Movies of proteins at work

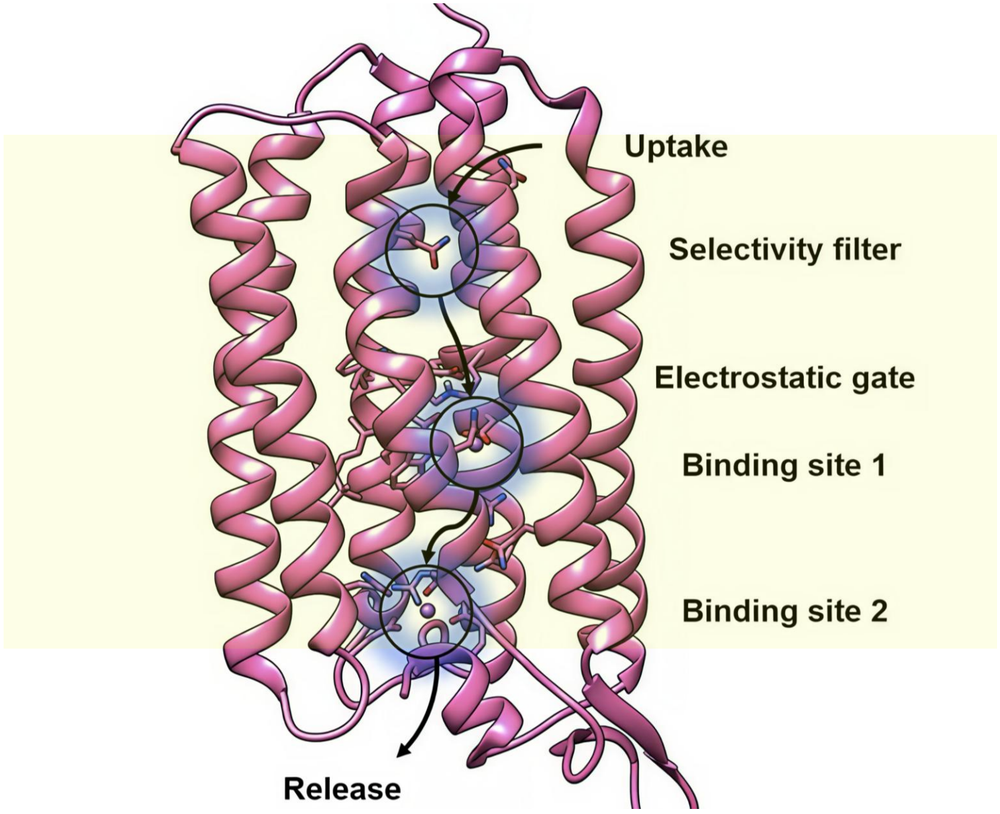
— and why they’re harder to make than they look

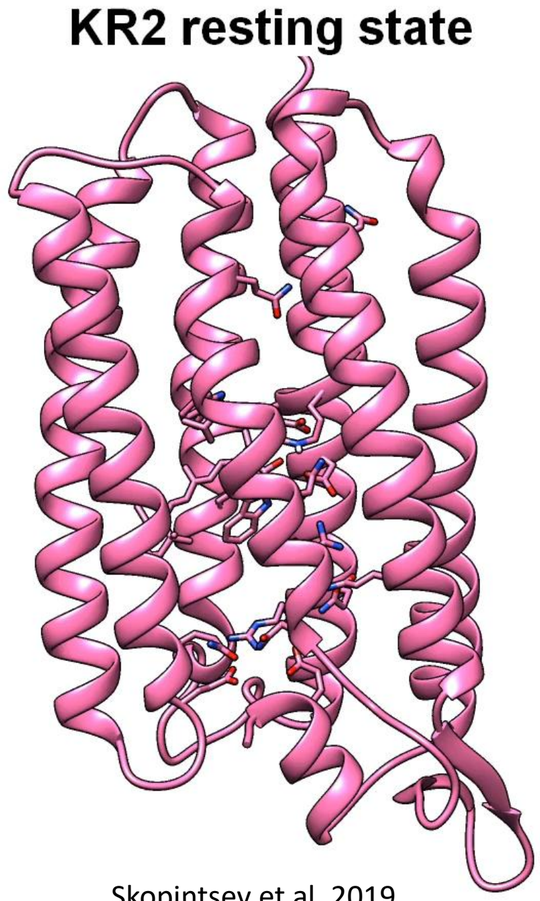
Tobias Weinert · Senior Scientist · Paul Scherrer Institute · PSI Center for Life Sciences

VII ALBA Users’ Meeting · 04.09.2026

Why movies?

A three-dimensional movie can explain function directly

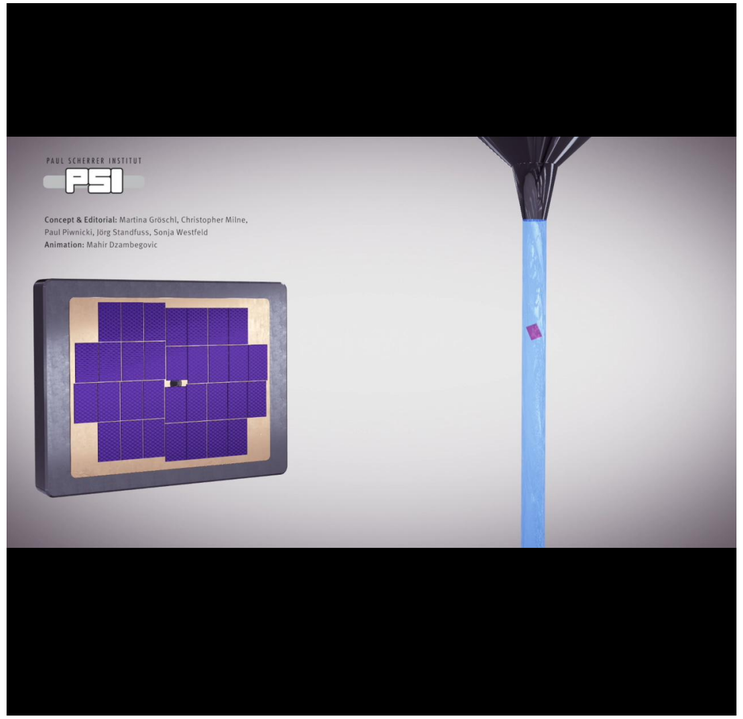




Skopintsev et al. 2019

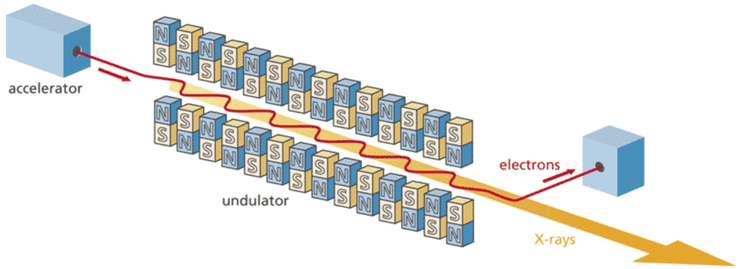
How can we make them?

Three things you need for a time-resolved experiment



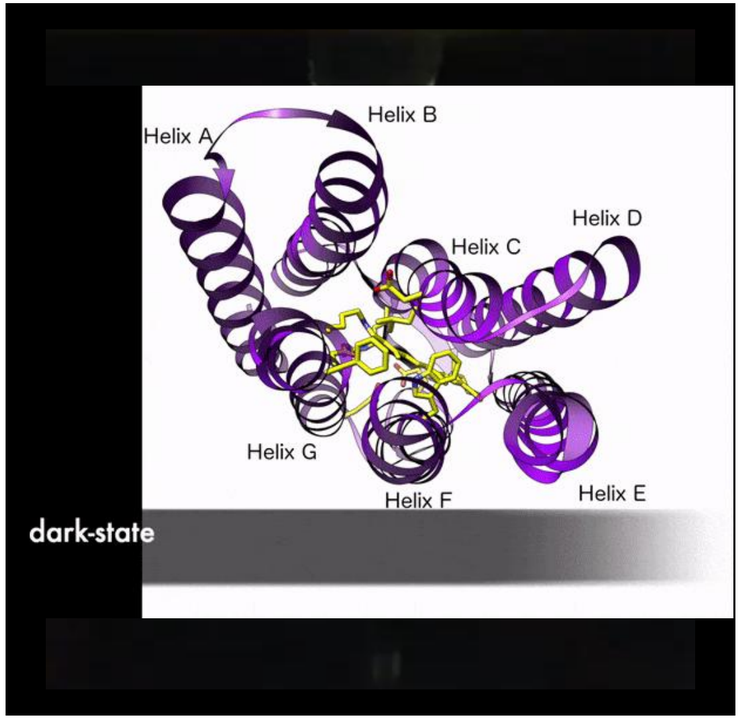
A trigger

Something that starts the reaction on command — a laser pulse for photoswitches, a rapid mix for enzymes.



A fast X-ray pulse, or a fast detector

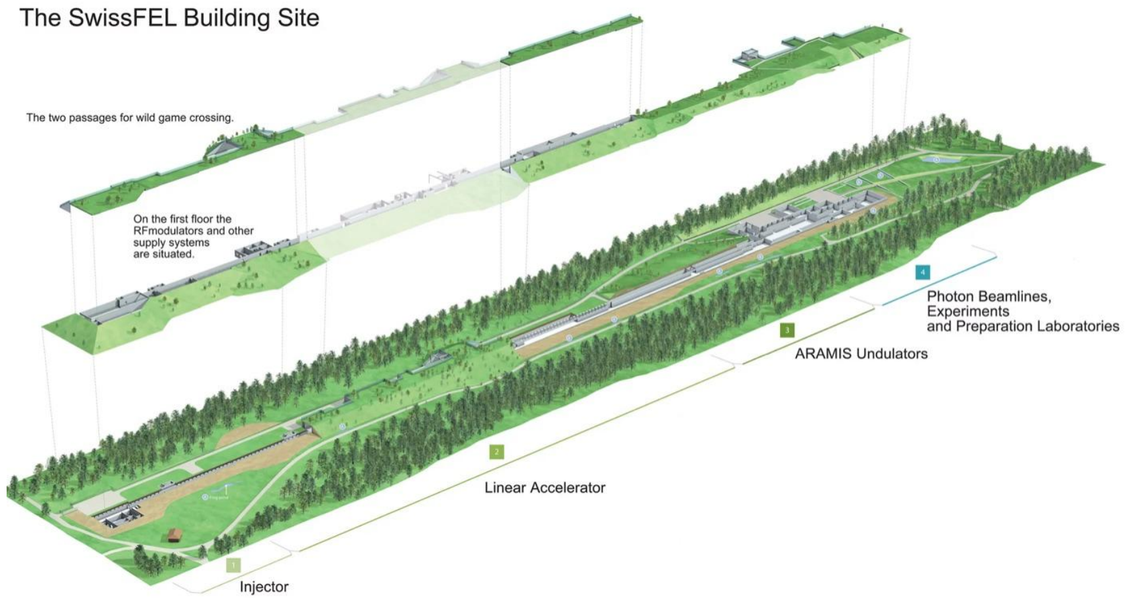
Pulses as short as femtoseconds freeze each moment like a strobe. One pulse, one snapshot.



Many crystals

Each crystal gives one shot — destroyed by the pulse at an XFEL, displaced out of the beam in serial millisecond crystallography. Thousands of fresh microcrystals flow through.

Two X-ray sources, two regimes



X-FEL · SwissFEL / SACLA / EuXFEL

Time resolution	femtoseconds
Shots per second	30–120 Hz, or MHz — scarce beam time
Best for	Ultrafast photochemistry, irreversible reactions

4th-gen synchrotron · SLS 2.0 / ESRF / ALBA (soon) / MAX IV

Time resolution	milliseconds
Shots per second	High throughput — abundant beam time
Best for	Slower processes, routine serial crystallography

PSI Center for Life Sciences

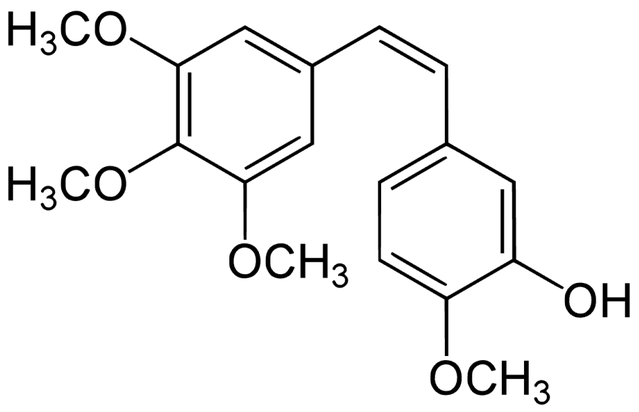
Not many proteins can be triggered by light. How do we “fix” that?

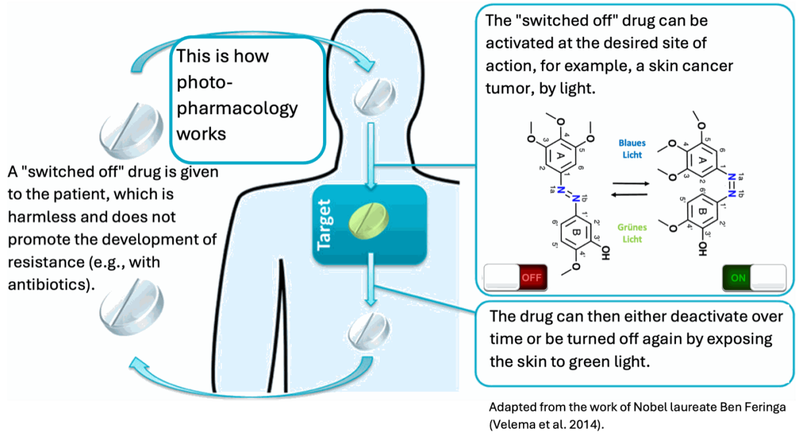
THE “EASY” WAY

A photoswitch · tubulin · femtoseconds to milliseconds

From bark to photoswitch — a cancer drug switched on by light

Replace the C=C bridge of combretastatin A4 with an azo group and you get **azo-CA4**: OFF in the dark (*trans* , inactive), ON at λ = 350–430 nm (*cis*, binds the colchicine site), and back at λ = 450–530 nm or spontaneously.





Adapted from the work of Nobel laureate Ben Feringa (Velema et al. 2014).

From the bark

The African bushwillow, *Combretum caffrum* — the natural source of the parent compound.

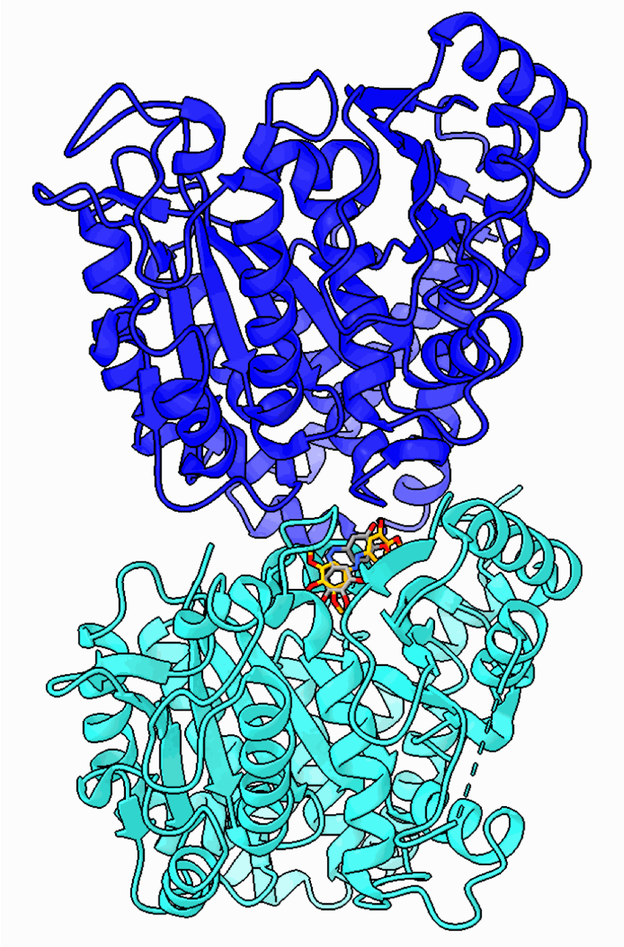
Combretastatin A4

The parent natural product: a potent tubulin binder at the colchicine site, and the scaffold the azo group replaces.

Why it matters

The drug is only active where the light is. Toxicity stays out of the rest of the patient.

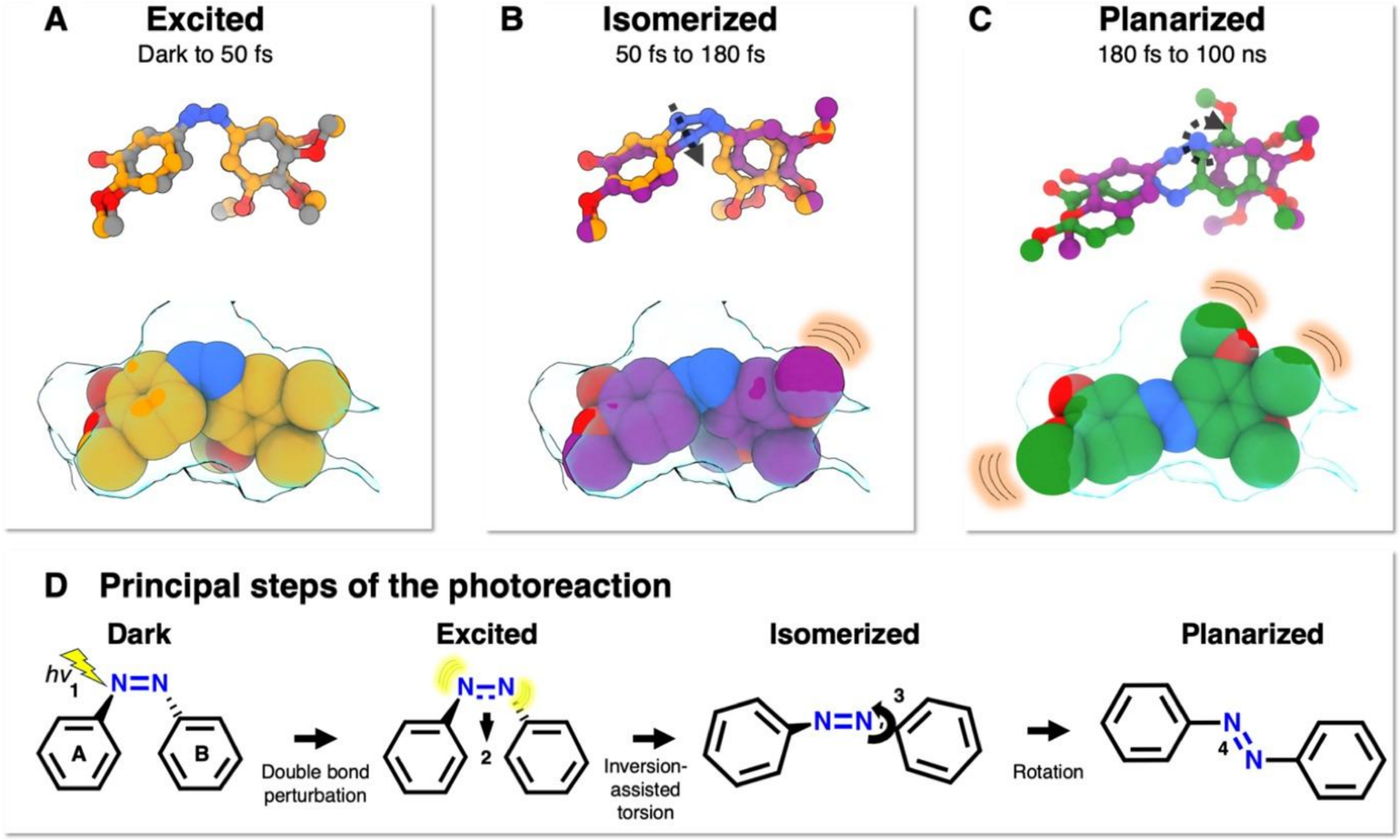
From photopharmacology to time-resolved serial crystallography



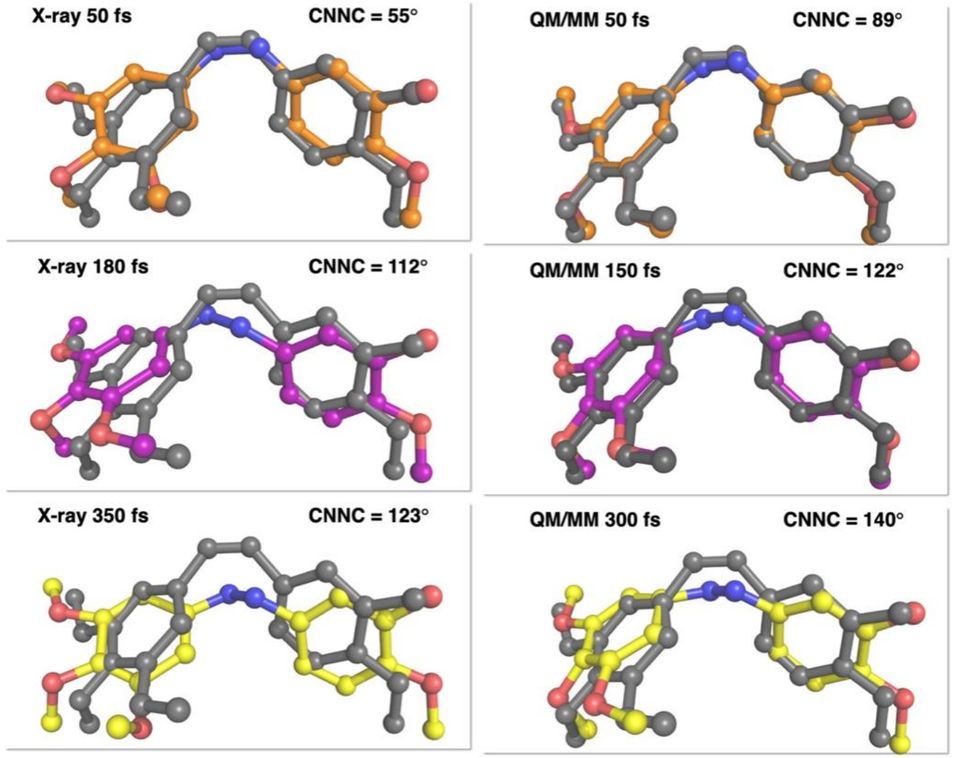
Green light resets every crystal to *trans* before it reaches the beam. The pump laser fires, the delay **Δ**t is scanned, and each time point is merged from thousands of crystals.

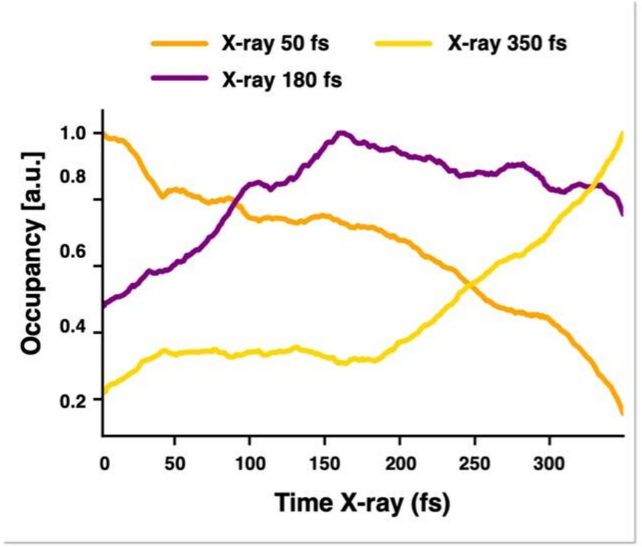
DELAYS MEASURED
–500 fs · –230 fs
450 fs — the fs scan
1 ps · 25 ps · 35 ps
125 ps · 1 ns · 10 ns
100 ns

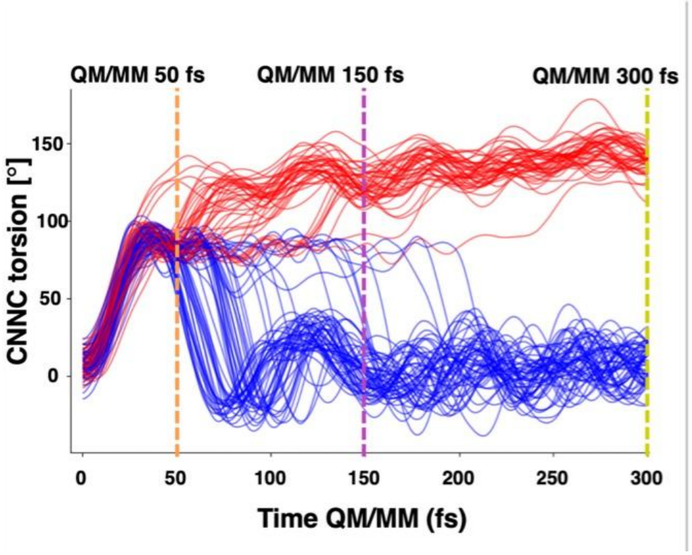
Chemistry — azobenzene isomerising inside a protein, femtoseconds to nanoseconds



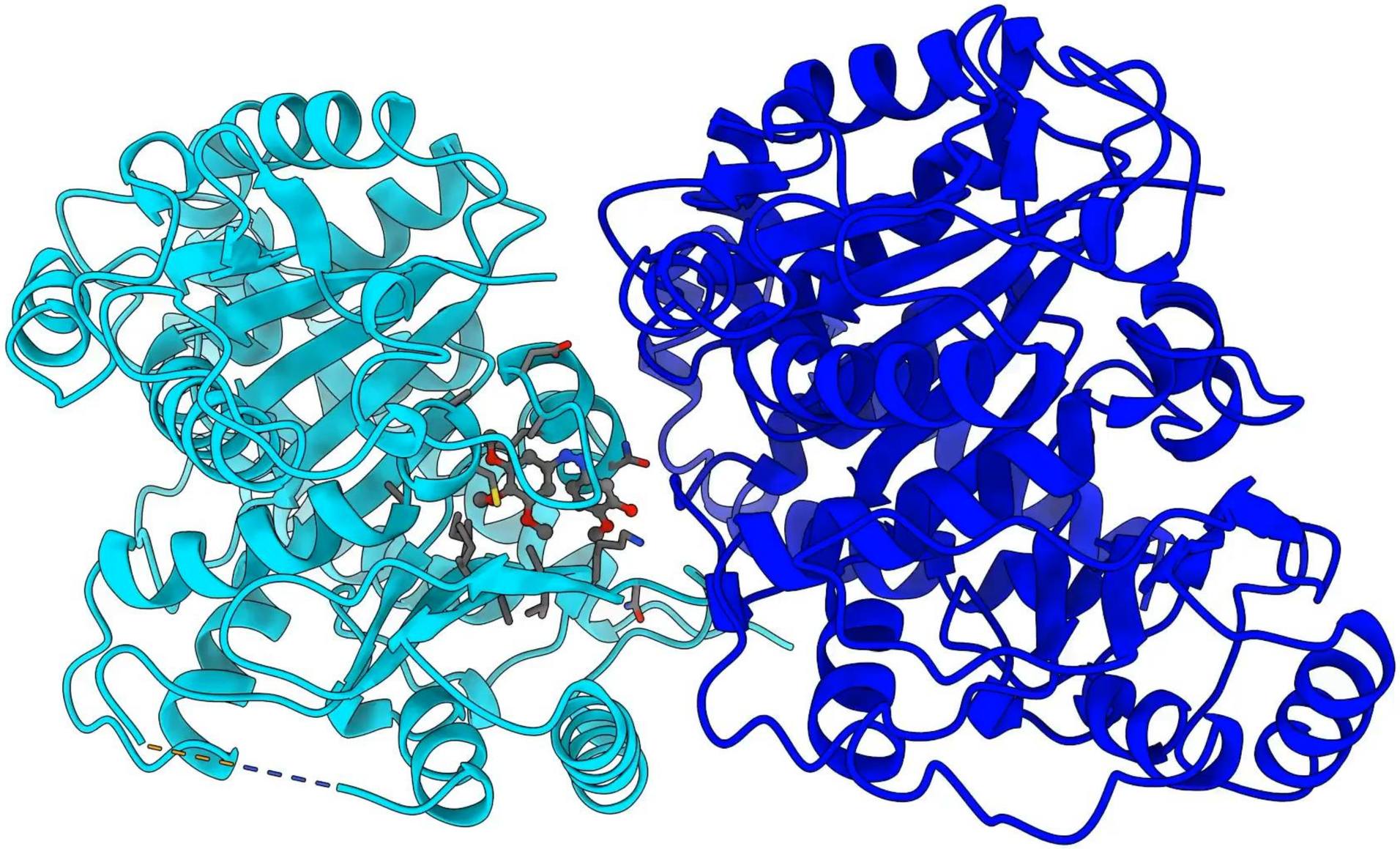
The femtosecond scan — states are already convoluted



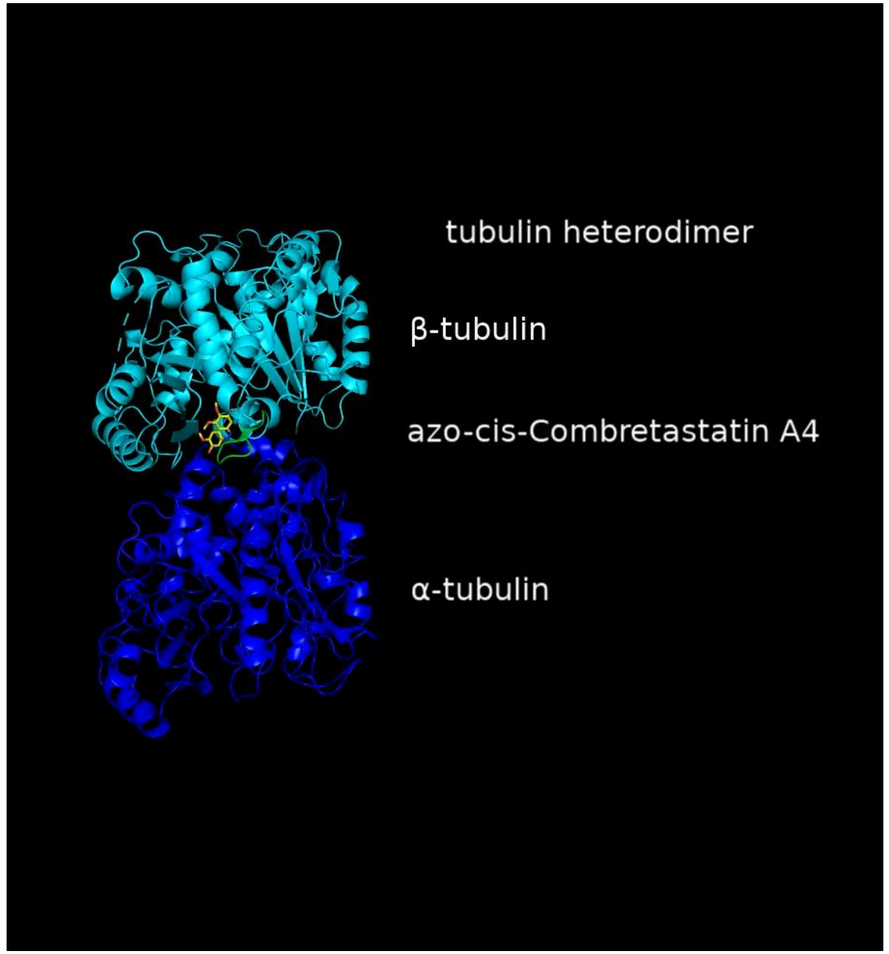


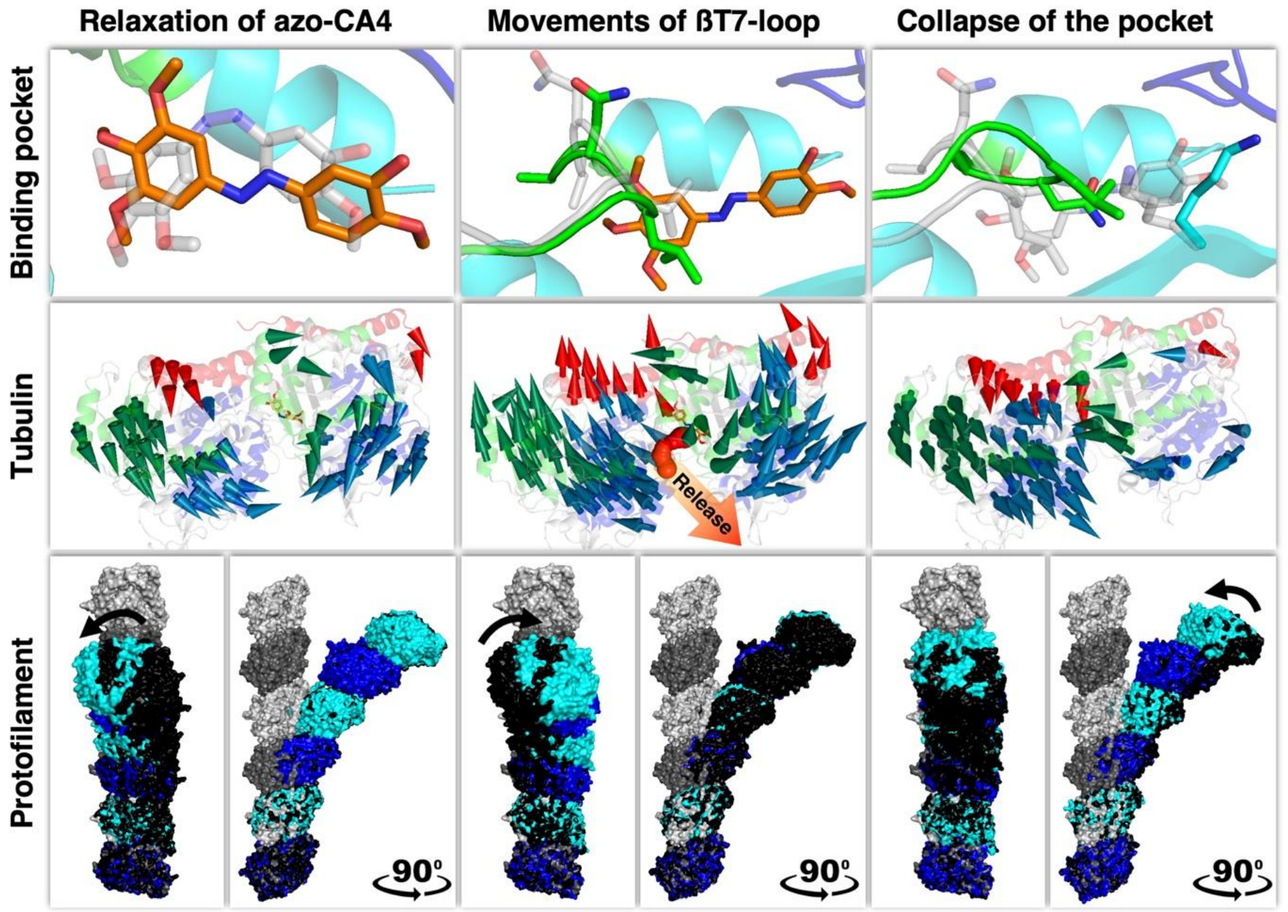


Watching Azobenzene Isomerization Inside a Protein



Pharmacology — watching a ligand unbind from its target





Wranik et al. 2023

Beautiful — and expensive

What we gain

- Femtosecond snapshots
- Atomic resolution at room temperature
- Undamaged structures
- Full reaction trajectories

What it costs

- XFEL beam time is extremely scarce
- Millions of crystals per dataset
- Complex sample delivery
- Data volumes of terabytes per hour

And it still assumes the reaction has a trigger, and that one state is present at a time. The next system has neither.

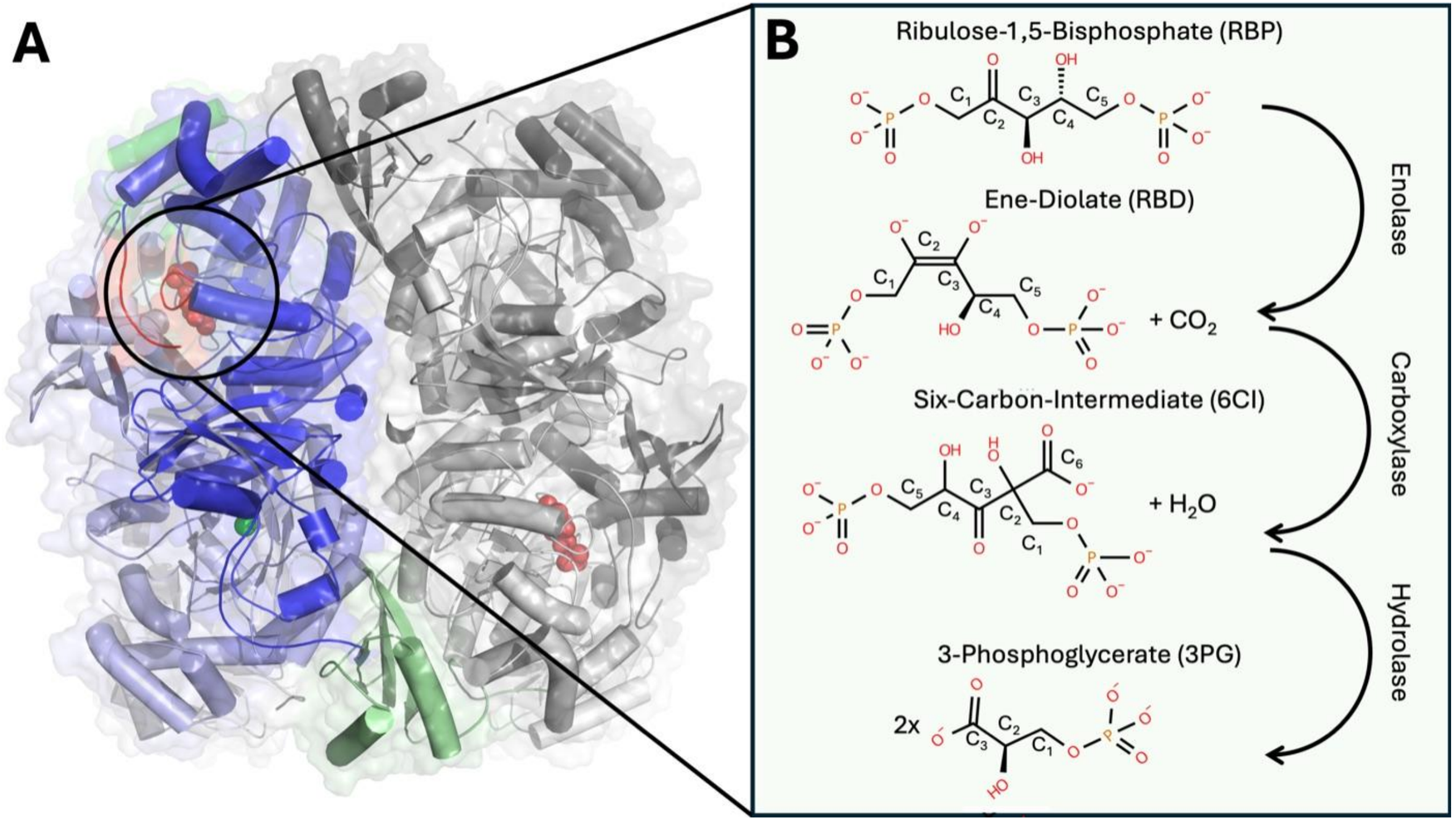
RuBisCO

THE “HARD” WAY

Four reactions · convolution of states · no trigger

Most of this work was done by **Georgii Khusainov**, who is in the room today.

RuBisCO — the most important enzyme on Earth

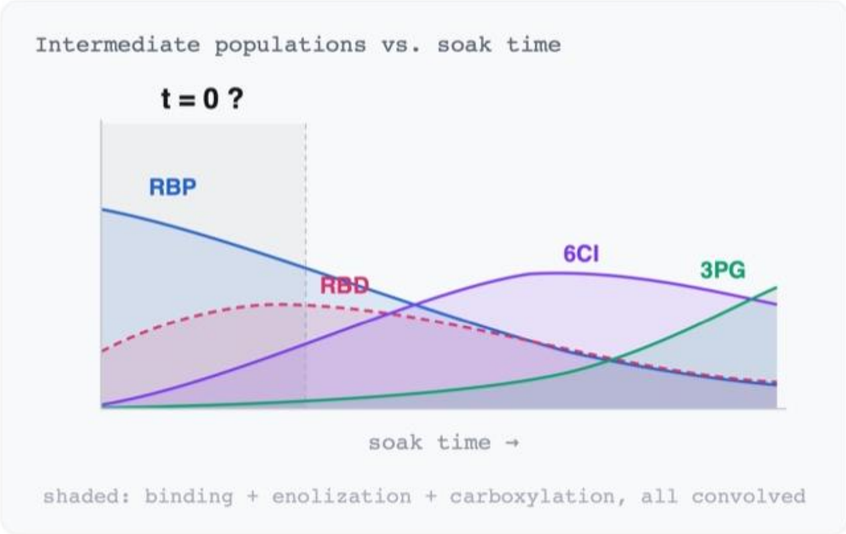


Convolution of states — usually ignored in crystallography

Four reactions run in series, and none of them waits for the last to finish. At no point in the time course is a single intermediate alone in the crystal — so every map is a weighted sum of states, and refining one model into it gives an average that matches nothing.

■ THE LIMITATION

The bottleneck: intermediates overlap, with no clean t = 0



01

States populate as mixtures

Every dataset is a superposition of intermediates — structures only emerge after heavy deconvolution.

02

No defined start, no tunable rate

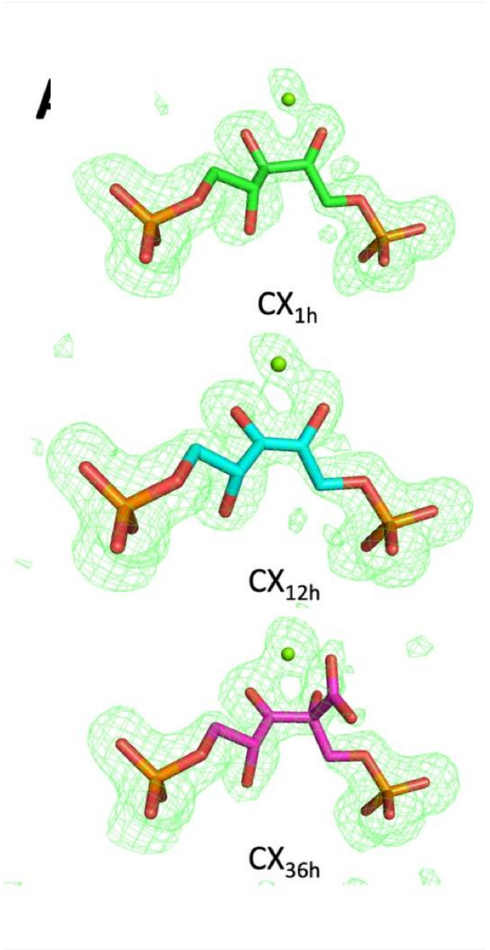
Binding, Mg^{2+} expulsion and carboxylation onset all overlap; t = 0 is indeterminate.

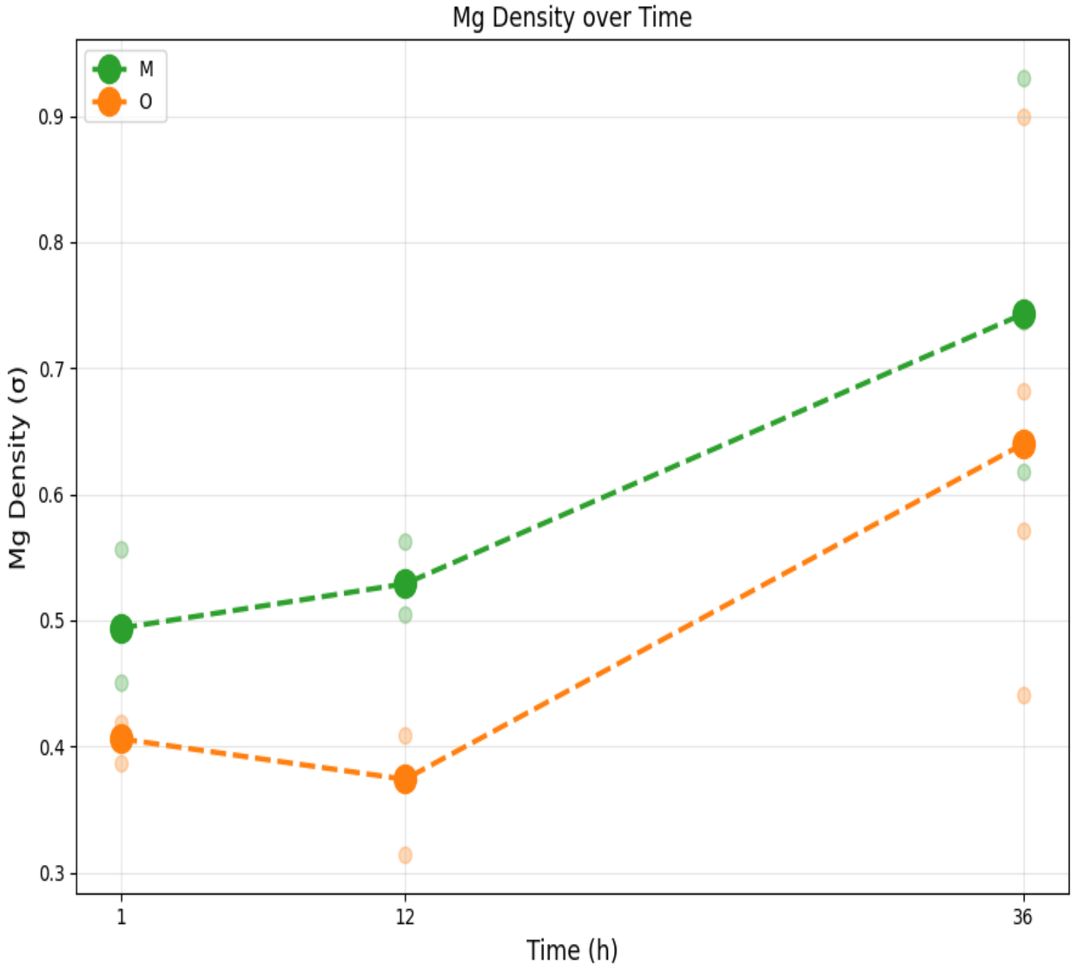
03

The ene-diolate stays hidden

RBD never exceeds ~40% occupancy — the pivotal catalytic state is the least certain of all.

Mg^{2+} leaves — and comes back





1. Starting Model (RBP, RBD, 6Cl, 3PG)

2. Initialization

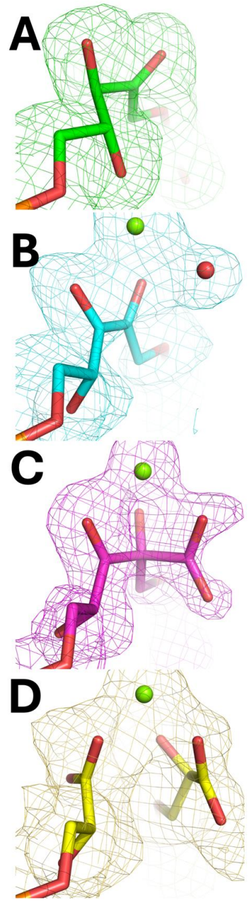
- Set occupancies to zero
- Set B-factors to active site average
- Generate omit maps for all time points (SX_{early}, CX_{1h}, CX_{12h}, CX_{36h}, SX_{late})

3. Partial Omit Map Recycling

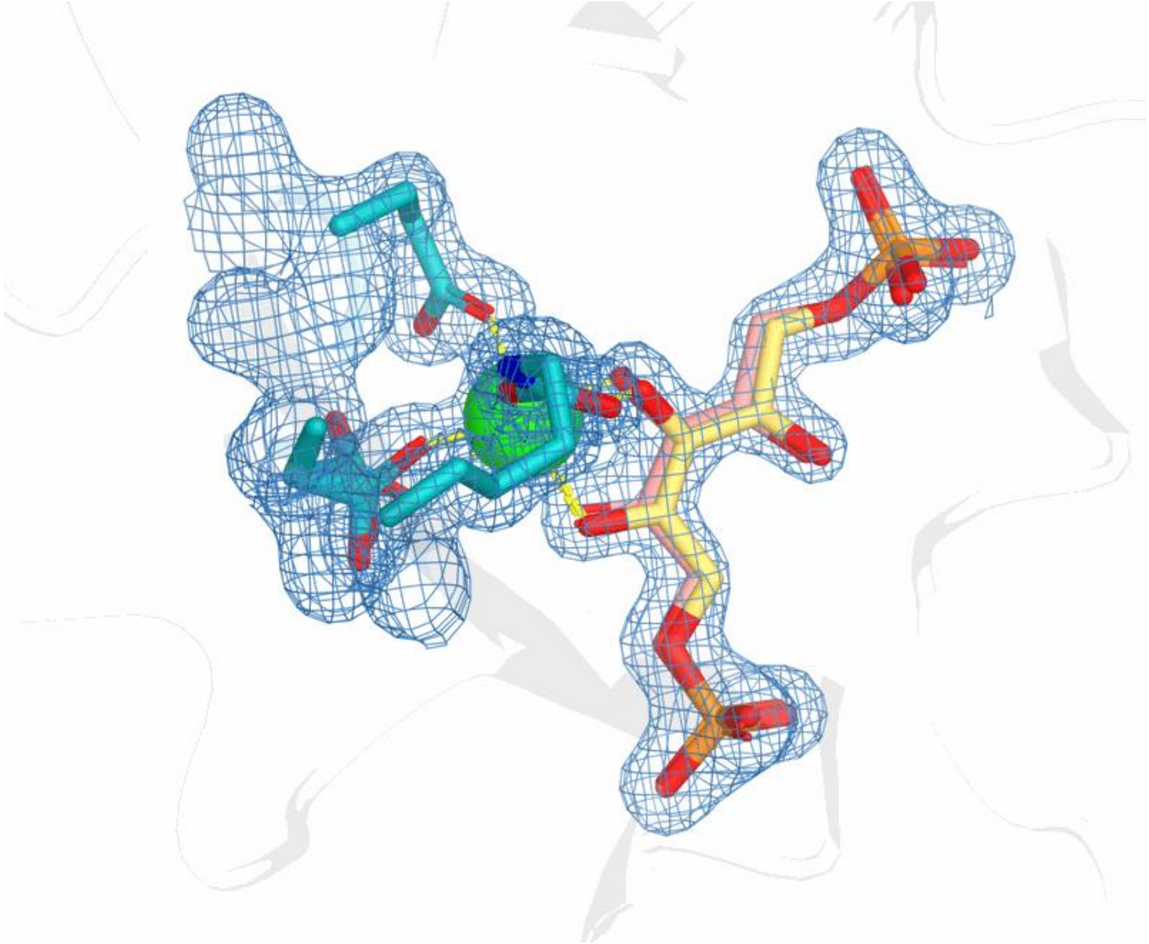
- **Model RBP (mainly present in SX_{early} and CX_{1h})**
- Adjust occupancy of RBD to remove its signal
- Generate partial omit map for RBP
- Fit RBP and update the model
- **Model RBD (mainly present in SX_{early} and CX_{1h})**
- Subtract RBP from earlier time points
- Generate residual omit map for RBD
- Fit RBD and update the model
- **Model 6Cl (mainly present in CX_{12h} and CX_{36h})**
- Subtract RBP & RBD optimized on earlier time points
- Generate residual omit map for 6Cl
- Fit 6Cl and update the model
- **Model 3PG (mainly present in CX_{36h} and SX_{late})**
- Generate RBP, RBD, and 6Cl optimized on earlier time points
- Compute residual omit map for 3PG
- Fit 3PG and update the model

4. Optimize occupancies across time points (minimizing F_o-F_c difference map signal)

5. Use final models for refinement



A trapped ene-diolate — and a coordination that did not fit



WHAT THE T-JUMP GAVE US

The difference maps were **flat** . The temperature jump did not drive the reaction.

WHAT THE DENSITY GAVE US INSTEAD

Two sp^2-hybridised carbons — the ene-diolate, the species we had already untangled from the cryo series. We were pleased with ourselves.

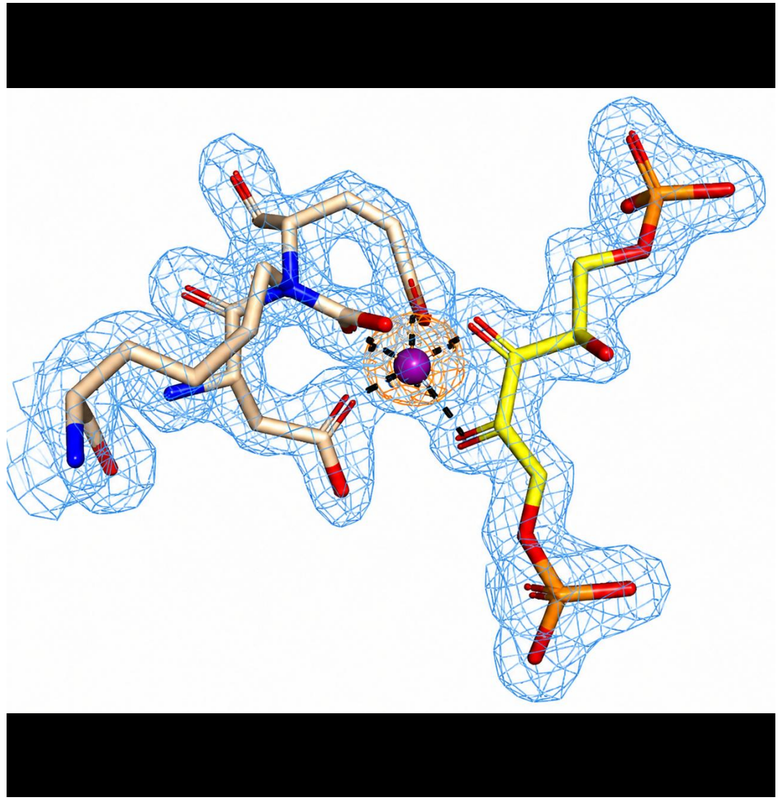
THEN THE METAL

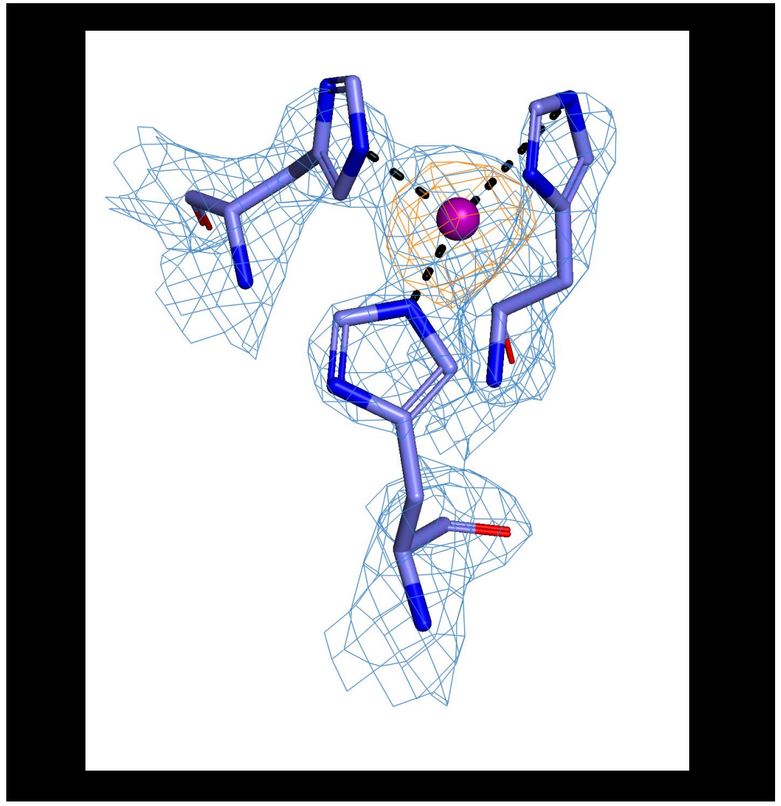
Here	Cryo model
5-fold	**6-fold**

The coordination did not match. So we went looking for alternative theories.

The alternative theory: Zn^{2+} — contaminant, or contributor?

We were above the zinc edge — and found a very strong anomalous peak **at the metal site** , which accounts for the five-fold coordination.





Two sites, not one. The strong peak sits at the metal site. In the subunits *not* occupied by substrate, a second peak appears at a nearby histidine cluster.

ESRF, without substrate, found that same adjacent site — there the fourth ligand is the Lys-carbamylate. We are still making sense of it.

PSI Center for Life Sciences

The technique is not yet routine.

ANALYSIS

PROBIXI

ACQUISITION

HATRX

CAN IT BE DONE AT ALL?

DeconX

REFINEMENT

TorchRef

ANALYSIS

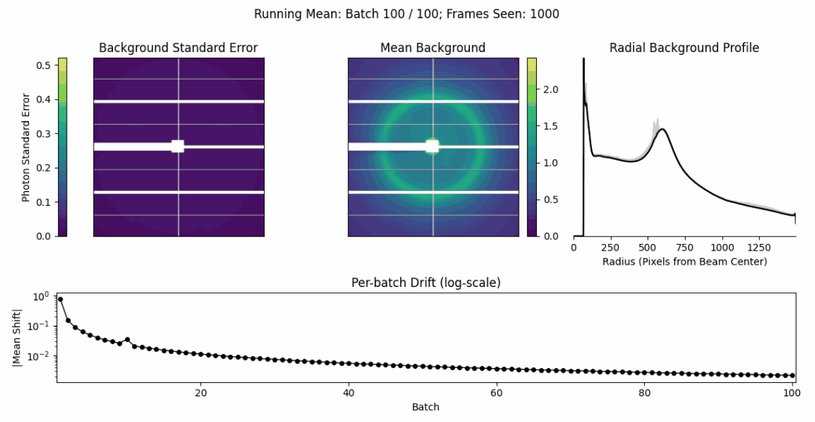
PROBIXI — peakfinding as anomaly detection against a learned noise prior

Learn per pixel what quiet looks like over a few seed frames, whiten the frame into a z-map, then force that map to be normal by searching the blend of pixel, radial and panel statistics that makes it so. Peaks are what is left over.

ONE PIECE OF MATHS

$$\log BF = -\log \kappa + \tfrac{1}{2} z^2 \left(1 - 1/\kappa^2\right)$$

Peak contrast κ, the prior, the mixture weights and the detection threshold are all solved on the seed sequence — none of them is tuned by hand.



C1C2 CHANNELRHODOPSIN · 8.72 MILLION FRAMES · 2.60 Å

METRIC	pf8 + XGANDALF	PROBIXI subset	PROBIXI
Indexed frames	3.8×10^4	3.8×10^4	**2.4×10^5**
Indexing rate (%)	0.44	—	**2.30**
R_{split} (%)	21.83	11.73	**5.08**
$CC_{½}$	0.979	0.980	**0.999**
⟨I/σ(I)⟩	2.04	4.90	**13.54**

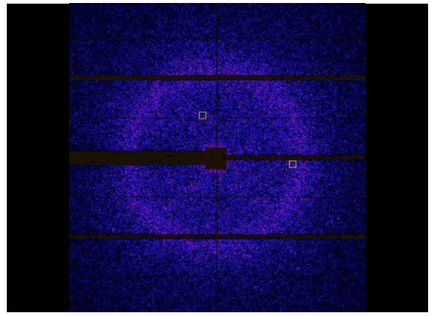
Merged with `partialator` under the xsphere partiality model; 2.60 Å is the nominal $CC_{½} > 0.3$ limit reached by both pipelines. The **subset** arm caps PROBIXI at the same frame count as the tuned pipeline — the gain is not just multiplicity.

Highest shell, R_{split}	Wall clock per crystal	Parameter tuning
58× lower	**0.87 s vs 8.82 s**	**0 h vs 89 h**

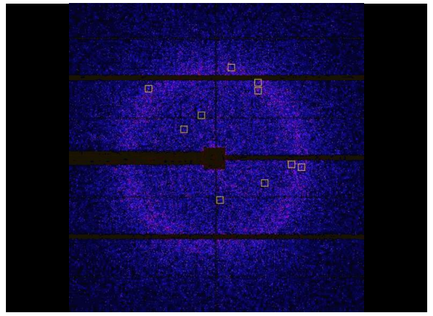
HATRX — encode time points before indexing, decode after merging

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix}$$

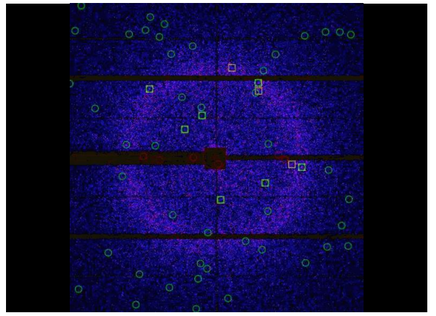
Sum by the pattern **S** before indexing; decode the merged intensities with $\mathbf{S}^{-1}$. Detector gating gives time resolution faster than the frame rate.



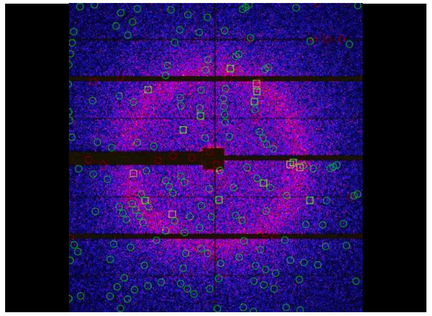
Will not index alone



Will index alone



Indexed conventionally



The HATRX sum indexes — both time points now contribute

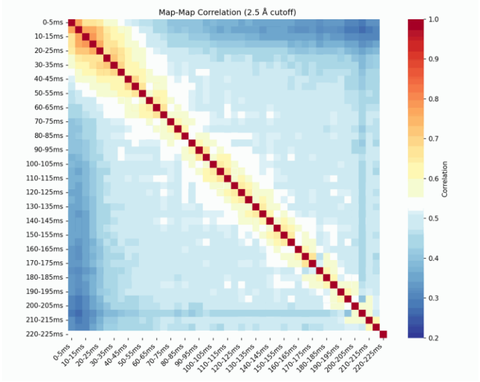


Fig. 18 — bR map–map correlation under HATRX, no scaling.

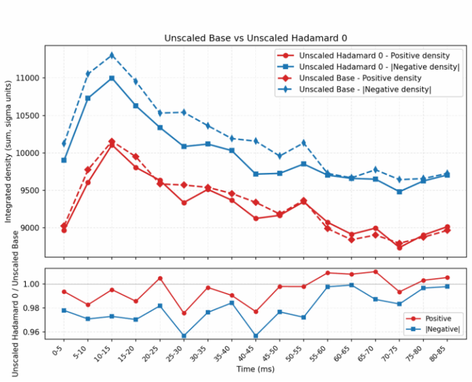


Fig. 19 — Signal along the opening helix, bR, unscaled: HATRX against canonical.

98 %

of the bR signal recovered where signal is already rich — and it begins to recover *more* where signal is poor.

A proof of principle: HATRX **recovers** the signal, it does not improve it. The gain is at low signal-to-noise — it is not a free lunch.

Can You Deconvolve States From Difference Electron Density Maps?

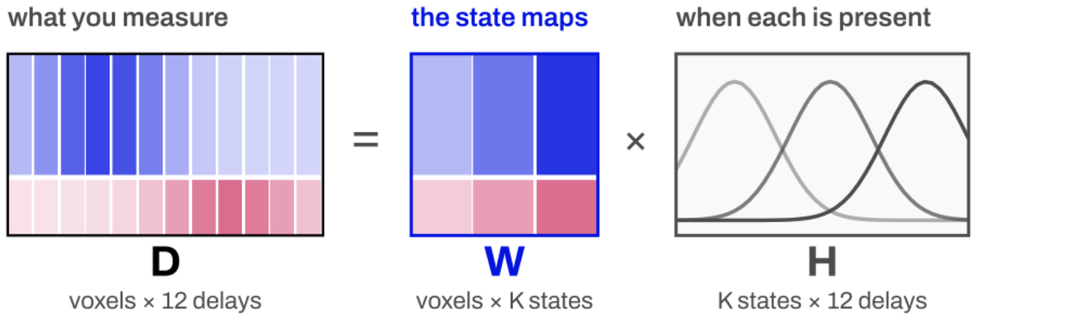
$$D = W\,H + E\,H \geq 0,\ \Sigma_k\, H[k,t] \leq 1$$

Build the mixtures from deposited PDB models so that **W** and **H** are known exactly. Then every unmixing method gets a number instead of an argument.

THE CORPUS — PDB AS OF 27.08.2026, 66 GATES OVER 5 TIERS

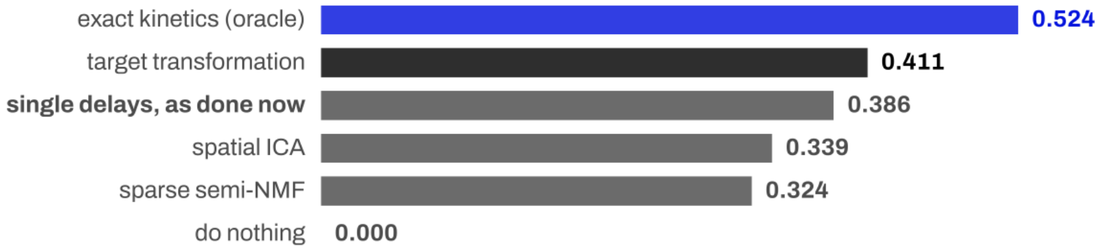
X-ray entries examined	**162 846**
Pass entry quality	**90 583**
Crystal-form families	**6 791**
Families that can pose 2+ states	**2 301**
Experiments passing every gate	**37 425**

185 distinct proteins · 426 crystal forms · ≈ 2.0×10^6 experiments supported. Bar length is log-scaled.



Deconvolution: recover W and H from D alone. Both are known here.

No blind method wins



PSI Center for Life Sciences

TorchRef

An open-source PyTorch framework for crystallographic refinement

---

H. P. Seidel · J. Standfuss · T. Weinert

https://doi.org/10.64898/2026.05.13.724821 (Just accepted yesterday)

As accurate as Phenix — orders of magnitude faster

< 1 %

median R-free gap to Phenix across the benchmark

0.3×

median wall-clock of an identical Phenix refinement

> 100×

faster structure factors on a GPU than CCTBX

1 000

PDB structures refined end-to-end, 1.0–3.0 Å

BENCHMARKED LIKE FOR LIKE

Perturbed starting models, matched protocols, then re-scored independently in REFMAC5. Geometry and ADP quality sit on par with Phenix.

Structure-factor calculation already beats CCTBX past two CPU threads, and runs more than twice as fast at four cores before the GPU pulls two orders of magnitude ahead.

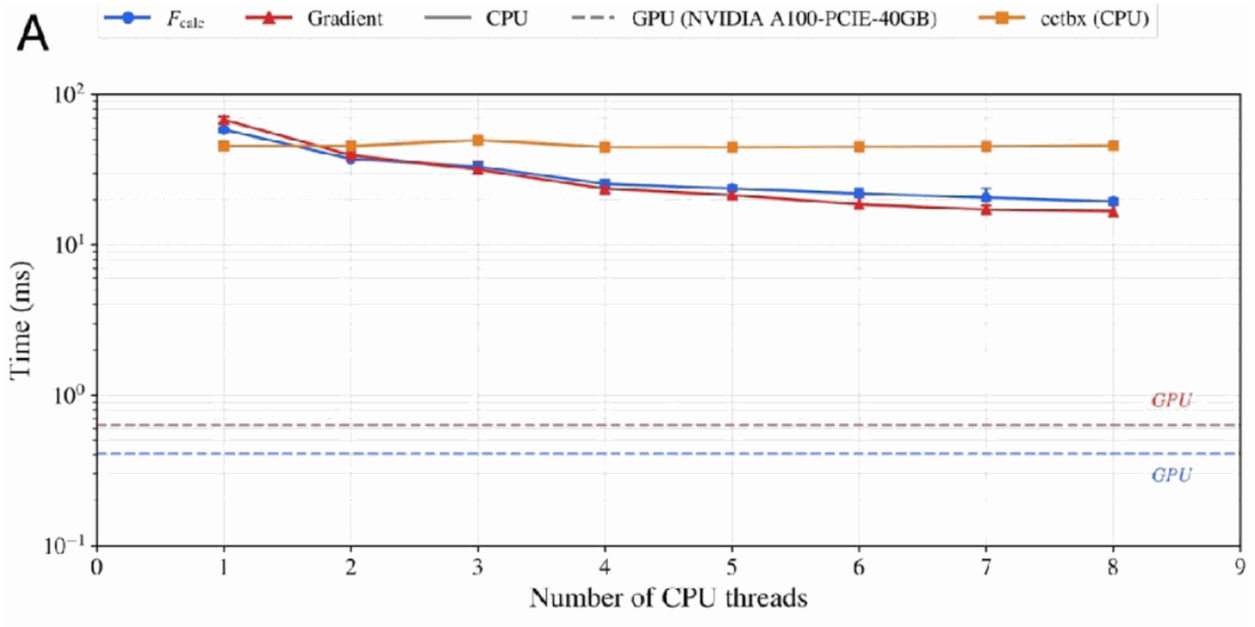


Fig. 20 — Wall-clock per F_{calc} evaluation against CPU threads; dashed lines mark GPU timings (NVIDIA A100).

Refining directly against the difference

End-to-end differentiability lets us fit the change itself — a combined dark + light model refined straight against the observed amplitude differences, with no extrapolation.

THE USUAL ROUTE — EXTRAPOLATED STRUCTURE FACTORS

Amplify a partially occupied state to mimic full occupancy. Noise grows as occupancy drops, dark-state phases leak in, and negative amplitudes need patching — or Bayesian shrinkage that biases back toward the dark state.

TORCHREF — DIRECT DIFFERENCE REFINEMENT

Fit **Δ**(dark → light) against the observed **Δ**F directly. Systematic errors common to both datasets cancel — and the whole target was about 200 lines of Python.

0.847

vs 0.735 extrapolated

Local correlation to the difference density around the ligand.

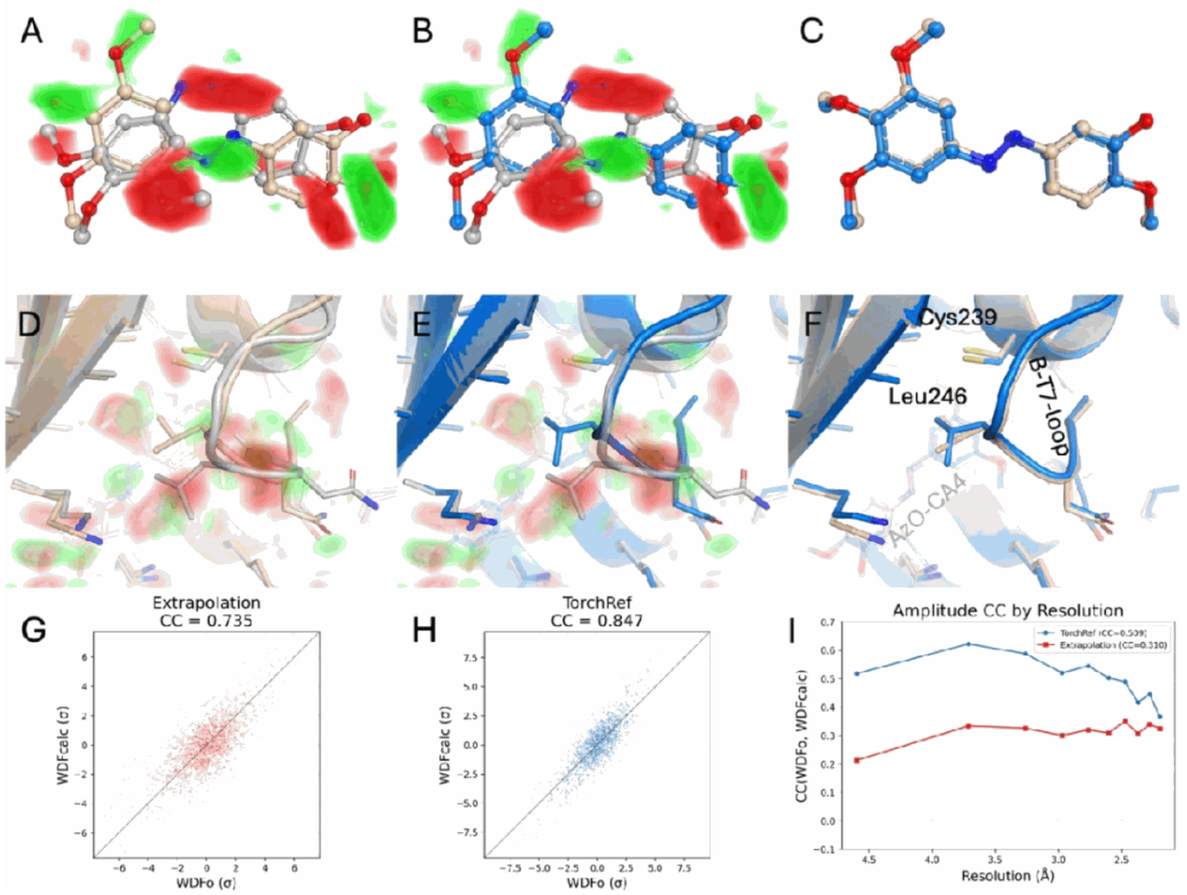


Fig. 21 — azo-CA4 (*cis*→*trans*) in αβ-tubulin, 7YYZ at 10 µs. The same ligand-unbinding system from earlier in this talk.

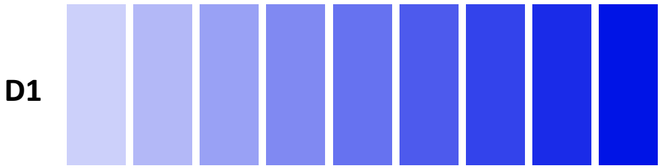
From difference to kinetic refinement — what RuBisCO needs

Two states generalise to many. The same multi-model, multi-dataset machinery refines a handful of conformational states simultaneously against an entire time series — letting their populations vary.

A FEW STATES	
Empty	apo enzyme
RBP	substrate bound
RBD	ene-diolate
6CI	carboxylated
3PG	product

×

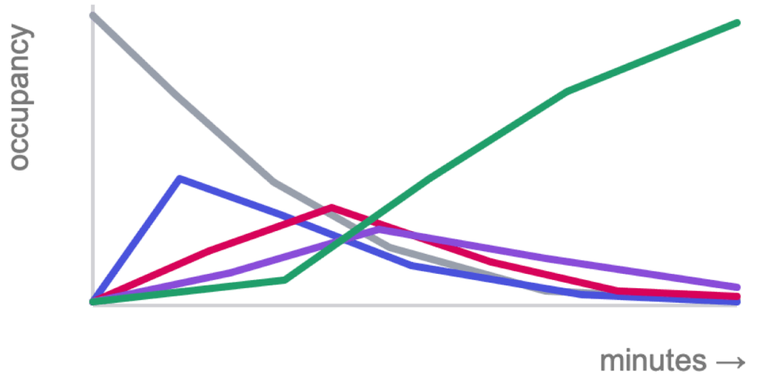
MANY DATASETS



reaction progress · minutes

→

RESOLVED OCCUPANCY



This is the fix for convolution — pulling apart the overlapping intermediates of the RuBisCO cycle, where Mg^{2+} leaves the active site and returns.

Work in progress. The difference refinement above is the simplest case of this: one light-activated state.

Acknowledgements

PSI Center for Photon Science

MX Group

John Beale · C. Y. Huang · Filip Leonarski · May Sharpe · Meitian Wang

ALVRA Group

Camilla Bacellar · Emma Beale · Florian Dworkowski · Chris Milne (now EuXFEL)

Pump Laser Group

Philip Johnson

PSI Center for Life Sciences

TR-SX Group

J. Standfuss · F. Stierli · H. P. Seidel · M. Kepa · R. O'Dea · A. Krishna Thodiyil · (P. Skopintsev) · (G. Khusainov) · (M. Wranik) · (Q. Bertrand) · (K. Starlinger)

Most of the RuBisCO work shown today is Georgii Khusainov's.

Steinmetz Group

Natacha Gaillard

Collaborators

Wachtveitl Group

Universität Frankfurt

Shapiro Group

TU Dortmund

MFX Team · LCLS

Special thanks to Sandra Mous

No X-rays — no structures.