GUARDRAILED, AUDITABLE AGENTIC WORKFLOWS FOR COMPUTATIONAL TOXICOLOGY VIA THE MODEL CONTEXT PROTOCOL (MCP)

Dr. Ivo Djidrovski,[1, 2]

1. Institute for Risk Assessment Sciences, Utrecht University, The Netherlands
2. in4r.ai

Background and Aim

Bottleneck:

Computational toxicology is fragmented across specialist tools, evidence databases, models, and regulatory knowledge sources. Writing justified safety narratives from these outputs remains slow, expert-intensive, and variable.

Aim:

To operationalize animal-free NGRA using ToxMCP: a guardrailed MCP infrastructure layer that connects trusted toxicology tools, structured evidence, and specialized LLM agents into transparent, reproducible workflows.

What ToxMCP does:

ToxMCP is a modular MCP infrastructure layer for agentic computational toxicology. It provides controlled access to modules for OECD QSAR Toolbox workflows, evidence registry, AOP/ontology mapping, exposure, PBPK/QIVIVE, read-across, mixtures, bioactivity PoD analysis, and decision reporting. Each module captures provenance and supports transparent, human-reviewed NGRA workflows.

The Model Context Protocol (MCP) in Action

A universal interface that connects AI agents with specialized tools and data sources.

ToxMCP
Modular · Auditable · Open Source

What are AI Agents?
AI agents are autonomous software entities that perceive, reason, plan, and take actions to achieve user goals.

What is MCP?
The Model Context Protocol (MCP) is an open standard that enables secure, structured communication between AI agents and external tools.

How it works
MCP provides a standardized **"USB-C for AI"**: discover tools, understand their capabilities, and exchange structured data.

YOUR AI AGENT (Orchestrator)

- Understand goal
- Plan steps
- Call MCP tools
- Analyze results
- Deliver answers

User Goal / Question
Tool Discovery & Schemas
Tool Invocation (Structured Request)
Structured Response (Results)
Final Answer & Report

MODEL CONTEXT PROTOCOL (MCP) HOST
The universal interface layer

- **Discover** — Find available tools & capabilities
- **Authenticate** — Secure access & permissions
- **Standardize** — Structured messages & schemas
- **Audit** — Logs, provenance & traceability

Connected MCP Servers (ToxMCP Ecosystem)

- **AOP MCP** — Adverse Outcome Pathways mapping & knowledge retrieval — MCP Server
- **OQT MCP** — OECD QSAR Toolbox automation & interpretation — MCP Server
- **EPA Comptox MCP** — Access to EPA CompTox data & dashboard — MCP Server
- **Exposure MCP** — Exposure data, scenarios & population context — MCP Server
- **Environmental Fate MCP** — Fate, transport & degradation predictions — MCP Server
- **PBPK MCP** — PBPK modeling & QIVIVE extrapolation — MCP Server

TOXICOLOGY TOOLS & DATA SOURCES

- Databases (e.g., AOP-Wiki, CompTox, ECHA)
- Software & Models (e.g., QSAR Toolbox, PBPK, Fate models)
- APIs & Web Services (e.g., EPA, OECD, public APIs)
- Literature & Evidence (curated studies, reports, guidelines)

CONSISTENT, AUDITABLE OUTPUT
Every workflow produces the same structured report: Executive Summary · Evidence & Justification · Assumptions & Uncertainties · Data & Model Traceability · Recommendations & Next Steps

- **Agent-Agnostic** — Works with any AI assistant or platform.
- **Secure & Controlled** — Tools are accessed through MCP with governance.
- **Modular & Extensible** — Add or swap tools via MCP without changing agents.
- **Transparent & Auditable** — Full provenance, logs, and human-in-the-loop review.
- **Open Source** — Community-driven, open standards & servers.

The Agentic Workflow

EXAMPLE WORKFLOW
From product use information to a NAM-based assessment using ToxMCP.

ToxMCP
Modular • Auditable • Open Source

- **Agent-Agnostic** — Works with any AI assistant or platform.
- **Secure & Controlled** — Tools are accessed through MCP with governance.
- **Modular & Extensible** — Add or swap tools via MCP without changing agents.
- **Transparent & Auditable** — Full provenance, logs, and human-in-the-loop review.
- **Open Source** — Community-driven, open standards & servers.

1 USER'S AGENT (ORCHESTRATOR)
User asks the question in natural language.

User Query: *"Assess the human health risks of Chemical X in a cosmetic leave-on lotion."*

Your preferred AI assistant (ChatGPT, Claude, Gemini, Copilot, Perplexity, etc.)

2 RETRIEVE PRODUCT USE INFORMATION
Agent retrieves product use information.

- **What happens:** Agent queries EPA CompTox to retrieve product use categories and conditions.
- **Result:** Product use:
  - Leave-on lotion
  - Consumer use
  - Dermal application
  - Specific use categories
- **EPA CompTox MCP:** Access to EPA CompTox Chemicals Dashboard and data (uses, functions, categories, endpoints)

3 DRAFT EXPOSURE SCENARIO
Agent develops an exposure scenario.

- **What happens:** Agent uses Exposure MCP to build a tailored exposure scenario for the product and population.
- **Result:** Exposure scenario:
  - Dermal exposure
  - Consumer population
  - Use frequency & amount
  - Duration & body site
- **Exposure MCP:** Exposure data, models and scenarios (consumers, occupational, situations, parameters)

4 RUN PROFILERS & MODELS
Agent runs OECD QSAR Toolbox profiler endpoints.

- **What happens:** Agent calls the OQT MCP to run profilers and retrieve endpoint predictions and mechanistic information.
- **Result:** Profiler outputs:
  - Physicochemical properties
  - Hazard predictions
  - Mechanistic alerts
  - Endpoint summaries
- **OQT MCP:** OECD QSAR Toolbox automation via Web API (profilers, endpoints, read-across, alerts)

5 SYNTHESIZE & DELIVER ASSESSMENT
Agent integrates results into an auditable assessment.

- **What happens:** Agent synthesizes all evidence into a NAM-based risk assessment with provenance and uncertainty.
- **Result:** Auditable Assessment:
  - Hazard characterization
  - Exposure assessment
  - Risk characterization
  - Uncertainties & data gaps
  - Recommendations
- **CONSISTENT OUTPUT (Same Structured Report Every Time)**
  - Executive Summary
  - Assessment Context
  - Methods & Tools Used
  - Findings & Evidence
  - Uncertainties & Data Gaps
  - Conclusions & Next Steps
  - References & Provenance

- **Governed** — Provenance captured at every step.
- **Auditable** — Logs, tool calls, and results tracked.
- **Human-in-the-Loop** — Expert review and decision support.
- **Open Source** — Transparent, community-driven infrastructure.

Results & Case studies:

Awesome !! Lets do a demo with chlorpyrifos :)

Worked for 3m 44s

Demo worked nicely.

OQT / QSAR Toolbox
Used CAS-linked Toolbox hit for chlorpyrifos:

```
Chlorpyrifos CAS 2921-88-2
ChemId: 9e81c32c-7867-2165-6a75-b60a2793c031
SMILES: CCOP(=S)(OCC)Oc1nc(Cl)c(Cl)cc1Cl
Toolbox: 4.8.2
```

Parent profiler highlights:

```
Protein binding by OASIS: phosphorylation / phosphate-phosphorothioate ester alerts +
Skin sensitisation by OASIS: SN2 thio/phosphates + SNAr activated heteroaryl alerts
DNA binding by OASIS: SN2 alkylation / alkylphosphates, alkylthiophosphates, alkylphos
Estrogen receptor: Non binder
DPRA Cys/Lys: Out of mechanistic domain
```

Metabolism simulation hit the important chemistry:

```
Rat liver S9: 9 metabolites
in vivo rat: 9 metabolites
neutral hydrolysis: 5 metabolites
skin: 1 metabolite
autoxidation: 0 metabolites
```

Key products found:

```
Chlorpyrifos oxon:
CCOP(=O)(OCC)Oc1nc(Cl)c(Cl)cc1Cl

Hydroxylated oxon:
CCOP(=O)(OCC)Oc1nc(Cl)c(Cl)c(O)c1Cl

3,5,6-trichloro-2-pyridinol / TCP:
Oc1nc(Cl)c(Cl)cc1Cl
```

AOP / CompTox
AOP-MCP mapped chlorpyrifos as a curated stressor to:

```
AOP:559 – Inhibition of acetylcholinesterase (AChE) leading to arrhythmias
AOP:490 – Co-activation of IP3R and RyR leading to reduced IQ / socioeconomic burden
AOP:202 – Topoisomerase II inhibitor binding leading to infant leukaemia
```

us which AOPs might be affected? Because it's not yet timing. Ah, you mean using reverse search using EPA compdocs? Yeah, for instance, or any other NUM data. Okay. So.

Worked for 5m 21s

Yes, exactly: this is a **reverse NAM-to-AOP search**.

For chlorpyrifos and chlorpyrifos-oxon, I would say:

"We start with the chemicals, not with the AOP. Then we ask: what NAM bioactivity exists, and which AOP key events do those data touch?"

Live CompTox / ToxCast Summary

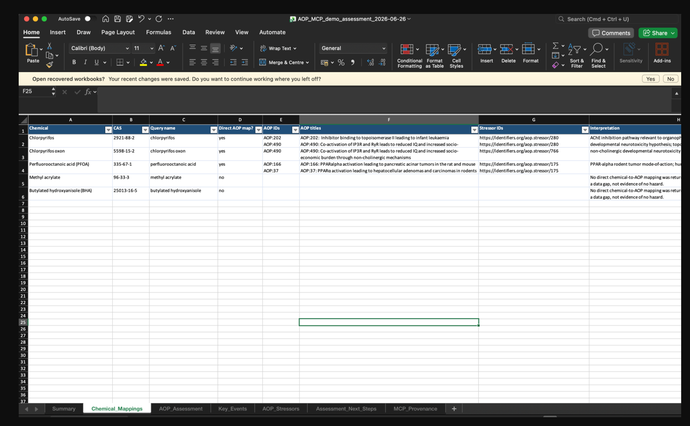
- **Chlorpyrifos** `DTXSID4020458` : 85 tested assay sets, 12 hit calls in summary; detailed high-confidence screen shows strong activity in **neurodevelopment**, **AChE/esterase**, **CYP/xenobiotic metabolism**, **nuclear receptor**, **cell cycle/apoptosis**, and **zebrafish developmental morphology** endpoints.
- **Chlorpyrifos-oxon** `DTXSID1030666` : 87 tested assay sets, 22 hit calls in summary; stronger and more potent than parent chlorpyrifos, especially for **AChE**, **pseudocholinesterase**, **acute neural network activity**, **developmental MEA**, and **zebrafish developmental endpoints**.

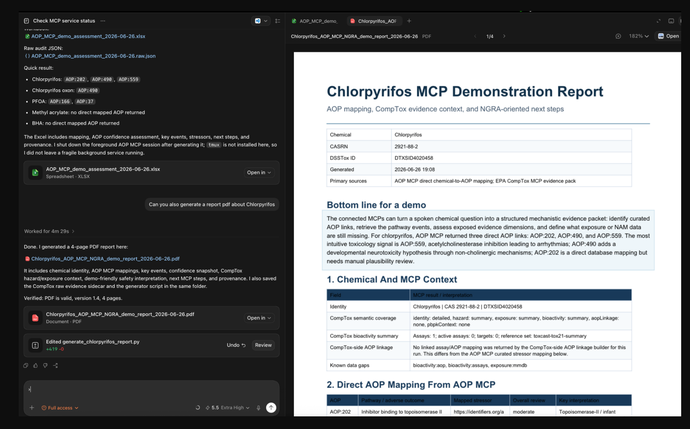
Most Relevant NAM Signals

NAM signal	Chlorpyrifos	Chlorpyrifos-oxon	AOP implication
AChE inhibition	strong	very strong, more potent	AOP:405, AOP:559
Neural network activity / MEA	active	very active at low concentration	developmental neurotoxicity AOPs
Zebrafish developmental morphology	active	active, more potent	developmental toxicity concern
CYP / metabolism	active	active	bioactivation, toxicokinetics, oxon formation
Oxidative stress / apoptosis / cell stress	active	active	AOP:500-like pathway
Thyroid / hormone transport signals, e.g. TTR/TPO	signals present	signals present	possible endocrine/developmental modifier, needs review

AOPs Likely Affected

AOP	Confidence	Why
AOP:405 OP-induced AChE inhibition leading to impaired cognitive function	High	Strong AChE NAM evidence; oxon is especially potent
AOP:559 AChE inhibition leading to arrhythmias	High for AChE MIE, endpoint is cardiac	Direct AOP-MCP stressor link plus AChE assays
AOP:490 IP3R/RyR co-activation leading to reduced IQ / socio-economic burden	Moderate	Direct AOP-MCP stressor link for chlorpyrifos and oxon, but NAM support is more indirect
AOP:475 glutamate receptor activation leading to learning/memory impairment	Moderate / literature-supported	2025 systematic review links CPF NAM evidence to glutamate/excitotoxicity mechanisms
AOP:499 MEK/ERK activation leading to learning/cognition deficits via neurotransmitter release	Moderate / literature-supported	NAM literature supports ERK/calcium/neurotransmitter disturbance
AOP:500 MEK/ERK activation leading to learning/cognition deficits via ROS and apoptosis	Moderate to strong for KEs	NAM data include oxidative stress, mitochondrial/cell stress, apoptosis
AOP:202 topoisomerase II leading to infant leukemia	Low for chlorpyrifos conclusion	It maps as a stressor, but evidence is much stronger for comparator chemicals like etoposide





Trust and transparency

Guideline Aware: OECD principles, reporting expectations, and workflow constraints are enforced across MCP modules through structured schemas, controlled tool calls, and predefined evidence requirements

Auditability: Every agent action, tool call, data source, assumption, and intermediate output is logged, versioned, and linked to evidence, making each assessment traceable and verifiable.

Cryptographic Provenance: Runs can be cryptographically signed and hashed, including inputs, tool versions, schemas, prompts, parameters, timestamps, and outputs, creating a tamper-evident audit trail.

Human-in-the-Loop: Expert review remains central. ToxMCP surfaces uncertainties, data gaps, assumptions, and decision points so that humans validate conclusions before regulatory or scientific use.

Scan for video demo and examples

Conclusions and next steps

Democratize toxicology tools: ToxMCP allows users to access specialist toxicology tools through their preferred AI assistant, lowering the barrier to transparent, evidence-based NGRA workflows.

Enable auditable assessment: Modular MCP servers capture tool calls, data sources, assumptions, uncertainties, and outputs, creating reproducible and cryptographically traceable assessment records.

Operationalize NAMs: By connecting QSAR, AOP, exposure, PBPK/QIVIVE, environmental fate, and regulatory evidence workflows, ToxMCP provides a practical infrastructure layer for animal-free safety assessment.

Operationalize NAMs: Expand validated MCP modules, align workflows with OECD reporting expectations, and test ToxMCP in real-world case studies with expert review and regulatory-facing documentation.