature to thoroughly study the specific relationship to date. Therefore, we employed Fisher's exact test to explore the reaction-target relationship, so as to provide convenience for understanding the mechanism of skin sensitization.

3.6. The network system

The reaction-substructure-compound-target-pathway network system in this study consisted of six sets of pairs. They were 121 compound-substructure pairs (Table S2), 12 substructure-reaction pairs (Table 1, Table S2), 154 compound-reaction pairs (Table S1, Table S4), 523 compound-target pairs (Table S2, Table S4), 52 target-pathway pairs (Table 3) and 19 reaction-target pairs (Table 4, Table S2). The schematic diagram of network system charted with Cytoscape was shown in Fig. 4.

The network system could be applied for the following purposes:

(1) As for new haptens with known structures, we could predict their protein binding reactions *via* the compound-substructure-reaction subnetwork, and predict

their skin targets and pathways via the compound-target-pathway subnetwork.

(2) As for new haptens with known protein binding reactions, their targets could be conveniently conjectured in virtue of the reaction-target relationship.

(3) The network system provided us with a good approach to comprehend the mechanism of allergy and immunity. In this study, the acquired dermatitis-related pathways referred to many kinds of biological processes, including metabolism, environmental information processes, cellular processes and the processes in organismal systems. This demonstrated that the process of allergy and immunity in the human body was intricate and study-worthy.

4. Discussion

4.1. Analysis of dermatitis-related targets

In this study, we discovered 33 dermatitis-related targets (Table 2). There were various high-frequency target families which were highly correlated with skin sensitization, such as CYP molecules, cluster of differentiation (CD) molecules, cytokines/chemokines, etc.

CYP molecules are heme-containing enzymes that catalyze the oxidative conversion of a series of predominantly lipophilic molecules into species that are usually more reactive and hydrophilic (Peiser et al., 2012). The CYP-dependent metabolism is the prerequisite for skin sensitization. Baron et al. (2001) deemed that dexamethasone induced the expression of CYP3A4 and benzoanthracene induced the increase of CYP1A1 level. In this study, CYP1A1 and CYP3A4 were deemed to be

involved in skin sensitization, and participated in the “metabolism of xenobiotics by cytochrome P450” pathway. CYP1A1 was regarded as one of the primary targets of the haptens with MA reactions. 96.8% of haptens in this study reacted with CYP3A4, which demonstrated the importance of CYP3A4 in KCs and antigen presenting cells.

CD molecules play a key role in the development of effective immune responses during the activation of lymphocytes. In the process of allergy and immunity, they generally act as effector T cells, which take effect throughout the sensitization phase and the elicitation phase (Honda et al., 2013). In this study, ABCB1 (CD antigen CD243), ABCG2 (CD antigen CD338), ACE (CD antigen CD143), ADAM17 (CD antigen CD156b), CCR4 (CD antigen CD194), CXCR1 (CD antigen CD181), DPP4 (CD antigen CD26), F3 (CD antigen CD142), ICAM1 (CD antigen CD54), ITGAL (CD antigen CD11a), KDR (CD antigen CD11a), NT5E (CD antigen CD73), PTPRC (CD antigen CD45), SELP (CD antigen CD62P), TLR2 (CD antigen CD282), TLR9 (CD antigen CD289) and VCAM1 (CD antigen CD106) were deemed to be involved in skin sensitization. They were mentioned in previous studies (Cottrez et al., 2015; Martin, 2012; Peiser et al., 2012; Ryan et al., 2007).

Cytokines and chemokines are vital in the migration and maturation of langerhans cells following skin sensitization (Ryan et al., 2007), including CC subfamily, CXC subfamily, IL family, TNF family, etc. Cottrez et al. (2015) and Lambrechts et al. (2010) reported multiple cytokines and chemokines modulated in sensitized skin. In this study, CCR4, CXCR1, IL2 and TNF were deemed to be

involved in skin sensitization, which was proven in previous studies (Cottrez et al., 2015; Lambrechts et al., 2010).

In addition, Lambrechts et al. (2010) regarded JUN, NFKB1 and NOS2 as gene markers in DCs unravel pieces of the skin sensitization puzzle. Cottrez et al. (2015) also reported BCL2, EPHX1, ICAM1, JUN and NOS3 as genes specifically modulated in sensitized skin. As a result, plenty of targets reacted with haptens and participated in allergic process. The haptens acted on various skin targets and accordingly induced pathways, so as to bring about ACD.

4.2. Analysis of dermatitis-related pathways

In this study, we discovered 12 dermatitis-related pathways (Table 4), which were subject to literature verification. Bønnelykke et al. (2015) declared that the process of allergic sensitization referred to multiple pathways, such as “Fc epsilon RI signaling pathway”, “T cell receptor signaling pathway”, “cytokine-cytokine receptor interaction” and “TGF-beta signaling pathway”. Lambrechts et al. (2010) also reported 15 pathways, such as “xenobiotic metabolism pathway”, “dendritic cell maturation”.

The skin sensitization mechanism refers to various pathways about metabolism, including xenobiotics metabolism, lipid metabolism, etc. “Metabolism of xenobiotics by cytochrome P450” and “steroid hormone biosynthesis”, which frequently occur in various biological events, are likely to participate in the process of allergy and immunity. In addition, the skin sensitization mechanism has been confirmed to be

associated with “T cell receptor signaling pathway” and “Toll-like receptor signaling pathway” (Bønnelykke et al., 2015; Honda et al., 2013). Their related pathways, namely “NF-kappa B signaling pathway” and “PI3K-Akt signaling pathway”, are deemed to be dermatitis-related. “HIF-1 signaling pathway” and “sphingolipid signaling pathway”, which are highly correlated with “PI3K-Akt signaling pathway”, are also deemed to be dermatitis-related. “TNF signaling pathway” is involved because TNFs induce multiple intracellular signaling pathways, including apoptosis and cell survival, as well as inflammation and immunity. “Apoptosis” and “cell adhesion molecules”, the indispensable pathways in the cellular processes, are also deemed to be involved in skin sensitization.

Therefore, among 12 pathways in this study, 11 pathways have been proven to be involved in skin sensitization. But “bile secretion” has not been proven to be dermatitis-related. Although no literature has reported a direct relationship between “bile secretion” and skin sensitization, studies have shown that the skin itching may be due to cholestasis (Bergasa, 2018), which indicated that “bile secretion” and skin sensitization had certain commonality in mechanism.

4.3. The application of the compound-substructure-reaction subnetwork

Studies have discovered that the haptens mainly acted as the following 5 roles: MA acceptors, Schiff base formers, S_N1/S_N2 electrophiles, S_NAr electrophiles and acylating agents (Aptula et al., 2005). According to the 12 key substructures of haptens identified in this study, we drew the following 5 conclusions: (1) The haptens

with MA reactions usually possess a double bond with electron-withdrawing substituent carbonyl group. (2) The haptens with SB reactions are usually reactive carbonyl compounds. (3) The haptens with S_N2 reactions usually possess leaving groups such as halogens and sulfonyl. (4) The haptens with Ac reactions usually possess the carbonyl and leaving groups. (5) The haptens with S_NAr reactions usually possess aromatic rings with leaving substituents. These conditions were consistent with the conditions under which protein binding reactions occurred.

There are three approaches to determine whether new haptens possess these key substructures. (1) Some endpoint-specific profiles, such as the “protein binding alerts for skin sensitization by OASIS v1.4” wizard of OECD QSAR Toolbox, can not only seek out the key substructures of haptens, but also identify the reaction mechanisms between haptens and proteins. (2) The SMILES codes of haptens could be converted into molecular structures using ChemDraw, then the structures can be observed directly. (3) If the reaction types of new haptens are identified, the key substructures can be obtained from reaction types based on the chemical reaction mechanism. We took safranal as an example and explained these three ways in detail in the Section 2 of the Supporting Information.

Therefore, we could predict the protein binding reactions of haptens according to their key substructures. As long as we knew the structures of new haptens, we could ascertain their key substructures and subsequently obtain their protein binding reactions. The clearness of protein binding reactions was beneficial to determining

experimental conditions and acquiring accurate prediction results of compound toxicology.

4.4. The application of the compound-target-pathway subnetwork

Studies have reported plenty of dermatitis-related targets and dermatitis-related pathways. But the question was how to determine the specific targets and pathways for new haptens. In this study, we had two suggestions for solving this problem: (1) As for new haptens, scholars can collect the known targets from databases and predict the potential targets *via* our bSDTNBI method. After the targets of haptens are obtained, the pathways of haptens can be enriched and acquired. (2) When scholars want to quickly predict targets for numerous compounds, the substructure-compound-target subnetwork and the reaction-target relationship are both serviceable to guide the determination of dermatitis-related targets. As long as we knew the substructures or the protein binding reactions of haptens, we could conjecture their dermatitis-related targets and subsequently enrich their dermatitis-related pathways.

In addition, the dermatitis-related targets and dermatitis-related pathways helped us comprehend the mechanism of skin sensitization. In this study, a sufficient number of haptens brought a sufficient number of skin targets, so as to clarify the pathways of skin sensitization. The skin sensitization mechanism could be thoroughly investigated with the help of various skin targets and pathways.

4.5. The comprehension of skin sensitization mechanism

As is known, the process of ACD is divided into two phases, namely the sensitization phase and the elicitation phase (Honda et al., 2013; Kimber et al., 2002). In the sensitization phase, haptens activate KCs and mast cells by means of innate immune systems. The activated DCs capture antigens, begin to mature, and migrate to the draining lymph nodes, and then the migrated DCs present antigen to naive T cells in draining lymph nodes. In the elicitation phase, KCs and mast cells are activated and produce a variety of chemical mediators when they are re-exposed to haptens. Infiltrated antigen-specific effector T cells are activated and pro-inflammatory cytokines and chemokines are produced. After that, effector T cells and regulatory T cells infiltrate the sites of inflammation and exert an inhibitory effect (Honda et al., 2013).

The network system in this study was applied to comprehend skin sensitization mechanism, not only at the level of protein binding reactions, but also at the level of targets and pathways. On one hand, the protein binding reactions of new haptens were speculated in virtue of the identification of key substructures. We determined 12 characteristic substructures and utilized them as bridges to connect haptens with protein binding reactions. So we could understand how haptens reacted with skin proteins from the view of chemical reactions. On the other hand, the targets and pathways were speculated in virtue of target prediction and pathway enrichment. We acquired 33 skin targets which acted as functional molecules to induce pathways, and accordingly enriched 12 pathways about skin sensitization. So we could understand

which targets the haptens interacted with and which pathways they induced, in order to understand the initiation of ACD.

We deemed that the haptens with same protein binding reaction tended to react with similar skin targets, and subsequently induce similar pathways. In other words, there should be a relationship between the protein binding reactions and the skin targets of haptens. Therefore, the reaction-target relationship explored *via* Fisher's exact test in this study was significant and convenient for understanding the mechanism of skin sensitization.

4.6. *Characteristics of our network system*

The detailed usage of our network system was provided in the Section 2 of the Supporting Information. The characteristics of our network system were concluded from the following three aspects:

(1) As for the prediction ability, the network system was applied to predict the allergic mechanism of haptens. As described in the Introduction section, there were four key mechanistic events of skin sensitization: the covalent binding of chemicals to endogenous proteins in the skin (molecular initiating event), the keratinocyte activation, the dendritic cell activation, and the proliferation of antigen-specific T cells. We developed the compound-substructure-reaction subnetwork to study the molecular initiating event, and developed the compound-target-pathway subnetwork to study the other three events. In addition, the relationships between the protein binding reactions of haptens and the skin targets were explored for the first time.

(2) As for the applicability domain, the network system was aimed at haptens. As for the prohaptens or prehaptens, the prediction results of the network system may be limited. If the network system was applied to predict the targets of prohaptens or prehaptens, the potential targets predicted would be likely to be some metabolic enzymes or oxidants. As for skin irritants, the network system is unavailable. The molecular initiating event of skin irritation is not necessarily the covalent binding of compounds to endogenous proteins in the skin. Because of the difference in the key events between skin sensitization and skin irritation, the prediction results of skin irritants might be inaccurate.

(3) As for the strategy for developing system, the strategy used in this study could be applied in many fields. The strategy for developing the compound-substructure-reaction subnetwork can be used to predict the protein binding reactions of all compounds, not just haptens. The bSDTNBI method for predicting the targets of compounds can be applied to study the chemical mechanism of action in many fields. Peng et al. (2019) used bSDTNBI method to study the mechanisms of drug-induced liver injury, while Wang et al. (2018) investigated the molecular mechanisms of herbal formula Tian-Ma-Gou-Teng-Yin for the treatment of Alzheimer's disease. Wang et al. (2017) studied the molecular mechanisms of *Polygonum multiflorum* Thunb-induced liver injury. These application examples could be helpful for researchers to build upon their work.

4.7. Comparison with other studies

Because an increasing number of people suffer from ACD, the study of skin sensitization mechanism has attracted a lot of attentions. Due to the drawbacks of traditional animal testing methods, many research groups have been devoted to develop *in silico* models.

On one hand, some *in silico* models were developed to identify the skin sensitization hazard. Luechtefeld et al. (2018) applied machine learning methods based on toxicological big data, and developed two kinds of RASAR (read-across structure activity relationship) models. The “Simple” RASAR model can be used to predict the risk from chemical analogues with known hazard data, while the “Data Fusion” RASAR model can be used to predict all available property data. Sanvido et al. (2018) proposed a quantitative risk assessment approach to address the sensitizing potential of plant protection products. This approach combines an agricultural exposure model and a methodology to derive a substance-specific threshold for skin sensitizers.

On the other hand, some *in silico* models were developed to study the biological processes of skin sensitization. Schwöbel and Klamt (2019) employed a new methodology called COSMOplex and constructed a skin mechanistic model, which captured the local interactions in the different parts of the skin on a molecular level. The model can cover changes of skin composition and permeation pathways in healthy and damaged skin. Our network system was developed for predicting the protein binding reactions, skin targets, and pathways related to ACD, and

subsequently identifying the associations among these different parts.

4.8. Future directions

We have to admit that our network system had the following limitations: (1) The network system was aimed at haptens, so it had poor predictive ability for prohaptens or prehaptens. (2) The network system cannot be applied to predict the skin sensitizing potential and potency of compounds.

In order to overcome these limitations, more researches need to be carried out and a comprehensive prediction network system needs to be developed in the future. (1) Different from haptens, prohaptens or prehaptens require metabolic activation or oxidation to react with skin proteins. So the targets of prohaptens or prehaptens are likely to be metabolic enzymes or oxidants. From this point of view, we could develop a new system, which is aimed at prohaptens and prehaptens. Then, the system in this study and the new system could be combined to distinguish haptens from prohaptens and prehaptens. (2) Our previous mechanism-based multi-classification models developed *via* machine learning methods (Di et al., 2019) can be put into the comprehensive system to identify the skin sensitizing potential and potency of compounds. Other non-animal assays (Di et al., 2019) can also be put to study the skin sensitization mechanism thoroughly.

5. Conclusions

In this study, a reaction-substructure-compound-target-pathway network system was proposed to predict the allergic mechanism of haptens. The

compound-substructure-reaction subnetwork was built to predict the protein binding reactions of new haptens, while the compound-target-pathway subnetwork was built to predict the skin targets and pathways of new haptens. The reaction-target relationship was explored to guide the determination of the reaction mechanisms of haptens. The network system can be applied not only to provide new ideas for mechanism research and chemical screening, but also to contribute to evaluating chemical safety and optimizing chemical structures. In addition, the network system can be combined with other models to study the hazard identification and biological processes of skin sensitization, which is helpful for understanding the skin immunity and resisting allergic contact dermatitis.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

We gratefully acknowledged the financial supports from the National Key Research and Development Program of China (Grant 2016YFA0502304), the National Natural Science Foundation of China (Grants 81273438 and 81673356), and the Shanghai Sailing Program (Grant 19YF1412700).

References

- Aptula, A.O., Patlewicz, G., Roberts, D.W., 2005. Skin sensitization: Reaction mechanistic applicability domains for structure-activity relationships. *Chem. Res. Toxicol.* 18, 1420-1426. <https://doi.org/10.1021/tx050075m>
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.L., 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, 115-119. <https://doi.org/10.1093/nar/gkh131>
- Baron, J.M., Höller, D., Schiffer, R., Frankenberg, S., Neis, M., Merk, H.F., Jugert, F.K., 2001. Expression of multiple cytochrome P450 enzymes and multidrug resistance-associated transport proteins in human skin keratinocytes. *J. Invest. Dermatol.* 116, 541-548. <https://doi.org/10.1046/j.1523-1747.2001.01298.x>
- Bergasa, N.V., 2018. The pruritus of cholestasis: From bile acids to opiate agonists: Relevant after all these years. *Med. Hypotheses* 110, 86-89. <https://doi.org/10.1016/j.mehy.2017.11.002>
- Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., Wiswedel, B., 2007. KNIME: The Konstanz Information Miner, in: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (Eds.), *Studies in classification, data analysis, and knowledge organization*. Springer, Berlin, pp. 319-326. https://doi.org/10.1007/978-3-540-78246-9_38

- Bønnelykke, K., Sparks, R., Waage, J., Milner, J.D., 2015. Genetics of allergy and allergic sensitization: common variants, rare mutations. *Curr. Opin. Immunol.* 36, 115-126. <https://doi.org/10.1016/j.coi.2015.08.002>
- Clippinger, A.J., Allen, D., Behrsing, H., Bérubé, K.A., Bolger, M.B., Casey, W., DeLorme, M., Gaça, M., Gehen, S.C., Glover, K., Hayden, P., Hinderliter, P., Hotchkiss, J.A., Iskandar, A., Keyser, B., Luettich, K., Ma-Hock, L., Maione, A.G., Makena, P., Melbourne, J., Milchak, L., Ng, S.P., Paini, A., Page, K., Patlewicz, G., Prieto, P., Raabe, H., Reinke, E.N., Roper, C., Rose, J., Sharma, M., Spoo, W., Thorne, P.S., Wilson, D.M., Jarabek, A.M., 2018. Pathway-based predictive approaches for non-animal assessment of acute inhalation toxicity. *Toxicol. in Vitro* 52, 131-145. <https://doi.org/10.1016/j.tiv.2018.06.009>
- Cottrez, F., Boitel, E., Auriault, C., Aeby, P., Groux, H., 2015. Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. *Toxicol. In Vitro* 29, 787-802. <https://doi.org/10.1016/j.tiv.2015.02.012>
- Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., McMorran, R., Wieggers, J., Wieggers, T.C., Mattingly, C.J., 2018. The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res.* 47, 948-954. <https://doi.org/10.1093/nar/gky868>
- Dearden, J.C., Hewitt, M., Roberts, D.W., Enoch, S.J., Rowe, P.H., Przybylak,

- K.R., Vaughan-Williams, G.D., Smith, M.L., Pillai, G.G., Katritzky, A.R., 2015. Mechanism-based QSAR modeling of skin sensitization. *Chem. Res. Toxicol.* 28, 1975-1986. <https://doi.org/10.1021/acs.chemrestox.5b00197>
- Di, P., Yin, Y., Jiang, C., Cai, Y., Li, W., Tang, Y., Liu, G., 2019. Prediction of the skin sensitising potential and potency of compounds via mechanism-based binary and ternary classification models. *Toxicol. in Vitro* 59, 204-214. <https://doi.org/10.1016/j.tiv.2019.01.004>
- Diderich, R., 2010. Tools for category formation and read-across overview of the OECD (Q) SAR Application Toolbox, in: M.T.D. Cronin, J.C. Madden (Eds.), *In Silico Toxicology: Principles and Applications*, The Royal Society of Chemistry Cambridge, UK, pp. 385-407. <https://doi.org/10.1039/9781849732093-00385>
- Ezendam, J., Braakhuis, H.M., Vandebriel, R.J., 2016. State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies. *Arch. Toxicol.* 90, 2861-2883. <https://doi.org/10.1007/s00204-016-1842-4>
- Ferrari, T., Cattaneo, D., Gini, G., Bakhtyari, N.G., Manganaro, A., Benfenati, E., 2013. Automatic knowledge extraction from chemical structures: The case of mutagenicity prediction. *SAR QSAR Environ. Res.* 24, 365-649. <https://doi.org/10.1080/1062936X.2013.773376>
- Fisher, R.A., 1922. On the interpretation of χ^2 from contingency tables, and the calculation of P. *J. R. Stat. Soc.* 85, 87-94. <https://doi.org/10.2307/2340521>

- Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., Overington, J.P., 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, 1100-1107. <https://doi.org/10.1093/nar/gkr777>
- Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2005. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis* 16, 157-202. <https://doi.org/10.1097/01206501-200512000-00002>
- Goebel, C., Aeby, P., Ade, N., Alépée, N., Aptula, A., Araki, D., Dufour, E., Gilmour, N., Hibatallah, J., Keller, D., Kern, P., Kirst, A., Marrec-Fairley, M., Maxwell, G., Rowland, J., Safford, B., Schellauf, F., Schepky, A., Seaman, C., Teichert, T., Tessier, N., Teissier, S., Weltzien, H.U., Winkler, P., Scheel, J., 2012. Guiding principles for the implementation of non-animal safety assessment approaches for cosmetics: skin sensitisation. *Regul. Toxicol. Pharmacol.* 63, 40-52. <https://doi.org/10.1016/j.yrtph.2012.02.007>
- Honda, T., Egawa, G., Grabbe, S., Kabashima, K., 2013. Update of immune events in the murine contact hypersensitivity model: toward the understanding of allergic contact dermatitis. *J. Invest. Dermatol.* 133, 303-315. <https://doi.org/10.1038/jid.2012.284>
- Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature*

Protoc. 4, 44-57. <http://dx.doi.org/10.1038/nprot.2008.211>

Jensen, B.F., Vind, C., Padkjaer, S.B., Brockhoff, P.B., Refsgaard, H.H., 2007. In silico prediction of cytochrome P450 2D6 and 3A4 inhibition using Gaussian kernel weighted k-nearest neighbor and extended connectivity fingerprints, including structural fragment analysis of inhibitors versus noninhibitors. *J. Med. Chem.* 50, 501-511. <https://doi.org/10.1021/jm060333s>

Kern, P.S., Gerberick, G.F., Ryan, C.A., Kimber, I., Aptula, A., Basketter, D.A., 2010. Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. *Dermatitis* 21, 8-32. <https://doi.org/10.2310/6620.2009.09038>

Kimber, I., Basketter, D.A., Gerberick, G.F., Dearman, R.J., 2002. Allergic contact dermatitis. *Int. Immunopharmacol.* 2, 201-211. [https://doi.org/10.1016/S1567-5769\(01\)00173-4](https://doi.org/10.1016/S1567-5769(01)00173-4)

Kleinstreuer, N.C., Hoffmann, S., Alépée, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Göbel, C., Kern, P.S., Klaric, M., Kühnl, J., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Strickland, J., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches. *Crit. Rev. Toxicol.* 48, 359-374. <https://doi.org/10.1080/10408444.2018.1429386>

Lambrechts, N., Vanheel, H., Hooyberghs, J., De Boever, P., Witters, H., Van Den Heuvel, R., Van Tendeloo, V., Nelissen, I., Schoeters, G., 2010. Gene markers in

- dendritic cells unravel pieces of the skin sensitization puzzle. *Toxicol. Lett.* 196, 95-103. <https://doi.org/10.1016/j.toxlet.2010.04.003>
- Liu, T.Q., Lin, Y.M., Wen, X., Jorissen, R.N., Gilson, M.K., 2007. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35, 198-201. <https://doi.org/10.1093/nar/gkl999>
- Luechtefeld, T., Marsh, D., Rowlands, C., Hartung, T., 2018. Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. *Toxicol. Sci.* 165, 198-212. <https://doi.org/10.1093/toxsci/kfy152>
- Martin, S.F., 2012. Allergic contact dermatitis: xenoinflammation of the skin. *Curr. Opin. Immunol.* 24, 720-729. <https://doi.org/10.1016/j.coi.2012.08.003>
- Mehling, A., Eriksson, T., Eltze, T., Kolle, S., Ramirez, T., Teubner, W., van Ravenzwaay, B., Landsiedel, R., 2012. Non-animal test methods for predicting skin sensitization potentials. *Arch. Toxicol.* 86, 1273-1295. <https://doi.org/10.1007/s00204-012-0867-6>
- OECD, 2013. Guidance document on developing and assessing adverse outcome pathways. Series on testing and assessment No. 184. Paris: OECD Publishing. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2013\)6&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en)
- Pawson, A.J., Sharman, J.L., Benson, H.E., Faccenda, E., Alexander, S.P.H., Buneman, O.P., Davenport, A.P., McGrath, J.C., Peters, J.A., Southan, C.,

- Spedding, M., Yu, W., Harmar, A.J., NC-IUPHAR, 2014. The IUPHAR/BPS guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res.* 42, 1098-1106. <https://doi.org/10.1093/nar/gkt1143>
- Peiser, M., Tralau, T., Heidler, J., Api, A.M., Arts, J.H.E., Basketter, D.A., English, J., Diepgen, T.L., Fuhlbrigge, R.C., Gaspari, A.A., Johansen, J.D., Karlberg, A.T., Kimber, I., Lepoittevin, J.P., Liebsch, M., Maibach, H.I., Martin, S.F., Merk, H.F., Platzek, T., Rustemeyer, T., Schnuch, A., Vandebriel, R.J., White, I.R., Luch, A., 2012. Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. *Cell. Mol. Life Sci.* 69, 763-781. <https://doi.org/10.1007/s00018-011-0846-8>
- Peng, Y., Wu, Z., Yang, H., Cai, Y., Liu, G., Li, W., Tang, Y., 2019. Insights into mechanisms and severity of drug-induced liver injury via computational systems toxicology approach. *Toxicol. Lett.* 312, 22-33. <https://doi.org/10.1016/j.toxlet.2019.05.005>
- Roth, B.L., Lopez, E., Patel, S., Kroeze, W.K., 2000. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* 6, 252-262. <https://doi.org/10.1177/107385840000600408>
- Ryan, C.A., Kimber, I., Basketter, D.A., Pallardy, M., Gildea, L.A., Gerberick, G.F., 2007. Dendritic cells and skin sensitization: biological roles and uses in hazard identification. *Toxicol. Appl. Pharmacol.* 221, 384-394. <https://doi.org/10.1016/>

j.taap.2007.03.006

Sanvido, O., Schmid, K., FitzGerald, R.E., Roth, N., Wilks, M.F., Bormann, P., Hopf, N.B., 2018. A quantitative risk assessment for skin sensitizing plant protection products: Linking derived No-Effect levels (DNELs) with agricultural exposure models. *Regul. Toxicol. Pharm.* 98, 171-183. <https://doi.org/10.1016/j.yrtph.2018.07.022>

Schwöbel, J.A.H., Klamt, A., 2019. Mechanistic skin penetration model by the COSMOperm method: Routes of permeation, vehicle effects and skin variations in the healthy and compromised skin. *Comp. Toxicol.* 11, 50-64. <https://doi.org/10.1016/j.comtox.2019.02.004>

Sturla, S.J., Boobis, A.R., FitzGerald, R.E., Hoeng, J., Kavlock, R.J., Schirmer, K., Whelan, M., Wilks, M.F., Peitsch, M.C., 2014. Systems toxicology: from basic research to risk assessment. *Chem. Res. Toxicol.* 27, 314-329. <https://doi.org/10.1021/tx400410s>

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigartyo, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Pontén, F., 2015. Tissue-based map of the human proteome.

- Science 347, 1260419. <https://doi.org/10.1126/science.1260419>
- Urbisch, D., Honarvar, N., Kolle, S.N., Mehling, A., Ramirez, T., Teubner, W., Landsiedel, R., 2016. Peptide reactivity associated with skin sensitization: The QSAR Toolbox and TIMES compared to the DPRA. *Toxicol. in Vitro* 34, 194-203. <https://doi.org/10.1016/j.tiv.2016.04.005>
- Wang, Y.Y., Li, J., Wu, Z.R., Zhang, B., Yang, H.B., Wang, Q., Cai, Y.C., Liu, G.X., Li, W.H., Tang, Y., 2017. Insights into the molecular mechanisms of Polygonum multiflorum Thunb-induced liver injury: a computational systems toxicology approach. *Acta Pharmacol. Sin.* 38, 719-732. <https://doi.org/10.1038/aps.2016.147>
- Wang, T., Wu, Z., Sun, L., Li, W., Liu, G., Tang, Y., 2018. A computational systems pharmacology approach to investigate molecular mechanisms of herbal formula Tian-Ma-Gou-Teng-Yin for treatment of Alzheimer's disease. *Front. Pharmacol.* 9: 668. <https://doi.org/10.3389/fphar.2018.00668>
- Wu, Z., Cheng, F., Li, J., Li, W., Liu, G., Tang, Y., 2017. SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief. Bioinform.* 18, 333-347. <https://doi.org/10.1093/bib/bbw012>
- Wu, Z., Lu, W., Wu, D., Luo, A., Bian, H., Li, J., Li, W., Liu, G., Huang, J., Cheng, F., Tang, Y., 2016. In silico prediction of chemical mechanism of action via an improved network-based inference method. *Br. J. Pharmacol.* 173, 3372-3385.

<https://doi.org/10.1111/bph.13629>

Yu, W., Gwinn, M., Clyne, M., Yesupriya, A., Khoury, M.J., 2008. A navigator for human genome epidemiology. Nat. Genet. 40, 124-125. <https://doi.org/10.1038/ng0208-124>

Figure legends:

Fig. 1. The flow chart of research procedures taking safranal as an example.

Fig. 2. Schematic diagram of bSDTNBI method.

Fig. 3. The diagram of compound-substructure associations. 154 haptens (1-154) were arranged on the vertical axis while 12 key substructures (I-XII) were arranged on the horizontal axis. The associations were colored and the others were blank.

Fig. 4. The reaction-substructure-compound-target-pathway association network schematic diagram charted with Cytoscape. The nodes were linked if they were associated.

Fig 1

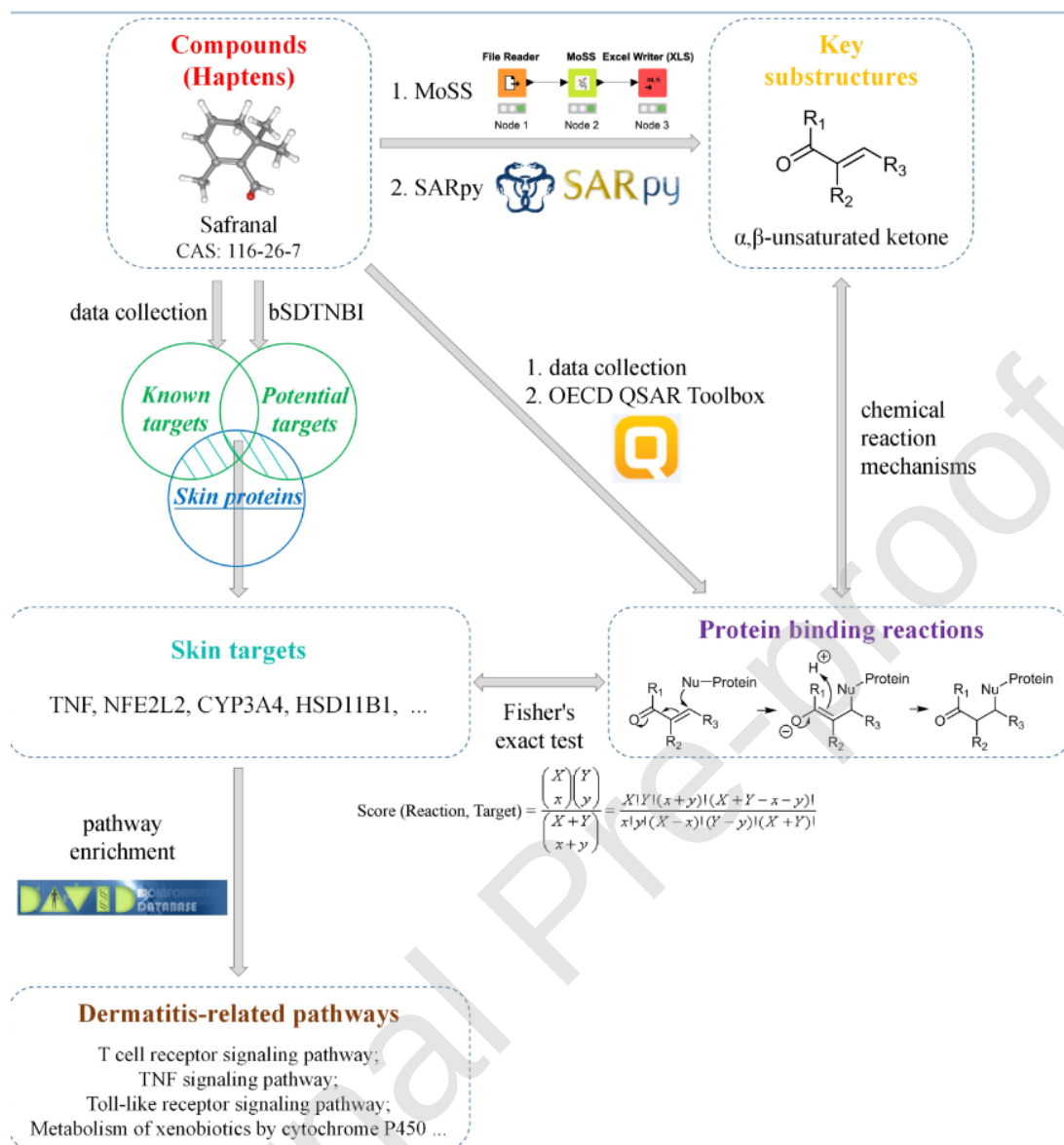


Fig 2

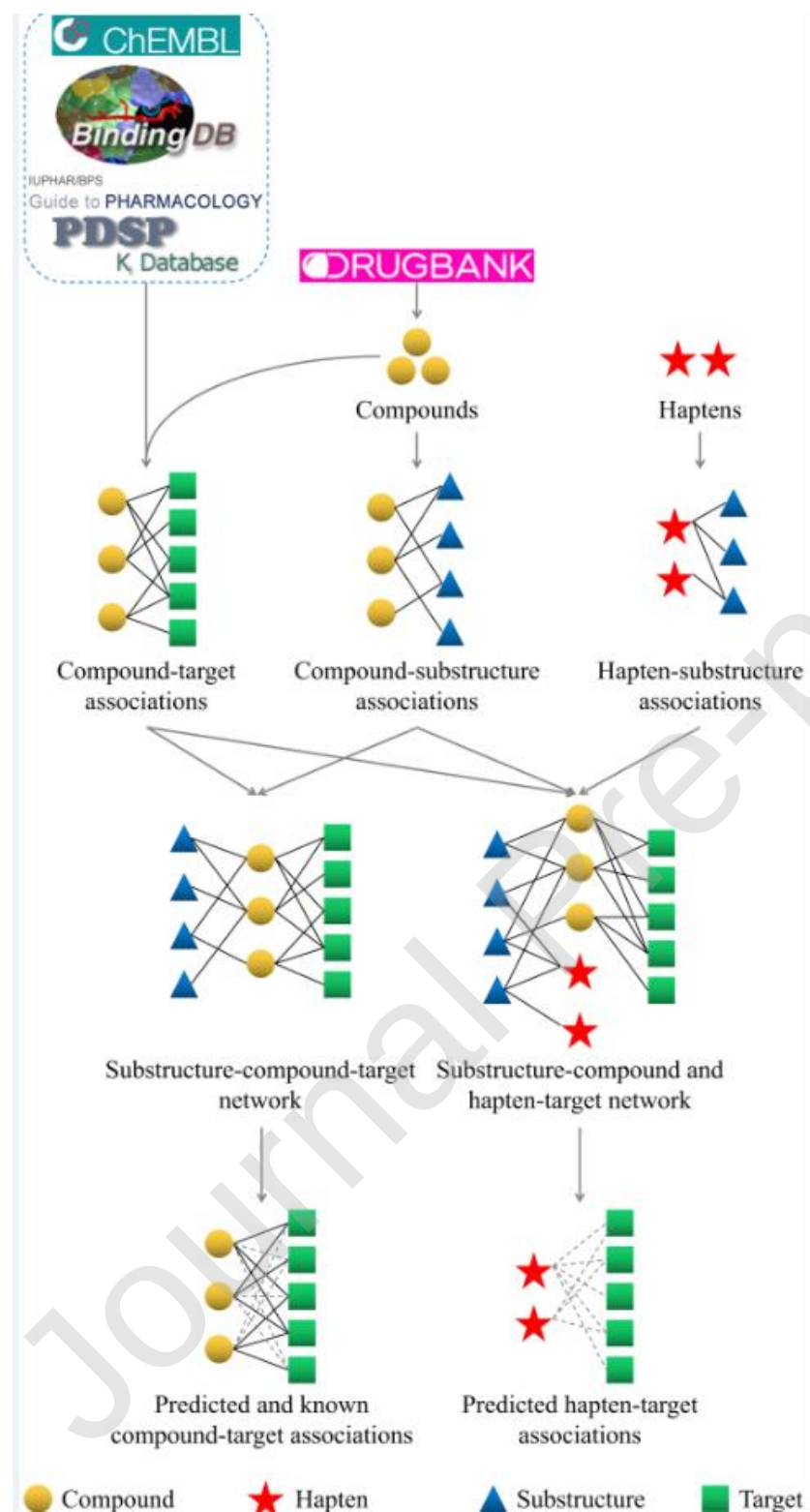


Fig 3

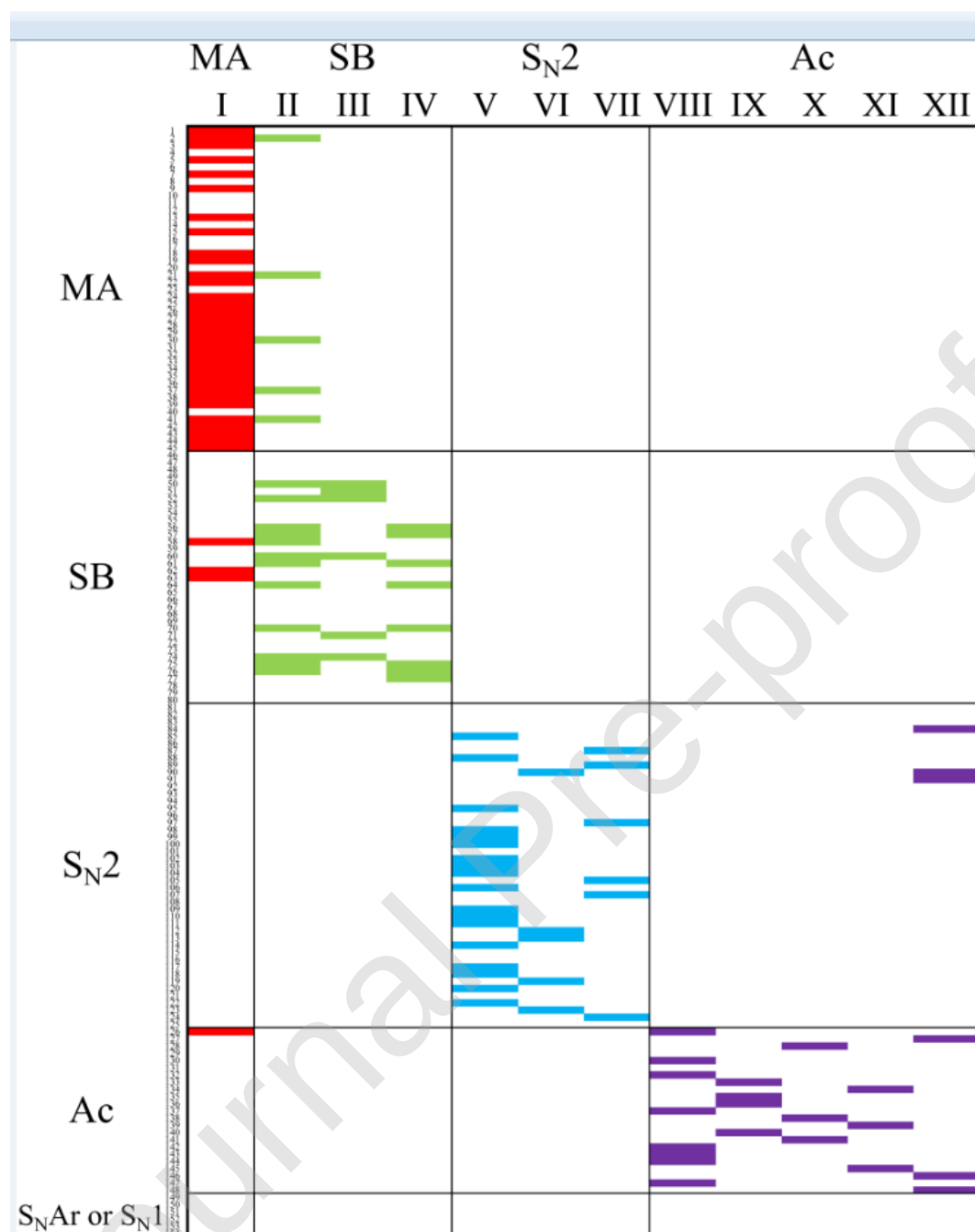


Fig 4

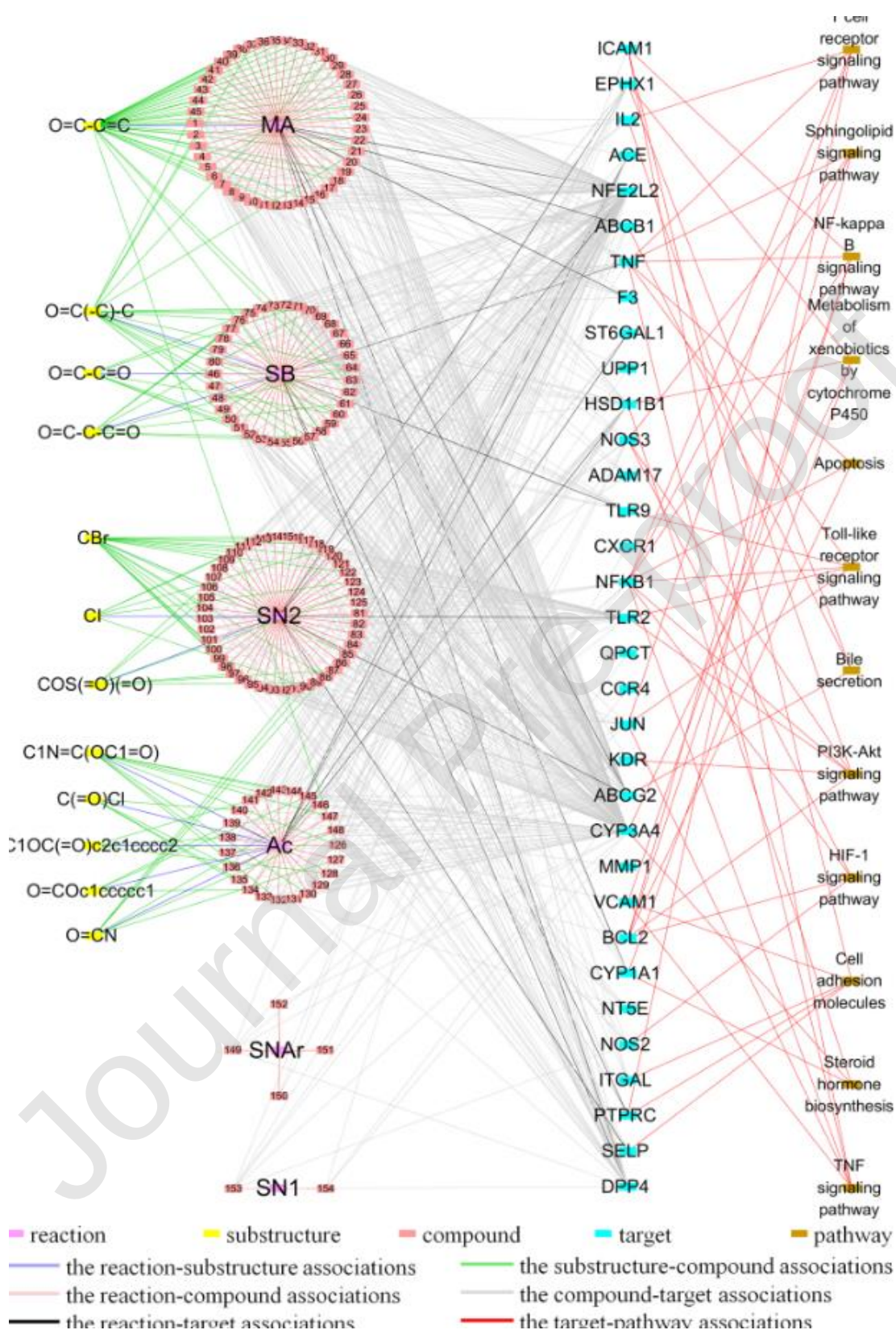
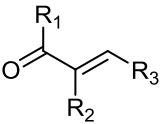
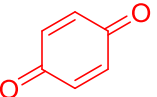
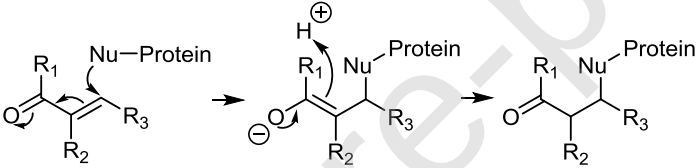
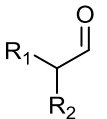
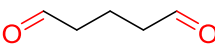
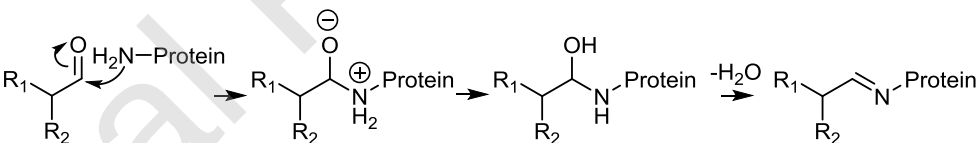
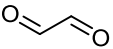
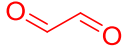
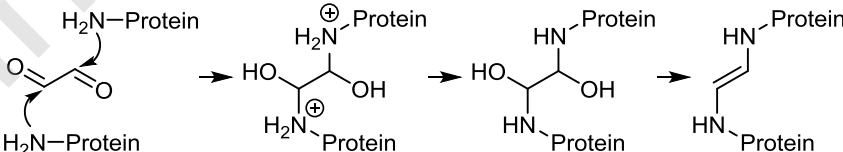
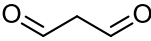
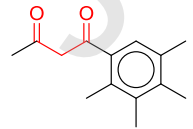
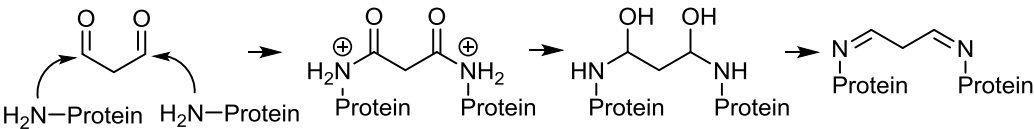
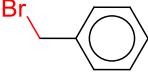
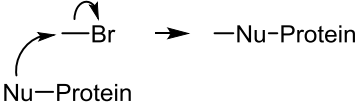
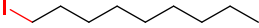
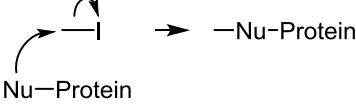
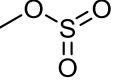
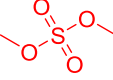
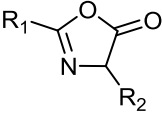
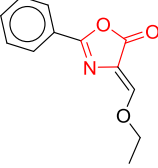
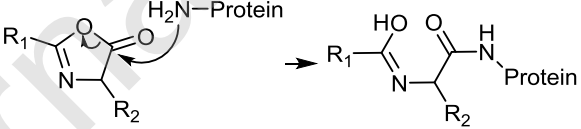
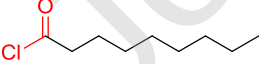
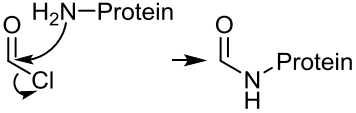
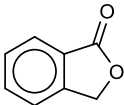
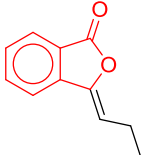
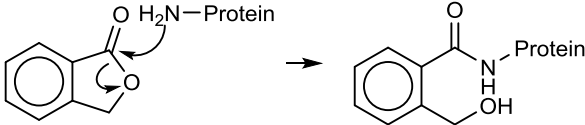
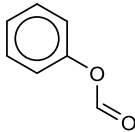
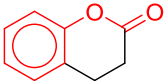
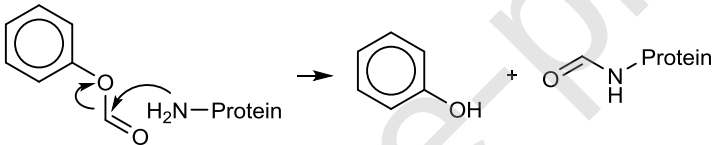
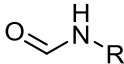
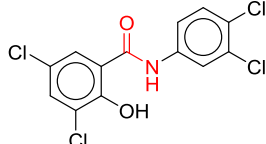
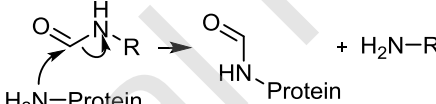


Table 1

List of the 12 key substructures and their representative haptens, protein binding reactions, reaction types, positive frequency (*PF*), negative frequency (*NF*), accuracy (*ACC*).

No.	Key substructures	Representative haptens	Protein binding reactions	Reaction types	Positive frequency (<i>PF</i>) (%)	Negative frequency (<i>NF</i>) (%)	Accuracy (<i>ACC</i>) (%)
I		 p-Benzoquinone (CAS:106-51-4) Extreme sensitizer		MA	73.33 (33/45)	3.67 (4/109)	89.19 (33/37)
II		 Glutaraldehyde (CAS:111-30-8) Strong sensitizer		SB	34.29 (12/35)	4.20 (5/119)	70.59 (12/17)
III		 Glyoxal (CAS:107-22-2) Moderate sensitizer		SB	17.14 (6/35)	0.00 (0/119)	100.00 (6/6)
IV		 1-(2',3',4',5'-Tetramethylphenyl)butane-1,3-dione Moderate sensitizer		SB	22.86 (8/35)	0.00 (0/119)	100.00 (8/8)

V		 Benzyl bromide (CAS:100-39-0) Strong sensitizer		S_N2	40.00 (18/45)	0.00 (0/109)	100.00 (18/18)
VI		 1-Iodononane (CAS:4282-42-2) Weak sensitizer		S_N2	11.11 (5/45)	0.00 (0/109)	100.00 (5/5)
VII		 Dimethyl sulfate (CAS:77-78-1) Strong sensitizer		S_N2	13.33 (6/45)	0.00 (0/109)	100.00 (6/6)
VIII		 4-Ethoxymethylene-2-phenyl-2-oxazolin-5-one (CAS:15646-46-5) Extreme sensitizer		Ac	34.78 (8/23)	0.00 (0/131)	100.00 (8/8)
IX		 Nonanoyl chloride (CAS:764-85-2) Moderate sensitizer		Ac	17.39 (4/23)	0.00 (0/131)	100.00 (4/4)

X				Ac	13.04 (3/23)	0.00 (0/131)	100.00 (3/3)
	3-Propylidenephthalide (CAS:17369-59-4) Moderate sensitizer						
XI				Ac	13.04 (3/23)	0.00 (0/131)	100.00 (3/3)
	3,4-Dihydrocoumarin (CAS:119-84-6) Moderate sensitizer						
XII				Ac	13.04 (3/23)	2.29 (3/131)	50.00 (3/6)
	Tetrachlorosalicylanilide (CAS:1154-59-2) Extreme sensitizer						

- (i) The substructures were regarded as key substructures if they simultaneously satisfied the following two criteria: PF > 0.1 and NF < 0.05.
- (ii) The associations between key substructures and haptens, as well as the associations between key substructures and protein binding reactions were listed in Table S2 of the supporting information.

Table 2

33 dermatitis-related targets and their main functions.

Target gene symbol (Target UniProt AC)	Target name	Target family	Main functions
ABCB1 (P08183)	ATP-binding cassette sub-family B member 1	ABC transporter superfamily	ATP binding
ABCG2 (Q9UNQ0)	ATP-binding cassette sub-family G member 2	ABC transporter superfamily	ATP binding
ACE (P12821)	Angiotensin-converting enzyme	Peptidase M2 family	Conversion of angiotensin
ADAM17 (P78536)	Disintegrin and metalloproteinase domain-containing protein 17	ADAM protein family	Conversion of tumor necrosis factor alpha
BCL2 (P10415)	B-cell CLL/lymphoma 2	Bcl-2 family	Apoptosis regulation
CCR4 (P51679)	C-C chemokine receptor type 4	G-protein coupled receptor 1 family	Antigen presentation; Signal transduction
CXCR1 (P25024)	C-X-C chemokine receptor type 1	G-protein coupled receptor 1 family	Antigen presentation; Signal transduction
CYP1A1 (P04798)	Cytochrome P450 1A1	Cytochrome P450 family	Xenobiotic metabolism; Antigen presentation
CYP3A4 (P08684)	Cytochrome P450 3A4	Cytochrome P450 family	Xenobiotic metabolism; Antigen presentation
DPP4 (P27487)	Dipeptidyl peptidase 4	Peptidase S9B family, DPPIV subfamily	Cell-mediated immune response; Signal transduction; Apoptosis regulation
EPHX1 (P07099)	Epoxide hydrolase 1	Peptidase S33 family	Compound metabolism
F3 (P13726)	Coagulation factor III	Tissue factor family	Signal transduction
HSD11B1 (P28845)	Corticosteroid 11-beta-dehydrogenase isozyme 1	Short-chain dehydrogenases/reductases (SDR) family	Signal transduction
ICAM1 (P05362)	Intercellular adhesion molecule 1	Immunoglobulin superfamily, ICAM family	Inflammation; Cell-mediated immune response; Signal transduction
IL2 (P60568)	Interleukin-2	IL-2 family	Antigen presentation; Cell-mediated immune response
ITGAL (P20701)	Integrin alpha-L	Integrin alpha chain family	Signal transduction
JUN (P05412)	Jun oncogene	BZIP family, Jun subfamily	Dermatological disorder; Cell-mediated immune response
KDR (P35968)	Vascular endothelial growth factor receptor 2	Protein kinase superfamily, Tyr protein kinase family, CSF-1/PDGF receptor subfamily	Regulation of endothelial cells
MMP1 (P03956)	Matrix metalloproteinase-1	Peptidase M10A family	Decomposition of extracellular matrix

NFE2L2 (Q16236)	Nuclear factor erythroid 2-related factor 2	BZIP family, CNC subfamily	Regulation of antioxidant protein expression
NFKB1 (P19838)	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	NF-kappa B/REL family	Dermatological disorder; Cell-mediated immune response; Signal transduction
NOS2 (P35228)	Nitric oxide synthase, inducible	NOS family	Antigen presentation
NOS3 (P29474)	Nitric oxide synthase, endothelial	NOS family	Antigen presentation
NT5E (P21589)	Ecto-5'-nucleotidase	5'-nucleotidase family	Conversion of adenosine monophosphate
PTPRC (P08575)	Receptor-type tyrosine-protein phosphatase C	Protein-tyrosine phosphatase family, Receptor class 1/6 subfamily	Signal transduction
QPCT (Q16769)	Glutamyl-peptide cyclotransferase	Glutamyl-peptide cyclotransferase family	Signal transduction
SELP (P16109)	P-selectin	Selectin/LECAM family	Inflammation; Cell adhesion
ST6GAL1 (P15907)	Beta-galactoside alpha-2,6-sialyltransferase 1	Glycosyltransferase 29 family	Signal transduction
TLR2 (O60603)	Toll-like receptor 2	Toll-like receptor family	Cell-mediated immune response; Signal transduction
TLR9 (Q9NR96)	Toll-like receptor 9	Toll-like receptor family	Cell-mediated immune response; Signal transduction
TNF (P01375)	Tumor necrosis factor	Tumor necrosis factor family	Dermatological disorder; Antigen presentation; Cell-mediated immune response; Signal transduction
UPP1 (Q16831)	Uridine phosphorylase 1	PNP/UDP phosphorylase family	Signal transduction
VCAM1 (P19320)	Vascular cell adhesion protein 1	VCAM family	Cell adhesion; Signal transduction

Table 3

12 dermatitis-related pathways and their p-values, related targets and main functions.

Pathway name	Entry	P-value	Related targets	Main functions	Category
NF-kappa B signaling pathway	hsa 04064	5.79E-04	VCAM1, ICAM1, TNF, BCL2, NFKB1	Signal transduction pathways; Regulation of NF-kappa B	(2)
T cell receptor signaling pathway	hsa 04660	9.79E-04	PTPRC, TNF, JUN, NFKB1, IL2	The pathway in immune system; Regulation of T cell receptors	(4)
Toll-like receptor signaling pathway	hsa 04620	1.22E-03	TNF, JUN, TLR2, NFKB1, TLR9	The pathway in immune system; Regulation of Toll-like receptors	(4)
TNF signaling pathway	hsa 04668	1.26E-03	VCAM1, ICAM1, TNF, JUN, NFKB1	Signal transduction pathways; Regulation of TNFs	(2)
Cell adhesion molecules	hsa 04514	3.55E-03	VCAM1, ICAM1, ITGAL, PTPRC, SELP	Signaling molecules and interaction	(3)
Metabolism of xenobiotics by cytochrome P450	hsa 00980	4.33E-03	CYP3A4, CYP1A1, HSD11B1, EPHX1	Xenobiotics biodegradation and metabolism; Regulation of CYP	(1)
HIF-1 signaling pathway	hsa 04066	8.92E-03	BCL2, NFKB1, NOS3, NOS2	Signal transduction pathways; Regulation of HIF-1	(2)
Sphingolipid signaling pathway	hsa 04071	1.63E-02	TNF, BCL2, NFKB1, NOS3	Signal transduction pathways; Regulation of sphingolipid	(2)
PI3K-Akt signaling pathway	hsa 04151	1.79E-02	BCL2, TLR2, NFKB1, NOS3, KDR, IL2	Signal transduction pathways; Regulation of PI3K-Akt	(2)
Steroid hormone biosynthesis	hsa 00140	2.78E-02	CYP3A4, CYP1A1, HSD11B1	Lipid metabolism; Biosynthesis and regulation of steroid hormones	(1)
Apoptosis	hsa 04210	3.14E-02	TNF, BCL2, NFKB1	The pathway about cell death; Programmed cell death	(3)
Bile secretion	hsa 04976	3.82E-02	EPHX1, ABCB1, ABCG2	The pathway in digestive system; Secretion of bile	(4)

- (i) The categories of pathways: (1) pathways about mechanism; (2) pathways in the environmental information processes; (3) pathways in the cellular processes; (4) pathways in organismal systems.
- (ii) The categories and entries of pathways were obtained from the website <https://www.kegg.jp/kegg/pathway.html>. These pathways were prefixed with “hsa” (Homo sapien).
- (iii) The pathways were considered to be critical when satisfying the following condition: p-value < 0.05.
- (iv) The abbreviations in this table were explained in ‘Abbreviations’ section.

Table 4

List of the 19 *Score* values which are less than 0.05.

Target gene symbol	Score values in reaction-target relationship			
	MA	SB	S _N 2	Ac
ABCB1	7.77E-04			
ABCG2	1.02E-07		6.05E-05	
ACE				4.07E-08
CYP1A1	2.40E-02			
DPP4	2.27E-02	8.25E-03		
EPHX1			7.58E-04	
F3	7.93E-03			
HSD11B1				2.36E-04
NFE2L2	1.07E-02		1.31E-02	4.26E-02
PTPRC		2.32E-03		
SELP			6.61E-03	
ST6GAL1				2.15E-02
TLR2			7.59E-11	
TLR9		3.61E-06		
TNF		3.86E-02		

(i) There were high correlations if the *Score* values < 0.05.

(ii) All the 132 *Score* values and their calculation processes were listed in Table S5 of the supporting information.

Supporting information

Section 1. The principle and application of bSDTNBI method

The balanced substructure-drug-target network-based inference (bSDTNBI) method developed by us was an integrated network and chemoinformatics tool for large-scale compound-target interactions (CTI) prediction and drug repositioning. The schematic diagram of bSDTNBI method was illustrated in Fig. 2. There were three steps to this method: (1) data collection and network construction; (2) prioritizing targets for known compounds; (3) predicting targets for new haptens.

(1) The bioactivity data for CTI pairs and the compound molecules were collected. As shown in Fig. 2, the bioactivity data for CTI pairs were collected from 4 databases, including ChEMBL Database (<https://www.ebi.ac.uk/chembl/>), Binding Database (<http://www.bindingdb.org/bind/index.jsp>), IUPHAR/BPS Guide to PHARMACOLOGY (<http://www.guidetopharmacology.org/>) and PDSP Ki Database (<https://pdsp.unc.edu/databases/kidb.php>). The compound molecules were extracted from the DrugBank database (<https://www.drugbank.ca/>). All the chemical structures were prepared with OpenBabel (http://openbabel.org/wiki/Main_Page), and the SMILES codes and molecular weights were obtained. After preparing the bioactivity data, we retained the items that met the following 5 criteria: (1) K_i , K_d , IC_{50} or $EC_{50} \leq 10 \mu M$; (2) the target protein could be described by a sole UniProt accession number; (3) the target protein was tagged as “reviewed” in the UniProt database (<https://www.uniprot.org/>); (4) the target protein was derived from Homo sapiens; (5) the molecular weight of compound ranged from 100-600 Daltons (Wu et al., 2016). Ultimately, the global CTI network was constructed with the prepared bioactivity data. In addition, the chemical substructures for molecules in each dataset were generated with the Fingerprint (KR), a kind of molecular fingerprint calculated by PaDEL-Descriptor (<http://www.yapcsoft.com/dd/padeldescriptor/>). After the data

preparation and chemical substructure generation, we obtained the known CTI bipartite networks and compound-substructure linkages, so as to develop a substructure-compound-target network.

(2) The substructure-compound-target network can be used to prioritize potential CTIs for known compounds in existing networks. The algorithms were explained in detail in our previous study (Wu et al., 2017) and briefly described here.

Denoting that $C = \{C_1, C_2, \dots, C_{N_C}\}$ is a set of N_C known compounds, $S = \{S_1, S_2, \dots, S_{N_S}\}$ is a set of N_S chemical substructures, and $T = \{T_1, T_2, \dots, T_{N_T}\}$ is a set of N_T targets. If C_i is linked with S_j , $M_{CS}(i,j) = 1$, otherwise, $M_{CS}(i,j) = 0$. If C_i is linked with T_j , $M_{CT}(i,j) = 1$, otherwise, $M_{CT}(i,j) = 0$. The substructure-compound-target network can be represented by an adjacency matrix A . A and W are two $(N_C + N_S + N_T)$ order square matrices. The A , W and the final resource matrix F were defined as follows:

$$A = \begin{bmatrix} 0 & M_{CS} & M_{CT} \\ M_{CS}^T & 0 & 0 \\ M_{CT}^T & 0 & 0 \end{bmatrix}$$

$$W(i, j) = \frac{A(i, j)}{\sum_{l=1}^{N_C+N_S+N_T} A(i, l)}$$

$$F = A \times W^k$$

where W^k is the transfer matrix and k is the number of resource-spreading processes. The value of $F(i, N_C + N_S + j)$ ($0 < i \leq N_C$, $0 < j \leq N_T$) is the score of C_i - T_j interaction.

(3) In the substructure-compound-(and hapten)-target network, new haptens can be regarded as special compounds without known targets. Let $H = \{H_1, H_2, \dots, H_{N_H}\}$ be a set of N_H new haptens. If H_i is linked with S_j , $M_{HS}(i,j) = 1$, otherwise, $M_{HS}(i,j) = 0$. The substructure-compound-(and hapten)-target network can be represented by an adjacency matrix A' . A' , B and W' are three $(N_H + N_C + N_S + N_T)$ order square

matrices. The A' , B , W' and the final resource matrix F' were defined as follows:

$$A' = \begin{bmatrix} 0 & 0 & M_{HS} & 0 \\ 0 & 0 & M_{CS} & M_{CT} \\ M_{HS}^T & M_{CS}^T & 0 & 0 \\ 0 & M_{CT}^T & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & M_{CS} & M_{CT} \\ 0 & M_{CS}^T & 0 & 0 \\ 0 & M_{CT}^T & 0 & 0 \end{bmatrix}$$

$$W' = \begin{cases} \frac{B(i, j)}{\sum_{l=1}^{N_H + N_C + N_S + N_T} B(i, l)} & \text{if } \sum_{l=1}^{N_H + N_C + N_S + N_T} B(i, l) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F' = A' \times W'^k$$

where W'^k is the transfer matrix and k is the number of resource-spreading processes. The value of $F'(i, N_H + N_C + N_S + j)$ ($0 < i \leq N_H$, $0 < j \leq N_T$) is the score of H_i-T_j interaction.

The performance of bSDTNBI method was evaluated by 10-fold cross validation and leave-one-out cross validation, which applied different methods to generate test sets and training sets (Wu et al., 2016; Wu et al., 2017). During 10-fold cross validation, 10% of CTIs were randomly separated from the substructure-compound-target network as the test set in turn. And the residual network, which contained the other 90% of CTIs and all compound-substructure associations, was applied as the training set. During leave-one-out cross validation, CTIs of each compound were separated from the substructure-compound-target network and used as the test set in turn. And the residual network, which contained the CTIs of the other compounds and all compound-substructures associations, was applied as the training set.

There were 4 parameters that needed to be tuned: k , α , β , γ . The parameter k stood for the number of resource-spreading processes. The parameter α was applied to

adjust the initial resource allocation of different node types. The parameter β was applied to adjust the weighted values of different edge types. The parameter γ was applied to adjust the influence of hub nodes. The performance of most models can be maximized when $\alpha = \beta = 0.1$, $\gamma = -0.5$, and $k = 2$. The performance of various models under different conditions was compared in our previous study (Wu et al., 2016). Under the optimal condition, the highest AUC value of Global-KR model in 10-fold cross validation was 0.961 ± 0.003 , while the highest AUC value of Global model in leave-one-out cross validation was 0.919 ± 0.144 . The prediction results of the models have been proven to be credible in several studies (Peng et al., 2019; Wang et al., 2017; Wang et al., 2018). Therefore, we applied bSDTNBI method to predict the targets of haptens in this study.

Section 2. The detailed usage of our network system

2.1. From haptens to key substructures

When the scholars intend to determine the key substructures of numerous haptens in batches, they can apply some endpoint-specific profilers, such as the “protein binding alerts for skin sensitization by OASIS v1.4” wizard of OECD QSAR Toolbox. After the SMILES codes of compounds were input into OECD QSAR Toolbox, the wizard can not only seek out the key substructures of compounds, but also identify the reaction mechanisms between compounds and proteins.

When the scholars intend to determine the key substructures of a few new haptens, they can convert their SMILES codes into molecular structures using ChemDraw and observe their structures directly. If the structures of haptens possess the 12 key substructures identified in this study (Table 1), they would be warned to be skin sensitizing. In addition, if the reaction types of new haptens have been identified, the key substructures can be obtained from reaction types based on the chemical reaction mechanism.

Taking safranal (CAS: 116-26-7) as an example, we can identify its key substructure in three ways. (1) After the SMILES code of safranal was input into OECD QSAR Toolbox, the profiler can not only determine that the key substructure of safranal was α,β -unsaturated ketone, but also judge that safranal acted as a MA acceptor. (2) After observing the structure of safranal by ChemDraw, we could easily find the presence of α,β -unsaturated ketone, and subsequently infer that safranal may undergo MA reaction. (3) Studies (Dearden et al., 2015) have confirmed that the reaction binding mechanism of safranal was MA reaction. According to the chemical reaction mechanism and the observation of structure, we could conjecture that the key substructure of safranal was α,β -unsaturated ketone.

2.2. From key substructures to protein binding reactions

Studies have discovered that the haptens mainly acted as the following 5 roles: MA acceptors, Schiff base formers, S_N1/S_N2 electrophiles, S_NAr electrophiles and acylating agents (Aptula et al., 2005). According to the 12 key substructures of haptens identified in this study (Table 1), we drew the following 5 conclusions: (1) If the haptens possess a double bond with electron-withdrawing substituent carbonyl group, they are supposed to undergo MA reactions. (2) If the haptens are reactive carbonyl compounds, they are supposed to undergo SB reactions. (3) If the haptens possess leaving groups such as halogens and sulfonyl, they are supposed to undergo S_N2 reactions. (4) If the haptens possess the carbonyl and leaving groups, they are supposed to undergo Ac reactions. (5) If the haptens possess aromatic rings with leaving substituents, they are supposed to undergo S_NAr reactions. These conditions were consistent with the conditions under which protein binding reactions occurred.

Therefore, we could predict the protein binding reactions of haptens according to their key substructures. As long as we knew the structures of new haptens, we could ascertain their key substructures and subsequently obtain their protein binding reactions. The clearness of protein binding reactions was beneficial to determining experimental conditions and acquiring accurate prediction results of compound toxicology.

2.3. From haptens to skin targets

When the scholars intend to obtain the skin targets of new haptens, they need to do three steps: collecting the known targets, predicting the potential targets, selecting the skin targets. Firstly, the known targets of haptens were collected from four databases (described in 2.3). Secondly, the potential targets were predicted *via* bSDTNBI method. After that, the targets of haptens, including known targets and potential targets, were obtained. Thirdly, the skin proteins were collected from three databases (described in 2.3), and were matched to the targets of haptens to obtain the skin targets of haptens.

2.4. The acquisition of pathways

Studies have reported plenty of pathways related to ACD. But the question was how to determine the specific pathways for new haptens. In this study, we had two suggestions for solving this problem: (1) As for new haptens, scholars can collect the known targets from databases and predict the potential targets *via* our bSDTNBI method. After the targets of haptens are obtained, the pathways of haptens can be enriched with DAVID v6.8. (2) When scholars want to quickly predict targets for numerous compounds, the substructure-compound-target subnetwork and the reaction-target relationship are both serviceable to guide the determination of dermatitis-related targets. As long as we knew the substructures or the protein binding reactions of haptens, we could conjecture their dermatitis-related targets and subsequently enrich their dermatitis-related pathways.

In addition, the dermatitis-related targets and dermatitis-related pathways helped us comprehend the mechanism of skin sensitization. In this study, a sufficient number of haptens brought a sufficient number of skin targets, so as to clarify the pathways of skin sensitization. The skin sensitization mechanism could be thoroughly investigated with the help of various skin targets and pathways.