X-Ray Crystallography

"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

Topics:

1. **Protein Data Bank (PDB)**

   Data mining and Protein Structure Analysis Tools

2. **Image Formation**

   Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

---

3. **X-Ray Crystallography (after NMR)**

   a) Crystal Growth – Materials / Methods

   b) Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit

   c) X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron

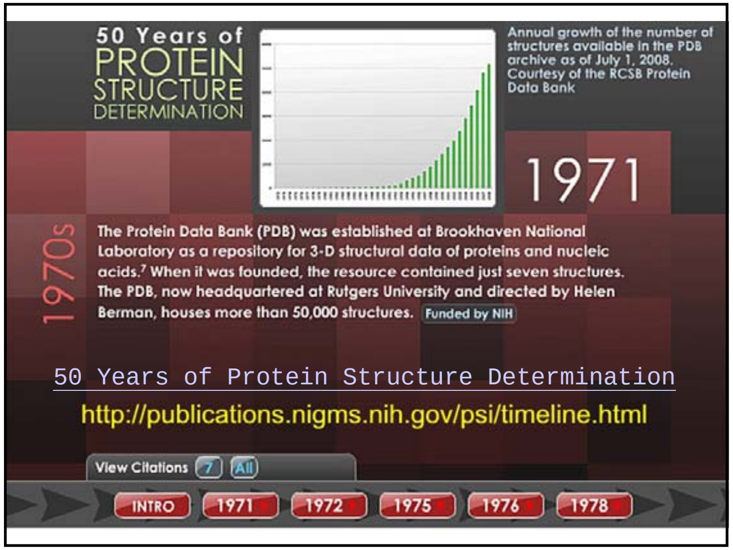
   d)Theory of Diffraction – Bragg's Law / Reciprocal Space

   e) Data Collection – Methods / Detectors / Structure Factors

   f) Structure Solution – Phase Problem: MIR / MR / MAD

   h) Refinements and Models

   i) Analysis and presentation of results



50 Years of PROTEIN STRUCTURE DETERMINATION

Annual growth of the number of structures available in the PDB archive as of July 1, 2008. Courtesy of the RCSB Protein Data Bank

1971

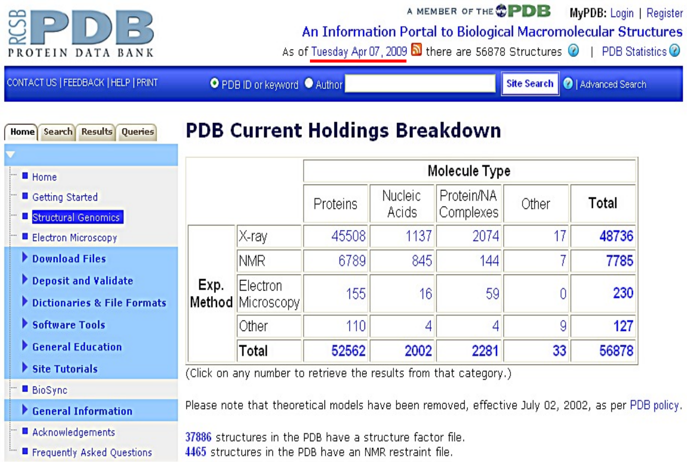
1970s

The Protein Data Bank (PDB) was established at Brookhaven National Laboratory as a repository for 3-D structural data of proteins and nucleic acids.[7] When it was founded, the resource contained just seven structures. The PDB, now headquartered at Rutgers University and directed by Helen Berman, houses more than 50,000 structures. Funded by NIH

50 Years of Protein Structure Determination

http://publications.nigms.nih.gov/psi/timeline.html

RCSB Protein Data Bank



An Information Portal to Biological Macromolecular Structures

As of Tuesday Apr 07, 2009 there are 56878 Structures | PDB Statistics

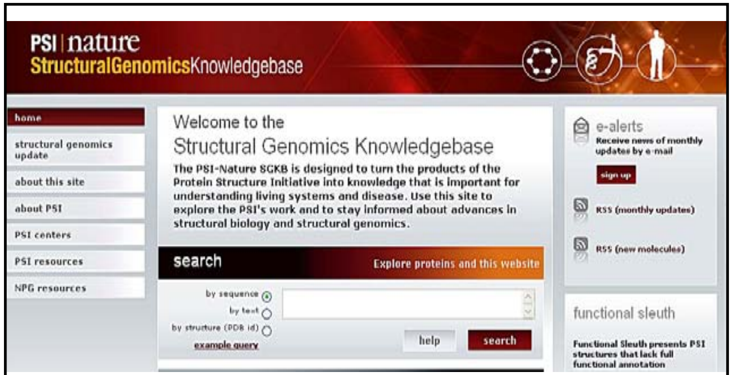
PDB Current Holdings Breakdown

		Molecule Type				
		Proteins	Nucleic Acids	Protein/NA Complexes	Other	**Total**
Exp. Method	X-ray	45508	1137	2074	17	**48736**
	NMR	6789	845	144	7	**7785**
	Electron Microscopy	155	16	59	0	**230**
	Other	110	4	4	9	**127**
	Total	**52562**	**2002**	**2281**	**33**	**56878**

(Click on any number to retrieve the results from that category.)

Please note that theoretical models have been removed, effective July 02, 2002, as per PDB policy.

37886 structures in the PDB have a structure factor file.
4465 structures in the PDB have an NMR restraint file.



Welcome to the
Structural Genomics Knowledgebase

The PSI-Nature SGKB is designed to turn the products of the Protein Structure Initiative into knowledge that is important for understanding living systems and disease. Use this site to explore the PSI's work and to stay informed about advances in structural biology and structural genomics.

Receive news of monthly updates by e-mail

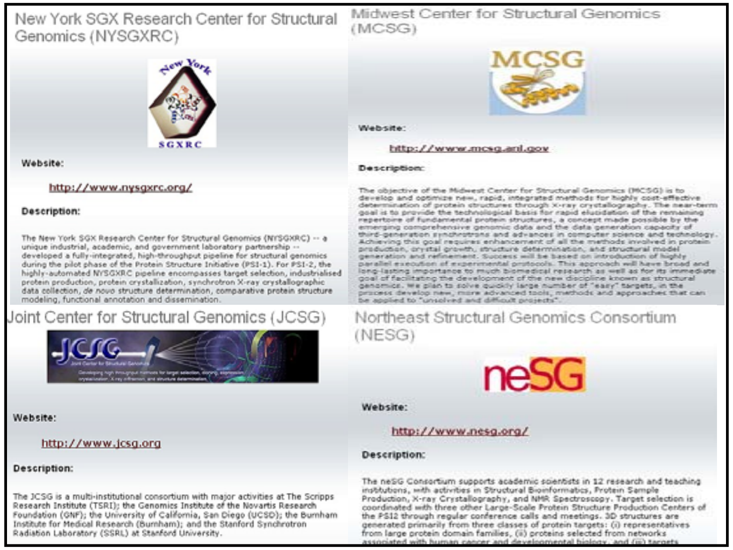
Functional Sleuth presents PSI structures that lack full functional annotation

4 - Large PSI Centers

2 - Modeling Centers

6 - Specialized Centers

Resouces Centers



New York SGX Research Center for Structural Genomics (NYSGXRC)

Website:

http://www.nysgxrc.org/

Description:

The New York SGX Research Center for Structural Genomics (NYSGXRC) -- a unique industrial, academic, and government laboratory partnership -- developed a fully-integrated, high-throughput pipeline for structural genomics during the pilot phase of the Protein Structure Initiative (PSI-1). For PSI-2, the highly-automated NYSGXRC pipeline encompasses target selection, industrialised protein production, protein crystallization, synchrotron X-ray crystallographic data collection, *de novo* structure determination, comparative protein structure modeling, functional annotation and dissemination.

Joint Center for Structural Genomics (JCSG)

Website:

http://www.jcsg.org

Description:

The JCSG is a multi-institutional consortium with major activities at The Scripps Research Institute (TSRI); the Genomics Institute of the Novartis Research Foundation (GNF); the University of California, San Diego (UCSD); the Burnham Institute for Medical Research (Burnham); and the Stanford Synchrotron Radiation Laboratory (SSRL) at Stanford University.

Midwest Center for Structural Genomics (MCSG)

Website:

http://www.mcsg.anl.gov

Description:

The objective of the Midwest Center for Structural Genomics (MCSG) is to develop and optimize new, rapid, integrated methods for highly cost-effective determination of protein structures through X-ray crystallography. The near-term goal is to provide the technological basis for rapid elucidation of the remaining repertoire of fundamental protein structures, a concept made possible by the emerging comprehensive genomic data and the data generation capacity of third-generation synchrotrons and advances in computer science and technology. Achieving this goal requires enhancement of all the methods involved in protein production, crystal growth, structure determination, and structural model generation and refinement. Success will be based on introduction of highly parallel execution of experimental protocols. This approach will have broad and long-lasting importance to much biomedical research as well as for its immediate goal of facilitating the development of the new discipline known as structural genomics. We plan to solve quickly large number of "easy" targets, in the process develop new, more advanced tools, methods and approaches that can be applied to "unsolved and difficult projects".

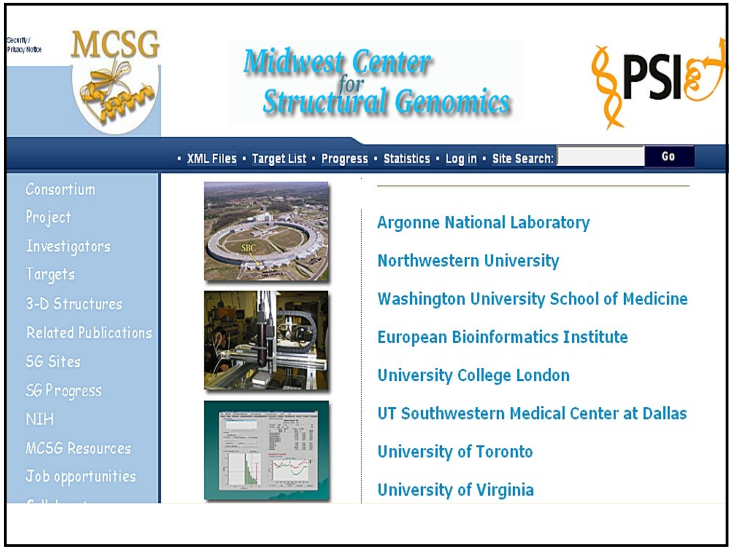
Northeast Structural Genomics Consortium (NESG)

Website:

http://www.nesg.org/

Description:

The neSG Consortium supports academic scientists in 12 research and teaching institutions, with activities in Structural Bioinformatics, Protein Sample Production, X-ray Crystallography, and NMR Spectroscopy. Target selection is coordinated with three other Large-Scale Protein Structure Production Centers of the PSI2 through regular conference calls and meetings. 3D structures are generated primarily from three classes of protein targets: (i) representatives from large protein domain families, (ii) proteins selected from networks associated with human cancer and developmental biology, and (iii) targets



Security / Privacy Notice
MCSG
Midwest Center for Structural Genomics
PSI
• XML Files • Target List • Progress • Statistics • Log in • Site Search:
Go
Consortium
Project
Investigators
Targets
3-D Structures
Related Publications
SG Sites
SG Progress
NIH
MCSG Resources
Job opportunities

Argonne National Laboratory

Northwestern University

Washington University School of Medicine

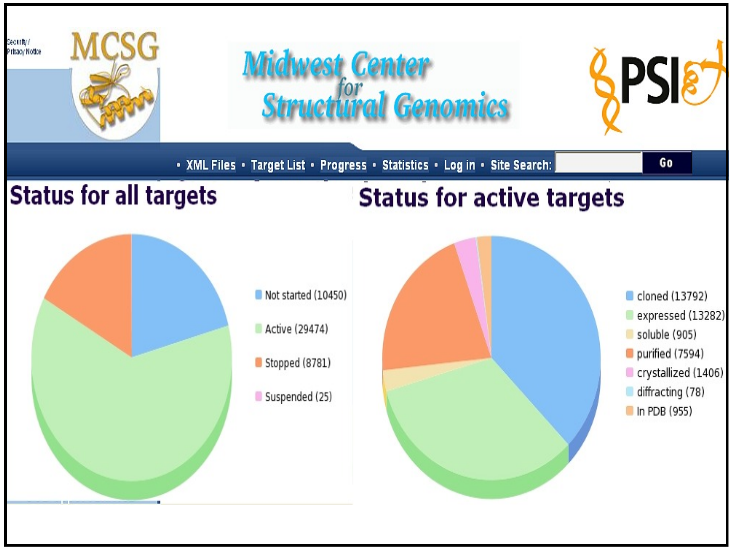
European Bioinformatics Institute

University College London

UT Southwestern Medical Center at Dallas

University of Toronto

University of Virginia



PMP | The Protein Model Portal

The Protein Model Portal (PMP) gives access to the various models that can be leveraged from PSI targets and other experimental protein structures by comparative modeling methods. The current release of the portal allows searching 7.6 million precomputed model structures provided by different partner sites, and provides access to various interactive services for template selection, target-template alignment, model building, and quality assessment.

- CSMP - Center for Structures of Membrane Proteins
- JCSG - Joint Center for Structural Genomics
- MCSG - Midwest Center for Structural Genomics
- NESG - Northeast Structural Genomics Consortium
- NMHRCM - New Methods for High-Resolution Comparative Modeling
- NYSGXRC - New York SGX Research Center for Structural Genomics

Analyze – structure (Ramachandran Plot) and biochemistry

Publish in leading biochemical or structural biology journal

Contribute results (coordinates, etc.) to PDB

****

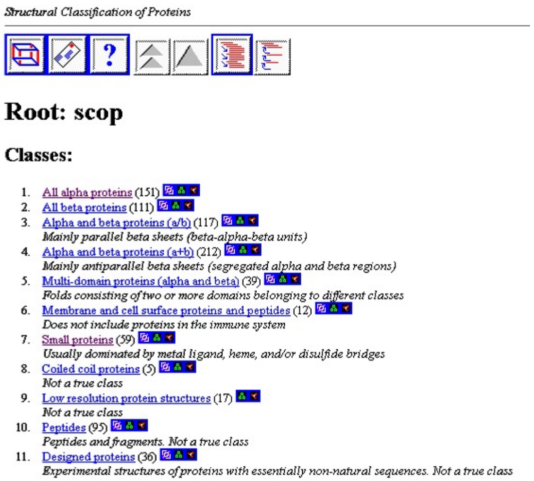
Data Mining

Visualization programs (Cn3D / RasMol / SwissPDBV / etc)

SCOP – Structural Classification of Proteins

CATH – Classification / Arch / Topology

SCOP Structural Classification of Proteins



Structural Classification of Proteins

Root: scop

Classes:

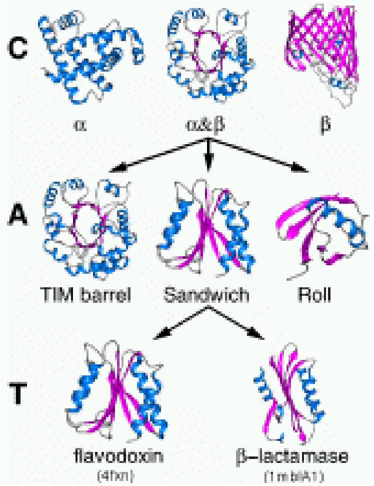
1. All alpha proteins (151)
2. All beta proteins (111)
3. Alpha and beta proteins (a/b) (117)
   Mainly parallel beta sheets (beta-alpha-beta units)
4. Alpha and beta proteins (a+b) (212)
   Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. Multi-domain proteins (alpha and beta) (39)
   Folds consisting of two or more domains belonging to different classes
6. Membrane and cell surface proteins and peptides (12)
   Does not include proteins in the immune system
7. Small proteins (59)
   Usually dominated by metal ligand, heme, and/or disulfide bridges
8. Coiled coil proteins (5)
   Not a true class
9. Low resolution protein structures (17)
   Not a true class
10. Peptides (95)
    Peptides and fragments. Not a true class
11. Designed proteins (36)
    Experimental structures of proteins with essentially non-natural sequences. Not a true class

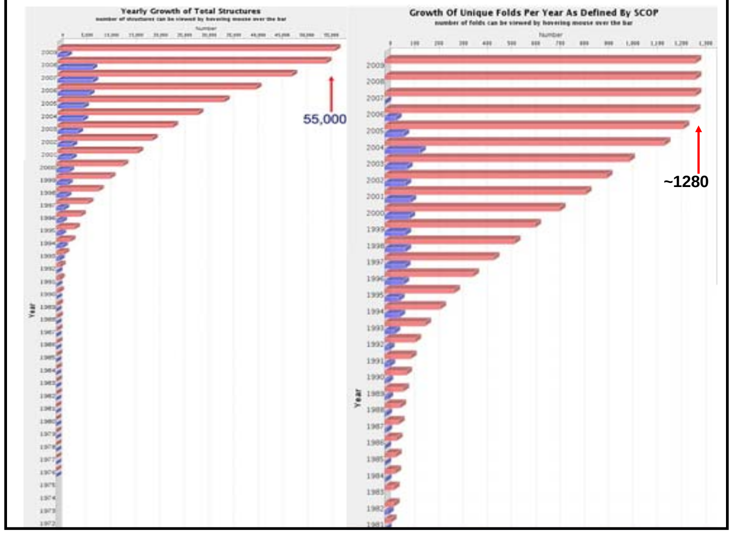
CATH - Protein Structure Classification

CATH is a novel hierarchical classification of protein domain structures, which clusters proteins at four major levels: Class (C), Architecture (A), Topology (T), and Homologous (H) Superfamily

Class, derived from secondary structure content, is assigned for more than 90% of protein structures automatically. **Architecture**, which describes the gross orientation of secondary structures, independent of connectivities, is currently assigned manually. The **topology** level clusters structures according to their topological connections and numbers of secondary structures. The **homologous superfamilies** cluster proteins with highly similar structures and functions. The assignments of structures to toplogy families and homologous superfamilies are made by sequence and structure comparisons.

CATH





X-Ray Crystallography

"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

Topics:

1. Protein Data Bank (PDB)

Data mining and Protein Structure Analysis Tools

2. Image Formation

Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

3. X-Ray Crystallography

a) Crystal Growth – Materials / Methods

b) Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit

c) X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron

d)Theory of Diffraction – Bragg's Law / Reciprocal Space

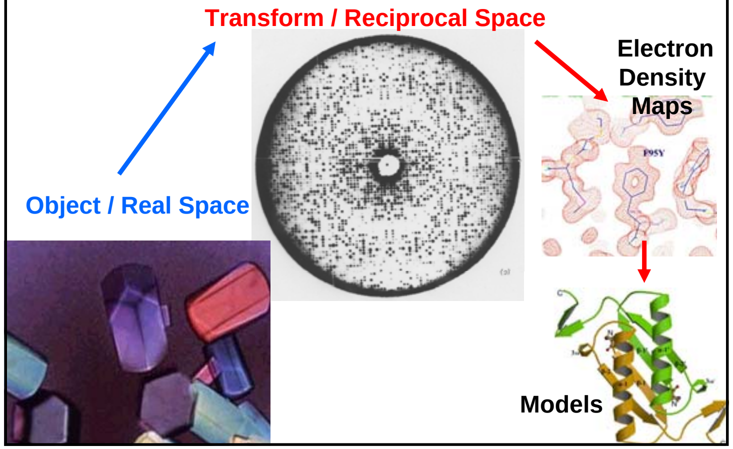
e) Data Collection – Methods / Detectors / Structure Factors

f) Structure Solution – Phase Problem: MIR / MR / MAD

h) Refinements and Models

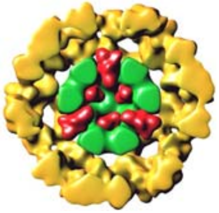
i) Analysis and presentation of results

Object → Transform → Image



"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

- Light Photography
  λ ~ 400 - 700 nm
- Electron Microscopy
  λ ~ 0.001 - 0.1 nm



- X-Ray or NMR
  λ ~ 0.1 nm

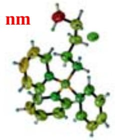


Image Formation

Abbe (~1873):

Limit Res. ~ λ/2

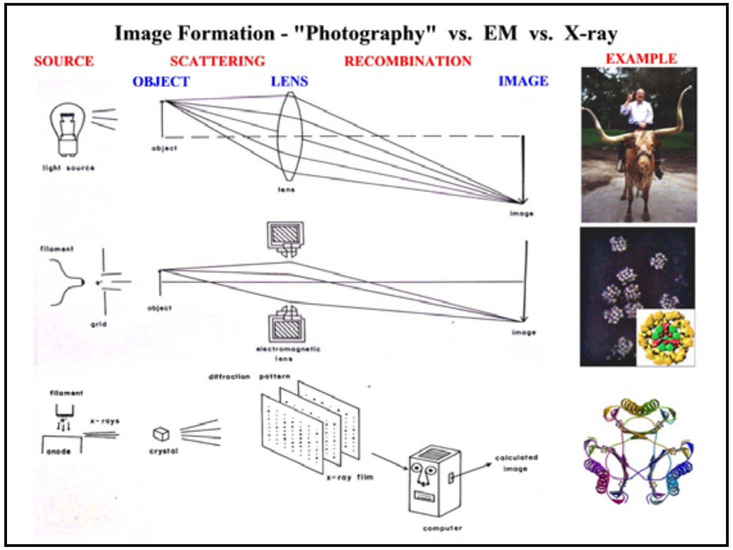
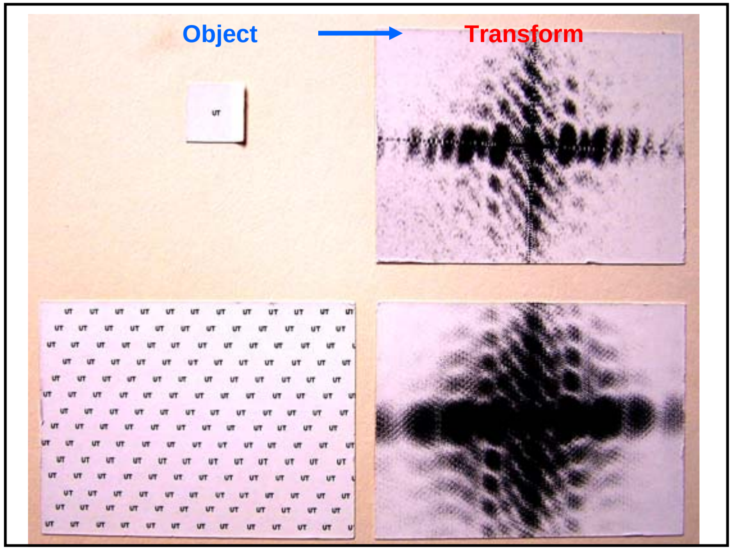
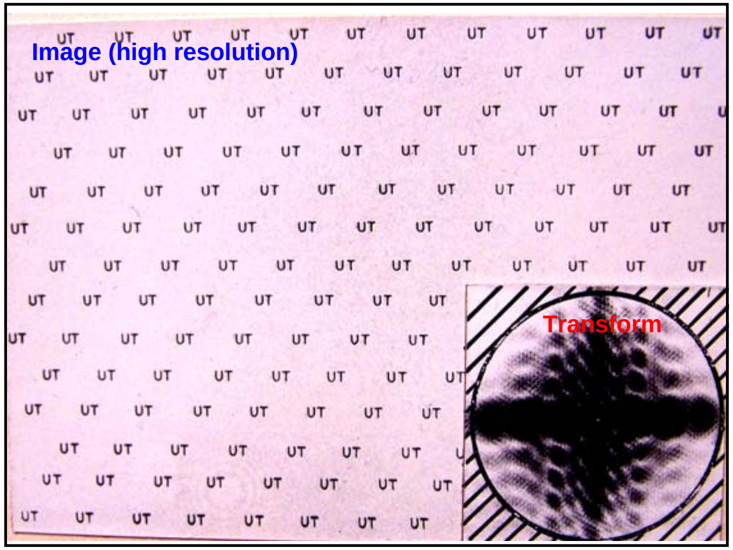
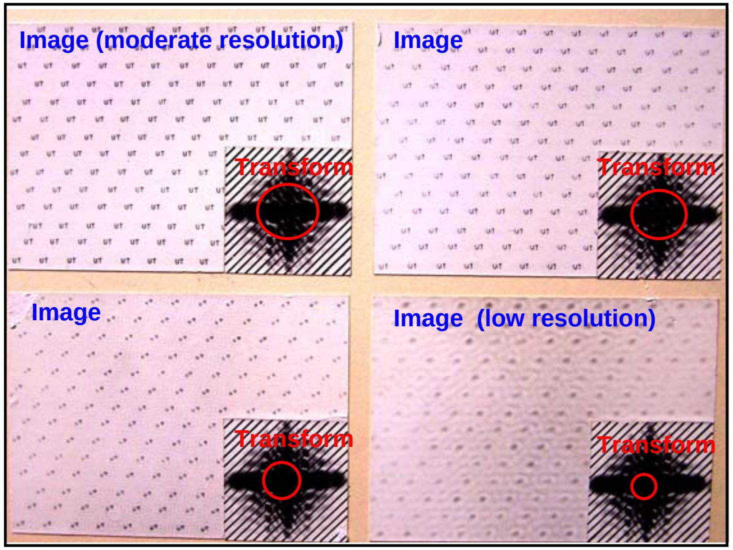


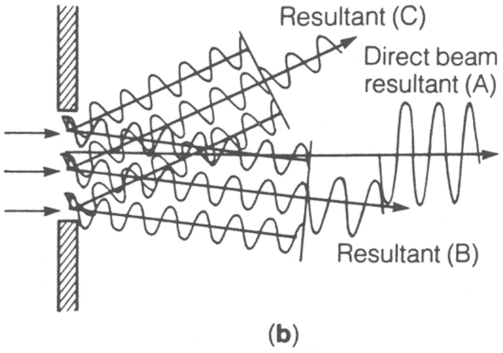
Image Formation - "Photography" vs. EM vs. X-ray





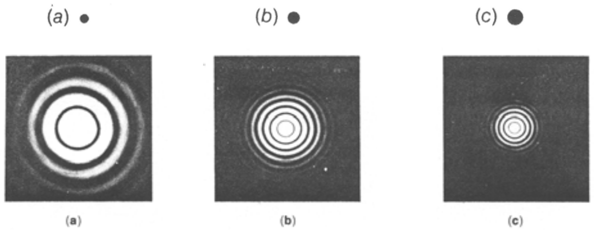


Transforms / Reciprocal Space

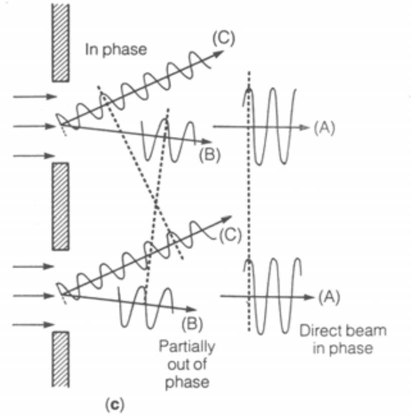


Transforms / Reciprocal Space

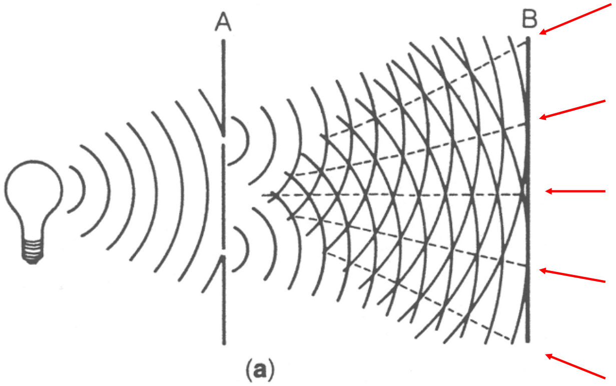
Different size holes



Transforms / Reciprocal Space

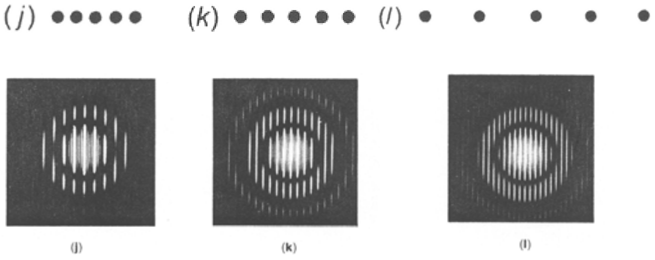


Transforms / Reciprocal Space



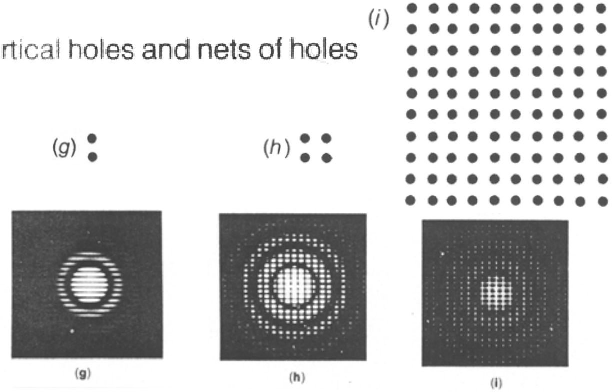
Transforms / Reciprocal Space

Five horizontal holes
with various spacings



Transforms / Reciprocal Space

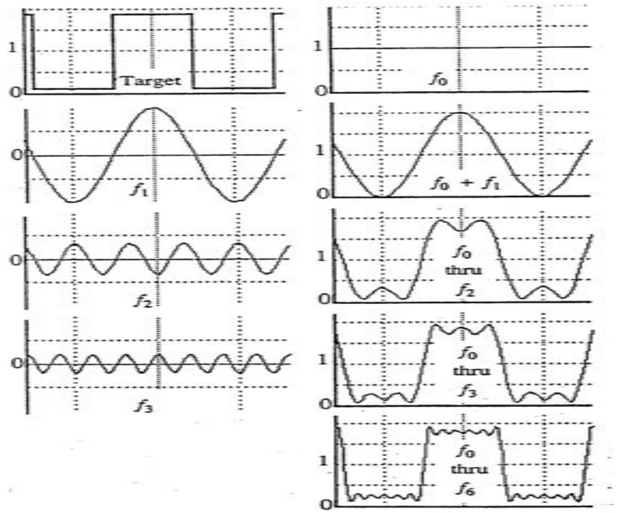
Vertical holes and nets of holes



Joseph Fourier / Fourier Series ~1808

Fourier series are named in honor of Joseph Fourier (1768-1830), who made important contributions to the study of trigonometric series, after preliminary investigations by Euler, d'Alembert, and Bernoulli. He applied this technique to find the solution of the heat equation, publishing his initial results in 1807, and publishing his Théorie analytique de la chaleur in 1822

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi t}{L} + \sum_{n=1}^{\infty} b_n \sin\frac{n\pi t}{L}$$



Fourier Series - a way of expressing functions in terms of an infinite series using the sum of sine and cosine functions.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi t}{L} + \sum_{n=1}^{\infty} b_n \sin\frac{n\pi t}{L}$$

If $f(t)$ is expanded in the range $-L$ to L (period = $2L$) so that the range of integration is $2L$, i.e. half the range of integration is L, then the Fourier coefficients are given by

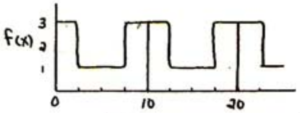
$$a_0 = \frac{1}{L}\int_{-L}^{L} f(t)dt$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(t)\cos\frac{n\pi t}{L}dt \quad b_n = \frac{1}{L}\int_{-L}^{L} f(t)\sin\frac{n\pi t}{L}dt$$

where $n = 1, 2, 3 \ldots$

Example:

$$f(x) = \begin{cases} 3 & \begin{matrix} 0<x<2.5 \\ 7.5<x<10 \end{matrix} \\ 1 & 2.5<x<7.5 \end{cases}$$



$$\Rightarrow a_n = \frac{1}{10}\int_0^{2.5}(3)\cos 2\pi\frac{nx}{10}dx + \frac{1}{10}\int_{2.5}^{7.5}(1)\cos 2\pi\frac{nx}{10}dx + \frac{1}{10}\int_{7.5}^{10}(3)\cos 2\pi\frac{nx}{10}dx$$

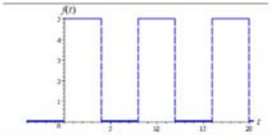
$$a_0 = \frac{1}{10}(7.5-0) + \frac{1}{10}(7.5-2.5) + \frac{1}{10}(30-22.5) = 2.0$$

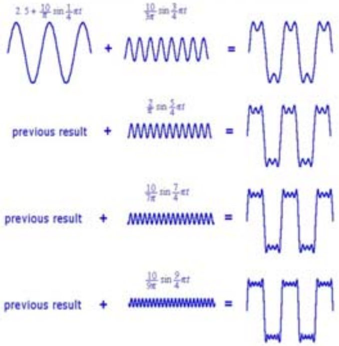
$$\underset{n\neq 0}{a_n} = \frac{1}{10}\left[\frac{3\cdot 10}{2\pi n}\sin\frac{2\pi nx}{10}\right]_0^{2.5} + \frac{1}{10}\left[\frac{10}{2\pi n}\sin\frac{2\pi nx}{10}\right]_{2.5}^{7.5} + \frac{1}{10}\left[\frac{3\cdot 10}{2\pi n}\sin\frac{2\pi nx}{10}\right]_{7.5}^{10}$$

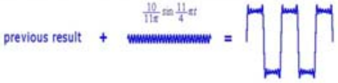
$$\Rightarrow \boxed{\begin{aligned} a_n &= \frac{1}{n\pi}\left[\sin\frac{n\pi}{2} - \sin\frac{3n\pi}{2}\right] \\ a_0 &= 2.0 \end{aligned}}$$

Fourier Series

Example - Square Wave

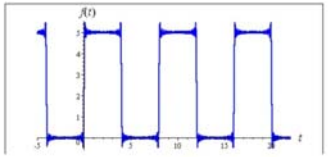






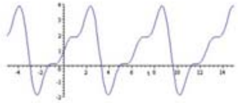
If we graph many terms, we see that our series is producing the required function. We graph the first 20 terms:

$$2.5 + \frac{10}{\pi}\sum_{n=1}^{20}\frac{1}{(2n-1)}\sin\frac{(2n-1)\pi t}{4}$$

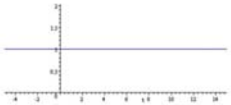


Example - Saw Tooth Function

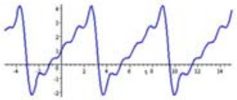
$f(t) = 1 + 2\sin t - \sin 2t + \frac{2}{3}\sin 3t$



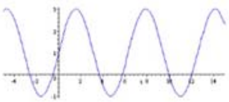
$f(t) = 1$ (first term of the series):



$f(x) = 1 + 2\sin t - \sin 2t + \frac{2}{3}\sin 3t - \frac{1}{2}\sin 4t + \frac{2}{5}\sin 5t + \ldots$

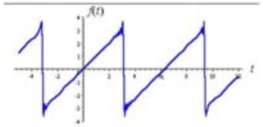


$f(t) = 1 + 2\sin t$ (first 2 terms of the series):

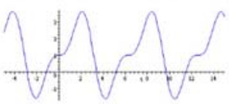


The graph of the first 40 terms is:

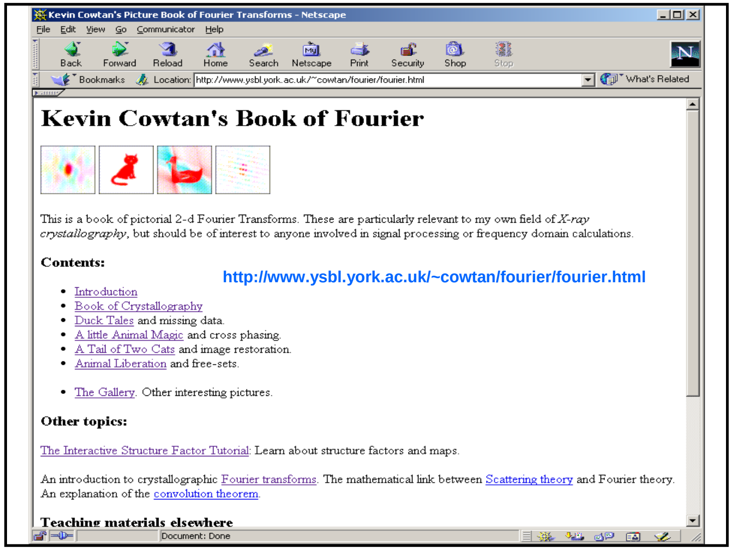
$$\sum_{n=1}^{40}\left(\frac{2}{n}(-1)^{n+1}\sin nt\right)$$



$f(t) = 1 + 2\sin t - \sin 2t$ (first 3 terms of the series):



Fourier Series Applet http://www.falstad.com/fourier/


Kevin Cowtan's Picture Book of Fourier Transforms - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html
What's Related
Document: Done

Kevin Cowtan's Book of Fourier

This is a book of pictorial 2-d Fourier Transforms. These are particularly relevant to my own field of *X-ray crystallography*, but should be of interest to anyone involved in signal processing or frequency domain calculations.

Contents:

http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

- Introduction
- Book of Crystallography
- Duck Tales and missing data.
- A little Animal Magic and cross phasing.
- A Tail of Two Cats and image restoration.
- Animal Liberation and free-sets.

- The Gallery. Other interesting pictures.

Other topics:

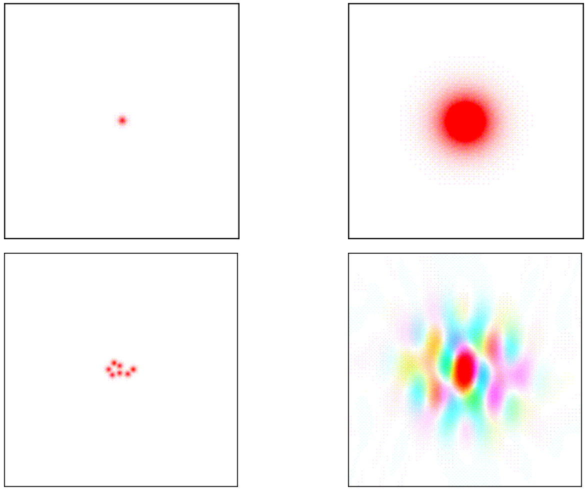
The Interactive Structure Factor Tutorial: Learn about structure factors and maps.

An introduction to crystallographic Fourier transforms. The mathematical link between Scattering theory and Fourier theory. An explanation of the convolution theorem.

Teaching materials elsewhere

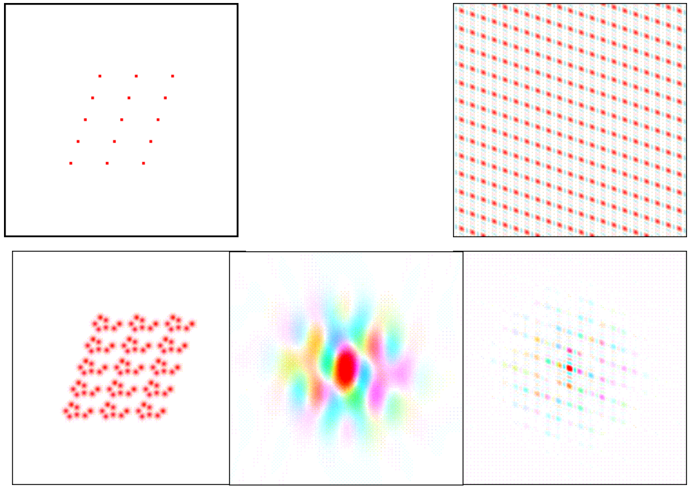
Object / Real Space

Transform / Reciprocal Space



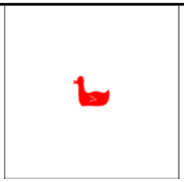
Object / Real Space

Transform / Reciprocal Space

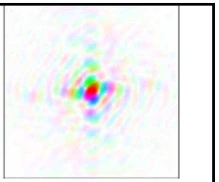


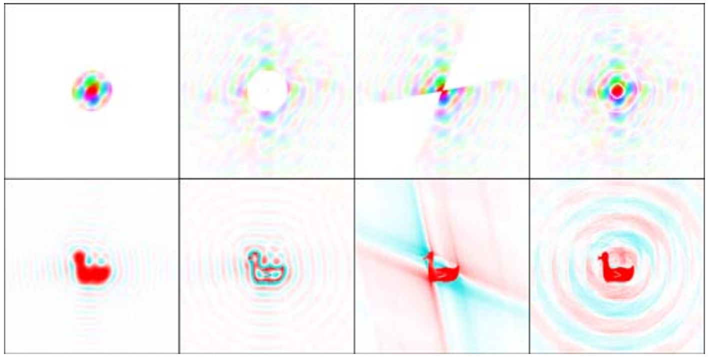
Objects – Transforms and Image Formation

A Duck



Transform of a Duck

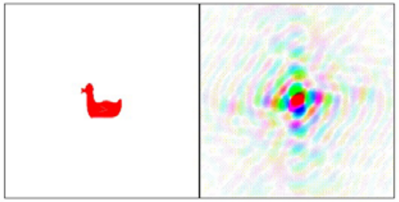




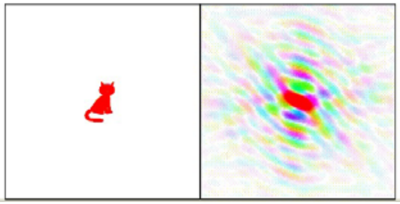
Kevin Cowtan's Book of Fourier

http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

Here is our old friend, the Fourier Duck, and his Fourier transform:



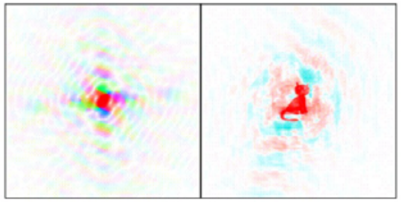
And here is a new friend, the Fourier Cat and *his* Fourier transform:



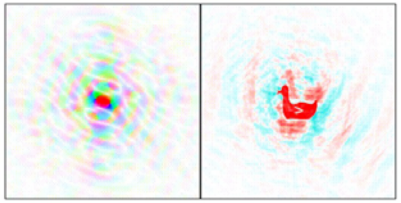
Kevin Cowtan's Book of Fourier

http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

Duck Transform Amplitudes + Cat Phases

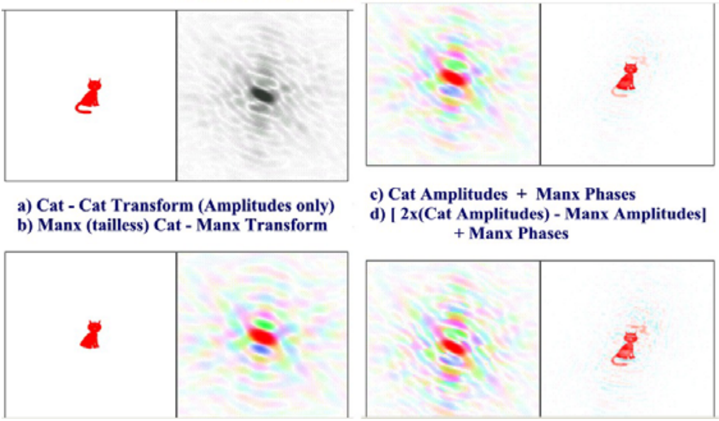


Cat Transform Amplitudes + Duck Phases



Kevin Cowtan's Book of Fourier

http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html



a) Cat - Cat Transform (Amplitudes only)
b) Manx (tailless) Cat - Manx Transform

c) Cat Amplitudes + Manx Phases
d) [2x(Cat Amplitudes) - Manx Amplitudes]
+ Manx Phases

X-Ray Crystallography

"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

Topics:

1. Protein Data Bank (PDB)

Data mining and Protein Structure Analysis Tools

2. Image Formation

Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

3. X-Ray Crystallography (after NMR)

a) Crystal Growth – Materials / Methods

b) Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit

c) X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron

d)Theory of Diffraction – Bragg's Law / Reciprocal Space

e) Data Collection – Methods / Detectors / Structure Factors

f) Structure Solution – Phase Problem: MIR / MR / MAD

h) Refinements and Models

i) Analysis and presentation of results