

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 04-08--2014		2. REPORT TYPE Journal Article		3. DATES COVERED (From - To) July 2013 - May 2014	
4. TITLE AND SUBTITLE Identifying Potential Protein Targets for Toluene using Molecular Similarity Search, <i>in Silico</i> Docking and <i>in Vitro</i> Validation				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Yaroslav G. Chushak, Richard R Chapleau, Jeanette S Frey, Camilla A Mauzy and Jeffery M Gearhart				5d. PROJECT NUMBER 7757	
				5e. TASK NUMBER HD	
				5f. WORK UNIT NUMBER 05/H0D1	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Henry M Jackson Foundation for the Advancement of Military Medicine, 2729 R Street, Wright Patterson AFB, Ohio 45433, United States Molecular Bioeffects Branch, Bioeffects Division, Human Effectiveness Directorate, 711th Human Performance Wing, Air Force Research Laboratory (711 HPW/RHDJ), 2729 R Street, Wright Patterson AFB, Ohio 45433, United States				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Molecular Bioeffects Branch Bioeffects Division Human Effectiveness Directorate 711th Human Performance Wing Air Force Research Laboratory Wright-Patterson AFB OH 45433-5707				10. SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHDJ	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER	
12. DISTRIBUTION AVAILABILITY STATEMENT Distribution A: Approved for public release; distribution unlimited. (PA Case No 88ABW-2014-4574, 26 September 2014)					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The main goal of this study was to develop an approach for predicting protein targets for chemical toxins using a molecular similarity search of toxin-target information collected in the Toxin and Toxin-Target Database. The developed method was used to identify new targets for toluene which could predict potential cellular toxicity and to validate the approach with <i>in vitro</i> laboratory studies. We obtained 101 potential targets for toluene from a molecular similarity search. Results were further analyzed using <i>in silico</i> molecular docking methods. The binding of toluene to two proteins, hemoglobin and serum albumin, was further validated by the measurement of binding using microscale thermophoresis. The measured binding constant between toluene and hemoglobin was 1.9 μ M, while albumin demonstrated toluene-induced aggregation. These results demonstrate the applicability of an exploratory <i>in silico</i> toxicity tool, based on a molecular similarity search and protein-ligand docking for identification of potential targets for chemical toxins.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF: U		17. LIMITATION OF ABSTRACT SAR		18. NUMBER OF PAGES 8	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	19a. NAME OF RESPONSIBLE PERSON C. Mauzy		
			19b. TELEPHONE NUMBER (Include area code) NA		



Cite this: *Toxicol. Res.*, 2015, **4**, 519

Identifying potential protein targets for toluene using a molecular similarity search, *in silico* docking and *in vitro* validation†

Y. G. Chushak,^{*a,b} R. R. Chapleau,^{a,b} J. S. Frey,^{a,b} C. A. Mauzy^b and J. M. Gearhart^{a,b}

The toxicity of chemicals greatly depends on their interaction with macromolecular targets. The main goal of this study was to develop an approach for predicting protein targets for chemical toxins using a molecular similarity search of toxin–target information collected in the Toxin and Toxin-Target Database. The developed method was used to identify new targets for toluene which could predict potential cellular toxicity and to validate the approach with *in vitro* laboratory studies. We obtained 124 potential targets for toluene from a molecular similarity search. Results were further analysed using *in silico* molecular docking methods. The binding of toluene to two proteins, hemoglobin and serum albumin, was validated by the measurement of binding using microscale thermophoresis. The measured binding constant between toluene and hemoglobin was 1.9 μM , while albumin demonstrated toluene-induced aggregation. These results demonstrate the applicability of an exploratory *in silico* toxicity tool, based on a molecular similarity search and protein–ligand docking, for identification of potential targets for chemical toxins.

Received 9th January 2015,
Accepted 2nd February 2015

DOI: 10.1039/c5tx00009b

www.rsc.org/toxicology

1 Introduction

The adverse outcome pathway (AOP) is a new framework proposed recently for toxicology and risk assessment.¹ It links the chemical exposure of an organism to a series of events on different biological levels that result in an adverse outcome effect. The AOP starts with the Molecular Initiating Event (MIE) that describes the interaction of a chemical with its macromolecular targets.² Although significant amounts of information about the biological activities of toxins and toxin–target interactions have already been collected using experimental methods and stored in a variety of public databases, such as the Toxin and Toxin-Target Database (T3DB),³ ChEMBL,⁴ and Comparative Toxicogenomics Database (CTD),⁵ a great deal of specific information still remains missing. Computational methods provide a means to fill this data gap and provide focused target hypotheses for future mechanistic experiments.

Several methods for computational target prediction (or target fishing) have been developed and summarized in com-

prehensive reviews.^{6,7} These methods can be divided broadly into two groups: methods that are based on existing published data or *de novo* prediction methods. The former employ the methods of chemical similarity search, data mining and machine learning approaches and uses the currently known small molecule–protein relationships to predict novel potential activities. These methods can be applied to a variety of targets that have a set of active ligands, including G-protein coupled receptors and ion channels. The latter apply target fishing methods that use structure-based approaches with protein–ligand docking of small molecules to hundreds or thousands of proteins. These methods can identify novel protein targets that are not elaborated in databases, but require known 3D structures of solved proteins; thus membrane proteins are usually excluded from consideration due to difficulty in pure crystallization forms and subsequently their lack of experimentally determined structures.

The presented study aimed to explore the application of a molecular similarity search to identify new protein targets for toluene. Toluene, a volatile organic compound (VOC), is commonly used as a solvent in commercial products. The main toxicological effect from toluene exposure is on the nervous system. Acute exposure to toluene vapours may cause drowsiness, headache, nausea, visual changes, muscle spasm, dizziness, and loss of coordination.^{8,9} Occupational studies have shown that chronic toluene exposure can result in nervous system effects such as a reduction in thinking, memory, and muscle control,^{10,11} as well as some losses in hearing¹² and

^aHenry M Jackson Foundation for the Advancement of Military Medicine, 2729 R Street, Wright Patterson AFB, Ohio 45433, USA. E-mail: Chushak@yahoo.com

^bMolecular Bioeffects Branch, Bioeffects Division, Human Effectiveness Directorate, 711th Human Performance Wing, Air Force Research Laboratory (711 HPW/RHD), 2729 R Street, Wright Patterson AFB, Ohio 45433, USA

†Electronic supplementary information (ESI) available: A complete list of protein targets from a molecular similarity search. See DOI: 10.1039/c5tx00009b

colour vision.¹³ Although the toxicity and health effects of toluene are well studied and documented,¹⁴ a mechanistic understanding of toluene toxicity is still limited. Despite the fact that T3DB listed 38 protein targets for toluene, they are different subunits of 4 proteins: cardiac sodium ion channel,¹⁵ NMDA-glutamate receptor,¹⁶ gamma-aminobutyric acid type A receptor¹⁷ and neuronal nicotinic acetylcholine receptor.¹⁸ The interaction of toluene with sodium ion channels may be responsible for the arrhythmogenic effect of toluene while its interaction with neuronal receptors can be related to toluene effects on the nervous system. The ChEMBL database lists 8 targets for toluene but only 5 of these targets are human proteins: acetylcholinesterase, arachidonate 12-lipoxygenase, pre-amin-A/C, thyroid hormone receptor and thyroid stimulating hormone receptor. The interaction of toluene with acetylcholinesterase was obtained from a QSAR analysis¹⁹ while all other interactions were extracted from the PubChem BioAssay data set. This limited information regarding the interaction of toluene with protein targets and the importance of protein binding in its biokinetic disposition have motivated us to explore the application of molecular similarity search to expand the list of potential protein targets for toxins.

The general idea underlying the molecular similarity approaches is that two similar chemical molecular structures are likely to have similar target-binding profiles. For solving this chemical comparison question, we developed a protocol that utilizes the OpenEye programs ROCS and EON²⁰ to perform molecular similarity searches on the T3DB database. The ROCS program performs 3D shape and chemical features (called 'color') alignments. The similarity is ranked based on the TanimotoCombo score. The second OpenEye program, EON, uses a field-based Tanimoto scoring to measure and compare the electrostatic potential of two small molecules. ROCS is considered to be one of the best programs for molecular similarity searches²¹ and has recently been used for drug repurposing analysis involving drug–target information from the Drug Bank.²² Although for simple molecules as, in our case, toluene, a similarity search using 2D methods will produce the same results, but for more complex molecules, 3D methods outperform 2D molecular similarity methods.²¹

Recently, the information collected in T3DB (<http://www.t3db.org>) was exploited using a support vector machine and random forest methods for prediction of multiple toxin–target interactions related to cardiovascular diseases.²³ We performed a molecular similarity search using toxin–target information collected in the T3DB database and obtained 124 potential protein targets for toluene. Thirty-seven of these targets had available X-ray 3D structures in the Protein Data Bank. Therefore, toluene binding was further analysed using the protein–ligand docking program AutoDock and the docking results predicted significant binding of toluene to six proteins: cytochrome P450 2E1, histone H3.2, serum albumin, histone H4, hemoglobin and DNA polymerase. To partially validate our *in silico* search, we tested the binding of toluene to hemoglobin and serum albumin (HSA) using microscale thermophoresis (MST).

2 Materials and methods

2.1 Database preparation

The most comprehensive, open information about toxic substances and toxin–target interaction is collected in the Toxin and Toxin-Target Database.³ Currently (as of August 2014), the database contains information on about 3053 toxins linked to 1670 protein targets with 37 084 toxin–target associations. The information in the database was manually extracted from over 5454 sources including electronic databases, government documents, books and scientific publications. Several criteria were considered for the inclusion of chemical compounds as 'common' toxins into the database, such as: (1) presence in the home, the environment or the workplace with a recorded medical consequence of acute reaction, injury or death; (2) identification as hazardous in relatively low concentrations (<1 mM for some, <1 μ M for others); and (3) appearance in multiple toxin/poison lists provided by government agencies or the toxicological and medical literature.³ Performing a molecular similarity search requires the structure of chemicals that are defined in T3DB by SMILES string. Some toxins in the database do not have a SMILES description, such as ricin, which is a protein, or asbestos, which is a mineral. Therefore, our study omitted toxins without SMILES descriptions. The final compilation used for the molecular similarity search contained 2743 toxins and 1204 protein targets.

2.2 Molecular similarity search

OpenEye Scientific Software tools²⁴ were used to generate a set of 3D conformations for toxins from their SMILES strings. Initially, the program FLIPPER was used to enumerate stereocenters in a molecule. The output file was sent to the program FIXPKA that used a rule-based system to set the molecule to its most energetically favourable ionization state for pH = 7.4. Next, the MMFF94 atomic partial charges were assigned to the molecule using the program MOLCHARGE. OMEGA (version 2.4.6) was used to generate a maximum of 50 low-energy conformers for all toxins. The generated conformers comprised a database for molecular similarity search. A similar approach was used to prepare a query molecule except that a single low-energy conformer generated by OMEGA was used. It has been shown²⁵ that using a low-energy conformation has no impact on the alignment performance. A molecular similarity search was performed using ROCS (version 3.1.2) and EON (version 2.1.0). ROCS is a shape comparison program that uses atom-centred Gaussian functions to represent molecular shape. Additionally, it allows comparison of the chemical features of molecules called 'color'. To evaluate a molecule's similarity, we used a TanimotoCombo scoring function with a maximum score of 2, representing a perfect shape and chemical features match (molecule matching with itself). The closer the score is to 2, the better is the match with the query molecule. The aligned molecules from ROCS were analysed with the program EON that compares electrostatic fields calculated from the Poisson–Boltzmann equation. The similarity was scored by

ET_combo that takes into account both shape and electrostatic match.

2.3 Protein–ligand docking

All protein X-ray 3D structures were obtained from the Protein Data Bank (<http://www.rcsb.org/pdb/>). These structures are the most accurate and can be used for molecular docking. Target flexibility is one of the main difficulties in protein–ligand docking. We addressed this problem by storing within the dataset several entries for the same protein co-crystallized with the different ligands. In total, 240 files with the structures of 37 proteins were downloaded from PDB and used for docking screening. The identification and characterization of binding pockets were performed using the program Fpocket.²⁶ The Fpocket program maintains a list of cofactors and keeps them as part of the protein image during pocket detection. The detected pockets are scored and ranked based on several criteria such as the volume of the pocket, polarity, hydrophobicity, and druggability. We selected up to 8 top-ranked pockets for each protein to employ in docking. The protein–ligand docking was performed using AutoDock software developed in the Molecular Graphics Lab at the Scripps Research Institute. The recently released version of AutoDock Vina²⁷ has an improved local search routine and allows the use of multicore/multi-CPU computer systems. To prepare the protein structures for docking, the ligand and all water molecules were removed, while cofactors were kept as part of the binding pocket. Polar hydrogen and charges were added using the prepare_receptor4.py script from AutoDockTools4.²⁸ The toluene molecule was prepared for docking using the prepare_ligand4.py script from AutoDockTools4.

2.4 Hemoglobin purification

Human hemoglobin was purified using a modification of the method of Sun and Palmer.²⁹ Whole blood (450 mL) containing the anticoagulant acid citrate dextrose was centrifuged at 3716g for 30 minutes at 4 °C. Plasma was removed and discarded while the remaining red blood cell pellet was re-suspended in a 0.9% isotonic saline solution and centrifuged at 3716g for 30 minutes at 4 °C. The wash process was performed three times until the supernatant was clear. After the final wash, the supernatant was removed and the red blood cell pellets were re-suspended with roughly three volume equivalents of cold 3.75 mM phosphate buffer (PB, pH = 7.2) and stored frozen. The suspension was thawed on ice and centrifuged at 3716g at 4 °C for 30 min. The supernatant was filtered through cheesecloth and centrifuged two additional times at 3716g at 4 °C for 30 min. The remaining solution was concentrated using a 4 mL 10 KD centrifugal filter (Millipore), and the retentate was filtered through a 0.45 µm syringe filter (Corning).

Hemoglobin was purified from this clarified supernatant by anion exchange chromatography using a Mono Q 4.6/100 PE (GE Healthcare) on an ÄKTA Explorer FPLC system (GE Healthcare). The lysate was loaded onto the column using a low salt binding buffer (20 mM Tris-HCl pH 8.2). A linear gradient

from 100% low salt buffer to 75% high salt buffer (20 mM Tris-HCl, pH 8.2, 0.2 M NaCl) was generated in 5 column volumes. This was followed by a step gradient to 100% high salt buffer. Following anion exchange chromatography, a HiPrep 26/10 desalting column (GE Healthcare) was used to perform a buffer exchange into 1× PBS, pH 7.5. The hemoglobin was concentrated to 1.5 mg mL⁻¹ using an Amicon Centrifugal Filter Unit (Millipore). SDS-polyacrylamide and native gel electrophoresis experiments of the purified human hemoglobin were performed under standard conditions. Analysis of purified hemoglobin using SDS and native polyacrylamide gel electrophoresis (PAGE) indicated that the protein was >98% pure, and in monomer form composed of alpha and beta units (data not shown).

2.5 Microscale thermophoresis

A NanoTemper Technologies Monolith NT.115 GR was used to observe changes in thermophoretic mobility caused by toluene binding to proteins. Purified hemoglobin and human serum albumin (Sigma Aldrich) were adjusted to 1 mg mL⁻¹ in PBS for labelling with DyLight 650 (Pierce) according to the manufacturer's instructions, except that an extra centrifugal purification step was added to remove excess free dye that interferes with MST measurements. For MST measurements, protein labels were held constant (hemoglobin: 8.75 nM, HSA: 2.57 nM) while toluene levels were diluted serially 1:1 descending from 200 µM in MST Buffer (NanoTemper Technologies) with 0.2% DMSO using a 384-well microplate. Proteins and toluene were incubated at room temperature for 5 minutes prior to loading into hydrophilic treated glass capillaries (NanoTemper Technologies) pre-loaded in a capillary tray and reading using the red LED at 50% power and 80% IR laser power. Each measurement consisted of a 5 second baseline, a 30 second laser pulse, and a 5 second relaxation prior to moving to the next capillary. Capillaries were measured only once and each binding experiment was repeated in triplicate.

Normalized thermophoresis time traces were converted to bound fractions according to eqn (1) and plotted as a function of compound concentration using GraphPad Prism 5.

$$f_{\text{bound}} = \frac{(F_{\text{max}} - F_{\text{obs}})}{(F_{\text{max}} - F_{\text{min}})} \quad (1)$$

where F_{max} = normalized thermophoresis for the unbound state, F_{min} = normalized thermophoresis for the bound state, and F_{obs} = normalized thermophoresis for a given binding reaction.

3 Results and discussion

The molecular similarity search was used to identify new potential protein targets for toluene by utilizing information about the toxin–target interaction collected in the T3DB. T3DB is a toxin-centred database that provides a list of biomolecular targets for a specific toxin. That information was used to create a target-centered library. For each of the 1204 targets,

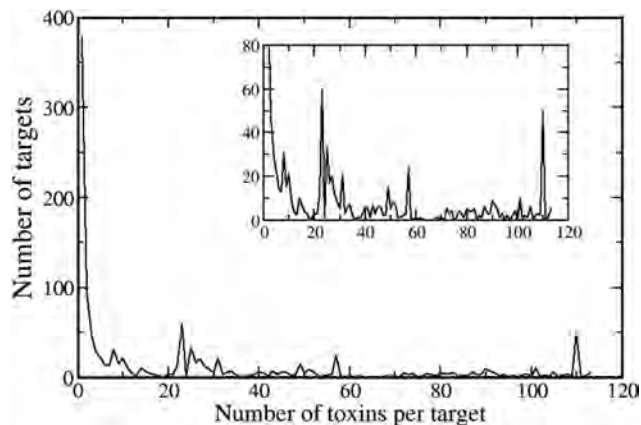


Fig. 1 Distribution of the number of targets vs. the number of toxins per target in T3DB. The inset shows the detailed view of target distribution on a lower scale.

we created a file with a SMILES description of toxins that interact with that target. Fig. 1 shows the distribution of the number of toxins per target in T3DB.

There are 379 targets that interact just with one toxin and 95 targets with two known toxins. On the other side, five targets (vascular cell adhesion protein 1, urokinase plasminogen activator surface receptor, C-X-C motif chemokine 10, C-C motif chemokine 2 and nuclear receptor subfamily 1 group I member 2) are affected by 113 toxins, one protein (peroxisome proliferator-activated receptor gamma) interacts with 112 toxins, and 50 proteins interact with 110 toxins. Furthermore, targets that interact with 113 and 112 toxins have 25 common toxins such as bisphenol A.

Multiple conformations of the toxins for each protein target were generated using the program OMEGA, creating a database for a molecular similarity search to identify potential targets for chemical toxins. Initially, the program ROCS was used to perform shape and chemical feature matches and then aligned molecules were sent to EON to perform electrostatic field comparisons. The similarity of molecules was ranked with the TanimotoCombo (TC) score that has a maximum value of 2 with a combination of values of 1 for shape and 1 for electrostatic similarity. Usually, $TC > 1.4$ is used to identify similar molecules.^{22,30} Fig. 2 shows some of the molecules in the T3DB database that were identified as similar to toluene. The toluene molecule is presented as green sticks and a grey surface; analogous molecules are presented by grey sticks. The green sphere in the centre of molecules represents a colour feature, in our case a ring. Table 1 presents similar molecules to toluene in the T3DB database with combo scores in the range of 1.5–2.0. We also included the number of targets for these molecules since some of the molecules do not have known targets in T3DB.

We identified 124 potential targets for toluene using a molecular similarity search of T3DB. They represent several families of proteins (calcium-transporting ATPases, sodium/potassium-transporting ATPase, cytochrome P450, DNA topo-

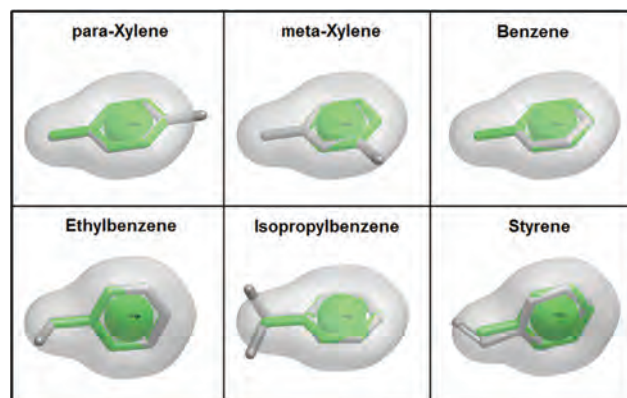


Fig. 2 Similarity of toluene to analogues in the T3DB database. The toluene molecule is presented as green sticks and a grey surface; analogous molecules are presented by grey sticks.

Table 1 Similar molecules to toluene in the T3DB database ranked by EON TanimotoCombo score

T3DB ID	Name	TanimotoCombo score	Number of targets
T3D0185	<i>para</i> -Xylene	1.943	2
T3D0863	<i>meta</i> -Xylene	1.941	2
T3D0006	Benzene	1.932	95
T3D0099	Ethylbenzene	1.927	12
T3D3468	Isopropylbenzene	1.889	0
T3D0864	<i>ortho</i> -Xylene	1.844	2
T3D0271	Styrene	1.769	1
T3D1779	Bromobenzene	1.676	9
T3D0635	Azulene	1.602	2
T3D1776	Benzyl bromide	1.578	0
T3D1809	<i>m</i> -Xylyl bromide	1.545	0
T3D1808	<i>p</i> -Xylyl bromide	1.54	0
T3D1807	<i>o</i> -Xylyl bromide	1.537	0

isomerases, histone proteins, and tubulin), different subunits of multimer proteins (cytokine receptor, DNA polymerase, hemoglobin and ribonucleoside-diphosphate reductase) and single proteins such as fatty acid-binding protein or serum albumin. The majority of the potential targets came from the similarity of toluene to benzene (95), ethylbenzene (12) and bromobenzene (9). Some of the targets interact with multiple toxins, thus confirming the hypothesis that similar molecules are likely to have a similar target-binding profile. For example, the cytokine receptor is affected by benzene, styrene and azulene. The complete list of potential targets is presented in the ESI.†

To analyse the binding of toluene to potential targets we performed protein–ligand docking. Computational docking is a common method used to identify small-molecule ligands that bind to proteins.³¹ It is usually used to screen a library of small molecules in order to find chemicals that bind to a specific protein receptor. In our approach, we screened the library protein molecules to find receptors with the highest binding affinity to toluene. Out of 124 protein targets identified by molecular similarity, 37 proteins had available X-ray 3D structures in the Protein Data Bank. Protein structures deter-

mined using X-ray scattering are the most accurate 3D structures and are suitable for molecular docking. In total, 240 files with protein structures were downloaded from the PDB and used for protein–ligand docking. It is widely accepted that ligand binding can drastically alter the functional binding pocket or receptor's conformation.³² To account for such conformational flexibility, we used several files containing solved 3D structure for the same protein if they were available in PDB. Table 2 shows the top proteins with a significant binding with toluene (binding energy $\Delta G < -6$ kcal mol⁻¹). The binding energy to similar toxins is also presented for comparison. The binding energies of toluene and *para*-xylene are very similar with a tendency of *para*-xylene to slightly lower binding energy to some proteins. On the other hand, benzene has higher binding energy (or lower binding affinity) to all the studied proteins.

The interaction of toluene with cytochromes P450 was obtained due to the similarity of toluene to xylenes that inhibited the activity of cytochromes P450 in lung of rats following inhalation of *m*-xylene.³³ It was shown that cytochrome P450 isozymes are responsible for the metabolism of toluene in human liver,³⁴ which can explain the high binding affinity of toluene to cytochrome P450. Occupational exposure to toluene has been shown to induce cytochrome P450 2E1 expression as assessed by mRNA content in peripheral lymphocytes,³⁵ further supporting our results of the “blind” *in silico* assessments.

All other proteins in Table 2 were obtained as possible targets for toluene due to the similarity to benzene. Epigenetic modifications such as histone acetylation and methylation play a crucial role in transcriptional regulation of genes and abnormal acetylation/deacetylation can lead to the development of cancer and other aberrant forms of homeostasis and pathologies. It was shown that benzene and its reactive metabolites attack multiple lysine residues within a single histone that can result in chromatin structural changes and gene expression.³⁶ The effect of toluene on histone acetylation was examined in different areas of the rat brain by immunofluorescence using antibodies that specifically recognize the acetylated form of histones H3 and H4.³⁷ It was found that histone H3 becomes readily acetylated in toluene exposed rats in the Nucleus Accumbens area of the brain while H4 acetylation levels remain similar to control rats. On the other hand, in the ventral tegmental brain region, H3 acetylation levels were unaffected by toluene inhalation, while H4 acetylation was strongly reduced from normal levels.³⁷ These experimental

results indicate that toluene can directly interact with histones H3 and H4 or indirectly through posttranslational modification by affecting histone acetyltransferases and histone deacetylases. Our computational results indicate possible binding of toluene to histones H3 and H4.

Serum albumin plays an important role in toxicology and drug development as a major transporter of chemicals to different tissues. The binding affinity of drugs and metabolites to serum albumin is directly related to the distribution of chemicals throughout the body as well as to the fraction of unbound chemicals available for biological processes and toxic effects.³⁸ The interaction of benzene and its metabolites with serum albumin and hemoglobin is well studied and documented.^{39–41} Benzene oxide, a metabolite of benzene, interacts with cysteine residues in both hemoglobin and albumin to form protein adducts^{39,40} that have been used as biomarkers of exposure to benzene.⁴¹ To the best of our knowledge, there are no data regarding the binding of toluene to human serum albumin (HSA). Although toluene is widely used in the crystallization of hemoglobin,⁴² we are not aware of any other effect of toluene on hemoglobin. However, exposure to toluene changes gene expression of both hemoglobin⁴³ and albumin⁴⁴ in the rat brain. Therefore, these two proteins were selected for further analysis of toluene binding.

Analysis of the predicted binding pose of toluene to hemoglobin (Fig. 3) suggested that toluene is located between the heme group HEM142 and a hemoglobin hydrophobic pocket formed by PHE33, PHE43, PHE46 and LEU48. It interacts with the hemoglobin through CG and CD1 atoms of PHE43, CG atoms of PHE46 and CD1 and CE atoms of HIS58 residues. Models of toluene binding show close contact with three atoms of the heme group HEM142: CBC, C1D and C2D (nomenclature is based on 1Y01.pdb). The location of toluene in close proximity to the heme group is in agreement with the results of NMR relaxation time measurements by Novak *et al.*⁴⁵ that showed that interaction of toluene and human hemoglobin occurred in the vicinity of the heme iron atom. The docked location of toluene in hemoglobin is different from the location of two toluene molecules observed in the crystallized form of hemoglobin (PDB entry 2DN1.pdb).⁴⁶ Furthermore, the predicted binding affinity to toluene in the present location is almost an order of magnitude higher compared to the position in 2DN1.pdb (33.4 μ M vs. 130 μ M).

Structural analysis of complexes of human serum albumin with a variety of drug and toxin molecules identified two

Table 2 Proteins with significant binding to toluene as predicted by AutoDock Vina. The binding energy to similar toxins is presented for comparison

Protein name	UniProt ID	Toluene ΔG (kcal mol ⁻¹)	<i>para</i> -Xylene ΔG (kcal mol ⁻¹)	Benzene ΔG (kcal mol ⁻¹)
Cytochrome P450 2E1	P05181	-7.1	-7.3	-6.3
Histone H3.2	Q71DI3	-6.7	-6.5	-5.9
Serum albumin	P02768	-6.2	-6.9	-5.4
Histone H4	P62805	-6.2	-6.1	-5.2
Hemoglobin subunit alpha	P69905	-6.1	-6.6	-5.4
DNA polymerase subunit gamma-2	Q9UHN1	-6.1	-6.6	-5.4

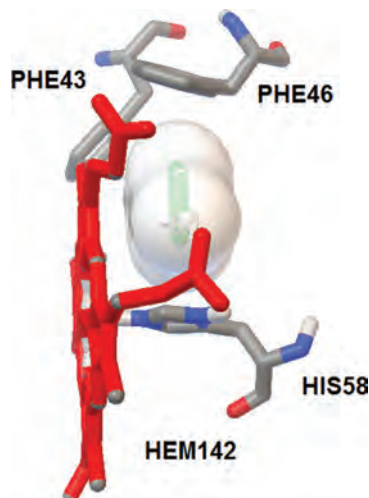


Fig. 3 Predicted mode of toluene binding to hemoglobin. Toluene (white sphere) is located between the heme group HEM142 (red) and a hemoglobin hydrophobic cavity formed by HIS58, PHE46 and PHE43.

primary and numerous secondary binding sites on the protein.⁴⁷ Analysis of the predicted binding pose of toluene to albumin showed that albumin residues ALA26, TYR30, LEU66, PHE70, LEU250 and LEU251 form a hydrophobic core of the binding pocket that is located in close proximity to the primary binding site 1. It also suggested that toluene forms classical parallel π - π stacking with the phenyl group of PHE70 residue of albumin (Fig. 4) that can lead to a lower binding energy of toluene to albumin compared to hemoglobin.

The binding of toluene to hemoglobin and albumin was experimentally evaluated using microscale thermophoresis measurements. It was found that toluene interacts with both proteins *in vitro* (Fig. 5). Our results indicate that the dissociation constants K_d calculated by AutoDock Vina underestimate the true affinity of the proteins for toluene. For

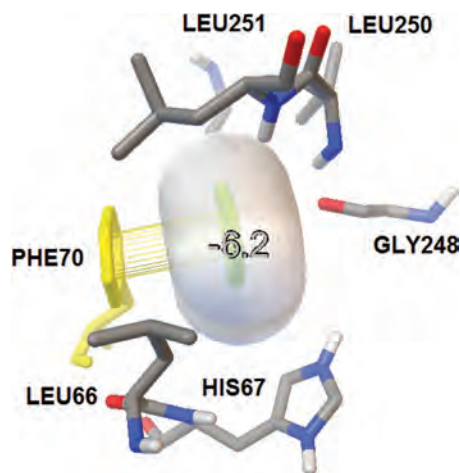


Fig. 4 Predicted pose of toluene binding to albumin. Toluene (white sphere) interacts with the phenyl group of PHE70 of albumin through a classical parallel π - π interaction. It also interacts with hydrophobic residues LEU66, LEU250 and LEU251, as well as with HIS67 and GLY248.

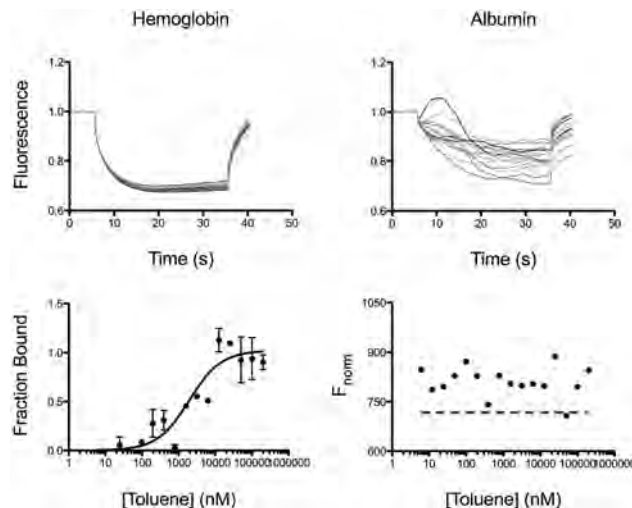


Fig. 5 Toluene binding to hemoglobin (left) and albumin (right). The MST fluorescence timetraces (top) were from a single representative experiment. The ordered gradient in the hemoglobin timetrace is indicative of a specific binding event with a calculable K_d . In contrast, the random signal in the albumin traces suggests nonspecific binding and induced aggregation. Error bars in the hemoglobin binding curve are from triplicate experiments (bottom left). The F_{norm} intensities (used to determine the bound fraction) for albumin and toluene are shown with the DMSO only control (bottom right, dots and dashed line, respectively).

example, we observed an *in vitro* binding affinity of 1.9 μM for hemoglobin and toluene at 25 $^{\circ}\text{C}$, whereas docking calculations predicted a dissociation constant of 33.4 μM . However, we were unable to calculate the dissociation constant for serum albumin and toluene. This was caused by aggregation induced by toluene binding to the protein (Fig. 5), a phenomenon frequently observed with HSA and small molecules.⁴⁸ At the lowest concentration tested, 6.57 nM, toluene initiated the aggregation of HSA. Neither control experiments of HSA diluted in a 0.2% DMSO containing buffer nor toluene diluted in the same buffer displayed any aggregation (data not shown due to non-incorporation of a fluorescent label). Therefore, we are able to conclude that HSA does indeed interact with toluene; whether or not this is a specific interaction remains to be elucidated.

A recent analysis of bovine serum albumin (BSA) binding to a variety of organic compounds, including benzene and toluene, by measuring the albumin-water and octanol-water partition coefficients, showed that the high binding affinity of BSA to small low-polarity molecules cannot be explained by nonspecific binding.⁴⁹ The dissociation constant for BSA and toluene, calculated according to eqn (12) and Table 1 in ref. 49, was estimated to be 81.3 μM , which is close to the value of 28.2 μM predicted by our docking calculations.

Thermophoresis is an equilibrium method for determining the binding affinity and does not require immobilization of either binding partner. It does, however, require the use of a fluorescent label, and a previous study has shown that data obtained from MST are comparable to those of label-free

methods such as isothermal titration calorimetry or the gold-standard immobilized method of surface plasmon resonance.⁵⁰ Although the values we observed for binding affinities are different from those predicted, when compared to other methods of biophysical analyses the differences seen between MST data vs. predicted binding were inconsequential. Therefore, our docking methodology clearly correlates well with *in vitro* results, suggesting that this *in silico* approach for toxicology target identification is a valuable tool for novel compound screening.

4 Conclusions

New protein targets for toluene were identified by performing a molecular similarity search of toxin–target information collected in the T3DB database. We obtained 124 potential targets that represented several families of proteins, different subunits of multimer proteins and single proteins. The obtained results were additionally analysed using molecular docking. Docking results show significant binding of toluene to 6 proteins: cytochrome P450 2E1, histone H3.2, serum albumin, histone H4, hemoglobin and DNA polymerase. We performed *in vitro* measurements of toluene binding to hemoglobin and serum albumin using microscale thermophoresis and confirmed the predicted binding. These results demonstrate the applicability of a molecular similarity search and protein–ligand docking for the identification of potential targets for chemical toxins.

However, this approach is limited to the amount of information collected in T3DB or other databases. Currently (as of August 2014), T3DB has information about 3053 toxins but the PubChem database has a collection of more than 42 000 000 chemicals. For many chemicals that we tested, we determined no hits from the similarity search. Secondly, T3DB has information about 1670 protein targets, but the Protein Data Bank has structures for more than 6000 human proteins. Finally, even for toxins and targets that are in the database, not all connections are added and annotated. For example, the ChEMBL database lists acetylcholinesterase (AChE) as a target for toluene. Although that information was obtained from a QSAR analysis,¹⁹ several studies showed that toluene at a concentration of 900 ppm inhibits AChE activity by up to 60%.^{51,52} But our molecular similarity search did not identify AChE as a target for toluene because there is no interaction of toluene or similar compounds with AChE in the T3DB database. A complementary, *de novo* approach for protein target prediction that does not depend on existing information about the toxin–target interaction is currently under development in our laboratory.

Acknowledgements

This research was funded by the US Air Force Aerospace Toxicology Program and by the Defense Threat Reduction Agency, Joint Science and Technology Office.

This paper has been approved for public release with unlimited distribution.

References

- 1 M. Vinken, *Toxicology*, 2013, **312**, 158–165.
- 2 S. Gutsell and P. Russell, *Toxicol. Res.*, 2013, **2**, 299–307.
- 3 E. Lim, A. Pon, Y. Djoumbou, C. Knox, S. Shrivastava, A. C. Guo, V. Neveu and D. S. Wishart, *Nucleic Acids Res.*, 2010, **38**, D781–D786.
- 4 A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, *Nucleic Acids Res.*, 2012, **40**, D1100–D1107.
- 5 A. P. Davis, C. G. Murphy, R. Johnson, J. M. Lay, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B. L. King, M. C. Rosenstein, T. C. Wiegiers and C. J. Mattingly, *Nucleic Acids Res.*, 2013, **41**, D1104–D1114.
- 6 J. L. Jenkins, A. Bender and J. W. Davies, *Drug Discovery Today Technol.*, 2006, **3**, 413–421.
- 7 A. Koutsoukas, B. Simms, J. Kirchmair, P. J. Bond, A. V. Whitmore, S. Zimmer, M. P. Young, J. L. Jenkins, M. Glick, R. C. Glen and A. Bender, *J. Proteomics*, 2011, **74**, 2554–2574.
- 8 D. Echeverria, L. Fine, G. Langolf, A. Schork and C. Sampaio, *Br. J. Ind. Med.*, 1989, **46**, 483–495.
- 9 P. J. Bushnell, W. M. Oshiro, T. E. Samsam, V. A. Benignus, Q. T. Krantz and E. M. Kenyon, *Toxicol. Sci.*, 2007, **99**, 181–189.
- 10 S. Y. Tsai, J. D. Chen, W. Y. Chao and J. D. Wang, *Environ. Res.*, 1997, **73**, 146–155.
- 11 R. V. Winchester and V. M. Madjar, *Ann. Occup. Hyg.*, 1986, **30**, 307–317.
- 12 T. C. Morata, A. C. Fiorini, F. M. Fischer, S. Colacioppo, K. M. Wallingford, E. F. Krieg, D. E. Dunn, L. Gozzoli, M. A. Padrao and C. L. Cesar, *Scand. J. Work. Environ. Health*, 1997, **23**, 289–298.
- 13 M. Zavalic, Z. Mandic, R. Turk, A. Bogadi-Sare and D. Plavec, *Am. J. Ind. Med.*, 1998, **33**, 297–304.
- 14 Agency for Toxic Substances and Disease Registry (ATSDR), *Toxicological profile for Toluene*, U.S. Department of Health and Human Services, Public Health Service, Atlanta, GA, 2000.
- 15 S. L. Cruz, G. Orta-Salazar, M. Y. Gauthereau, L. Millan-Perez Pena and E. M. Salinas-Stefanon, *Br. J. Pharmacol.*, 2003, **140**, 653–660.
- 16 A. S. Bale, M. D. Jackson, Q. T. Krantz, V. A. Benignus, P. J. Bushnell, T. J. Shafer and W. K. Boyes, *Toxicol. Sci.*, 2007, **98**, 159–166.
- 17 C. J. Meulenberg and H. P. Vijverberg, *Toxicology*, 2003, **190**, 243–248.
- 18 A. S. Bale, C. T. Smothers and J. J. Woodward, *Br. J. Pharmacol.*, 2002, **137**, 375–383.
- 19 M. G. Cardozo, Y. Iimura, H. Sugimoto, Y. Yamanishi and A. J. Hopfinger, *J. Med. Chem.*, 1992, **35**, 584–589.

- 20 T. S. Rush, J. A. Grant, L. Mosyak and A. Nicholls, *J. Med. Chem.*, 2005, **48**, 1489–1495.
- 21 G. Tresadern, D. Bemporad and T. Howe, *J. Mol. Graphics Modell.*, 2009, **27**, 860–870.
- 22 M. D. AbdulHameed, S. Chaudhury, N. Singh, H. Sun, A. Wallqvist and G. J. Tawa, *J. Chem. Inf. Model.*, 2012, **52**, 492–505.
- 23 W. Zhou, C. Huang, Y. Li, J. Duan, Y. Wang and L. Yang, *Toxicology*, 2013, **304**, 173–184.
- 24 *OpenEye Scientific Software: Santa Fe, NM*, 2014.
- 25 J. Kirchmair, S. Distinto, P. Markt, D. Schuster, G. M. Spitzer, K. R. Liedl and G. Wolber, *J. Chem. Inf. Model.*, 2009, **49**, 678–692.
- 26 V. Le Guilloux, P. Schmidtke and P. Tuffery, *BMC Bioinformatics*, 2009, **10**, 168.
- 27 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
- 28 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.
- 29 G. Sun and A. Palmer, *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.*, 2008, **867**, 1–7.
- 30 L. C. Blum and R. van Deursen, *J. Comput. Aided Mol. Des.*, 2011, **25**, 637–647.
- 31 G. Pujadas, G. M. Vaque, A. Ardevol, C. Blade, M. Salvado, M. Blay, J. Fernandez-Larrea and L. Arola, *Curr. Pharm. Anal.*, 2008, **4**, 1–19.
- 32 A. May, F. Sieker and M. Zacharias, *Curr. Comput. Aided Drug Des.*, 2008, **4**, 143–153.
- 33 J. W. D. Foy, D. M. Silverman and R. A. Schatz, *J. Toxicol. Environ. Health, Part A*, 1996, **47**, 135–144.
- 34 H. Kim, R. S. Wang, E. Elovaara, H. Raunio, O. Pelkonen, T. Aoyama, H. Vainio and T. Nakajima, *Xenobiotica*, 1997, **27**, 657–665.
- 35 A. Mendoza-Cantú, F. Castorena-Torres, M. Bermúdez De León, B. Cisneros, L. López-Carrillo, A. E. Rojas-García, A. Aguilar-Salinas, M. Manno and A. Albores, *Environ. Health Perspect.*, 2006, **114**, 494–499.
- 36 K. E. Williams, T. A. Carver, J. L. Miranda, A. Kautiainen, J. S. Vogel, K. Dingley, M. A. Baldwin, K. W. Turteltaub and A. L. Burlingame, *Mol. Cell. Proteomics*, 2002, **1**, 885–895.
- 37 S. L. Sánchez-Serrano, S. L. Cruz and M. Lamas, *Neurosci. Lett.*, 2011, **489**, 142–147.
- 38 T. Peters Jr., *All about albumin: biochemistry, genetics, and medical applications*, Academic Press, New York, 1995.
- 39 W. E. Bechtold, J. D. Sun, L. S. Birnbaum, S. N. Yin, G. L. Li, S. Kasicki, G. Lucier and R. F. Henderson, *Arch. Toxicol.*, 1992, **66**, 303–309.
- 40 W. E. Bechtold, J. K. Willis, J. D. Sun, W. C. Griffith and T. V. Reddy, *Carcinogenesis*, 1992, **13**, 1217–1220.
- 41 Agency for Toxic Substances and Disease Registry (ATSDR), *Toxicological profile for Benzene*, U.S. Department of Health and Human Services, Public Health Service, Atlanta, GA, 2007.
- 42 M. F. Perutz, *J. Cryst. Growth*, 1968, **2**, 54–56.
- 43 J. E. Royland, P. R. S. Kodavanti, J. E. Schmid and R. C. MacPhail, *Toxicol. Sci.*, 2012, **126**, 193–212.
- 44 S. D. Hester, A. F. Johnstone, W. K. Boye, P. J. Bushnell and T. J. Shafer, *Neurotoxicol. Teratol.*, 2011, **33**, 521–529.
- 45 R. F. Novak, M. Dershwitz and F. C. Novak, *Mol. Pharmacol.*, 1979, **16**, 1046–1058.
- 46 S. Y. Park, T. Yokoyama, N. Shibayama, Y. Shiro and J. R. Tame, *J. Mol. Biol.*, 2006, **360**, 690–701.
- 47 J. Ghuman, P. A. Zunszain, I. Petitpas, A. A. Bhattacharya, M. Otagiri and S. Curry, *J. Mol. Biol.*, 2005, **353**, 38–52.
- 48 P. Rondeau, G. Navarra, V. Militello and E. Bourdon, in *Protein Aggregation*, ed. D. A. Stein, Nova Science Publishers, Hauppauge NY, USA, 2011, pp. 139–159.
- 49 S. Endo and K. U. Goss, *Chem. Res. Toxicol.*, 2011, **24**, 2293–2301.
- 50 M. Ramakrishnan, F. A. de Melo, B. M. Kinsey, J. E. Ladbury, T. R. Kosten and F. M. Orson, *PLoS One*, 2012, **7**, e40518.
- 51 M. Korpela and H. Tähti, *Toxicol. In Vitro*, 1988, **2**, 135–139.
- 52 F. Arduini, F. Ricci, I. Bourais, A. Amine, D. Moscone and G. Palleschi, *Anal. Lett.*, 2005, **38**, 1703–1719.