7. Science Opportunities in the Life Sciences
(see also: sections 3.4, 6.1, 8.3, 8.6, 9.3, 9.4, & appendix 2)

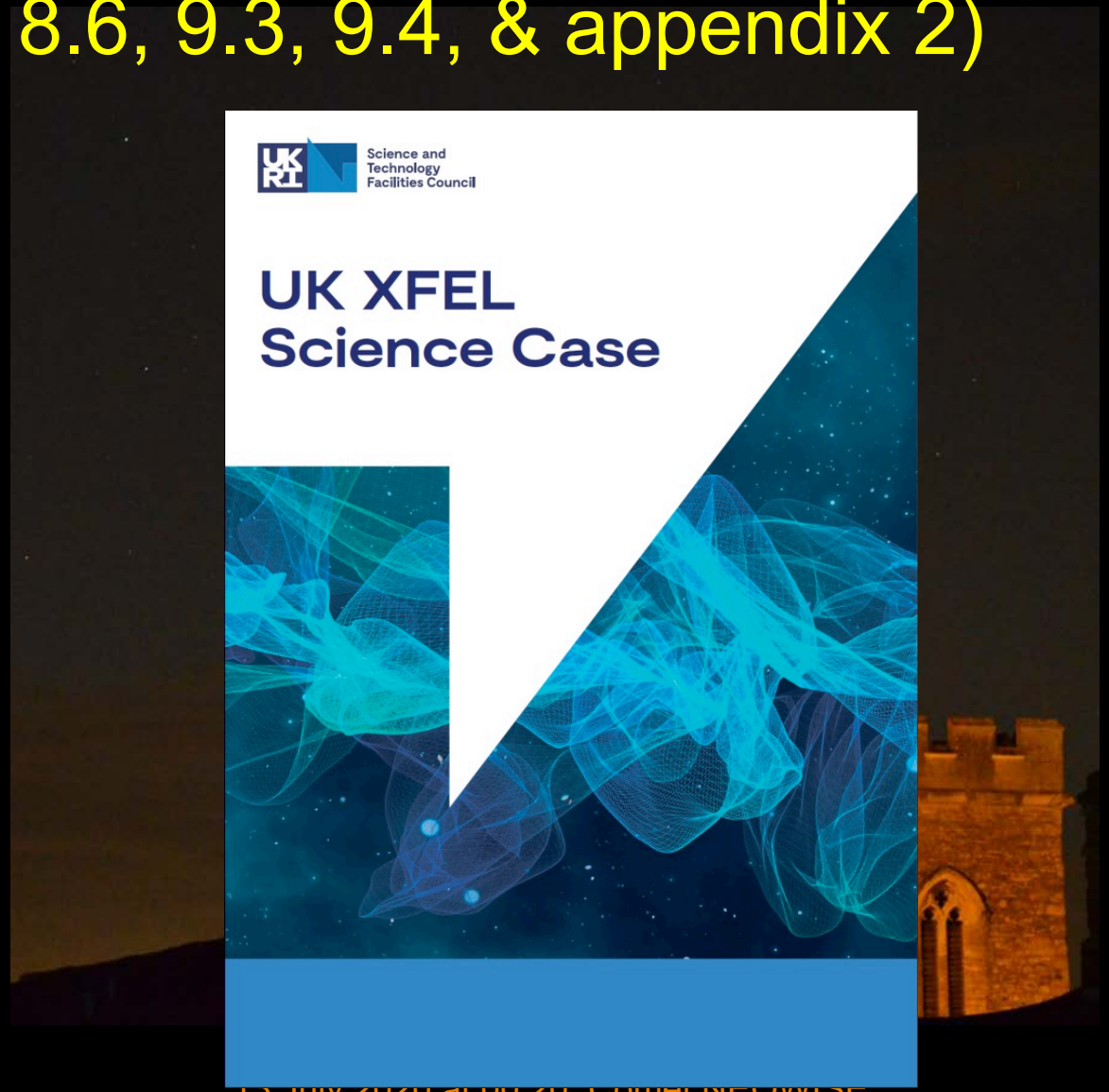
Allen M. Orville, Ph.D.
XFEL Hub at Diamond Light Source
allen.orville@diamond.ac.uk

Jasper van Thor, Ph.D.
Imperial College London
j.vanthor@imperial.ac.uk

Xiaodong Zhang, Ph.D.
Imperial College London
xiaodong.zhang@imperial.ac.uk

et al
thank you to many!

Project Lead: Jon Marangos (ICL)	STFC Project Champion: John Collier (CLF)



13 July 2020 at 00:20, Comet NEOWISE
St Nicolas' Church (1170), Abingdon-on-Thames

Most MX diffraction data is measured at synchrotrons
from a sample held at 100 K …

… a bit warmer than the
surface of ***Pluto***

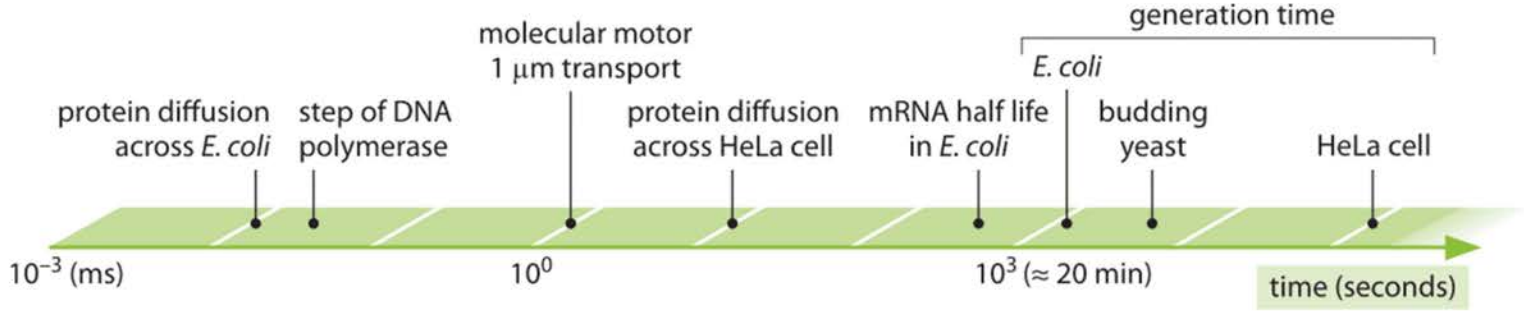
(33 – 55 K; average 44 K) …

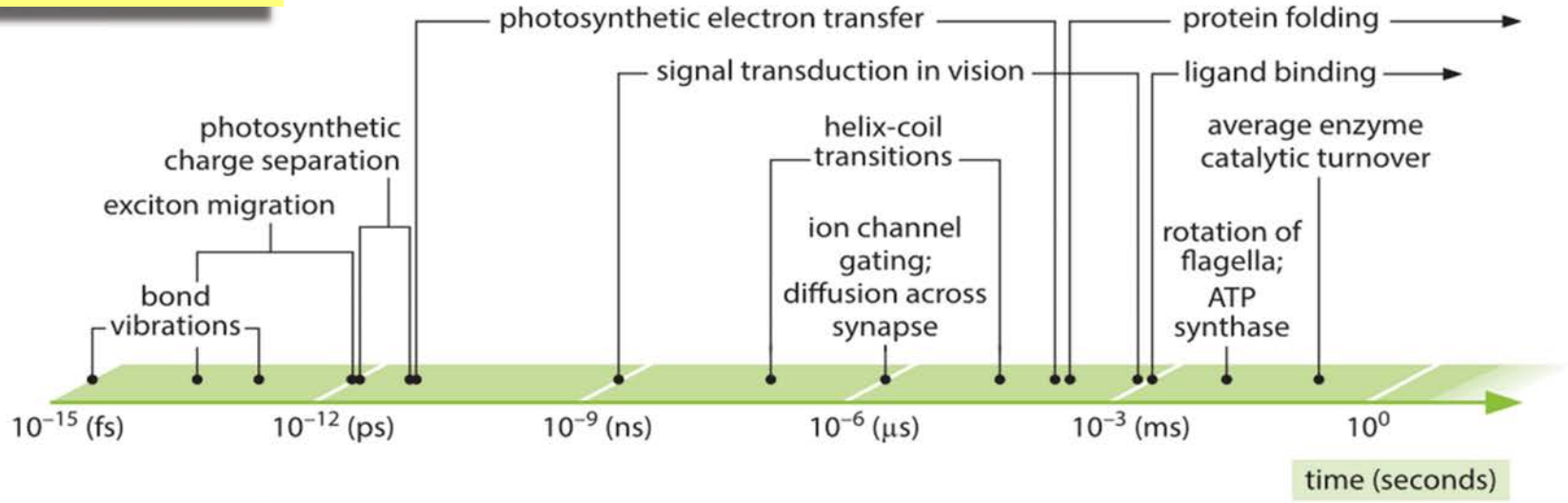
… and so a ***remaining frontier challenge***
in structural biology is to determine
time-resolved (crystal) structures
directly from systems engaged in catalysis (function),
at physiological temperature and pressure
…… ***on earth***

24 Dec 1968
NASA / Apollo 8
AS08-14-2383

Figure 7.2: A comparison of X-ray sources and types of phenomena they are used to study.

Biochemical processes span orders of magnitude in time.





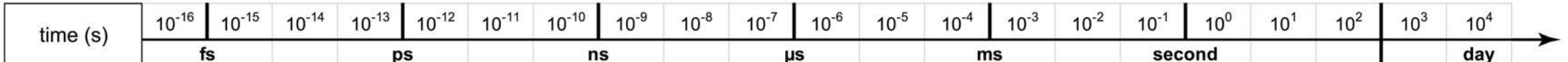
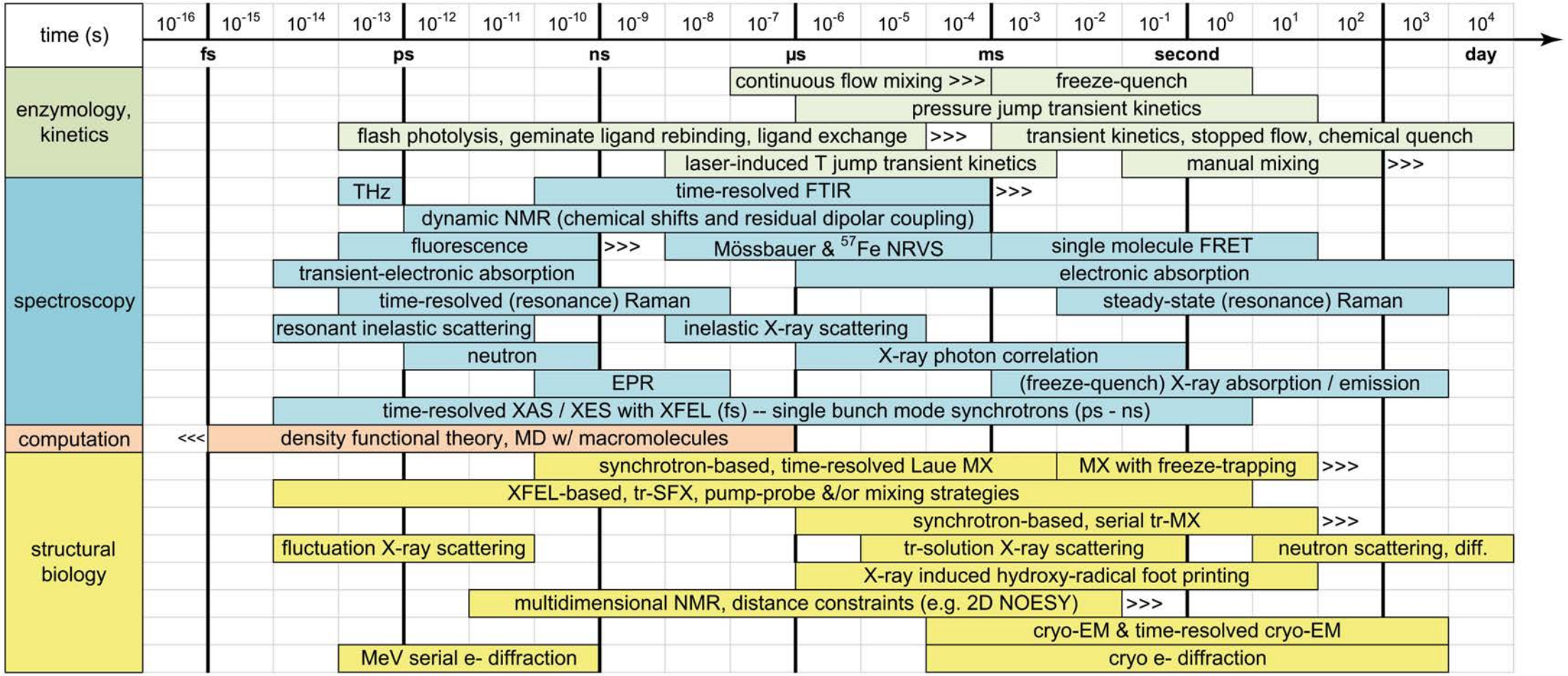


Figure 7.2: A comparison of X-ray sources and types of phenomena they are used to study.

Complementary methods address particular features & time domains.



PDB History

Structural Biology Highlights

2024
2023
2022
1990
1989 — First **NMR structure** released in PDB: protein BDS-I (*Driscoll et al.*, 1989)
8
7
6
5
4
3
2
1 — First **B-DNA** structure determined
0
9
8
7
6
5
4
3 — (Kim *et al.*, 1973; Robertus *et al.*, 1974)
2
1971 — PDB established
1970

Section 8.6 Industrial inspirations from deeper insights into biology: Pharma to clean energy

>90% of all New Medical Entities (NMEs or drugs) approved by FDA since 2010 used PDB models

- 93% (184/210) have relevant structures in PDB
- 6% (13/210) have no known molecular target
- 5,914 unique PDB structures linked to NMEs
- ~ all targets released by PDB
- Median time between PDB deposition and FDA approval > 10 years
- ~ $600 million invested (~ $100,000 / structure)
- > $100 billion public NIH funding (estimated 20% of NIH budget) + private-sector = > $700 billion

Westbrook et al (2020) *Drug Discov Today 25*, 837-850

Westbrook and Burley (2019) *Structure 27*, 211-217

PDB Replacement Cost > US $15.6 billion

Conservative cost
(assuming US$100,000 / atomic model)

Atomic Models Released by the PDB

	15 July 2019	21 July 2020	Δ	% Δ
X-ray cryst.	**137,265**	**147,909**	**10,644**	**7.75**
NMR spec.	**12,679**	**13,042**	**363**	**2.68**
Cryo-EM	**3,465**	**5,368**	**575**	**54.92**

wwPDB consortium (2019) *Nucleic Acids Res* 47: D520-D528

Cryo-EM growth *-vs-* X-ray crystallography *-vs-* diversity of life on earth (yes please, all of the above!)

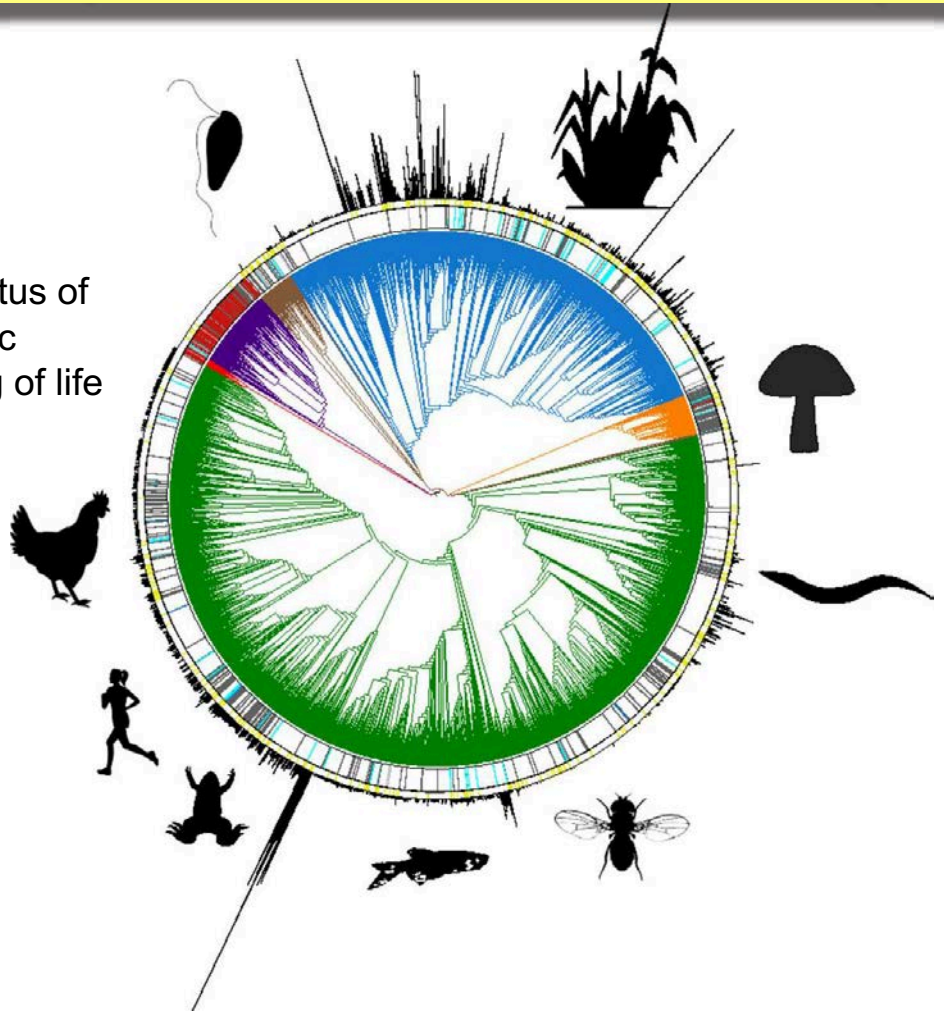


Figure 7.1: Current status of the genomic sequencing of life

Figure 9.3: Growth of the PDB archive from Xray and cryo-EM methods

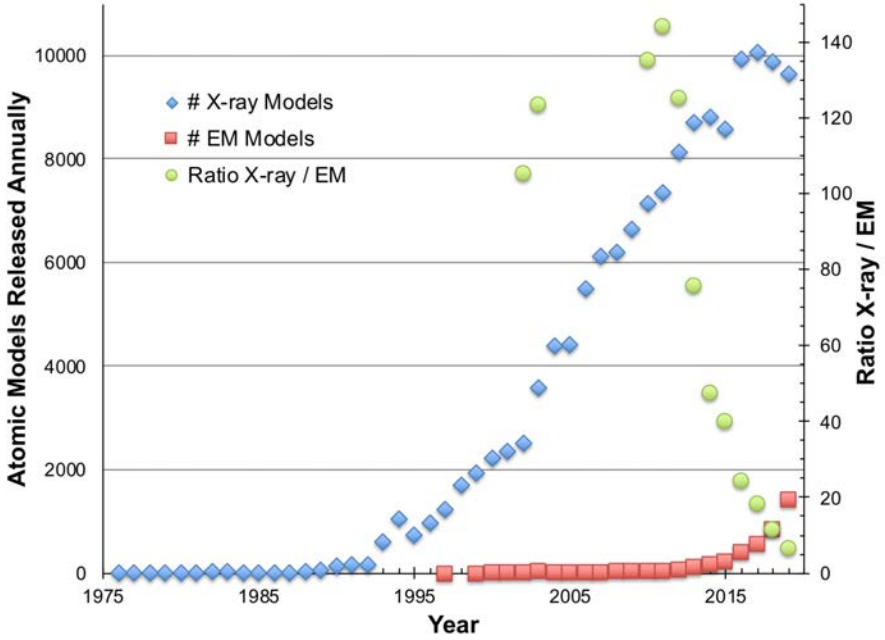


Table 7.1: KEGG GENES Annotation Statistics (as of 17 May 2020) [a]

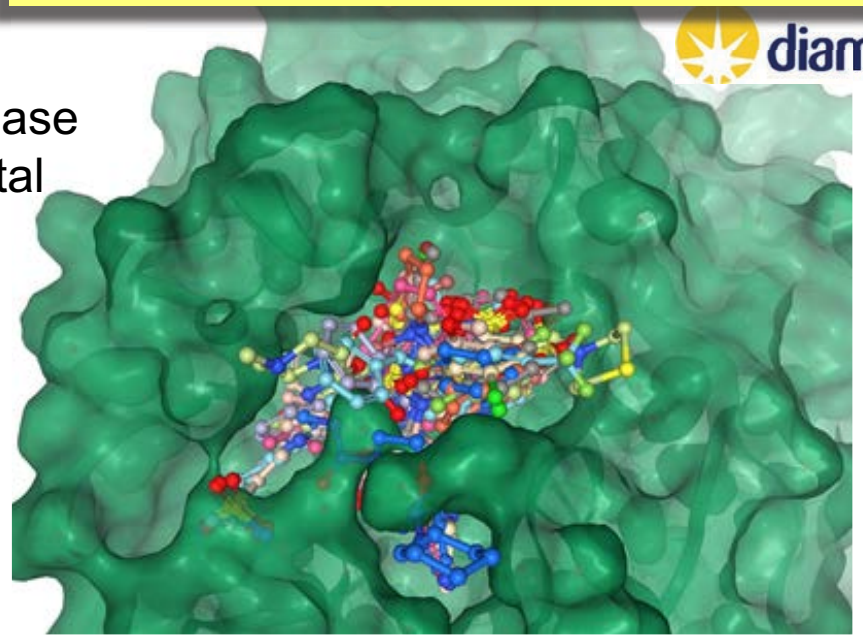
Category	Protein-based genes		RNA-based genes		Pathway linked genes	Enzyme genes with EC numbers
	All genes	KO assigned genes [b]	All genes	KO assigned genes [b]		
KEGG organisms	30,224,913	15,635,407	628,182	347,049	8,700,385	6,750,695
Brassica napus	96,972	31,875	65	64	17,171	13,133
Homo sapiens	19,855	14,264	2,641	336	8,039	3,363
Saccharomyces cerevisiae	6,002	3,793	415	394	2,399	1,311
Escherichia coli	4,240	3,169	179	152	1,697	1,302
Viruses	367,122	10,122	5,500	24	N/A	5,235
Addendum	3,973	3,881	N/A	N/A	N/A	3,064

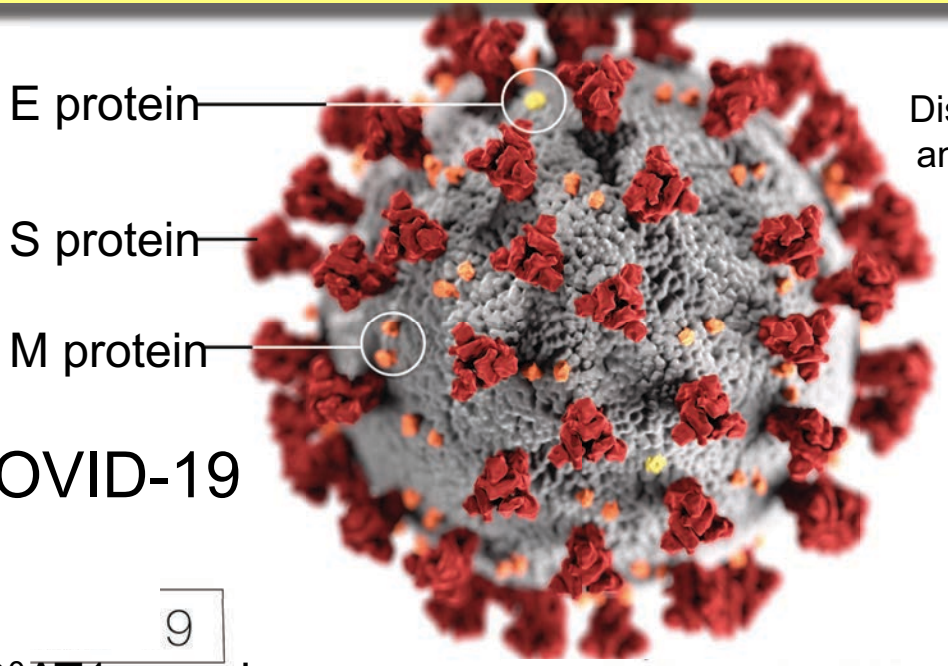
[a] Source: www.genome.jp/kegg/docs/genes_statistics.html
[b] Annotated with the KEGG Orthology (KO) system; the basis for cross-species annotation in KEGG. The set of genes in the genome that can be mapped to KEGG reference pathways and BRITE reference hierarchies to generate organism-specific pathways and hierarchies.

Figure 8.11: Some morphology of the COVID-19 virus

Covid-19 main protease (M^{pro}) crystal structures

XChem fragment screening results



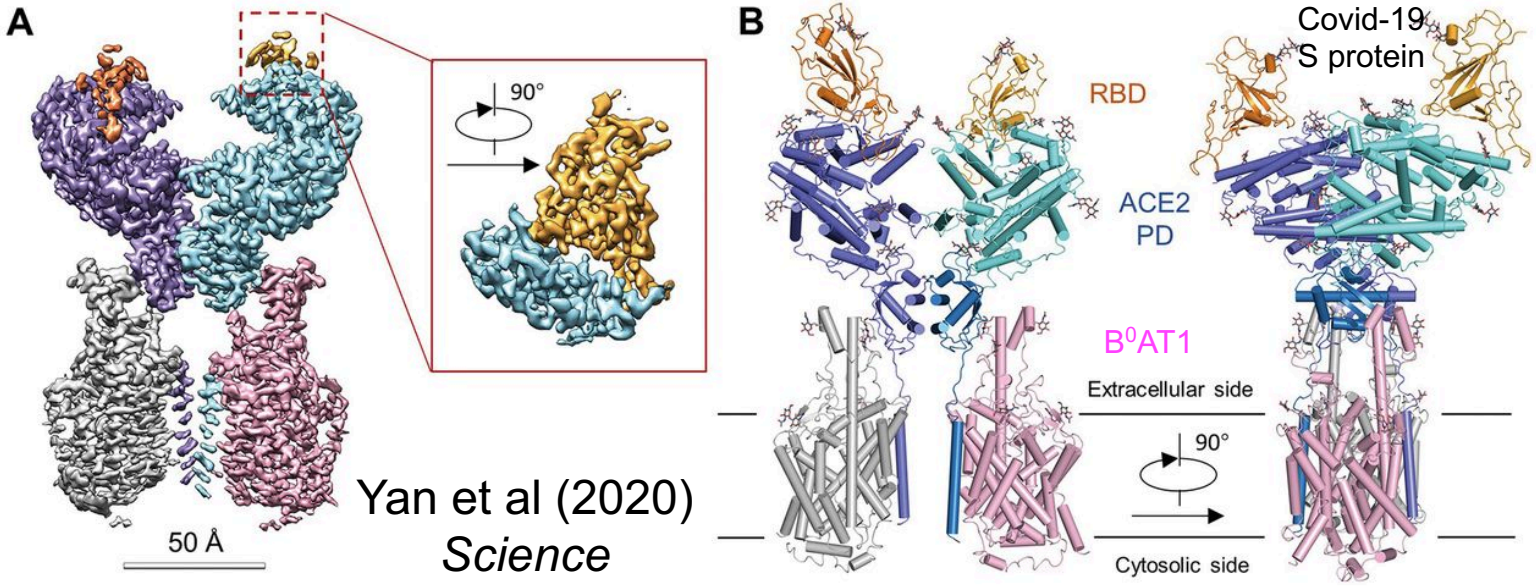


COVID-19

Centers for Disease Control and Prevention (CDC)

9

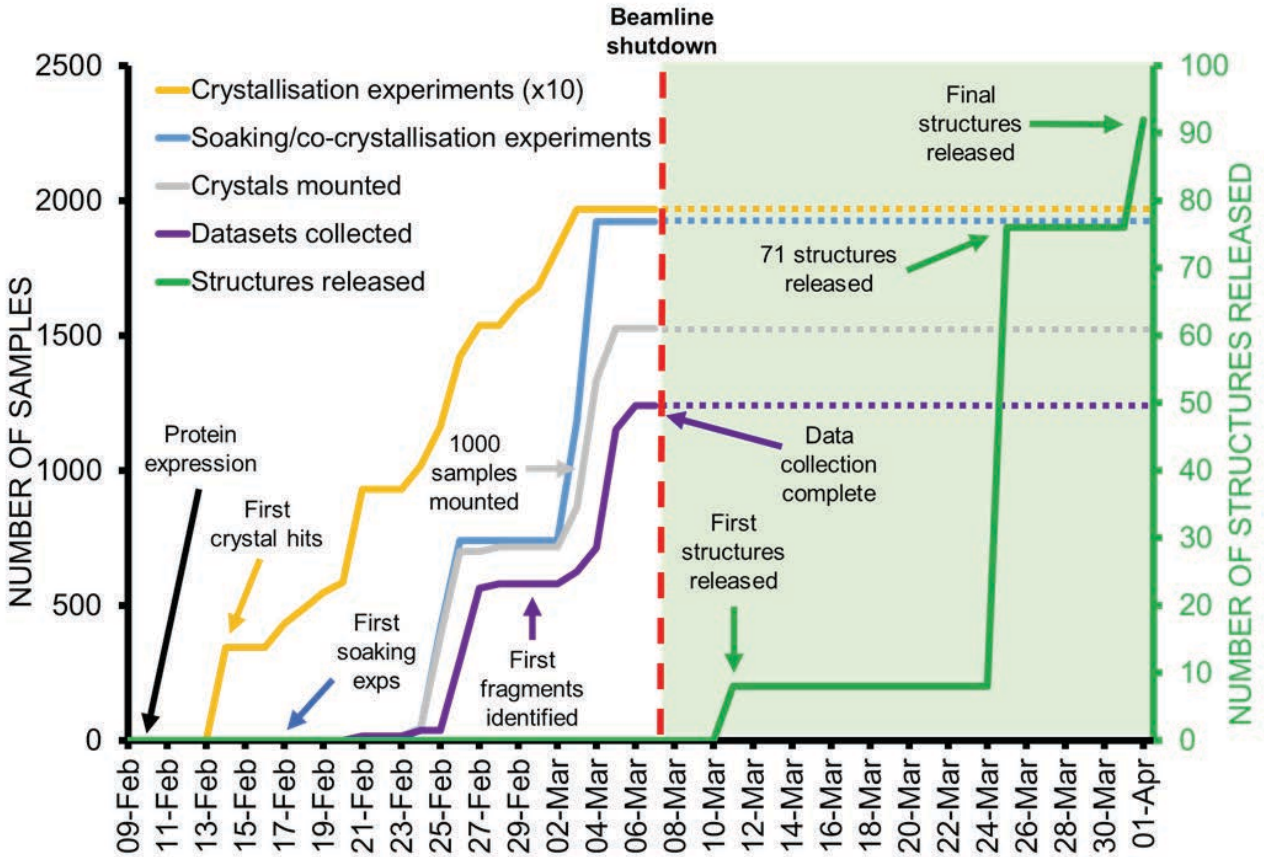
Cryo-EM structure of the Covid-19 S protein : human ACE2 : B^0AT1 complex

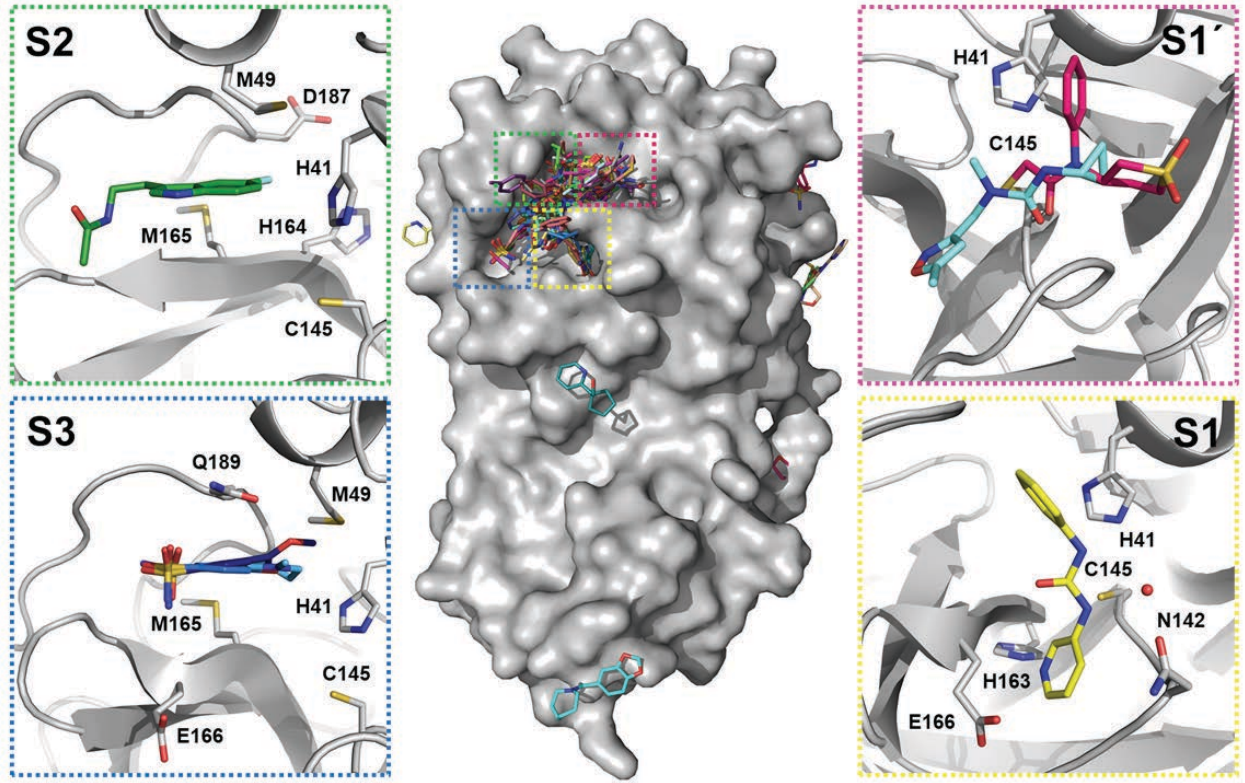


Yan et al (2020) *Science*

Diamond / XChem and the COVID-19 Main Protease (M^{pro})

- Viral RNA encodes two open reading frames, generates two polyproteins pp1a and pp1ab
- These polyproteins produce most of the proteins of the replicase-transcriptase complex
- Processed by two viral proteases: Papain-like protease (**PL^{pro}**) and 3C-like protease (main protease (**M^{pro}**); both are primary target for antiviral drug development
- Diamond / XChem were early into this R&D effort with fragment-based library screening





Douangamath *et al* (submitted)

Global structural biology response to Covid-19

Atomic Models in PDB (22 July 2020 release date)	Number	Sample Temperature	Resolution Range (average) Å
Cryo EM Total	47	~ 100 K (all cryo)	2.5 – 3.84 (3.31)
X-ray Crystallography Total	250	295 – 98 (11 room temp)	0.95 – 4.36 (1.89)
Diamond Light Source	**125**	**100 K**	**1.25 – 4.36 (1.78)**
Total Number Released	299	100 K	0.95 – 4.36 (2.13)

LCLS proposals awarded XFEL beamtime in 2020 for Covid-19 R&D via rapid access process		
P173 MFX (H. DeMirici et al)	16 – 19 Aug	Structural dynamics of SARS-CoV-2 3-Chymothrypsin-like Protease and its Inhibitor Complexes
P171 MFX (M. Schmidt et al)	21 – 24 Aug	Room Temperature Structure and Inhibition of the Coronavirus SARS CoV-2 Main Protease
P172 CXI (P. Fromme et al)	28 Aug – 01 Sep	Time-resolved serial femtosecond crystallography studies on the endonuclease NendoU protein of SARS-CoV-2
P175 MFX (A. Orville et al)	18 – 21 Sep	Time-resolved SFX of Covid-19 proteins including M-pro
P178 CXI (B. Hogue et al)	25 – 29 Sep	Coronavirus Viroporin Structural Studies

G-Protein Coupled Receptor (GPCR) are critical to human health

Thousands of Ligands - Chemical Diversity

↓

>800 Different Human Receptors
(largest family in human genome)
Share 7TM Fold But Diverse Structural Features

↓

Dozens of transducers

CNS
(**acetylcholine**, **dopamine**, serotonin, **opioid**, cannabinoid, etc)

Cardiovascular
(**adrenergic**, **adenosine**, purinergic, angiotensin, etc)

Immune
(**histamine**, **cytokine**, **sphingosine**)

Metabolic
(parathyroid, THS, calcitonin, glucagon etc.)

Reproductive
(oxytocin, gonadorelin, etc.)

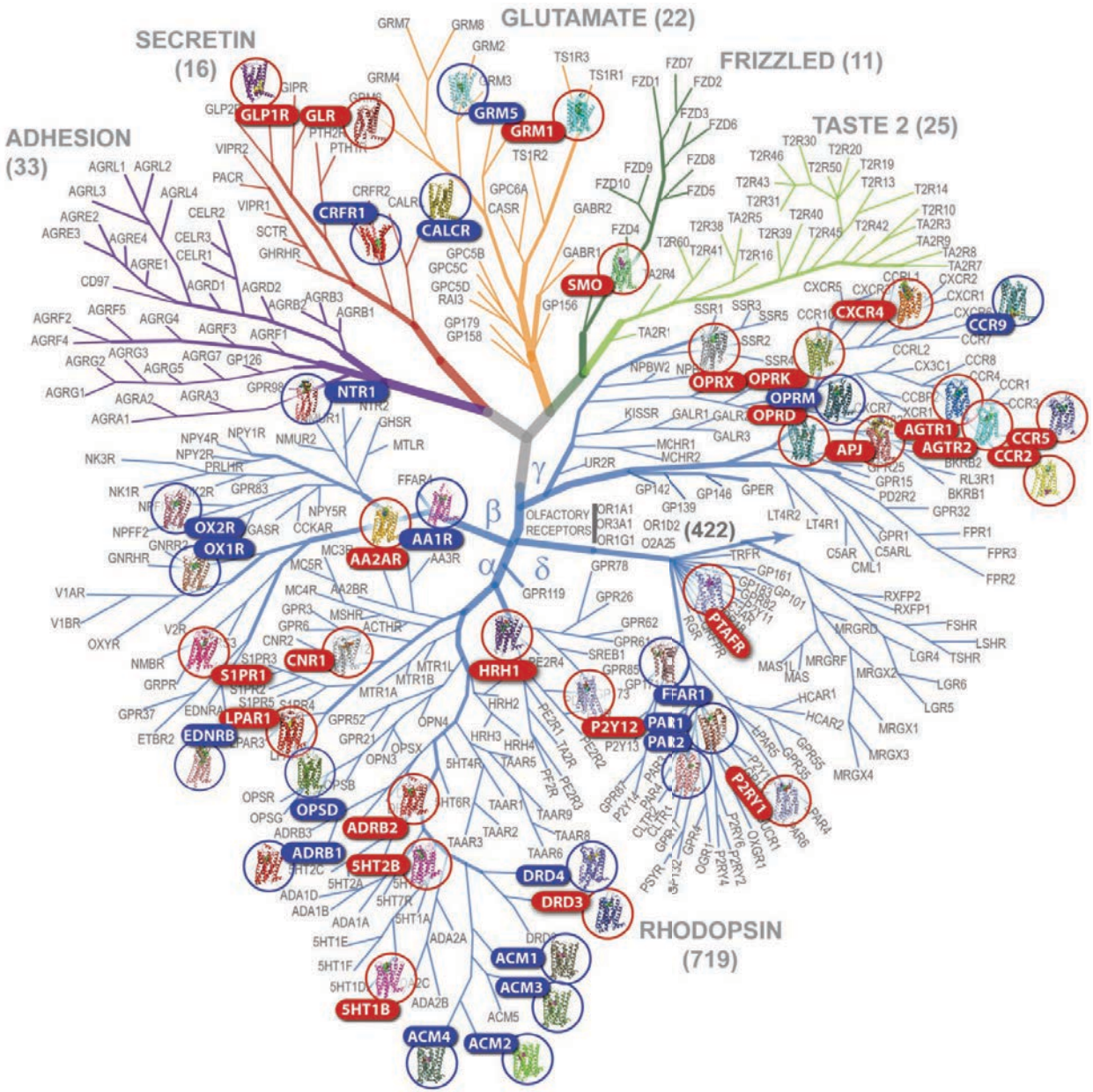
Current GPCR Drugs:

- **~40%** of all prescribed drugs
- Both agonists & antagonists
- GPCRs (2%) make **>25%** of all established clinical targets

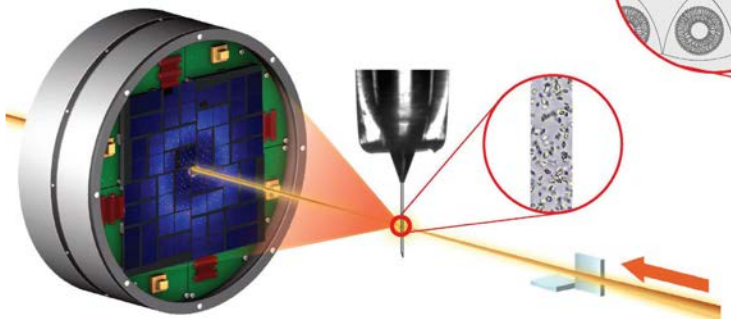
New GPCR Drugs Needed:

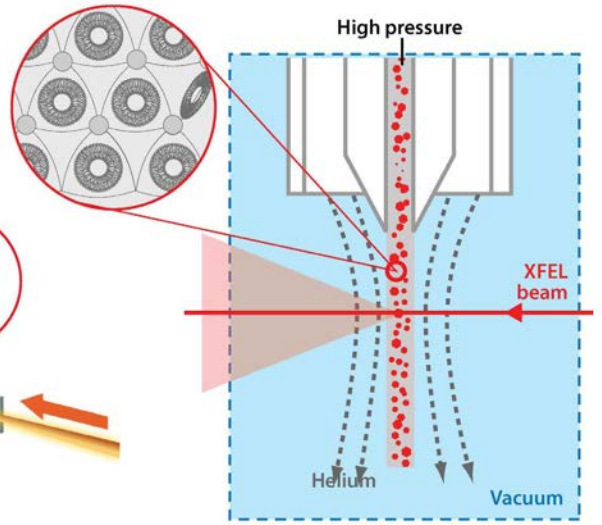
- oral bioavailability
- subtype selectivity
- functional selectivity
- allosteric modulation
- **~100** new preclinical targets (new subtypes, new families, orphans)

Figure 8.12: G-Protein Coupled Receptor (GPCR) crystal structures – most by XFEL methods

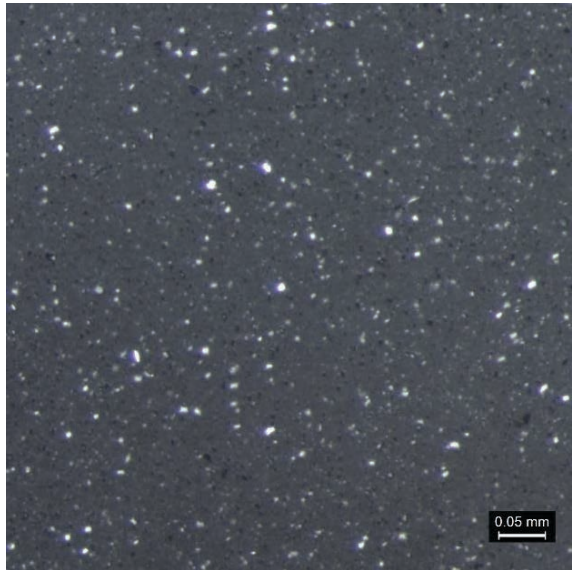


LCP viscous media extruder for SFX data collection





Micro-crystals for XFELs



Macro-crystals for synchrotrons

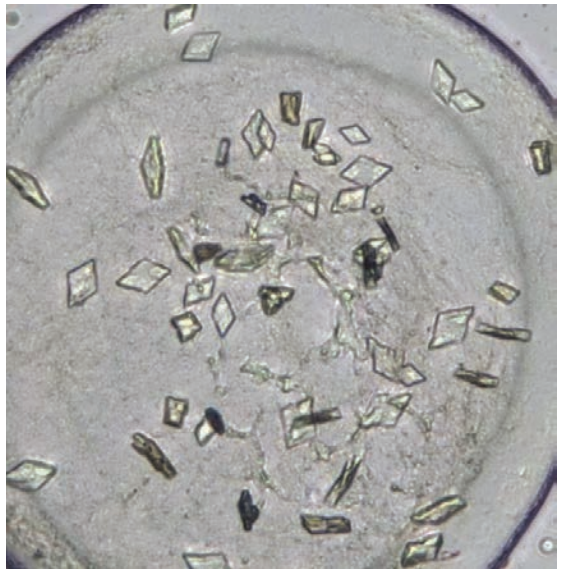


Figure 7.8: Concepts of time-resolved structural biology

$$E + S_1 \rightleftarrows E{\bullet}S_1 + S_2 \rightleftarrows E{\bullet}S_1{\bullet}S_2 \rightleftarrows [\, I_1 \rightarrow I_2 \rightleftarrows I_3 \rightarrow \ldots] \rightarrow E{\bullet}P \rightleftarrows E + P$$

Traditional MX: synchrotrons, macro-crystals, 100 K, resting state **E**, soaked **E•S$_1$** or **E•P**; soaking or crystallization lacks function and dynamics: >90% structures PDB / year

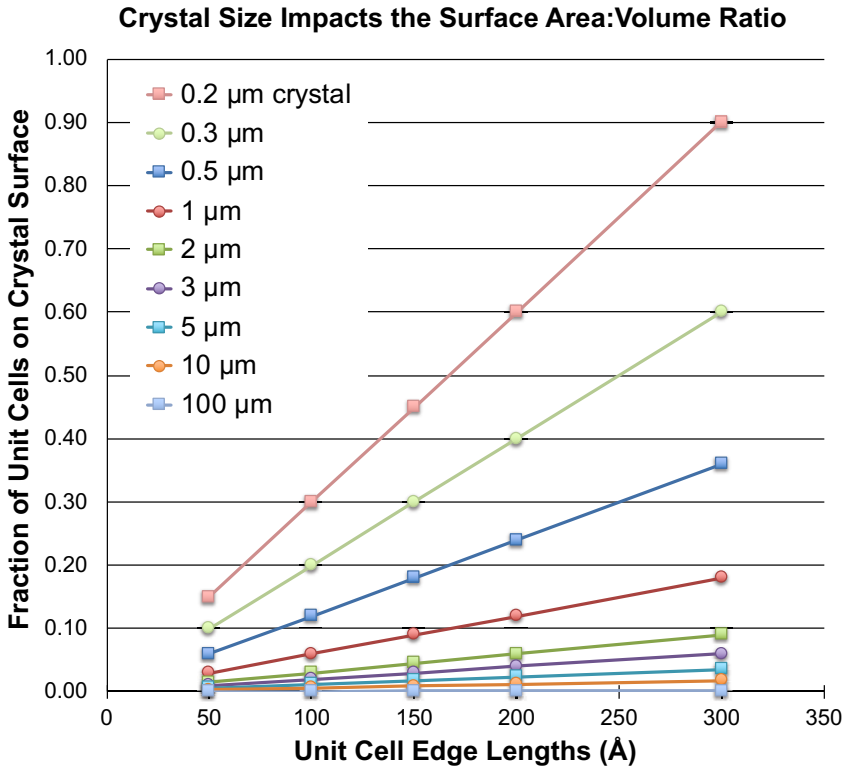
Cryo-EM: in solution, low Temp or freeze-quench ≈ ms time resolution, complements & benefits from MX, class averages, limited dynamics, no spectroscopic confirmation

Serial MX at XFELs & Diamond (SFX & SMX)

- study entire reaction cycles at room temp & pressure
- XFEL fs pulse ≈ bond vibrations, photo-active reactions
- No radiation-induced damage to reactive intermediates
- DLS / VMXi ≈ μs time resolution with mixing strategies
- μ-crystal slurries ≈ atomic & electronic structural data

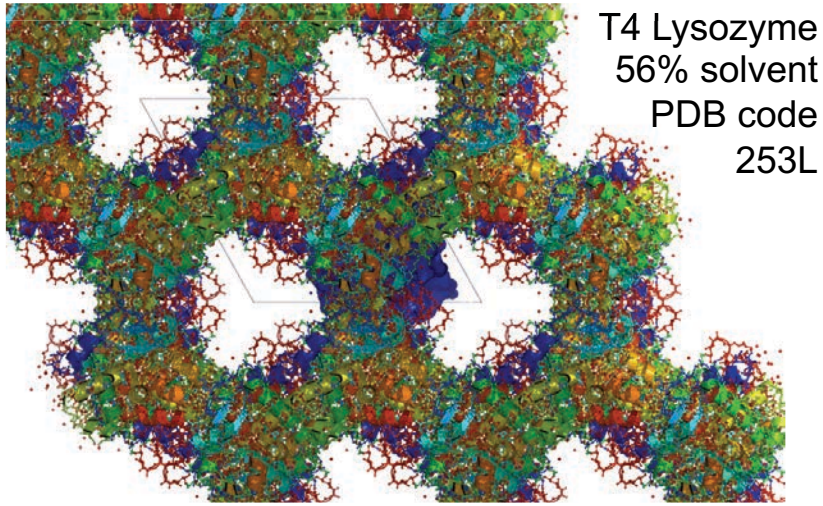
Entering an era of *dynamic* structural biology… a concept, a set of tools, to collect as much data as possible from every sample and X-ray pulse, enables atomic resolution "movies" of macromolecules engaged in catalysis

AIM: Within 5 – 10 years, routine molecular movies via serial MX at all XFELs & synchrotrons



The driving hypothesis for generalized time-resolved serial µMX (Section 7.3 & Figure 8.4)

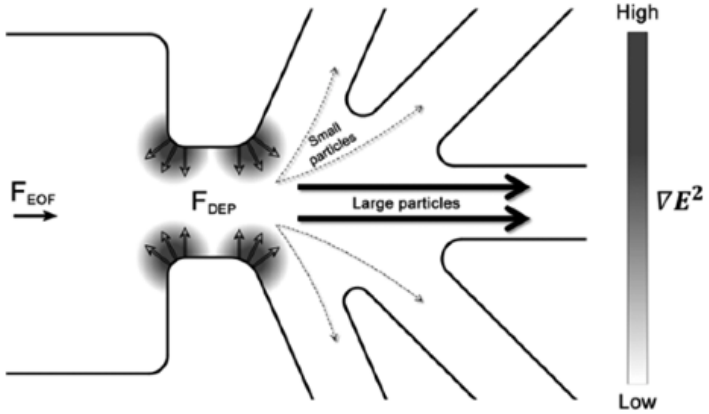
- use **enzyme microcrystals** (~2x2x2 µm^3 and smaller)
- substrate(s) diffusion ≈ µm / µs, **will equilibrate** in ~ µs – ms
- average enzyme reaction in solution is ~ 60 ms
- Thus, **many times faster than typical reaction cycle**



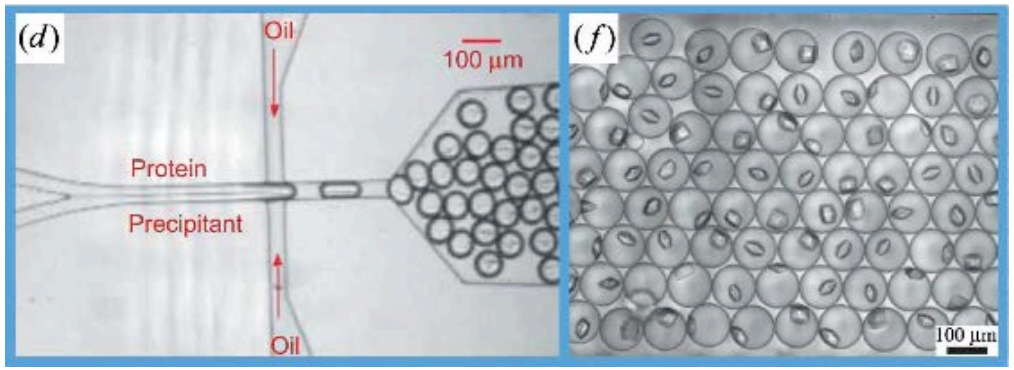
T4 Lysozyme
56% solvent
PDB code
253L

Examples of Producing Homogeneous Slurries via:

Dielectrophoretic Sorting
(Abdallah et al (2013) ACS Nano 7, 9129-9137)



Crystallization in Emulsion Droplets
(Heymann et al (2014) IUCrJ 1, 349-360)

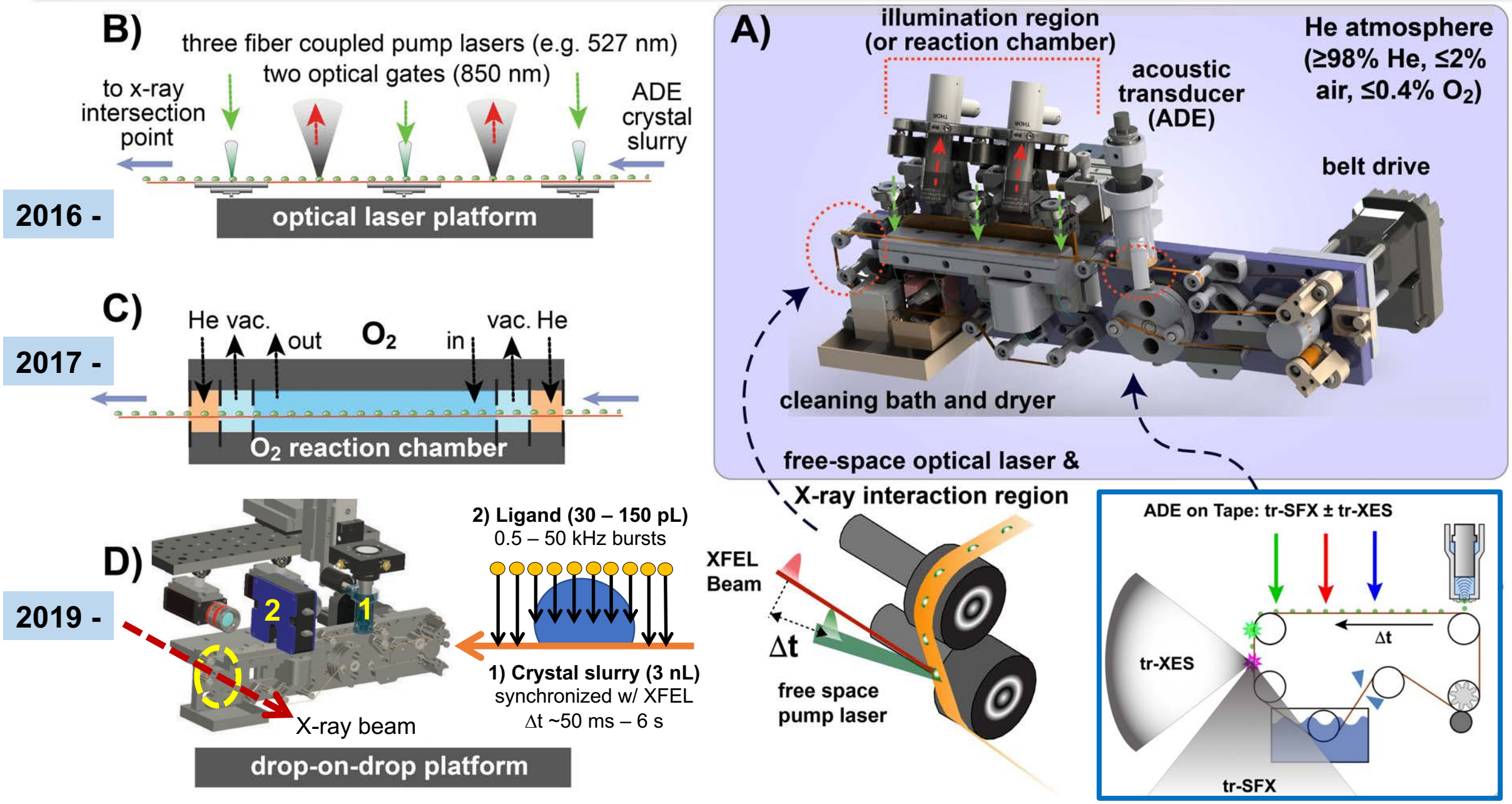


Some key considerations:

- Space group / Crystal packing
- Lattice channels / Access to active site(s)
- Viscosity / Ligand diffusion rate
- pH / Ions / Co-substrate(s)
- Dynamic change(s) -vs- Lattice packing constraints

Schmidt, M. (2017)
Methods Mol Biol 1607, 273-294

Figure 7.4: Acoustic tape drive system for time-resolved SFX and XES experiments



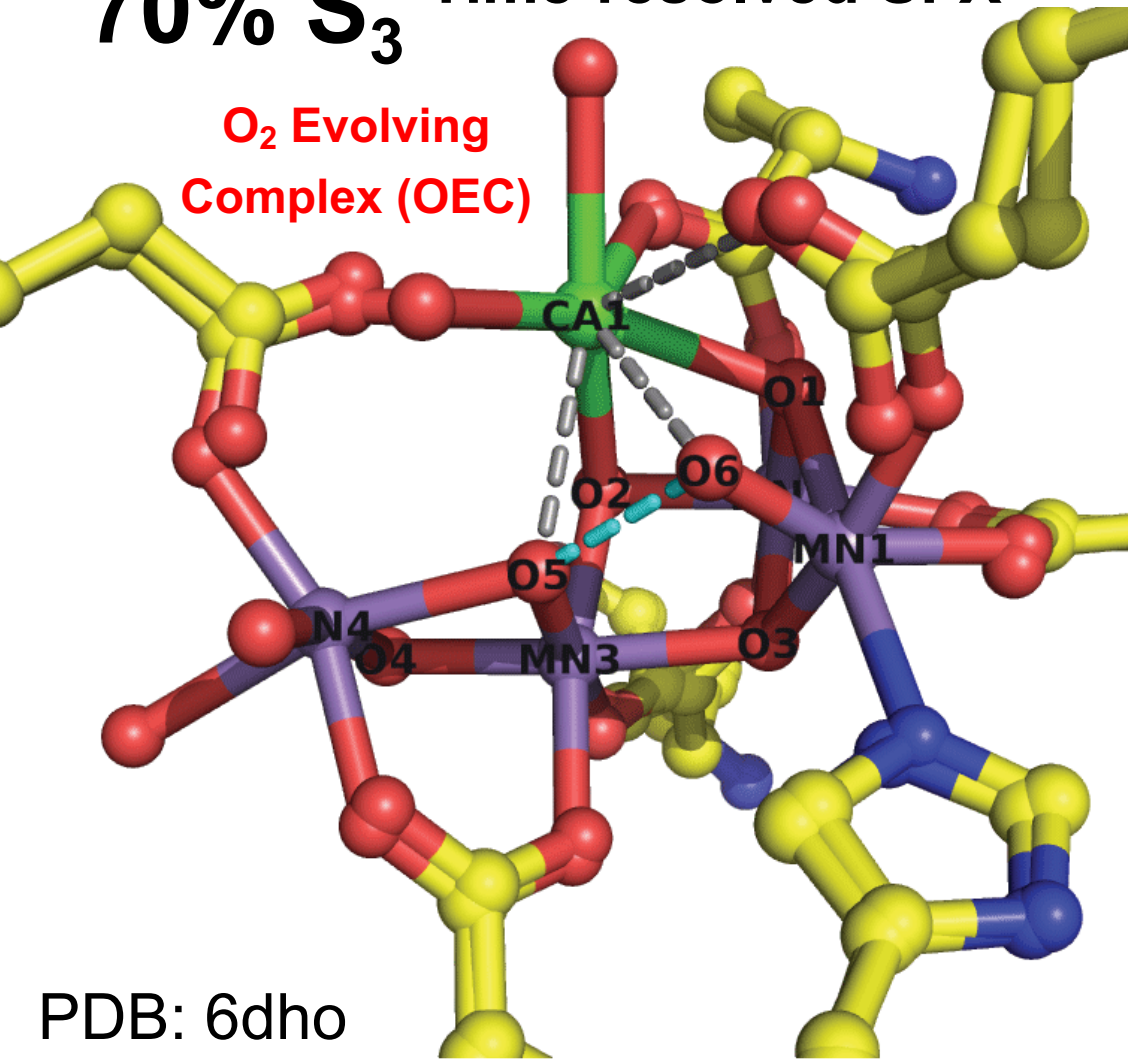
2 Flashes:

(30% S_2)

70% S_3

Figure 7.5: Correlated time-resolved SFX and XES results from photosystem II

Time-resolved SFX

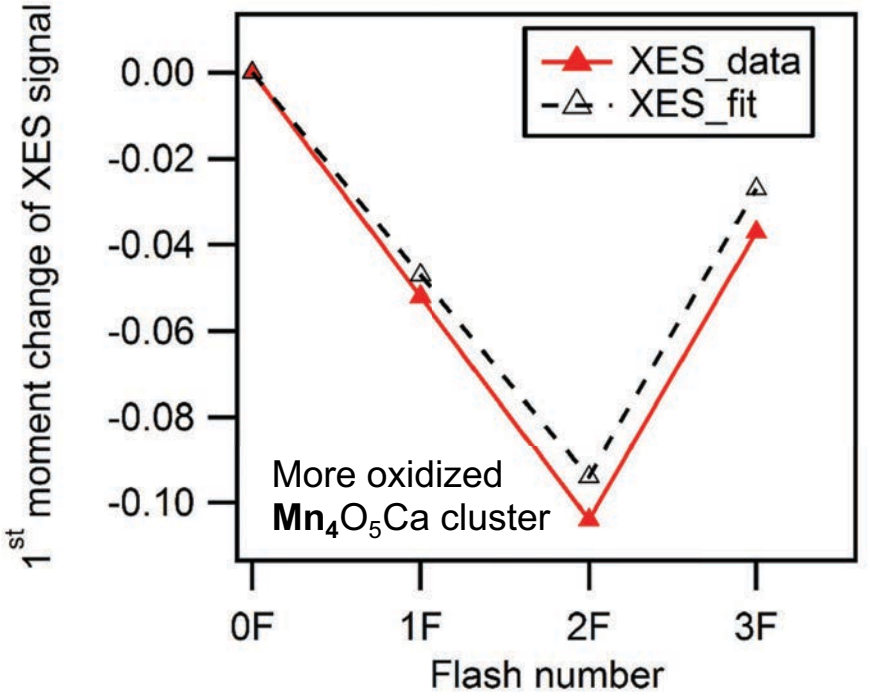


PDB: 6dho

Kern et al (2018) *Nature* 563: 421-435

Ibrahim et al (2020) *PNAS* USA 117, 12624-12635

Time-resolved X-ray emission spectroscopy (XES)



OEC

Poised in S_3

O6 new solvent

hν

O5 + O6 → O_2 ?

Membrane inlet mass spectroscopy

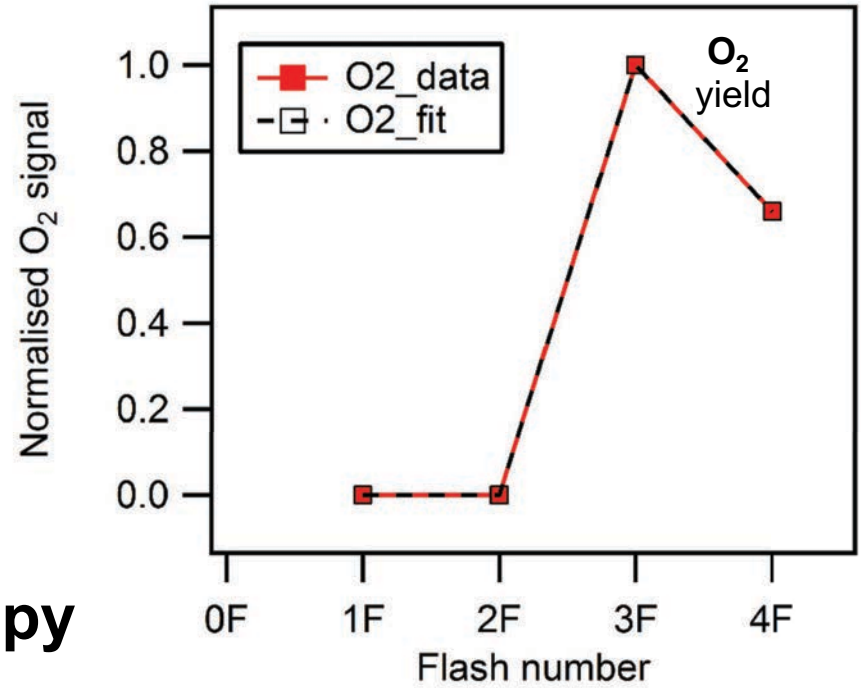
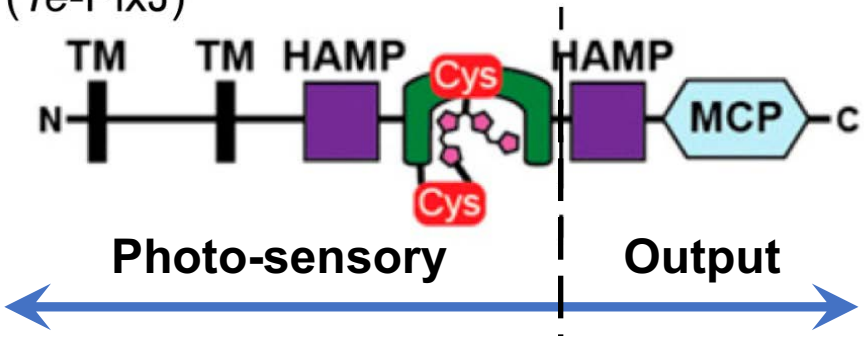


Figure 7.3: Two phytochrome photosensors: enablers of Optogenetics

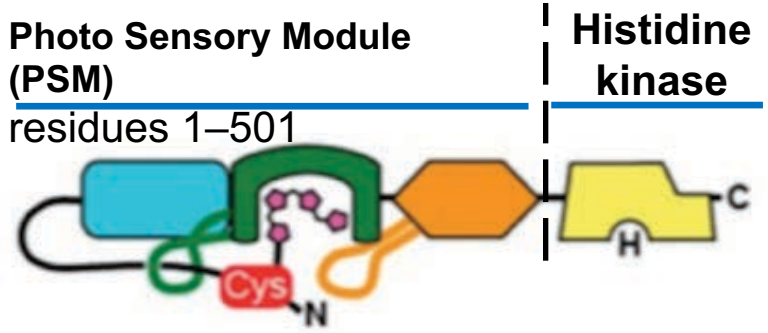
Cyanobacteria

Thermosynechococcus elongatus photoreceptor (*Te*-PixJ)

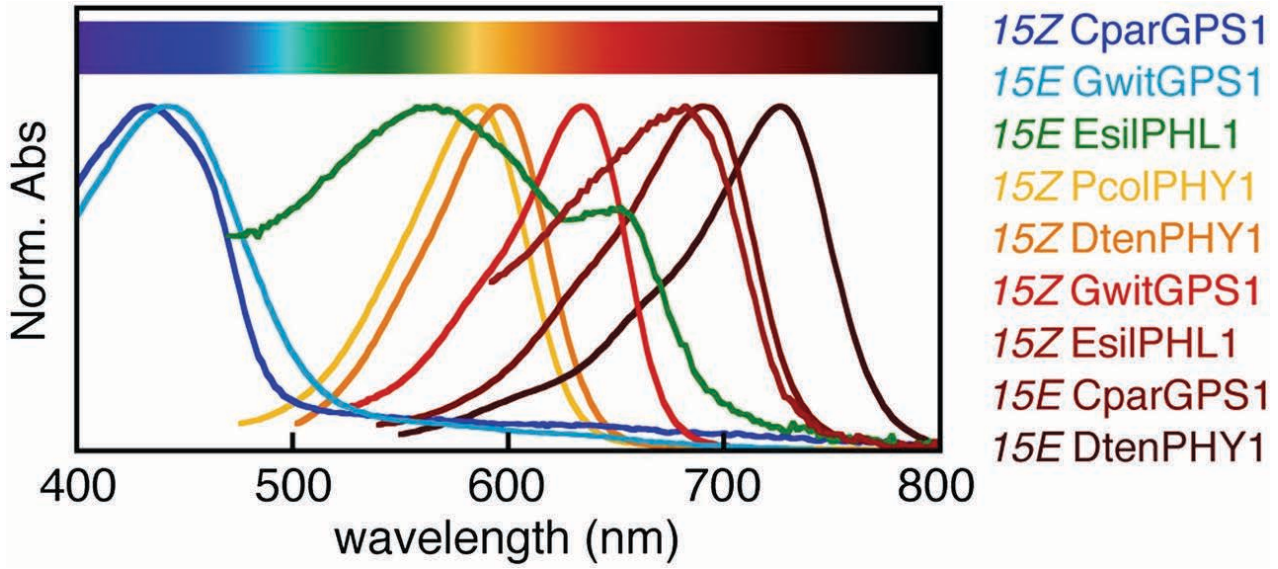


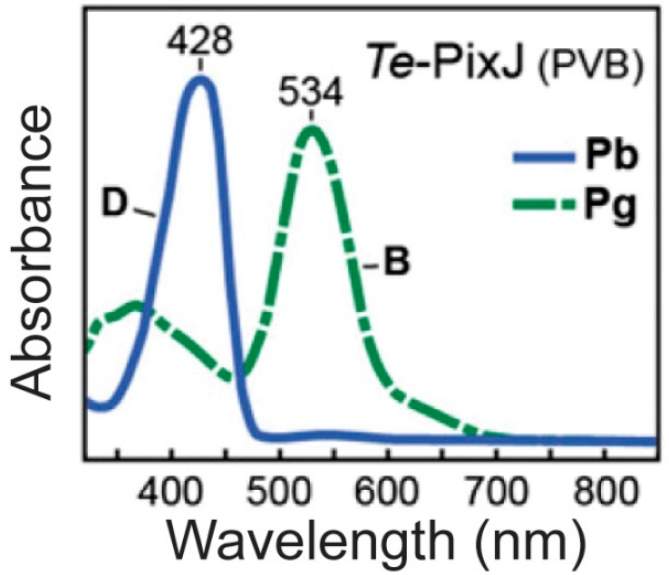
Protobacteria

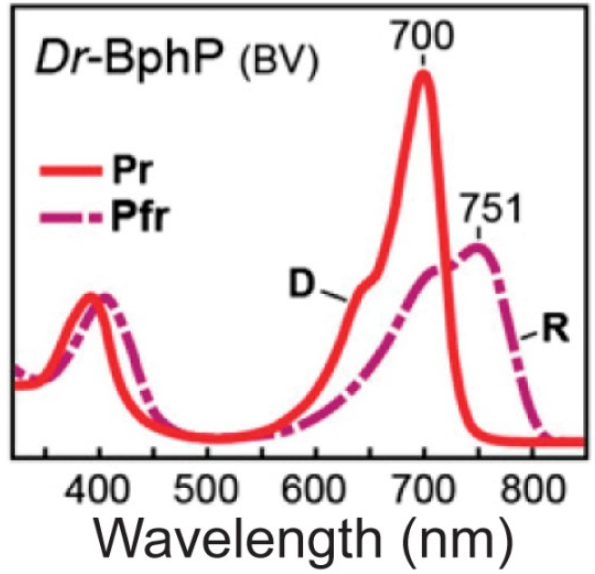
Deinococcus radiodurans phytochrome BphP (*Dr*-BphP)



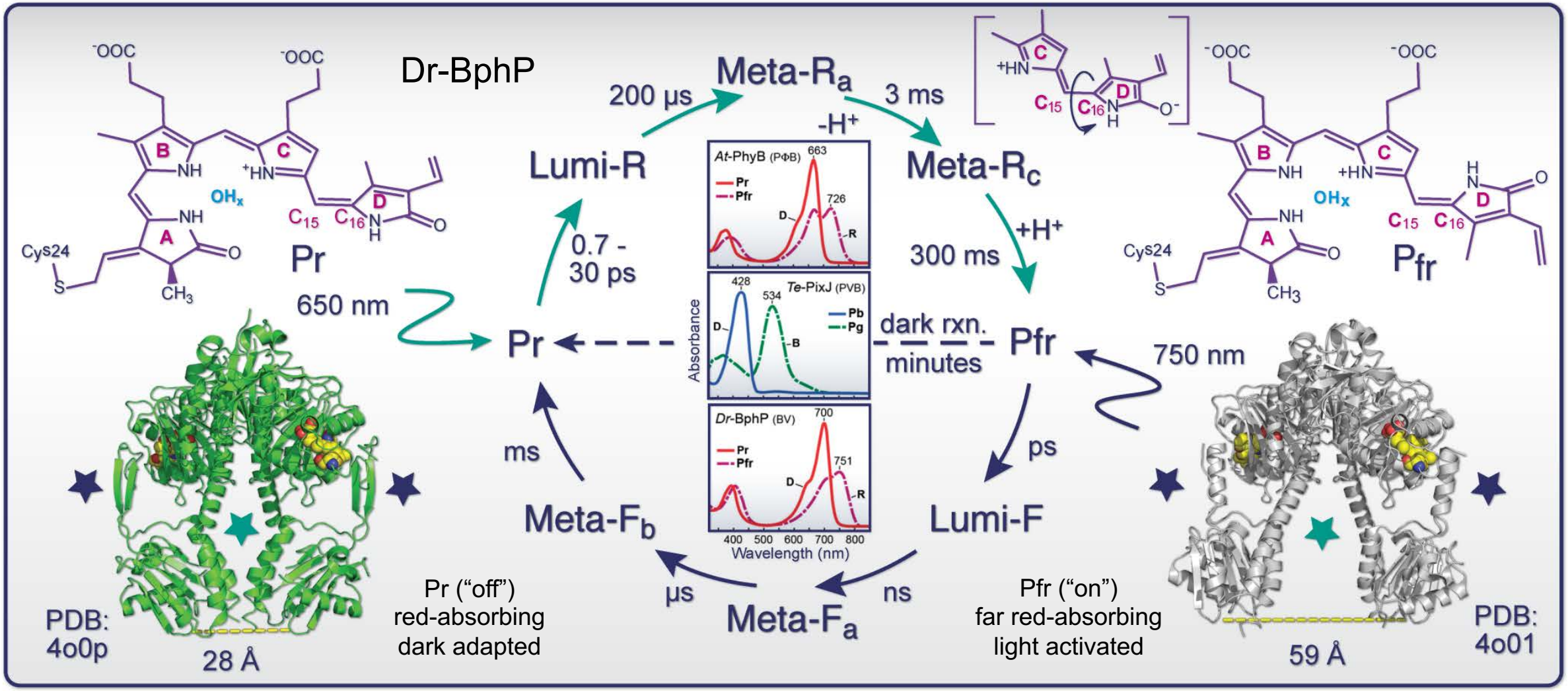
Chromophore Binding Domains (CBD)
residues 1–321







Dr-BphP photocycle and crystal structures suggest large conformational changes



Takala et al (2014) *Nature* **509**: 245-248; Burgie et al (2016) *Structure* **24**: 448–457; Burgie et al (2020) *Proc Nat Acad Sci* USA ***117***, 300-307

Single Particle Imaging (SPI) at physiological temperature

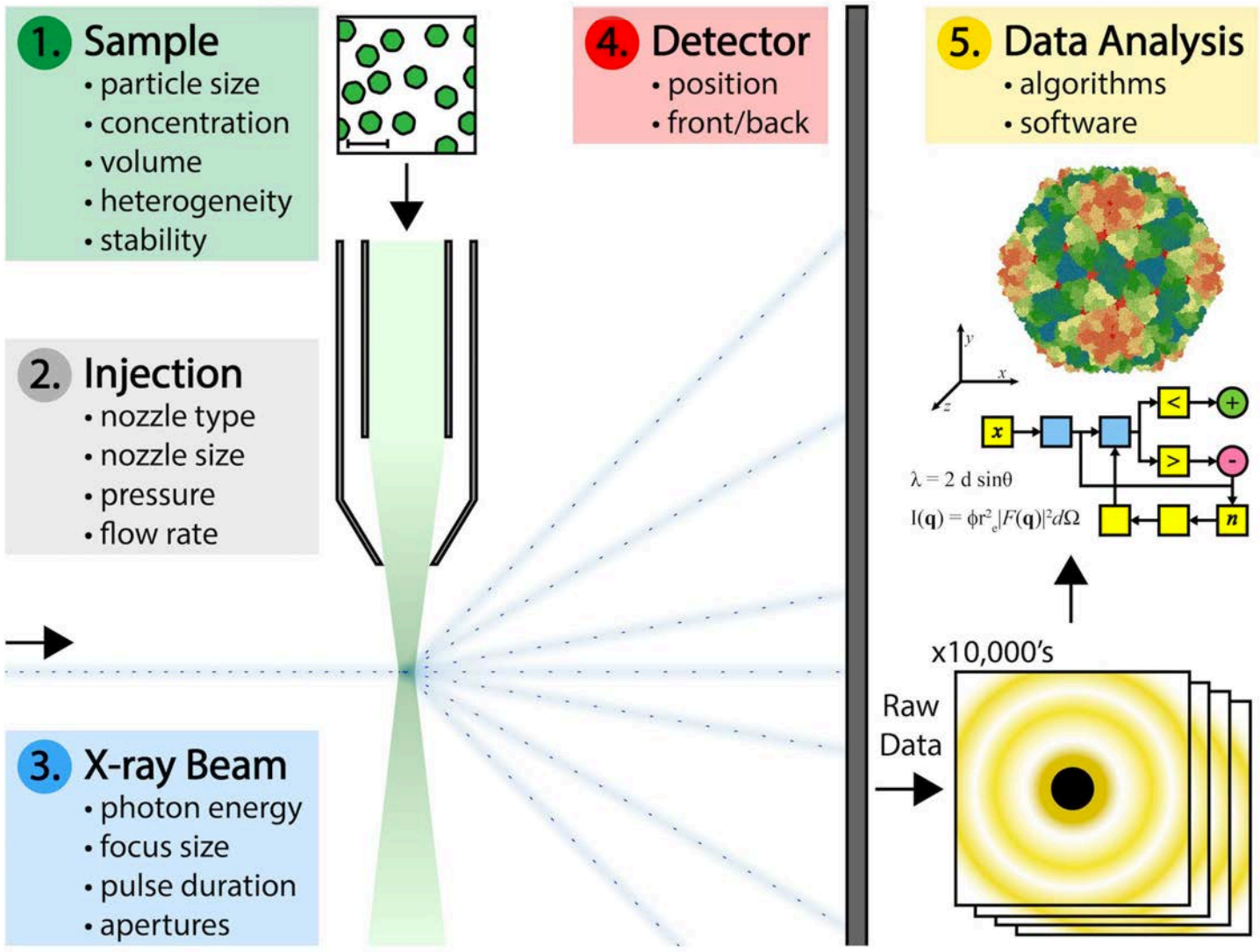
1. Sample
- particle size
- concentration
- volume
- heterogeneity
- stability

2. Injection
- nozzle type
- nozzle size
- pressure
- flow rate

3. X-ray Beam
- photon energy
- focus size
- pulse duration
- apertures

4. Detector
- position
- front/back

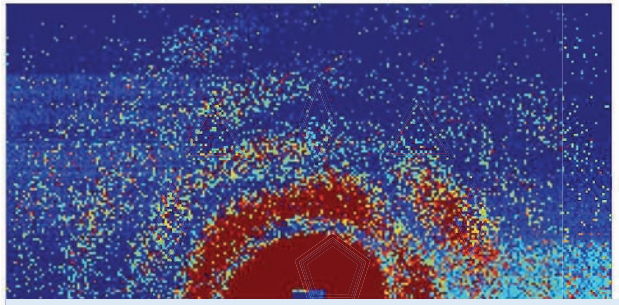
5. Data Analysis
- algorithms
- software



A raw snapshot PR772 virus icosahedral symmetry

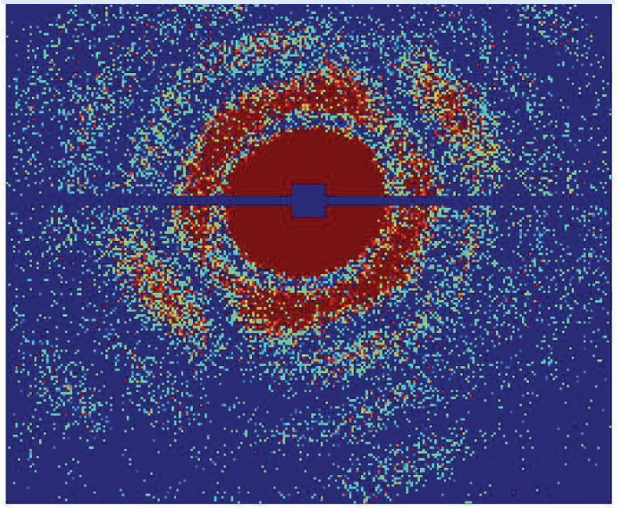
Same snapshot after background correction

Ideal test samples are often highly symmetric



Are very large, noisy datasets with modest spatial resolution of scientific value?

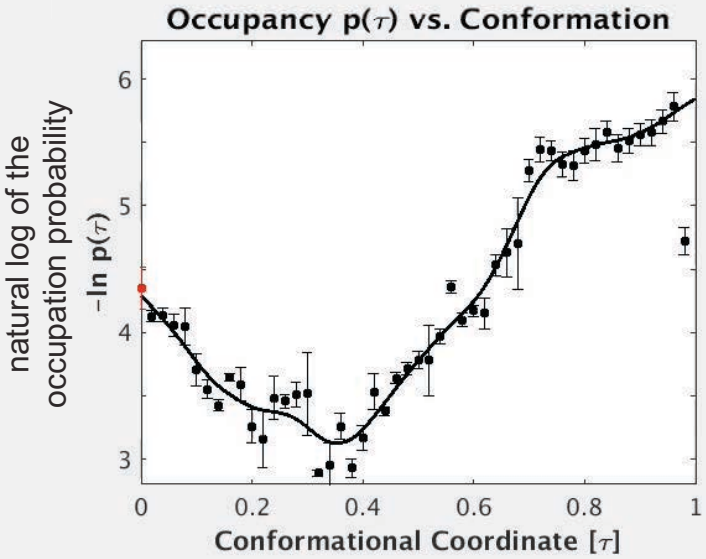
The answer is very likely yes.

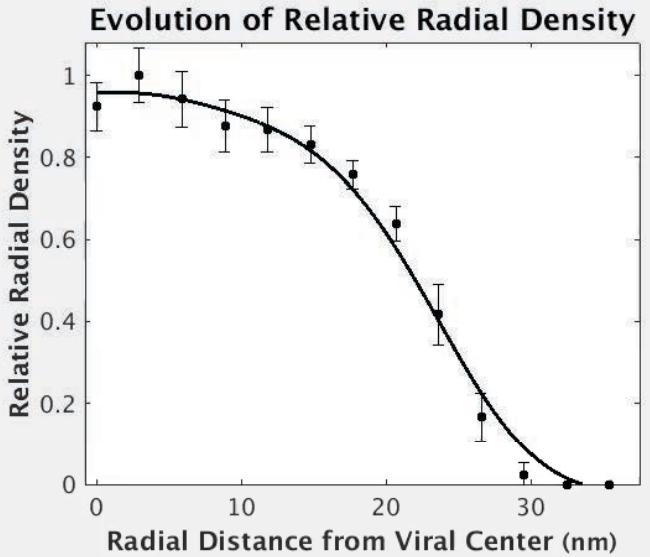


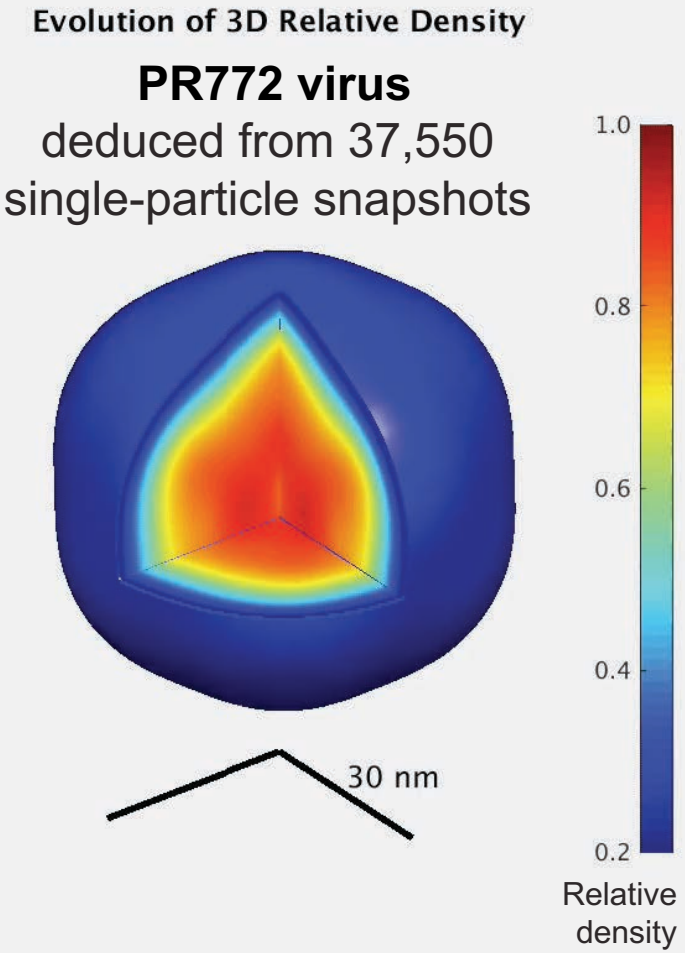
Yellow = asymmetric unit

NSF BioXFEL Science and Technology Center

SPI 3D conformational movie ***with*** imposed icosahedral symmetry

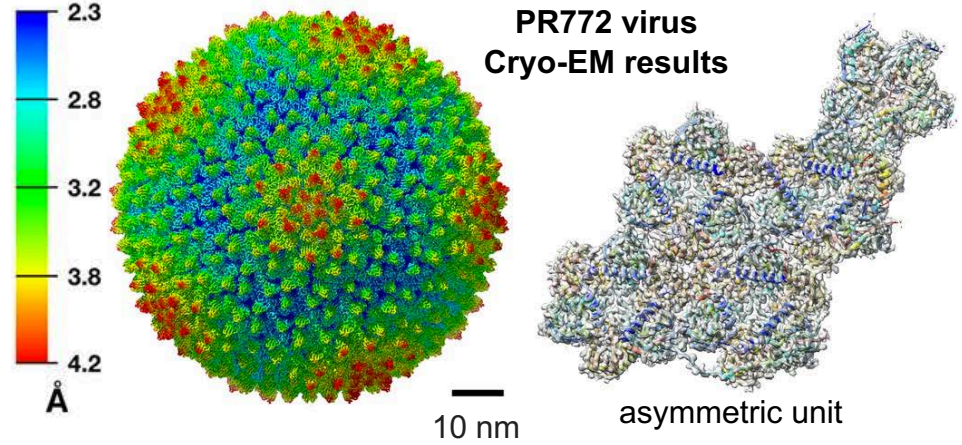


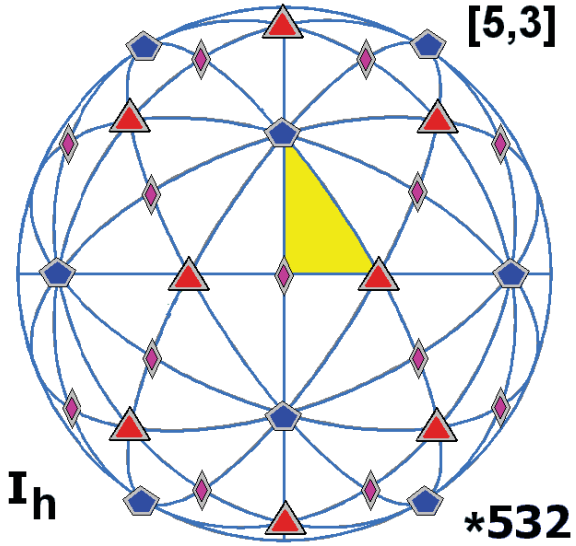




Hosseinizadeh et al (2017)
Conformational landscape of a virus by single-particle X-ray scattering
Nat Methods 14, 877-881

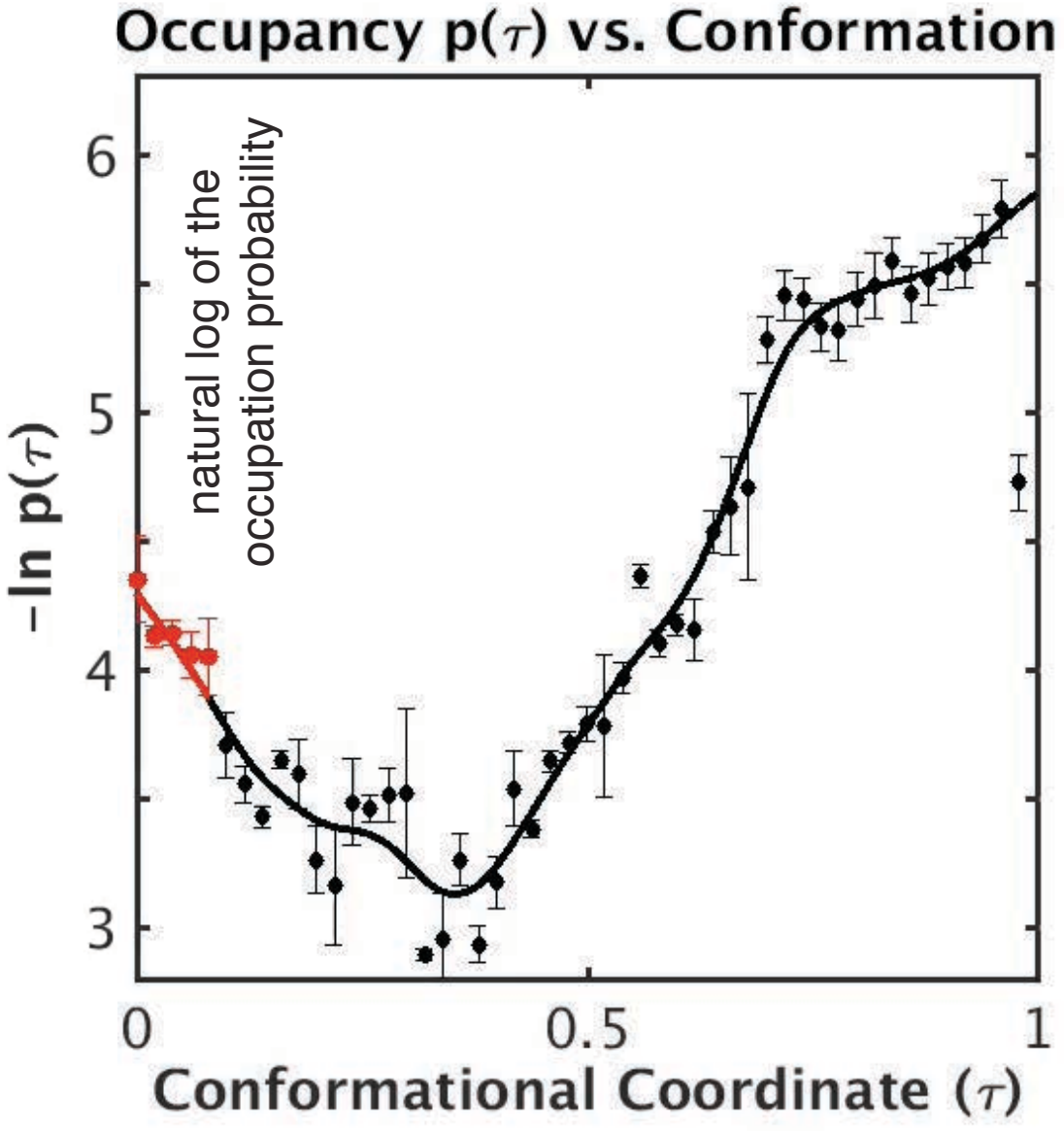
Reddy et al (2019)
Elife 8 doi: 10.7554/eLife.48496

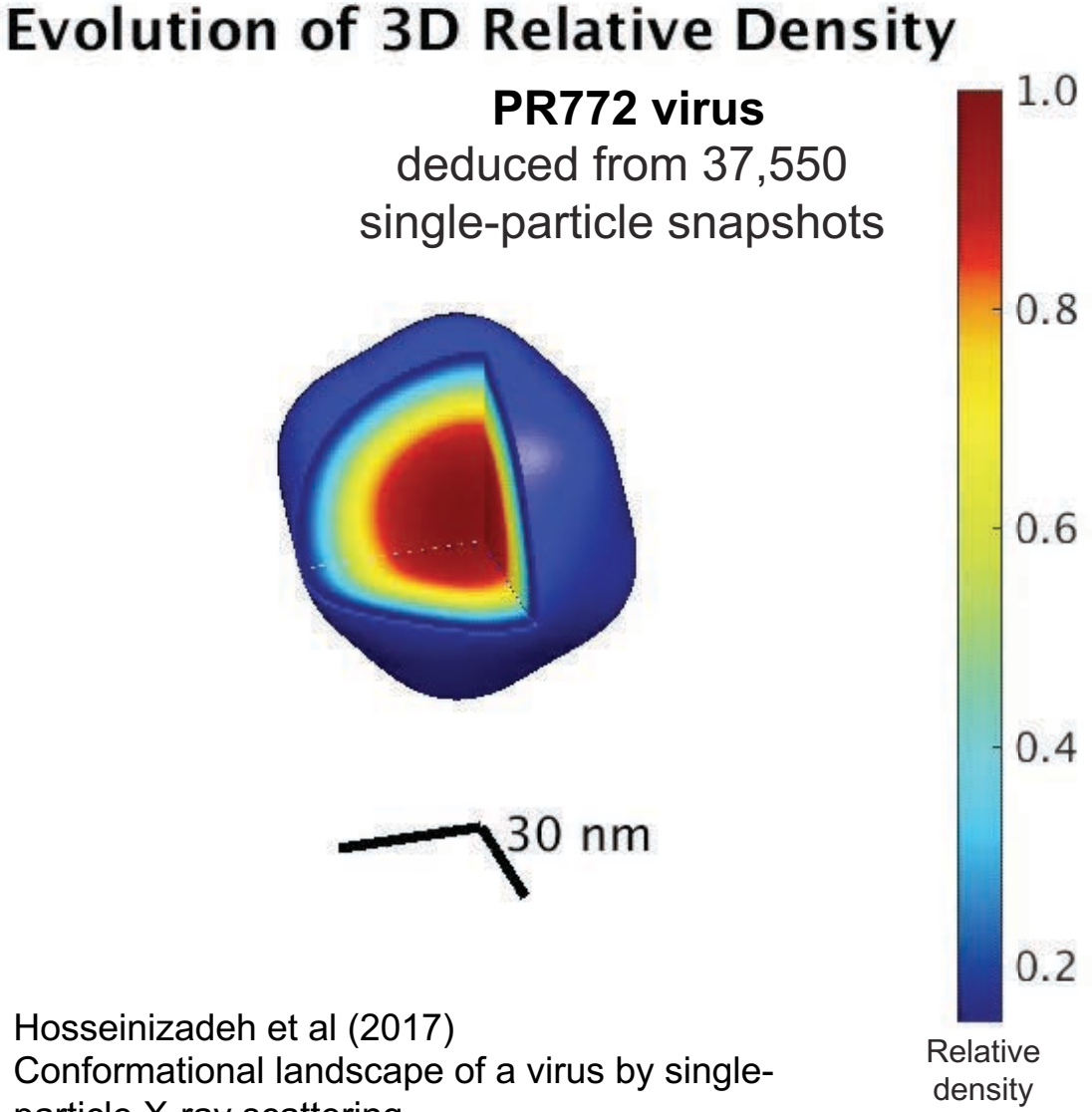




Yellow = asymmetric unit

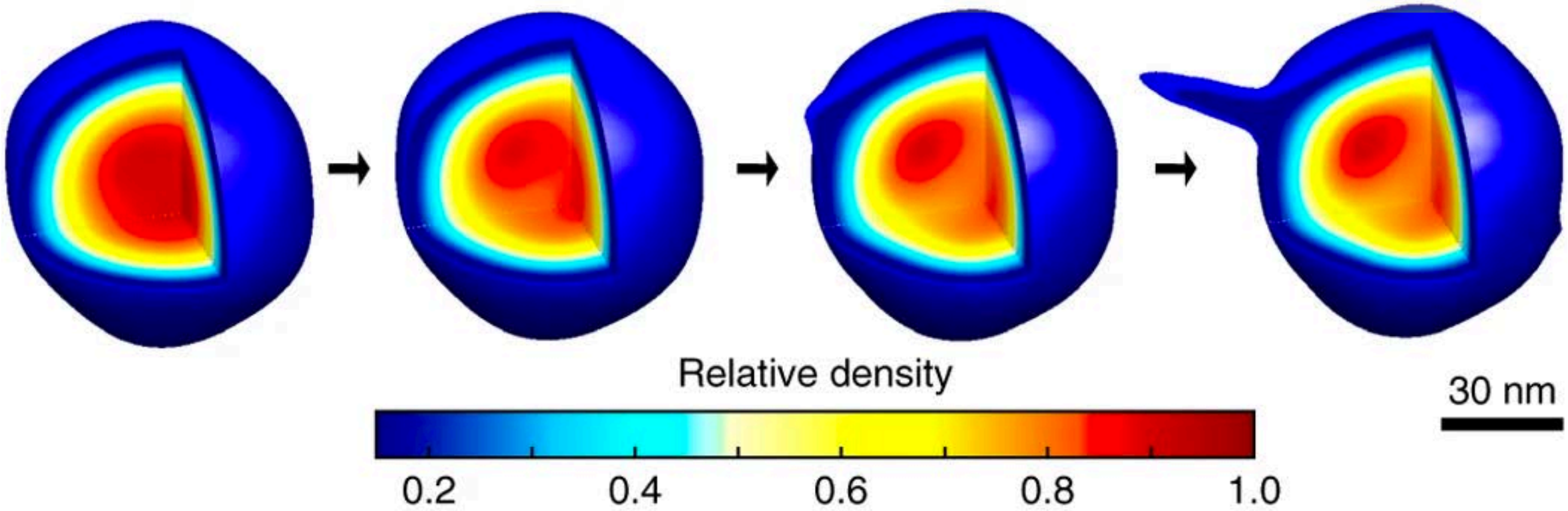
SPI 3D conformational movie ***without*** imposed icosahedral symmetry





Hosseinizadeh et al (2017)
Conformational landscape of a virus by single-particle X-ray scattering
Nat Methods 14, 877-881

Figure 3.7: 3D structures revealed by conformational analysis of 37,550 single-particle X-ray snapshots of the PR772 virus grouped by conformational parameter



The last four frames of a 50-frame movie showing the conformational changes in the PR772 virus. The movie was compiled from experimental single-particle XFEL snapshots. Note the accumulation of viral content near the fivefold portal, from which a tubular structure emerges.

Hosseinizadeh, et al. (2017)
Conformational landscape of a virus by single-particle X-ray scattering
Nat Methods 14, 877-881

Ourmazd, A. (2019)
Cryo-EM, XFELs and the structure conundrum in structural biology
Nat Methods 16, 941-944

Summary for Life Sciences

Activities happening now

- **XFEL Hub at Diamond** focusing on life science applications
  Concepts & preliminary data (Diamond &/or XFELs) → proposals for XFEL beamtime → SFX data collection → data analysis → report(s) → follow-on R&D
- **Travel assistance** to UK life scientists awarded XFEL beamtime
- **BAG access** "Dynamic Structural Biology at Diamond & XFELs"
  I24 & VMXi with fixed targets, LCP / viscous media extruder, on-demand acoustic injectors; pump-probe & mixing for time resolved studies

Activities with completion within ~ 1 – 5+ years

- Serial MX at Diamond / VMXi and at Kinetic MX at Diamond II
- Collaboration with SwissFEL for SFX sample delivery at Cristallina
- Collaboration with European XFEL for SFX sample delivery at SPB/SFX

Prospects for UK XFEL and longer-term outlook

- Biology is a large and high-impact area at all synchrotron and XFEL facilities
- The strongest current cases of XFEL use in the life sciences include SFX, time-resolved SFX and time-resolved single particle imaging (SPI)
- Dynamic structural biology & molecular movies of function will become routine at Diamond, Diamond II & XFELs
- A frontier opportunity: extend SPI methods of biomolecules in solution, to enable studies of nearly all dynamic processes with high temporal and spatial resolution.

Figure 9.4: Interactions between the XFEL Hub at Diamond and the UK life science communities