

Publishing reproducible, interactive binding constant analyses using FAIR data principles

James Wilmot^a, Varvara Efremova^a, Prof. Pall Thordarson^{a,b}

^a UNSW School of Chemistry, The Australian Centre for Nanomedicine, ^b The UNSW RNA Institute



opendata.studio

Closing the gap between analysis and publication

The problem

Currently there is a large gap between running an analysis and publishing it online or in a journal. Published workflows and their artefacts (such as visualisations and tables) are often lacking full reproducibility.

Readers are left guessing:

- How was this generated?
- What algorithm was used?
- What environment was it run in?

Researchers also lack the time and resources to create user-friendly and well-documented workflows, leading to knowledge loss and "reinventing the wheel".

Capturing diverse workflows for publication

To achieve fully Findable, Accessible, Interoperable and Reusable (FAIR) web publication of interactive workflows, we developed a set of tools alongside a workflow packaging specification. (See section "What are datakits?")

Our tooling allows users to run and publish fully FAIR workflows as embeddable web components. (see Figure 1 below)

Proof of concept: Bindfit

In 2016, we released Bindfit — a web application for interactive analysis and publication of binding constant calculations. Every Bindfit workflow can be saved and published via a link with a unique identifier.

Since 2016, it has gained wide adoption with over 70,000 sessions and nearly 500 citations on Scopus.



Example of a published analysis in Bindfit
(Takemasa & Nozaki 2024)

app.supramolecular.org/bindfit/view/b24c676b-5fc5-4774-8caa-9dc1267ee771

Publishing a binding constant analysis datakit

Fork

Fork the [opendatastudio/bindfit-datakit](https://github.com/opendatastudio/bindfit-datakit) repository to preserve provenance

Ingest data

1. Load data and automatically describe schema
`dk load data data/input.csv`

Configure algorithm

1. Choose fit model
`dk set model nmr1to2`
2. Set initial parameter guesses
`dk set params k11.init 1.44`

Run algorithm

`dk run`

Visualise

Visualise facets of the analysis

1. View fit graph
`dk view fitGraph`
2. View calculated parameter table
`dk show params`

Publish

1. Create a release of your datakit workflow and mint a DOI with Zenodo

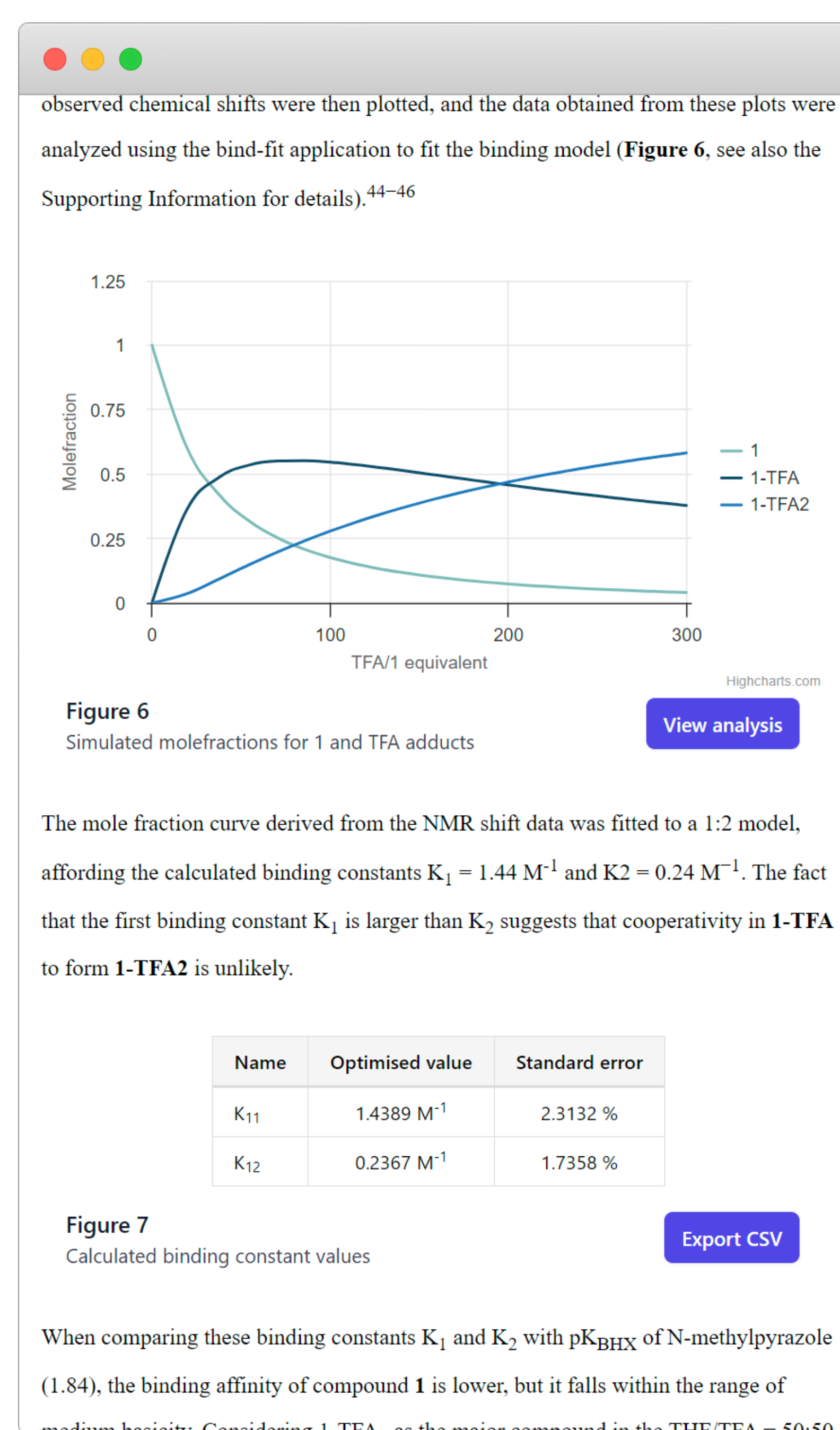


Figure 1: Embeddable workflows in an example publication

What are datakits?

In order to encapsulate analysis workflows, we developed "datakits" - an open-source JSON-based data standard extending the Frictionless Data specification. (See frictionlessdata.io)

A datakit describes:

- the analysis algorithm and its execution environment
- saved run states from algorithm executions
- input and output data, along with configurable options
- visualisations of data, including graph and table specifications
- user interface definitions.

The opendata.studio project

Currently released:

- a binding constant analysis tool packaged as a datakit
- a command line tool for modifying and running datakits
- documentation for running and implementing custom datakits.

In the pipeline:

- embeddable web components for datakit visualisations
- web application with local browser execution of datakits currently in development

If you're interested in using datakits or collaborating, please get in touch! hello@opendata.studio

Further reference



Datakit and command line documentation
docs.opendata.studio



Command line tool
[opendatastudio/cli](https://github.com/opendatastudio/cli)



Bindfit datakit
[opendatastudio/bindfit-datakit](https://github.com/opendatastudio/bindfit-datakit)



UNSW
SYDNEY