



Structural Biology with X-FEL and ERL: expectations and limitations

Soichi Wakatsuki

**Structural Biology Research Center, Photon Factory, IMSS,
KEK, Tsukuba, Japan**

Outline

I. Structural genomics and structural biology

- **New national project of Protein 3000**
- **Target oriented structural genomics on protein modification and transport of proteins**

II. Limitations of the current methods

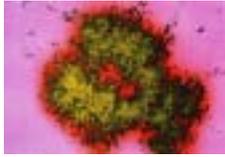
- **Beam line development and high-throughput R&D**

III. Key issues in realizing single particle and nano-crystal structural biology



Part I Role of Structural Biology & Structural Genomics

Genome

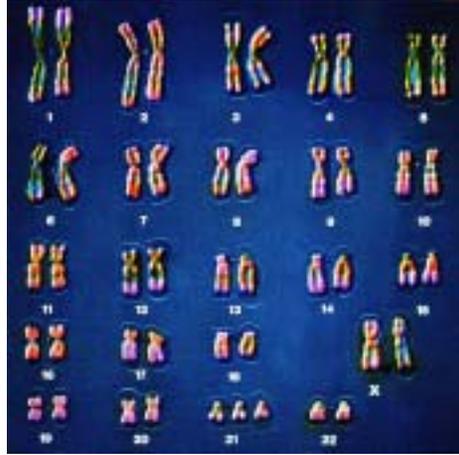
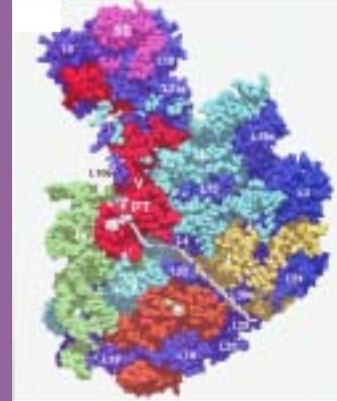


DNA RNA

Proteins

Molecular machines
(ribosomes, enzymes)

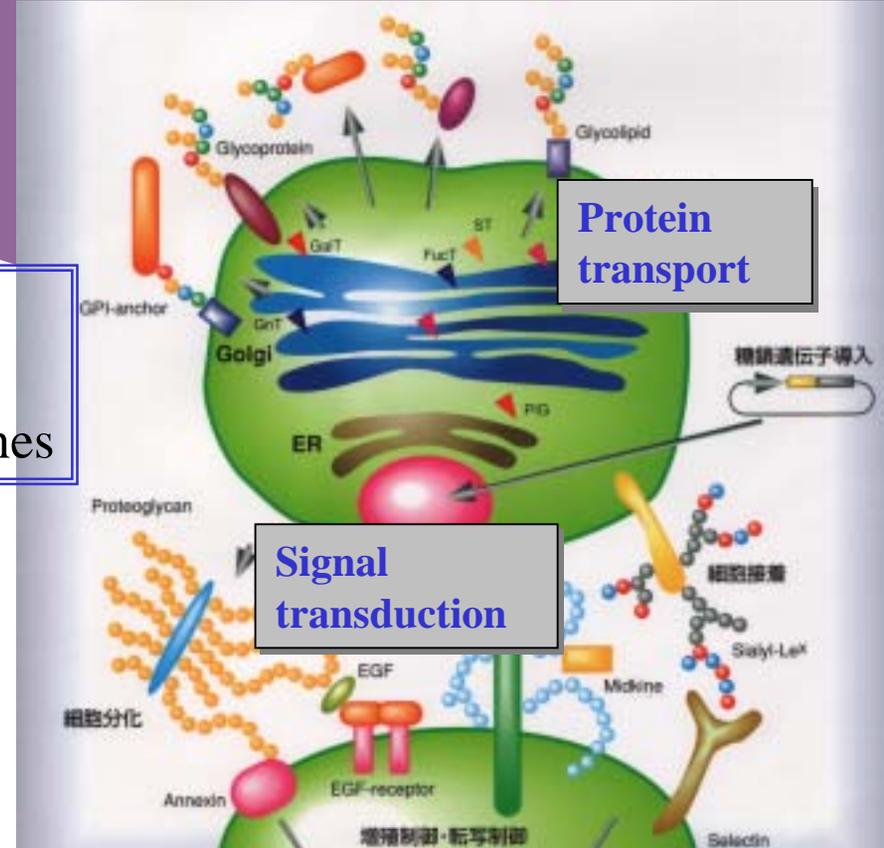
Information network



Atomic resolution analysis
Protein structure and dynamics
Interactions between molecular machines

Biology
(Basic research)

Medicine, drug design
(industrial applications)



Currently 35 Structural Genomics Projects Worldwide

International Structural Genomics Organization

<http://www.isgo.org>

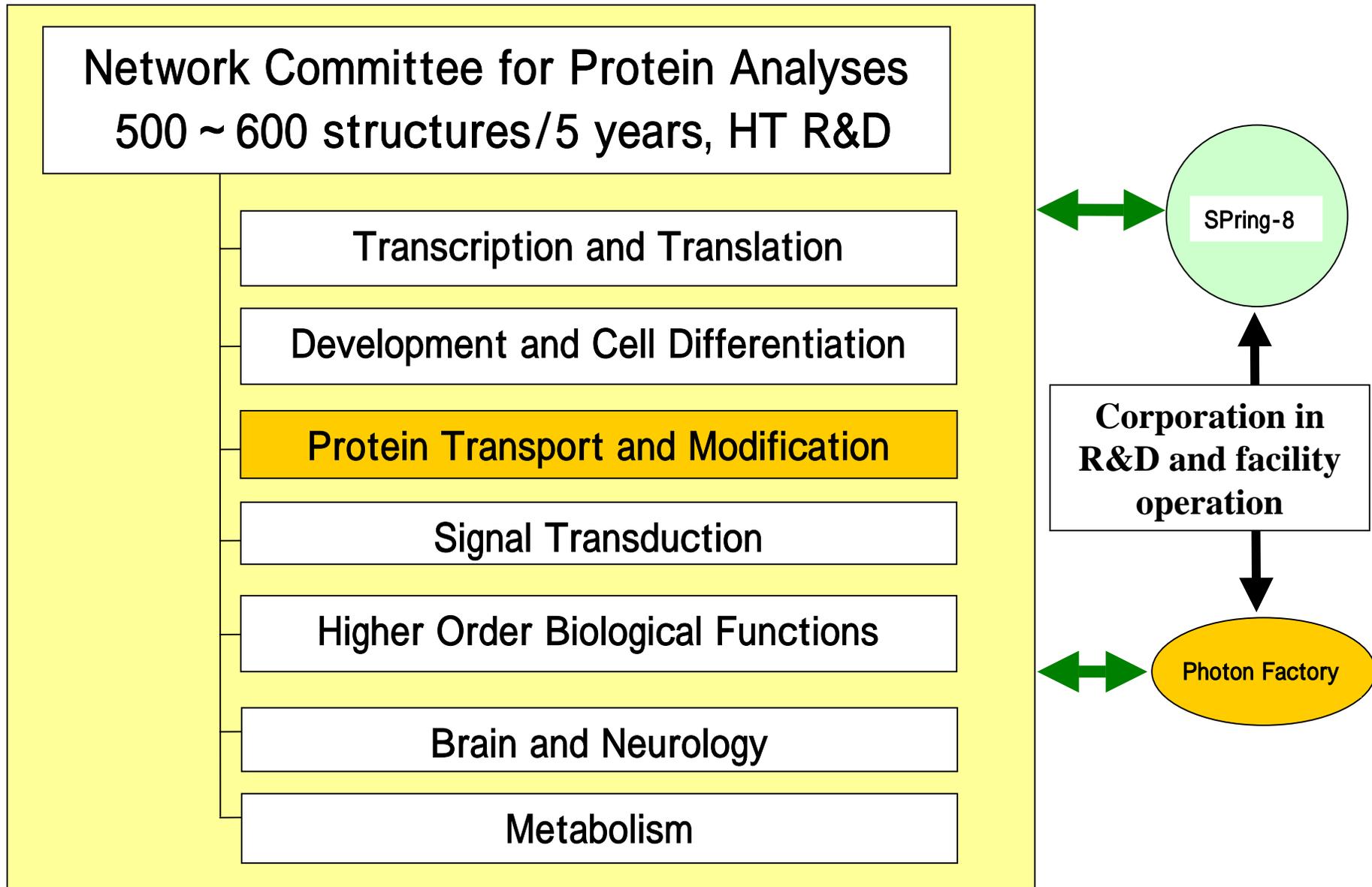
Structural Genomics and Proteomics Project list

-Worldwide Initiatives -

[Australia](#) / [Canada](#) / [EU](#) / [France](#) / [Germany](#) / [Japan](#) / [Korea](#) / [Switzerland](#) / [UK](#) / [USA](#)

Australia:	3 planned	Canada	4
EU	1	France	4
Germany	1 & 1 planned	Japan	10
Korea	1	Switzerland	1
UK	3	USA	10

Target oriented structural genomics consortia of universities and national institutes





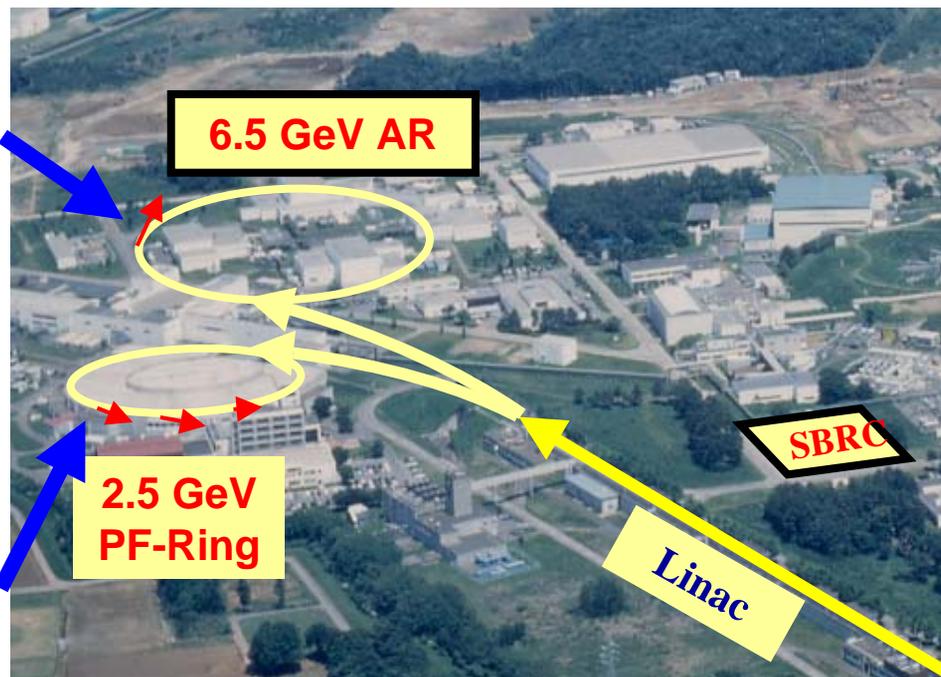
KEK-PF Structural Biology Research Center



AR NW12



PF BL5



Crystallization
200,000 trials/day

Newly extended building

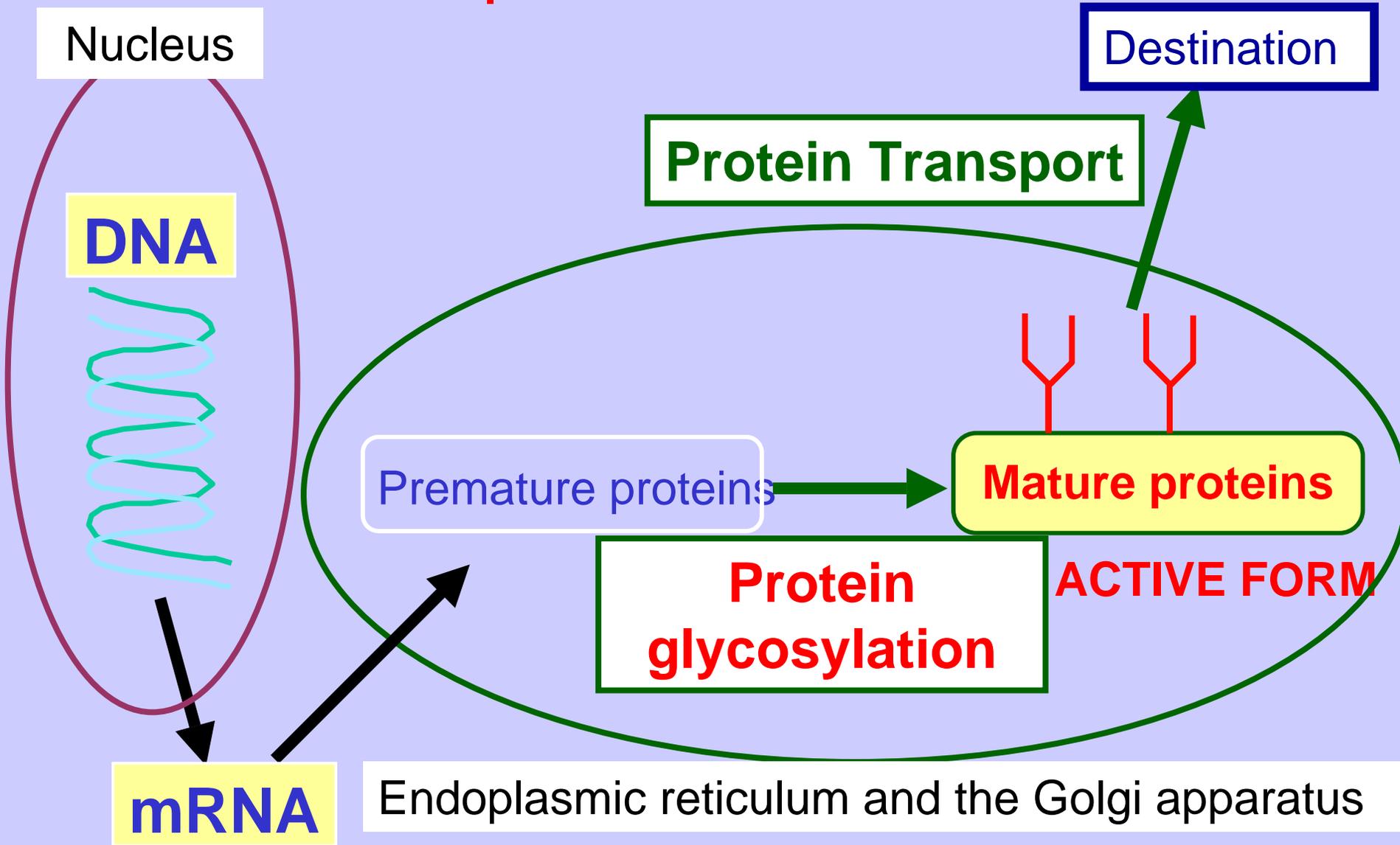
Next generation detector:
X-ray-HARP

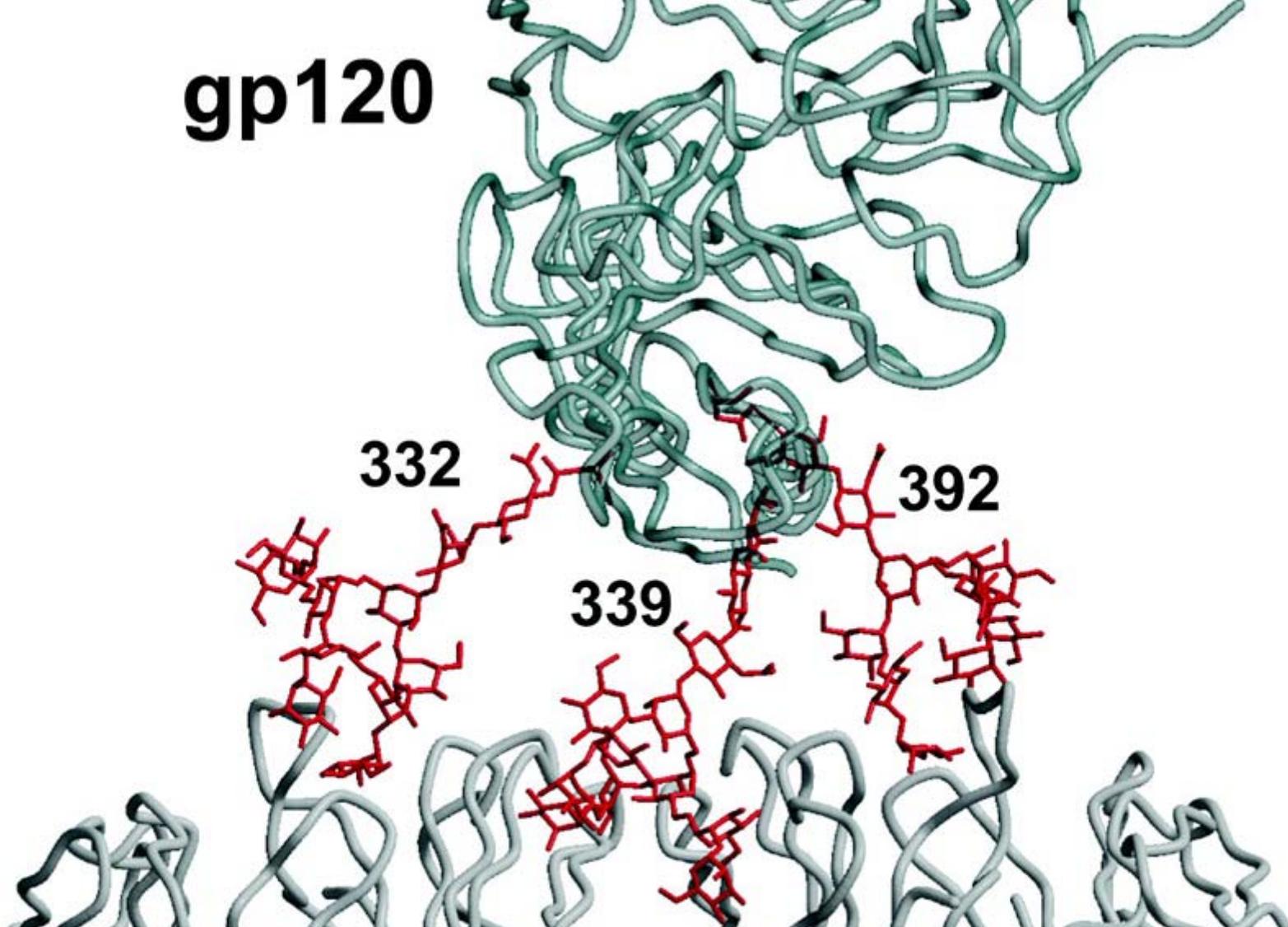
Electron beam scans the prods by applying voltages to the cathode and gate electrodes

Post-translational modification and transport of proteins

Micro-manipulator

More than half of our proteins are modified with oligosaccharides (glycosylated) to become mature proteins and transported to correct destinations





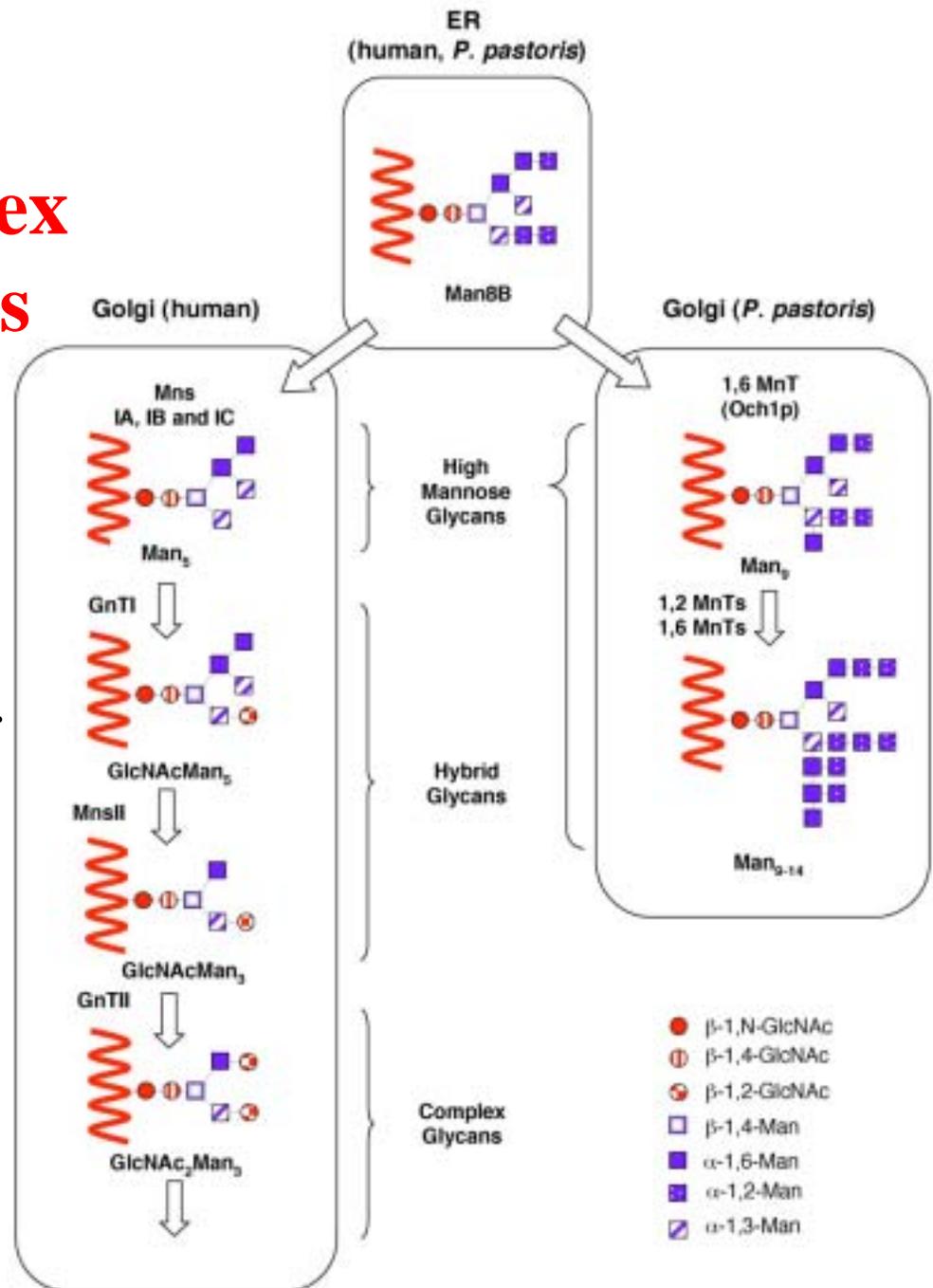
Model of 2G12 glycan recognition of gp120. On the basis of our model, three separate $\text{Man}_9\text{GlcNAc}_2$ moieties, shown in red (two in the primary combining sites and one in the V_H/V_H' interface), potentially mediate the binding of 2G12 to gp120. D.A. Calarese et al., *Science* 2003 June 27; 300: 2065-2071.

Production of Complex Human Glycoproteins in Yeast

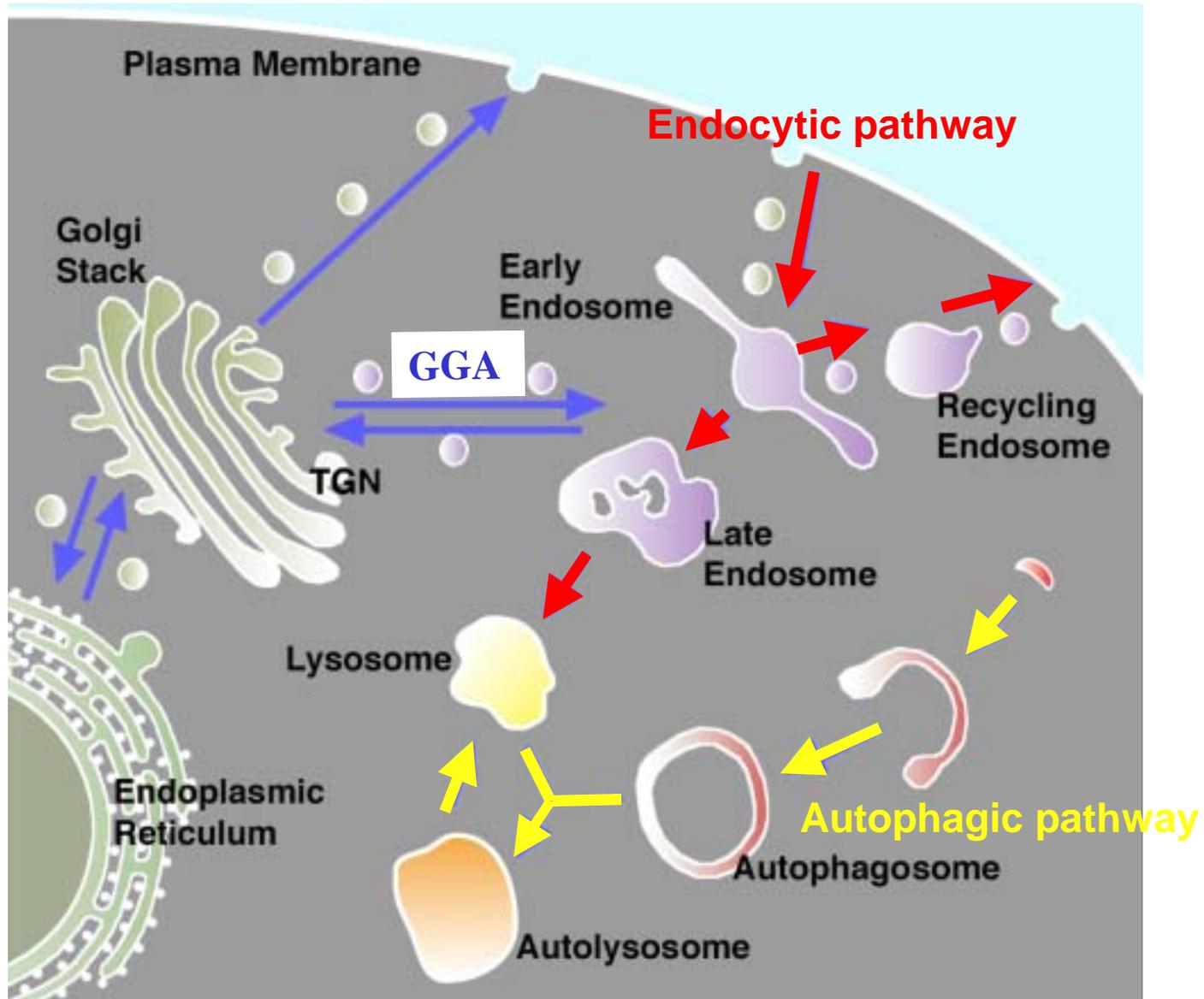
Stephen R. Hamilton et al.

Science 2003 301: 1244-1246.
29 Aug 2003

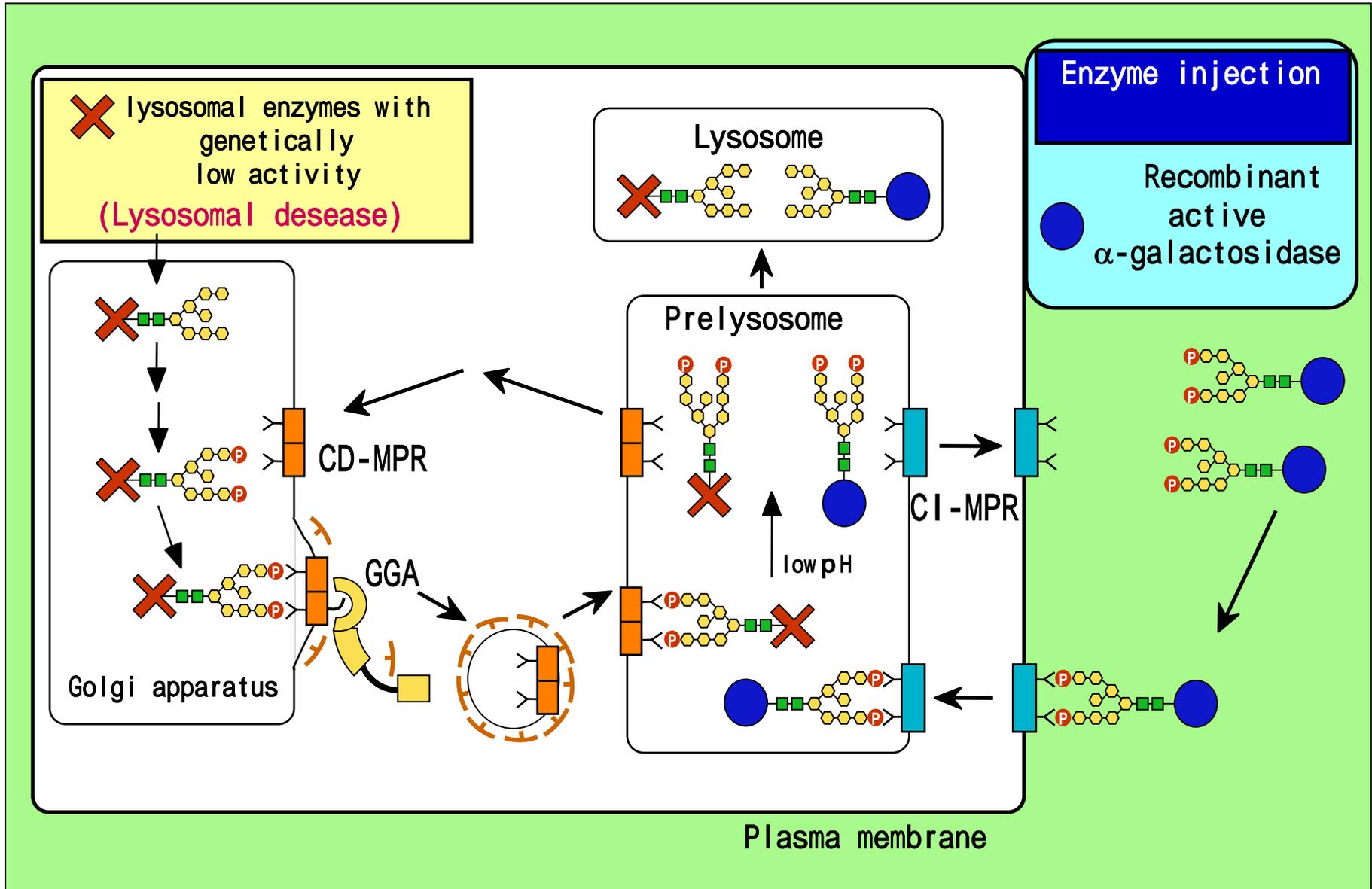
New frontiers of highly effective drugs



Lysosomal Function Depends on Membrane Traffic



Treatment of lysosomal diseases

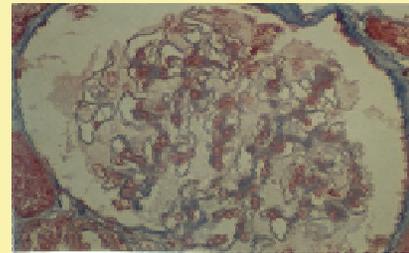


Fabry Disease and Enzyme Replacement Therapy

Fabry disease : A disease caused by mutation of α -galactosidase gene, which degrades enzymatic activity of the hydrolase in lysosome leading to accumulation of glycolipids



Enlargement of the heart

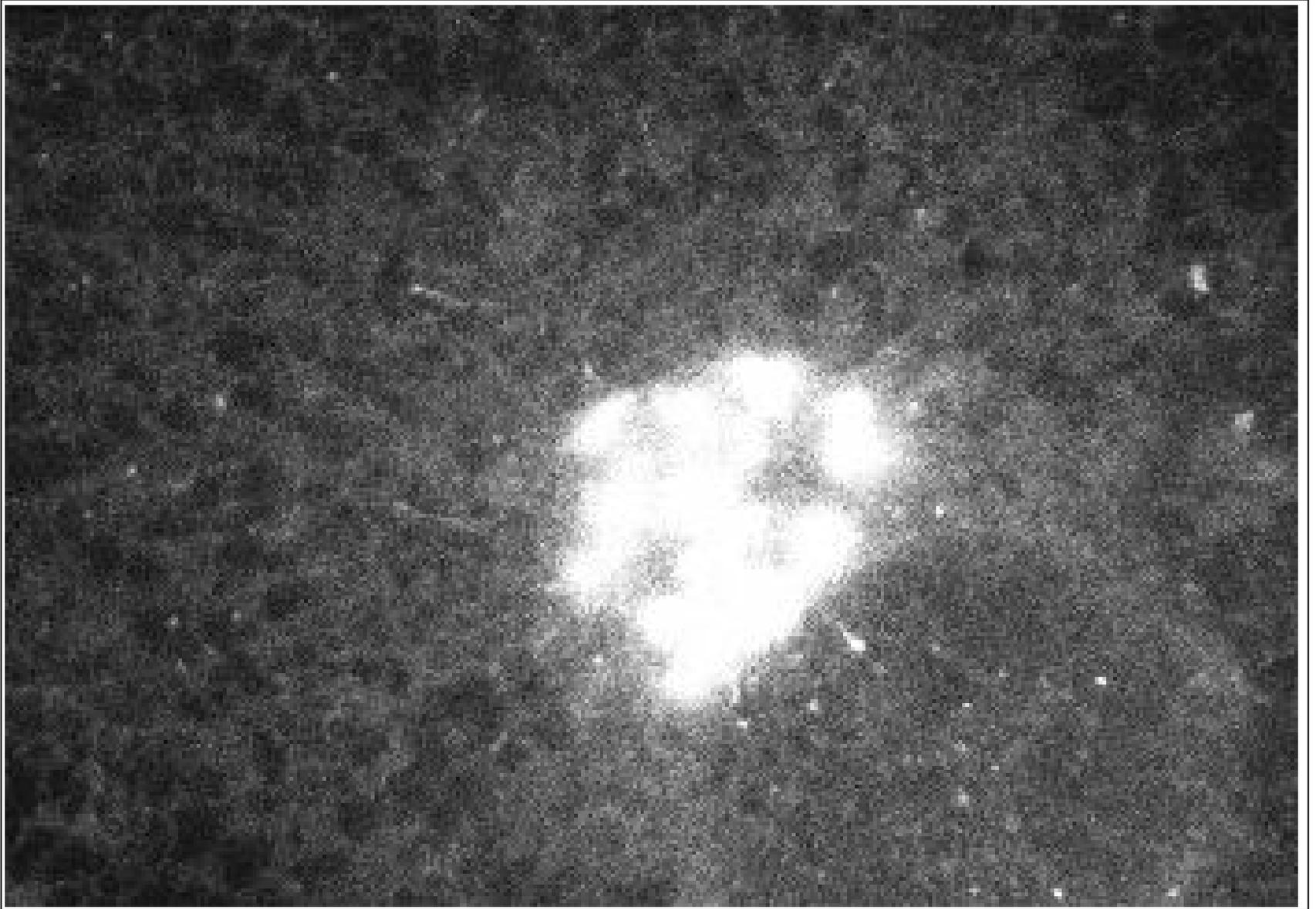


Accumulation of glycolipids in glomerulus

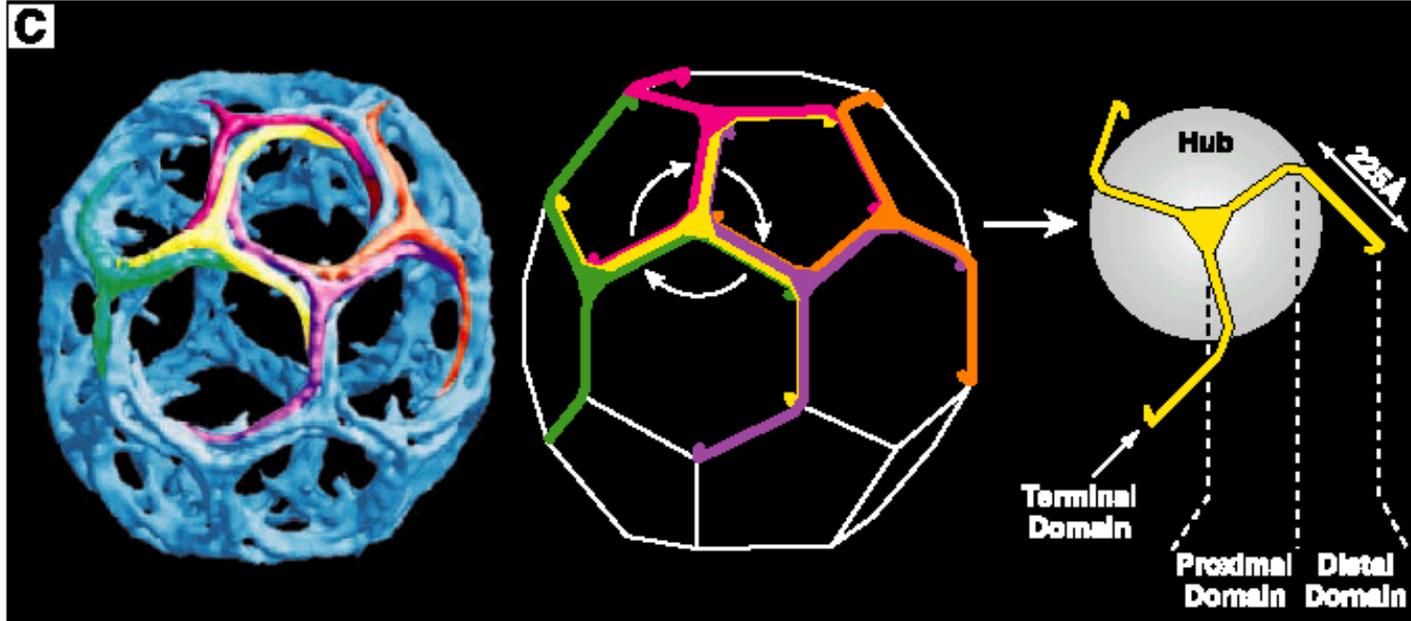
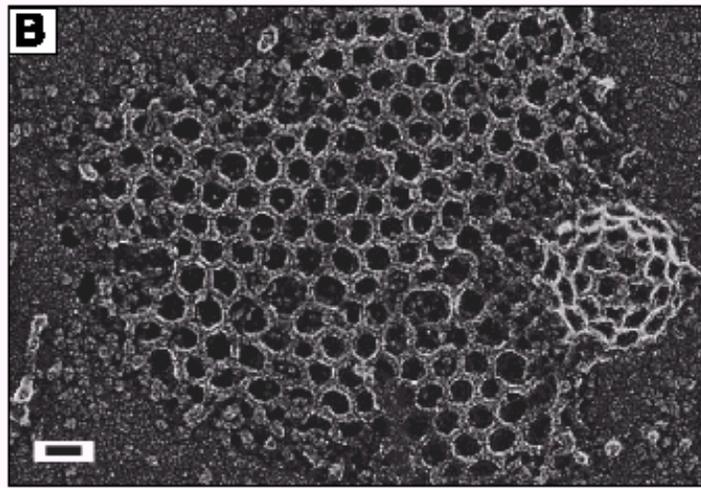


Angiokeratoma

Vesicle transport from the ER to the Golgi apparatus



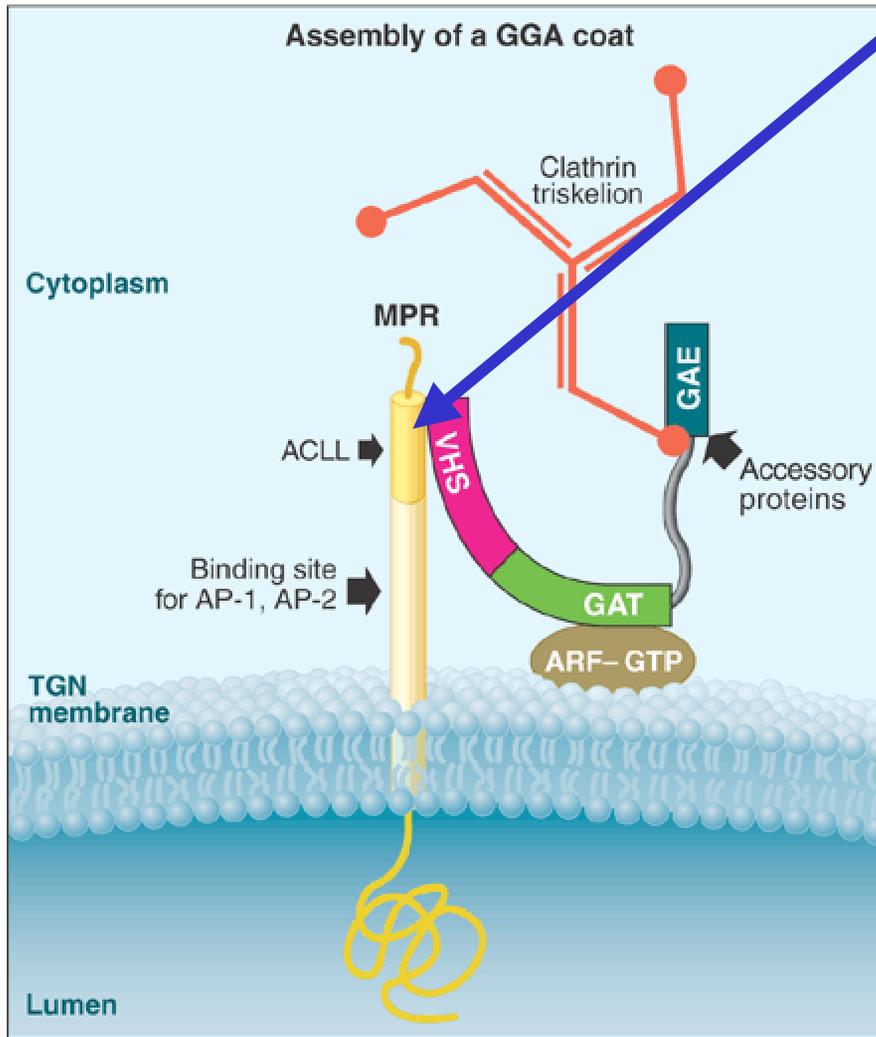
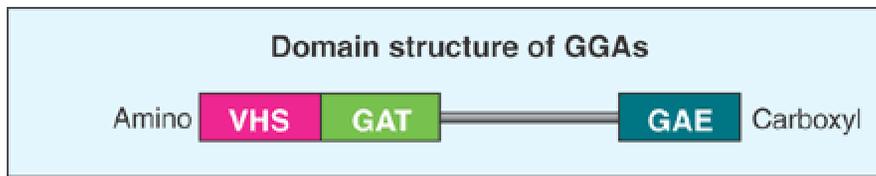
Lippincott-Schwartz, J. (1998) MBC 9, 1617



Clathrin
movie by
Allison
Bruce,
Harvard
University

M. Marsh and H. T. McMahon, *Science*, 1999, Vol. 285, 215

<http://www.hms.harvard.edu/news/clathrin/>



ACLL (acidic dileucine) motif

ACLL Peptides recognized by GGA1-VHS domain

LRP3	-MLEASDDEALLVC
CD-MPR	-EESERDDHLLPM
CI-MPR	-SFHDDSDDLLHI
Sort (WT)	-GYHDDSDDLLLE
Sort (DD/NN)	-GYHNSDEDLLE
Sort (S/A)	-GYHDDADEDLLE
Sort (S/D)	-GYHDDDDDLLLE
Sort (DED/NQN)	-GYHDDSNQNLLE
Sort (LL/AA)	-GYHDDSDEDAEE
β -secretase	-QHDDFADDISLLK

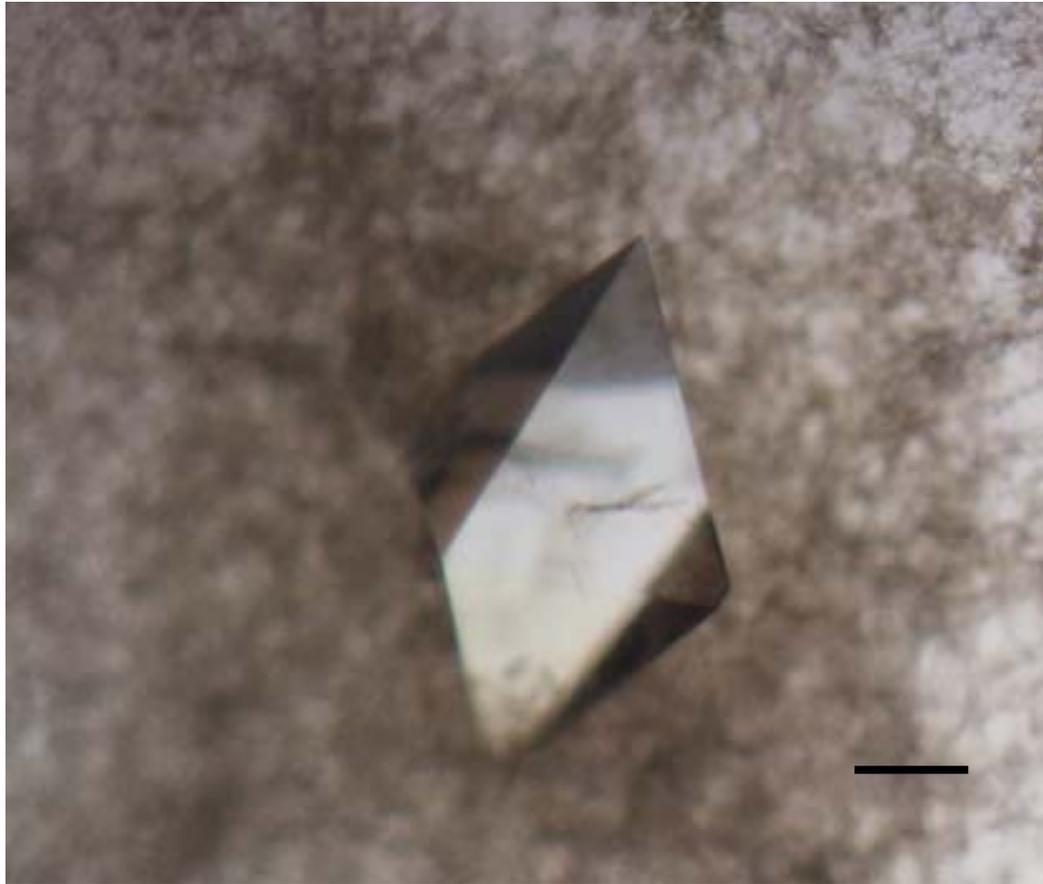
Red: acidic residues

Blue: leucine pairs

Purple: serine residues that can be phosphorylated by CK-II

Takatsu et al, J. Biol. Chem. 276, 28541-28545

**Crystal of
Human GGA1
VHS domain**



bar = 0.1 mm

Crystallization method: hanging drop vapor diffusion
Protein conc.: 13 mg / ml
Precipitant: 17 % (w/v) PEG3350, 0.2 M KH_2PO_4
Buffer: 0.1 M Tris-HCl (pH 7.5)
Temperature: 20 °C

Monday 5 PM, 13 August, complex crystals FedExed to ALS
Wednesday 1 PM, 15 August, 1.8Å data set collected at ALS!

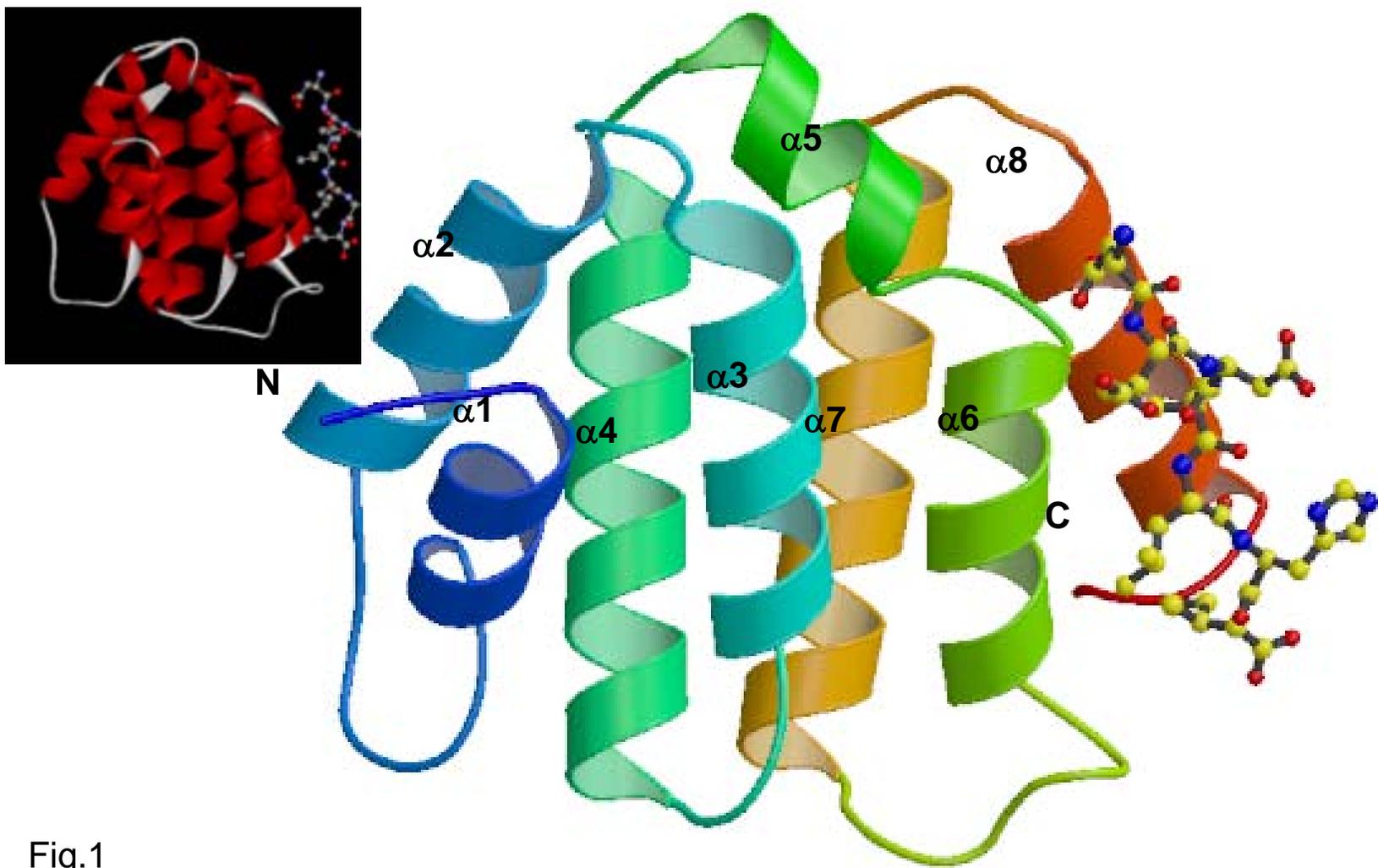


Fig.1
Ribbon diagram of VHS domain of human GGA1 complex with M6PR peptide. The peptide molecule is shown as a ball-and-stick model colored according to atom type (nitrogen, blue; carbon, yellow; oxygen, red).

Structural basis for acidic-cluster-dileucine sorting-signal recognition by VHS domains

Saurav Misra^{*}, Rosa Puertollano[†], Yukio Kato[†], Juan S. Bonifacino[†] & James H. Hurley^{*}

^{} Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA*

[†] Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

pp 933-937

Structural basis for recognition of acidic-cluster dileucine sequence by GGA1

Tomoo Shiba^{*†‡}, Hiroyuki Takatsu^{‡§}, Terukazu Nogi^{*}, Naohiro Matsugaki^{*}, Masato Kawasaki^{*}, Noriyuki Igarashi^{*}, Mamoru Suzuki^{*}, Ryuichi Kato^{*}, Thomas Earnest^{||}, Kazuhisa Nakayama[§] & Soichi Wakatsuki^{*}

^{} Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan*

[†] Foundation for Advancement of International Science (FAIS), Tsukuba, Ibaraki 305-0062, Japan

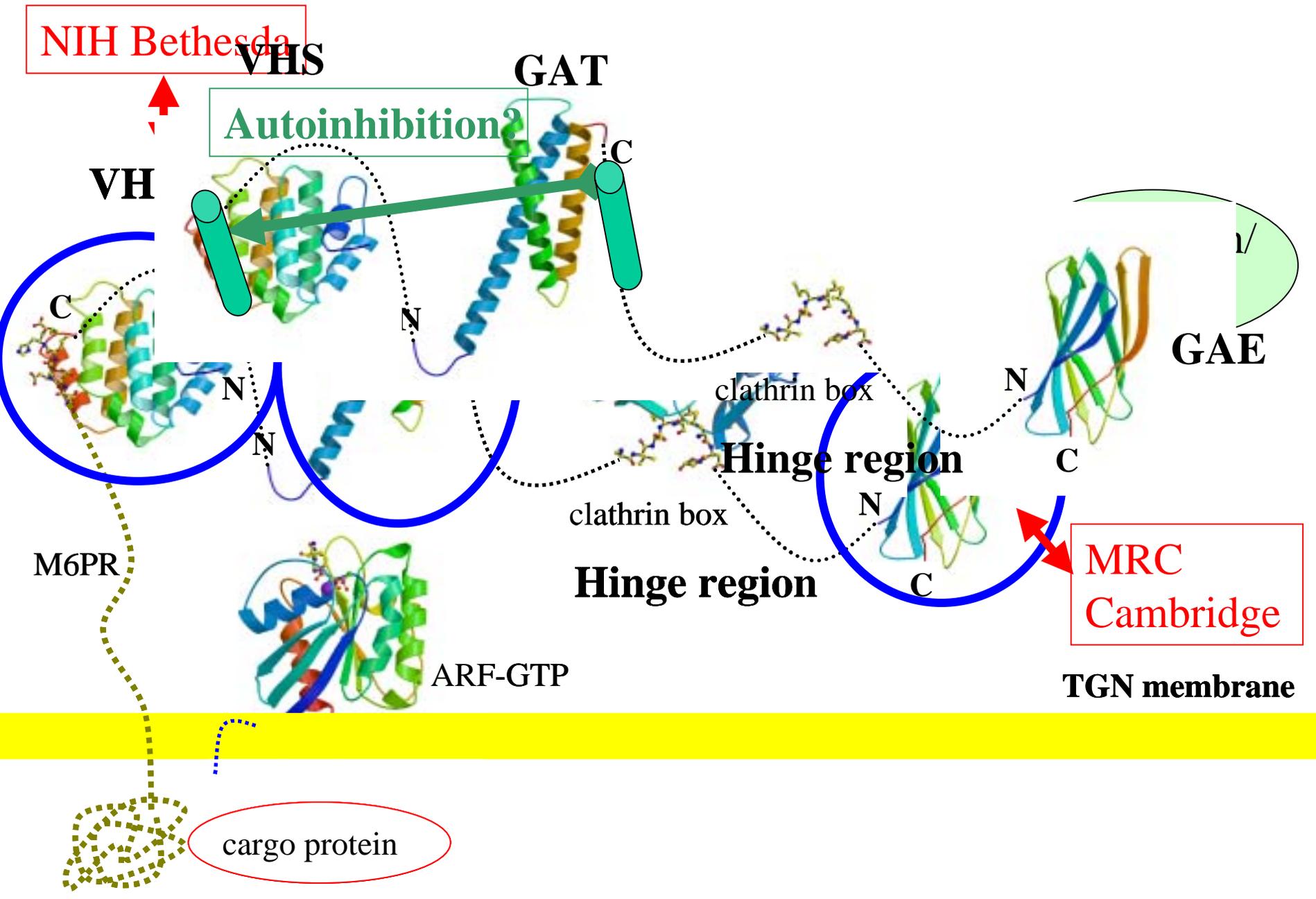
[§] Institute of Biological Sciences and Gene Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan

^{||} Advanced Light Source, Berkeley, Berkeley Center for Structural Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

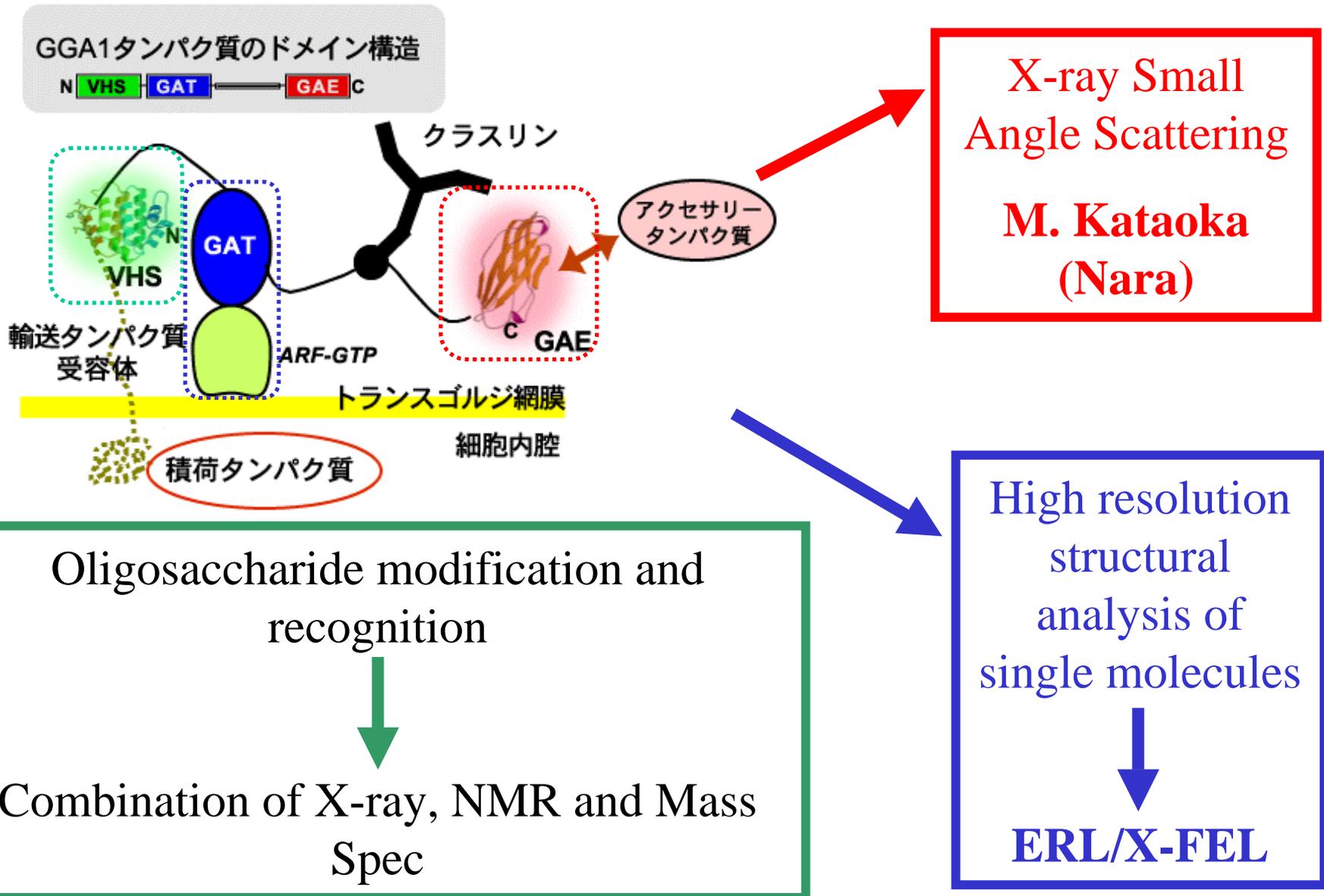
[‡] These authors contributed equally to this work

pp 937-941

Model for the assembly of GC



Future Directions of structural and functional analyses of posttranslational modification and intracellular transport



GGA domains are far apart! May be impossible to crystallize the entire protein? What else can we do?

Small Angle X-ray scattering?

Energy Recovery Linac (Next generation SR)?

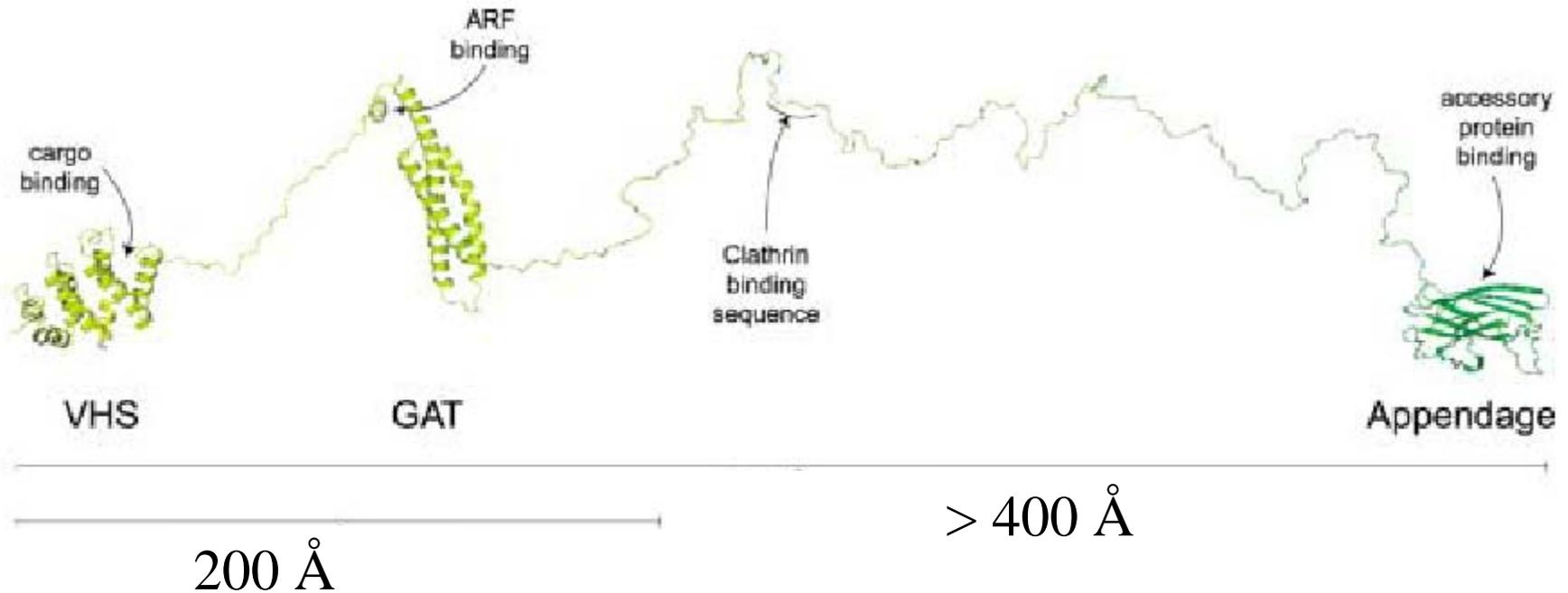
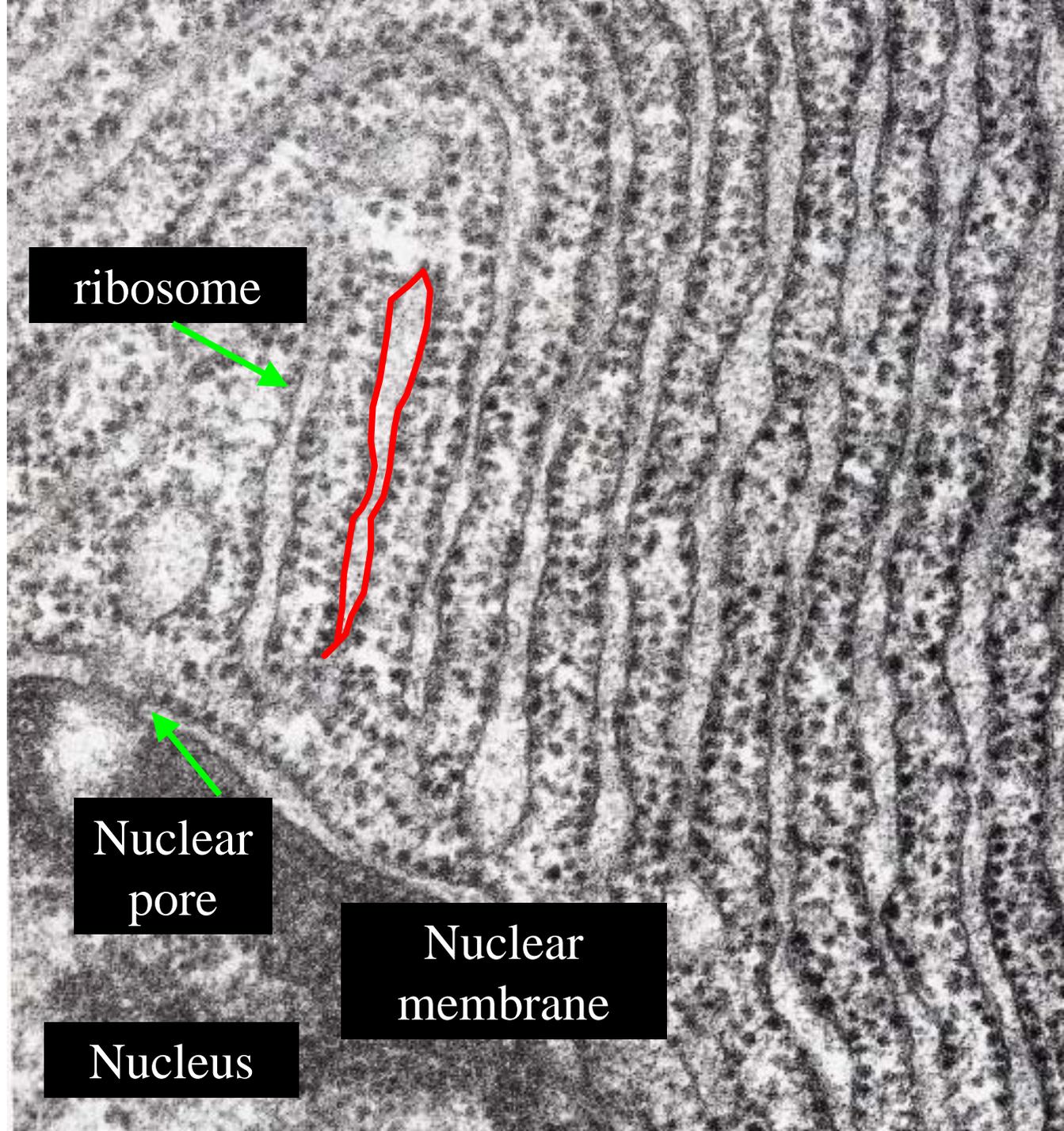


Figure 6. Model Showing the Overall Structure and Dimensions of the Full-Length GGA1 Protein

•From B.M. Collins et al., *Developmental Cell*, Vol 4, 321-332, March 2003

イヌすい臓の
外分泌細胞

B. Alberts et
al. 「Molecular
Biology of the
Cell」



ribosome

Nuclear
pore

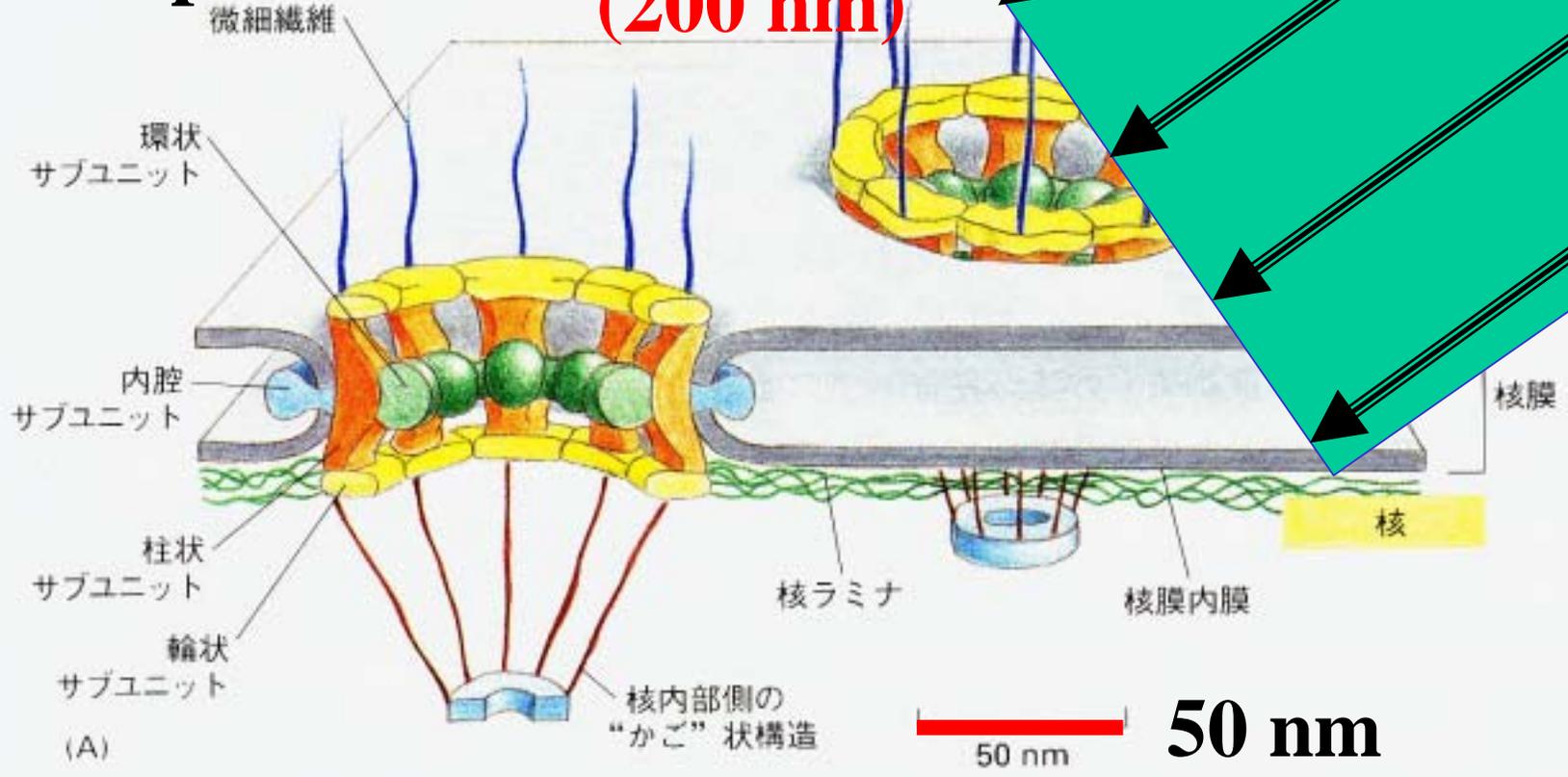
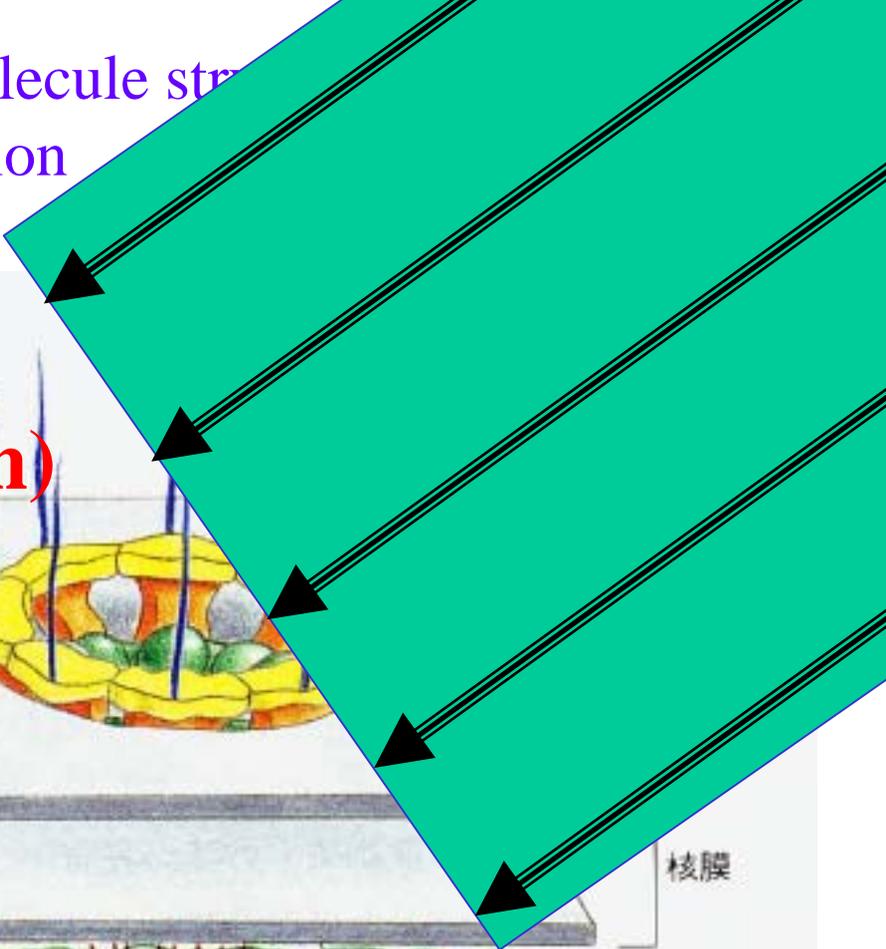
Nucleus

Nuclear
membrane

Future of structural biology: Single molecule structural resolution

Nuclear Pore Complex

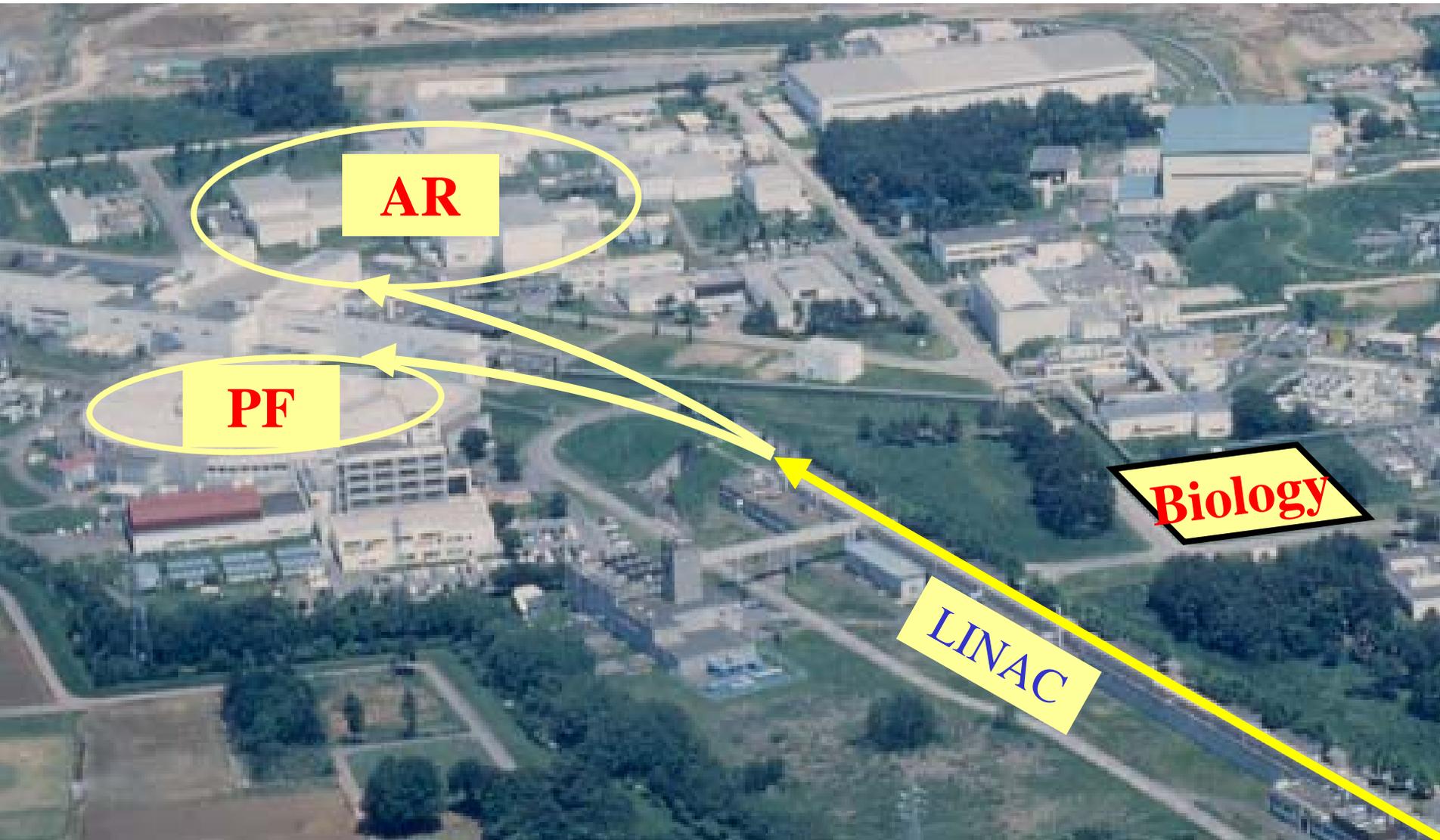
X-ray beam (200 nm)



B. Alberts et al. "Molecular Cell Biology"

PART II. Limitations of the current methods

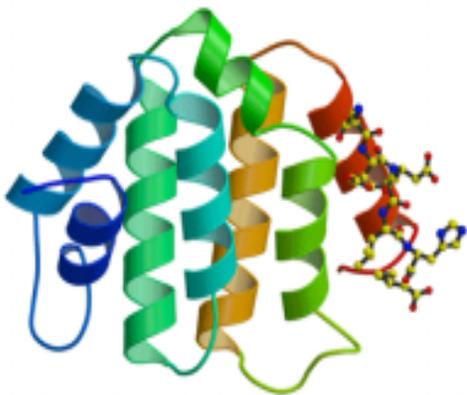
- Beam line development and high-throughput R&D



Future: automated/integrated system



Expression and purification



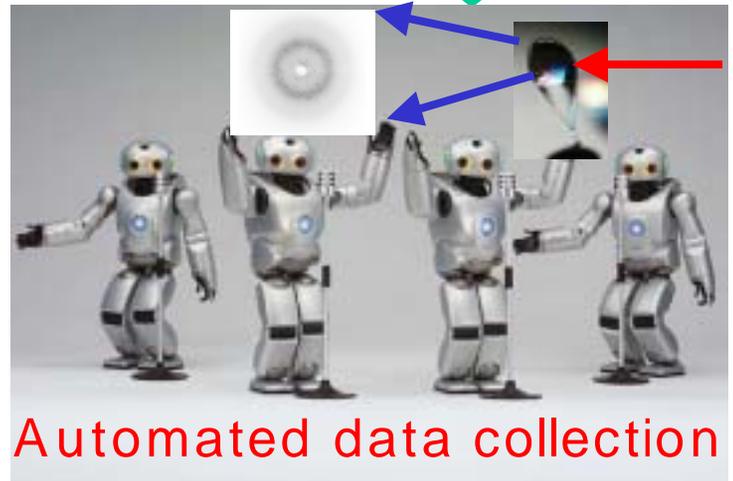
Crystallization robot

Crystallization



Crystal harvesting robot

Crystal harvesting



Automated data collection

Mouting & data collection



Data analysis

Protein Crystallization and crystal observation robot system to be completed in Autumn 2003

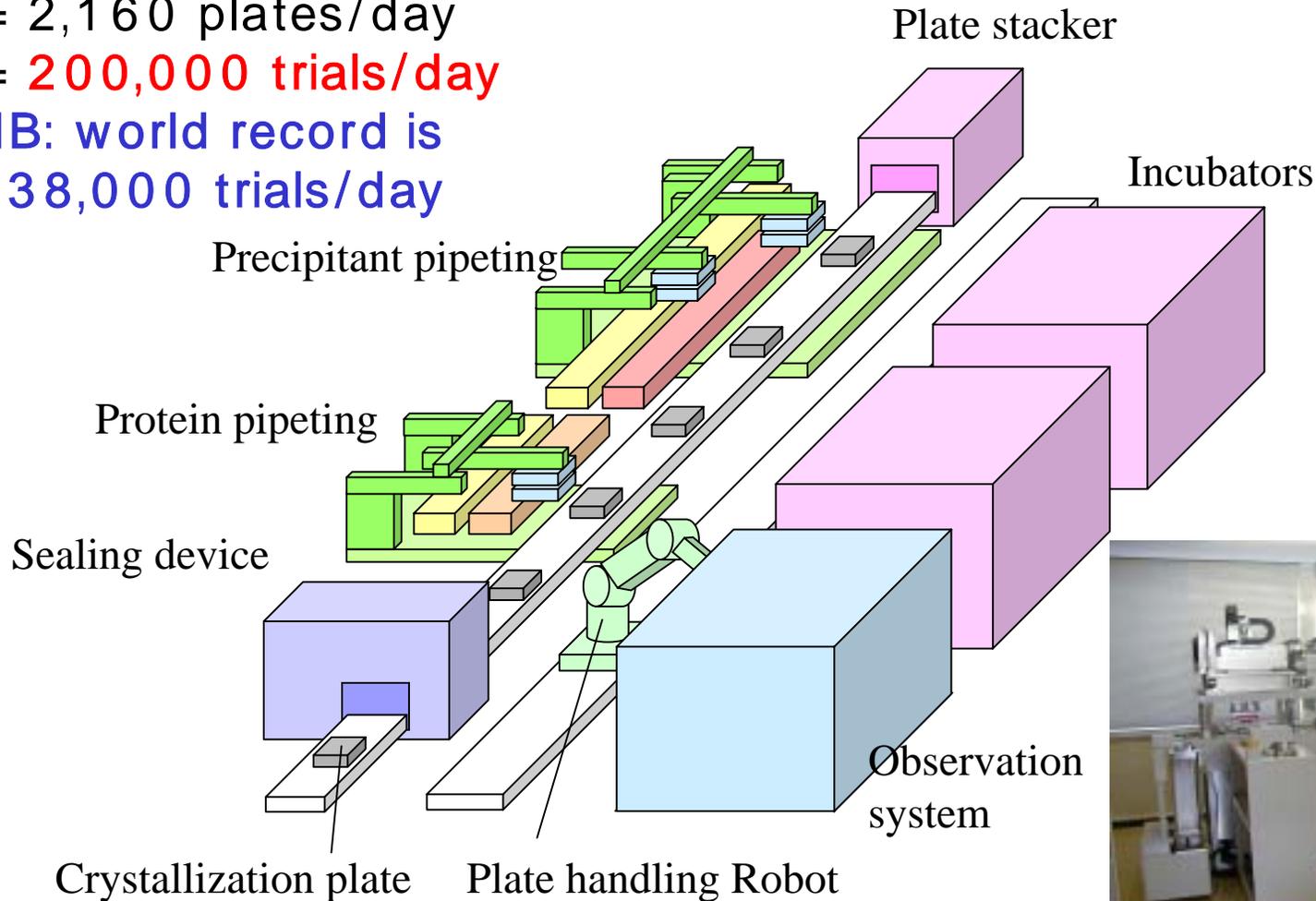
Target: 1 plate/40 sec

= 2,160 plates/day

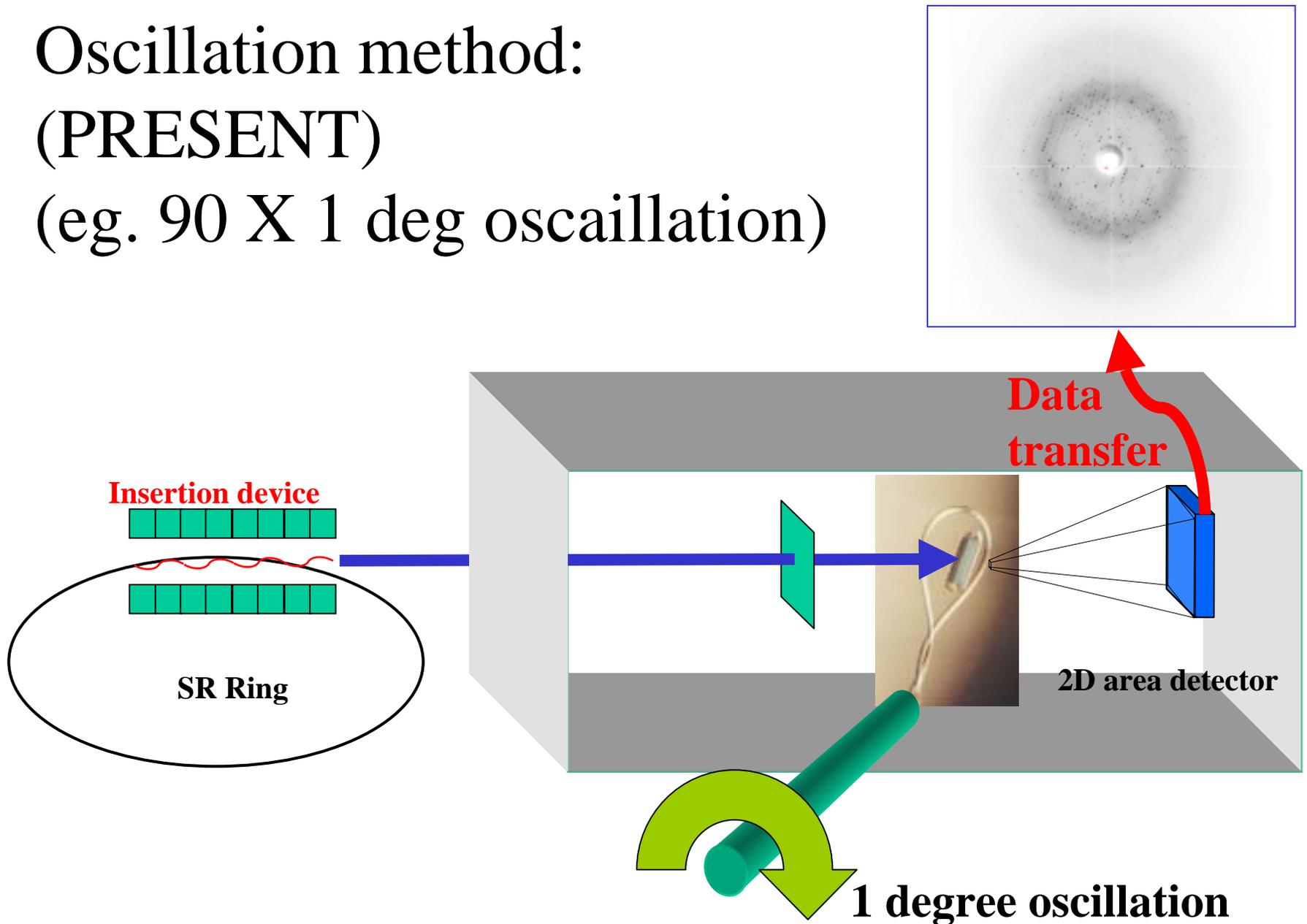
= **200,000 trials/day**

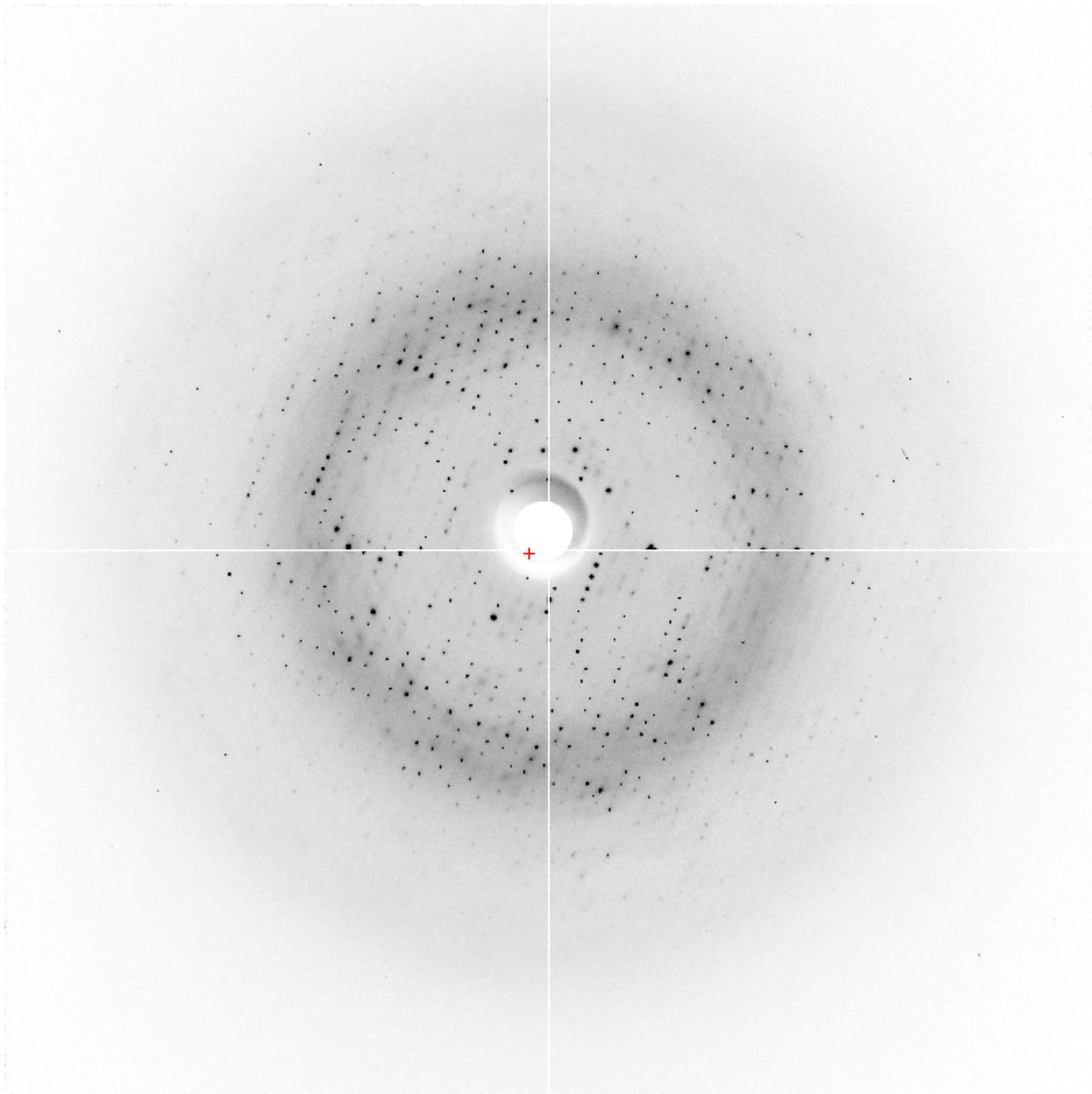
NB: world record is

138,000 trials/day



Oscillation method:
(PRESENT)
(eg. 90 X 1 deg oscillation)





Phase Determination using Multiple Anomalous Dispersion

Methionine

Sulfur \longrightarrow Selenium (SeMet)

FVTASYNVGYPAYGAKFLNNDTLLVAGGGGEGNGIIPNKLTV

LRVDPTKDTEKEQFHILSEFALEDNDDSPTAIDASKGIILVGCNENSTKITQGKGNKH

LRKFYDKVNDQLEFLTSVDFDASTNADDYTKLVYISREGTVAAIASSKVPAIMRIID

PSDLTEKFEIETRGEVKDLHFSTDGKVVAYITGSSLEVI STVTGSCIARKTDFDKNWS

LSKINFIADDTVLI AASLKKGKGI VLT KISIKSGNTSVLR SKQVTNRFGITSM DVDM

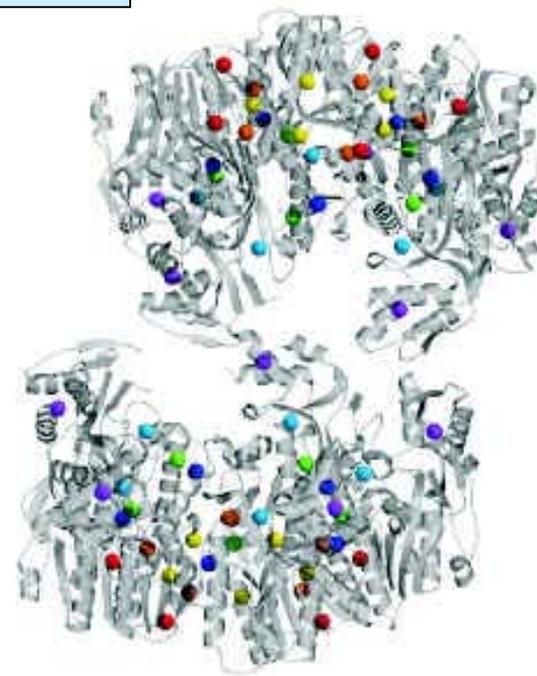
KGELAVLASNDNSIALVKLKDLSMSKIFKQAH SFAITEVTISP DSTYVASVSAANTI H

I I KLPLNYANYTSMKQKISKFFTNFILIVLLSYILQFSYKHNLHSM LFN YAKDNFLTK

RDTISSPYVVEDDLHQTTLFGNHGTKTSVPSVDSIKVHGVHETSSVNGTEVLC TESNI

INTGGAEFEITNATFREIDDA

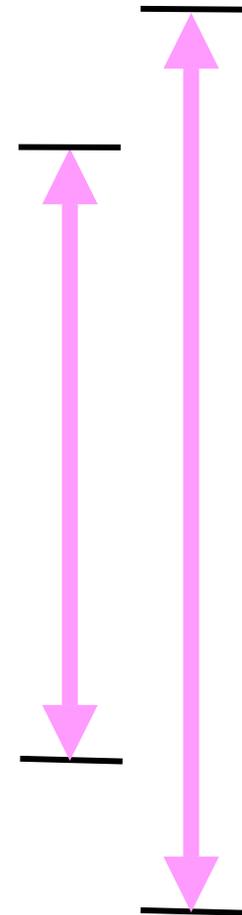
6 methionines out of 471 residues



A ribbon representation of a protein with 70 selenium atoms superimposed in colour. Equivalent selenium atoms from molecule to molecule are coloured the same.

The absorption edges of the elements frequently used in protein crystallography

Xe	0.3587Å	34.57keV
U	0.7223Å	17.16keV
Br	0.9202Å	13.48keV
Pb	0.9511Å	13.08keV
Se	0.9795Å	12.66keV
Hg	1.0093Å	12.29keV
Au	1.0402Å	11.92keV
Pt	1.0722Å	11.57keV
Zn	1.2837Å	9.66keV
Cu	1.3808Å	8.98keV
Sm	1.6625Å	7.46keV
Fe	1.7433Å	7.11keV



Phase determination using MAD

(Multiple Anomalous Dispersion)

Bijvoet pairs

$$|F^+| \equiv |F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}| \quad |F^-| \equiv |F_{\bar{h}kl}| = |F_{h\bar{k}\bar{l}}|$$

Bijvoet Differences

$$\Delta F = |F^+| - |F^-|$$

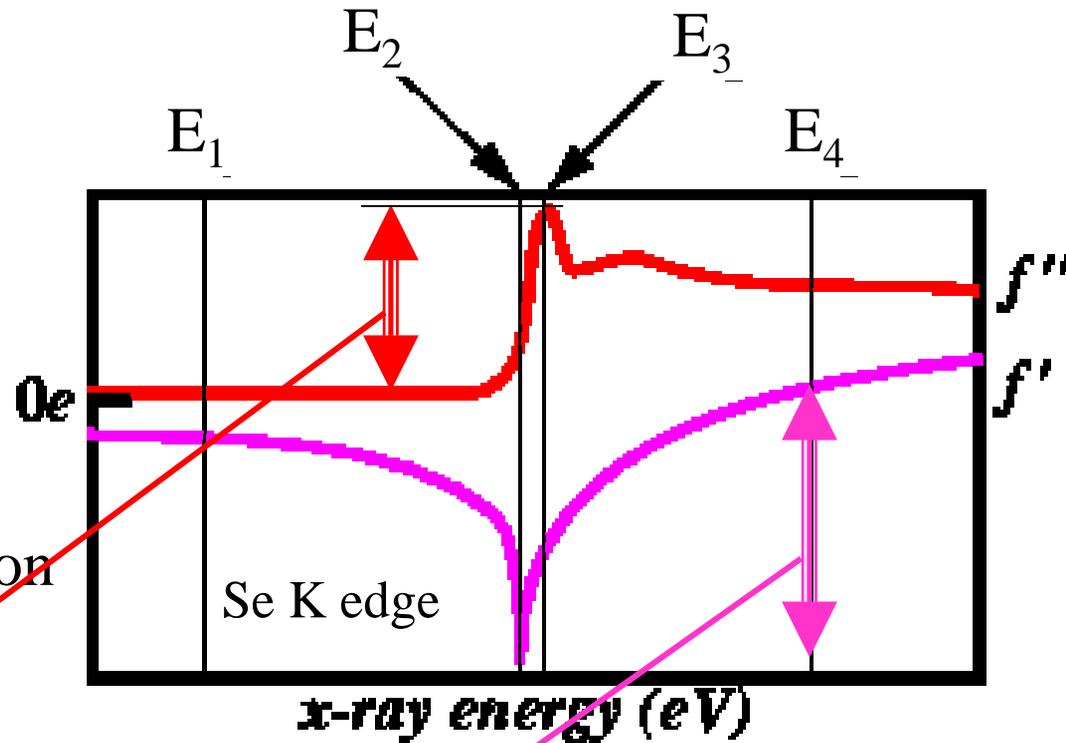
MAD experimental maps:

Anomalous difference Patterson

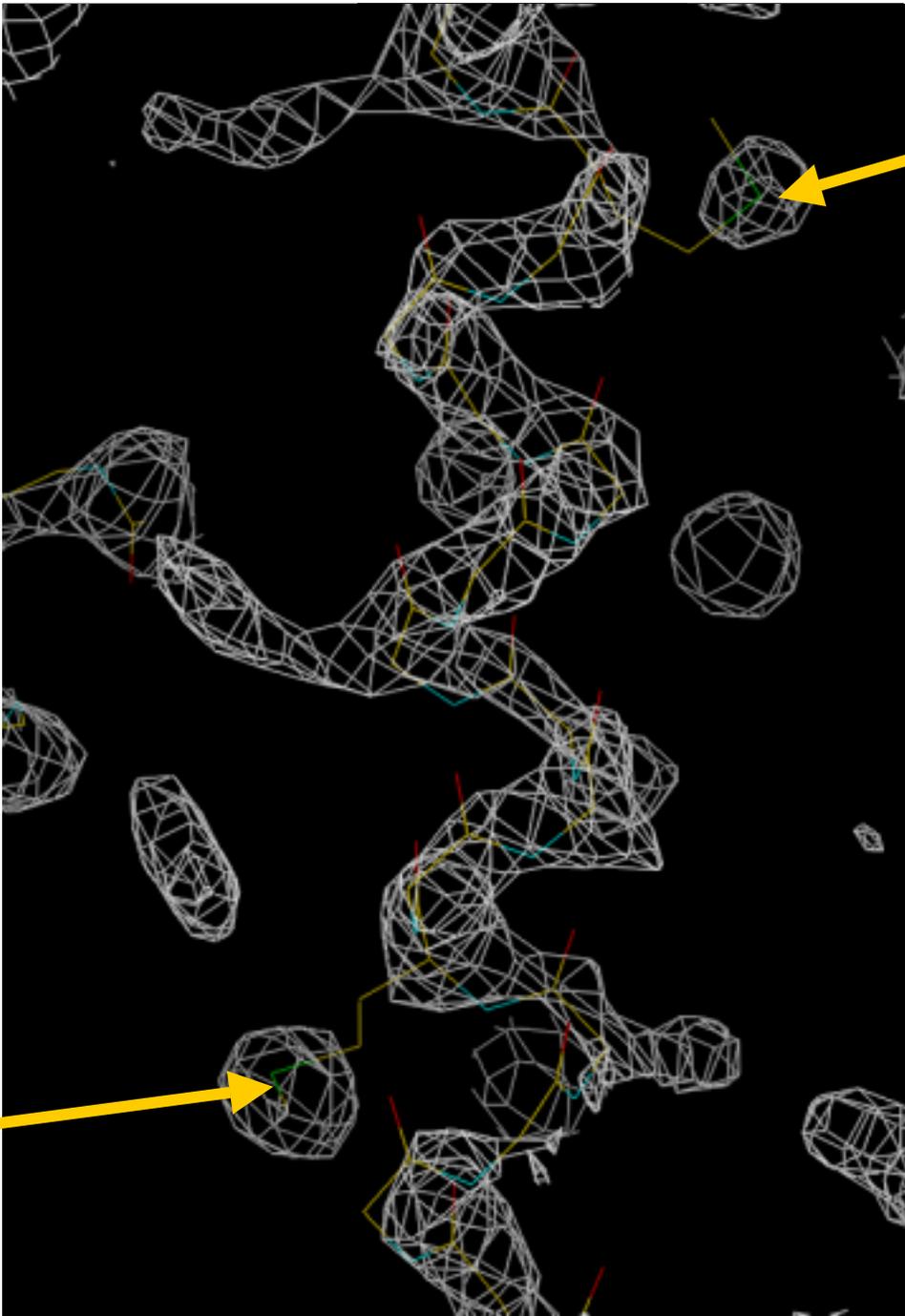
$$\Delta F_{E_3}^2 \text{ (at the peak)}$$

Dispersive Patterson

$$F_{E_4}^2 - F_{E_2}^2 \text{ (remote - inflection point)}$$

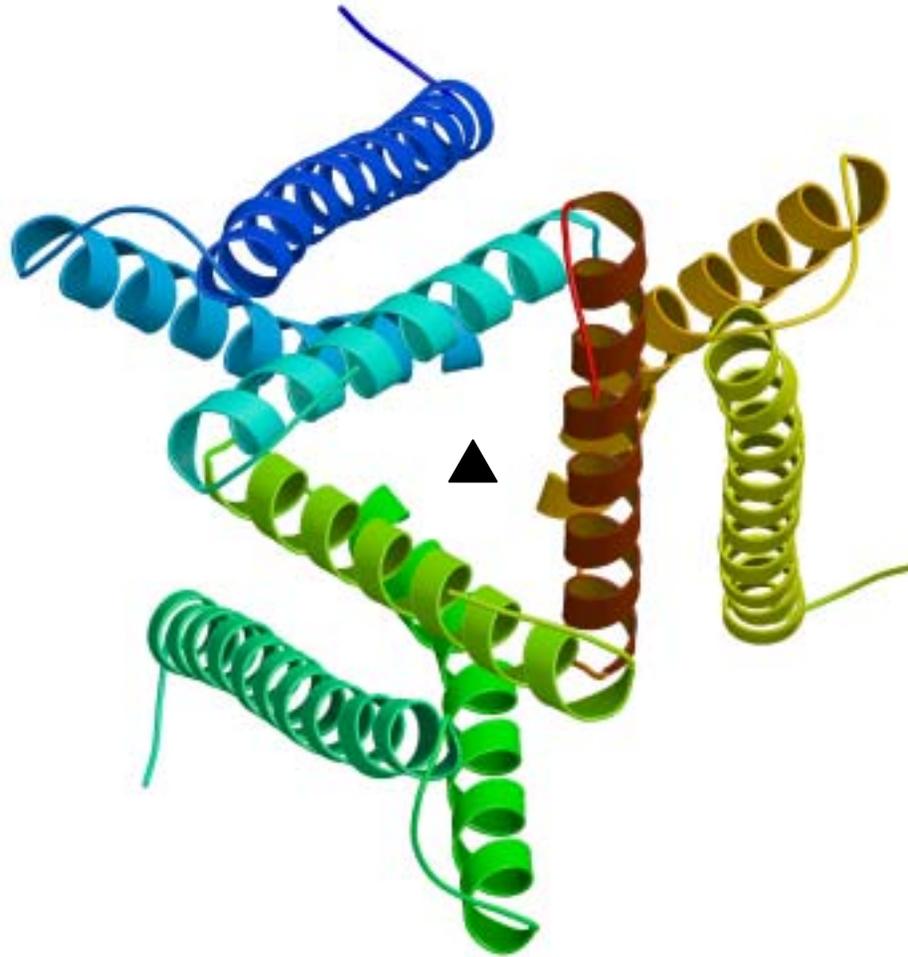


$$f(\lambda) = {}^0f + f'(\lambda) + if''(\lambda)$$



Met94

Met84



Ribbon diagram of trimer of GGA1 GAT domain

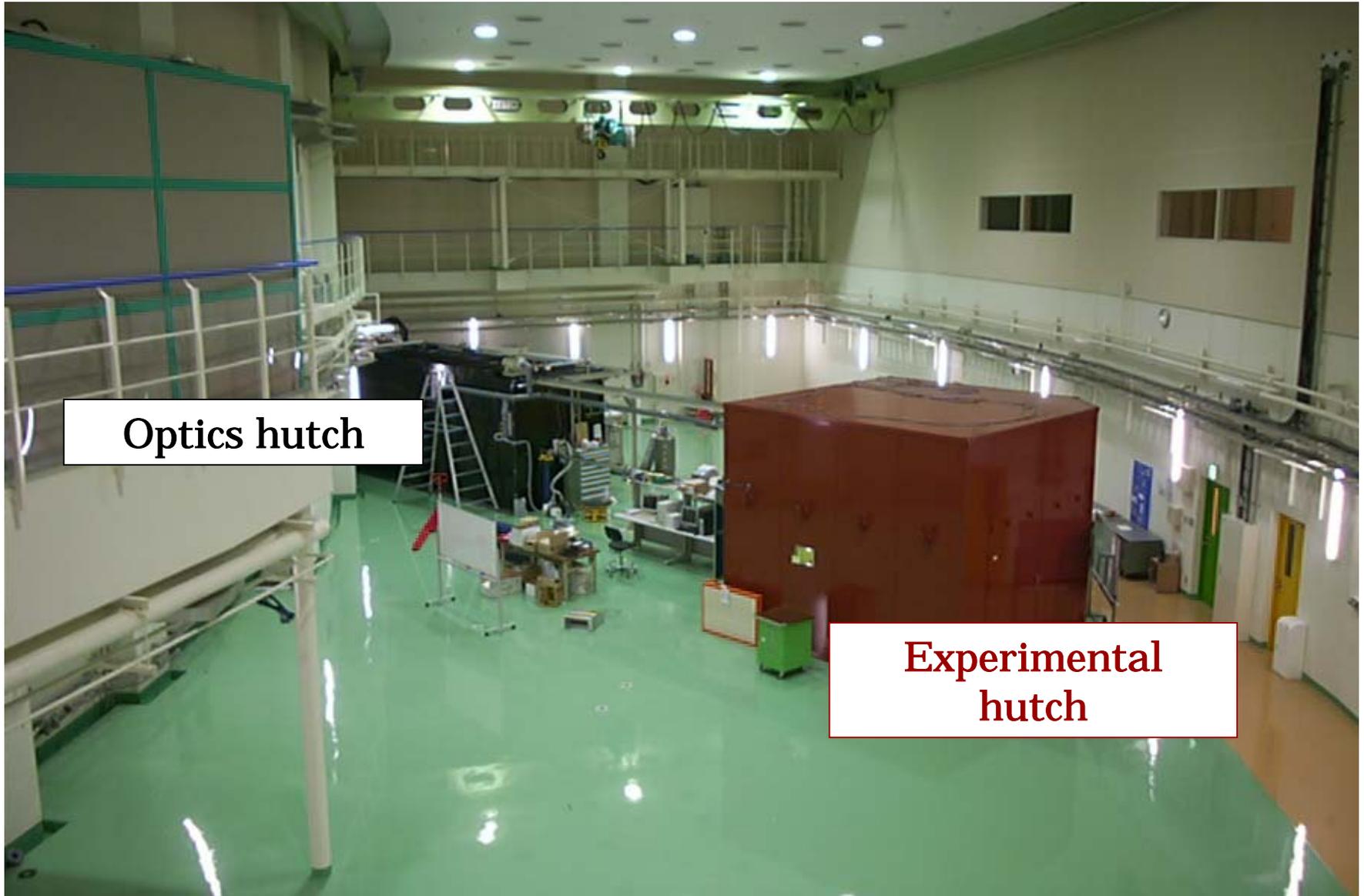
Triangle is threefold axis.



One of the fastest MAD beam lines in the world



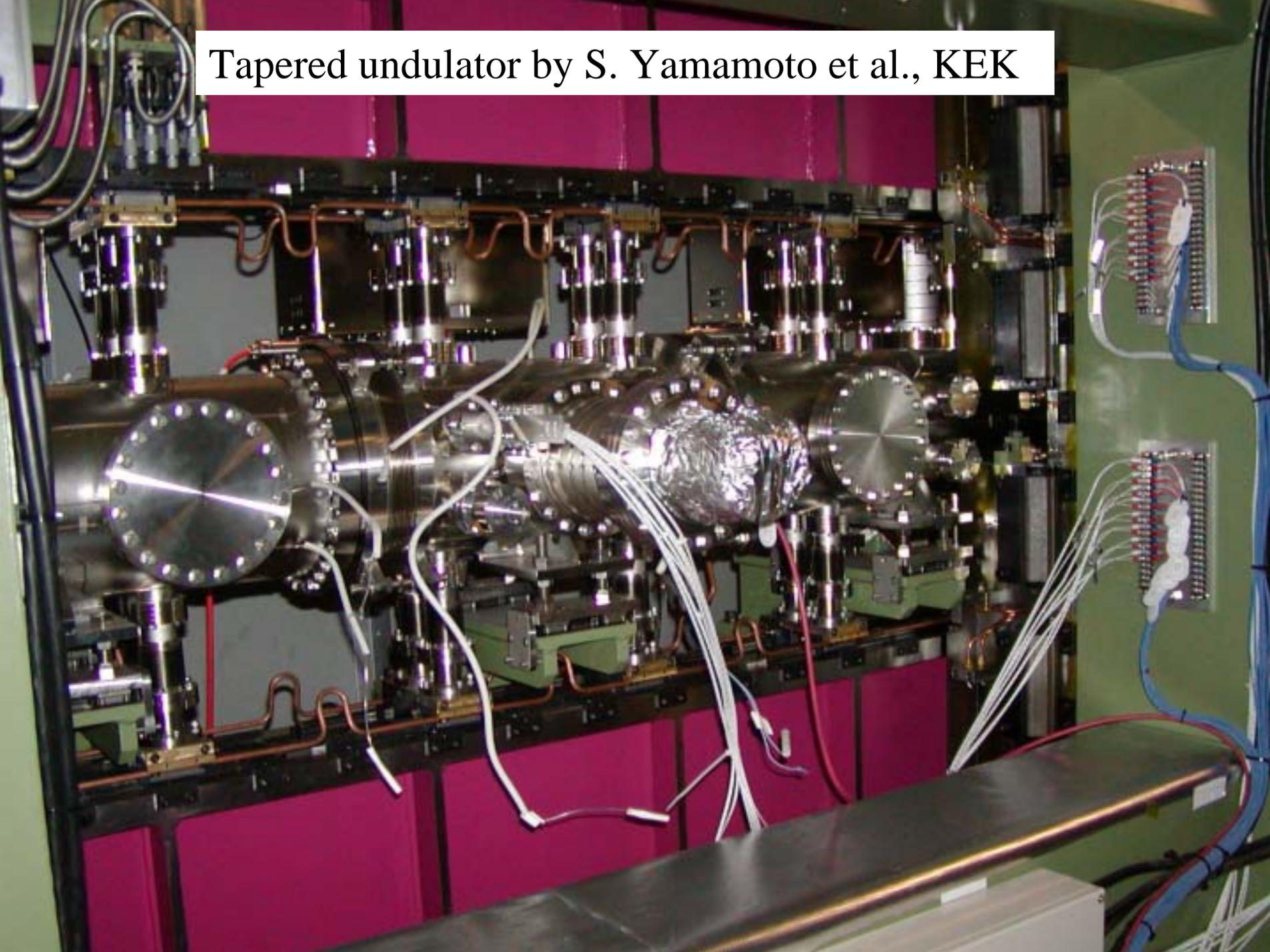
PF-AR NW12



Optics hutch

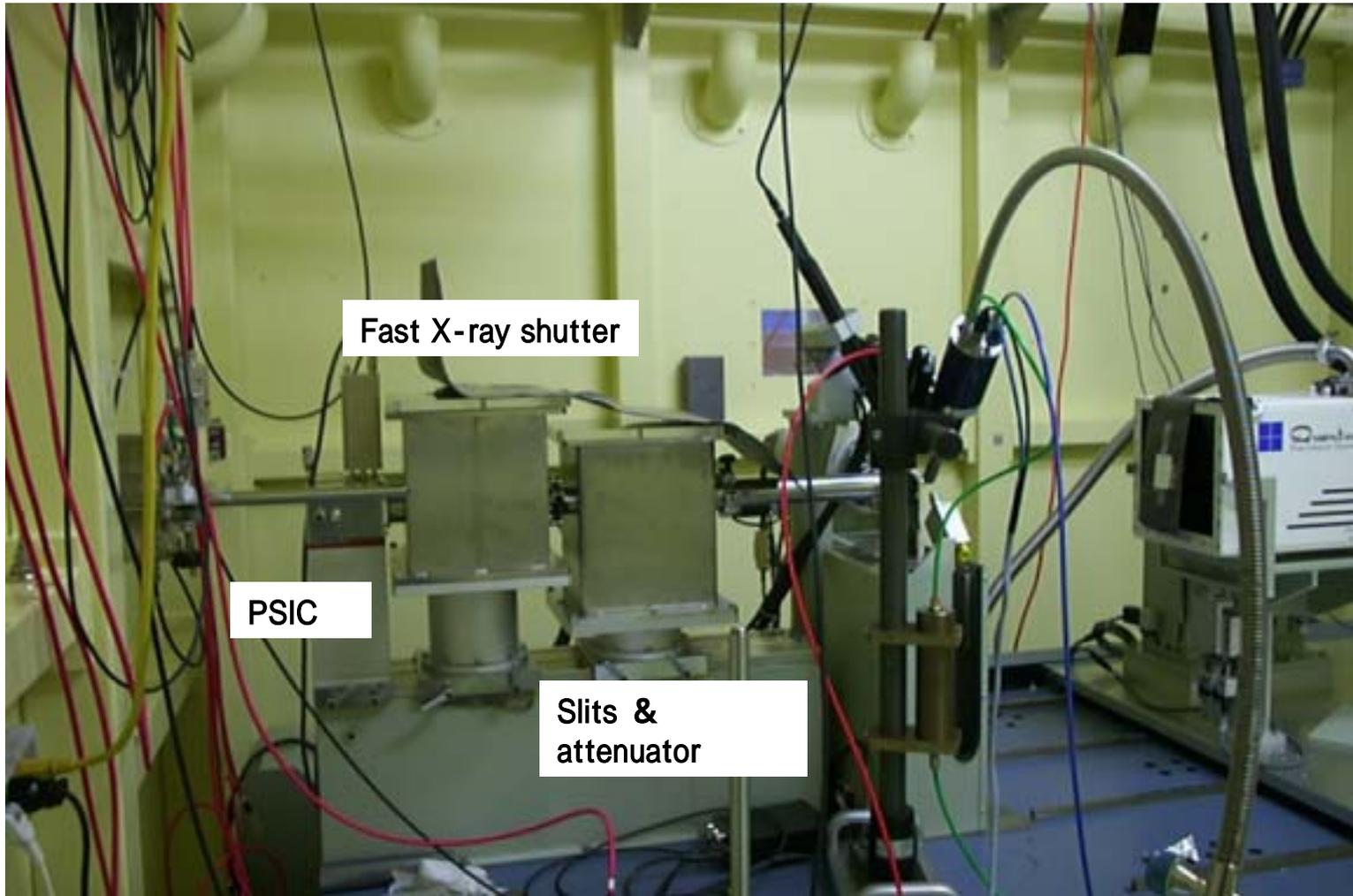
Experimental hutch

Tapered undulator by S. Yamamoto et al., KEK



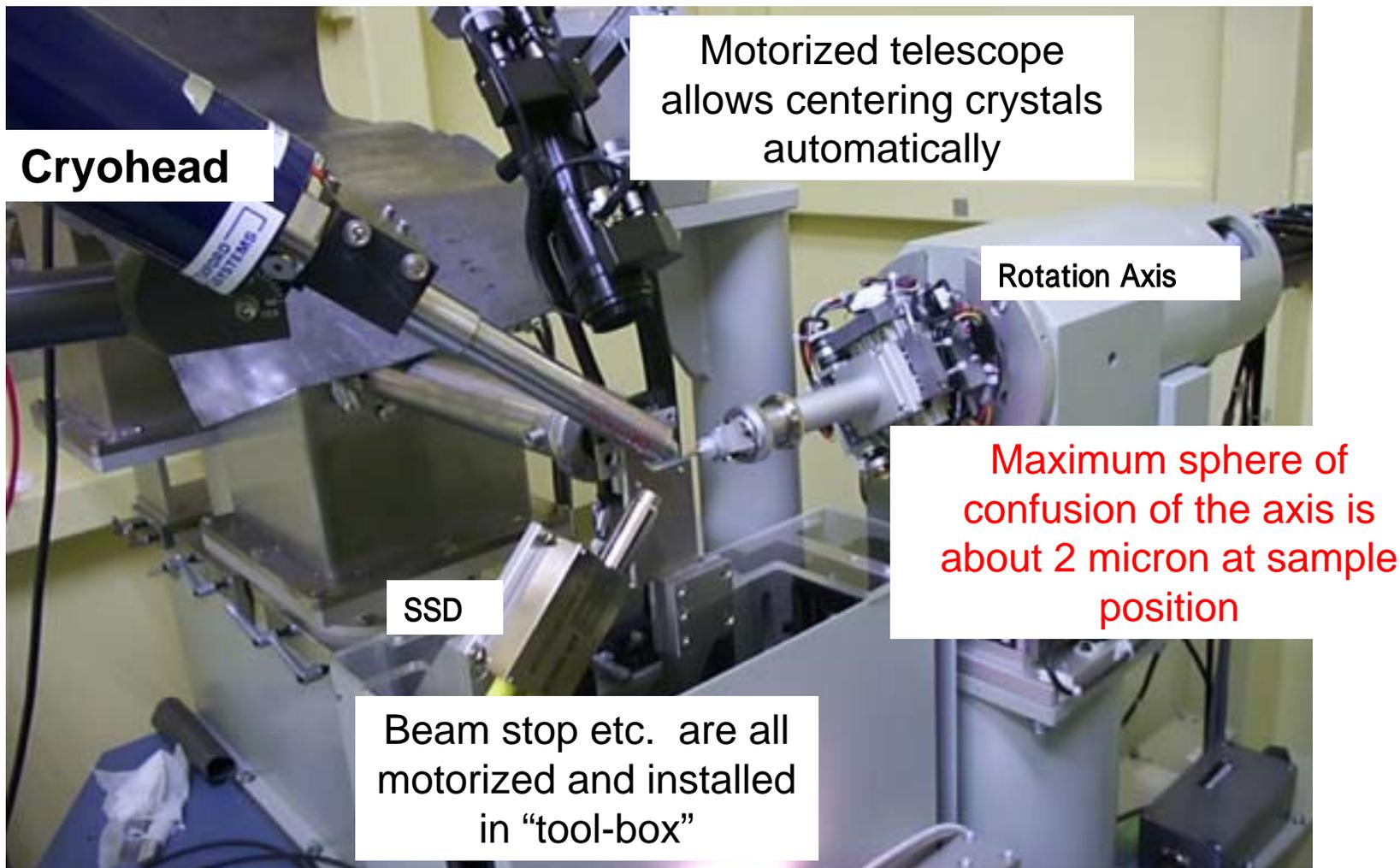
NW12 is one of the fastest MAD beam lines in the world.

Total data collection time (min): 10 to 30 min



NW12 allows for data collection from very small crystals.

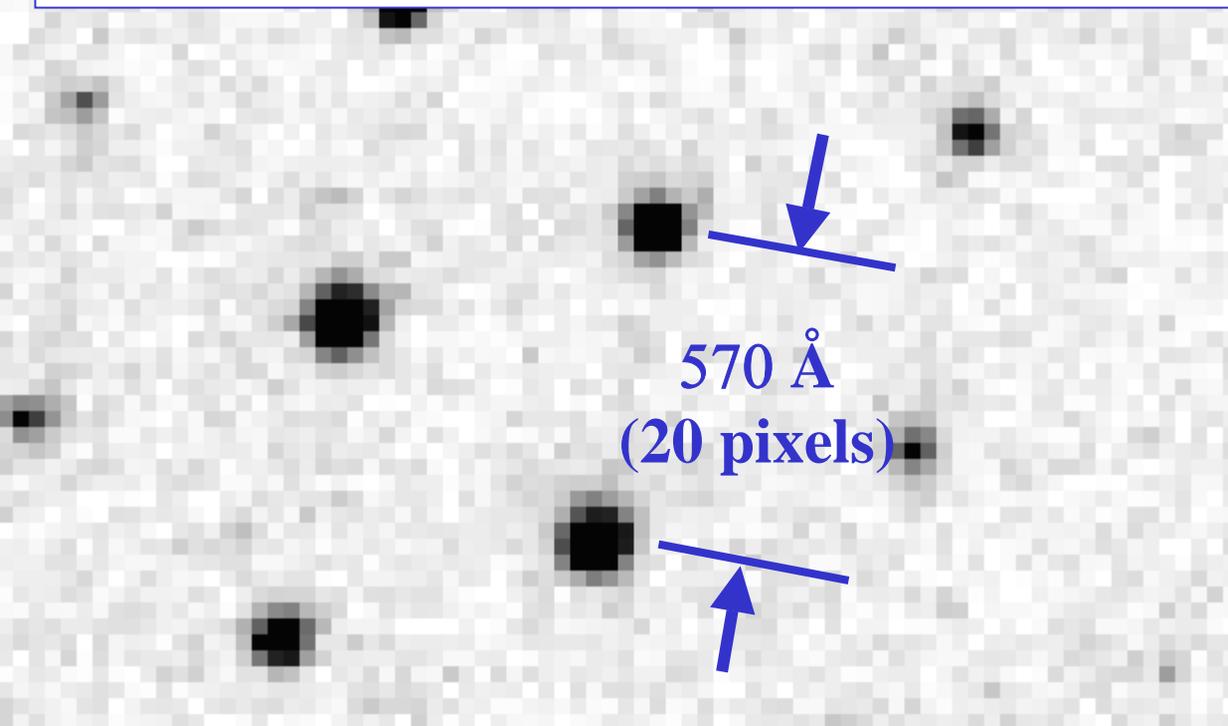
Tool box makes manual crystal mounting easy.



Limits of third generation synchrotron sources

**(1) The largest unit cell
dimension : 2000 Å**

(2) Smallest crystals: a few μm^3



Blue Tongue Virus

Core Particle 1

P2₁2₁2

755 X 796 X 825 Å³

ID14/EH3

Crystal to IP: 1250 mm

Wavelength 0.918Å

Oscillation 0.1 deg

Exposure 100 sec

Pixel size 100 μm

Beam size 100 μm

FWHM 181 μm

Part III

Key issues in realizing single particle/nano-crystal structural biology

Averaging

in crystal

crystallography

in computer

conventional

single molecule analysis

Expectation from the 4th generation synchrotron

Future of structural biology : **Single molecule or nanocrystals**
structural analysis at atomic resolution

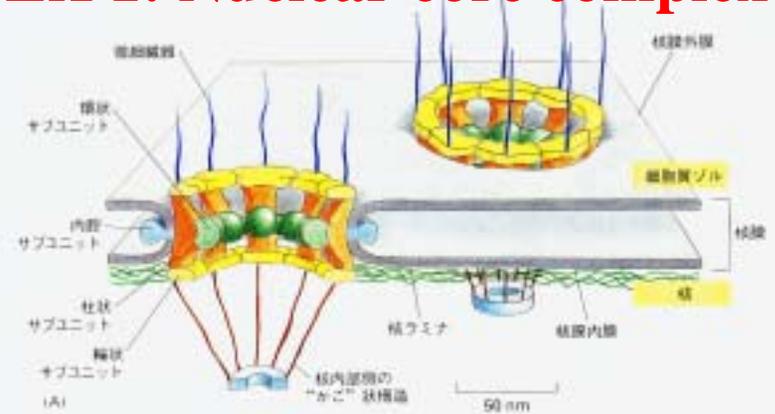
Properties

Low cost operation of multiple beam lines owing to the energy recovery

Brilliance: 1000 to 10000 times of the 3rd generation synchrotrons (ESRF, APS, Spring-8)

Very short pulse length : 100 femto seconds (1/100)

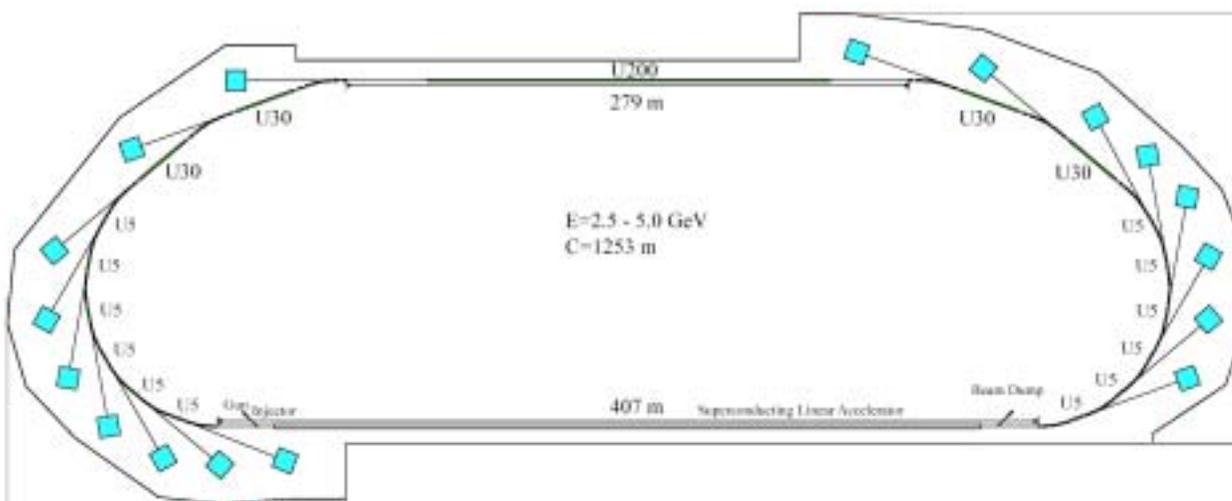
EX 1: Nuclear core complex



EX 2: Nano crystals 10 to 100 nm size



*Weak signal and
many more images
to collect*



Critical conditions for successful single molecule structural analyses

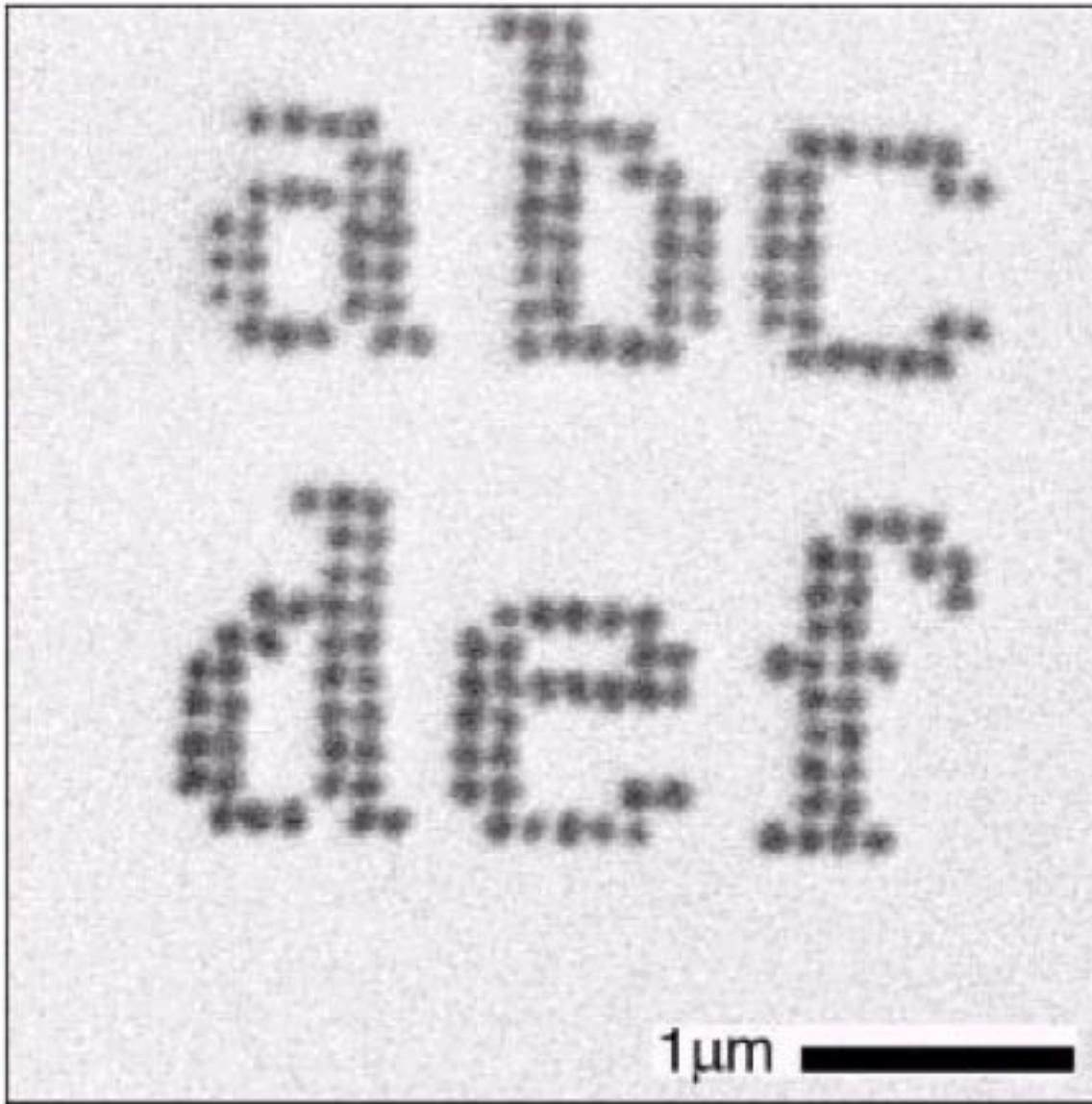
- Determination of orientation of single molecules
- Phase determination using the oversampling method
- Radiation damage
- Sample manipulation
- Sufficient S/N ratio from averaging single molecules

Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens

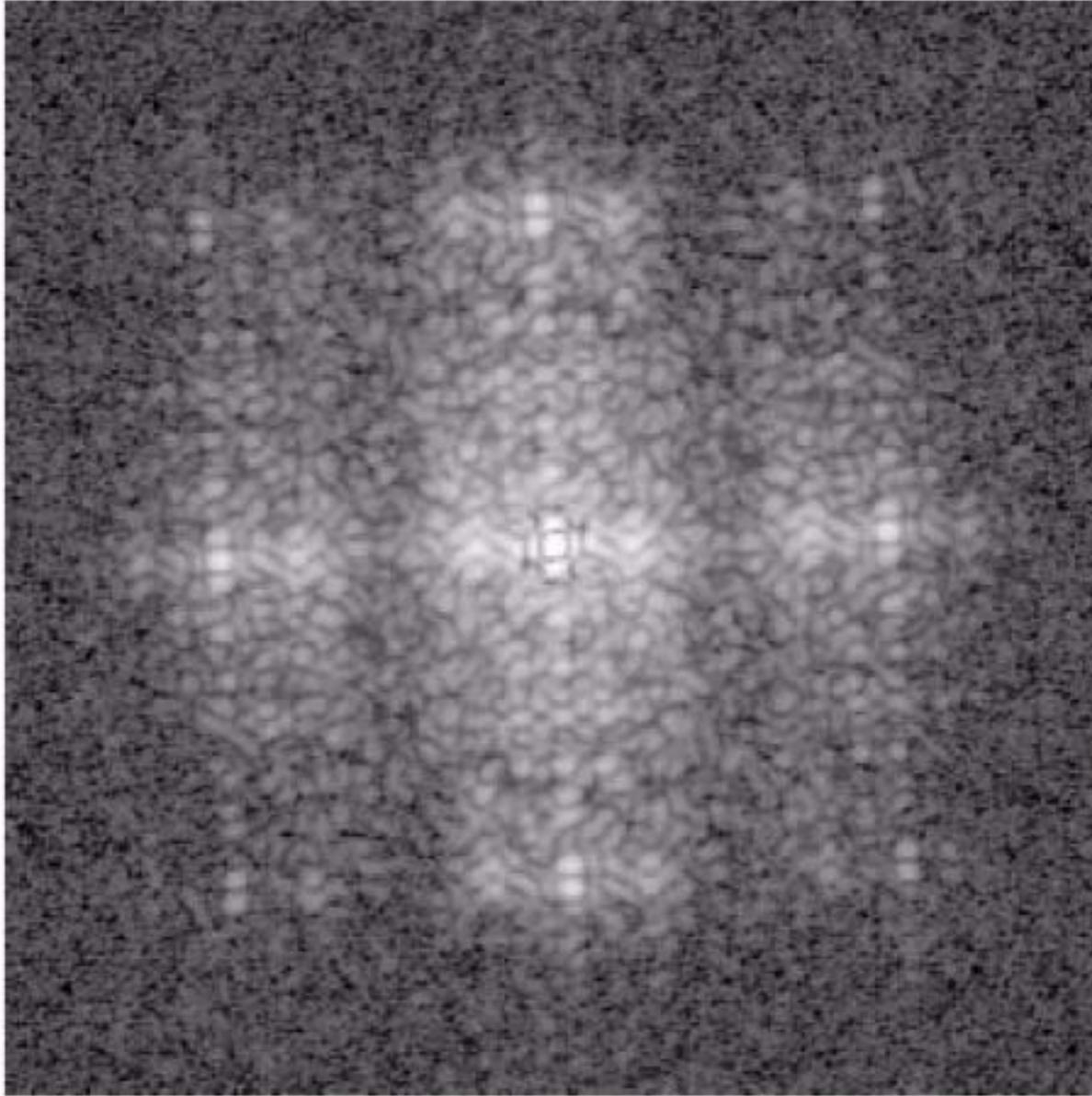
**Jianwei Miao^{*}, Pambos Charalambous[†], Janos Kirz^{*}
& David Sayre^{*‡}**

^{} Department of Physics and Astronomy, State University of New York, Stony Brook, New York 11794-3800, USA*

[†] Kings College, Strand, London WC2R 2LS, UK



A scanning electron microscope image of the specimen. The specimen was fabricated by depositing gold dots, each ~ 100 nm in diameter and 80 nm thick, on a silicon nitride membrane.

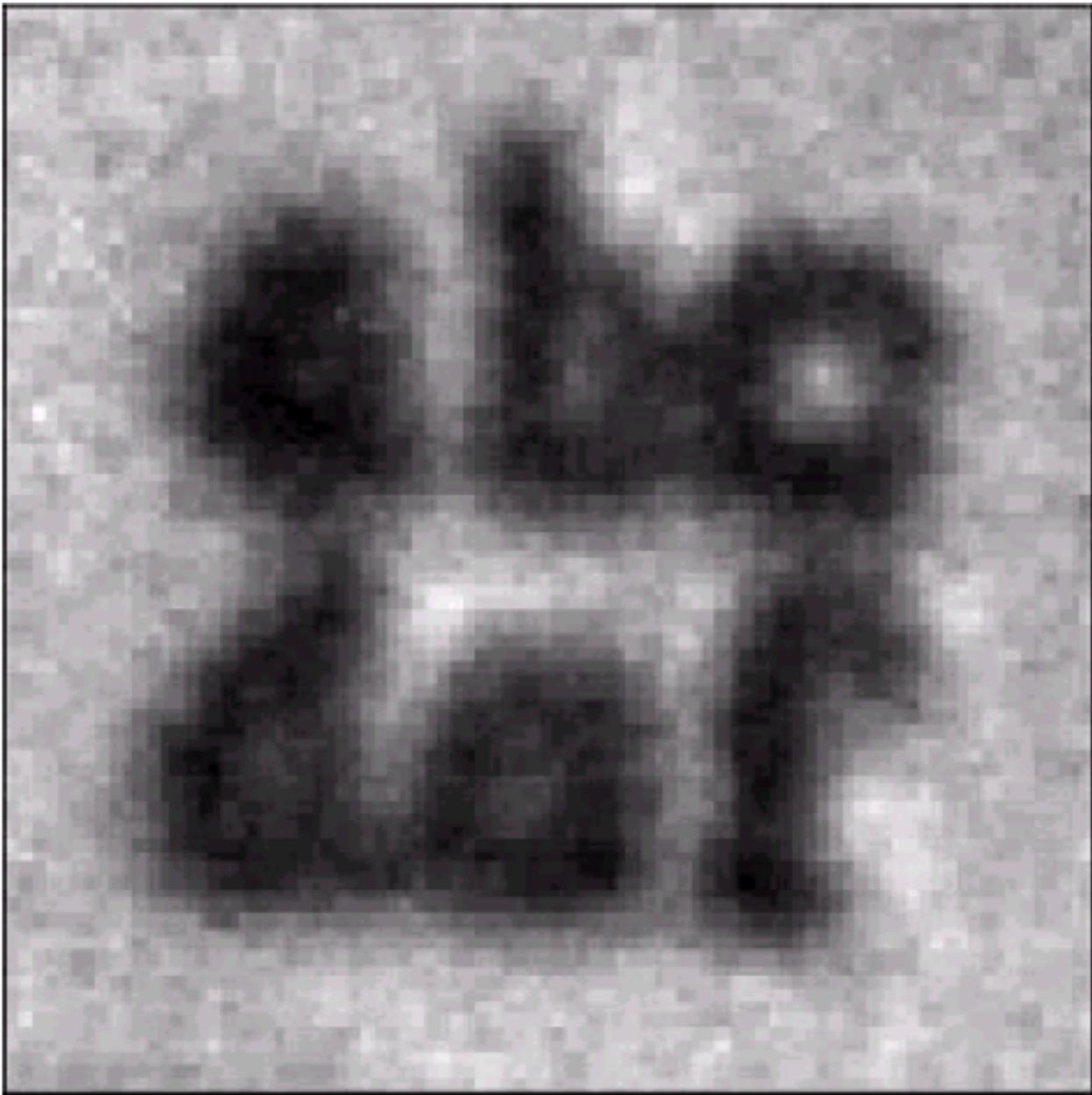


$\lambda=1.7$ nm, X1A NSLS

CCD: 512 x 512

(24 μ m pixel)

A diffraction pattern of the specimen (logarithmic intensity scale). The central 15-pixel-radius circular area is supplied by the squared magnitude of the Fourier transform of the optical microscope image.



An optical microscope image of the specimen.

Structure
amplitudes



$$F(\mathbf{k}) = \sum_{\mathbf{x}=0}^{N-1} f(\mathbf{x}) \exp(2\pi i \mathbf{k} \cdot \mathbf{x} / N) \quad (1)$$

Electron density



$$|F(\mathbf{k})| = \left| \sum_{\mathbf{x}=0}^{N-1} f(\mathbf{x}) \exp(2\pi i \mathbf{k} \cdot \mathbf{x} / N) \right| \quad (2)$$

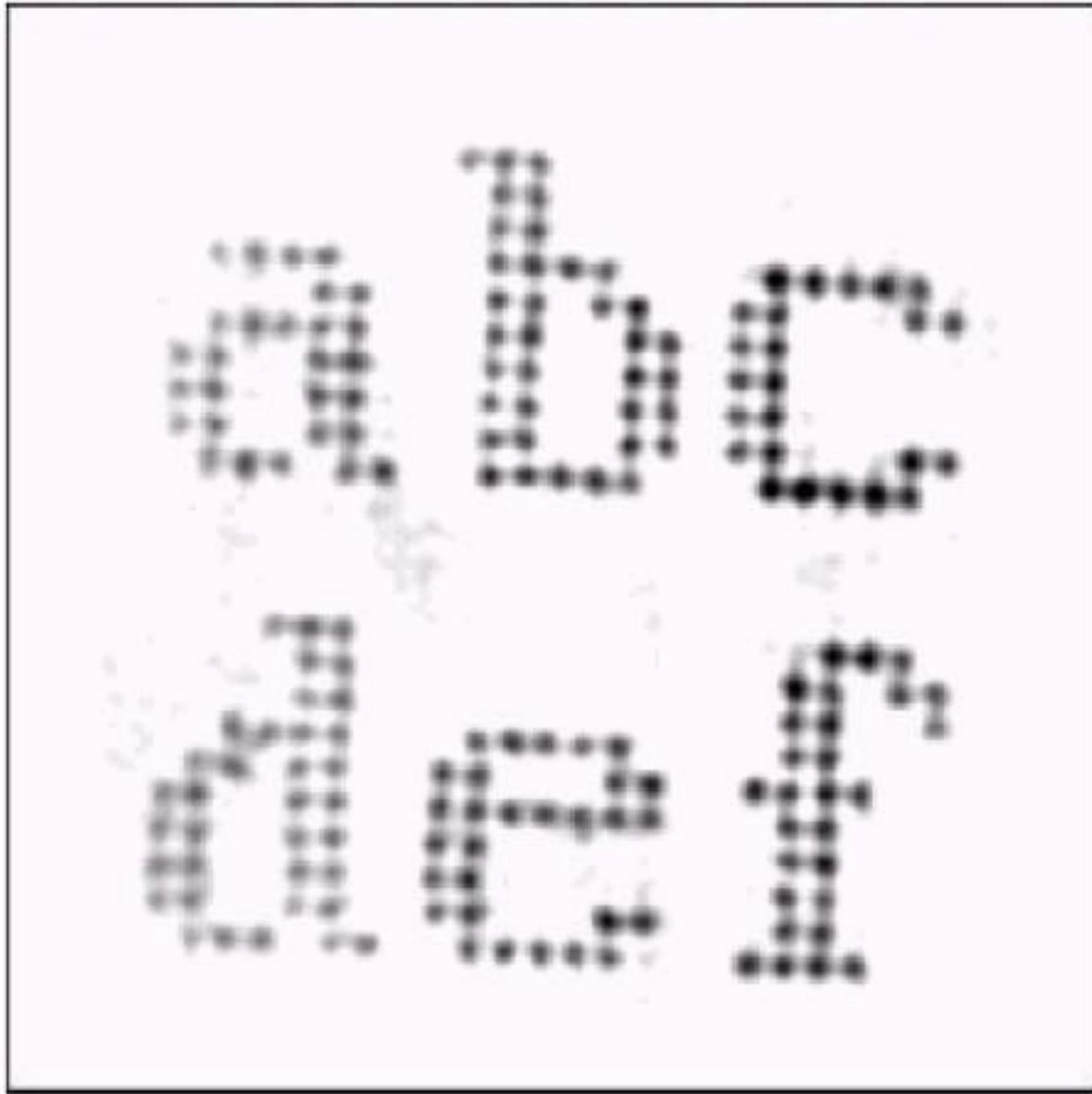
Eq. (2)	No of equations (A)	No of unknown variables (B)	(B) / (A) underdeter mined
f(x): real	$N^3/2$	N^3	2

Determination of Phases

- Given the magnitude of a Fourier transform sampled at the Bragg density, the phase problem is underdetermined by a factor 2.
- Thus at least in principle, oversampling the magnitude of a Fourier transform by factor of $2^{1/3}$ (=1.26) in each dimension is necessary to retrieve the phase of a 3D object

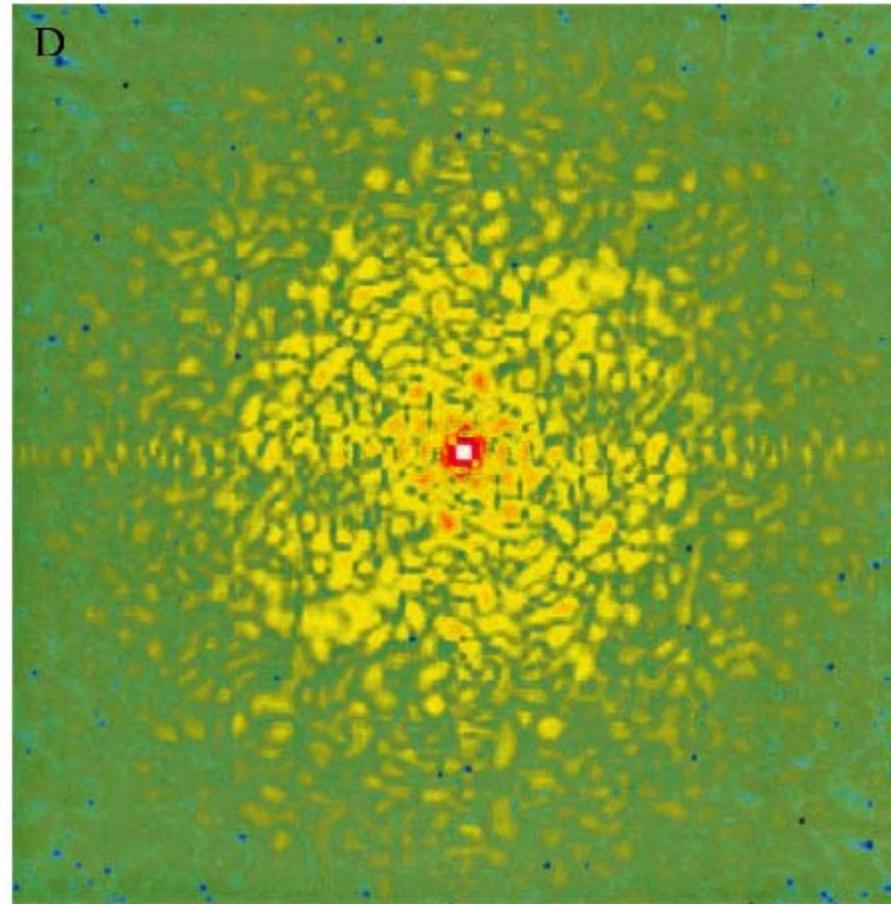
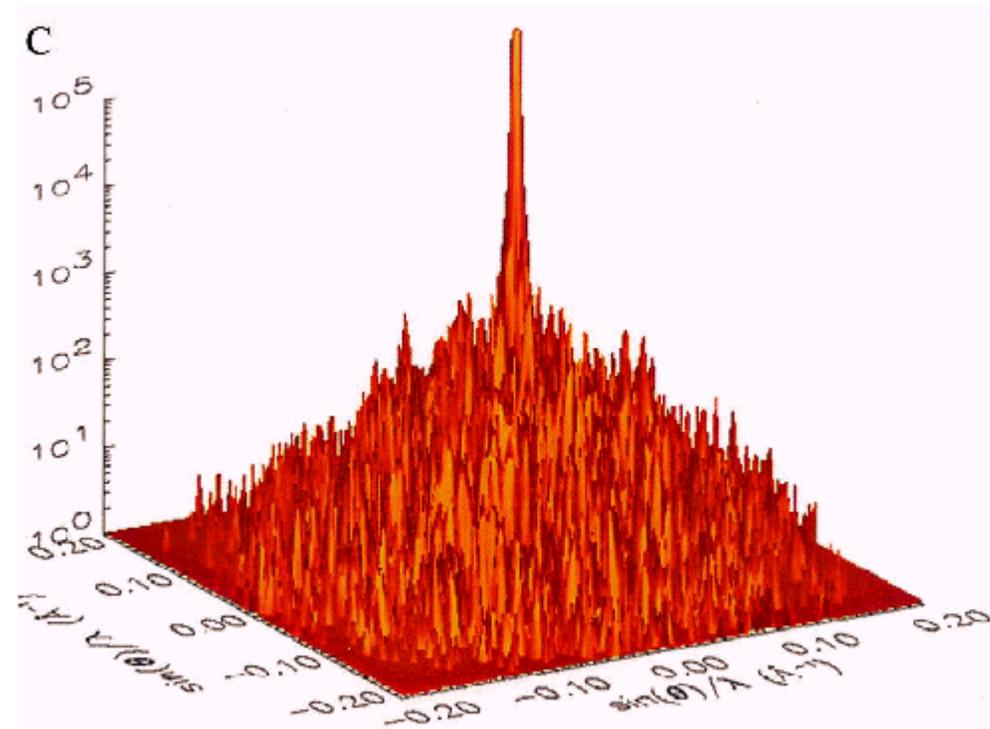
Phase Problem

- 1) Decrease the number of unknown-valued pixels based on the knowledge of the object – finite support; **density outside the protein is zero** (Protein Crystallography → Density Modification)
- 2) **Oversampling by factor 2**

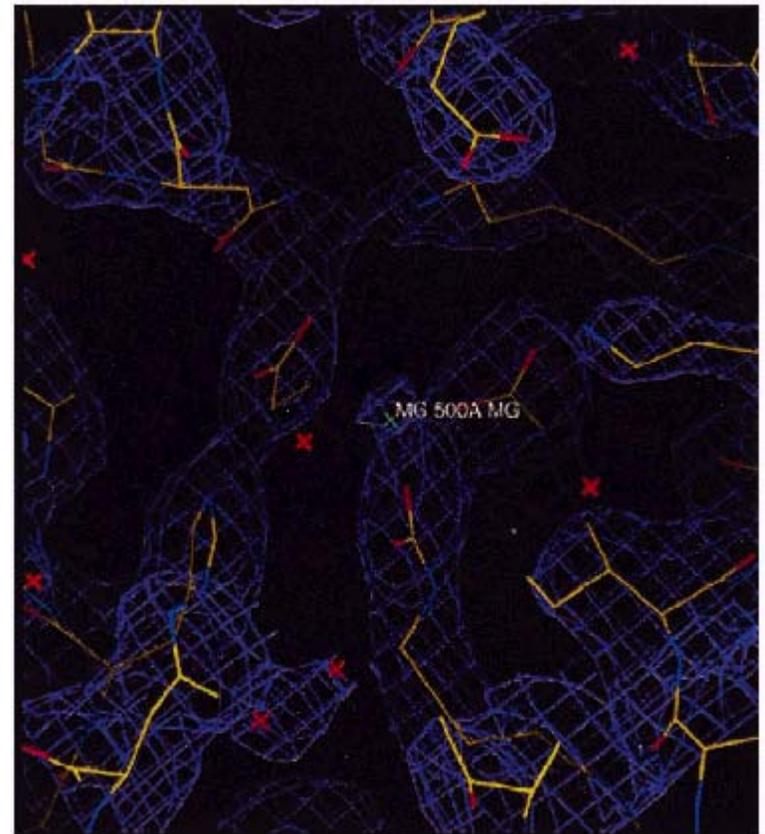
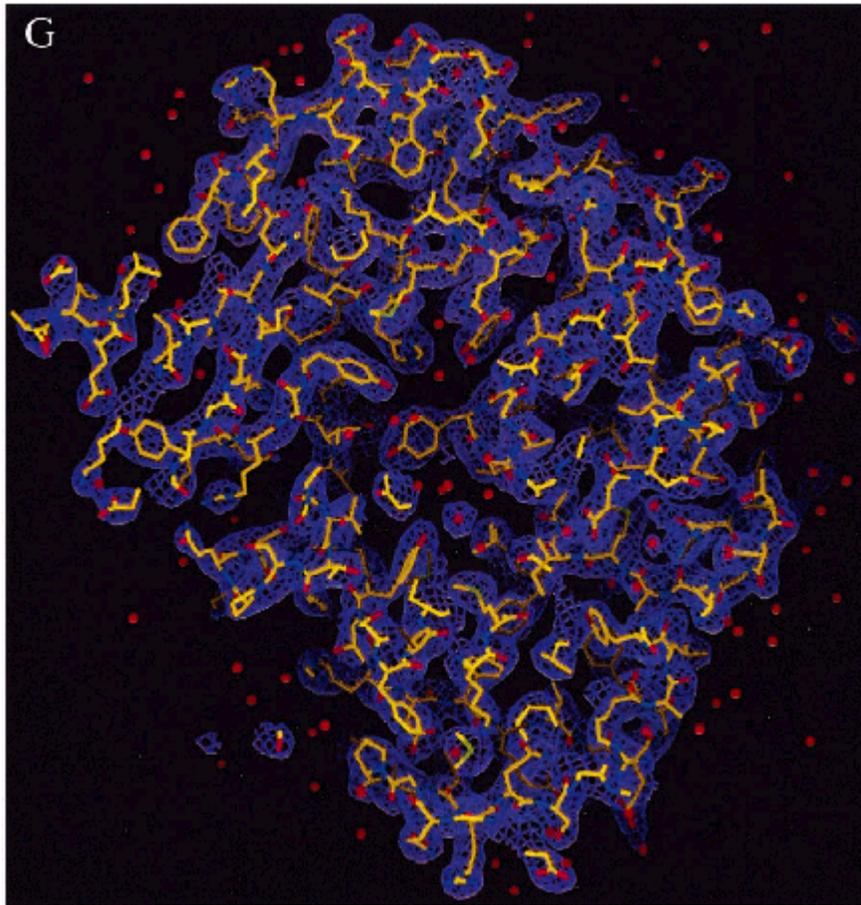


The specimen image as reconstructed from the diffraction pattern

3D structural determination of single rubisco molecules utilizing a simulated X-FEL and direct phase retrieval by the oversampling technique. Miao, Hodgson, Sayer, PNAS, June 5, 2001, vol. 98, 6641-645



3D structural determination of single rubisco molecules utilizing a simulated X-FEL and direct phase retrieval by the oversampling technique. Miao , Hodgson , Sayer , PNAS , June 5, 2001, vol. 98, 6641 · 645



Simulated diffraction images

X-ray energy:	12 keV
Integrated X-ray intensity	3×10^{12} (3.8×10^6 per \AA^2)
Detector	100 mm by 100mm (128 by 128 pixel)
Sample to detector distance:	100 mm
Resolution limit at the rim of the detector:	2.2 Angstroms

NO BACKGROUND WAS TAKEN INTO ACCOUNT although they are aware of sources of various errors.

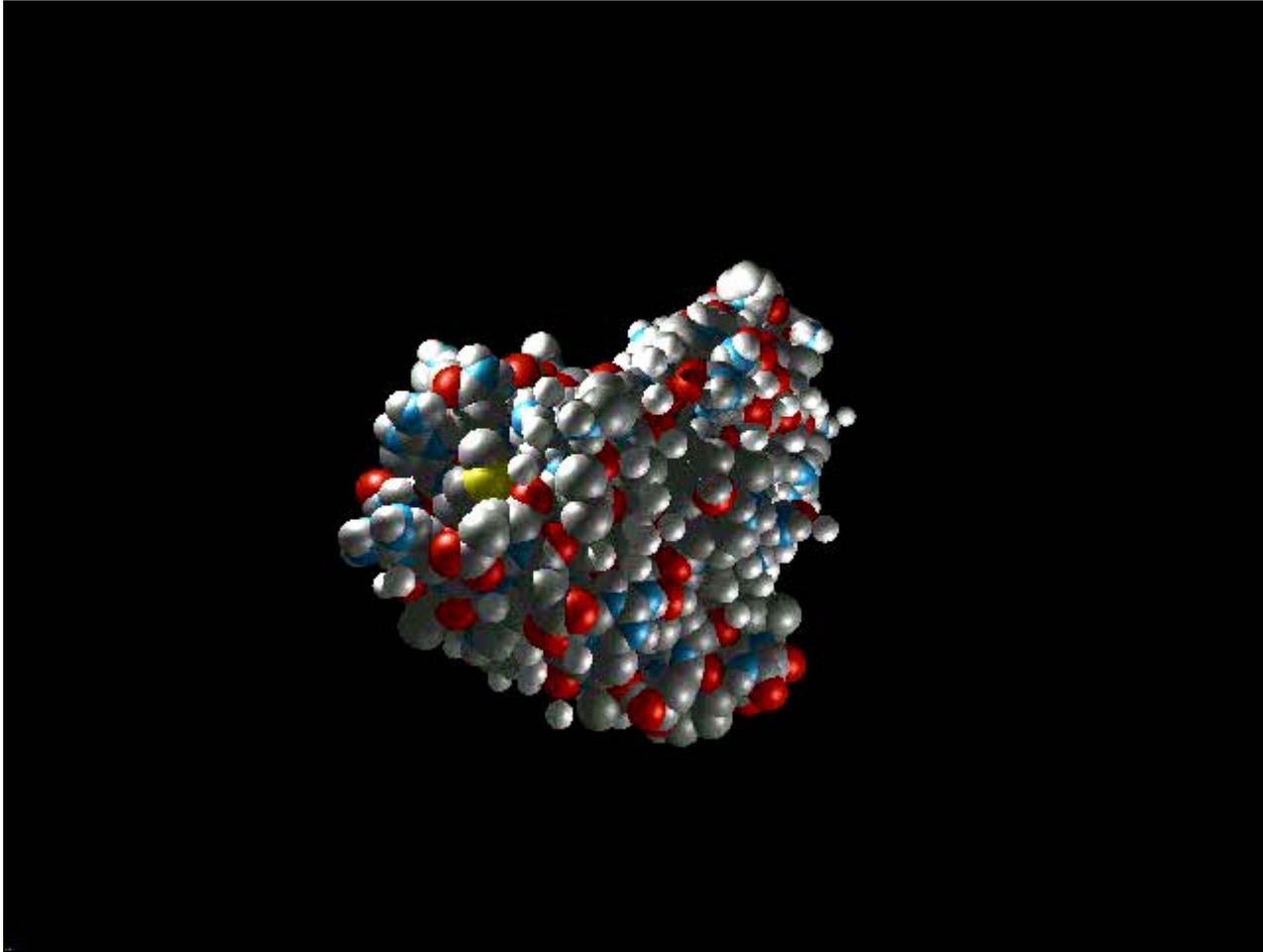
Potential for biomolecular imaging with femtosecond X-ray pulses

**Richard Neutze^{*}, Remco Wouts^{*}, David van der Spoel^{*}, Edgar Weckert^{†‡}
& Janos Hajdu^{*}**

^{} Department of Biochemistry, Biomedical Centre, Box 576, Uppsala University,
S-75123 Uppsala, Sweden*

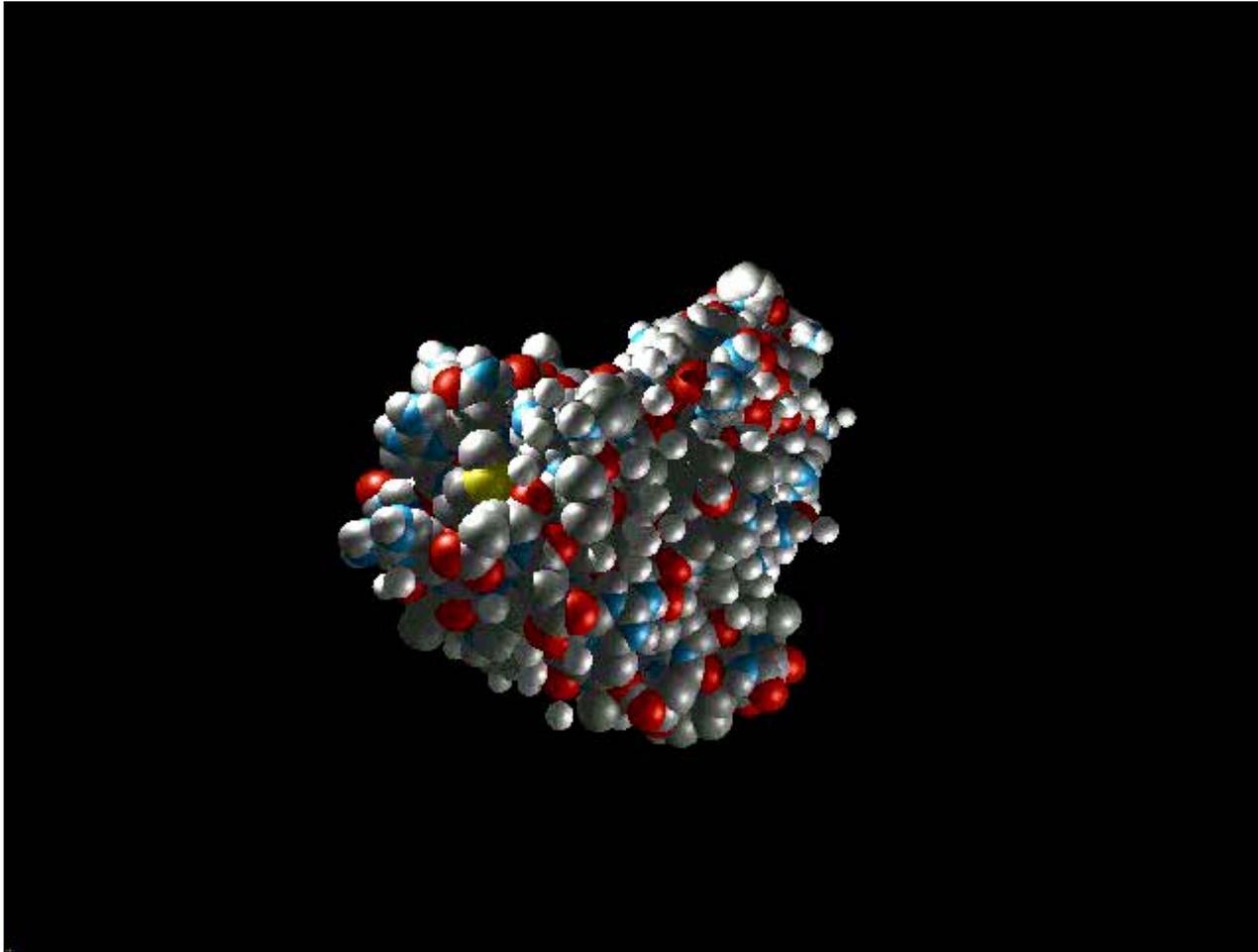
*[†] Institut für Kristallographie, Universität Karlsruhe, Kaiserstrasse 12, D-76128,
Germany*

200 fsec X-ray pulse onto lysozyme single molecule



200 fsec pulse, 3×10^{11} photons in 100 nm spot. Radiation damage interferes with atomic positions and the atomic scattering factors

100 fsec X-ray pulse onto lysozyme single molecule



100 fsec pulse, 3×10^{11} photons in 100 nm spot. One could record data prior to onset of significant radiation damage

Manipulation of nano-scale samples

Sphere of frustration (Paul Sigler, Yale)

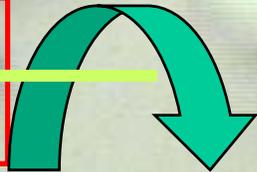
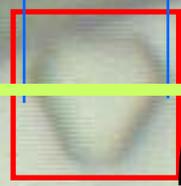
- Visualization of samples smaller than 100 nm
 - Goniometer with sphere of confusion better than 10 nm
- Electrospray techniques no control of sample rotation
 - Vitrous ice-EM techniques tilt angle limits
 - Immobilization of protein molecule(s) onto a substrate using multiple chemical bonds.
 - Laser tweezers

NANO crystals

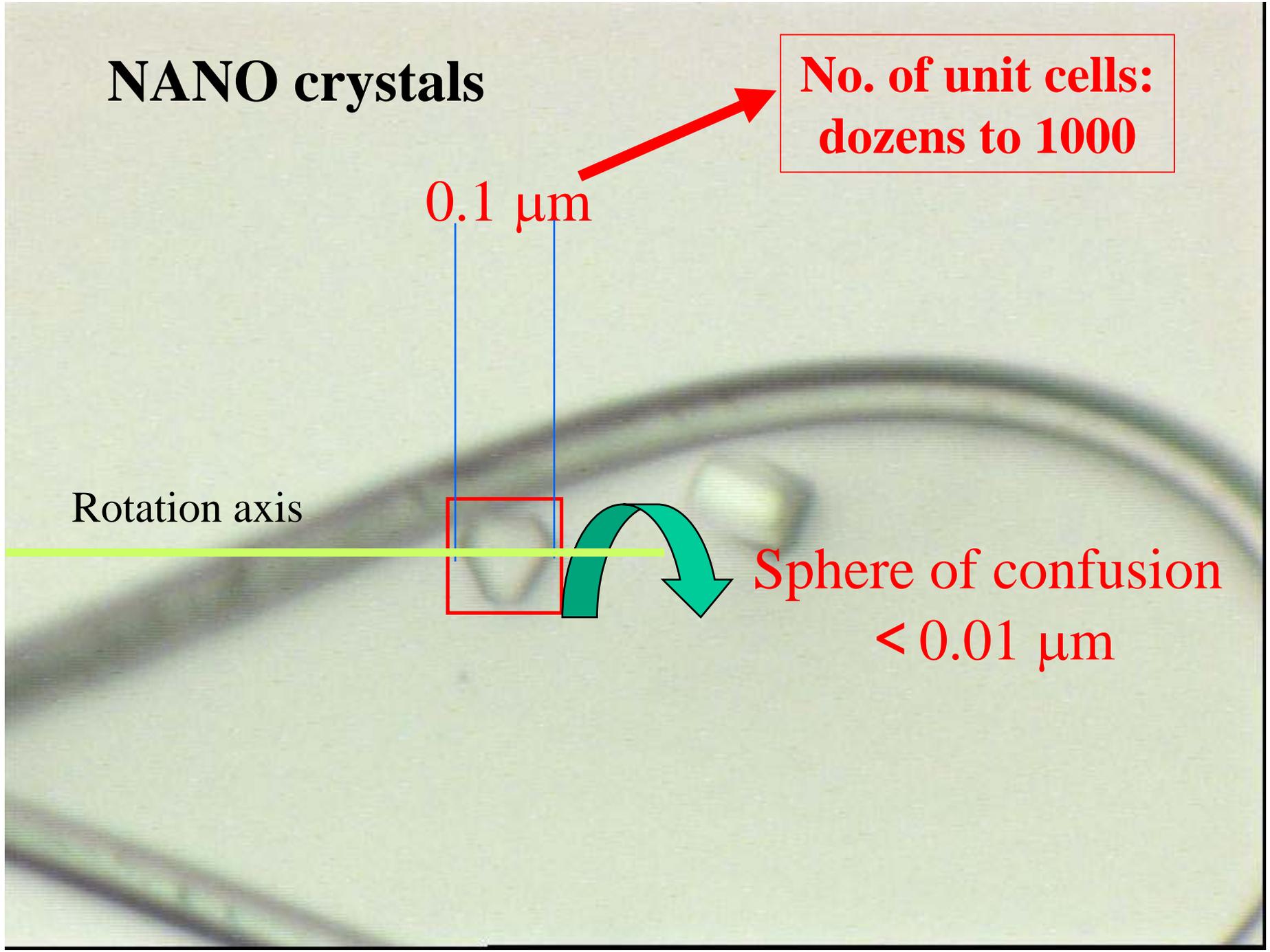
No. of unit cells:
dozens to 1000

0.1 μm

Rotation axis



Sphere of confusion
< 0.01 μm



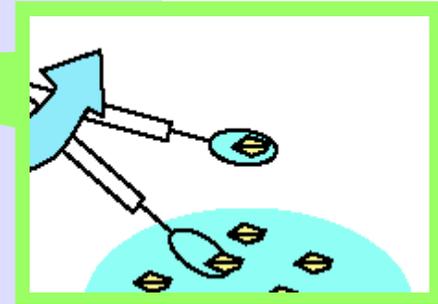
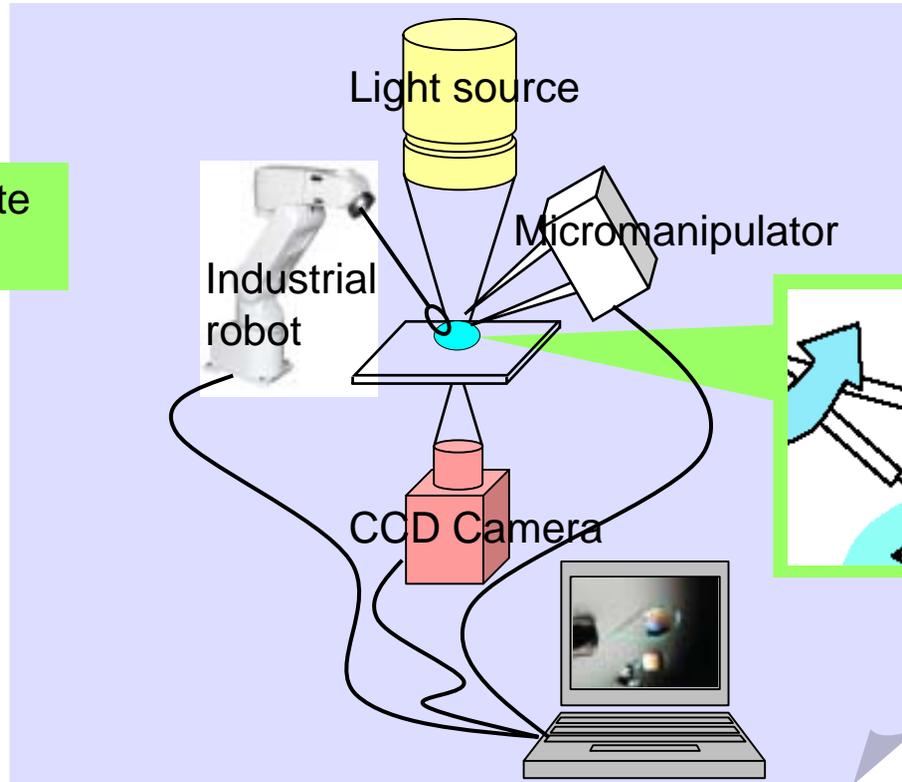
Carrying Robot



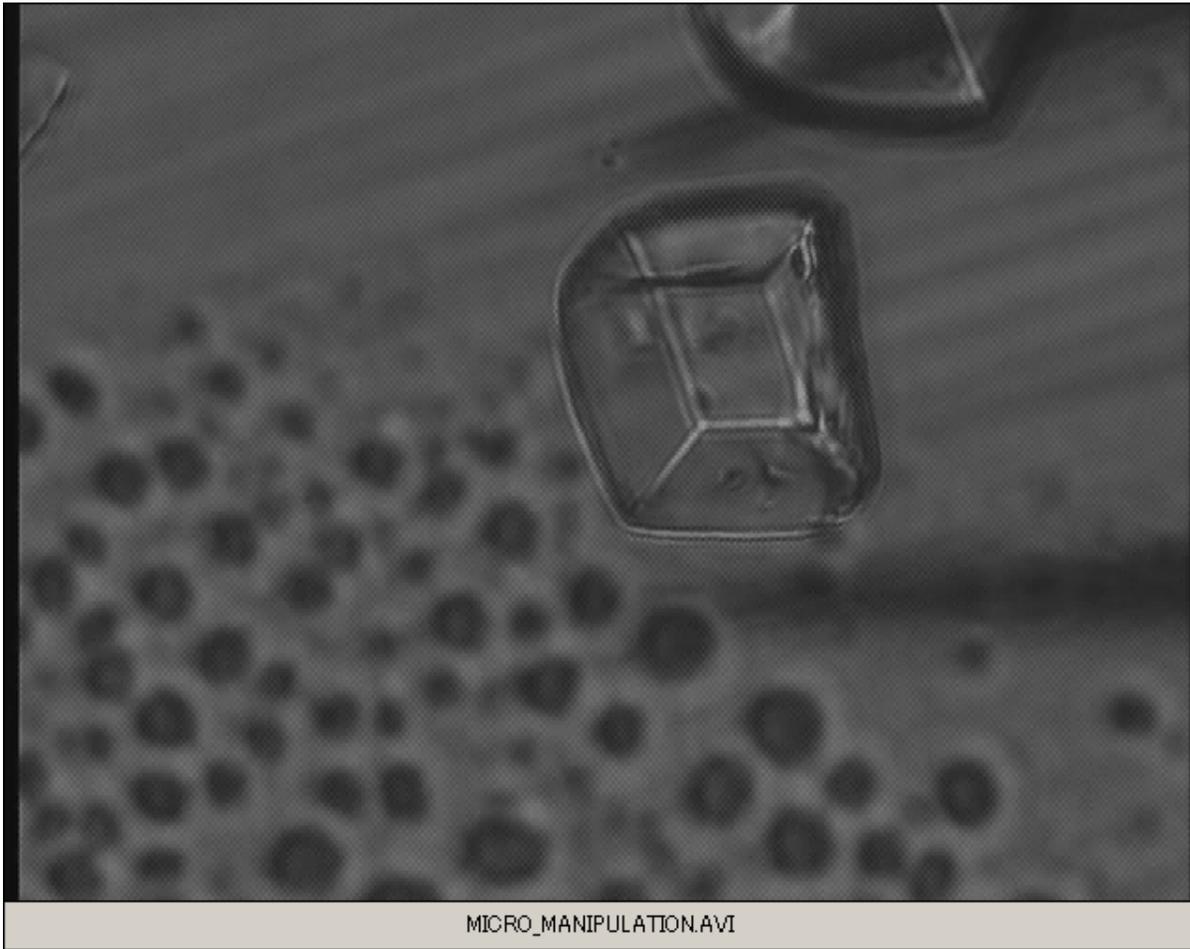
Loops

Decides an appropriate size of a loop

Crystal Harvesting System

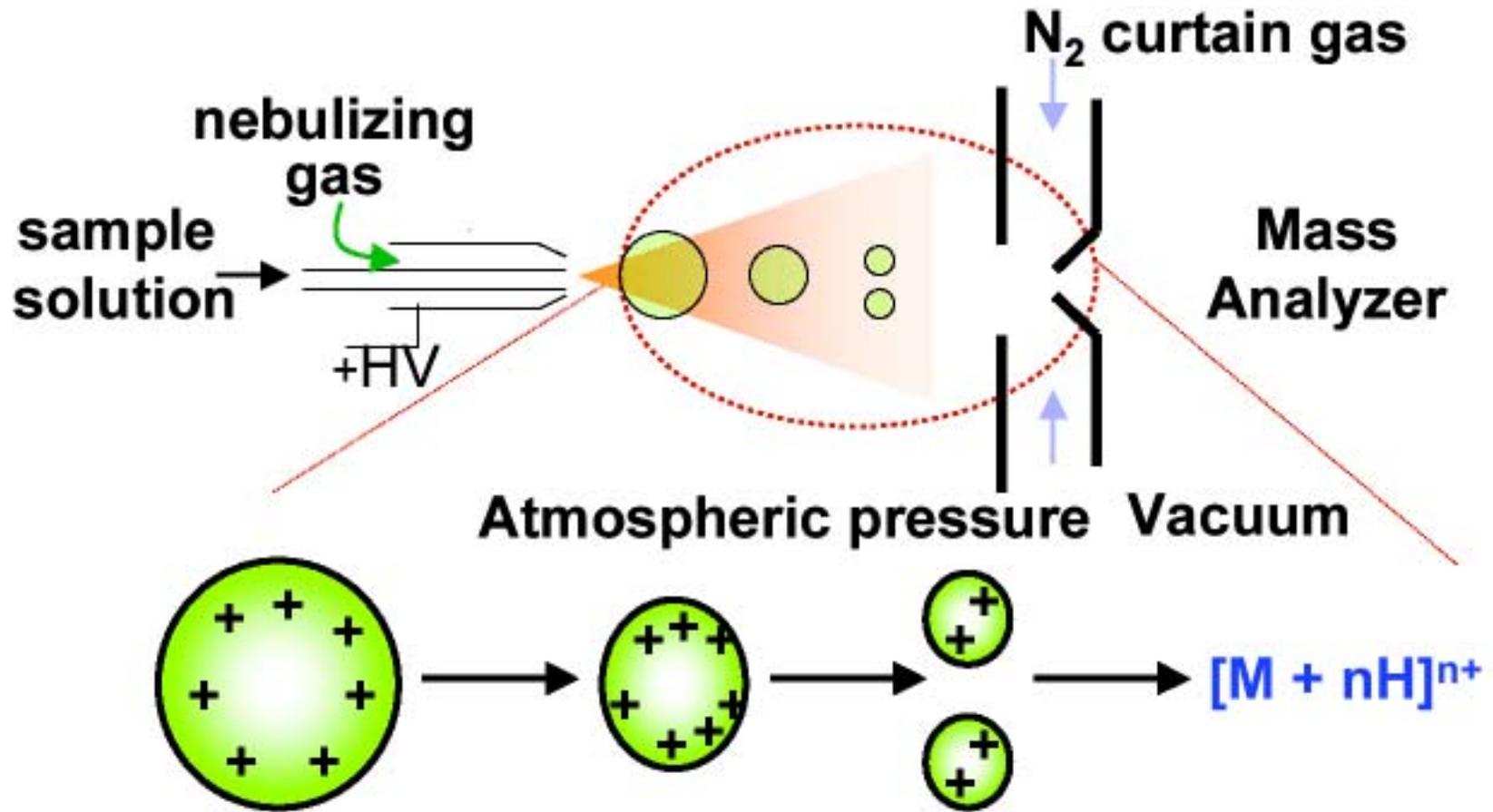


Mounting/Data Acquisition System



↑ ↑ ↑
Micromanipulation (international patent to be submitted,
Wakatsuki & Tanikawa)

Electrospray Ionization (ESI)



1. Solvent evaporation
2. Coulombic repulsion

Averaging is necessary anyway for atomic structure determination!

1. in crystal conventional crystallography
2. in computer single molecule analysis

Simulation: effect of the water molecules

Parameters in for the structure amplitudes calculation

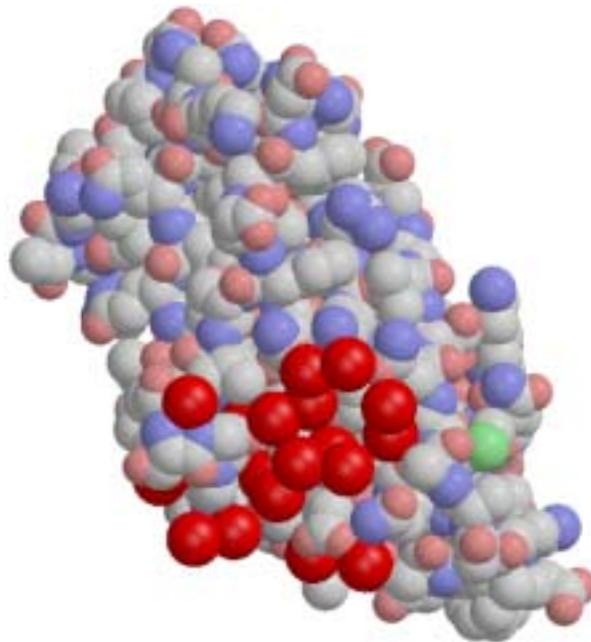
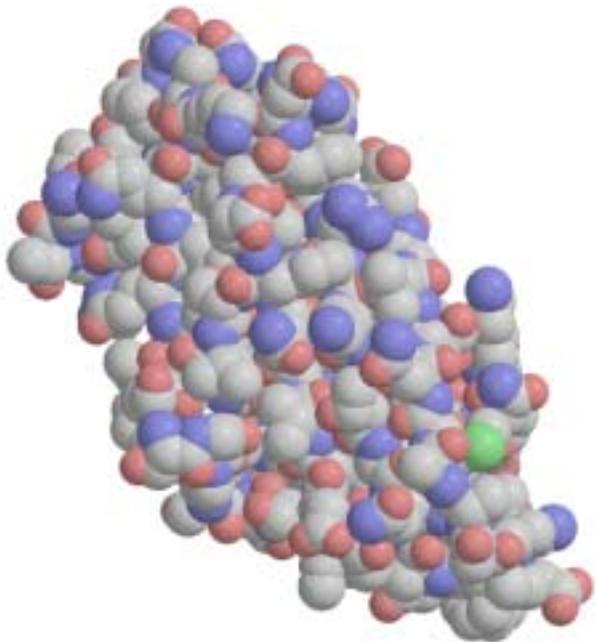
Resolution range: 30.0 – 2.5 Å
Spacegroup: P1
Unit cell: $a = b = c = 700 \text{ Å}, \alpha = \beta = \gamma = 90^\circ$

Coordinate: γ 1-ear domain complex with γ -synergin peptide (PDB ID:1UI4)
(from the crystal of the peptide tagged protein)

Resolution: 1.85 Å
Number of residues: 124
Number of water molecules: 140

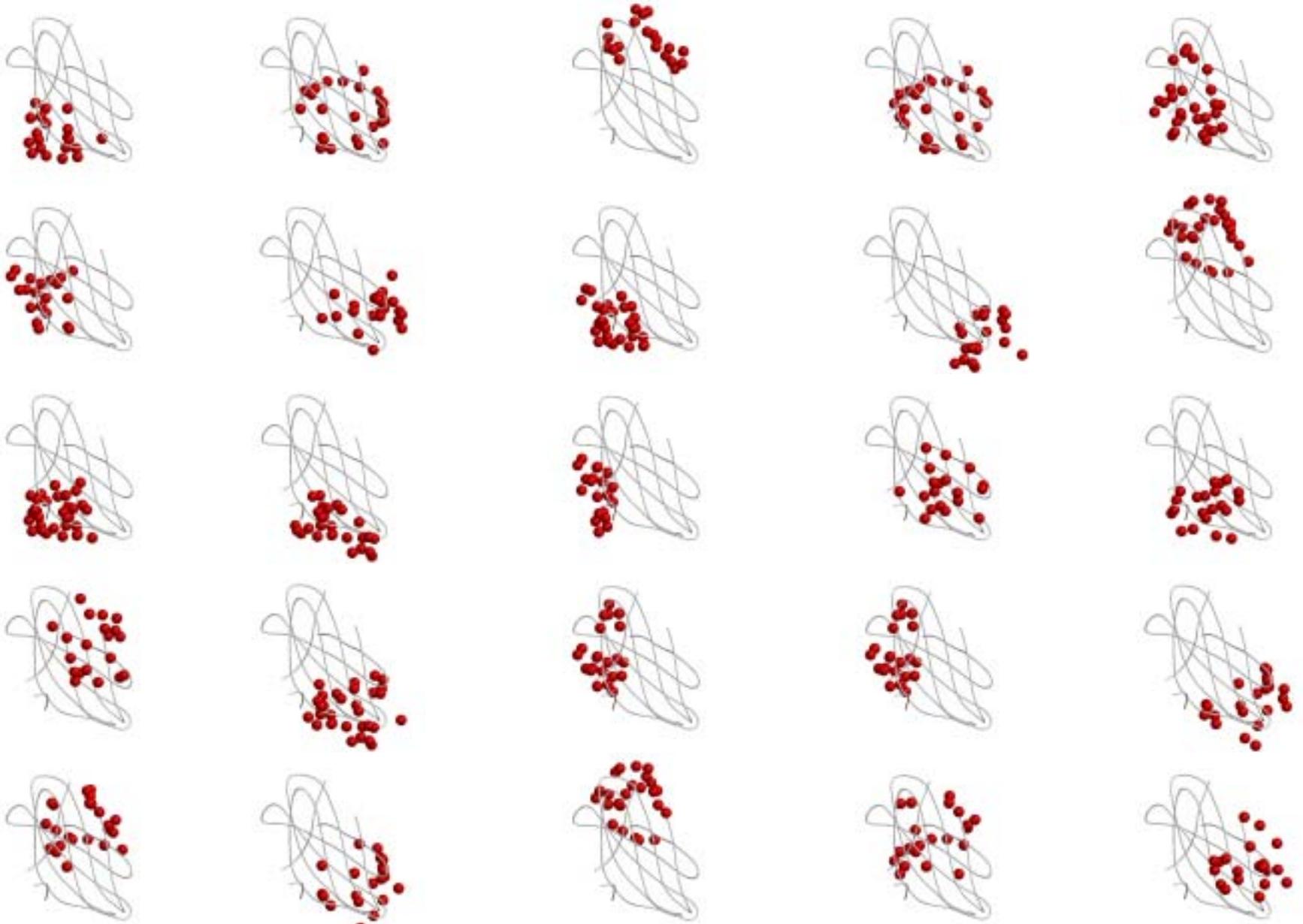
Protein only

Protein and 17 water molecules

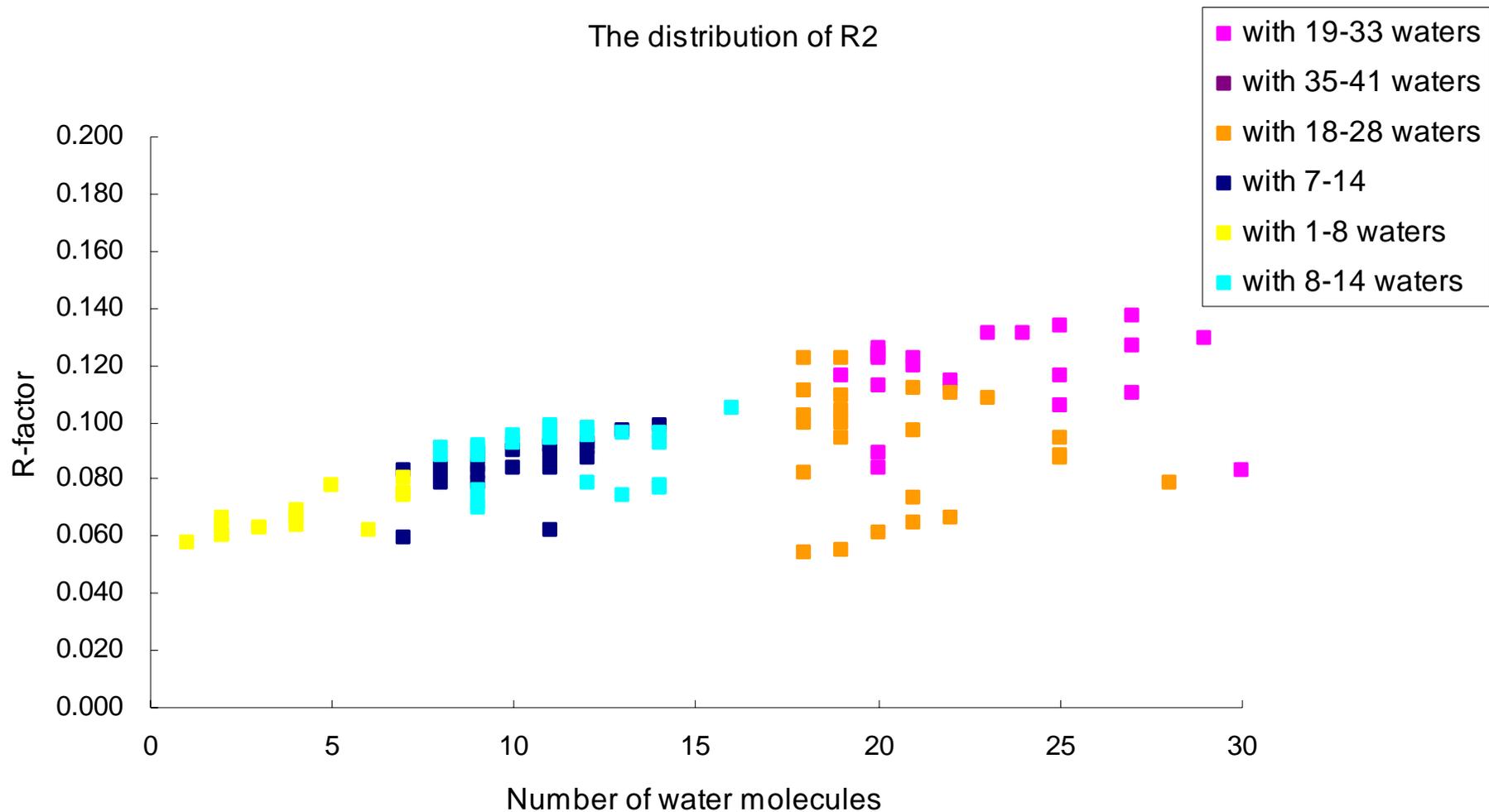


$$R1 = \frac{\sum_h \| F_{\text{protein-water}} - F_{\text{protein}} \|}{\sum_h \| F_{\text{protein}} \|}$$

Twenty-five coordinates were generated so that each coordinate contains the protein molecule and 19 ~ 33 water molecules which are within a sphere of 12 Å radius. The sphere was created randomly.



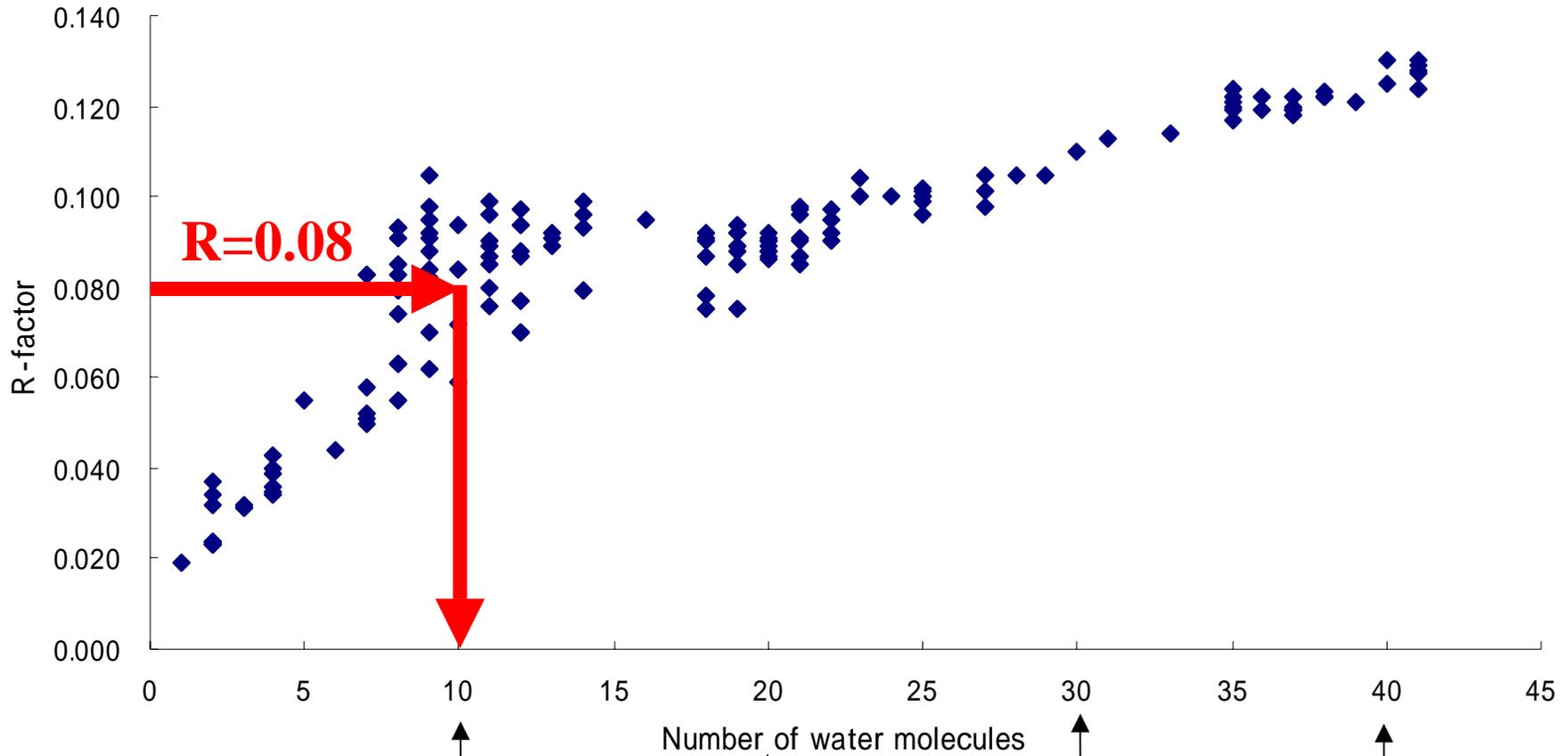
Difference between structures with different number of water molecules



Deviation from the protein structure without water

Water content should be less than 1 %.

The distribution of R1



% in No. of
electrons

1.2%

2.3%

3.5%

4.7%

XFEL REQUIREMENTS FOR THE PLANNED BIOLOGY EXPERIMENTS
Studies on large structures

	J. Hajdu et al.	S.W.'s comments
Bandwidth	0.2%	Could be much broader for experiments without anomalous signals.
Pulse length	230 fs and then shorter (while maintaining high dose/pulse)	Much longer for non-dynamic experiments
Pump-probe?	May be possible (delay pump-probe) Synchronization requirement: around 200 fs	Then short pulse
Polarization (hor/vert)	Not relevant	Must be known
Pulse-to-pulse fluctuations	Not a factor (normalization)	BIG problem since it will be almost impossible to normalize.

Comparison of various types of the coherent X-ray sources (by N. Kulipanov)

	ESRF storage ring	LCLS linac	MARS
Wavelength, nm	.1	.15	.1
Electron energy, GeV	6	14	5.4
Average current, A	.2	3×10^{-8}	10^{-3}
Peak current, A		3.4×10^3	1
Relative energy spread		2×10^{-4}	1×10^{-5}
Emittance, nm ϵ_x ϵ_z	4 2.5×10^{-2}	3×10^{-2}	3×10^{-3}
Undulator period, cm	4.2	3	1.5
Undulator length, m	5	100	150
Coherent flux, photon/s	6×10^{12}	6×10^{14}	7×10^{13}
Bandwidth	10^{-2}	10^{-3}	10^{-4}
Average brightness, ph/s/mm ² /mrad ² /0.1% BW	10^{20}	6×10^{22}	3×10^{23}
Peak brightness, --/--		5×10^{33}	3×10^{26}
Transverse size of source (standard deviation), μm	σ_x 350 σ_y 8	9	10
Radiation transverse divergence (standard deviation), μrad	σ_x' 13 σ_y' 3	2	1

PF-ERL Main Parameters

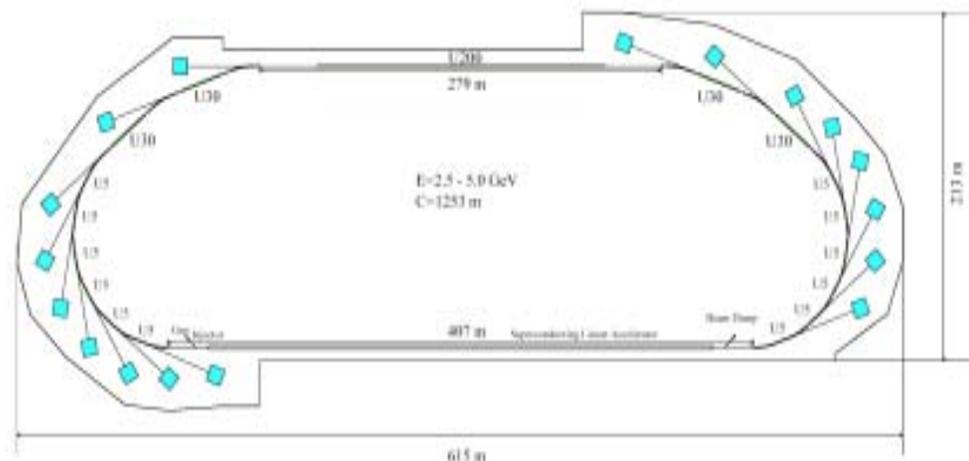
Beam Energy	2.5 ~ 5.0 (GeV)
Injection Energy	10 (MeV)
Circumference	1253 (m)
Beam Current	~100 (mA)

Normalized Emittance	~0.1 (μmrad)
Horizontal Emittance	~10.0 (pmrad) at 5.0 GeV
Vertical Emittance	~10.0 (pmrad) at 5.0 GeV

Energy Spread	~ 5×10^{-5}
Bunch Length	1 (ps) ~ 100 (fs)

RF Frequency	1.3 (GHz)
ACC. Gradient	~20 (MV/m)

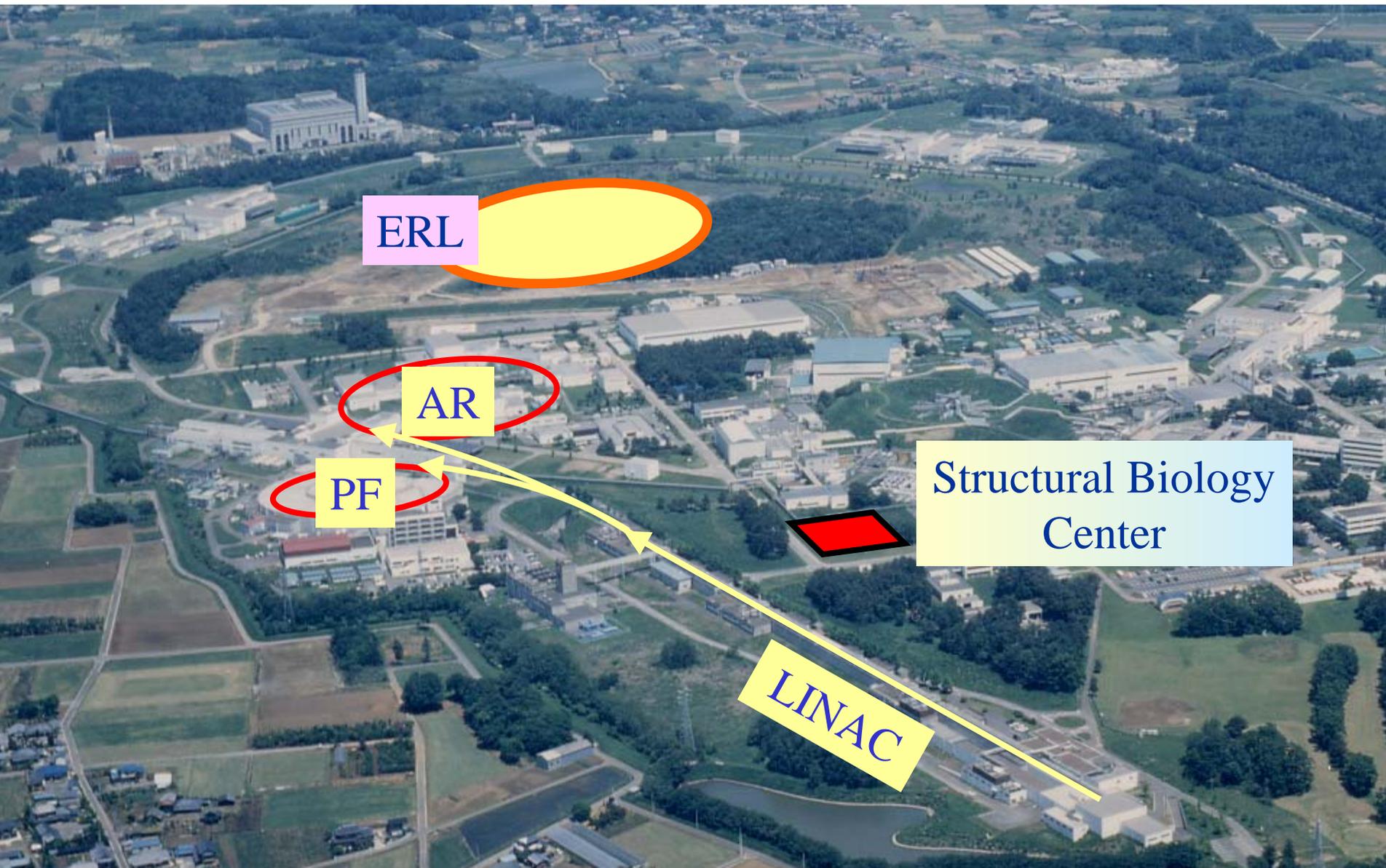
Long Undulator	200 (m) x 1
Middle Undulator	30 (m) x 4
Short Undulator	5 (m) x 12



SUMMARY

1. There will a very large number of biological problems that will require single molecule or nanocrystal structural studies
2. Radiation damage due to X-FEL pulse(s) still a problem
3. Nano crystals (smaller than 1 μm cube)
 - Very promising as a way to overcome the limitations of the 3rd generation synchrotron sources
4. Single molecule analysis
 - Phase determination using oversampling seems possible
 - Sample manipulation and determination of sample orientation requires a lot of work
5. Storage ring type SR facilities will not be replaced by XFEL -- cohabilitation

Future Plan for Synchrotron Sources at KEK



ERL

AR

PF

LINAC

Structural Biology
Center

PF Structural Biology Research Center

- Ryuichi Kato (Assoc. Prof.)
- Mamoru Suzuki (Research Assoc.)
- Noriyuki Igarashi (Research Assoc.)
- Naohiro Matsugaki (Research Assoc.)
- Masato Kawasaki (Research Assoc.)
- Masahiko Hiraki (Research Assoc.)
- Minora Nagai (Robotics technician)
- Tomoo Shiba (Post-doc)
- Shinsuke Hiramoto (Post-doc)
- Tamami Uejima (Staff scientist)
- Tadashi Sato (Post-doc)
- Satoshi Hirano (Post-doc)
- Michio Inoue (Ph.D. student)
- Yusuke Yamada (Ph.D. student)
- Yurii Gaponov (Scientific programmer)
- Leo Chavas (Ph.D. student ~Oct 2002, EMBL, Grenoble, France)

Vacancies

- Staff scientists (a few)
- Post-docs (4~6)
- Ph.D students (1~2)
- Engineers
- Technicians

