"From Protein Toxins to Applied Toxicological Testing" virtual workshop identifies the need for a bioinformatic framework to assess novel food protein safety

Patricia A. Bauman, Andrew C. Doxey, Ivano Eberini, Emir Islamovic, Florence Jungo, Colton Kessenich, John Kough, Mansi Krishan, Luca Palazzolo, Laura Privalle, Chester Rodriguez, Karla Satchell, Andre Silvanovich, Lucilia Pereira Mouriès

PII: S0273-2300(22)00033-2

DOI: https://doi.org/10.1016/j.yrtph.2022.105146

Reference: YRTPH 105146

To appear in: *Regulatory Toxicology and Pharmacology*

Received Date: 2 February 2022

Accepted Date: 16 February 2022

Please cite this article as: Bauman, P.A., Doxey, A.C., Eberini, I., Islamovic, E., Jungo, F., Kessenich, C., Kough, J., Krishan, M., Palazzolo, L., Privalle, L., Rodriguez, C., Satchell, K., Silvanovich, A., Mouriès, L.P., "From Protein Toxins to Applied Toxicological Testing" virtual workshop identifies the need for a bioinformatic framework to assess novel food protein safety, *Regulatory Toxicology and Pharmacology* (2022), doi: https://doi.org/10.1016/j.yrtph.2022.105146.

Title:

"From Protein Toxins to Applied Toxicological Testing" Virtual Workshop Identifies the Need for a Bioinformatic Framework to Assess Novel Food Protein Safety

Authors: Bauman Patricia A.[1], Andrew C. Doxey[2], Ivano Eberini[3]; Emir Islamovic[4], Florence Jungo[5], Colton Kessenich[6], John Kough[7], Mansi Krishan[8], Luca Palazzolo[3], Laura Privalle[4], Chester Rodriguez[9], Karla Satchell[10], Andre Silvanovich[6], Lucilia Pereira Mouriès[11]*.

Affiliations:

[1] Syngenta Crop Protection, LLC, 9 Davis Drive, Research Triangle Park, NC 27709, USA
[2] University of Waterloo, Department of Biology, 200 University Ave. W, Ontario, Canada
[3] Dipartimento di Scienze Farmacologiche e Biomolecolari, Università degli Studi di Milano, Via Balzaretti 9, 20133, Milano, Italy
[4] BASF Corporation, 407 Davis Drive Tech 3, Morrisville, NC 27560, USA
[5] SIB Swiss Institute of Bioinformatics - Swiss-Prot group, 1 rue Michel Servet – CH 1211 Geneva 4, Switzerland
[6] Bayer Crop Science, 700 Chesterfield Parkway West, Chesterfield, MO 63017, USA
[7] US Environmental Protection Agency, Office of Pesticides Program, 1200 Pennsylvania Avenue, NW, Washington DC 20460, USA
[8] Becton, Dickinson, and Company (BD), 21 Davis Drive, Research Triangle Park, NC 27709, USA; and Society of Toxicology (SOT) Food Safety Specialty Section
[9] Duke University, DUMC 2914 Durham, NC 27710, USA; and Society of Toxicology (SOT) Food Safety Specialty Section
[10] Northwestern University Feinberg School of Medicine, 303 E. Chicago Avenue, Ward 6-205, Chicago, IL 60611, USA
[11*] Health & Environmental Sciences Institute (HESI), 740 15th Street NW, Suite 600, Washington, DC, 20005, USA

***Corresponding author email:** lmouries@hesiglobal.org

Abstract: On October 21–22, 2020 the HESI (Health and Environmental Science Institute) Protein Allergens, Toxins, and Bioinformatics Committee, and the Society of Toxicology Food Safety Specialty Section co-hosted a virtual workshop titled “From Protein Toxins to Applied Toxicological Testing”. The workshop focused on the safety assessment of novel proteins contained in foods and feeds, was globally represented by over 200 stakeholder attendees, and featured contributions from experts in academia, government and non-government organizations, and agricultural biotechnology developers from the private sector. A range of topics relevant to novel protein safety were discussed, including: the state of protein toxin biology, modes and mechanisms of action, structures and activity, use of bioinformatic analyses to assess the safety of a protein, and ways to leverage computational biology with *in silico* approaches for protein toxin identification/characterization. Key outcomes of the workshop included the appreciation of the complexity of developing a definition for a protein toxin when viewed from the perspective of food and feed safety, confirming the need for a case-by-case hypothesis-driven interpretation of bioinformatic results that leverages additional metadata rather than an alignment threshold-driven interpretation, and agreement that a “toxin protein database” is not necessary as the bioinformatic needs for toxin detection may be accomplished by existing databases such as Pfam and UniProtKB/Swiss-Prot. In this paper, a path forward is proposed.

Keywords: Protein toxin, safety assessment, bioinformatics, toxin database, mechanism of toxicity.

Abbreviations:

AoP: Adverse Outcome Pathway

CPD: Cysteine Protease Domain

EFSA: European Food Safety Agency

F3S: Food Safety Specialty Section

GO: Gene Ontology

HESI: Health and Environmental Science Institute

MARTX: Multifunctional auto-processing repeats-in toxins

NGO: Non-Governmental Organization

PATB: Protein Allergens, Toxins, and Bioinformatics

Pfam: Protein family database

POI: Protein of Interest

SNARE: soluble N-ethylmaleimide-sensitive factor attachment protein receptor

SOT: Society of Toxicology

TAS: Toxin-Antitoxin System

1. **Introduction**

The HESI **Protein Allergens, Toxins, and Bioinformatics (PATB) Committee**, in collaboration with the **Society of Toxicology (SOT) Food Safety Specialty Section (F3S)**, co-hosted a virtual workshop on 21–22 October, 2020 titled *"From Protein Toxins to Applied Toxicological Testing"* in the context of safety assessment of novel foods and feeds. The workshop had a dual purpose: (i) to learn about the status of the science of protein toxins biology from experts in the field, and (ii) to inform the development of new approaches for the assessment of potential protein toxicity in novel food and feed. Specifically, the workshop focused on the translation of knowledge to practical applications toward assessing the safety of new proteins in foods and feeds using *bioinformatic analyses*, and evaluating the value, applicability, and limitations of existing protein toxin databases and *in silico* tools for risk assessment.

HESI, in collaboration with SOT F3S, planned the workshop to create an open, peer-to-peer forum where scientists from academia, government, the agricultural biotechnology industry, NGOs, and other strategic stakeholders in the field could discuss and work together to improve approaches to the safety assessment of novel foods and feeds. The first two days were open to the global scientific community, and the third day was structured as a round-table discussion, with a globally representative subset of experts and attendees. The interactive format of this scientific program was intended to elicit participant feedback on current data gaps, research needs, and the potential for integration of new data into the food safety assessment process. Two hundred participants from over 20 countries and 4 continents registered for this event, demonstrating the global relevance of this topic.

This paper presents the key outputs of the meeting, including current methods used to evaluate the safety of novel proteins in foods, an overview of known bacterial protein toxins, and ways in which bioinformatics may be used to identify new protein toxins. In addition, it identifies a need for

developing a harmonized framework or consensus approach for assessing the potential toxicity of novel proteins.

2. Brief Overview of the State of the Science, Protein Toxins Biology

2.1 Assessing allergenicity vs. toxicity of newly expressed proteins

A key goal of the workshop was to discuss bioinformatic analyses that may help to discern or predict potential protein toxicity. These discussions built upon HESI's expertise in developing the *in silico* tool COMPARE (van Ree et al., 2021), www.comparedatabase.org, to help evaluate potential allergenicity by enabling comparisons of amino acid sequences to those in a peer-reviewed protein allergen database.

The use of sequence comparisons to assess potential allergenicity versus toxicity was presented by Dr. Andre Silvanovich (Bayer CropScience), with differences between protein allergens and toxins highlighted. Among those differences are the link between toxicity and protein function in the source organism, that is lacking in the case of allergens; the indiscriminate effects of a toxin compared to the restriction of allergic responses to those who are genetically predisposed to allergy; and the availability of many *in vitro* and *in vivo* functional assays to evaluate toxicity compared to the lack of such tests for allergy. Due to the relative lack of options for predicting allergy, assessors benefit from sequence comparison searches against a curated allergen database as a first step in a weight-of-evidence approach to assessing protein safety (Codex, 2009; EFSA, 2010), which includes bioinformatic analysis. The assessment of protein toxicity also incorporates bioinformatics analyses into its weight of evidence approach. This is especially useful early in product development to provide confidence that a gene selected for use in a biotechnology product will not encode a protein with toxicity against mammals. Amino acid sequence similarities to toxins can be readily addressed using existing publicly available sequence databases such as UniProtKB/Swiss-Prot and GenBank, where toxins are typically identified as such in the functional descriptions of the proteins.

The process of interpreting bioinformatic information during the evaluation of potential toxicity differs from that of allergenicity in that specific alignment thresholds have been described in Codex for addressing allergen sequence similarity, but such thresholds are not globally harmonized for toxins, and need closer scrutiny to determine relevance. For example, toxins may share sequence similarities over domains that define structural architecture or provide proper amino acid juxtaposition for functionality. Architectural elements (e.g., α-helices or β-pleated sheets,) may be shared without implied toxicity. It is also the case that toxicity described in the database may not be relevant to all organisms. For example, a multi-component pore forming toxin may not cause toxicity in a non-target organism if its requisite assembly does not occur.

An overview of the current approaches to the evaluation of protein safety, focusing on proteins of interest (POIs) in food crops and novel food ingredients was discussed by Dr. Laura Privalle (BASF Corporation) and Dr. Ray Matulka (Burdock Group Consultants), respectively. Dr. Privalle emphasized that the safety assessment of POIs in food crops created using biotechnology begins at the earliest stages of product development. When a trait is identified as having potential value, its protein sequence is screened by simple bioinformatic analyses to identify potential similarity with sequences of known protein allergens and toxins. If similarity is observed in early product development, a decision is made to either accept potential requirements for more complex explanations or studies or to discontinue further development of that trait. As product development proceeds, additional data are generated, including a detailed characterization of the newly encoded protein with respect to exposure, stability, functionality which may include oral toxicity testing in animals. Complete characterization of any predicted and/or unpredicted phenotypes attributable to the trait of interest are also performed as part of consumer and environmental safety assessments. During these evaluations, updated bioinformatic analyses may be requested, although for toxicological assessment these provide little benefit once a conclusion of safety has been reached for the POI.

A similar process for assessing the safety of proteins in novel food additives was described by Dr. Ray Matulka. In this case, both the protein and the components of the food additive are characterized, and a rationale for adding this component to food is required. Quality control parameters such as the presence of potential environmental or microbial toxins or contaminants associated with the production method, or the raw materials are evaluated. Some features of the safety analysis depend on the intended use of the food additive. If the protein is produced by microbial fermentation, then the production microbe is also characterized. Additional considerations include the evaluation of dietary exposure and potential allergenicity and toxicity.

2.2 Modes of action of protein toxins

The discussion of protein toxicity began by agreeing on the definition of a protein toxin. For the purposes of the workshop, the following definition was adopted: *"proteins that interact selectively with one or more biological molecules in another organism (the target organism), initiating pathogenesis (leading to an abnormal, generally detrimental state) in the target organism"* (Palazzolo et al., 2020).

Using bacterial toxins as models, Dr. Karla Satchell (Northwestern University) presented an overview of effectors, delivery systems and targets as essential components of toxicity. Bacteria can coordinate the expression, secretion (delivery) and stability of effectors that cause toxicity in target organisms. The delivery of bacterial toxins to target organisms typically requires passage through channels in bacterial and host cell membranes. Therefore, the formation of channels is often coupled to toxin delivery. Multiple secretory pathways may be used to deliver the toxin from its source to its target. Pore-forming proteins that deliver toxins can also act as toxins by causing efflux of essential ions (Collier & Young, 2003; Peng et al., 2019). These proteins can occur as monomers requiring assembly into pores (*Bacillus thuringiensis*), multi-component toxins with separate functional domains (*Bacillus anthracis* anthrax

144 toxin; *Vibrio cholerae* cholera toxin), or multidomain toxins that require processing for activation
145 (*Cornyebacterium diphtheriae* diphtheria toxin; *Clostridium difficile* toxins (Dutta et al., 2010; Ganguly et
146 al. 2014; Pruitt and Lacy, 2012). Target organism recognition can occur by receptor binding and is often
147 followed by receptor-mediated endocytosis as a first delivery step. The effector is typically translocated
148 into endosomes, then proteolytically cleaved to release active effector proteins. Effectors enzymatically
149 modify host cell targets and/or disrupt biochemical or cellular processes in a myriad of ways that change
150 the host cell biology. The activity of any given effector may also be influenced by other effectors,
151 resulting in an additive/synergistic or antagonistic effect. Certain bacterial toxins contain multiple
152 domains that perform the roles of target organism recognition, toxin delivery, and toxin effector.
153 Multifunctional Auto-processing Repeats-in ToXins (or MARTX proteins), are large secreted bacterial
154 proteins that are systems of toxins and multi-effector molecules (Gavin and Satchell, 2015; Satchell
155 2015). MARTX proteins commonly contain 2 to 5 effector proteins with an auto-processing cysteine
156 protease domain (CPD). Certain bacteria can exchange effector domain sequences, making the number
157 of available effector combinations dynamic and providing a mechanism to evolve new toxin pathologies.
158 The varied strategies for producing and delivering toxic effectors underscores the importance of
159 environmental context in determining toxicity, and the difficulties of predicting toxic effects if
160 information is limited or made by sequence comparisons alone.

161 Dr. Andrew Doxey (University of Waterloo) discussed the use of bioinformatics to predict protein toxins,
162 focusing on the example of neurotoxins produced by *Clostridium botulinum and C. tetani* (Doxey et al.
163 2018). Botulinum toxins are composed of a metalloprotease light chain (LC) and a heavy chain (HC)
164 which contains translocation and receptor-binding domains. Botulinum toxins target proteins at
165 neuronal synapses, cleaving SNARE proteins and disrupting neurotransmission. Dr. Doxey's team was
166 interested in understanding why such a complex and mechanistically sophisticated family of toxins could
167 be so taxonomically restricted, and whether similar toxins might be produced by other species. Their

approach involved building a statistical profile of existing botulinum-like toxins, and then scanning databases for other proteins matching that profile. Once candidate sequences were identified, additional information was gathered by analyzing conserved functional amino acid residues, motifs, and protein domains; structural modeling; phylogenetic analysis, and by investigating the genomic neighborhoods of the candidate genes for similarities to botulinum toxin gene clusters. They identified and characterized candidate "BoNT-like" toxins from three non-*Clostridium* species (Mansfield et al. 2015, 2019; Zhang et al. 2018), highlighting similarities and differences between the candidates and botulinum toxins. Dr. Doxey noted that the species (hosts) targeted by the predicted toxins and their potential enzymatic substrates may be difficult to predict, and that this represents an important area of future work for bioinformatic toxin prediction.

3. **Bioinformatics and Computational Biology Tools**

3.1 Overview of available tools

Ms. Florence Jungo (SIB Swiss Institute of Bioinformatics - Swiss-Prot group) presented an overview of the UniProtKB/Swiss-Prot database and other resources. The UniProtKB/Swiss-Prot knowledgebase (https://www.uniprot.org/) is a freely available resource for the scientific community, which provides a wealth of information on many proteins including toxins from all kingdoms. In each UniProtKB/Swiss-Prot entry, one may find protein sequences and features such as cleavage sites, posttranslational modifications, etc.; taxonomy and lineage of source species; and annotations regarding function, tissue specificity, subcellular location, and interactions with other proteins. Links to other databases and source references are also available. In addition, ontologies (Gene Ontology [GO] terms and Swiss-Prot keywords) are provided by UniProtKB.

There are also several specialized databases that focus on venoms and are either organism- or target-specific [see for example ConoServer (http://www.conoserver.org/) and Kalium (https://kaliumdb.org/)]. In addition to these databases, VenomZone (https://venomzone.expasy.org/) is a web-portal that provides an overview of venoms from the six major venomous taxa (snakes, scorpions, spiders, cone snails, insects and sea anemones). VenomZone provides descriptions of venom activity, target proteins and pharmacology, and links to a range of relevant resources.

3.2 Regulatory insights into protein risk assessment: in silico tools

Following the discussion of databases, a regulatory approach to using bioinformatic analyses for assessing protein safety was presented. Dr. Anna Lanzoni, from the European Food Safety Agency (EFSA), focused on the risk assessments of dietary proteins, including POIs used in genetically modified food crops, animals and microorganisms, and novel proteins used in other foods and feeds (EFSA GMO Panel Guidances 2006, 2011, 2012, 2017). Among the data included are the potential toxicological, allergenic, and adjuvant properties of the POI, including its potential involvement in non-IgE-mediated immune reactions such as in celiac disease. Among the methods used for data analyses are *in silico* methods, *in vitro* assays, and more complex *in vivo* studies in certain cases. Most relevant to the workshop were the role and uses of *in silico* tools in the risk assessment, especially their potentials for predicting hazard pertaining to toxicology, allergenicity, adjuvanticity, and non-IgE-mediated immune reactivity. The need to align scientific developments with advances in technology to deliver informative and realistic assessments was highlighted, and the evolution of assessments of potential allergy was contrasted with the evolution of the evaluation of potential toxicity, which is less advanced. Toward this end, EFSA commissioned the exploration of existing *in silico* protein toxicity prediction methods that might be useful in the assessment of potential toxicity.

Dr. Luca Palazzolo presented an outcome of this exploration, including an overview of *in silico* methods that are currently available and might be useful in predicting protein toxicity (Palazzolo et al. 2020). The work included conducting a comprehensive literature search regarding proteins causing adverse effects in humans and animals. Data collected included descriptions of the biochemical, functional and structural properties of the identified proteins, including the types of adverse effects that could occur, and molecular signatures (e.g. motifs, domains) and roles in the context of Adverse Outcome Pathways (AOP). Also included was a description and evaluation of available *in silico* resources, including one database and four protein toxicity prediction servers (Table 1). To manage the data retrieved, a Python-based software (TOXAPEX) was created which queries the UniProtKB database and interprets the .xml response by building an object-oriented database. From this object-oriented database, many Excel-compatible tables can be generated to report and summarize all the entries retrieved by a query. Two collections were established: the Main Collection consists of 6964 entries describing proteins with an associated well-recognized toxin activity; the Toxin-Antitoxin System (TAS) Collection contains 627 entries that describe bacterial toxins that primarily interact with their antitoxins but can interact with proteins of a target organism if the toxin-antitoxin equilibrium is disrupted.

As toxins are very different in terms of mode of action, size or structure, the protein toxicity prediction tools target different types of proteins. Thus, NTXpred (Saha & Raghava, 2007a) focuses on animal neurotoxins, BTXpred (Saha & Raghava, 2007b) on bacterial toxins, ClanTox (Naamati, G., et al. 2009) on animal toxins, while ToxinPred (Gupta et al. 2013) assesses whether peptides (less than 51 residues) from bacteria or animals are toxins. The KNOTTIN database (Postic et al. 2018) and its prediction tool are also described. Knottins are small, cysteine-rich proteins which contain at least three disulfide bonds, one of which crosses the macrocycle formed by the two other disulfides bonds and the interconnecting backbone. The tool evaluates if the protein is a knottin but not a toxin. However, since many knottins

237 (estimated to 80%) are toxins it could be of value. Various information on these resources, including
238 accuracy, is summarized in Table 1.

239

240 **Table 1:** Protein toxicity databases and prediction tools

Tool name and URL	Summary
NTXpred http://crdd.osdd.net/raghava/ntxpred/	• Reliably identifies cellular target and molecular function of neurotoxins • Accepts input in a variety of sequence formats • Yields 4 classifications including a yes-no assessment of whether the protein is a neurotoxin.
BTXpred http://crdd.osdd.net/raghava/btxpred/	• Yields a yes-no assessment of whether the protein is a bacterial toxin • Describes toxin type and function • Identifies non-toxins reliably • Less able to discriminate bacterial versus animal toxins • Fair sensitivity, poor accuracy in one of two statistical analyses.
The KNOTTIN database http://www.dsimb.inserm.fr/KNOTTIN/	• Identifies knottins • Distinguishes knottins from non-knottin proteins Does not predict toxicity, since many toxins are knottins • Good overall accuracy.
ClanTox http://www.clantox.cs.huji.ac.il	• Identifies animal toxins • Results are reported as “Toxin-like”, “Probably toxin-like”, "Possibly

	toxin-like" and “Probably not toxin-like”. • At the highest level of stringency, the method is insensitive • Displays excellent specificity and fair accuracy.
ToxinPred https://webs.iiitd.edu.in/raghava/toxinpred/index.html	• Uses an SMV (Support Vector Machine) and sequences from different databases to build prediction models. • The SMV output is then combined using MEME (Multiple Em for Motif Elicitation) and Motif Alignment & Search Tool (MAST) results to yield toxin/non-toxin assessments. • Shows good accuracy and specificity but only fair sensitivity.
ConoServer http://www.conoserver.org	• ConoServer is a database specializing in the sequence and structures of conopeptides • Provides information about conotoxin amino acid sequences, nucleotide sequences and structural information • Supports a range of options that include sequence- and structure-based searches.

4. Round Table discussion: Key opportunities, challenges, and proposed solutions to improving the assessment of protein toxicity

The focus of day 3 of the workshop was an open discussion of how toxin assessment might be approached or consolidated. It opened with the general consensus of developing a common set of integrated tools for *in-silico* testing. Although acknowledged as a desirable end point for all stakeholders, the complexities of such a task were quickly outlined. These included but were not limited

to: 1) complexities in defining toxins in a single repository; 2) complications in applying bioinformatics algorithms, which were developed to measure evolutionary similarities of proteins and are being applied to predict toxicity; and 3) complexity in defining thresholds for similarity. Consequently, the consensus opinion was that bioinformatics remain a key part of the weight of evidence approach to safety assessment that is further informed by a hypothesis driven approach which considers contextual information about the POIs' specific family. Currently, bioinformatic analysis would benefit from additional tools (decision frameworks) to allow a more comprehensive and predictive determination of toxicity. However, if a protein has been subjected to *in vivo* toxicology testing and shown to be safe, repeated bioinformatics analyses are not necessary.

The question of the need for a toxin database for safety assessments was also addressed. Workshop participants acknowledged that content is already collected, and curated to a large degree, in other repositories such UniProtKB/SwissProt or the protein family database Pfam. The consensus was that there is no need for a new database of curated toxins. Instead, it was suggested that efforts might be directed toward developing a harmonized framework or consensus approach to assessing potential toxicity. Such a framework was not fully defined in this workshop, but participants argued that it would likely include elements such as a tiered evaluation of bioinformatic data. Such information would include: 1) sequence similarity across the protein family to any known toxic or non-toxic members; 2) any knowledge of history of safe use for the host or source organisms; 3) probable exposure; and 4) a consolidation of terminology and approach for various stages of product development. This framework would interrogate the public databases that are already available, and in the context of predicting "toxicity", procedures could be developed that facilitate the use of literature, existing databases and repositories to further aid in the interpretation of sequence similarity results and other bioinformatic information obtained.

Challenges remain for further elaborating the framework in detail. Questions that need to be considered include what type of identification of potential hazard or non-hazard would drive the need to advance to the next tier of evaluation. The workshop participants concluded that resources and time should be devoted to fleshing out the framework process through collaborative discussions with stakeholders in multiple domains including those represented in this workshop.

In conclusion, the joint workshop found that using bioinformatics tools can provide useful information for the assessment of protein toxicity, but the development of a new additional toxin database is unnecessary. Distinguishing between toxic and non-toxic proteins is a more nuanced - and context-dependent process. It was recommended that the focus of future efforts in this field should be directed at the development of a framework for the bioinformatic assessment of proteins to guide navigation across existing public databases using available tools. Such efforts will require collaboration across sectors (regulatory, academic, and product developers) to the benefit of all stakeholders.

5. Key learnings and needs identified

Key learnings include:

- Toxicity requires effectors, delivery systems and susceptible targets;
- Currently bioinformatic analysis can identify sequence similarity but not predict toxicity because contextual information is required for the determination of relevance;
- An additional “protein toxin database” for bioinformatic searches is ***not*** needed since high quality public databases addressing toxin activity are already available;
- A harmonized framework or consensus approach to using bioinformatic tools and interrogating available public databases is needed to aid in the interpretation of sequence alignments.

Disclaimer:

The opinions, findings or conclusions recorded here are those of the individual workshop participants and do not necessarily represent the views of participants' organizations, the planning committee, HESI, or SOT FS3.

Funding:

This activity was supported by the Protein Allergens, Toxins and Bioinformatics (PATB) technical committee of the Health and Environmental Sciences Institute (HESI). The Food Safety Specialty Section (FS3) of the Society of Toxicology (SOT) collaborated with HESI PATB, helped plan the workshop, provided scientific inputs, speaker recommendations, and disseminated the workshop announcement to SOT FS3 membership. SOT FS3 did not provide any funding support for this workshop.

COI Statement:

Drs. Doxey, Eberini, Jungo, Kough, Palazzolo, Pereira Mouriès and Rodriguez have no conflict of interests to declare. Dr. Krishan was Immediate Past President of the SOT Food Safety Specialty Section (2020-2021) at the time of this workshop and represented the SOT Food Safety Specialty Section during the event. Dr. Satchell has significant financial interests in *Situ Biosciences*, a contract research organization that conducts research unrelated to this work. She holds patents on use of bacterial toxins as therapeutics.

PB is an employee of Syngenta Crop Protection LLC, a crop protection company. EI And LP are employees of BASF, a crop protection and chemical company. CK and AS are employees of Bayer CropScience, a crop protection company.

About HESI:

HESI is a non-profit institution whose mission is to collaboratively identify and help to resolve global health and environmental challenges around risk assessment and safety through the engagement of scientists from academia, government, industry, NGOs, and other strategic partners. The **HESI PATB Committee** is dedicated to advancing the scientific understanding of the relevant parameters defining allergenic proteins and protein toxins, to help establish scientifically sound methodologies for evaluating the potential allergenicity and/or the potential toxicity of novel proteins in the context of safety assessments of novel foods and feeds. Within PATB, the ***Protein Toxins Task Force*** has been established with the objective to address the later component of PATB's mission.

About the Food Safety Specialty Section (FSSS) of the Society of Toxicology (SOT):

The Food Safety Specialty Section (FSSS) is a specialty group within the Society of Toxicology; it was formed in March, 1993 to provide a forum for the interaction of toxicologists and other professionals involved in food safety. The purpose of this specialty section is to provide a vehicle where state-of-art research involving food safety and regulations can be communicated and to serve as a scientific resource for critical issues involving food safety.

Acknowledgements:

The authors gratefully acknowledge the contributions of all speakers, the input from workshop participants of all sectors and geographical areas for enriching the discussions, as well as the SOT F3S for their partnership in organizing this meeting. The authors also thank former HESI staff, Ms. Carolina Morell-Perez, for her support at various stages of the workshop planning.

Supplementary materials (see "Appendix" on next page)

References:

Codex Alimentarius Commission, 2009. Foods derived from modern biotechnology. 2nd ed. Joint FAO/WHO Food Standards Programme. Rome, Italy: 1-85.

Collier RJ, Young JA. Anthrax toxin. 2003. Annu Rev Cell Dev Biol. 19:45-70. doi:10.1146/annurev.cellbio.19.111301.140655

Doxey AC, Mansfield MJ, Montecucco C. 2018. Discovery of novel bacterial toxins by genomics and computational biology. *Toxicon*. 147:2-12. doi:10.1016/j.toxicon.2018.02.002

Dutta S, Mazumdar B, Banerjee KK, Ghosh AN. 2010. Three-dimensional structure of different functional forms of the Vibrio cholerae hemolysin oligomer: a cryo-electron microscopic study. J Bacteriol. 192(1):169-178. doi:10.1128/JB.00930-09

EFSA Panel on Genetically Modified Organisms, 2006. Guidance document for the risk assessment of genetically modified microorganisms and their derived products intended for food and feed use by the Scientific Panel on Genetically Modified Organisms (GMO) [1]. EFSA Journal 2006; 4(7):374, 121 pp. doi:10.2903/j.efsa.2006.374

EFSA Panel on Genetically Modified Organisms (GMO); Scientific Opinion on Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal 2011; 9(5): 2150. [37 pp.] doi:10.2903/j.efsa.2011.2150.

EFSA Panels on GMO and AHAW; 2012. Scientific Opinion on the Guidance on the risk assessment of food and feed from genetically modified animals and animal health and welfare aspects. EFSA Journal 2012; 10(1):2501. [43 pp.] doi:10.2903/j.efsa.2012.2501

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), Naegeli, H, Birch, AN, Casacuberta, J, De Schrijver, A, Gralak, MA, Guerche, P, Jones, H, Manachini, B, Messéan, A, Nielsen, EE, Nogué, F, Robaglia, C, Rostoks, N, Sweet, J, Tebbe, C, Visioli, F, Wal, J-M, Eigenmann, P, Epstein, M, Hoffmann-Sommergruber, K, Koning, F, Lovik, M, Mills, C, Moreno, FJ, van Loveren, H, Selb, R and Fernandez Dumont, A, 2017. Guidance on allergenicity assessment of genetically modified plants. EFSA Journal 2017; 15(5):4862, 49 pp. https://doi.org/10.2903/j.efsa.2017.4862

EFSA Panel on Genetically Modified Organisms (GMO); 2010. Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. EFSA Journal, 8: 1700. [168 pp.]. https://doi.org/10.2903/j.efsa.2010.1700

Ganguly S, Mukherjee A, Mazumdar B, Ghosh AN, Banerjee KK. 2014. The β-prism lectin domain of Vibrio cholerae hemolysin promotes self-assembly of the β-pore-forming toxin by a carbohydrate-independent mechanism. J Biol Chem. 289(7):4001-4008. doi:10.1074/jbc.M113.522284

Gavin HE, Satchell KJ. 2015. MARTX toxins as effector delivery platforms. Pathog Dis. 73(9):ftv092. doi:10.1093/femspd/ftv092

Gupta, S., Open Source Drug Discovery Consortium, et al., 2013. In silico approach for predicting toxicity of peptides and proteins. PLoS One, 8(9):e73957. doi: 10.1371/journal.pone.0073957.

Liu, B., et al., 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 47(D1): D687-D692. https://doi.org/10.1093/nar/gky1080.

Mansfield, M.J., et al., 2015. Botulinum neurotoxin homologs in non-Clostridium species. FEBS Lett. 589(3): 342-8. doi: 10.1016/j.febslet.2014.12.018.

Mansfield, M.J., et al., 2019. Bioinformatic discovery of a toxin family in *Chryseobacterium piperi* with sequence similarity to botulinum neurotoxins. Sci Rep., 9(1):1634. doi: 10.1038/s41598-018-37647-8.

Naamati, G., et al., 2009. ClanTox: a classifier of short animal toxins. Nucleic Acids Res., 37(Web Server issue):W363-8. doi: 10.1093/nar/gkp299.

Palazzolo, L., et al., 2020. Literature search – Exploring in silico protein toxicity prediction methods to support the food and feed risk assessment. EFSA supporting publication 2020: 17(7): EN-1875. 89 pp https://doi.org/10.2903/sp.efsa.2020.EN-1875

Peng W, de Souza Santos M, Li Y, Tomchick DR, Orth K. 2019. High-resolution cryo-EM structures of the E. coli hemolysin ClyA oligomers. PLoS ONE 14(5): e0213423. https://doi.org/10.1371/journal.pone.0213423

Postic, G., et al., 2018. KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. Nucleic Acids Res., 46(D1):D454-D458. doi: 10.1093/nar/gkx1084.

Pruitt R. and Lacy D. 2012. Toward a structural understanding of Clostridium difficile toxins A and B. Frontiers in Cellular and Infection Microbiology. 2:28. https://doi.org/10.3389/fcimb.2012.00028

Saha, S., Raghava, G.P., 2007a. Prediction of neurotoxins based on their function and source. In Silico Biol. 7(4-5): 369-87.

Saha, S., Raghava, G.P., 2007b. BTXpred: prediction of bacterial toxins. In Silico Biol., 7(4-5): 405-12.

Satchell KJF. 2015. Multifunctional-autoprocessing repeats-in-toxin (MARTX) Toxins of Vibrios. Microbiol Spectr. 3(3):10. 1128/microbiolspec.VE-0002-2014. doi:10.1128/microbiolspec.VE-0002-2014

van Ree, R., et al., 2021. The COMPARE Database: A Public Resource for Allergen Identification, Adapted for Continuous Improvement. Frontiers in Allergy, 2: 39. https://doi.org/10.3389/falgy.2021.700533

Zhang, S., et al., 2018. Identification of a Botulinum Neurotoxin-like Toxin in a Commensal Strain of *Enterococcus faecium*. Cell Host Microbe, 23(2): 169-176.e6. doi: 10.1016/j.chom.2017.12.018.

Bauman et al. *"From Protein Toxins to Applied Toxicological Testing" Virtual Workshop Identifies the Need for a Bioinformatic Framework to Assess Novel Food Protein Safety*

Manuscript Highlights

- Safety assessment of novel food proteins includes an evaluation for potential toxicity.
- Toxicity is a complex process: involves effectors, delivery systems and susceptible targets.
- *In silico* analysis can identify sequence similarity but contextual information is required for relevance.
- High quality public databases addressing toxin activity are already available and applicable.
- A framework to interrogating these databases is needed to aid interpretation of sequence alignments.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Drs. Doxey, Eberini, Jungo, Kough, Palazzolo, Pereira Mouriès and Rodriguez have no conflict of interests to declare.

Dr. Krishan was Immediate Past President of the SOT Food Safety Specialty Section (2020-2021) at the time of this workshop and represented the SOT Food Safety Specialty Section during the event.

Dr. Satchell has significant financial interests in *Situ Biosciences*, a contract research organization that conducts research unrelated to this work. She holds patents on use of bacterial toxins as therapeutics.

Dr. Bauman is an employee of Syngenta Crop Protection LLC, a crop protection company. Drs. Islamovic and Privalle are employees of BASF, a crop protection and chemical company. Drs. Kessenich and Silvanovich are employees of Bayer CropScience, a crop protection company.