Where do molecular models come from?

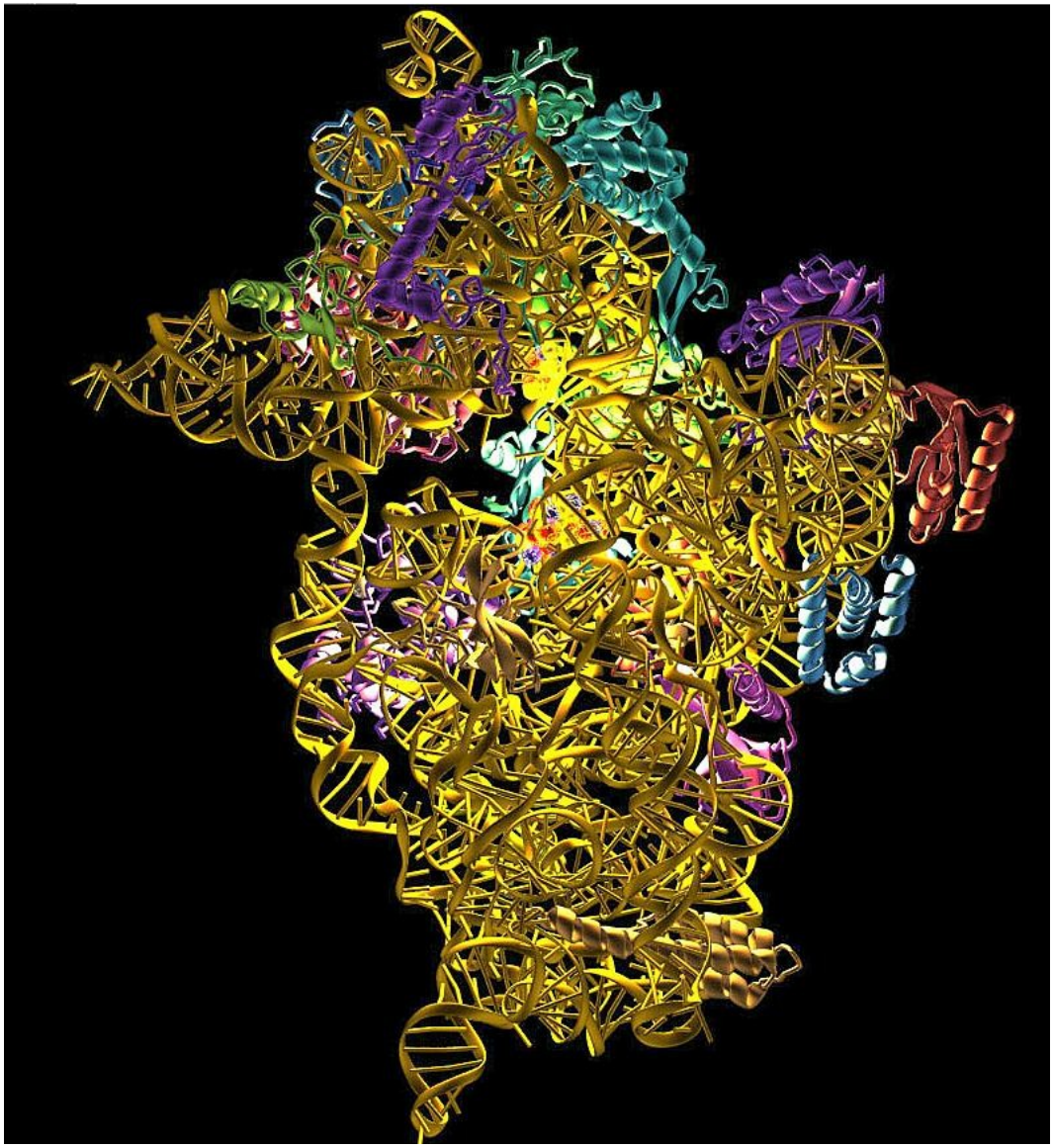
Tom Huxford

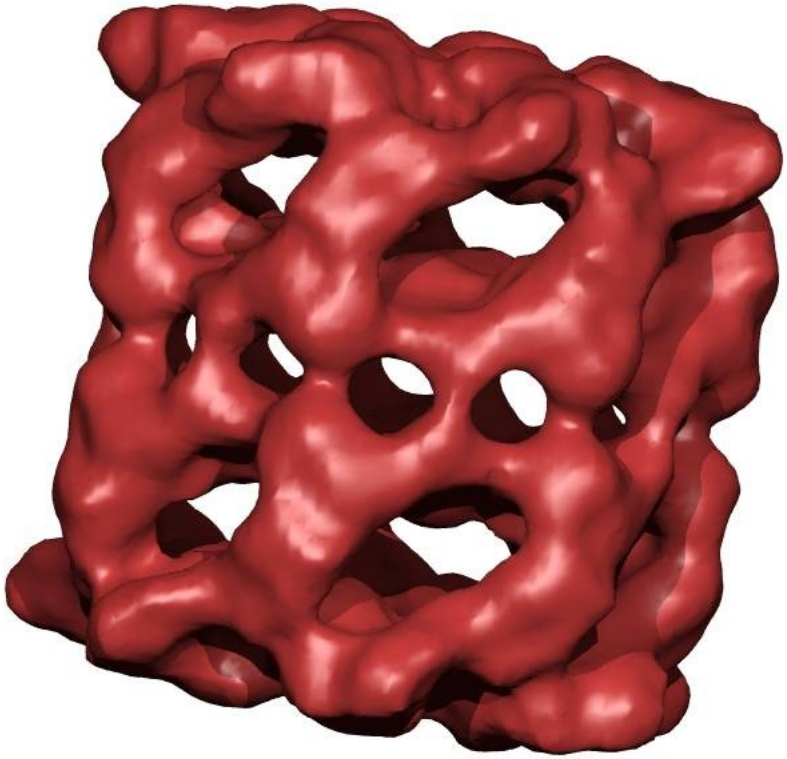
Department of Chemistry & Biochemistry

San Diego State University

Friday, February 27, 2009

What is meant by "model" anyway?





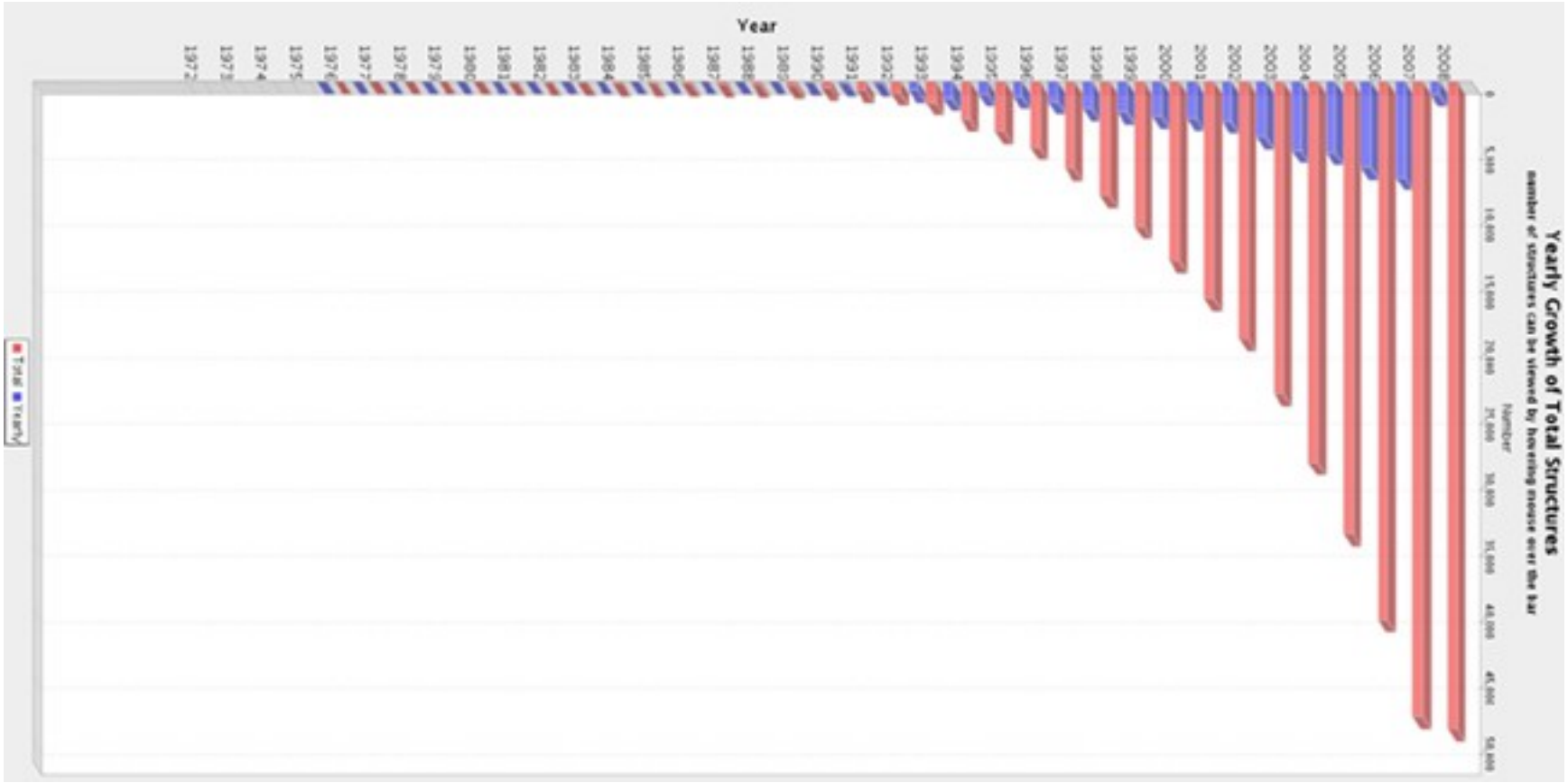
The model

- **Our best representation of a system given all of the experimental data available**
- **Models can be of differing values depending upon what they are based (Garbage in, garbage out)**
- **Structural biologists use experimentally derived high resolution data to generate models of extremely high quality**

Molecular models in computational biology

- **X-ray crystallography**
  - **Advantages: Subject size, resolution**
  - **Disadvantages: Sample preparation**
- **NMR spectroscopy**
  - **Advantages: Solution structure, resolution**
  - **Disadvantages: Subject size, sample preparation**
- **Cryoelectron Microscopy**
  - **Advantages: Size of samples, easy preparation**
  - **Disadvantages: Resolution, relies on symmetry**

The increasing number of protein structures



The number of reported macromolecular x-ray crystal structures continues to increase at a logarithmic rate.

The Protein Data Bank

RCSB PDB PROTEIN DATA BANK

A MEMBER OF THE PDB MyPDB: Login | Register

An Information Portal to Biological Macromolecular Structures

As of Tuesday Feb 24, 2009 there are 56066 Structures | PDB Statistics

CONTACT US | FEEDBACK | HELP | PRINT PDB ID or keyword Author Site Search | Advanced Search

Home Search

- Home
- Getting Started
- Structural Genomics
- Download Files
- Deposit and Validate
- Dictionaries & File Formats
- Software Tools
- General Education
- Site Tutorials
- BioSync
- General Information
- Acknowledgements
- Frequently Asked Questions

Quick Tips :

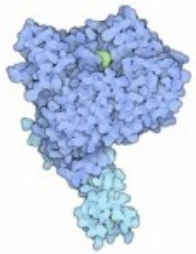
Use the RCSB RSS (Really Simple Syndication) feed for an updated list of new structures as soon as they are available.

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

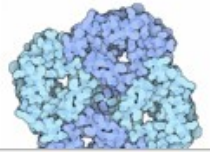
Molecule of the Month: Auxin and TIR1 Ubiquitin Ligase



Plants, like animals, have hormones that deliver chemical messages between distant cells. Charles Darwin and his son discovered this over a century ago--they noticed that if they shined a light on the tips of grass shoots, the stems bend to bring the entire shoot towards the light. Somehow, a message was being sent from the tip down to the stem. You might also have observed the action of hormonal signals in plants: when you prune a tree to make it more bushy, you are modifying the traffic of plant hormones. Both of these effects are caused by the phytohormone auxin.

Read more ... Previous Features

PSI Featured Molecule: Aquaglyceroporin



Researchers at the PSI CSMP have revealed the mechanism of the dual-specificity aquaglyceroporin from the major parasite that causes malaria.

Read more from PSI SGKB Previous Features

News

- Complete News
- Newsletter
- Discussion Forum
- Job Listings

24-February-2009

Exhibition at the Biophysical Society Meeting

The RCSB PDB will participate in the exhibition at the 53rd Annual Meeting of the Biophysical Society (February 28 - March 4) in Boston, Massachusetts. More >>

http://www.rcsb.org-the current home of the Protein Data Bank. Note that there are 56,066 structures available as of 1:14 p.m. this afternoon.

X-ray crystallography-History

- **William Röntgen (1845-1923)**
  - **Discovery of x-rays, 1895 (Nobel Prize in Physics, 1901)**

- **Braggs, von Laue, and others formulate theory and practice of x-ray diffraction by inorganic crystals, 1910-1920**

- **J.D. Bernal and Dorothy Crowfoot Hodgkin theorize and then illustrate the diffraction of x-rays by protein crystals, 1934**

X-ray crystallography-History

- **First Protein Structure, 1959**
  - **Myoglobin (John Kendrew, Nobel Prize in Chemistry, 1962-shared with Max Perutz)**
- **First Enzyme Structure, 1965**
  - **Lysozyme**
- **First Membrane Protein Structure, 1985**
  - **Photosynthetic Reaction Center (Diesenhoefer, Huber, Michel, Nobel Prize in Chemistry, 1988)**

Protein structure determination by X-ray crystallography

- **Derive a source of material for study**
  - **Cloning, expression, purification**
- **Grow single crystals of subject**
  - **Crystallization**
- **Collect x-ray diffraction data**
  - **X-ray source, experimental design**
- **X-ray diffraction data processing**
  - **Convert spots to numbers**

Protein structure determination by X-ray crystallography

- **Solve the "phase problem"**
  - **MIR**
  - **MR**
  - **MAD/SAD**
- **Build and refine model**
  - **Computer graphics workstation**
- **Analyze structure**
  - **Propose and then test structure-based functional hypotheses**

Derive a source of material for study

- **Many biologically active factors are present in extremely limited quantities within cells**
- **Recombinant DNA technology and exogenous protein expression systems**
  - ***E. coli* bacteria with recombinant plasmid DNA**
  - **Yeast (*S. cerevisiae* and *P. pastoris*)**
  - **sf9 insect cells infected by recombinant baculovirus**
  - **Mammalian cells (HEK 293) stably transformed**

Grow single crystals of subject

- **Purity AND homogeneity of sample**
- **Supersaturate sample under mild conditions of pH, temperature, solvent**
- **Leave undisturbed**
- **Watch, think, experiment!**

Grow single crystals of subject

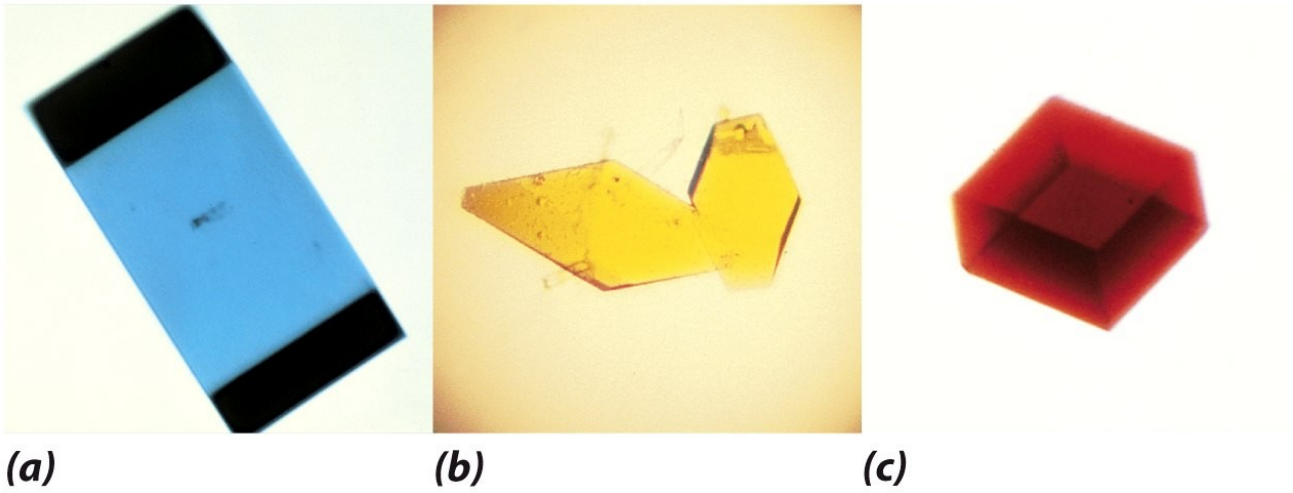
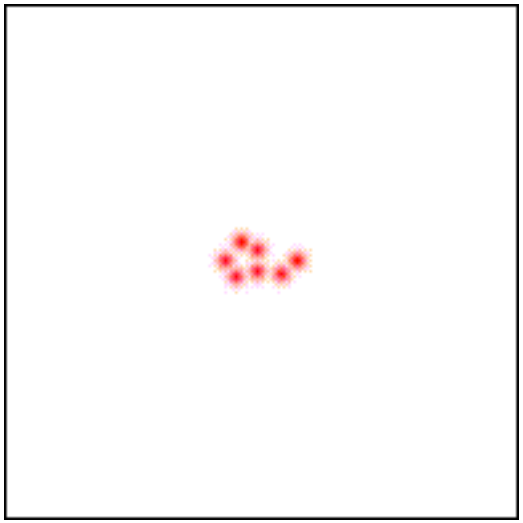


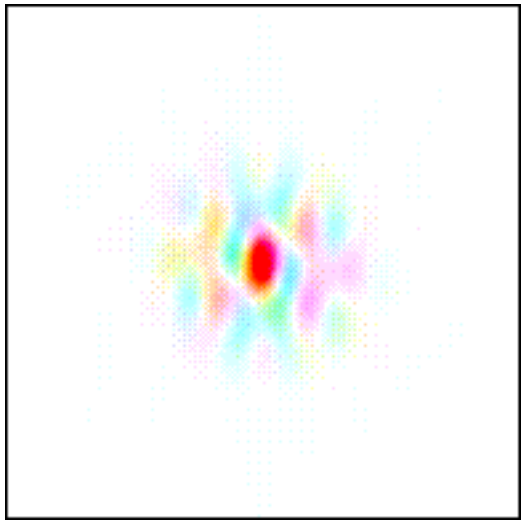
Figure 6-21 Fundamentals of Biochemistry, 2/e

Collect x-ray diffraction data

- **A well ordered crystal will diffract a monochromatic beam of x-rays**
- **Crystals must be harvested and prepared (stabilization, cryo preservation) for data collection**
- **Goal of data collection is a complete set of high quality (accuracy and resolution) diffraction intensities**

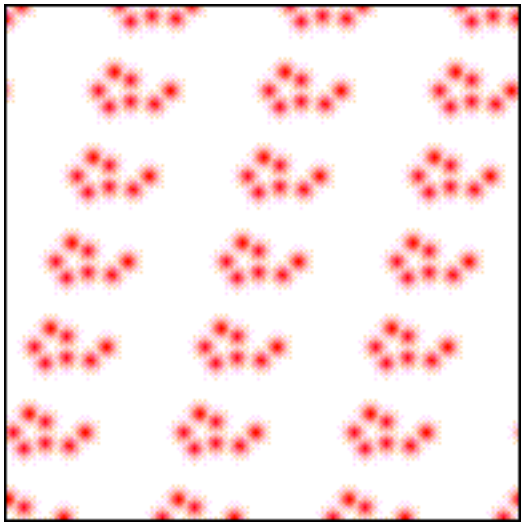
X-ray diffraction

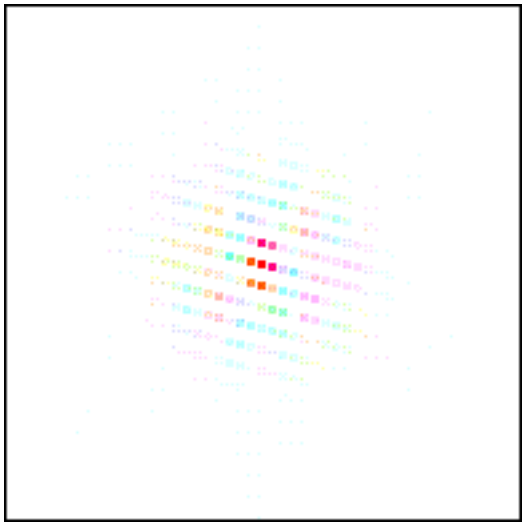




A diffraction pattern is a convolution of the continuous scattering of radiation by a molecule…

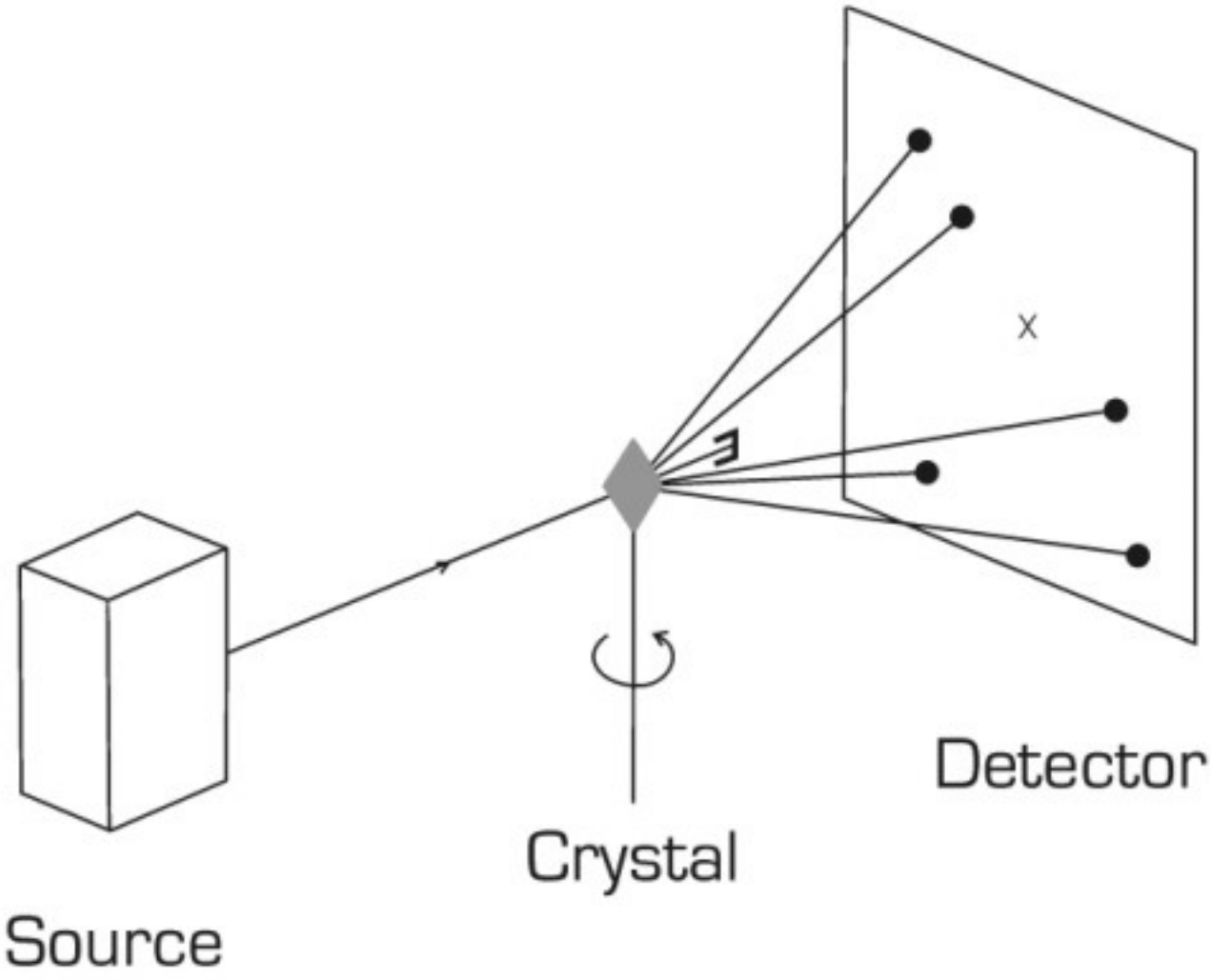
X-ray diffraction





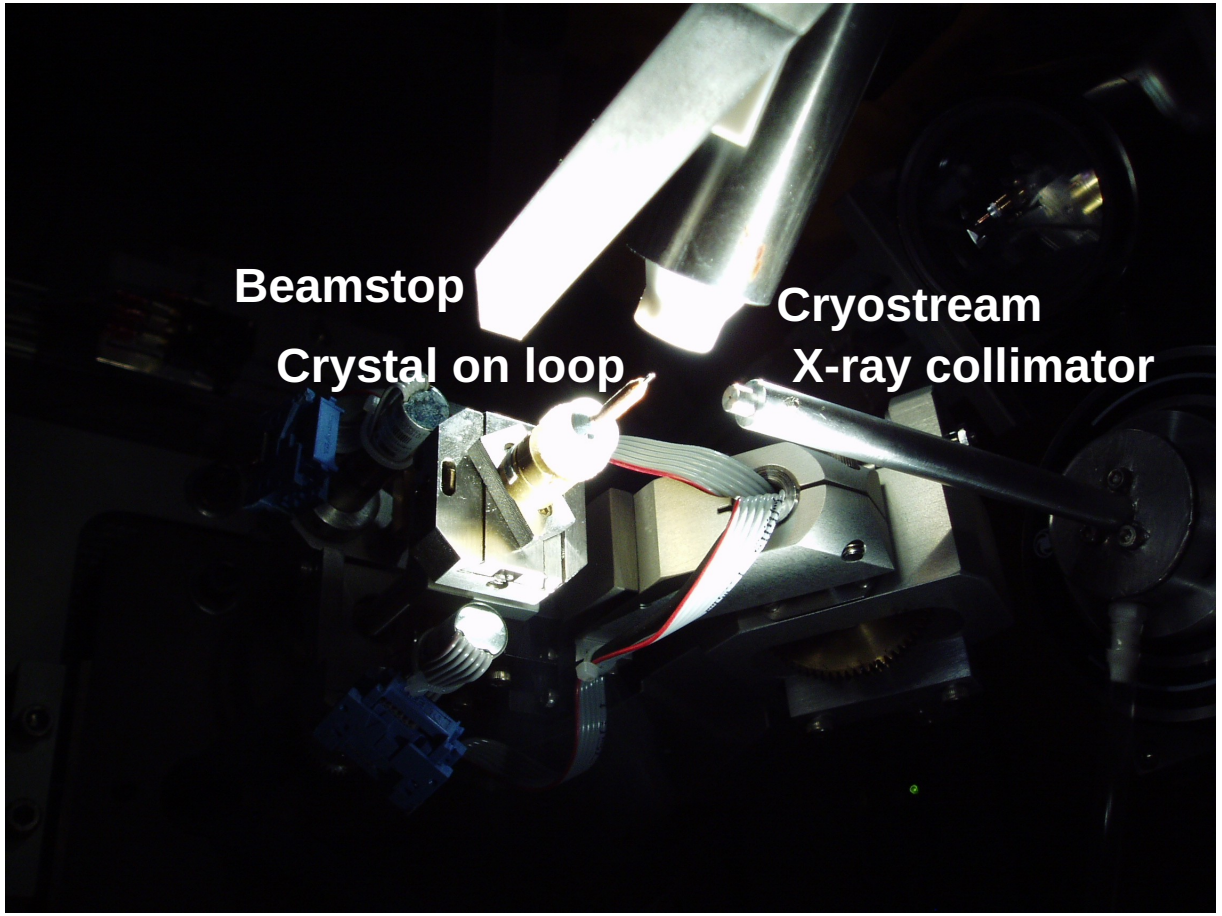
A diffraction pattern is a convolution of the continuous scattering of radiation by a molecule…
…sampled at discrete points in reciprocal space

Collect x-ray diffraction data

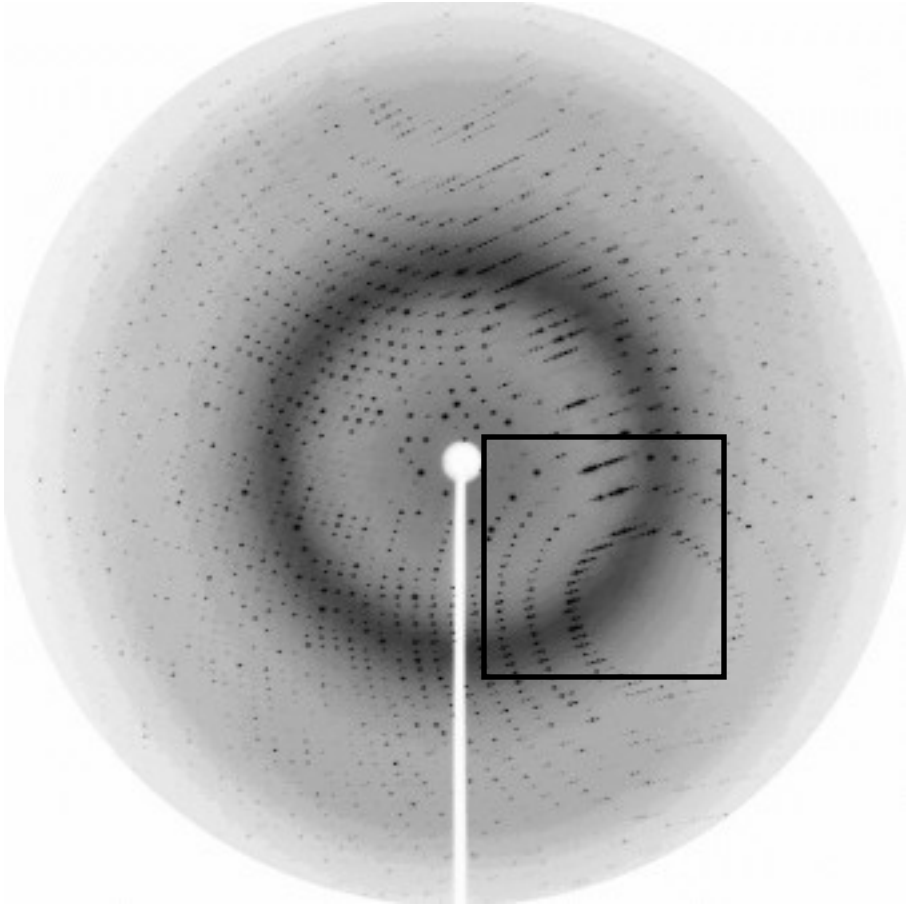


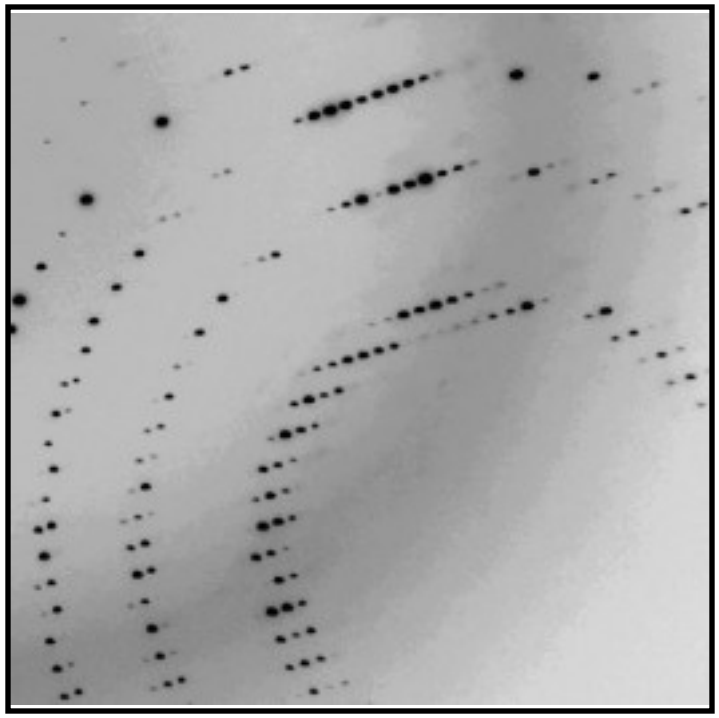
X-ray diffraction data collection. Improved data resolution correlates with greater distance from the center of the x-ray beam.

Grow single crystals of subject



Collect x-ray diffraction data





This data was collected in oscillation mode. The crystal was rotated by 1° as it was exposed to the x-ray beam and all of the diffracted rays were measured on an image plate detector...

X-ray diffraction data processing

- **X-ray data is a stack of digital images covered with reflections (spots)**
- **Each reflection can be given a three digit address (*hkl*) that relates to the size of the crystal unit cell**
- **Each reflection also exhibits a distinctive intensity value**

X-ray diffraction data processing

```
184.079   184.079   184.079    90.000    90.000    90.000 p23
  3   1   2  1832.3    29.3
  3   2   1   513.4     8.3
  3   2   2   574.8     8.5
  3   3   0   628.3    13.1
  3   3   1  1221.4    17.9
  3   3   2   170.1     2.6
  3   3   3  2134.3    54.0
  4   0   0   722.7    21.4
  4   0   1   186.3     4.3
  4   0   2   202.5     5.3
  4   0   3     2.7     0.3
  4   1   0   247.2     4.9
  4   1   1  3000.6    50.5
  4   1   2     1.5     0.1
  4   1   3    25.7     0.5
  4   2   0     0.7     0.1
  4   2   1   603.1     9.4
  4   2   2   561.1     8.3
  4   2   3    82.9     1.3
  4   3   0   118.8     2.6
  4   3   1   134.7     2.1
  4   3   2    50.7     0.8
  4   3   3   345.1     5.2
  4   4   0   245.2     5.2
```

The output of a successful diffraction experiment is an indexed list of diffraction intensities. The space group symmetry and unit cell geometries are derived from the reflection (spot) positions. Once a unique Miller index (h,k,l) is assigned to each reflection (spot) then a value for intensity is measured. The values on the left are the unit cell constants (a,b,c,α,β,γ) and symmetry (P23). Then the values of intensity are listed for the first 24 reflections

Structure solution

- An electron density map can be built as a summation of structure factors (F_{hkl})--That's all there is to it!
- F_{hkl} is a vector quantity with both magnitude (amplitude) and direction (phase)
- The amplitude, $|F_{hkl}|$ is equal to the square root of the measured intensity I_{hkl}
- The phase (an angular value between 0 and 2π) of F_{hkl}, however, is not directly measured in a diffraction experiment

Solve the "phase problem"

The goal: an equation that describes the electron denisty that repeats itself periodically throughout the crystal.

$$\rho(x,y,z) = (1/V)\Sigma \boldsymbol{F}_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Where ρ is electron density in real space (x,y,z) and
F is the "structure factor" for each reflection (spot) hkl

Solve the "phase problem"

The goal:

$$\rho(x,y,z) = (1/V)\Sigma \boldsymbol{F}_{hkl} e^{-2\pi i(hx+ky+lz)}$$

In a diffraction experiment, we measure $\boldsymbol{I}_{hkl}$.

$$SQRT(\boldsymbol{I}_{hkl}) = |\boldsymbol{F}_{hkl}|$$

$$\boldsymbol{F}_{hkl} = |\boldsymbol{F}_{hkl}| e^{i\alpha hkl}$$

So, all we need is α for each spot and we can build an electron density map!

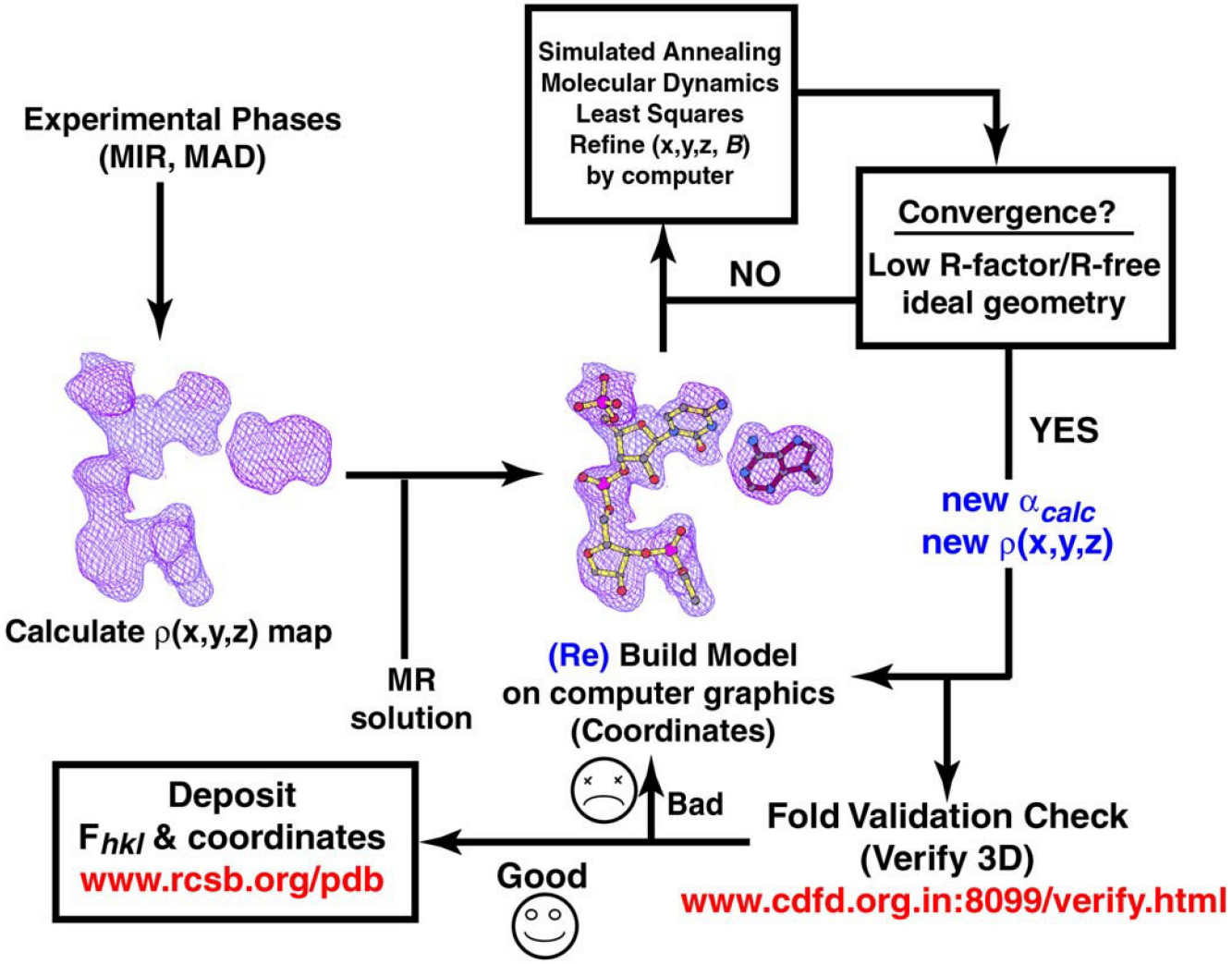
Solve the "phase problem"

- An initial estimate of α_{hkl}, or the "phase"of each reflection intensity, must be determined experimentally by one of three methods
- Multiple isomorphous replacement (MIR)
  - Crystals that are chemically identical but contain heavy atoms (derivatives) are used to determine initial estimates for phase
- Molecular replacement (MR)
  - A structure of a related protein is fitted into the experimental unit cell and estimated phase values are calculated from the model
- Multiwavelength anomalous dispersion (MAD)
  - A crystal containing "anomalous scatterers" is shot at different wavelengths to measure an accurate estimate of phase

Build and refine model

- **With a good estimate of phases and accurately measured intensities one can construct a high quality electron density map**
- **One then builds the amino acid sequence into the map**
- **The iterative process of refinement involves improving phases to generate maps that correlate better with data**

Crystallographic Refinement Flowchart



R-factor

- After each cycle of model building and refinement, a calculation of the error in the observed and calculated structure factors is measured.

- $$R = \frac{\Sigma \| |F_{obs}| - |F_{calc}| \|}{\Sigma |F_{obs}|}$$

This is known as the R-factor

Problems with R-factors

- During refinement, a crystallographer can “cheat” but putting atoms in every single blob of electron density that appears in a map whether or not it really represents anything in the crystal
- This quickly raises the number of fitted parameters (atomic positions) relative to the constant number of observables (reflections) and drives down R-factors

R-free

- Prior to any refinement a significant amount (5-10%) of reflection data are removed at random and not used during refinement
- The addition of new atoms and/or changes in model structure is accepted only if it results in a lowering of R-factor measured against both the working set (R-work) and the test set (R-free) of reflections

Build and refine model

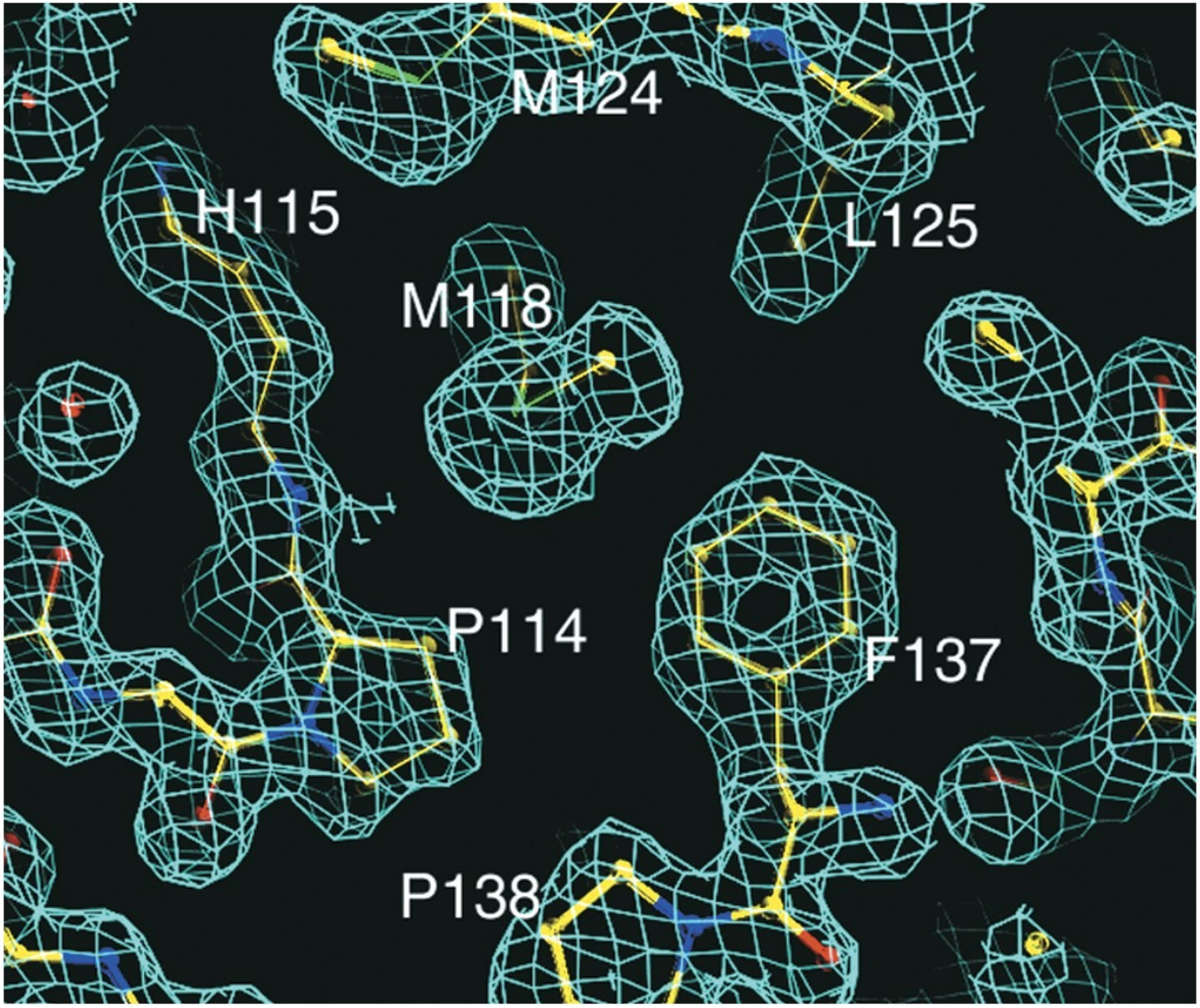


Figure 6-23 Fundamentals of Biochemistry, 2/e

Build and refine model

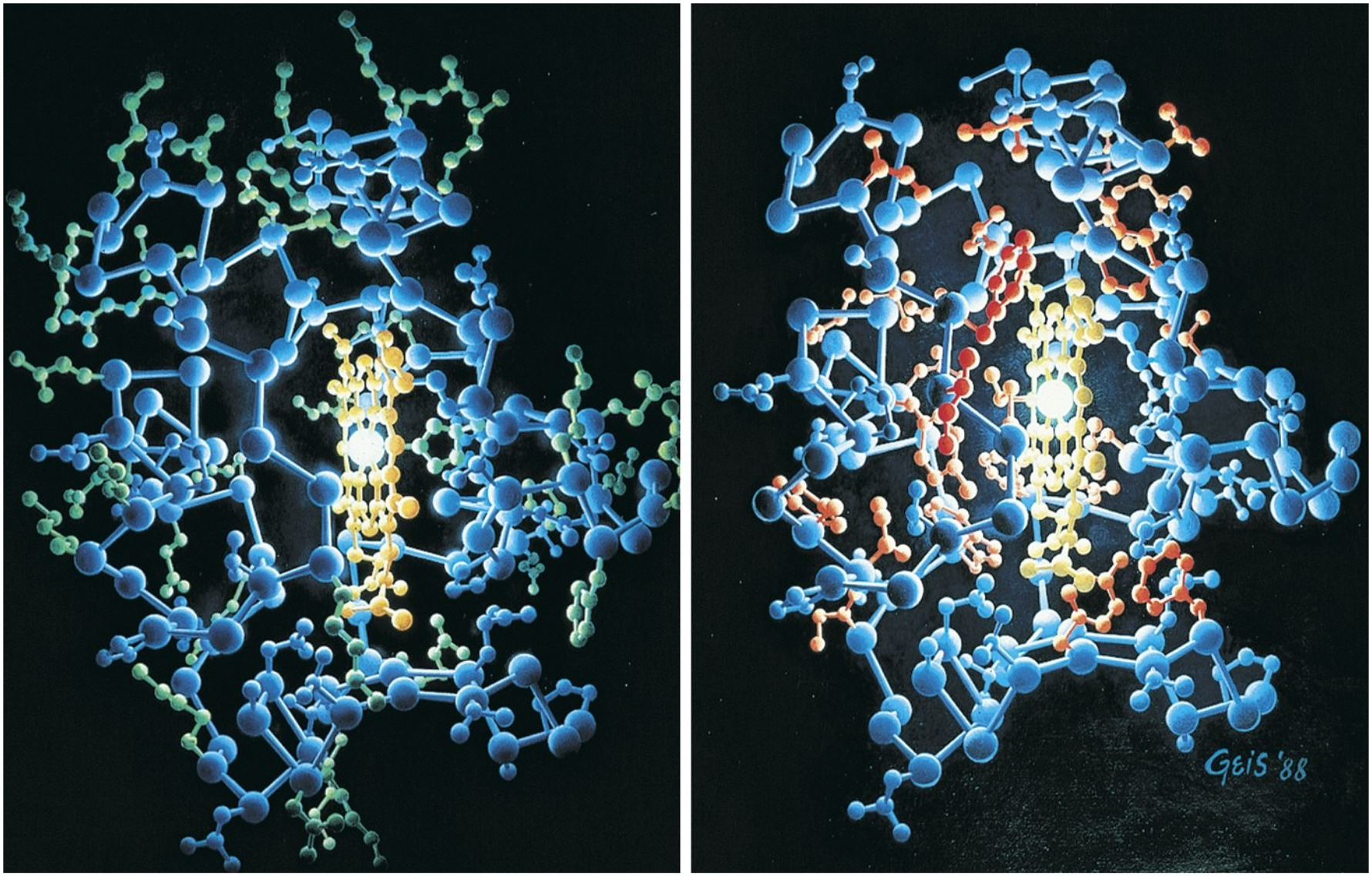


Figure 6-27 Fundamentals of Biochemistry, 2/e

Analyze structure

- **What does the structure suggest about function? Mechanism of biological action?**
- **What testable hypotheses are immediately suggested by the structure?**
- **Get to work!**

Where computation can improve on current methods

- **Streamline the entire process-allow programs to talk to one another**
  - **PHENIX**

- **Computational methods for solving phase problem**
  - **Direct methods**
  - **Sulfur anomalous scattering**
  - **Structure prediction as a means of generating molecular replacement probes**

Read more about it

- **Many excellent web-based tutorials**
  - **http://phillips-lab.biochem.wisc.edu/tools.html**
  - **http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html**

- **Gale Rhodes, “Crystallography Made Crystal Clear” Academic Press**