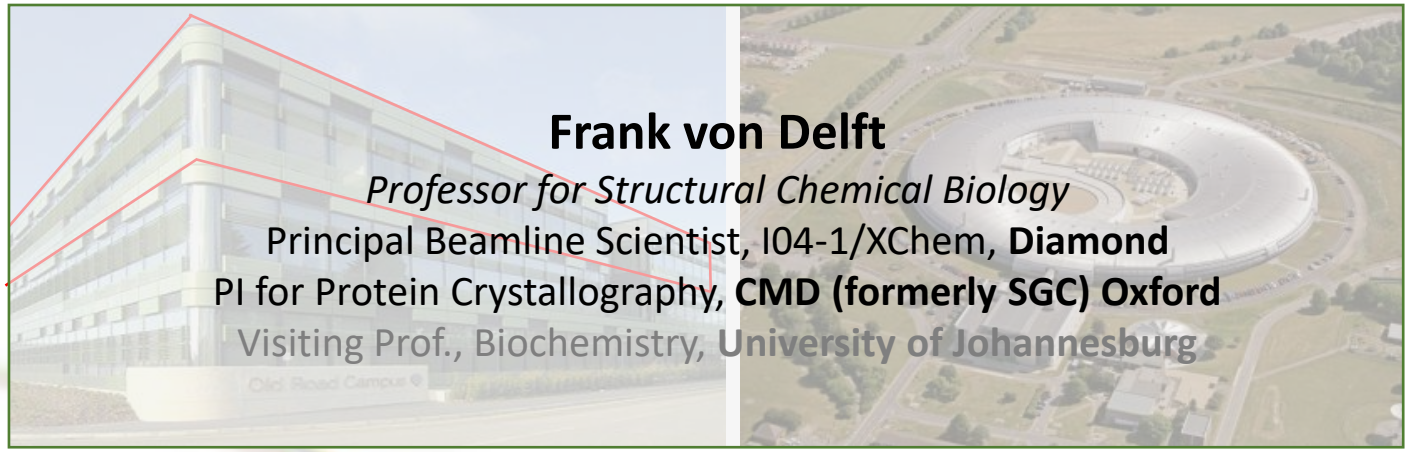
Moonshot lessons for open science: sustaining a collaboration platform for real-time data and idea sharing in targeting SARS-CoV-2 Main Protease

How Data Infrastructure Should Accelerate Open Science

ASAP Open Science Forum – 19 Oct 2022



Frank von Delft

Professor for Structural Chemical Biology

Principal Beamline Scientist, I04-1/XChem, **Diamond**

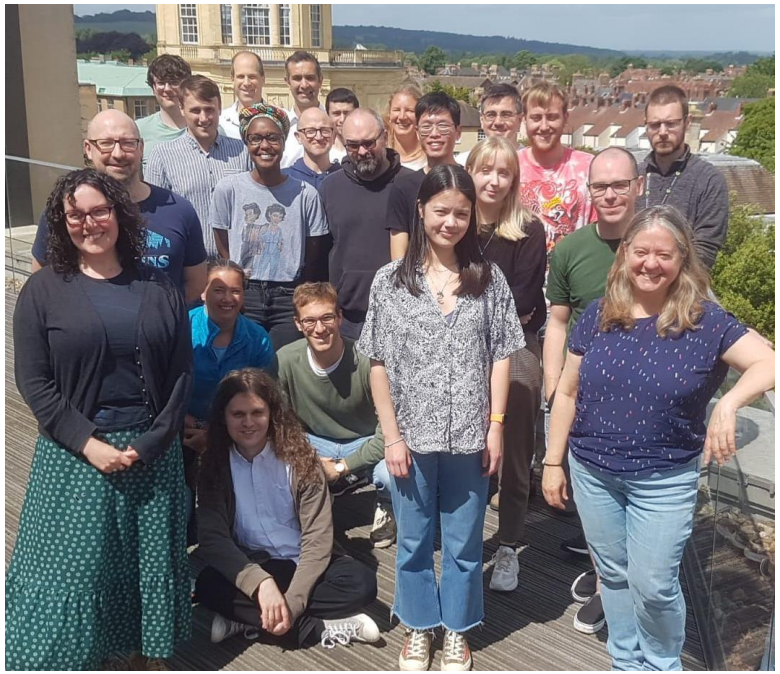
PI for Protein Crystallography, **CMD (formerly SGC) Oxford**

Visiting Prof., Biochemistry, **University of Johannesburg**

Acknowledgements

CMD-Oxford: PX group
- Lizbe Koekemoer
- Mike Fairhead
- Nathan Wright
- Beth MacLean
- Andrew Thompson
- Ruben Sanchez
- Haitem Elmassoudi
- ** Mpho Makola*
- ** Conor Wild*
- ** Harold Grosjean*
- ** Anna Carbery*
- ** Stephanie Wills*
- ** Melissa Marx*
- ** Jack Scantlebury*
- ** (Mingda Ye)*

Diamond: beamline I04-1
- José Brandaõ-Neto
- Daren Faeron
- Louise Dunnett
- Charlie Thompson
- Warren Thompson
- (Rachael Skyner)
- (Tyler Gorrie-Stone)



CMD (U. Oxford):
Annette von Delft
Brian Marsden
Ellie Williams
Oleg Fedorov
Paul Brennan
Kilian Huber
Alex Bullock
…

Diamond:
Industrial liaison
- Alex Dias
- Ailsa Powell
- Elizabeth Shotton

Masses of people…
- MX Village
- GDA team
- Controls
- Scientific software

COVID Moonshot

ASAP
ACCESSIBLE ANTIVIRALS TO PREVENT PANDEMICS

U.Oxford - Statistics
- Charlotte Deane

U.Oxford - Chemistry
- Chris Schofield

U.Oxford – Biochem
- Nicole Zitzmann
- Phil Biggin

U.Oxford – Dunn School
- Ivan Ahel

U.Leeds
- Adam Nelson

U.Warwick
- Chris Dowson

U.Sussex
- John Spencer

Weizmann Institute
- Nir London

UCSF
- James Fraser

U.Cambridge
- Tom Blundell
- Dave Spring

U.Leiden
- Monique Mulder
- Nick Pearce

U.Barcelona
- Xavier Barril

UCL
- Mat Todd

SIMM, Shanghai
- Caiguang Yang

U.Kent
- Dave Brown

UCT – H3D
- Kelly Chibale
- Greg Basarab

SSRF, Shanghai
- Jianhua He

U.Newcastle
- Martin Noble

Sloan-Kettering
- John Chodera

U. Edinburgh
- Julien Michel

UCB
- Jiye Shi

Informatics Matters
- Tim Dudgeon

M2M Solutions
- Vladimir Psenak

Mitegen
- Ben Apker

Enamine
- Vladimir Ivanov

PostEra, Cambridge
- Alpha Lee

Exscientia
- Anthony Bradley

Janssen Pharaceuticals
- Edgar Jacoby

GlobalPhasing Ltd
- Gerard Bricogne

LifeArc
- Peter Coombs

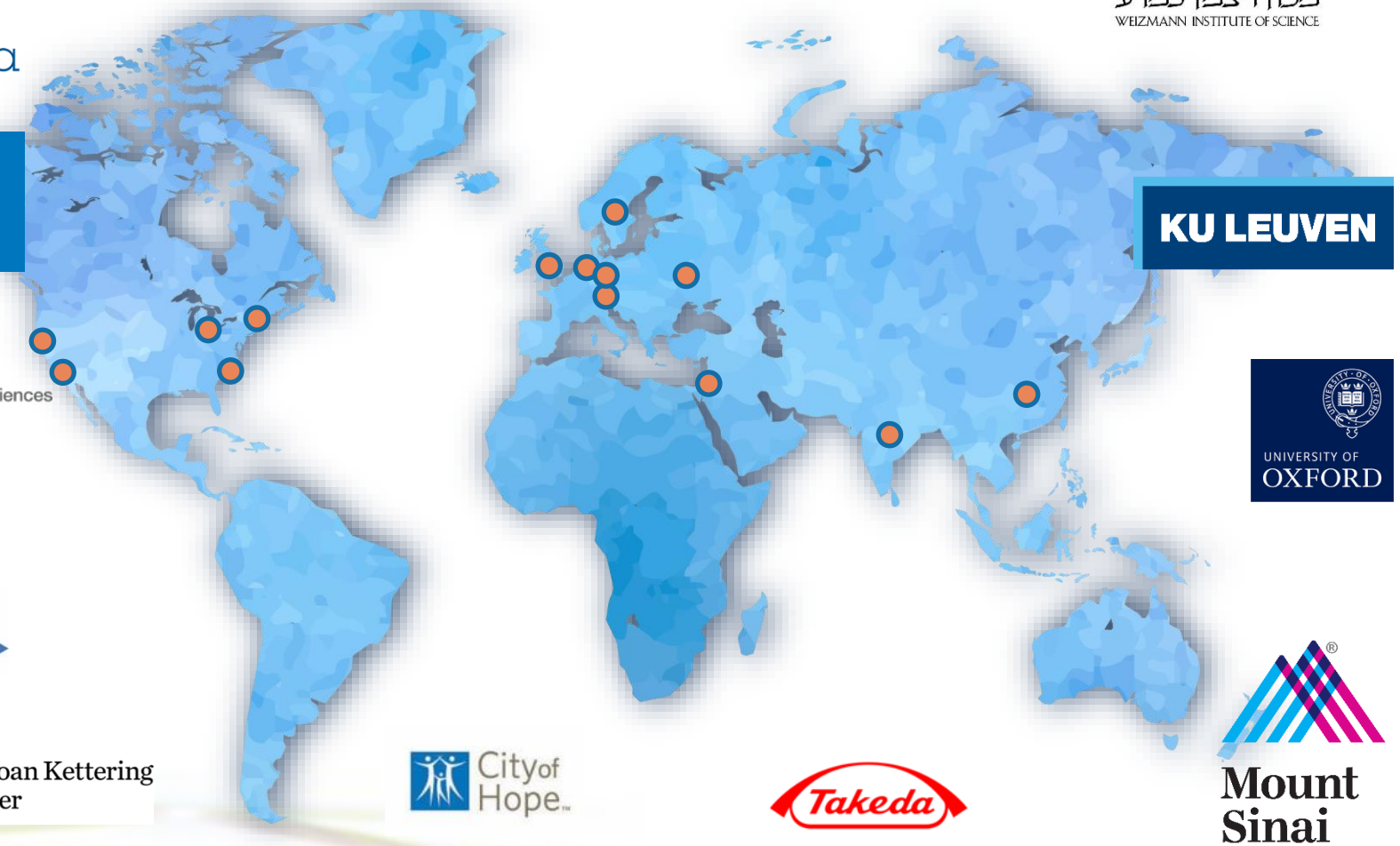
MedChemica
- Ed Griffin

Daren Fearon

Warren Thompson

Acknowledgements - Moonshot

>180 contributors world-wide



Daren Fearon	Diamond Light Source Ltd; Research Complex at Harwell
Louise Dunnett	Diamond Light Source Ltd; Research Complex at Harwell
Alice Douangamath	Diamond Light Source Ltd; Research Complex at Harwell
Alex Dias	Diamond Light Source Ltd; Research Complex at Harwell
Ailsa Powell	Diamond Light Source Ltd; Research Complex at Harwell
Jose Brandao Neto	Diamond Light Source Ltd; Research Complex at Harwell
Rachael Skyner	Diamond Light Source Ltd; Research Complex at Harwell
Warren Thompson	Diamond Light Source Ltd; Research Complex at Harwell
Tyler Gorrie-Stone	Diamond Light Source Ltd; Research Complex at Harwell
Martin Walsh	Diamond Light Source Ltd; Research Complex at Harwell
David Owen	Diamond Light Source Ltd; Research Complex at Harwell
Petra Lukacik	Diamond Light Source Ltd; Research Complex at Harwell
Claire Strain-Damerell	Diamond Light Source Ltd; Research Complex at Harwell
Halina Mikolajek	Diamond Light Source Ltd; Research Complex at Harwell
Sam Horrell	Diamond Light Source Ltd; Research Complex at Harwell
Lizbé Koekemoer	University of Oxford
Tobias Krojer	University of Oxford
Mike Fairhead	University of Oxford
Beth MacLean	University of Oxford
Andrew Thompson	University of Oxford
Conor Francis Wild	University of Oxford
Mihaela D. Smilova	University of Oxford
Nathan Wright	University of Oxford
Annette von Delft	University of Oxford
Carina Gileadi	University of Oxford
Victor L. Rangel	School of Pharmaceutical Sciences of Ribeirao Preto
Chris Schofield	University of Oxford
Tika R. Malla	University of Oxford
Anthony Tumber	University of Oxford
Tobias John	University of Oxford
Ioannis Vakonakis	University of Oxford
Anastassia L. Kantsadi	University of Oxford
Nicole Zitzmann	University of Oxford
Juliane Brun	University of Oxford
J. L. Kiappes	University of Oxford
Michelle Hill	University of Oxford
Finny S. Varghese	Radboud University Medical Center
Ronald P. van Rij	Radboud University Medical Center
Gijs J. Overheul	Radboud University Medical Center
Susana Tomásio	Collaborative Drug Discovery
Charlie Weatherall	Collaborative Drug Discovery
Mariana Vaschetto	Collaborative Drug Discovery
Hannah Bruce Macdonal	Memorial Sloan Kettering Cancer Center
John D. Chodera	Memorial Sloan Kettering Cancer Center
Dominic Rufa	Memorial Sloan Kettering Cancer Center

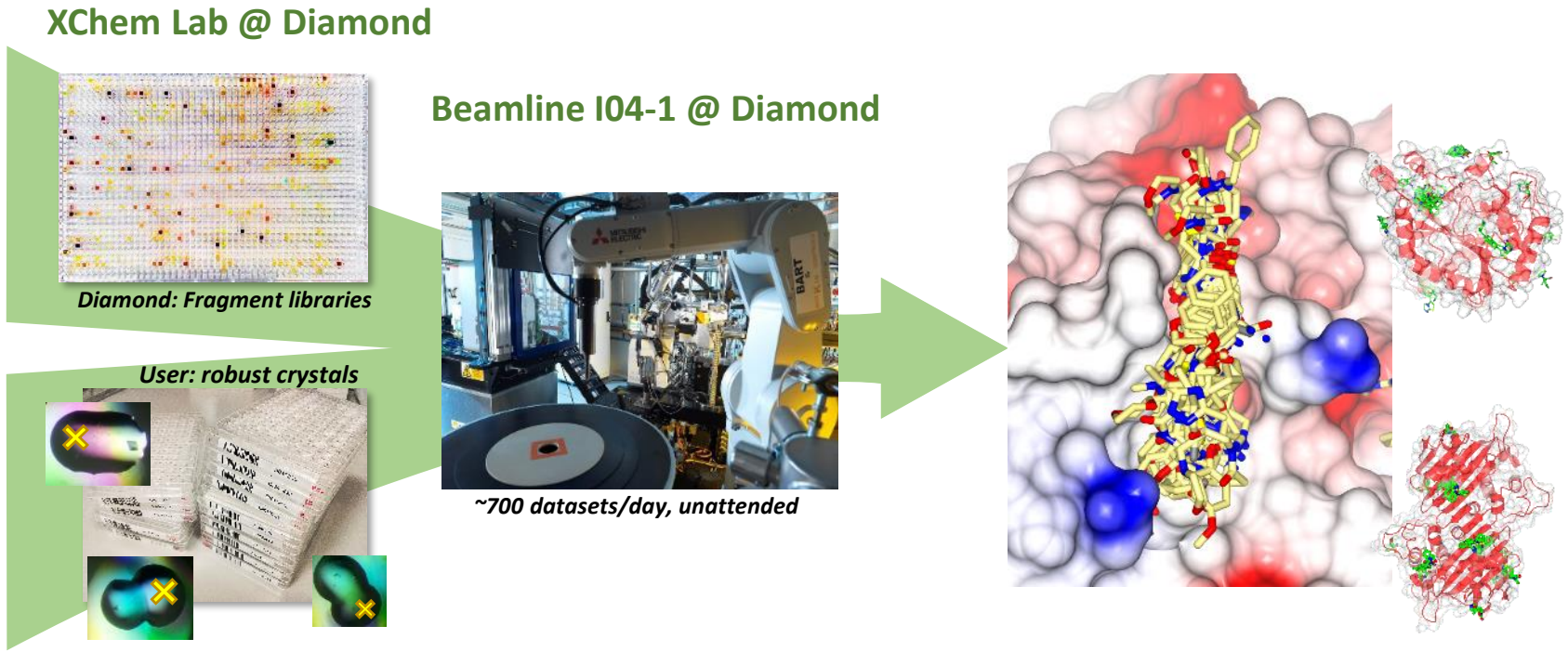
https://tinyurl.com/covid-moonshot-authors

Open data – version 0.1

Zip up and deposit

zenodo

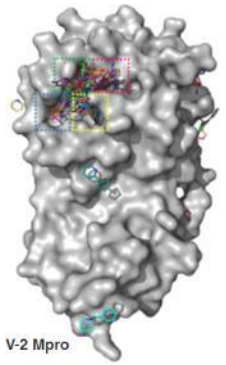
XChem@Diamond: routine access to gold-standard but out-of-reach crystal-based fragment screening for world-wide community



- Fragment screening directly in crystals – order-of-magnitude faster
- Crystal growth, soaking, harvesting, data collection, data processing – *>1000 xtals/wk, 2-3 users/wk*
- Routine users since 2016 – *>200k crystals in 3 years, >5000 hits for >150 targets*
- Discovery projects: **>130** academic, **>40** industry, **>40** collaboration (SGC, Moonshot, AMR)
- Similar platforms being established at 5 synchrotrons world-wide

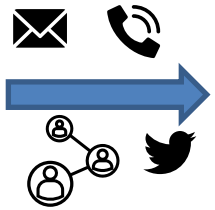
COVID Moonshot – spontaneous international collaboration

Mpro: Validated, Essential, Conserved



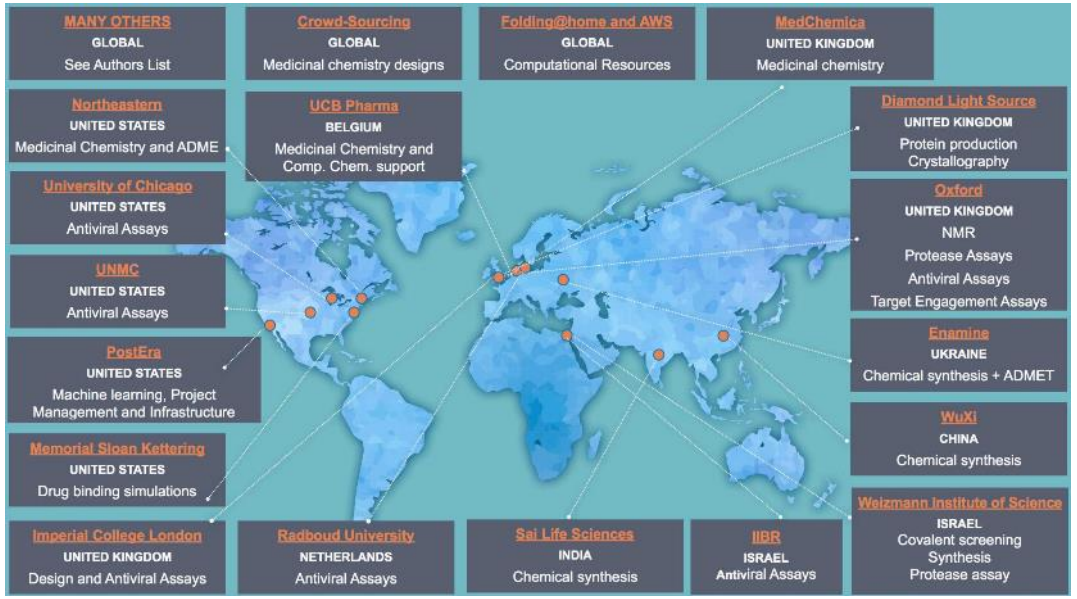
Fragment screen

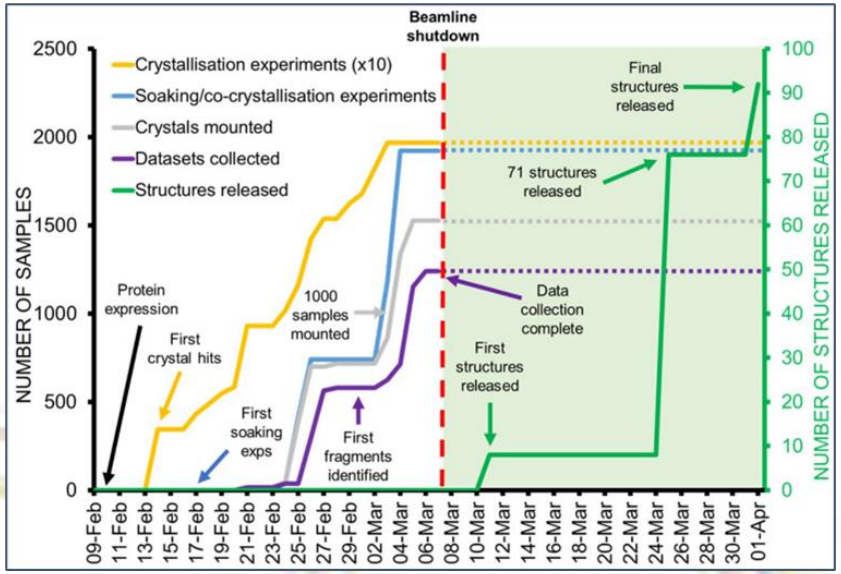
(Pre-lockdown)



>40 groups

Expertise pro bono - Experiments at cost - Philanthropic funding

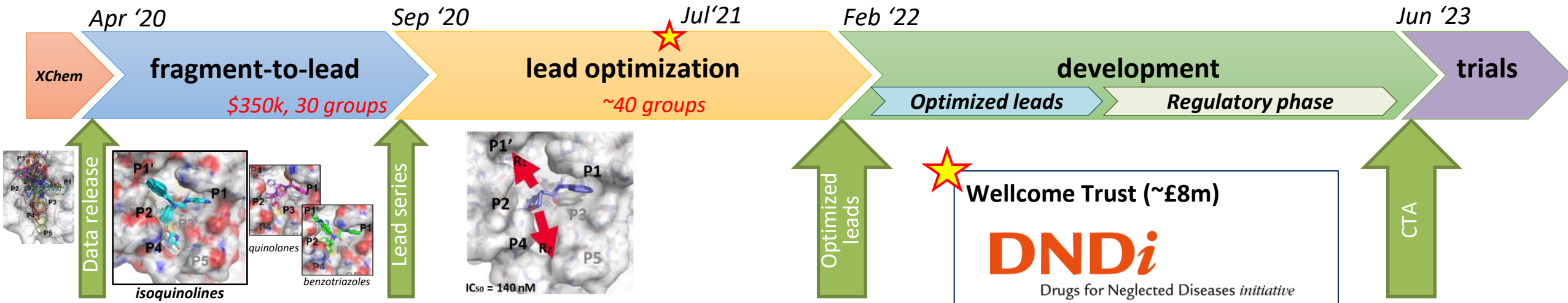




COVID Moonshot initiative (ongoing)

Unique strategy: <u>work fully open to enable rapid global availability</u>

- data and compound designs immediately public
- no IP filed, generic drug "straight from pipeline"
- Clinical strategy and downstream development under consideration



<u>6 months</u>: 3 lead series
200nM enzyme inhibition
cellular antiviral activity
(philanthropic funding)

<u>Current status</u>: *Lead candidates*
potent but simple
good safety data
oral exposure (blood levels)

Wellcome Trust (~£8m)

DNDi
Drugs for Neglected Diseases *initiative*

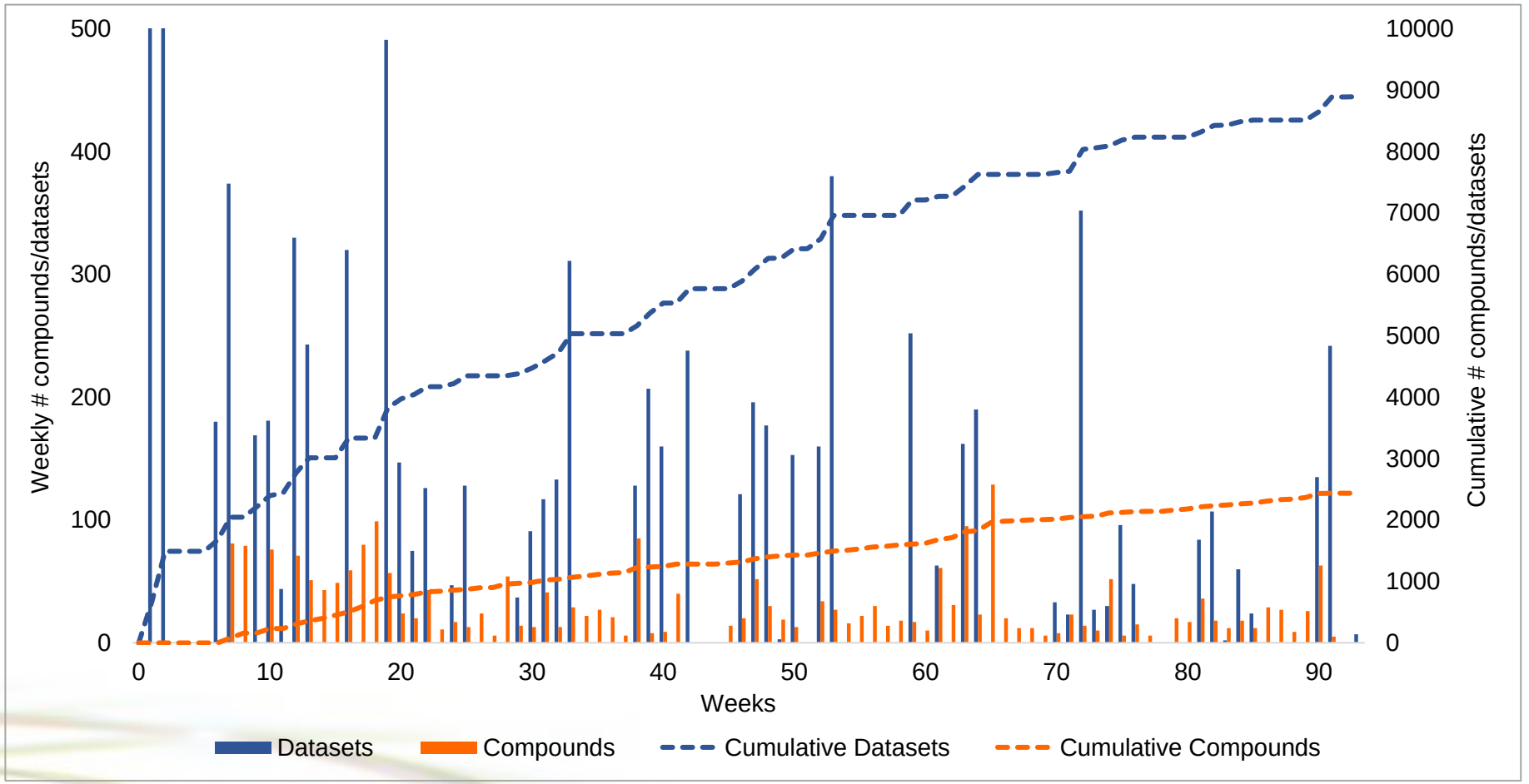
- Deliver oral & safe antiviral
- Rapid & equitable access
- Accelerated timelines

Funding
- Late lead-op, up to CTA

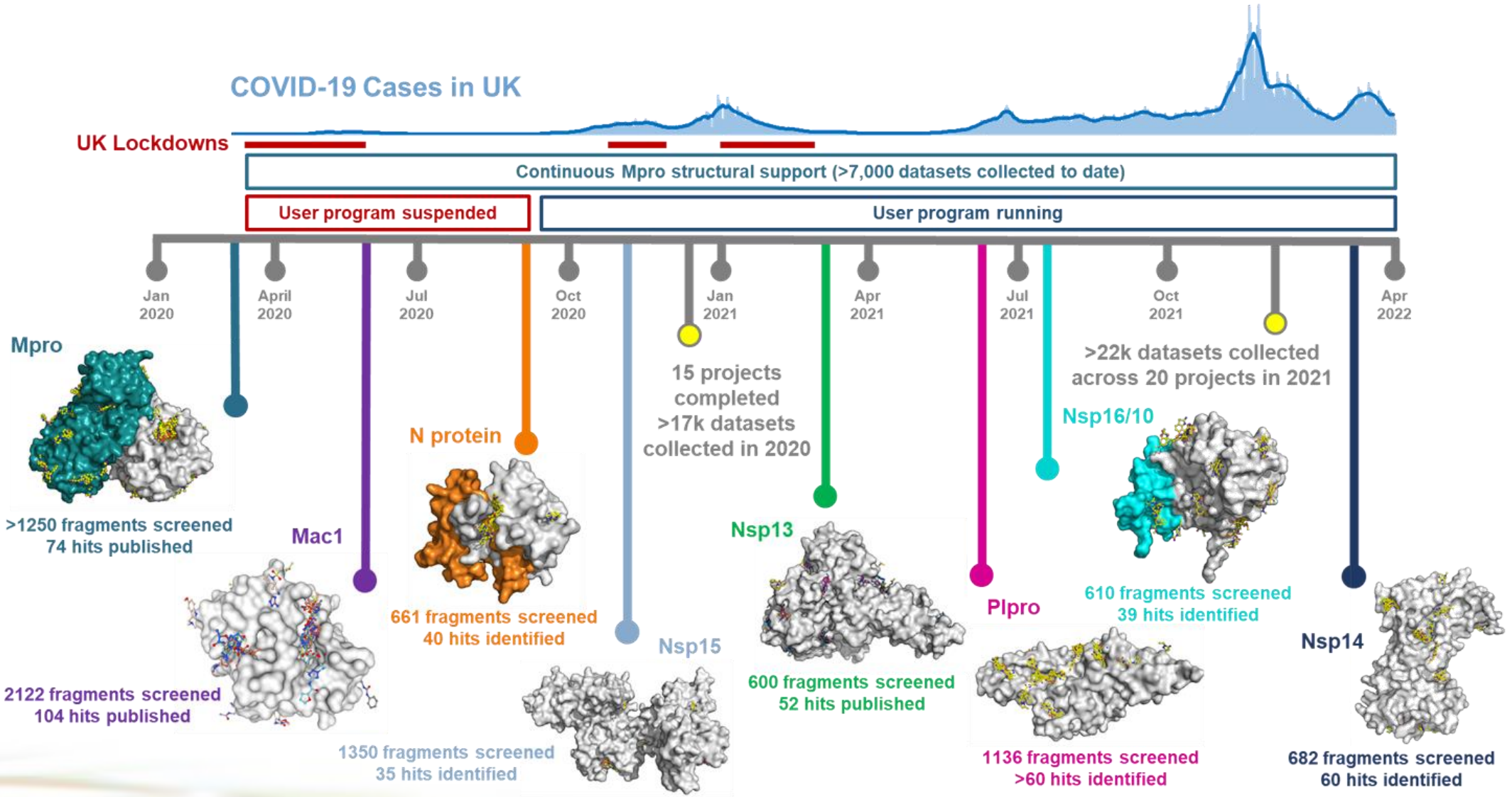
Moonshot structural data flood – all compounds tried, ~40% worked

- Goal: fast-as-possible turnaround to design team
- Goal: always available alongside assay data (selectively achieved)

Daren Fearon
(Diamond)



COVID XChem: 9 international projects during lockdown



The first data dumps: pre-preprint by webpage…

diamond | Coronavirus Science

Home | For Scientists | For Journalists | For the Public | For Staff | Diamond Website

In This Section

- Main protease structure and XChem fragment screen
  - COVID MoonShot - Taking fragments to impact
  - Electron density evidence
  - Downloads
  - Highlights on progress
  - Credits
  - FAQ
- Nsp3 macrodomain ADP-ribosyl hydrolase and XChem fragment screen
- New scientific animations
- Rapid Access
- Research Areas
- Our collaborators

Main protease structure and XChem fragment screen

Summary

To contribute to the global effort to combat COVID-19, Diamond has been able to solve a new structure of the SARS-CoV-2 main protease (M^{Pro}) at high resolution (PDB ID: 6YB7), and complete a large XChem crystallographic fragment screen against it (detailed below). Data have been deposited with the PDB, but we are making the results available immediately to the world on this page; additional work is ongoing, and updates will be continually posted here in coming days and weeks.

This work builds on the sensationally fast crystal structure of M^{Pro} at 2.16 Å in complex with a covalent inhibitor, released in January this year by Prof Zihe Rao (6LU7, published here, described here). We thus ordered the synthetic gene and cloned the full length protein as previously described for the SARS main protease (Xue et al 2007). This yielded crystals of the unliganded enzyme that diffracted to high resolution (1.25 Å) on beamline I04-1, in a different space group to the inhibitor complex, and the structure was determined and refined rapidly. ***Critically, this showed it had the active site empty and solvent accessible - perfect for fragment screening.***

So it proved: the first 600-crystal experiment could be completed in 72 hours, through growing large numbers of crystals, optimising the soaking conditions, soaking and harvesting all 600 crystals and completing the data collection run on beamline I04-1. The hits from this initial run and other details were pre-released on March 6th.

By the 24th of March, the initial 1500-crystal experiment was complete, and the results made publicly available. Screening additional libraries throughout April brought the total number of active site fragments to 71, with 48 fragments binding covalently (full timeline here and download page here). This was an exceptionally large screen which yielded a remarkably rich readout, with vast opportunities for fragment growing and merging.

We have already triggered computationally-driven follow-up work internally, and externally joined forces to launch a fully-open crowdsourcing and crowdfunding initiative – the COVID Moonshot - to establish urgently the shortest route possible to clinical impact by maximally exploiting the readout - you can help, read more here.

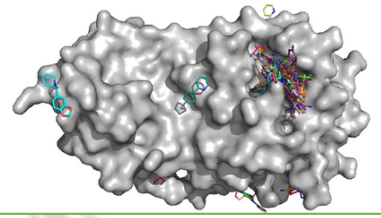
On the 11th of May, the first biochemical and structural data from Moonshot compounds was released and by the 12th of June over 500 compounds had been tested, demonstrating that the design-make-test process is fully in place.

XChem fragment screen

The initial screen encompassed multiple fragment libraries: the DSI-poised library, MiniFrags (Astex) FragLites & Peplites (CRUK Newcastle Drug Discovery Unit (Newcastle University)), York3D (University of York), SpotFinder and heterocyclic electrophilic fragment library (Hungarian Academy of Sciences) and an electrophilic fragment library designed and pre-screened by mass spec at the Weizmann Institute (see below).

There were 74 hits of high interest - data and extensive details are here, and some interactive views here:

- 23 non-covalent hits in the active site
- 48 covalent hits in the active site
- 3 hits in the dimer interface, one in a calculated hotspot



diamond | Coronavirus Science

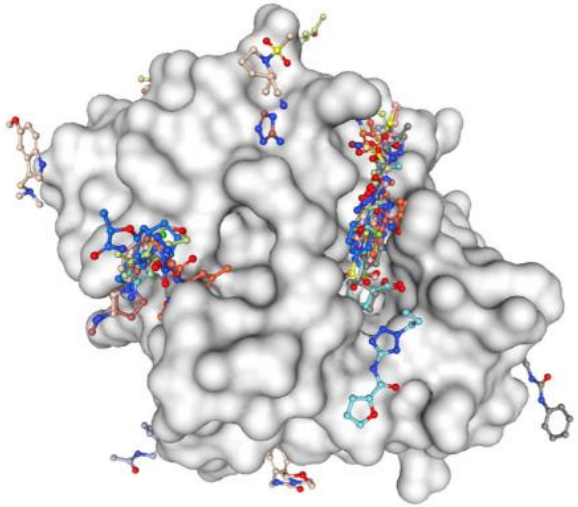
Home | For Scientists | For Journalists | For the Public | For Staff | Diamond Website

In This Section

- Main protease structure and XChem fragment screen
- Nsp3 macrodomain ADP-ribosyl hydrolase and XChem fragment screen
  - Highlights on progress
  - XChem Fragment Screen
  - International collaboration
  - Electron density evidence
  - Interactive views and downloads
  - Credits
  - FAQ
- New scientific animations
- Rapid Access
- Research Areas
- Our collaborators

Nsp3 macrodomain ADP-ribosyl hydrolase and XChem fragment screen

Summary

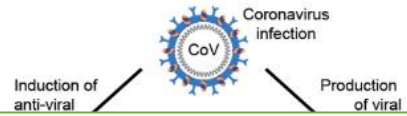


In light of the ongoing coronavirus (COVID-19) pandemic, the Ivan Ahel Laboratory at the University of Oxford joined together with the XChem team at Diamond Light Source to contribute to the current global efforts discovering and developing much-needed novel antiviral therapeutic possibilities.

For this, the Ivan Ahel team has been able to rapidly solve the SARS-CoV-2 macrodomain structure to near atomic resolution, and then, together with Frank von Delft's XChem team, completed a large crystallographic fragment screen against this target.

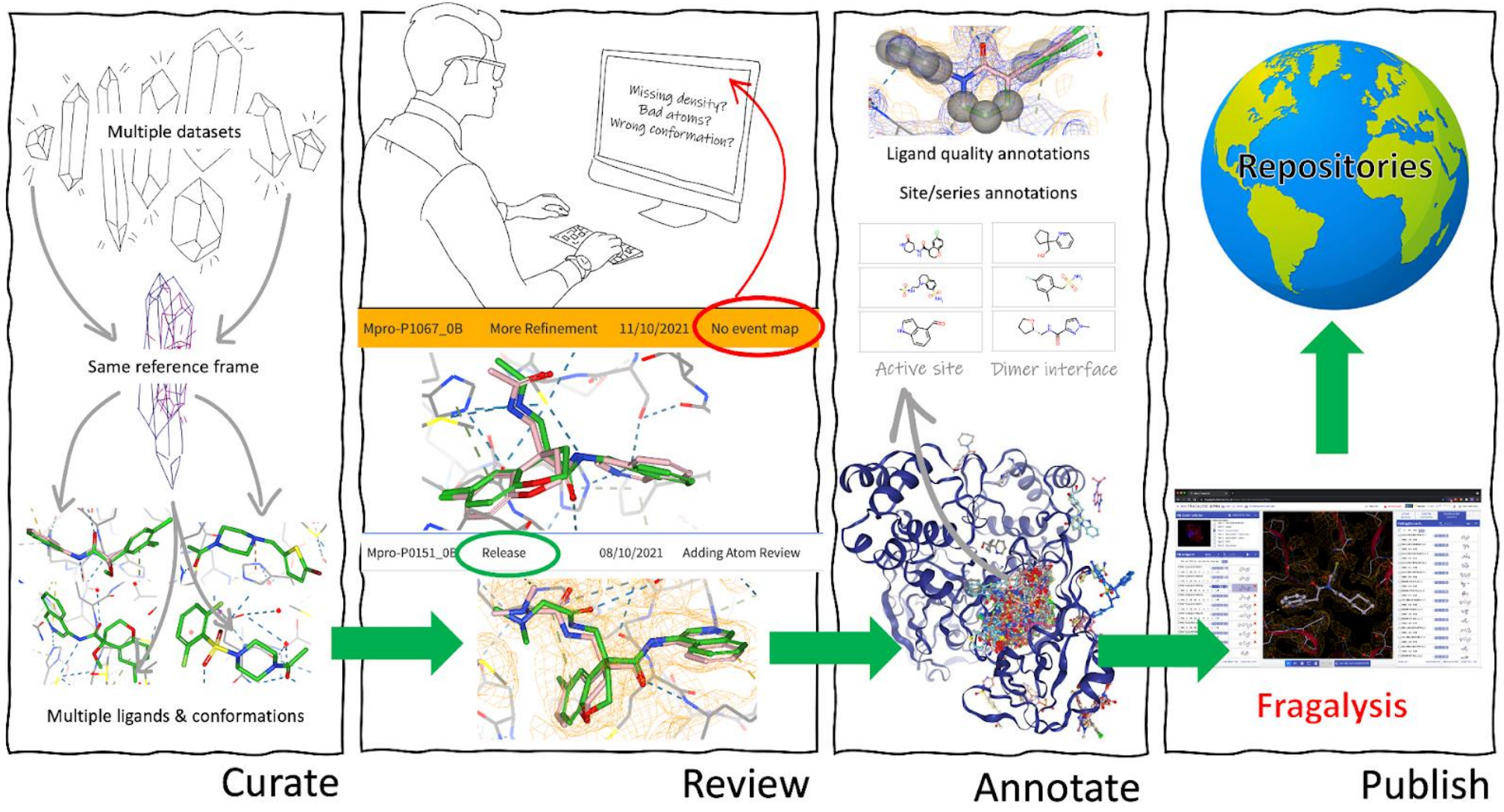
The macrodomain (also called Macro X domain) is a 150 amino acid protein module with (ADP-ribosyl)-hydrolase activity that is a part of the SARS-CoV-2 multidomain protein nsp3. ADP-ribosylation is a reversible post-translational modification of proteins synthesised by the PARP family of enzymes, and regulates many pathways in human cells, including the DNA damage response and antiviral defences. Several of the human PARPs such as PARP10, PARP13 and PARP14 act specifically as antiviral proteins to generate an antiviral environment and prevent virus replication. In contrast, the viral macrodomain removes the ADP-ribosylation modifications, thereby enabling the virus to counteract the PARP-induced innate immunity (Fehr et al, Trends Microbiol, 2018; Figure 1).

The macrodomain enzyme represents a promising drug target for the treatment of coronavirus infections, since macrodomain-deficient viruses (including all studied coronaviruses, alphaviruses and hepatitis E virus) are unable to replicate in human cells (eg. Fehr et al, mBio, 2016). Nevertheless, no inhibitors for this enzyme have been developed to date.



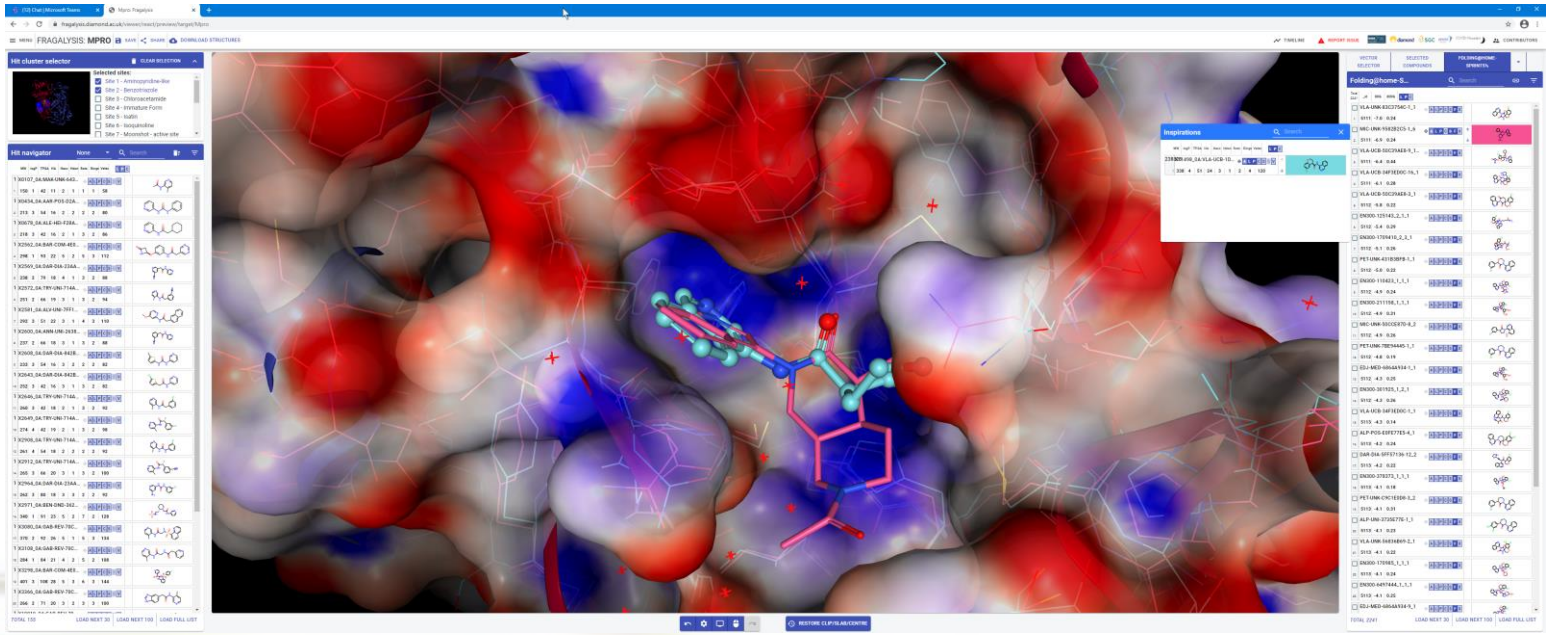
https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html

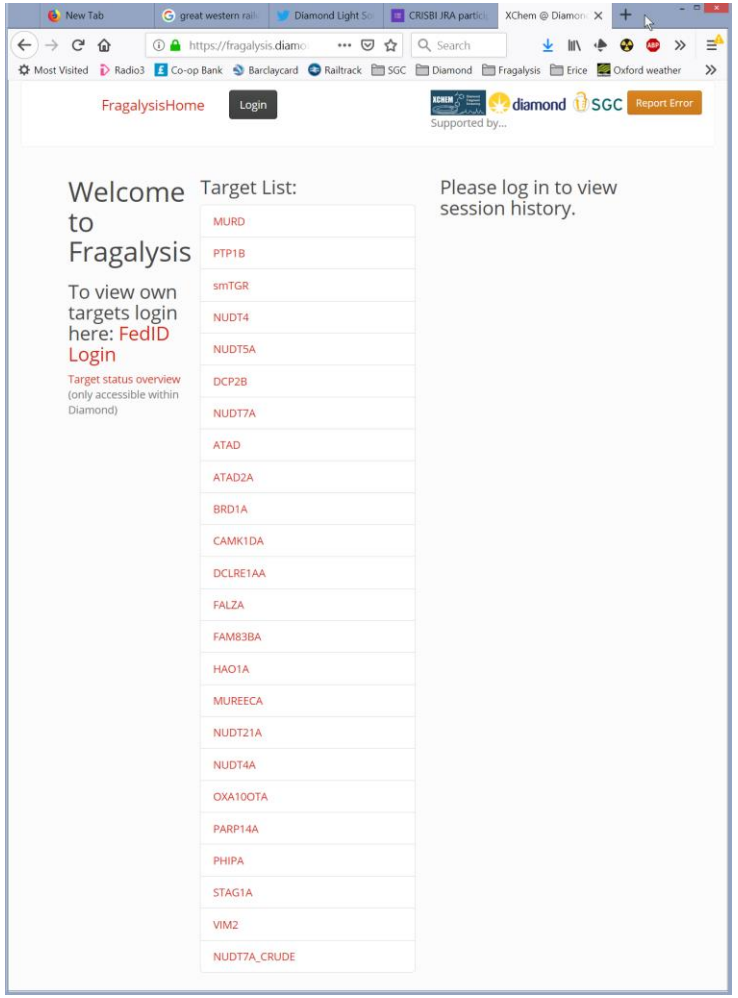
Achieving 3D Data: Curate, Review, Disseminate - Fragalysis



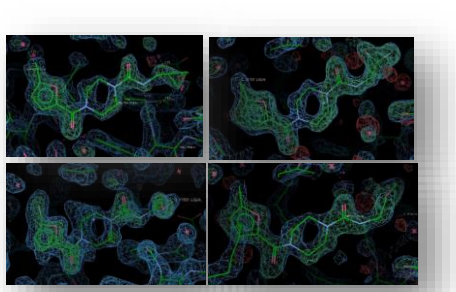
Fragalysis Cloud : 3D data sharing and collaboration

- https://fragalysis.diamond.ac.uk
- Collaboration tool for 3D data – ligand-centric
- Rigorously engineered Cloud stack – **deployed elsewhere**
- Scientifically ***alpha/beta***
- Lots of open data (>30 targets - SGC & COVID) – **all FAIR**

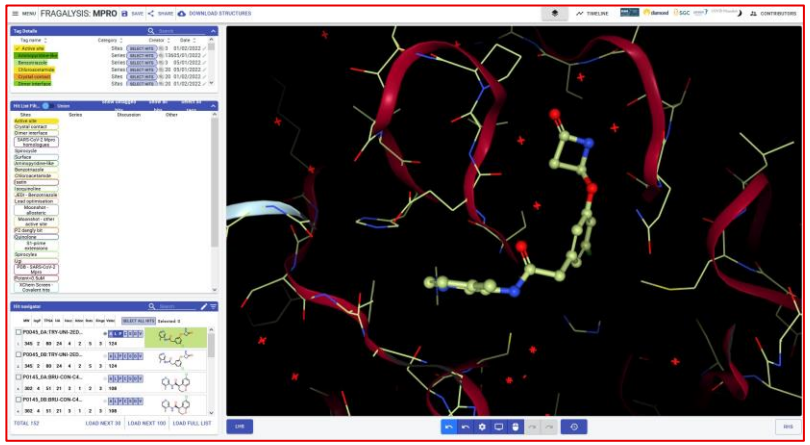




Fragalysis goal: supports sharing and analysis



Review and release data

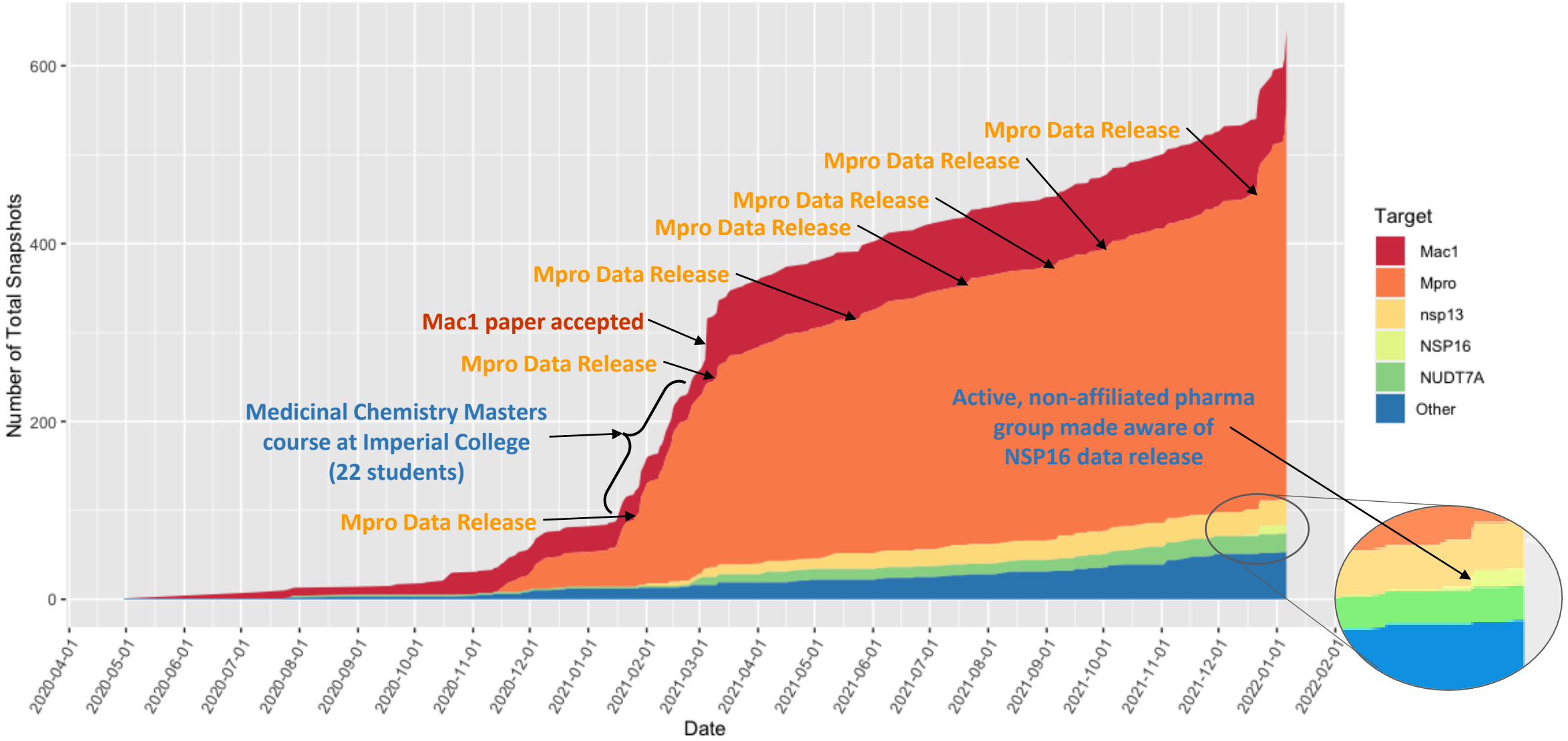


Data snapshot

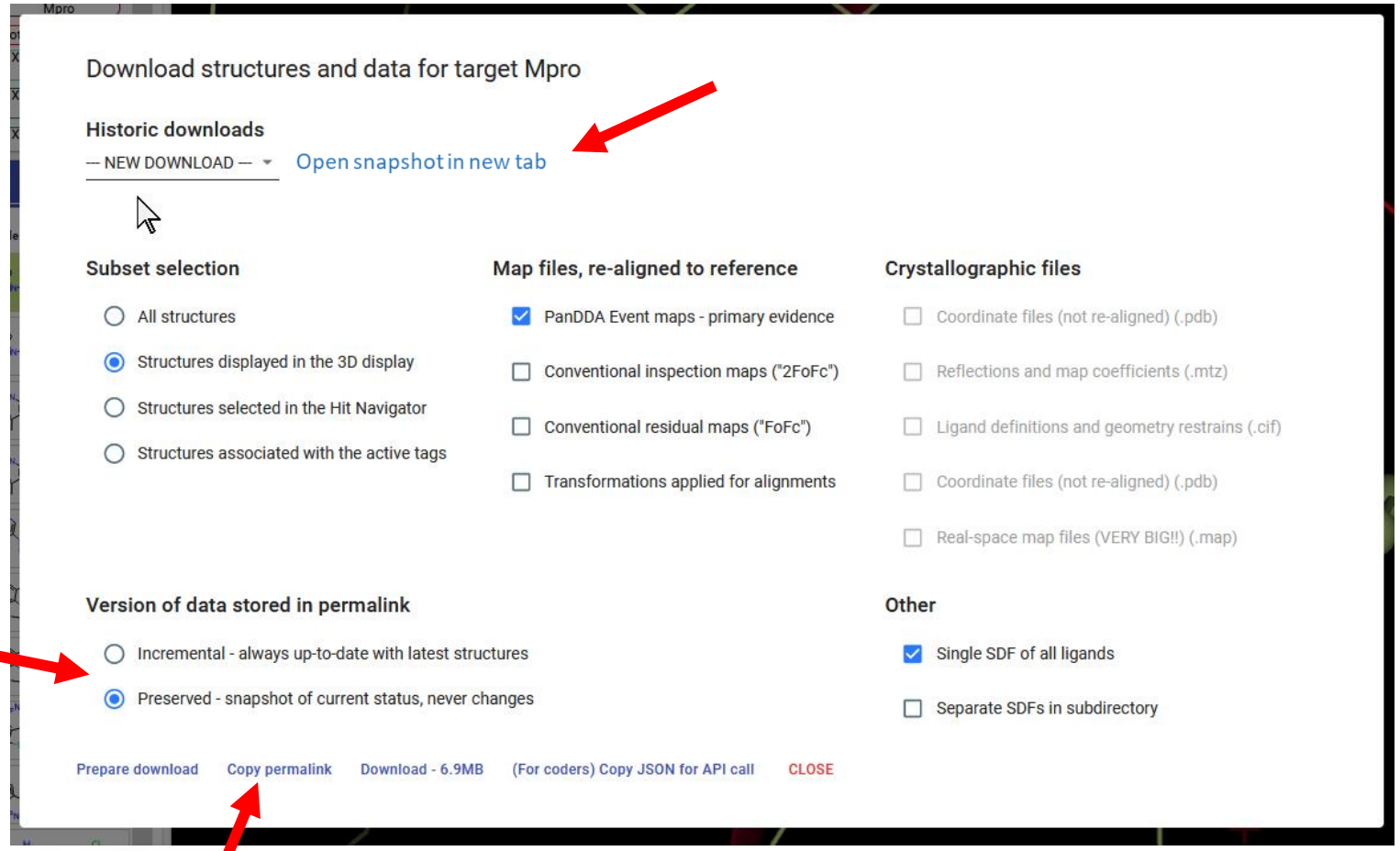
Data download

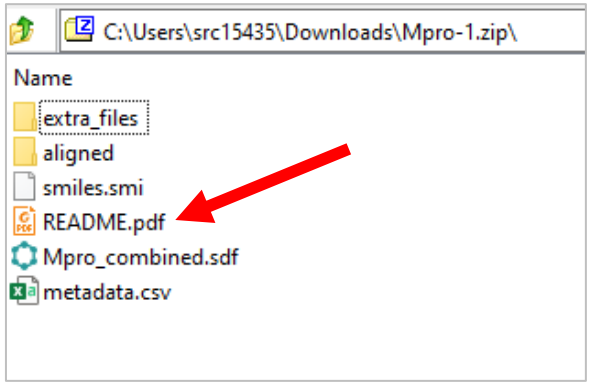
Computation

Timeline of snapshots suggests: both useful and in demand



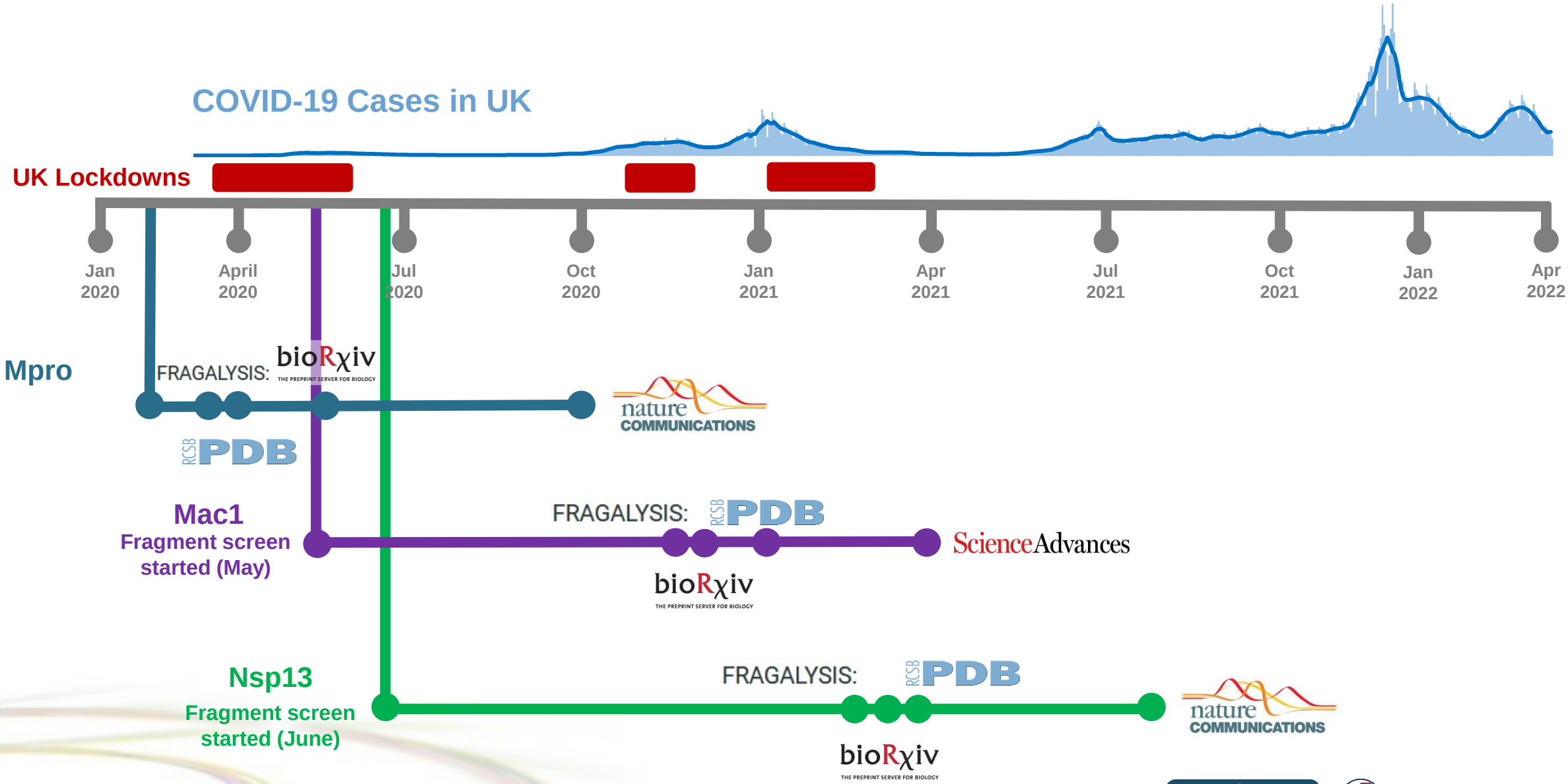
Fragalysis downloads: FAIR and self-documenting





https://fragalysis.diamond.ac.uk/viewer/react/download/tag/15139fac-7c08-423d-b04e-46d820e4aef9

Publication lags dissemination lags data generation…



Context-free dissemination creates problems

Protein Sci. 2022 Apr;31(4):784-786. doi: 10.1002/pro.4271. Epub 2022 Feb 21.

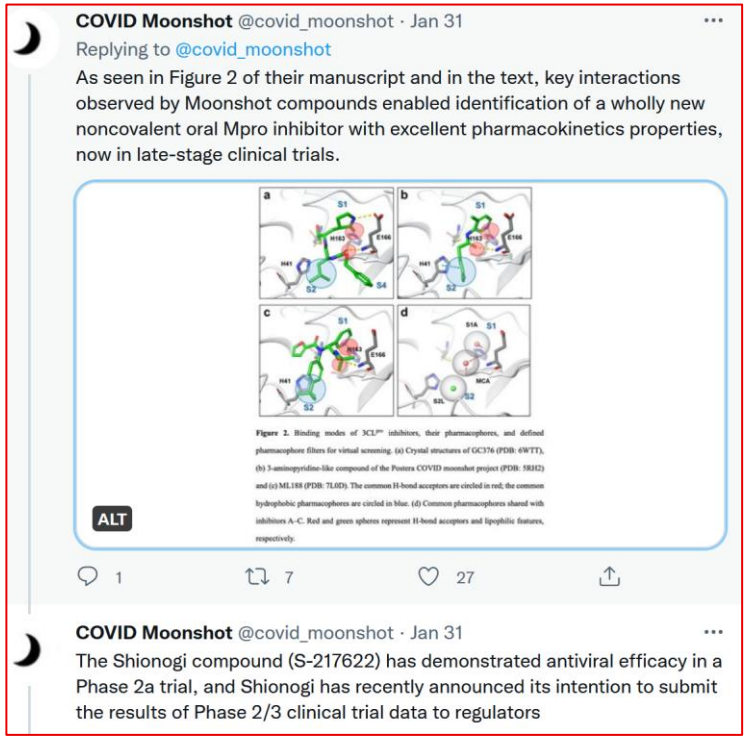
Group depositions to the Protein Data Bank need adequate presentation and different archiving protocol

Mariusz Jaskolski [1 2], Alexander Wlodawer [3], Zbigniew Dauter [3], Wladek Minor [4], Bernhard Rupp [5 6]

£8m Moonshot

£80m ASAP

£800m Shionogi project



Psychology of publication…

- Why can journals still charge?
  - Polemically: we're fools (see what Robert Maxwell did)
  - But maybe: we're trained to be scared/worried

- Why did preprints shoot up in pandemic?
  - Urgency finally trumped fear???

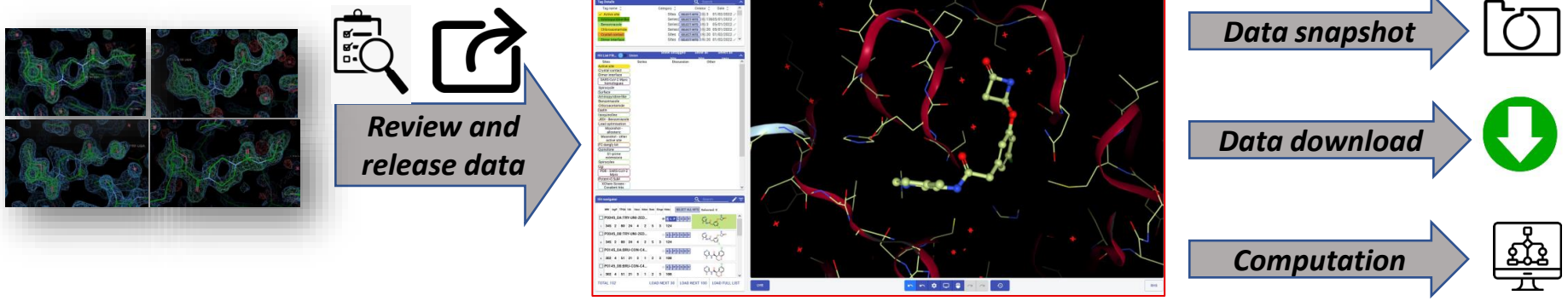
- Data release: my 20 years of Structural genomics: release always scary
  - JCSG, SGC, TEPs, Moonshot

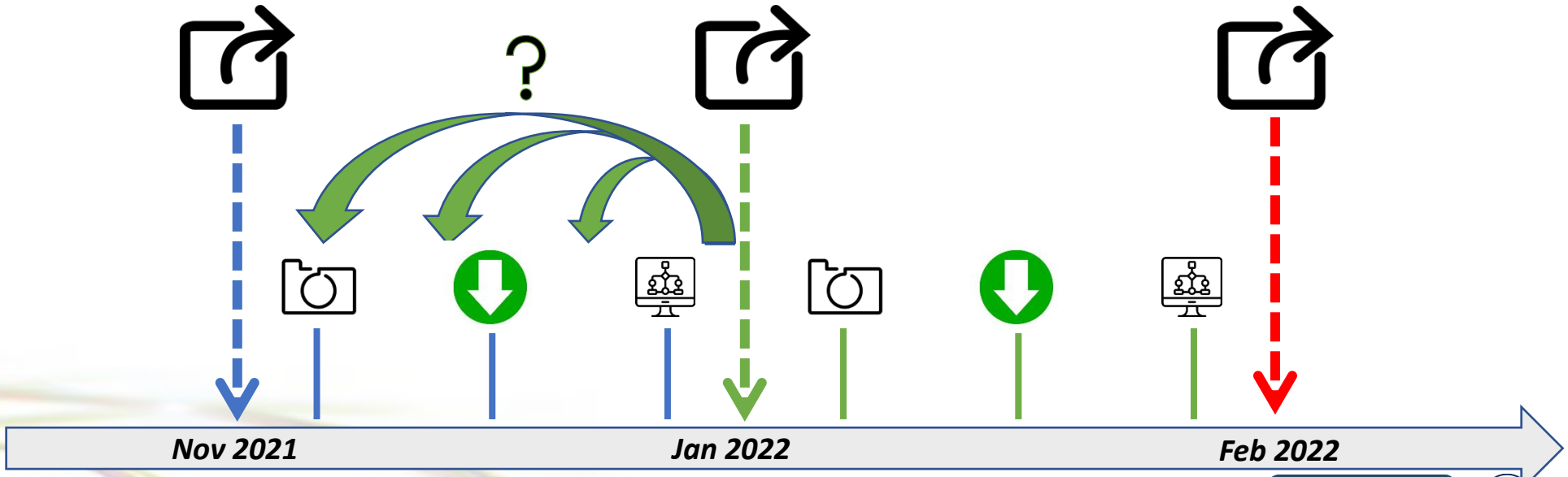
Psychology of publication: fear

Data release

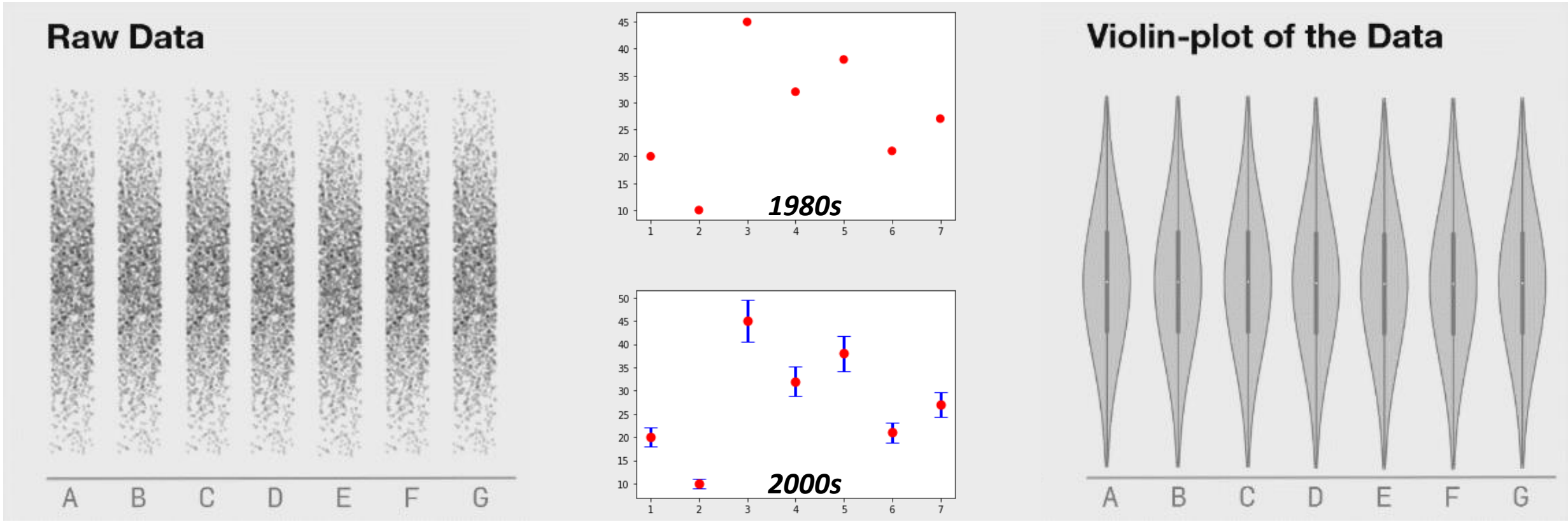
ILLUSTRATION: JOHN S. DYKES (WSJ)

Challenge: Juggling consistency vs usage



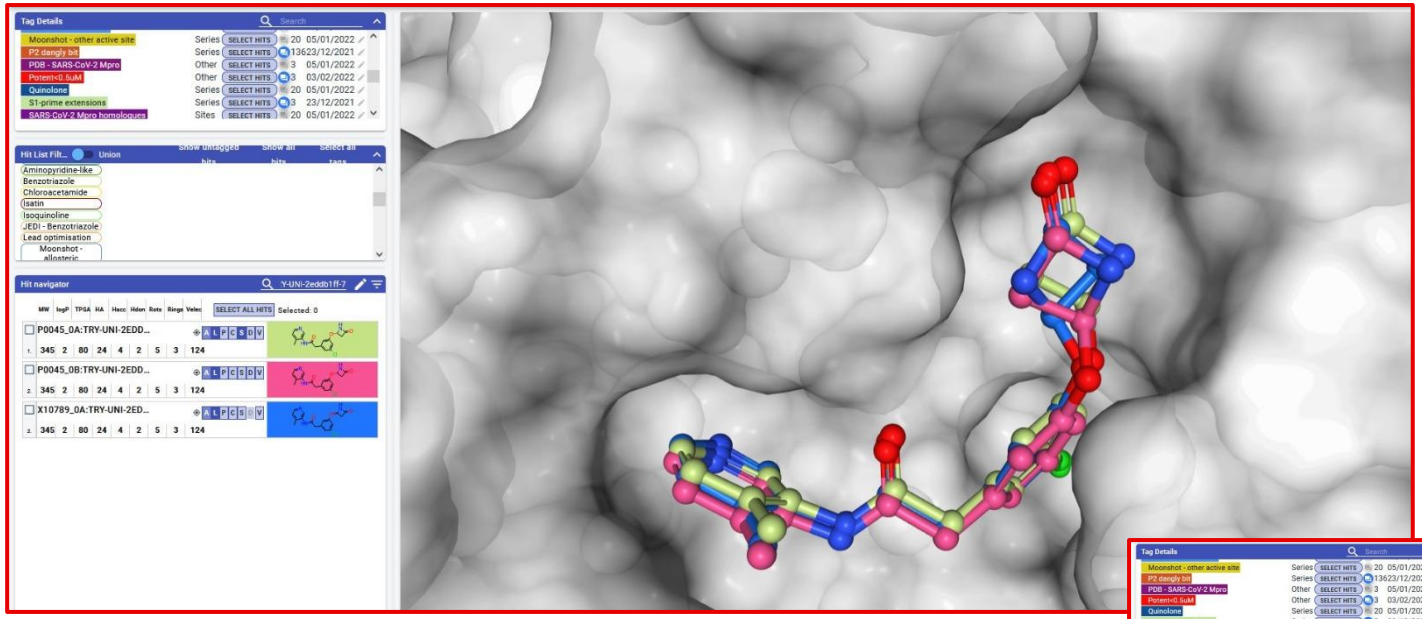


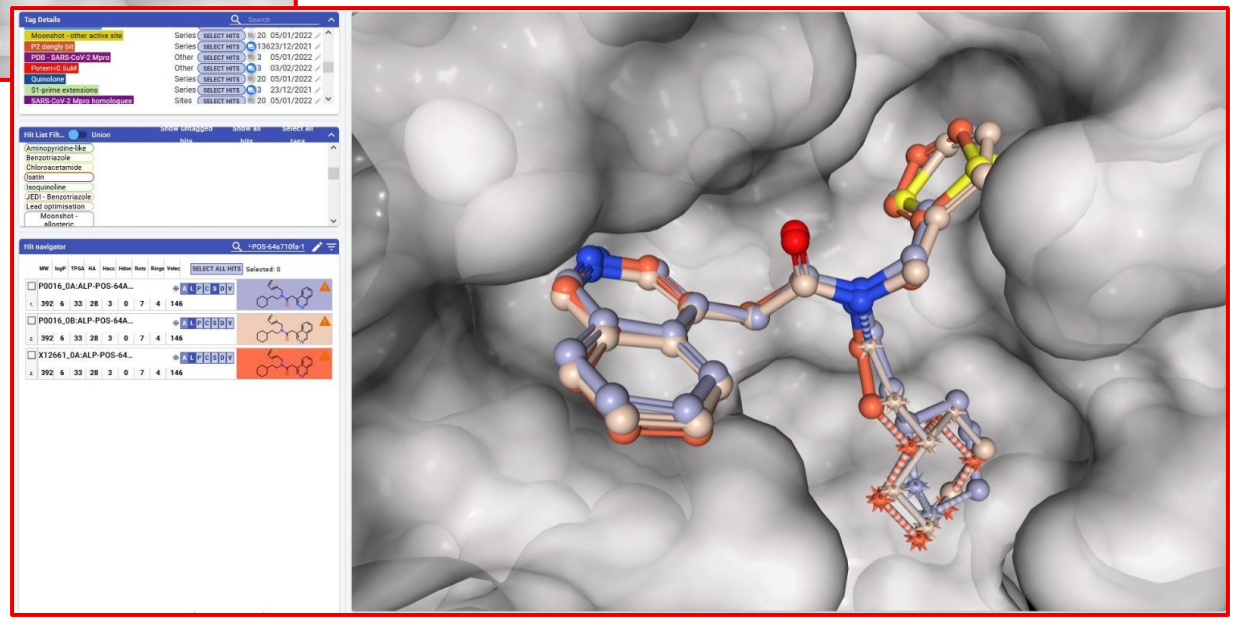
Evolution of what's acceptable in 2D plots



2020s

Structural data can/should be *SO* much richer…





Open question 2: Discoverability

- ASAP will generate a LOT of structural data
- How can people know what's there?

- Googleable snapshots?
- Web 2.0 technologies?
- Page-rank approach?
- Social media approach ("maximize eyeballs")?

Open question 3: Generate better data?

- Exploit annotation & exploration
  - On-platform analysis
  - Capture of conclusions
  - Sign-off mechanism
  - XYZ-located discussion threads

- Annotations can be better – cf. web content annotation (missing blogpost reference…)
  - Mistreat annotators as low-skilled ➔ poor data, poor "AI"
  - Embrace annotators as high-skilled ➔ "good" AI becomes achievable

- Converting to reality
  - Rapid assimilation of annotations
  - Make implications/outcomes visible to annotator
  - Implement interrogation of model/rules
  - Give agency to edit/review

- Is this what Fragalysis should support? - **A big ask…**