COMPUTATIONAL SYSTEMS BIOLOGY TO EXPLORE ENVIRONMENTAL CHEMICAL TOXICITY HAZARDS

Audouze K[1], Grandjean P[2,3]

[1]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark; [2]Institute of Public Health, University of Southern Denmark, Odense, Denmark; [3]Department of Environmental Health, Havard School of Public Health, Boston, MA, USA.

Introduction

Computer-based modeling has been recommended as part of a new approach to predictive toxicology. In its report on "Toxicity Testing in the 21st Century", the National Research Council called for development of new approaches to human health risk assessment that would rely, in part, on computer-based models, rather than animal testing and epidemiology (National Research Council 2007). Moreover, the European Union promulgated a regulatory initiative for the Registration Evaluation Authorization, and Restriction of Chemicals (REACH). REACH requires alternative methods to animal testing to evaluate the chemical safety and/or risk assessment. These methods include computer based strategies, which may involve emerging strategies i.e. systems biology.
Systems biology is the study of an organism, viewed as an integrated, dynamic and interacting network of genes, proteins and biochemical reactions which gives rise to life. Instead of analyzing individual components or aspects of the organism, systems biologists focus on all the components and the interactions among them, all as part of one system. These interactions are ultimately responsible for an organism´s form and functions. Accordingly, systems biology may help to elucidate complex networks of genetic interactions that lead to toxicity of chemicals. Systems biology is possible due to recent advances within the "omics revolution". This provides the data and tools to study biological systems comprehensively, e.g. by transcriptomics, toxicogenomics, metabolomics or proteomics. These techniques enable detection of a large number of components on biological systems in parallel, which is required for the adoption of the systems approach.
To determine the usefulness of an integrated computational systems biology approach, we analyzed a case study involving the pesticide dichlorodiphenyltriichloroethene (DDT), its isomers and metabolites within the aim to ascertain their possible links to relevant adverse effects. DDT compounds are of high interest as humans are daily exposed to various persistent food contaminants such as DDTs. The potential effect of these environmental food contaminants on human health is a major concern and their mechanism of action is often not completely understood and can be associated to toxic effects. There is thus a need to improve our understanding of the underlying mechanism of action of chemicals and the biological pathways they perturb to fully evaluate their impacts on human health.

Materials and methods

Chemical-protein association networks:
For each DDT isomer and its metabolites, individual network of human chemical-protein associations were generated after data extraction using ChemProt[1], a disease chemical biology database that includes both binding data and gene expression data.
Protein complexes:
Within the aim to expand the list of relevant proteins linked to chemical, each network was enriched with other proteins to create protein complexes related to each chemical. A high confidence human interaction was explored to include know first order protein-protein interaction partners and other first order proteins participating in the same pathways[2]. Then follow protein enrichment using a protein-protein association network (P-PAN)[3]. The P-PAN is a knowledge- and data-driven method developed using toxicogenomics information.
Integration of disease information:
To identify associated dysfunctions and diseases linked to DDT compounds, protein-disease annotations were integrated into protein complexes. Two different sources of protein-disease information were explored: the

Online Mendelian Inheritance in Man database (OMIM) and the Comparative Toxicogenomics Database (CTD). OMIM is a highly reliable compilation of genetic variants from medical and genetics publications. CTD contains both *direct* and *inferred* gene–disease relationships and may therefore provide less certain associations. While direct associations have been documented in experimental models or through epidemiological studies, inferred relationships are established through indirect evidence. Thus, if chemical A is associated with disease B and chemical A has a curated interaction with protein C, then protein C has a direct relationship with disease B. This integrative computational systems biology approach allow to identify potential novel molecular mechanisms of action of chemicals, meaning deciphering unknown links between a chemical and a protein, which may lead to a toxic effect.

Results and discussion:

A total of 38 human proteins were extracted for *p,p'*-DDT, 83 for *o,p*'-DDT, and 18 for the *p,p*'-dichlorodiphenyldichloroethene (*p,p'*-DDE) metabolite from ChemProt. All three lists of proteins were used to create three human protein networks. Data extraction for other DDT compounds/metabolites did not lead to relevant information. Therefore the data analysis was concentrated on the three chemicals mentioned. Human protein complexes for the three chemicals were generated, by determining protein-protein interactions and protein-protein associations linked to each protein network. A total of 175 proteins was identified for *p,p'*-DDT, 187 for *o,p*'-DDT and 52 for *p,p'*-DDE. Disease enrichment was done on each protein complex. Although there were differences between the two data sources for linkage to diseases, integrated results predicted most diseases were linked to the two DDT isomers. Asthma was uniquely linked with *p,p´*-DDT. Several reproductive and neurobehavioral outcomes were linked to all three compounds studied although only *o,p´*-DDT was associated with autism. Several cancer types were connected to all three substances.

To illustrate our findings, male infertility disorders, including hypospadias and cryptorchidism, were predicted for *p,p'*-DDT, *o,p´*-DDT and *p,p'*-DDT. Genes linked to these male phenotypes included the androgen receptor (AR) listed for hypospadias. Although more genes were identified from the CTD, the statistical significance decreased. For cryptorchidism, both AR and gonadotropin-releasing hormone 1 were predicted. While primarily linked to the diseases mentioned, these genes also suggested that DDTs in humans might have endocrine disrupting effects. Looking at the literature, antiandrogen effects of *p,p'*-DDT and *p,p'*-DDE have been demonstrated experimentally[4]. Epidemiological studies showed an association between the anogenital distance in boys and both *p,p'*-DDT and *p,p'*-DDE concentrations in maternal serum[5], while another only found a significant association for the latter[6]. It is not clear whether *o,p'*-DDT plays any role in this respect. Most studies on cryptorchidism and hypospadias had limited statistical power or focused only on *p,p'*-DDE[7]. In regard to semen quality, a cross-sectional study of pesticide sprayers currently using DDT showed negative effects associated with the current serum concentration, especially for *p,p'*-DDT[8], while another study in infertile men showed that the sum of all *p,p'* isomers was negatively associated with the sperm concentration[9]. Again, much additional information is available on *p,p'*-DDE.

Based on current research data, differences in numbers of connected proteins may be due to deficient data. However, the parent DDT compounds appear responsible for more disease connections than the metabolites. Overall, our findings demonstrate that the systems biology approach is feasible and could have an important role in considering potential causal associations derived from toxicology and epidemiology studies. If the effects are more likely due to an unmeasured parent compound, or an unmeasured metabolite, then toxicokinetic calculations may need to be applied to generate a more appropriate exposure measure that reflects the amount of the active substance. Although our approach is of course based on currently knowledge and may therefore have overlooked some linkages, the results show that the DDT compounds examined, although chemically related, have tertiary structures, gene expression profiling and binding properties that deviate sufficiently from one another to predict outcomes that differ substantially. The large differences would unlikely be due to differences in the amount of information available. Thus, the major parent compound, *p,p'*-DDT, would seem to be much more potent in regard to adverse effects than the isomers and metabolites.

The usefulness and validity of the computational approach is likely to improve, as more information becomes available, including ´omics´ studies, gene-environment interaction studies, and more chemical-protein data. Further, the results of the disease-compound association analysis will improve in the future as newer, more complete and curated data will become available to expand and fine-tune the protein-disease data. In addition, studies like the present one will help validating the *in silico* findings, and cumulated experience will help

interpreting such analyses in light of possible unknown interactions and absent dose-effect relationships. Thus, the results obtained with the DDT compounds serve as an illustration of the potential use of computational predictions in toxicology, epidemiology, and environmental health research.

Acknowledgements: This work was supported by the Villum Kann Rasmussen Foundation, the Innovative Medicines Initiative Joint Undertaking (IMI-JU) for the eTOX project (115002) and the National Institute for Environmental Health Sciences (ES 012199 and ES 013692).

References:

1. Taboureau O, Nielsen SK, Audouze K, Weinhold N, Edsgard D, Roque FS, et al. (2011); *Nucleic Acids Res* 39(Database issue): D367-372
2. Lage K, Karlberg EO, Storling ZM, Olason PI, Pedersen AG, Rigina O, et al. (2007); *Nat Biotechnol* 25(3): 309-316.
3. Audouze K, Juncker AS, Roque FJ, Krysiak-Baltyn K, Weinhold N, Taboureau O, et al. (2010); *PLoS Comput Biol* 6(5): e1000788.
4. Gray LE, Ostby J, Furr J, Wolf CJ, Lambright C, Parks L, et al. (2001); *Hum Reprod Update* 7(3): 248-264.
5. Longnecker MP, Gladen BC, Cupul-Uicab LA, Romano-Riquer SP, Weber JP, Chapin RE, et al. (2007); *Am J Epidemiol* 165(9): 1015-1022.
6. Torres-Sanchez L, Zepeda M, Cebrian ME, Belkind-Gerson J, Garcia-Hernandez RM, Belkind-Valdovinos U, et al. (2008); *Ann N Y Acad Sci* 1140: 155-162.
7. Brucker-Davis F, Wagner-Mahler K, Delattre I, Ducot B, Ferrari P, Bongain A, et al. (2008); *Hum Reprod* 23(8): 1708-1718.
8. Aneck-Hahn NH, Schulenburg GW, Bornman MS, Farias P, de Jager C. (2007); *J Androl* 28(3): 423-434.
9. Messaros BM, Rossano MG, Liu G, Diamond MP, Friderici K, Nummy-Jernigan K, et al. (2009); *Environ Res* 109(4): 457-464.