Inventing Prediction for Regulation: The Development of (Quantitative) Structure-Activity Relationships for the Assessment of Chemicals at the US Environmental Protection Agency

Henri Boullier

INSERM, Institut de Recherche Interdisciplinaire sur les enjeux Sociaux (IRIS), France/henri.boullier@inserm.fr

David Demortain

INRA, UMR 1326 Laboratoire Interdisciplinaire Sciences Innovations Sociétés (LISIS), France

Maurice Zeeman

Retired Senior Scientist, U.S. Environmental Protection Agency, USA[1]

Abstract

In policies targeting environmental and health hazards, an effort is frequently made to anticipate and avert more or less probable adverse events. In this context, computerized models are often portrayed as superior knowledge tools, for their capacity to extrapolate from existing data and predict hazards. This paper looks at the historical development and use of such models in regulation, with the specific example of structure-activity relationships (SARs) in the regulation of new industrial chemicals at the US Environmental Protection Agency (EPA). It asks how evidential culture(s) in a regulatory organization change, in particular how new methods and forms of knowledge find their place alongside others to forge regulatory decisions. The development and application of, first, a qualitative approach to structure-activity relationships, and then of quantitative models, show that the EPA had the necessary autonomy to imagine and adjust a method emerging in the research environment to respond to regulatory needs. This can be understood from a coproductionist perspective, if adjusted to take into account the bureaucratic knowledge that mediates the imagining and application of prediction in regulatory practice.

Keywords: Structure-activity relationships, QSAR, SAR, modelling, prediction, Environmental Protection Agency, regulatory knowledge, industrial chemicals, regulation

Introduction

In policies targeting environmental and health hazards, an effort is frequently made to anticipate uncertain adverse events, calculating the probability of their occurrence in the future (Sare-

witz and Pielke Jr, 1999; Nelson et al., 2008)[2]. A number of scientific disciplines have taken on this ambition to predict risks, developing such tools as computational models, and associated software technologies, to do so. In the past decades, these tools and the underpinning practices have expanded, to be more and more routinely incorporated in the processes of risk assessment. Computational knowledge is used alongside experimental evidence and situated observations of risk, to better extrapolate from existing data and predict safety issues. Computational models help bring the future of human risk to bear on present decisions (Adams et al., 2009; Montgomery, 2017; Rajan, 2009).

This paper asks how model-based predictions come to constitute a routine form of knowledge for regulatory agencies. The use of computational models to predict risks is a case of change in the forms of evidence that a regulatory agency uses. This change affecting the way in which an organization knows risk and make decisions is puzzling in itself. First, regulatory agencies are generally constrained to use certain kinds of evidence, under the influence of legal frameworks and of representations of what is credible scientific knowledge. There are norms that define what counts as regulatory knowledge (Demortain, 2017). Second, preferred forms of knowledge tend to institutionalize in the organization. They materialize by roles, identities and boundaries that are difficult to change thereafter. Third, the knowledge that regulatory agencies generate is generally subjected to rigorous trials of credibility and deconstruction games, in courts and in other arenas (Jasanoff, 1990; Hilgartner, 2000). An agency is seldom in a position to impose the validation of new methods or claims itself – validation being the key question when it comes to using models to make predictions (Oreskes, 1998). Because of these circumstances, one can assume that a science-based regulatory organization has a limited capacity to choose and evolve new forms of regulatory knowledge, or to shift from one kind of science to another. How then have computational methods emerged as a form of science-for-policy in an organization marked by high level of constraints on the demonstration of risk? How have these computational methods grown into an element of the evidential cultures practiced in a regulatory agency?

These questions are applied to the case of models of structure-activity relationships (SAR) at the US Environmental Protection Agency. In the past three decades, such models have been developed and applied in greater amounts in EPA's regulatory assessment of chemicals. In SAR, 'structure' refers to the molecular structure of the chemical under consideration, while 'activity' stands for the biological activity, including toxicity that the substance may cause in the body. The correlation between the structure of a molecule and the toxicity it causes in the human body can be established qualitatively by a chemist or toxicologist with experience in toxicity testing, simply by looking at the structure of the substance, identifying within it a particular element or 'fragment' that, in an experimental study that was seen in the past, caused some kind of toxicity. A quantitative structure-activity relationship, or QSAR model, is a statistical correlation between a chemical property that is common to a class of chemical substances (e.g. solubility) and a frequent biological effect of that class, as established in animal experiments. Once a correlation is established, it can be used as a benchmark to infer the potential toxicity of a chemical for which no test data is available, without further (or with limited amount of) animal experimentation, if this chemical has the same structure-related property as the class of chemicals for which the correlation has been established. Such 'in silico' methods (as opposed to studies performed 'in vivo' or on animals[3], and those conducted 'in vitro', on cultured cells) are often presented as an alternative to animal experiments because modelling is future-oriented, predictive, and is not affected by uncertainties surrounding extrapolation of results in animals to future, human conditions. It is the US Environmental Protection Agency (EPA) that early on invested most resources in the development of SAR tools and models, turning it into an accepted and credible way of knowing the hazards of new industrial chemicals, now used across the world and particularly in Europe, as Thoreau and Laurent show in this very issue (Thoreau and Laurent, forthcoming). The agency formalized these methods as part of the implementation of

the Toxic Substances Control Act (TSCA) – an act that did not foresee the use of computational methods, but only authorized the agency to conduct or request "epidemiologic studies, serial or hierarchical tests, in vitro tests, and whole animal tests" (Anonymous, 1976: 2007)[4].

The history provided in this paper shows that the agency did not only import quantitative modelling from the outside, namely from the pharmaceutical industry, but actively constructed an original, qualitative technique of anticipation of risk based on the consideration of molecular structures by experts. It made use of quantitative models, derived from large, validated sets of experimental data produced in house only gradually, after designing this qualitative way of making structure-activity correlations. So, predictive modelling gained credibility as a regulatory tool very progressively and only because the agency designed an initially limited, contextual use of predictions to prioritize substances, as a complement to other kinds of data and studies used to assess risks. This invention can only be understood if we take into account the organizational context by which the emerging scientific order – the supposed capacity to predict safety quantitatively, without experiment — meets and interacts with the political or legal order — the requirement to make decisions without data imposed on the EPA by the law-makers.

This research makes use of three different sources of information. The first are the official numerical archives of the EPA, from which we retrieved several dozens of documents produced by the agency about structure-activity relationships. Second, we interviewed nine officials of the agency who were, or still are, in charge of structure-activity assessment in its Office of Toxic Substances (OTS), or of research on structure-activity relationships on the side of the Office of Research and Development (ORD). Third, we relied on the deep knowledge of the agency and of relevant events, methods, decisions and people of one of us (Maurice Zeeman), who was in charge of the Environmental Effects Branch (EEB) in the Office of Pesticides Pollution and Toxics between 1988 and 1997, with responsibility for the supervision of the group of QSAR specialists of this office. In the remainder of the paper, we begin by giving some background concerning today's importance and use of SAR. We then move to the history of what happened at the EPA during the 1980s and 1990s, before engaging in a discussion and analysis of this history, bringing out the forms of bureaucratic knowledge pertaining to the EPA, which explain how it could shape and apply structure-activity thinking in regulatory work.

Evidential cultures and the regulation of risk

The rise of modelling as a practice underpinning judgement about chemical hazards is a problem in regulatory knowledge that concerns the formation and use of particular bodies of evidence to justify a decision and intervention. There are different ways of producing evidence of a risk, and of the need for intervention, just like there are different ways of producing a proof in any "pure" scientific discipline or in fundamental research. One way to characterize this conflict is to distinguish between the various ways of producing regulatory evidence of a risk, or 'evidential cultures' that coexist or compete within the same regulatory regime to constitute the norm of objective regulatory knowledge, particularly in controversial policy environments (Böschen, 2009; Demortain, 2013). The notion of 'evidential culture', first articulated by sociologist Harry Collins (1998) as part of his ethnographic study of gravitational wave research, broadly refers to strategies and criteria that frame the collective validation of knowledge.

For our purpose, we will draw from the framework advanced by Böschen (2009, 2013), who distinguishes among four different evidential cultures in chemical regulation — restrictive, holistic, instrumental and evaluative. According to Böschen, a restrictive evidential culture, first, rests primarily on experimental methods and on the possibility, in controlled laboratory settings, to verify toxicity at a given endpoint in an animal model, and establish causality between a dose of chemical and that endpoint. That culture takes form in the context of a regulatory regime which names and concentrates on these endpoints and individual chemical objects. The ambition to establish proofs for such causalities, and the high evidence threshold[5], has its drawback, namely the

reduction of the phenomenon being evaluated (e.g. toxicity in a human population is reduced to toxicity measured in a limited population of rats). In the holistic culture that is typical of ecotoxicology, experiments may be combined with other tests and knowledge evaluating the phenomena at other scales of biological organization. The interest is less in discovering simplified causal relationships than in capturing complex interactions between elements of an ecosystem – which means that this culture is less operational, and its agents less frequently consulted for making individual regulatory decisions on a given chemical, than for framing, delimiting an issue in the first place. The third, instrumental culture is oriented towards the development and use of instruments to detect and produce data to estimate the existence of a problem in a new context (embodied by analytical chemistry in the case of environmental chemicals). The disciplines of environmental medicine, or the exercise of hazard assessment, embody a fourth, evaluative kind of culture, where the epistemic goal is less to explain and precisely predict – the evidence threshold is not high – but to address practical problems as they arise, evaluating them against the background of other problematic situations to determine a level of response (Böschen, 2013: 77-80).

Historians, philosophers and sociologists of science alike agree on the fact that modelling is a practice of mediating between experimentation and theory, of creating a fit between the data that surge from experimentation or observation, and available theory. They allow "experimenting on theory" (Dowling, 1999: 261) with constantly renewed sets of data (Morgan, 1999), gluing one and the other (Sismondo, 2006) to be able to use conclusions emerging from experiments to learn about untested situations. This is what makes models more or less useful, possibly a more relevant benchmark for evaluating them than "truth" (Box, 1979; Sismondo, 1999; Zeeman and Mayo-Bean, 2009; Wambaugh, 2014). Models thus have a possible role in each culture. A restrictive culture, for instance, incorporates statistical causal models, to establish links between the experimentally measured variables and the tested object. Analogical models are necessary to apprehend the complexity of systems, and contemplate the relationships between parts of this system. Instruments of detection and measurement cannot function outside ontological models and classifications, that define, delimit or demarcate the thing being measured. Finally, an evaluative culture oriented towards the definition of practical solutions, will employ analogical, physical or statistical models to be able to simulate the effects of a given change in the system. The introduction of structure-activity thinking at the EPA means that evidential cultures in use in the organization evolve either in the direction of a restrictive culture (using quantitative models to find correlations between two reductively considered thing, a molecule and a given toxicity endpoint), or an evaluative one (using analogies to produce signals of safety, and justify a pragmatic decision to further test a chemical).

The coproduction perspective developed by Sheila Jasanoff (2004) offers a way to analyze such change. From that perspective, the evidential culture evolves under the influence of two mutually influencing dynamics: changes in what defines the regulatory order, or more simply the regulatory regime that the agency operates; changes in what counts as valid scientific knowledge in the corresponding scientific fields. In other words, changes in the evidential culture of an organization like the EPA is coproduced by two emerging changes in the realm of the law and in the realm of scientific knowledge. The making of structure-activity correlations at the EPA would be the result, from this perspective, of an emerging legal order in which the agency is requested to make decisions about risks in the absence of data, interacting with an emerging scientific discipline of quantitatively predicting toxicity problems in chemicals. Both comforted one another in the constitution of a credible regulatory knowledge culture.

Coproduction, however, does not take place in the abstract, but in concrete organizational conditions, that need to be taken into account to understand the particular kind of modelling that was identified, chosen and encultured at the EPA. In one of the rare references to this organizational dimension in the coproductionist literature, Jasanoff mentions that the "making of institutions" — the emergence of "tried-and-true

repertoires of problem-solving", or "administrative routines" that provide ready-made solutions to political problems and controversies — is one of the "pathways" of coproduction (Jasanoff, 2004: 40; see Waterton and Wynne, 2004 also Hunt and Shackley (1999) had a few years earlier noted that what they call "bureaucratic knowledge" — the heuristic guides, aids or frameworks that help an organization achieve politically feasible and legitimate outcomes — plays an important role in defining what emerges as science-for-policy, at the junction of science and law (that they called, respectively, academic knowledge and fiducial knowledge). Bureaucratic knowledge designates the heuristics that are shared across the organization, to know how to form a final and credible decision of the organization as a whole, beyond and above the boundaries that separate the members of the organization, notably the specialists of various kinds of science, lawyers, political decision-makers (Bijker et al., 2009). It is specific to a regulatory organization, whose main ambition and challenge it is to precisely turn out credible decisions with a variety of knowledge bases and criteria, observable inside and outside the organization. This bureaucratic knowledge, existing or emerging, influences the definition of appropriate science-for-policy. It mediates the interpretation of scientific affordances and legal mandates to orient the definition of what kind of evidence is most appropriate.

In summary, we assume here that the forms of knowledge that are incorporated in the practice of regulatory organizations are the product of an organizational interpretation of legal constraints and scientific capacities; a process in which the bureaucratic knowledge of the organization in charge – its formalized experience of the coordination among participants in the formation of a decision — plays a key role. This bureaucratic knowledge was particularly important in forging a particular kind of structure-activity reasoning in the agency, mostly evaluative, and distinct from the kind of restrictive QSAR that was then emerging in the field of quantitative drug design. We now turn to the history of the development of structure-activity reasoning and computation models in the EPA. The discussion section then returns to the descriptions of these three orders of change — in the legal order, in science and in the bureaucratic knowledge pertaining to the EPA — to explain how structure-activity reasoning has become regulatory knowledge at the EPA.

Qualitative and quantitative structure-activity relationships at the EPA

Preparing for the review of new industrial chemicals without data (1973–1979)

A QSAR model is a statistical analysis (by regression or classification or else) of the biological activity of a group of two or more chemicals that have some structural similarity, as captured through a chosen descriptor of the chemical[6]. The modelling of causal relations between chemical properties and biological impacts is rooted in fundamental chemistry (Crum-Brown and Fraser, 1868; Meyer, 1899; Overton, 1899). The *quantitative* approach towards these correlations was pioneered by a Professor of Chemistry at Pomona College in California, Corwin Hansch, now known as "father" of computer-assisted molecule design[7].

The interest for QSAR modelling at EPA emerged in the mid-1970s, thanks to connections between the agency and this emerging work of computer-assisted drug design. At that time, the passage of the future TSCA was already under discussion. The proposition to have a dedicated status for the control of chemical substances emerged in the 1960s under the pressure of public interest groups, to comprise what will soon be known as the new social or risk regulation: regulatory regimes dedicated not to the control of markets and economic activities, but to the improvement of health, environment and working conditions (Harris and Milkis, 1989). The new Act was designed to cover the kinds of chemicals that were not already regulated via provisions applying to food additives, pesticides or medicines. A whole continent of industrial chemicals, many suspected to be toxic, had escaped the legislation in place (Vogel and Roberts, 2011). The Council on Environmental Quality suggested in a report in 1971 to develop new legislation to cover all of these chemicals, and to generate information about them in the first place, as many were simply not known or registered. It started being discussed

soon after the establishment of a federal environmental agency, the EPA, was decided, in 1970. It was one of the first Acts that the agency would be entirely in charge of, from the start, and applied the sort of holistic perspective that inspired the creation of a dedicated environmental agency.

During the final years of the discussion of the Act — it was finally adopted in 1976, after six years of negotiation — the EPA's newly OTS was starting to realize that it would eventually have to handle the rapid evaluation of large amounts of unknown, new chemical substances. During the negotiation of the Act, the chemical industry succeeded in convincing Congress and the executive to withdraw most of the requirements for mandatory testing from it. At the end of the day, implementing the Act appeared as a challenge for the agency that had to prove scientifically the existence of risks, to regulate products, without any possibility to execute or require scientific studies from the industry, even though tests were available and already routinely applied by corporations of the sector (Craeger, 2018). The officials of the EPA had to very quickly operationalize an approach to deal with the evaluation of new chemicals for which no testing and no data were going to be available (since there was no obligation for companies to do testing). The discussions revolved around the need for "identification" of chemicals and methods for "early warning". By the end of 1973, the EPA's toxic substances staff had already identified a significant body of scientific literature concerning structure-activity correlations and methodologies, without making any clear-cut decision as to the potential which these methods might have for helping EPA in its early warning activities (FRI, 1975, 1976). At the beginning of 1974, the view according to which key properties of substances, notably their toxicity can be derived from their structure was accepted (EPA, 1975).

Structure-activity interests in the EPA's Office of Toxic Substances (OTS) and Office of Research and Development (ORD)

The interest for the modelling of structure-activity correlations crystallized simultaneously in two separate places in the EPA. The benefits of using structure-activity correlations to formulate judgments about the safety of new chemicals became clearer as Joseph Seifter, a medical doctor by training and a pioneering pharmacologist, joined the EPA's OTS in 1978 from the George Washington University Medical School. Seifter had approached structure-activity work as part of his research in pharmacology and drug development. He contributed to establish the so-called 'structure-activity team' (SAT) in the OTS in 1979, to prepare for the incoming of the first new chemical 'pre-manufacture notifications' (PMN). These notifications applied to any new chemical, meaning a chemical substance that was not on the inventory of existing chemicals prepared by the OTS and published in July 1979 (Hepler-Smith 2019). After that date, there could now be a 'new chemical PMN' submitted to OTS's New Chemicals Program (NCP) by industry.

The interest for QSAR modelling also crystallized in another place in the EPA, the Office for Research and Development (ORD)[8]. Gilman Veith, a scientist in one of the laboratories of the ORD, the Environmental Research Laboratory (located in Duluth), got interested as early as in the mid-1970s by this notion of prediction of toxicity from chemical structures, and devoted enormous efforts to the development of that science[9] (Veith, 1981). In 1975, at about the same time as the EPA's toxics staff started to conceive of SAR as a possible approach to deal with the chemicals 'data gap' in the upcoming TSCA, he initiated the 'QSAR research program' of his laboratory (Bradbury et al., 2015: 17). He developed a clear vision of the necessary tools for the QSAR approach to become applicable and had the necessary leadership skills to have people work together, both in his lab and between his lab and the EPA's new OTS, to develop these tools.

One of the early projects that Veith and his group launched was the development of high-quality databases of experimental results, on which to compute correlations and estimations of toxicity. Two databases were developed. The first came to be known as ECOTOX, and was essentially a collection of experimental results presented in scientific journals. The second developed from dissatisfaction with literature-derived databases. The results collected from the literature are never fully comparable: even where two experimenters

test the same substance using a similar protocol (e.g. administration of a 50% concentrated dose of a substance to rats for 28 days), these protocols necessarily differ in some dimensions. The resulting database of toxicity measures is not of sufficient quality to compute robust mathematical correlations with chemical structures. Veith thus came to the conclusion that he would need to do the experiments at home, replicating a strictly identical protocol on a large number of substances, and collecting the result in an operational database. To do so, a large testing program was set up in 1981, in close cooperation with, and funded by, the OTS. It consisted in performing a short-duration toxicity test on a fish (96 hours, on fathead minnows), to derive the LC50 value of several hundred of chemicals.

3.3 Structure-activity in the regulatory practice of OTS

Only eight PMN were submitted to OTS in 1979, but their numbers grew quickly. There were almost 1,000 PMNs submitted to OTS by September 1981. 900 more PMNs had been submitted by September 1982, and another 1,400 PMNs by September 1983. The PMN process consequently resulted in an average of about 1,600 submissions per year, in 32 years of application[10].

The Act imposed a 90-day limit to the agency but did not impose the industry to provide any data to the EPA for it to do that estimation (other than data the company already has). By law, a PMN submission dossier includes the name of the chemical, a description of its structure, the production volume, methods of uses and disposal, estimates of human exposure, and any extant test data obtained. Nevertheless, in approximately 65% of cases, submissions by the industry did not include any substance-specific experimental data. The information the EPA got was the name of the chemical, a description of its molecular structure, the volume of production, the uses and disposal methods, and estimates of the number of people in the general population that will be in contact with the substance ('human exposure'). Only 45% of the dossiers included health test data, mainly for acute toxicity endpoints, genotoxicity test results or local irritation studies (EPA, 1984). There was little, if any, ecotoxicological or physical/chemical fate data submitted (e.g., Auer et al., 1990; Auer et al., 1994; Zeeman et al., 1993; Zeeman, 1995; Zeeman et al., 1995). If the EPA wants to get test data from the manufacturer, it has to make a risk-based case for it, an obviously difficult thing to do with minimal information in hand or, in regulatory science terms, in a 'data-poor' situation. The sheer volume of substances to assess, coupled with the absence of data, rendered the perspective of making predictions from structure-activity correlations in similar chemicals, particularly attractive, if not a necessity.

The first step in the PMN process designed by the OTS was determining that all necessary information has been included in the notification. This was followed by a series of three meetings (1. Chemistry Review and Search Strategy; 2. Structure Activity Team; 3. Exposure Analysis Meeting) which bring senior level expertise to bear on the questions of chemistry, hazard, and exposure within the first 15 days of the 90-day period available to EPA for the assessment of each new chemical. The 'Chemistry Review' meeting would be held between from days 8–12. 30% of substances, on average, would be left off the hook at this stage[11] during which the chemical identity of the substance is considered, the methods by which it is synthesized and the feedstocks used for the process, the physico-chemical properties of the substance. The remaining 70% would then be considered in the 'structure-activity meeting' between days 9 and 13 of the process.

The structure-activity work was prepared within the Health and Environmental Review Division (HERD) of the OTS, which provided the scientific and technical support for chemical assessment[12]. In the terminology of the paradigm set in the National Research Council report on risk analysis (NAS, 1983), then being institutionalized in the agency (Demortain, 2019), the structure-activity work contributed to the first step of the process, 'hazard assessment' – that is, the mere identification of toxicity problems, without measurement of their gravity, frequency or probability of apparition (the heart of the exposure assessment and risk characterization stages) (Zeeman and Gilford, 1993).

In structure-activity meetings, one chemist was assigned the task of summarizing the profile

of the substance. The physico-chemical properties were then considered, followed by the environmental fate of the substance, the health issues (metabolism, mutagenicity, neurotoxicity…), to conclude by consideration of the ecotoxicity of the substance. Inferences from what was known of the properties of a given chemical structure infused the work of the whole team.

Structure-activity meetings were not a meeting of modelers discussing mathematical models and numerical estimates. These were "professional judgment" meetings (EPA, 1984: 4), where people used their experience of the toxicity typically associated with a kind of chemical structure, to anticipate the safety issues that might arise from exposure to a new chemical with a structure that they deemed comparable, or sufficiently similar, to those for which they had prior experience of toxicity. These meetings served to elicit the views of experts. They were patterned, in effect, after the Delphi method — a collective forecasting method based on successive rounds of questioning of a group of experts, developed by the US Army forces — in which OTS people found inspiration. Auer, who headed the structure-activity team from 1979 to 1986, described it in this way:

> We had the first meeting. You have the [chemical's] structure and you have the little or no data that were available on it, and you just started going around the room. What do you think? You know, in your area of expertise, what can you offer about this chemical? Over time that evolved into a very regularized approach to decomposing the chemical, and then through SAR [structure activity relationships], putting it back together to tell the story. […] Pretty quickly, within probably the first year of the operation of this program, you had a regimen in place where you had done preliminary chemistry analysis. So, what kind of chemical is it? How does [the chemical] function in its use? … (CHF, 2010).

Of all the necessary resources to do QSAR modelling (availability of test data on the substance being examined, data on analogous substances, statistical methods to analyze them and 'professional judgment'), the knowledge and professional judgments of scientific assessors in the interpretation and integration of available information, was "the most critical in terms of the overall success of the evaluation effort" (EPA, 1984: 13). In regulatory practice, then, modelling was a mode of reasoning applied to accumulated experience, to form a hypothesis about the lack of safety of a substance in anticipation of any experiment or observation. This was, in essence, a qualitative kind of structure-activity analysis, based on knowledge gained from reading masses of experimental data published in the literature, an experience that toxicologists classed by chemical categories, themselves defined by molecular structures. This knowledge was deposited in people, and exercised by them during meetings in what became a particular kind of competence in making analogies between substances. Auer, again, recollects that

> there were smart people on the team who could say 'Jeez, this substance looks a lot like that case we had a year ago'[13]

such as Joe Seifter, recalled as

> one of the early practitioners of the concepts of forming categories of chemicals [and] looking for ranges of toxicity across a category, being sensitive to where the toxicity shifted in a category, and then, attempting to understand mechanistically what was going on to cause that shift. He was just a remarkable guy, encyclopedic knowledge. You could show him a structure and he could just tell you what kinds of things it was likely to do to a human. (CHF, 2010).

Charles Walker was another of these experts that came to the EPA from the *U.S. Fish and Wildlife Service* to help apply this practical analysis (Lipnick, 1998). According to their colleagues, the scientists of the OTS, Robert Lipnick, Richard Clements and Vincent Nabholz in particular, were said to have develop a great ability in that exercise of inferring possible toxicity issues from reading chemical structures over time. More people soon joined the team, extending this repository of embodied knowledge of structures and toxicity that compensated for the absence of experimental data in industry notifications. Paul Bickart, a Harvard-trained chemist with a very broad background, contributed his capacity to characterize chemi-

cals. Joseph Arcos, a university-based chemist had joined the SAT in 1979 already, bringing his vast knowledge of chemical carcinogenesis to the team. Adrian Albert, an Australian professor in medicinal chemistry, spent the summer of 1982 working with the SAT too.

Structure-activity considerations were not equally useful for all kinds of toxicity and endpoints, however. They were mostly useful for the environmental fate or ecotoxicological issues, for one simple reason: these were the issues for which experimental results or what Auer calls a "base set of data[14]" (generated through acute tests, such as fish tests) could be generated more quickly, assays being short and relatively less expensive than the tests on rodents used for human health outcomes. This analog chemical assessment was *not* the basis of final regulatory decisions. The final decisions resulted from a more complete risk assessment process and the vast majority of such decisions could be made without any testing needed. Some of the more difficult decisions required the consideration of data produced by the company *after* the initial structure activity meeting, and following the indications of the structure-activity team. But chemical analogues served to anchor the assessment that problems may arise from exposure to that substance. It was sufficient to meet the standard of proof established in the Act to justify requiring data from a company: that is, that it "can reasonably be determined or predicted" that the substance "may present an unreasonable risk of injury to health or the environment" (Anonymous, 1976: 2006).

3.4 Criticism, doubts and progress towards quantitative SARs

Although the structure-activity team grew over time, as notifications started to pour in, from around six to a dozen people, the (Q)SARs approach was not consensual. Skeptics of structure-activity inferences could be found either in the agency or outside, among academics of corresponding fields for instance. Several reviews of EPA work on structure-activity emerged just a few years after the initiation of the structure-activity team. EPA's use of SAR in reaching PMN hazard assessment conclusions soon started to be questioned by Congress, environmental groups and others (OTA, 1983; GAO, 1984; ACS, 1984) who point out the many uncertainties associated with the approach (EPA, 1984).

Adrien Albert, a professor at Australian National University specialized in structure-activity relationships who participated in several of the team meetings in 1982, undertook a review of the work of the EPA on structure-activity correlations. In his report, he noted that the EPA heavily relied on the professional expertise of scientific assessors, and on the exercise of relating a whole molecule to a class of chemicals for which adequate biological data exist, much more than on quantitative structure activity relationships, which was limited by the lack of toxicity data and was based on physical chemical property data, and QSAR descriptors (EPA, 1984).

At about the same time, a paper appeared in the journal Environmental Science and Technology, quoting Corwin Hansch questioning the use of SARs in a regulatory environment. The pioneer and "father" of quantitative SARs thought that the approaches employed by the EPA differed from his own for two reasons. Contrary to their applications in the pharmaceutical industry, the EPA did not focus on a single endpoint but multiple pathways, which made the objectives and contents substantially different. The lack of data also made the EPA approach very different: both experimental data for the health effects of PMN chemicals, but also more basic data of physical chemical properties were missing. As a consequence, Corwin Hansch judged that EPA's approach was useful but insufficient:

> While SARs can be very helpful to regulatory agencies in deciding which chemicals should be subject to special testing—EPA is doing this now—I believe that you cannot yet base regulations on SARs. In other words, SARs are not yet ready to use for confirming or denying market access to any given chemical, but they are of use, and are being used to guess which may be especially toxic or relatively safe (Anonymous, 1984).

Debates emerged as to whether the chemicals between which comparisons were made were sufficiently analogous, but also about the extent to which SARs could ever predict other forms of

toxicity than acute toxicity, for which there was more biological data available than the rest. Criticism over EPA's PMN review process also stemmed from the Government Accounting Office and the Congress' Office of Technology Assessment, for which uncertainties were pervasive in toxicity assessment and in experimental tests "on the exact product, not closely related chemicals, are necessary to ban or restrict production" (OTA, 1984: 77).

These reports, while critical of quantitative structure-activity modelling, were not altogether depreciative of the use of analogies between groups of chemicals to screen large numbers of chemicals, and select those on which more data would be requested, and that would undergo a closer review. They also confirmed that the approach was promising, and that it could be refined and reinforced in order to go beyond its initial uses. This is what happened subsequently in the OTS and in the lab of Gilman Veith. Besides routine evaluation work in the structure-activity team, the "QSAR folks" of the OTS, as they were sometimes called, put forces into the development of quantitative SARs and in tools to develop them faster. Robert Lipnick, a chemist turned "QSAR scholar" (Lipnick, 1985, 1991) started using QSAR models as early as in 1981. With the support of the chief of the Environmental Effects Branch (EEB) of the OTS, he set out to compare the results of the screening of 55 alcohols with a known QSAR model for narcosis that was published in the literature. The model allowed generating a value for the level of narcosis that predictably occurs for certain chemical structures. Comparing the alcohols with these chemical structures, a possibility emerged to actually say whether these alcohols would themselves produce narcosis (Lipnick et al., 1985). Lipnick was known as a more theory-oriented person, very much interested in researching and validating models and applying them to new substances.

Two other scientists in the EEB worked to develop quantitative structure-activity models. Despite this actual lack of ecotoxicological (and chemical fate) test data, they worked continuously for several years to increasingly develop and to then make use of many individual SAR and QSAR estimates. This team had managed to develop 13 QSAR models by the early 1980s (see Clements et al., 1993). By 1988 there would be 49 different QSARs for estimating the aquatic toxicity or bioconcentration potential for about 30 classes or subclasses of industrial chemicals produced (Zeeman et al., 1993), the majority of which by EEB scientists. All of them were published in the so-called QSAR manual in 1988 and 1994 (over 120 QSARs by then). Structure activity work could be refined by intense collaborations between ORD and OTS. ORD's "fathead minnow studies" continued in parallel and eventually covered 617 industrial chemicals in total[15]. The programmatic offices extended funds for the ORD to develop its databases, with the ORD making these results accessible to their scientists so that they could identify structure-activity correlations and computer programs for each correlation, to generate predicted toxicity values. Thanks in part to this database, yet more quantitative models were produced[16], all of which fed the structure-activity meetings.

A more systematic use of quantitative SARs models implied some tension among scientists of the EEB. Some of them, such as Robert Lipnick, seemed not willing to accept the use of QSARs that had not been somewhat rigorously evaluated. Others, like Vincent Nabholz or Richard Clements, were more the reluctant sponsors of creating and using whatever QSARs they found or developed that were able to provide them with some of the numerical ecotoxicity answers that were needed for use under the circumstances of the OTS PMN review process. The EEB's leadership managed the opposition by functionally separating these scientists that were more oriented to the pragmatic use of QSARs in regulatory practice, from the more theoretical QSAR folks, who were actually getting in the way of efficiently performing the ongoing regular NCP chemical assessments via QSAR. The regulation-oriented scientists got major control of the QSAR hazard assessment process of new chemicals, while the more theoretical-inclined ones were allowed to invest in the publication and further development of models.

These tensions were not necessarily sensed outside the EEB. The new chemicals program was not subject to intense political and legal scrutiny. Auer recalls that in normal days, as director of OTS, he would only meet with the people in charge of

PMNs three or four times a year, which denotes a low priority[17]. This relative protection of the space of EEB, coupled with the support from its hierarchy and the commitment of its scientists to improve structure-activity inferences, meant that progress towards increased quantification of SARs and consideration in regulatory evaluation of dossiers, did continue. Vincent Nabholz and Richard Clements made a continuous effort to revise EEB's structure-activity correlations, as valid new data were found. They assembled their models into the various QSAR Manuals (EPA, 1988; 1994b) and also created ECOSAR. The initial one was an internal notebook of 13 QSARs and it could be considered as the first QSAR Manual (EPA, 1988[18]). The second manual had 49 QSARs and it was published by OTS as an EPA document (EPA, 1988). The hope of the OTS, at this point in time, was to make the tool circulate. A revised version of the QSAR manual was completed in 1993, and this one constituted an important turning point. It was published by the toxics office as an EPA document (EPA, 1994b), and it stated that it was published to accompany the EPA release of the SAR software program called ECOSAR (see EPA, 1994a: 1). Until 1993, the QSAR manual had only been available in paper version. In 1994, the ECOSAR, which was a PC version of the manual, was then made available to the public. This revised version contained 42 chemical classes along with 120 QSARs.

In the 1990s, structure-activity work on hazard assessment still progressed in these circumstances, although the NCP resources declined[19]. After 1996, there was a hiatus in that EPA QSAR development. The EEB was dissolved in the 1997 reorganization of OPPT. The old EEB staff were then distributed amongst the various new Product Line Branches in the newly organized Risk Assessment Division[20], and thus many of the QSAR staff were now basically spread out and were basically on their own, with less support from management. The investment in SAR/QSAR continued, but those scientists that had been instrumental in its development no longer had a direct access to a management support system.

By that time, however, the New Chemicals Program, including in its use of QSAR, had been hailed a success by the chemical industry in the US. The review process had been formalized, with highly detailed manuals, guidelines and software tools, that helped understand how the process unfolded, and how each sort of information – chemical, toxicological or exposure-related – was used (EPA 1997). Many of the developments of the EPA/OTS of the preceding years were taken up in the OECD. When the OECD initiated the harmonization of SAR tools and models, it first undertook a validation exercise, to verify the accuracy of predictions that were made by the EPA staff. The conclusion of that exercise was that the SAR methods of the EPA "performed extremely well in predicting acute toxicity to fish and daphnia" (OECD 1994, cited in Zeeman, 1995: 712). In general, model-based predictions generated estimates of toxicity that were within an order of magnitude of those that were observed in animal experiments. The OECD validation of EPA's predictions was very much an endorsement of the work performed there since the 1970s to develop these tools, the "careful development and analysis of chemical categories", "the thoroughness and diligence in adding new data points to established categories", the high level of "refinement" of its predictive capabilities (OECD, 2007a: 28). Several important tools for SAR work, notably the concept of 'chemical category', were taken up by the OECD, recognizing the usefulness and applicability of what the EPA had developed internally (see EPA, 1993; cited and described in Zeeman, 1995, OECD, 2007; OECD, 2009).

Organizing the coproduction of science and law

The above history shows that the EPA did a lot around structure-activity in the area of chemicals hazard assessment, much more than any other organization, and earlier than anyone else too. It conceptualized the use of structure-activity correlations and judgments as a method to evaluate the hazardousness of chemicals, accumulated the experience in making such judgments, practicing them day by day on a very large number of products over a long period of time, integrating those judgments into a concrete decision-making machinery. It invested in a massive testing program to generate a database of experimental results, to perform the statistical analysis neces-

sary to the production of more models. It formalized and put into circulation several important tools for other organizations to be able to use these estimations or develop new ones in turn. Altogether then, the EPA brought much credibility to the translation of an initial chemistry theory into the practical regulatory assessment of chemicals.

The striking aspect this history is that computerized models were not abruptly introduced as a regulatory tool, and certainly not approached as a tool for definitive decision-making through prediction. Rather, it is structure-activity thinking that was first introduced as an element of knowledge in the gradual formation of decisions, buying time for the more progressive development and subsequent consideration of validated statistical models in the regulatory process. The agency initially fell back on a more modest approach to knowing and predicting the future, embodied by qualitative, non-statistical structure-activity thinking. It practically restricted the validity of the approach of establishing relationships between structure and activity, to the screening of large sets of substances, and to the triaging of those that needed further investigation from those that could be deemed reasonably safe. Making precise, quantified predictions on any one of these substances was a horizon, and promise, that EEB scientists and regulators in the Office of Toxic Substances gradually came closer to[21].

Several elements concur in explaining why the EPA emerged as a site of formalization of a new kind of regulatory knowledge, and of invention of ways of applying structure-activity thinking. Based on the discussion of science-policy coproduction dynamics in the first part of this article, it appears that an evidential culture takes form, first, in response to the political stabilization of criteria of decision and proof. An evidential culture is not made of "research science", but of this particular kind of science that is believed to be appropriate to inform a decision criteria (a given definition of what counts as a risk) and a standard of proof set in the law. Those criteria are defined not by the regulatory organization itself, but emerge from the power relationships among the actors that take part in the construction of the regulatory framework, and in its implementation and subsequent evolution. From this perspective, an evidential culture changes where and when risk criteria and standards of proof evolve, under the pressure of principals (Congresspersons that design the Act that the agency must implement), courts (who review decisions and confirm or change the actual criteria of safety that the agency is supposed to apply) and/or of the regulated industry.

In that case, it seems clear that an evolution in the ways of establishing a proof of the existence of chemical risks was in order, given the particular regulatory design of TSCA. As the first leader of the structure-activity team recalled ex-post facto, "necessity is the mother of invention" (CHF, 2010: 4): the EPA had no alternative in the face of the double-bind in the implementation of TSCA — an ambitious mission to rapidly review a great number of different chemicals, with limited scientific and legal means to do so. Computerized QSAR models and expert systems simply filled the gap left after the Act was emptied of any requirement for testing (Mayo et al., 2012; Craeger 2018). The particularly large number of substances to review created a strong pressure to apply new methods, even though reliable QSAR models were not yet in sight. Structure-activity thinking thus became the immediate solution, even though it was a new, embodied expertise only practiced by a handful of specialists who the EPA had managed to recruit or attract. At the same time, the criterion that was chosen to define safety (an "unreasonable risk" of injury), coupled with a low evidentiary threshold (the Act authorizes the agency to act if there is a "reasonable basis" to conclude about the existence of an "unreasonable risk") meant that the agency could use emerging, judgmental (as opposed to formal and quantitative) methods in its work. No one among courts, environmental groups or regulated businesses, contested the interpretation that the agency made of this criterion. In other words, the agency was both constrained and given some autonomy to search for new methods to document the risk. The context gave weight to non-testing approaches emerging in the world of pharmaceuticals development.

Second, an evidential culture is built on a set of methods, and representations of what can be known, and what may not or is not interesting to know. In this sense, changes in evidential cultures

depend on the production and availability of new research outside the organization, and on the capacity of the organization to translate and incorporate this research into its own expertise for its needs. What is noticeable in the present case is the network of relationships between the agency and external scientific groups, and the capacity of the EPA to attract experienced scientists from the world of computational drug design, and even more so, of qualitative structure-based toxicity prediction. Another noticeable aspect is that there were protected spaces in the organization, in which the necessary work to make modelling function could be envisioned and deployed. The group of QSAR folks inside OTS, alongside the ORD's ERL in Duluth and private scientific service companies, which the EPA intensely used, all dedicated to bettering the approach. Those different spaces were inter-linked by a network of QSAR people who consistently cooperated, sometimes in productive tensions, e.g. over the relative importance of the administrative imperative of availing of usable tools for deciding on substances, or screening them, and the more scientific, long-term ambition to produce reliable, fully validated statistical models. The OTS had among its staff the necessary level of scientific skills to manage these collaborations and profit from external productions. This means that the Office was, almost from the start of its existence, in a position to understand the challenges and difficulties of modelling, and aware of the technical developments to perform before embracing models fully.

One should add another factor, namely the autonomy of the new chemicals program. The very design of the NCP program in TSCA was a recipe for failure (thousands of substances to review, with little possibility for EPA to neither request data nor obligation to the industry to provide some), and the EPA leadership did not expect much from this program to start with. The New Chemicals program was also much less of a threat for companies than the Existing Chemicals program[22]. So, both the legal challenges and the political supervision from the higher echelons of the organization were limited, granting autonomy to the people inside the OTS to forge common rubrics of information and judgment, and ways of evolving decisions.

However, coproduction does not occur in the abstract, and structure-activity methods did not emerge spontaneously. The concrete form that modelling and prediction took in the agency was not quantitative modeling to start with, but a collective, human judgment about similarities between structures and of toxicity associated with structures. This collective judgement was formed during dedicated meetings, conceived of as a step in a regulatory sequence, organized around a dedicated team. This organizational materialization of structure-activity thinking can only be understood taking into account the autonomous bureaucratic knowledge present in the agency, which mediated both the legal requirements and scientific affordances.

Bureaucratic knowledge, in this case, covers several things: a form of procedural rationality, by which objective decisions are the product of a sequence of judgments formed on discrete bodies of information, applying different criteria (chemical analysis, then structure-based hazard assessment, followed by risk), some scientific, others more readily political. This, in essence, is the heart of the "risk analysis framework" that the agency and a suite of expert bodies started to formalize and apply in those years, faced as it was with massive controversies about the proper use of science, and suspicions about the distortion of evidence by political appointees (Demortain, 2019). There was, in those years, an important kind of bureaucratic experience of what were the correct ways of coordinating scientists and decision-makers in the organization, across the boundary that separates or should separate them (Bijker et al., 2009). EPA's specific bureaucratic knowledge also includes the notion that, in the presence of enduring uncertainties and disagreements — uncertainties that no single method or discipline could lift — regulatory decisions would only seem objective if they were based on a set of converging expert judgments, a style of decision formation that was captured in the Delphi method, that inspired the people of the OTS. It covers the infrastructural knowledge of individual chemicals, of their uniqueness and mutual resemblances, accumulated in people who make up this

"molecular" bureaucracy (Hepler-Smith 2019). Bureaucratic knowledge is, finally, the experience of the credibility of this mode of making decisions in the interaction with audiences that evaluate the agency – scientists, courts and regulated businesses among others, as opposed to the making of decisions based on non-validated quantitative models. Each of these elements demonstrate the existence of an autonomous bureaucratic knowledge of the people of the Office of Toxic Substances, that decisively influenced the interpretation of the constraints posed by TSCA, and the capacities emerging from the scientific world, to give shape to, and anchor, an original and credible way of making structure-activity correlations.

Conclusion

In this paper, we have looked into a case of formalization of a kind of predictive knowledge in policy and regulatory practice. The main aspect of the history of the use of structure-activity thinking and predictions at the EPA is that it was a history of experimenting, ascertaining the value of this kind of knowledge for implementing a complex, challenging public program of reviewing hundreds of substances at once. Structure-activity predictions at the EPA was a developing discipline throughout the period, only getting to validated models after applying a qualitative form of structure-activity method, and investing in the gradual development of a database of good enough quality to produce reliable, usable quantitative models in a not too distant future.

Today, the development and use of models like structure-activity relationships in chemical regulation has become almost systematic. For many commentators, their adoption is the logical result of recent advances in biology and biotechnology, along with numerous controversies over the relevance of animal models (NRC, 2007). Such interpretations, however, fail to take into account the set of determinants that are necessary to make sticky epistemic and evidential cultures evolve in an organization faced with multiple constraints. It overlooks the particularity of this history: namely the fact that the EPA, as an organization that needs to forge credible demonstrations of the risk to convince audiences that are necessarily critical of its assessments and decisions, first opted more pragmatically for an evidential culture, in which prediction takes for the form of an analogical reasoning that helps define objects that can legitimately be subjected to further review and investigation. It approached the promises of predicting risks that underpin the development of quantitative, computational models of toxicity with great caution.

We have outlined a framework to analyze this particular case, expanding the coproductionist perspective to include an organizational component that seemed crucial to make sense of the particularity of this case: the fact that the change in the way of knowing and proving risks for regulatory decision, emerged from the capacity of the EPA to articulate a new method, adapted to its constraints, and in a sense to innovate. New methods and ways of proving the existence of a risk emerge as a new culture in a regulatory organization, if, first, a regulatory regime emerges in the environment that commands new practices and ways of making decisions. In the present case, the sub-regime of the new chemicals program, with its distinct and relatively weak criteria of decisions and demonstration, delimited a new space of regulatory work. Second, this culture will take form in the presence of a capacity to connect with the research environment to find methods that may respond to regulatory needs and uncertainties. Third, and perhaps most importantly, the methods that are chosen and the form they take – in this case, a qualitative form before a quantitative one – derives from the need to foster a mode of cognitive coordination in the agency. The knowledge that is incorporated in the agency must provide references of the risks and the uncertainties that various members of the decision process can share and build on, sequentially. In this case, we find no hierarchy of knowledge, or dominant standard of proof. No superiority is granted to modelling techniques or computational models, as a tool to handle uncertainties and compensate for the limits of animal experiments. We see an articulation of evidences, invented in situ.

The particular historical case of predictive, structure-activity knowledge at EPA, teaches us

something more general about future-oriented expertise. Foreknowledge, or in this case scientific methods that claim to have a greater predictive power, is specific in this that its credibility is more difficult to establish. It is knowledge of uncertain, yet-to-come objects. Its credibility and value can only be established in the longer run, as and when these objects finally take form and make the demonstration of its relevance. As a form of knowledge that is both highly pertinent for policy-making, yet also less immediately authoritative than more realist knowledge, its production depends on the autonomy of that a given organization can find, to be able to try, verify and evaluate different ways of making decisions. The paradox of predictive organizations is that of being under a lot of constraints, not least to have to face many uncertainties, and of having the capacity to gain autonomy from these constrains, to design forms of knowledge and decision-making. Ways of predicting risks emerge from what science allows and what the law requests, decisively mediated by the knowledge of how to coordinate people and their knowledge to produce a decision, in an at least partly autonomous agency.

References

Adams V, Murphy M and Clarke AE (2009) Anticipation: Technoscience, life, affect, temporality. *Subjectivity* 28(1): 246–265.

Anonymous (1976) An Act To regulate commerce and protect human health and the environment by requiring testing and necessary use restrictions on certain chemical substances, and for other purposes. Oct. 11, 1976. Public Law 94-469, 94th Congress.

Anonymous (1984) Structure activity relationships: Are they predictive of toxicity and potential regulatory tools? *Environmental Science and Technology* 18(9): 285–286.

Auer CM, Nabholz JV and Baetcke KP (1990) Mode of action and the assessment of chemical hazards in the presence of limited data: use of structure-activity relationships (SAR) under TSCA, Section 5. *Environmental Health Perspectives* 87: 183–197.

Auer CM, Zeeman M, Nabholz JV and Clements RG (1994) SAR—the US regulatory perspective. *SAR and QSAR in Environmental Research* 2(1–2): 29–38.

Bijker W, Bal R and Hendriks R (2009) The paradox of scientific authority: The role of scientific advice in democracies. Cambridge and London: The MIT Press.

Bowker G and Star SL (1999) Sorting things out: classification and its consequences. Cambridge and London: MIT press.

Box GEP (1979) *Robustness in the Strategy of Scientific Model Building*. Wisconsin University Madison Mathematics Research Center.

Bradbury SP, Russom CL, Schmieder PK, Schultz TW, Diderich R and Auer CM (2015) Advancing Computational Toxicology in a Regulatory Setting: A Selected Review of the Accomplishments of Gilman D. Veith (1944–2013). *Applied In Vitro Toxicology* 1(1): 16–25.

Boullier H (2019) Toxiques légaux. Comment les firmes chimiques ont mis la main sur le contrôle de leurs produits. Paris : La Découverte.

Böschen S (2009) Hybrid regimes of knowledge? Challenges for constructing scientific evidence in the context of the GMO-debate. *Environmental Science and Pollution Research* 16(5): 508–520.

Böschen S (2013) Modes of Constructing Evidence: Sustainable Development as Social Experimentation—The Cases of Chemical Regulations and Climate Change Politics. *Nature and Culture* 8(1): 74–96.

CHF (2010) *The Toxic Substances Control Act from the perspective of Charles M. Auer*. Oral history program, Chemical Heritage Foundation.

Clements RG, Nabholz JV, Johnson DW and Zeeman M (1993) The Use and Application of QSARs in the Office of Toxic Substances for Ecological Hazard Assessment of New Chemicals. In: Landis W, Hughes J and Lewis M (eds) *Environmental Toxicology and Risk Assessment*. ASTM STP 1179, Philadelphia: American Society for Testing and Materials, pp. 56–64.

Collins HM (1998) The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves. *American Journal of Sociology* 104(2): 293–338.

Craeger A (2018) *Testing Toxics: Tools, Standards, and Corporate Data in the History of Chemicals Regulation. Conference Pervasive Powers: Corporate Authority in the Shaping of Public Policy*. University Paris-Dauphine, Paris, 14-16 June 2018.

Crum-Brown AC and Fraser TR (1868) On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *Journal of Anatomy and Physiology* 2(2): 224–242.

Demortain D (2013) Regulatory Toxicology in Controversy. *Science, Technology & Human Values* 38(6): 727–748.

Demortain D (2017) Expertise, Regulatory Science and the Evaluation of Technology and Risk: Introduction to the Special Issue. *Minerva* 55(3):139–159.

Demortain D (2019) The Science of Bureaucracy. Risk Decision – Making and the US EPA. Cambridge and London: MIT Press.

Dowling D (1999) Experimenting on Theories. *Science in Context* 12(2): 261–273.

EPA (1975) Papers of a Seminar on Early Warning Systems for Toxic Substances. Conference Report. Office of Toxic Substances, US Environmental Protection Agency, Washington D.C. EPA-560/1-75-003. July 1975.

EPA (1981) *State-of-the-art report on structure activity methods development*. Prepared by Environmental Research Laboratory, for the Office of Toxic Substances, US Environmental Protection Agency, Washington D.C.

EPA (1984) *Proposed Design for a Retrospective Study of PMN Health Hazard Predictions*. Memorandum from John A. Moore, Assistant Administrator for the Office of Pesticides and Toxic Substances, to Terry Yosie, Director of the Science Advisory Board, US Environmental Protection Agency, Washington D.C.

EPA (1988) *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships, Volume 1*. Office of Toxic Substances, US Environmental Protection Agency, Washington, DC. EPA-560/6-88-001. July 1988.

EPA (1993) Categories Provide Guidelines for Premanufacture Notice Submitters. *OPPT's Chemicals in Progress Bulletin* 14(2): 35-36. US Environmental Protection Agency, Washington D.C. EPA-745-N-93-001. June 1993.

EPA (1994a) ECOSAR: *Computer Program for Estimating the Ecotoxicity of Industrial Chemical Based on Structure Activity Relationships, User's Guide*. Office of Pollution Prevention and Toxics, US Environmental Protection Agency, Washington D.C. 748-93-R-002. February 1994.

EPA (1994b) *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure Activity Relationships 2nd Edition*. Office of Pollution Prevention and Toxics, US Environmental Protection Agency, Washington D.C. EPA-R-93-001. February 1994.

EPA (1997) *Chemistry Assistance Manual for Premanufacture Notification Submitters*. Office of Pollution Prevention and Toxics, EPA, Washington DC. EPA 744-R-97-003. March 1997.

EPA (2015) *Summary of Accomplishments in the New Chemicals Program*. Available at: http://www.epa.gov/oppt/newchems/pubs/accomplishments.htm (accessed 15.3.2019).

FRI (1975) Structure activity correlation bibliography, Prepared by the Franklin Institute Research Laboratories for the Office of Toxic Substances, PB-240-658, Philadephia, Pennsylvania.

FRI (1976) *Analysis and trial application of correlation methodologies for predicting toxicity of organic chemicals*. Technical report requested by the Office of Toxic Substances from the Franklin Institute Research Laboratories, EPA/560-1-76-006.

GAO (1984) A*ssessment of New Chemical Regulation Under the Toxic Substances Control Act*. General Accounting Office, Washington, DC (USA). Office of the Comptroller General.

Harris R and Milkis S (1989) *The politics of regulatory change: a tale of two agencies*. London and New York: Oxford University Press.

Hilgartner S (2000) *Science on stage: Expert advice as public drama*. Stanford: Stanford University Press.

Hepler-Smith E (2019) Molecular Bureaucracy: Toxicological Information and Environmental Protection. *Environmental History* 0(2019): 1-27.

Hunt J and Shackley S (1999) Reconceiving Science and Policy: Academic, Fiducial and Bureaucratic Knowledge. *Minerva* 37: 141–164.

Jasanoff S (2004) *States of Knowledge: The Co-Production of Science and Social Order*. London and New York: Routledge.

Knorr-Cetina KD (2011) Financial analysis: Epistemic profile of an evaluative science. In: Lamont M (ed) *Social knowledge in the making*. Chicago: University of Chicago Press, pp. 405-442. Lipnick RL (1985) A perspective on quantitative structure-activity relationships in ecotoxicology. *Environmental Toxicology and Chemistry* 4(3): 255–257.

Lipnick RL (1991) Outliers: their origin and use in the classification of molecular mechanisms of toxicity. *The Science of the Total Environment* 109–110: 131–153.

Lipnick RL (1998) *Correlative vs Mechanistic QSAR Models for Review of Industrial Chemicals Under TSCA*. Conference, Dayton, Ohio.

Lipnick, RL, Bickings CK, Johnson DE and Eastmond DA (1985) Comparison of QSAR predictions with fish toxicity screening data for 110 phenols. In Aquatic Toxicology and Hazard Assessment: Eighth Symposium. ASTM International.

Lipnick RL, Johnson DE, Gilford JH, Bickings CK and Newsome LD (1985) Comparison of fish toxicity screening data for 55 alcohols with the quantitative structure-activity relationship predictions of minimum toxicity for nonreactive nonelectrolyte organic compounds. *Environmental Toxicology and Chemistry* 4(3): 281–296.

Mayo-Bean K, Moran K, Meylan B et al. (2012) *Methodology Document for the ECOlogical Structure-Activity Relationship Model (ECOSAR*) Class Program: Estimating Toxicity of Industrial Chemicals to Aquatic Organisms using ECO- SAR Class Program (Ver. 1.1). Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency.

Meyer H (1899) Zur Theorie der Alkoholnarkose - Erste Mittheilung. Welche Eigenschaft der Anästhetica bedingt ihre narkotische Wirkung? *Archiv für Experimentelle Pathologie und Pharmakologie* 42(2): 109–118.

Montgomery CM (2017) From standardization to adaptation: Clinical trials and the moral economy of anticipation. *Science as Culture* 26(2): 232-254.

Morgan MS and Morrison M (1999) *Models as mediators: Perspectives on natural and social science.* Cambridge: Cambridge University Press.

Nelson N, Geltzer A and Hilgartner S (2008) Introduction: the anticipatory state: making policy-relevant knowledge about the future. *Science and Public Policy* 35: 546–550.

NRC (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Committee on Toxicity, National Research Council, Washington D.C.: National Academies Press.

NTP (2002) *Current Directions and Evolving Strategies*. Washington D.C.: National Toxicology Program, US Department of Health and Human Services.

OECD (1994) *US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relations (QSARS)*. OECD Environmental Monographs N°88.

OECD (2007a) Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals, ENV/JM/MONO(2006)25, Feb. 15, 2007.

OECD (2007b) *Guidance on Grouping of Chemicals.* OECD's EHS Series on Testing and Assessment No. 80. ENV/JM/MONO(2007)28; Sept. 26, 2007.

OECD (2009) *Guidance Document for Using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals*. OECD's EHS Series on Testing and Assessment No. 102. ENV/JM/MONO(2009)5; Feb. 24, 2009.

Oreskes N (1998) Evaluation (not validation) of quantitative models. *Environmental Health Perspectives* 106: 1453–1460.

OTA (1983) *The Information Content of Premanufacture Notices*. Background paper, Congress of the United States, Office of Technology Assessment.

Overton E (1899) Über die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutlichen Ursachen und ihre Bedeutung für die Physiologie. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich* 22: 189–209.

Powell MR (1999) *Science at EPA: Information in the Regulatory Process*. Washington D.C.: Resources for the Future.

Rajan KS (2006) *Biocapital: The constitution of postgenomic life*. Durham: Duke University Press.

Sarewitz D and Pielke RJr (1999) Prediction in science and policy. *Technology in Society* 21: 121–133.

Schultz TW (2014) Gilman D. Veith (1944–2013) – A dedication. *SAR and QSAR in Environmental Research* 25(4): 249–251.

Sismondo S (1999) Models, Simulations, and Their Objects. *Science in Context* 12(02): 247–260.

Thoreau F and Laurent B (forthcoming) Situated expert judgment: QSAR models and transparency in the European regulation of chemicals. *Science & Technology Studies*.

Veith G (1981) *State of the art of structure-activity methods development.* Environmental Research Laboratory, Office of Research and Development, US EPA.

Vogel SA and Roberts JA (2011) Why the toxic substances control act needs an overhaul, and how to strengthen oversight of chemicals in the interim. *Health affairs* 30(5): 898–905.

Wambaugh J (2014) "Building fit-for-purpose pharmacokinetic models", In Vitro to In Vivo Extrapolation for High Throughput Prioritization and Decision Making Webinar Series.

Waterton C and Wynne B (2004). Knowledge and political order in the European Environment Agency. In: Jasanoff S (ed) *States of Knowledge: The Co-Production of Science and Social Order*. London and New York: Routledge, pp. 87-108.

Zeeman M (1995) Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act (TSCA). In: Rand GM (ed) *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate And Risk Assessment*. Boca Raton: CRC Press, pp. 703–715.

Zeeman M, Auer CM, Clements RG, Nabholz JV and Boethling RS (1995) US EPA regulatory perspectives on the use of QSAR for new and existing chemical evaluations. *SAR and QSAR in Environmental Research* 3(3): 179–201.

Zeeman M and Gilford J (1993) Ecological Hazard Evaluation and Risk Assessment under EPA's Toxic Substances Control Act (TSCA): An Introduction. In: Landis W, Hughes J, Lewis M (eds) *Environmental Toxicology and Risk Assessment – 1st Volume*. ASTM STP 1179. Philadelphia: American Society for Testing and Materials, pp. 7–21.

Zeeman M and Mayo-Bean K (2009) *Development and Use of (Q)SAR Tools by U.S.* EPA/OPPT to Assess Industrial Chemicals, Invited talk, Annual SETAC congress, New Orleans.

Zeeman M, Nabholz JV and Clements RG (1993) The Development of SAR/QSAR for Use Under EPA's Toxic Substances Control Act (TSCA): An Introduction. In: Gorsuch J, Dwyer J, Ingersoll C and La Point TW (eds)

Environmental Toxicology and Risk Assessment - 2nd Volume. ASTM STP 1216. Philadelphia: American Society for Testing and Materials, pp. 523–539.

Notes

1 Disclaimer: the views expressed by Dr. Zeeman are those of the author and this article has not been reviewed by the U.S. EPA and it should not be considered to reflect the views or policy of the U.S. EPA.

2 This article is based on empirical research performed in the framework of the project INNOX - Innovation in Expertise. Modeling and simulation as Tools of Governance, supported by the French Agence Nationale de la Recherche (Grant n°ANR-13-SOIN-0005), between 2014 and 2017, and coordinated by David Demortain (innox.fr).

3 For the last 40 years, the 'gold standards' of toxicological testing for humans have been the 13-week and the 2-year rodent bioassays along with other whole-animal studies (NTP, 2002).

4 The main use of QSAR modeling is performed within the pharmaceutical industry, as part of the early screening of candidate molecules or leads for toxicity problems. This aspect is outside the scope of this paper.

5 The level of certainty that is to be reached before one goes out making a claim or publishing something presented as true (see Collins, 1998).

6 As an example, we can cite an early EPA publication that introduced a QSAR model for estimating the LC50 for industrial alcohols, ethers, alkyl halides and benzene derivatives (EPA, 1981). At the heart of the model, one finds the following equation: Log 1= 1.17 + 0.94 log P, where log P is the logarithm of the n-octanol/water partition coefficient, the structure-related property of the chemical with which the toxicity in question (LC50, the concentrations of the chemical in air or water that kills 50% of the test animals during the observation period) is correlated. Running such a formula produces a numerical estimation of toxicity. Structure-activity theory allows inferring that a chemical that is the member of/included in the class of substances for which the formula has been developed (similarity according to the chosen descriptor for constituting the class; in the above case log P or the n-octanol/water partition coefficient), is likely to have a similar level of toxicity.

7 His equations summarizing structure-activity relationships are frequently nicknamed "Hansch equations". Hansch and other chemists have developed these equations and pushed for their use in the pharmaceutical industry throughout the 1970s and 1980s, leading to a situation where pharmaceutical companies R&D departments now routinely have "computational drug design" units that collaborate with other groups such as toxicologists, who perform the initial tests of the toxicity and effects of a substance on animals.

8 The EPA is organized in offices. These offices are headed by an "assistant administrator" who, like the EPA Administrator, is a presidential appointee. There are two types of offices: programmatic offices, which are created to implement a particular act (on toxic chemicals; on air quality; on water quality; on pesticides; and so on); and non-programmatic offices or services, such as the office of the general counsel or the Office for Research and Development, which is basically the scientific arm of the agency. ORD counts several dozen laboratories and more than 500 staff. It develops science for programmatic offices, but also follows its own internally defined research programs. Coordination between program offices and the ORD, or the responsiveness of the latter to the needs of program offices, is a recurrent issue in the history of EPA (Powell, 1999).

9 Veith had a PhD in water chemistry from the University of Wisconsin. He joined the Environmental Research Laboratory of the EPA in 1972, developing work on bioaccumulation of chemicals and pesticides in the environment. Veith did show some understanding of the specific goals and constraints

of the regulatory work, specifically the needs and challenges of implementing the new chemicals program for the OTS, and of the need for applicability of methods in regulatory evaluation of products (as opposed to ever more refined and sophisticated methods and results) (Schultz, 2014).

10 As of September 2010, the OTS had received a total of 50,449 submissions, more than the total number of substances included in the EU REACH program (EPA, 2015).

11 This means that they could in effect legally be manufactured. However, there was always a significant proportion of new chemicals that made it through the entire NCP PMN process, but that for a variety of reasons were apparently never actually manufactured (i.e., no notice of commencement of manufacture was received by the EPA, and thus they were not put on the TSCA inventory as an existing chemical in commerce).

12 OTS, HERD and EEB no longer exist. OTS was renamed the Office of Pollution Prevention and Toxics (OPPT) in around 1992. The major reorganization of that office in 1997 resulted in the morphing of HERD into what became the Risk Assessment Division (RAD).

13 Source: interview with the authors.

14 Source: interview with the authors.

15 On this, see Bradbury et al. 2015: "its usefulness in QSAR modelling can mainly be credited to the strategic approach taken in the development of the database. The express purpose of the fathead minnow database was to build relevant and reliable QSAR models based on data that covered a wide range of structure space and thereby a wide range of possible modes of toxicity. All toxicity tests were conducted in the same laboratory following standard test methods. Both the dilution water and fish used were from a single source. Chemicals used were of the highest purity, with all treatment concentrations measured under stringent data quality objectives. By controlling for these factors, variability in the test results was minimized and thereby increased confidence that variation in toxicity was related to variation in chemical structure and associated toxicological properties." (Bradbury et al., 2015: 19)

16 There is a list of the 49 SARs in the OTS QSAR Manual (EPA, 1988: ix). Thirty-one (31) of them cite their "Source" as being developed by EEB scientists (Vincent Nabholz and/or Richard Clements, etc.), and four of them list their "Source" as publications of the laboratory of Gil Veith in ORD.

17 Source: interview with the authors.

18 Two manuals developed in 1984 and 1996 were never published.

19 There was a decrease of almost 40% in NCP funding and a decrease of about 33% in NCP staffing between 1990 and 1995, even though the number of PMNs received seemed to trend upward.

20 Previously called HERD

21 When one launches the ECOSAR program on a computer, a special warning appears in a window saying "it is a screening-level tool", and that "Estimated values should not be used when experimental (measured) values are available".

22 The New Chemicals program eventually managed to process dozens of thousands of PMN, but the most profitable chemicals were the existing ones, for which TSCA is often analyzed as a failed statute in terms of decisions actually made on controlling existing toxic chemicals (Vogel and Roberts, 2011; Boullier, 2019).