



CRICOS: 00233E | TEQSA: PRV12076

“Smart Compounds: Overcoming Challenges

Jan Hettenhausen, Rebecca Lang and Julie Toohey





ACKNOWLEDGEMENT OF COUNTRY

Griffith University acknowledges the people who are the Traditional Custodians of the land. We pay respect to the Elders, past and present, and extend that respect to all Aboriginal and Torres Strait Islander peoples.

Speakers



Rebecca Lang
Operations Manager
Compounds Australia



Jan Hettenhausen
Senior Developer and Tech Lead
eResearch & Specialised Platforms



Julie Toohey
Business Analyst
Compounds Australia & eResearch
and Specialised Advisory

Overview: Compounds Australia

**Australia's only
dedicated biodiscovery
sample management
facility**



Compounds Australia, located within the Griffith Institute for Drug Discovery, provides cost-effective sample management and logistics throughout the drug discovery pipeline.

By securely storing and curating sample libraries, we allow translational health researchers worldwide to outsource their sample and natural product logistics and focus on results.

Through specialist infrastructure and expertise, we're pursuing a vision to transform sample management practice in Australia, in support of smarter, faster and more efficient drug discovery.



1.5 million compounds

We curate libraries of more than 1.5 million samples. Open library samples are conveniently searchable via our custom web platform CASPeR.



Servicing 150+ drug discovery projects annually

Compounds Australia supports drug discovery and life science research conducted by our member organisations across the biotech and pharmaceutical sectors.



4 dedicated robotics platforms

Compounds Australia employs a custom suite of robotics and liquid handling infrastructure that sets the benchmark for sample management practice in Australia.

eResearch Services



Research software
+
software development



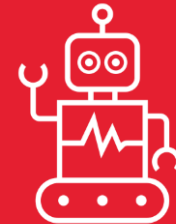
Research computing services



Computing training
+
support



Research data services
+
storage



Emerging technologies



Research infrastructure support



Collaboration
+
communication tools

[eResearch Services \(griffith.edu.au\)](http://griffith.edu.au)

What is Compound Management?

Compound Management and Logistics is a key enabling technology for early-stage drug discovery projects

Valuable chemical resource preserved

Quality compounds

Access to limited samples

Tracking of all samples throughout storage life-span

Chemistry focus

Samples supplied in vials, deepwells, microtubes



Small molecule libraries
Natural product libraries
Purchased commercial libraries

Compound Management & Logistics

Storage, curation & plate handling



Expertise & Infrastructure for storage, nanolitre/microlitre plate processing, data handling & manipulation

Biology & Screening focus

Samples provided in multi-format 96, 384 & 1536 well microplates

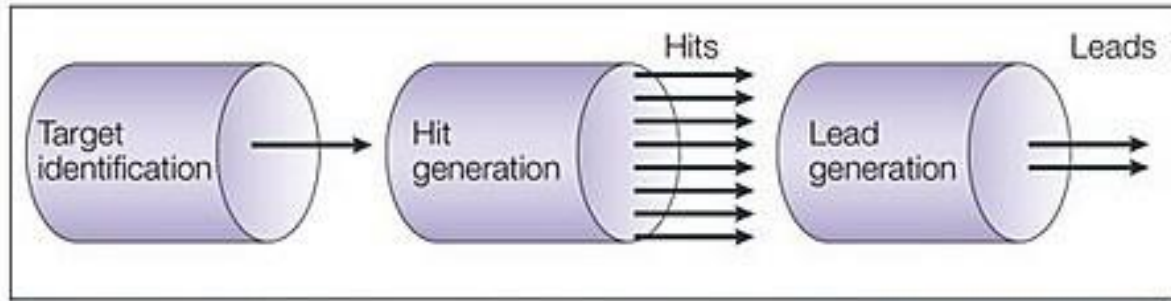


Cherrypicking, dose response & combination layouts provided

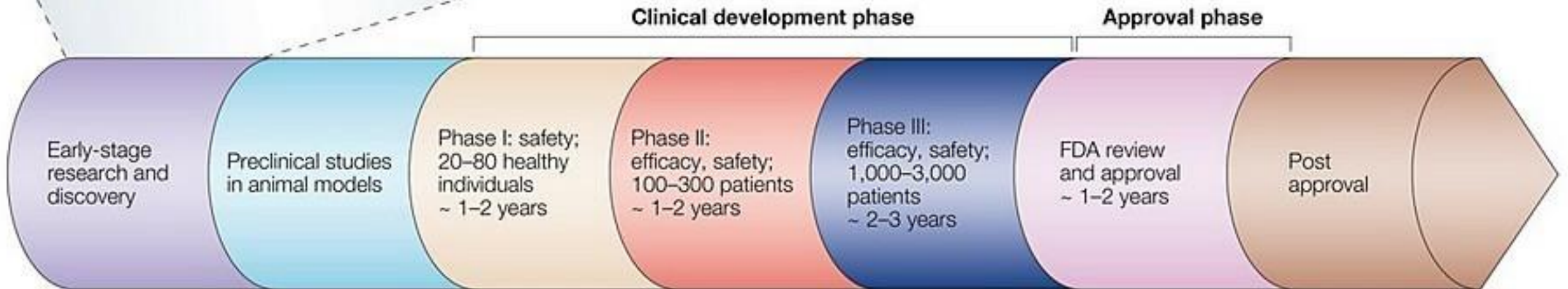
DRUG DISCOVERY

Compound Libraries
High Throughput Screening
Medical

Drug Discovery Pipeline



**COMPOUNDS
AUSTRALIA**



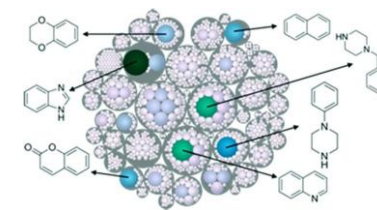
**Preclinical
Development**

*Efficacy
Safety and Toxicity*

Source: Computational Biology Lab <http://dlab.cl/>

Open Access Collections – Available to all members

- Open Scaffold collection - ~34,000 compounds – 1,200 core scaffold groups
- Open Drug collection
 - 8,500 compounds
 - Various vendors
 - Annotated into several collections based on known activity profiles or target (i.e. Anti-Cancer, Kinase inhibitors, Epigenetic modulators, Approved drugs, etc)
- Open Academic collection - >40,000 compounds
 - >10,000 compounds – provided from Academic researchers
 - CSIRO is the majority stakeholder of this collection >30,000 compounds
- NatureBank
 - >18,000 extracts and >90,000 fractions
- Fragment Collections
 - Commercially Available Fragment Library sets:
 - ~1280 Novel Fragment Diversity Collection
 - ~950 Electrophilic Covalent Probes
 - 1,000 compounds – purchased UniAdelaide Fragment collections
- Australian Lead Identification Consortium Library (ALIDC)
 - 330,000 Lead-like diversity collection – designed to represent the chemical space of over 1 million diverse lead-like commercial available compounds from multiple reliable vendors



What is/was CASPeR?

Compounds Australia Structure PoRtal

CASPeR provided researchers with a searchable interface to access available information on sample properties on various Open Access compound/sample libraries.

CASPeR was a bespoke interface built in 2014 to allow members to search available samples on sample properties and identify samples of interest.

compoundsaustralia.com/casper



About CASPeR

The Compounds Australia Structure PoRtal provides structural information on our Open Access Collections On-Demand



Browse Our Compounds

Filter the collection by chemical properties and library types or browse the entire library



Find Compound Structures

Search the Open Access library by using the SMILES code of a known compound or drawing the compound structure

Phone: +61 (07) 3735 4448

Email: compounds-australia@griffith.edu.au

CASPeR



Need for CASPeR to evolve

2017-2018 Compounds Australia's Collections grew and expanded; as did our member base:

CASPeR started to display performance issues and had inadequate functionality for the growing sample types, creating significant obstacles for client uptake and use.

Shortfalls of CASPeR: Older technology, solutions unreliable (MarvinJS Integration), search functionality limited – SMILES string search for similar samples, Known “linked samples” (I.e. Scaffold ID), Known Sample Identified Number (SN)

Performance issues & inadequate search capability - **Impeded Uptake and Use by Members**

Important to have an efficient system to serve member networks better (add additional functionality and directly integrate known sample information to publicly accessible online databases (E.g. PubChem, DrugBank, ChEMBL, etc))

Time to generate a solution that was adaptable, configurable and could be built upon to pave way for future research and development opportunities.

Why we decided to start over?

- Maintainability
 - CASPeR needed its own VM
 - PHP 5.x is long dead and even newer RedHat versions don't support it
 - MarvinJS discontinued support for the version used in Casper
- Sustainability of the software we build
 - Sometimes starting over is simply cheaper and easier
 - Technology ecosystems evolve and what used to be bespoke is often commodity now

The State of the Art in Compound Databases

- PubChem / ECDB / ChemSpider / ChEMBL
- Search
 - Substructure
 - Superstructure
 - Similarity
 - Identifiers
- Filtering by Chemical Properties
- Visual editing of Structures
- 2D/3D Visualisation of Structures

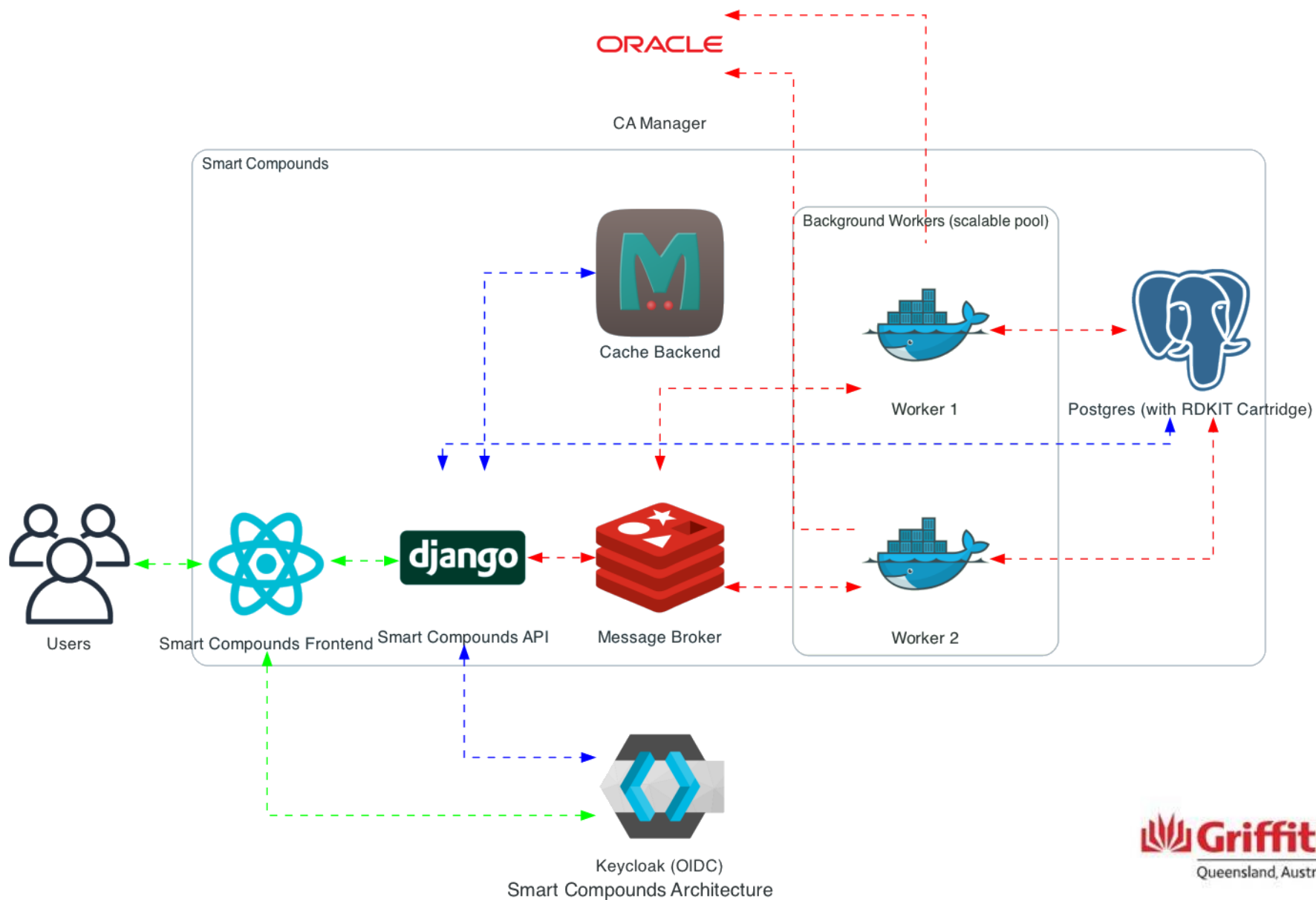




Solution: Smart Compounds

- Challenges:
 - Navigating the landscape of OSS Chemistry Toolkits
 - Building a Sustainable Product
 - Searching efficiently when the searches are quite costly
- Solution:
 - Aligned with the rest of our tech stack
 - Django / Restframework, React, Docker, Keycloak
 - Computation happens where the data is
 - RDKit Postgres Cartridge
 - Ketcher, Indigo and WASM
 - Shift towards commodity components

Smart Compounds Architecture



Enter SMILES Code

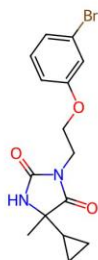
Similarity Search
 Substructure Search
 Superstructure Search
 Full-Text Search



87255 Results

Page 1 of 8726

Order By



SN00769310

SMILES Code: Brc1cccc(OCCN2C(=O)NC(C)(C3CC3)C2=O)c1

InChi: InChi=1S/C15H17BrN2O3/c1-15(10-5-6-10)13(19)18(14(20)17-15)7-8-21-12-4-2-3-11(16)9-12/h2-4,9-10H,5-8H2,1H3,(H,17,20)

InChiKey: NUKNTZISZUVGAI-UHFFFAOYSA-N

Molecular Formula: C₁₅H₁₇BrN₂O₃

Molecular Weight: 353.21108 g/mol

Chiral Centres: 3

HBA: 5 **HBD:** 1 **LogP:** 2.8

Rotatable Bonds: 5

Lipinski Filter: Yes; 0 violations

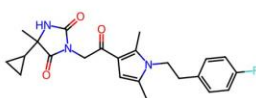
Scaffold ID: SC000256

QCL Sample Number: SN00769310

SCAFFOLD

Enlarge

[Details](#)
[Similar Compounds](#)
[Structure Search](#)
[Related Records \(Scaffold ID\)](#)



SN00769311

SMILES Code: Fc1ccc(cc1)CCn1c(C)cc(C(=O)CN2C(=O)NC(C)(C3CC3)C2=O)c1C

InChi: InChi=1S/C23H26FN3O3/c1-14-12-19(15(2)26(14)11-10-16-4-8-18(24)9-5-16)20(28)13-27-21(29)23(3,17-6-7-17)2-5-22(27)30/h4-5,8-9,12,17H,6-7,10-11,13H2,1-3H3,(H,25,30)

InChiKey: ODLKSEJDRXJNE-UHFFFAOYSA-N

Molecular Formula: C₂₃H₂₆FN₃O₃

Molecular Weight: 411.469243 g/mol

Chiral Centres: 4

HBA: 6 **HBD:** 1 **LogP:** 3.7

Rotatable Bonds: 7

Lipinski Filter: Yes; 0 violations

Scaffold ID: SC000256

QCL Sample Number: SN00769311

SCAFFOLD

Enlarge

[Details](#)
[Similar Compounds](#)
[Structure Search](#)
[Related Records \(Scaffold ID\)](#)

LIBRARY

- Academic
- Drug
- Fragment
- Scaffold

Help

SAMPLE IDENTIFIERS

No Identifiers Selected

Add Identifier(s) Add

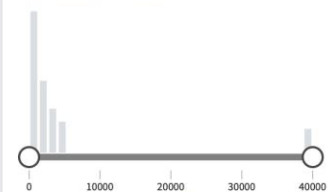
Help

LIPINKSI'S RULE OF FIVE

- No violations
- ≤ 1 violation
- ≤ 2 violations
- ≤ 3 violations
- ≤ 4 violations

Help

MOLECULAR WEIGHT



Min -1 Max 40000



What have we learned?

- Research project planning and strategy focus on innovation over maintenance
- Especially in the eResearch space we need to plan more for sustainability of our projects. That includes:
 - Tech Stack
 - Maintainability
 - Clear Expectation of the Project Lifecycle
- Technology Ecosystems evolve fast and what was once a bespoke system may be commodity now
 - Sometimes starting over is a better option than maintaining an old system
 - Maintaining an old product tends to increase its technical debt over time.
 - Starting over may even allow for a much better product with less effort.



How we build Sustainable Products

- We have standardised our Tech Stack
- We use testing and continuous delivery to simplify long term maintenance
- We discuss timeframes of the project before we even start any work. Clear expectations matter.
- And if it makes sense, we rebuild from scratch



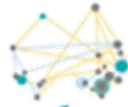
Wind up

- CA has an emphasis on **continual commitment to serving the research community** which SC supports.
- SC provides an accessible and efficient platform for searching compounds Australis open access compound sample libraries, promoting faster and more efficient drug discovery.
- Future State: **Anticipated enhancements and development direction SC might hold biological data; target data of screening data being available.**
- CA will be launching Smart Compounds late October.
- If anyone is interested in collaborating or learning more, please get in touch directly via program contact information, come see us afterwards or contact compoundsaustralia@griffith.edu.au.
- Alternatively, if you are attending AusBiotech Conference in Brisbane on 1-3 November, please drop by to the Griffith: Compound Australia (co-exhibiting with Life Science Queensland) booth to learn more about Smart Compounds.

Questions



Transform compound management in Australia- smarter, faster and more efficient drug discovery



- Unique in Southern Hemisphere
- Serve as a central depository for small molecule compounds synthesized or isolated in Australia
- Provide industry standard microtube/microplate storage solutions
- Providing quality compounds in standard and customised 96, 384, 1536 assay ready formats
- Promote collaborative drug discovery and foster exposure of the unique Australian small molecule collection to industry