

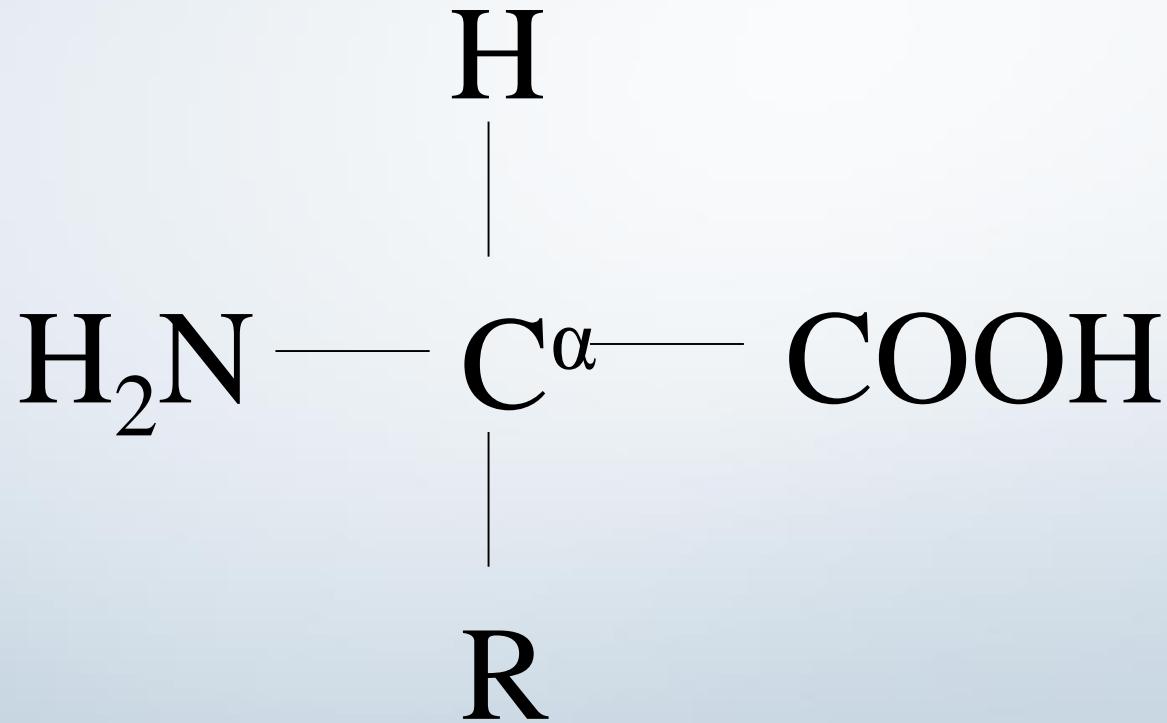


Protein Structure Basics

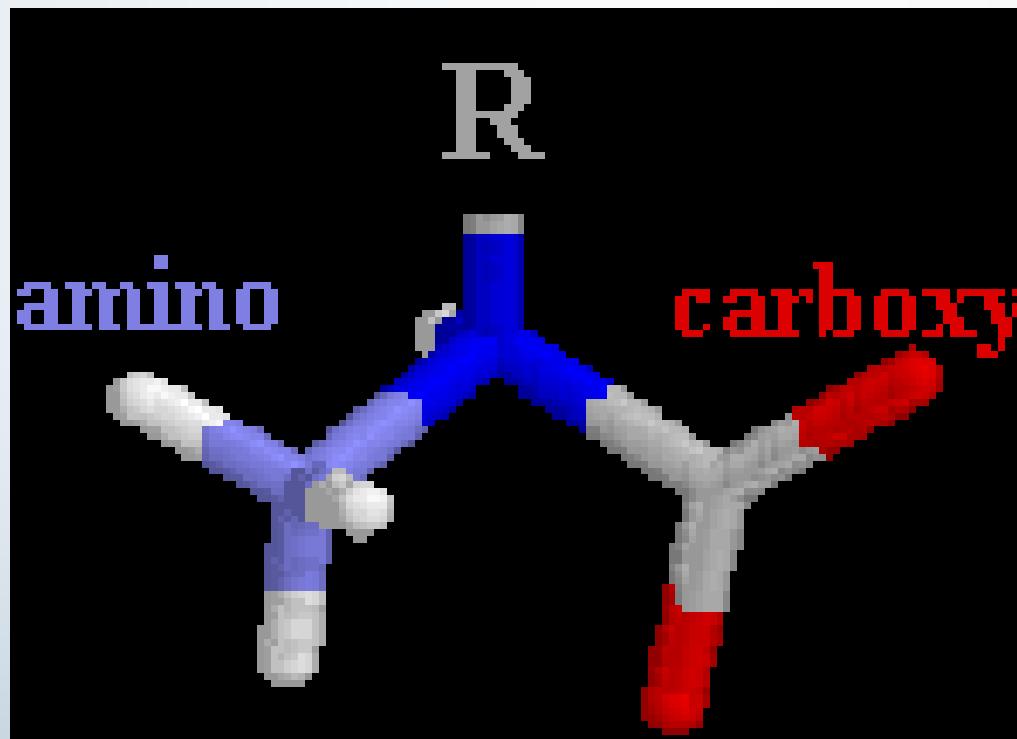
Yazdan Asgari

2020

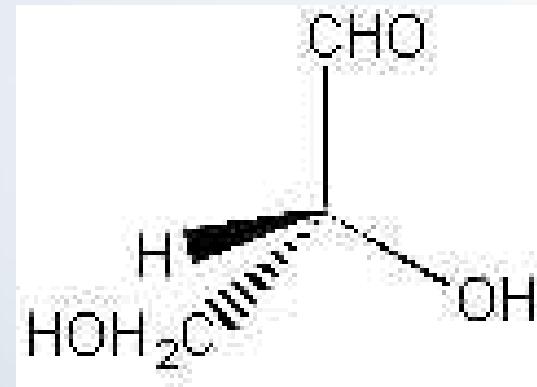
General Amino Acid Structure



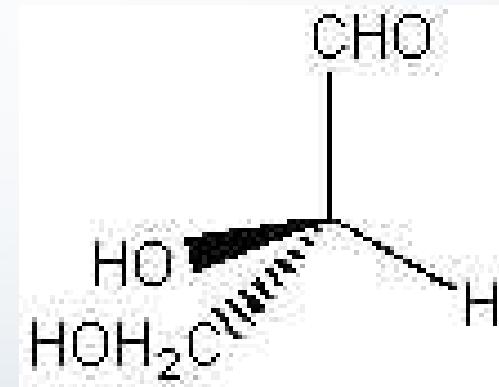
General Amino Acid Structure



Chirality: Glyceraldehyde

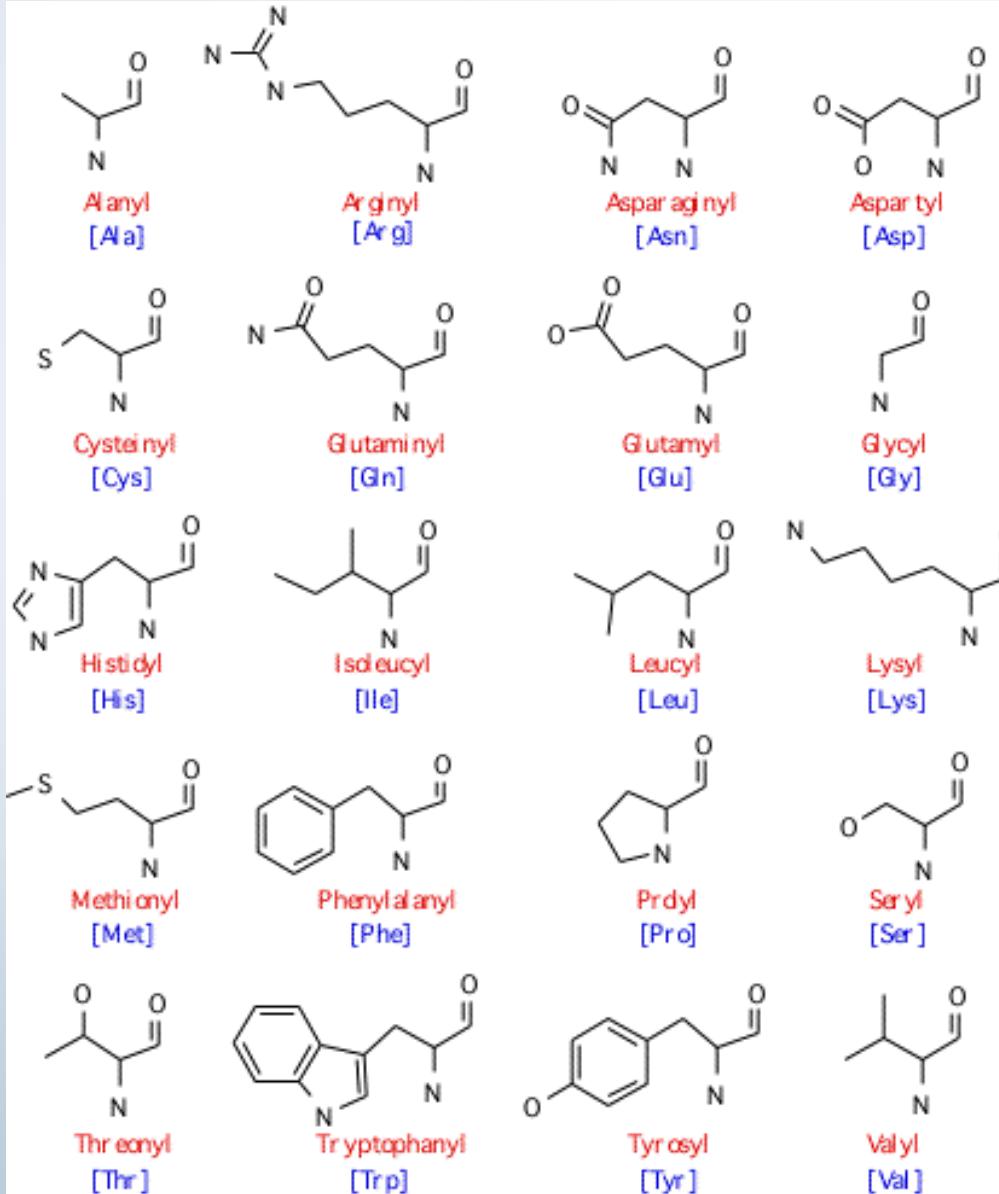


D-glyceraldehyde

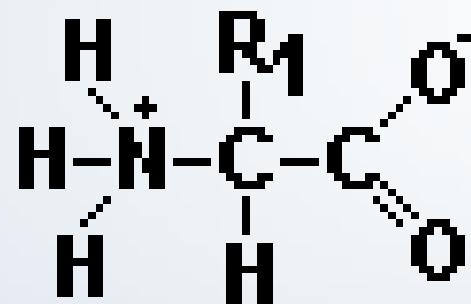


L-glyceraldehyde

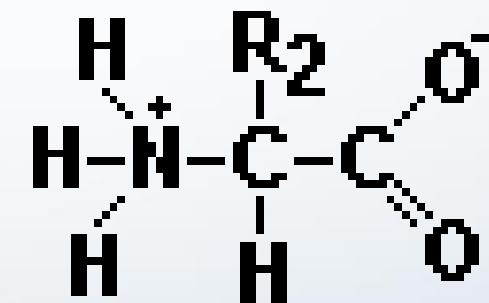
20 Naturally-occurring Amino Acids



Peptide Bond Formation

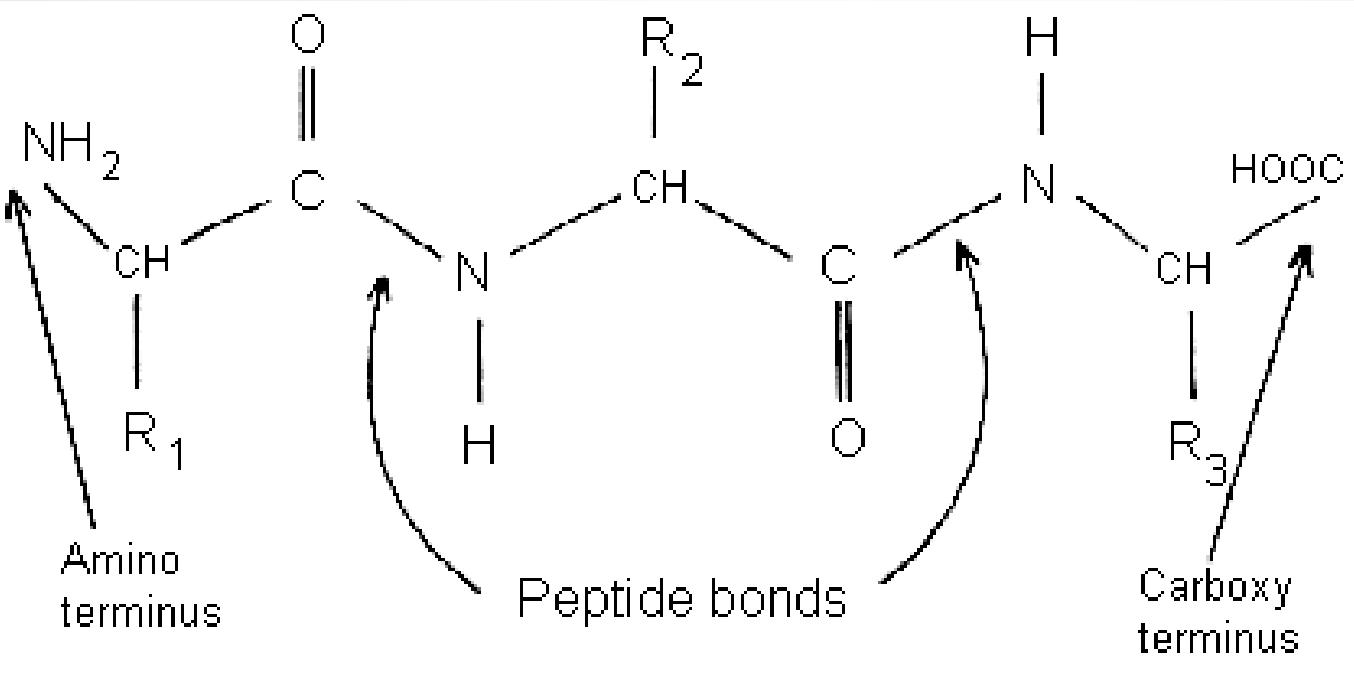


amino acid 1



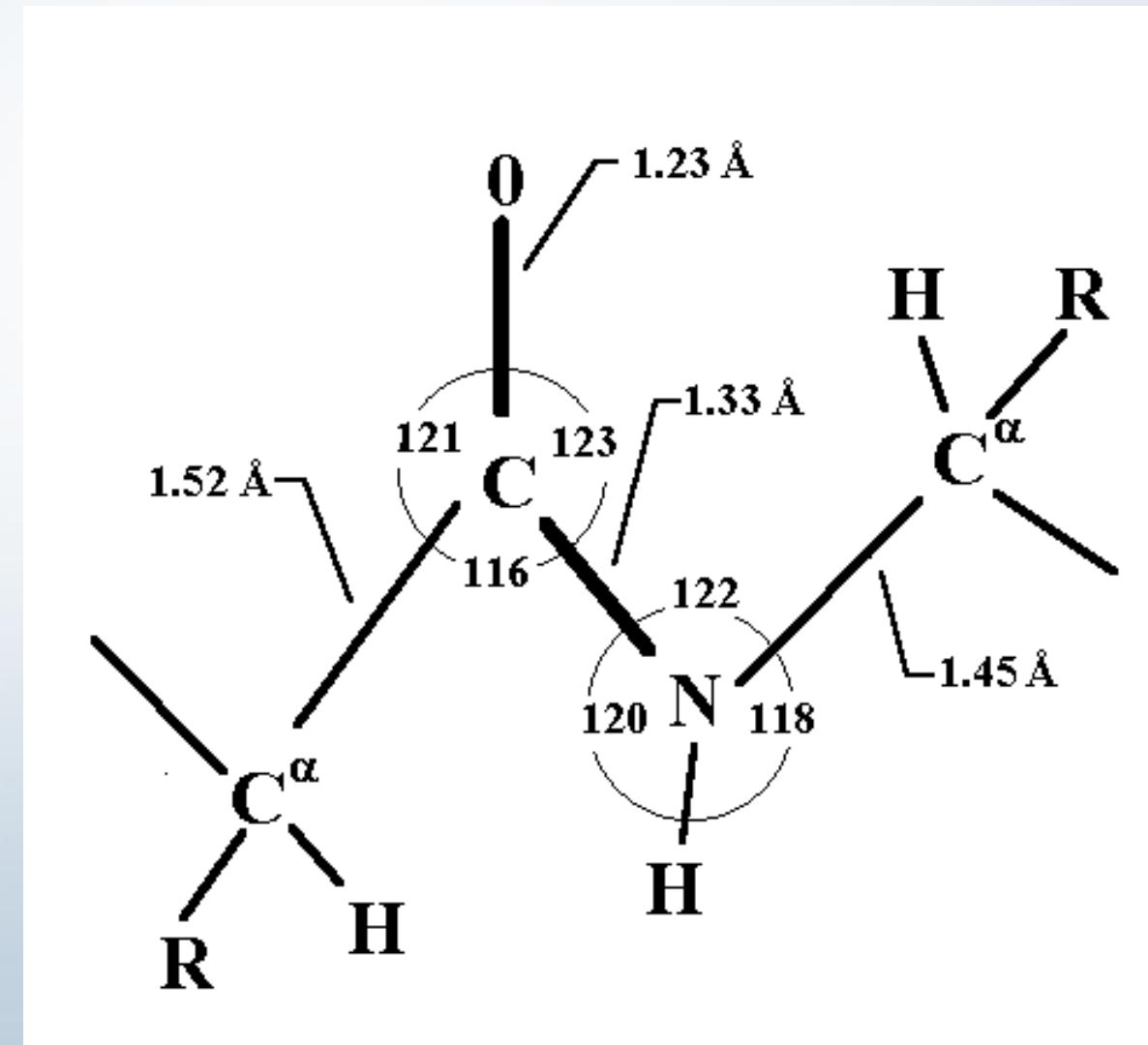
amino acid 2

Peptide Chain



Peptide Bond

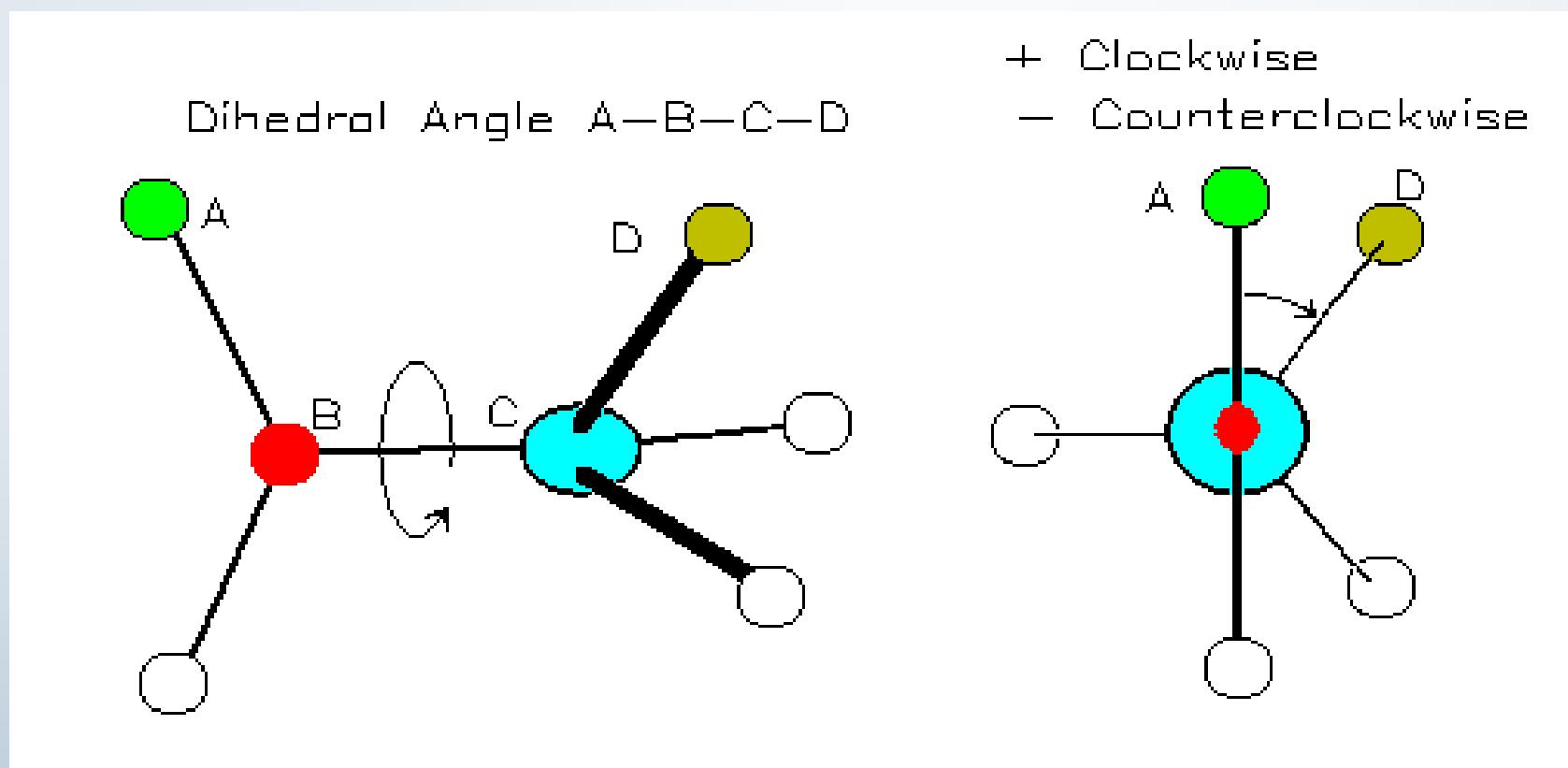
- Joins amino acids
- 40% double bond character
 - Caused by resonance
 - Results in shorter bond length
 - Double bond disallows rotation
- Bond rotation determines protein folding, 3D structure



Protein Conformation Framework

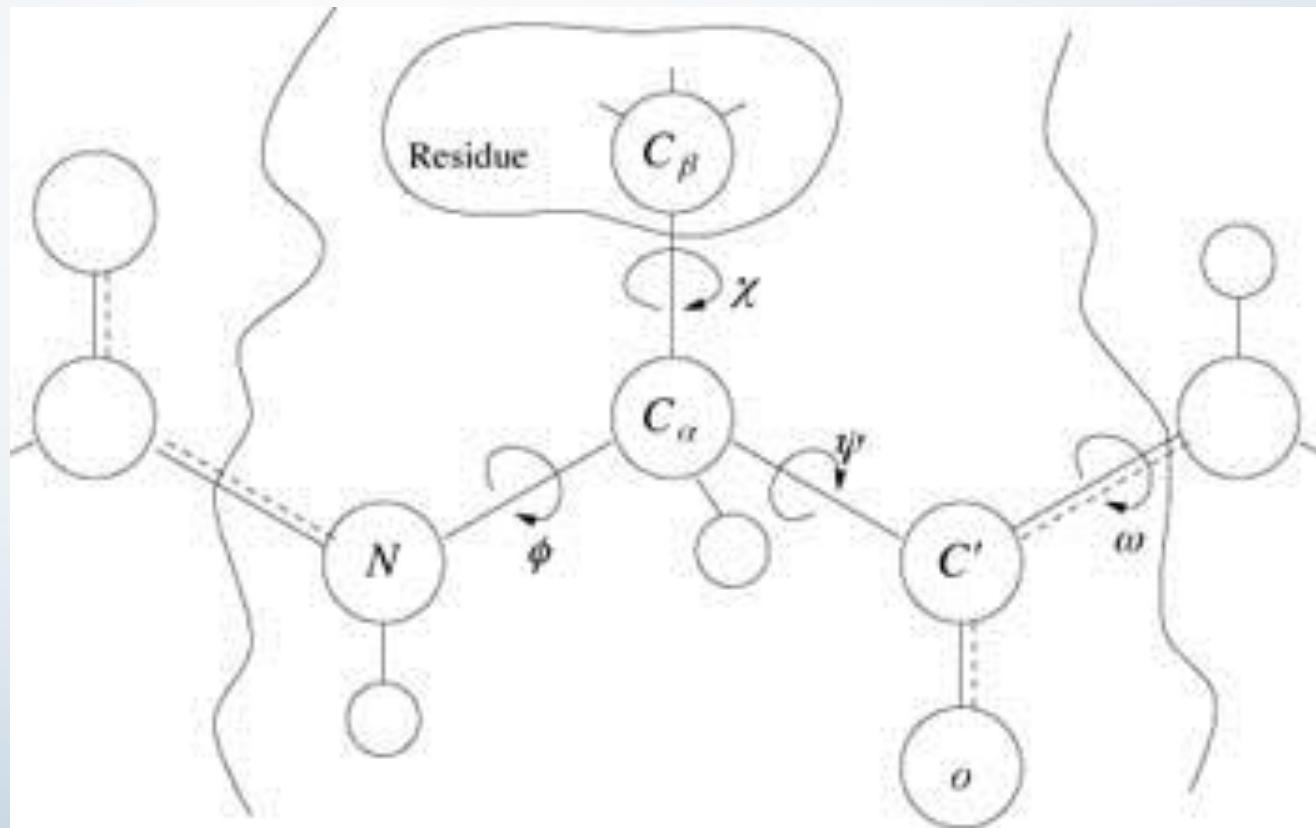
- Bond rotation determines protein folding, 3D structure
- Torsion angle (dihedral angle) τ
 - Measures orientation of four linked atoms in a molecule: A, B, C, D
 - τ_{ABCD} defined as the angle between the normal to the plane of atoms A-B-C and normal to the plane of atoms B-C-D
 - Three repeating torsion angles along protein backbone: ω , φ , ψ

Protein Conformation Framework



Backbone Torsion Angles

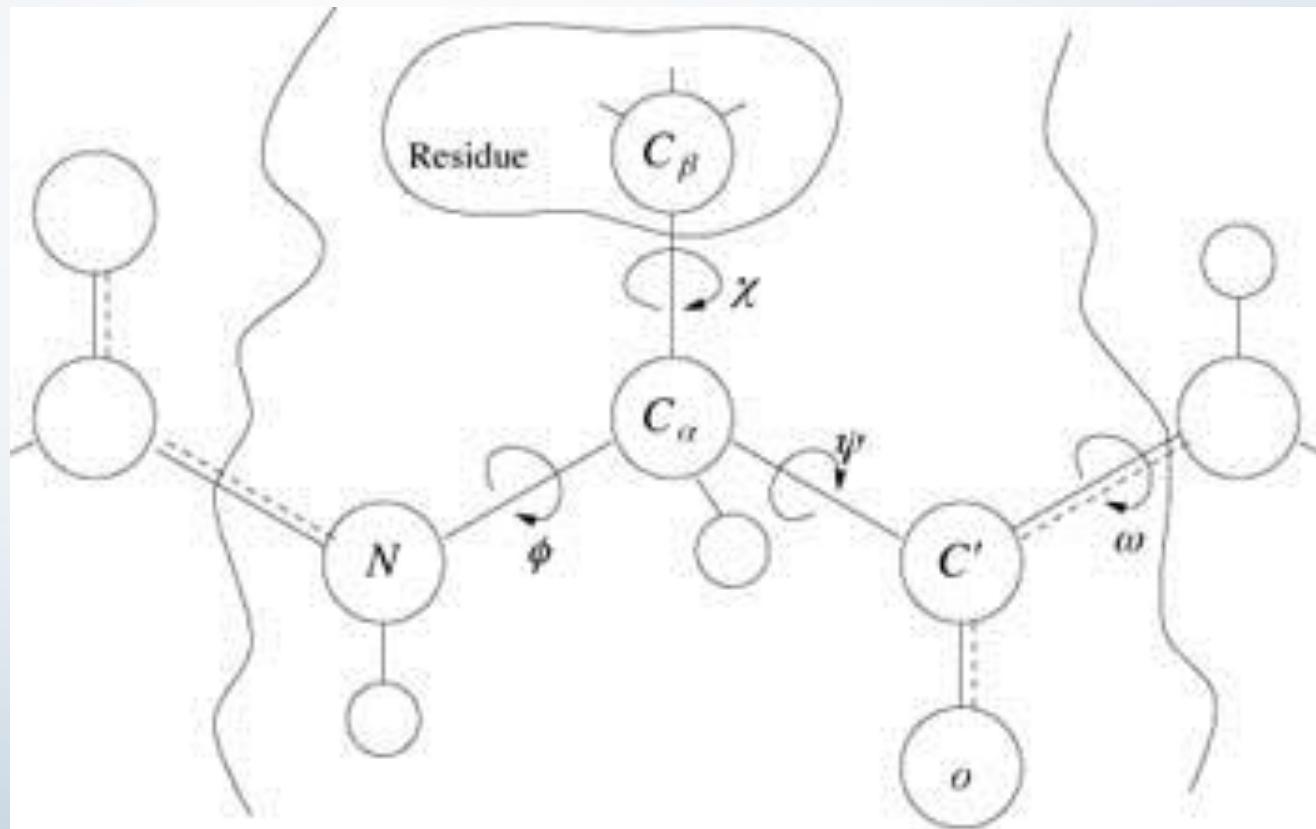
- Dihedral angle ω : rotation about the peptide bond, namely $C^\alpha_1-\{C-N\}-C^\alpha_2$



Backbone torsion angles of a protein

Backbone Torsion Angles

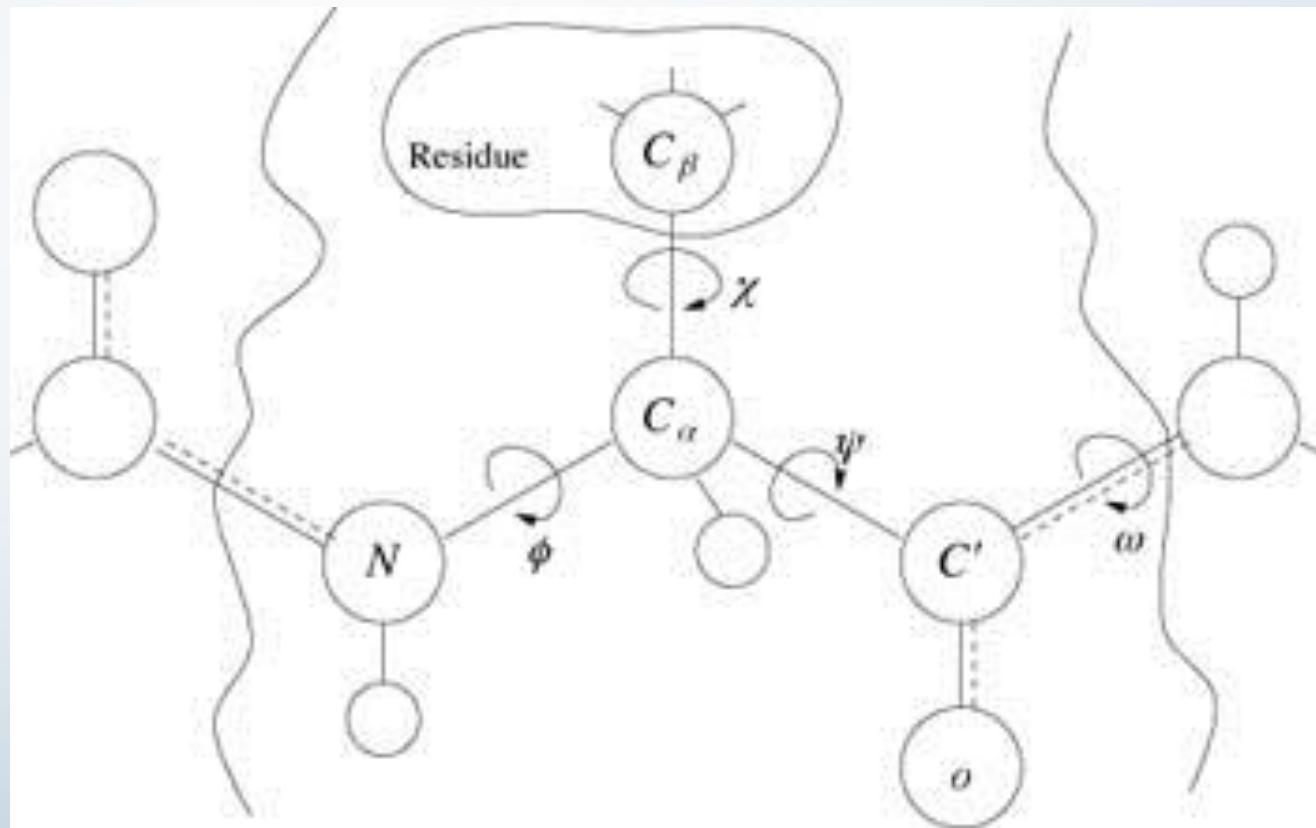
- Dihedral angle ϕ : rotation about the bond between N and C^α



Backbone torsion angles of a protein

Backbone Torsion Angles

- Dihedral angle ψ : rotation about the bond between C^α and the carbonyl carbon

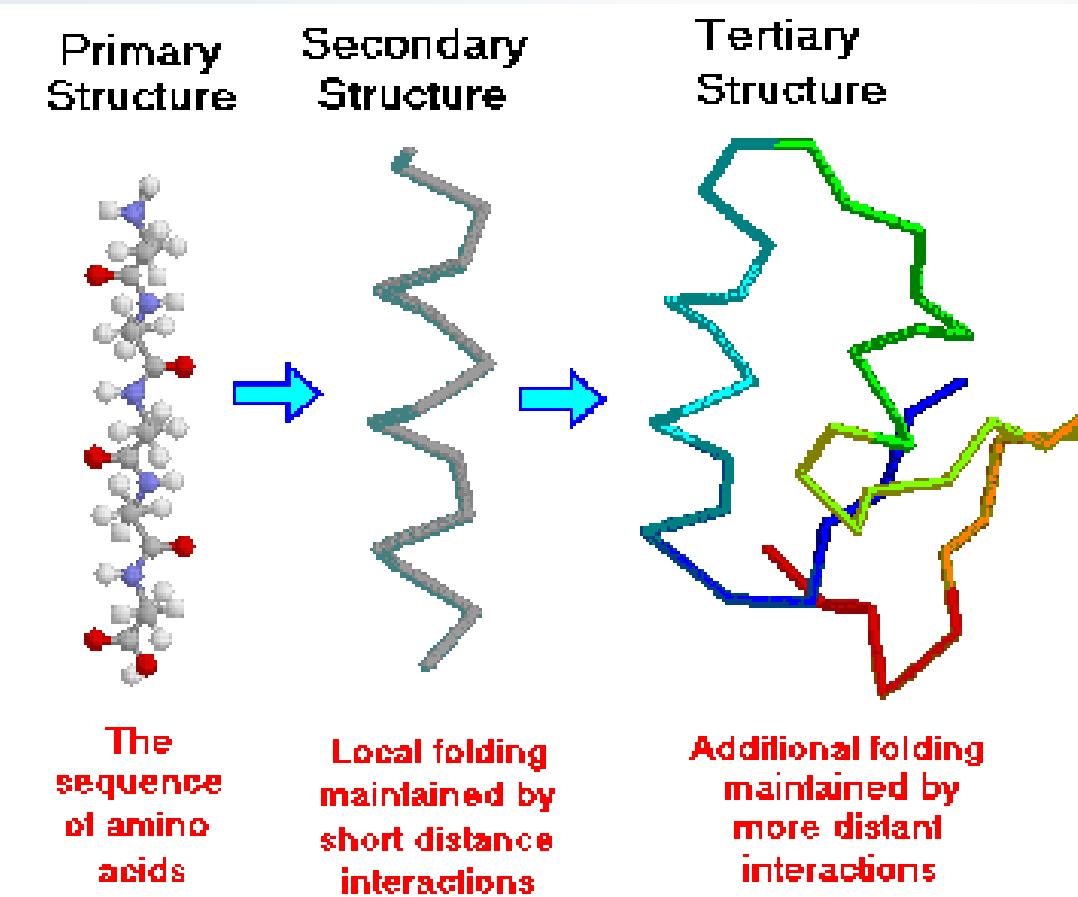


Backbone torsion angles of a protein

Backbone Torsion Angles

- ω angle tends to be planar (0° - cis, or 180° - trans) due to delocalization of carbonyl pi electrons and nitrogen lone pair
- φ and ψ are flexible, therefore rotation occurs here

Protein Structure



Backbone Torsion Angles

- However, ϕ and ψ of a given amino acid residue are limited due to steric hindrance
- Only 10% of the area of the $\{\phi, \psi\}$ space is generally observed for proteins
- First noticed by G.N. Ramachandran

G.N. Ramachandran

- Used computer models of small polypeptides to systematically vary ϕ and ψ with the objective of finding stable conformations
- For each conformation, the structure was examined for close contacts between atoms
- Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii
- Therefore, ϕ and ψ angles which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone

Regular Secondary Structure Conformations

Secondary Structure Element	Torsional Angle (°)		Residue/turn	Translational distance per residue (Å)
	ϕ	ψ		
α helix	-57	-47	3.6	1.50
3_{10} helix	-49	-26	3.0	2.00
π helix	-57	-70	4.4	1.15
Parallel β strand	-139	+135	2.0	3.20
Antiparallel β strand	-119	+113	2.0	3.40
Poly(Pro) I	-83	+158	3.3	1.90
Poly(Pro) II	-78	+149	3.0	3.12

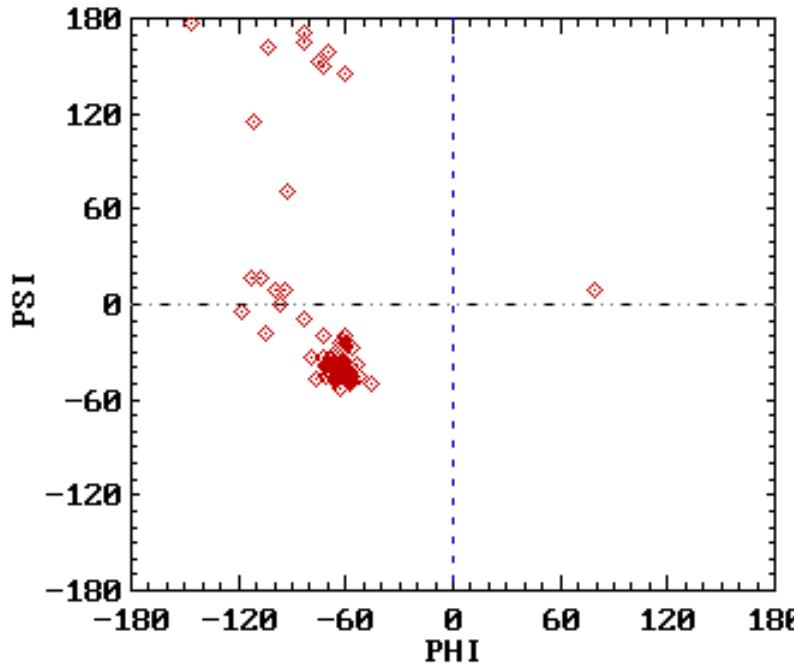
Ramachandran Plot

- Plot of ϕ vs. ψ
- Repeating values of ϕ and ψ along the chain result in regular structure
- For example, repeating values of $\phi \sim -57^\circ$ and $\psi \sim -47^\circ$ give a right-handed helical fold (the alpha-helix)
- The structure of cytochrome C-256 shows many segments of helix and the Ramachandran plot shows a tight grouping of ϕ , ψ angles near -50, -50

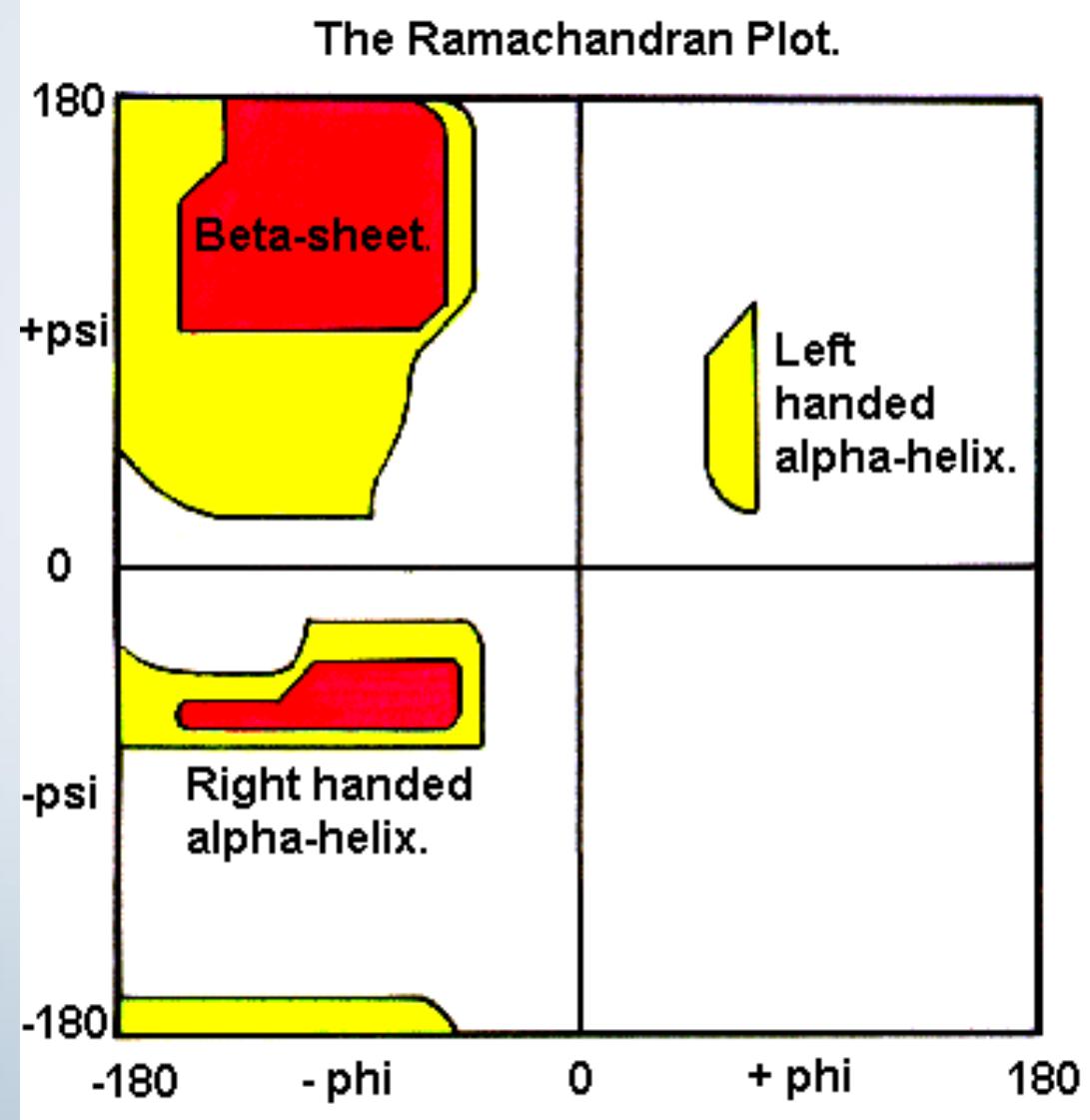
The structure of cytochrome C-256 shows many segments of helix and the Ramachandran plot shows a tight grouping of φ , ψ angles near -50,-50



alpha-helix



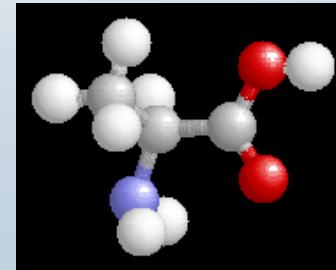
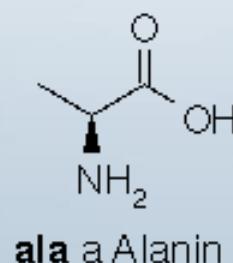
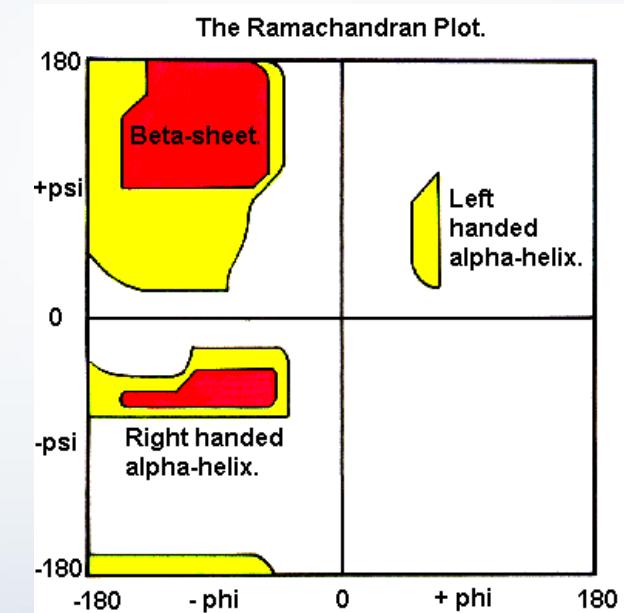
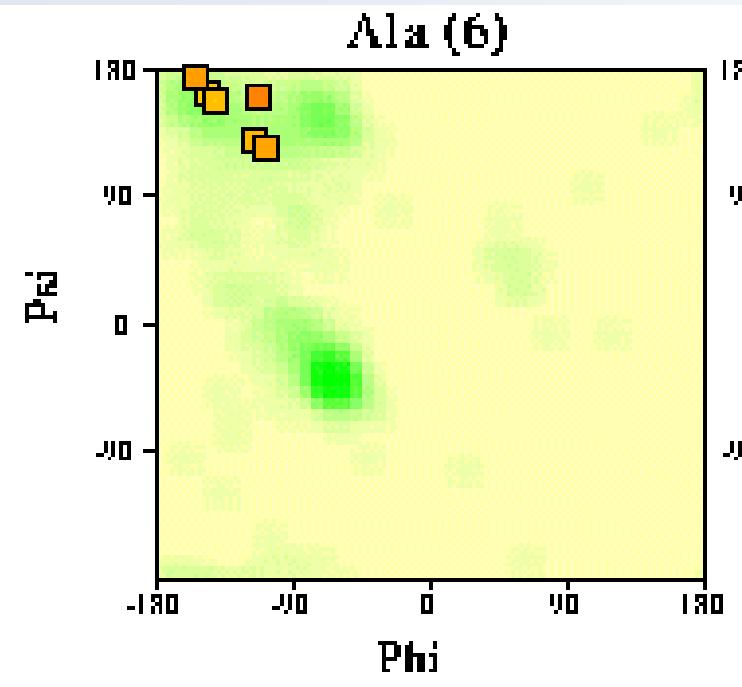
cytochrome C-256
Ramachandran plot



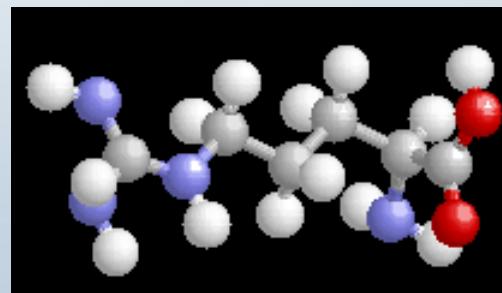
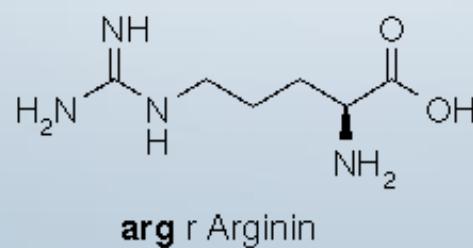
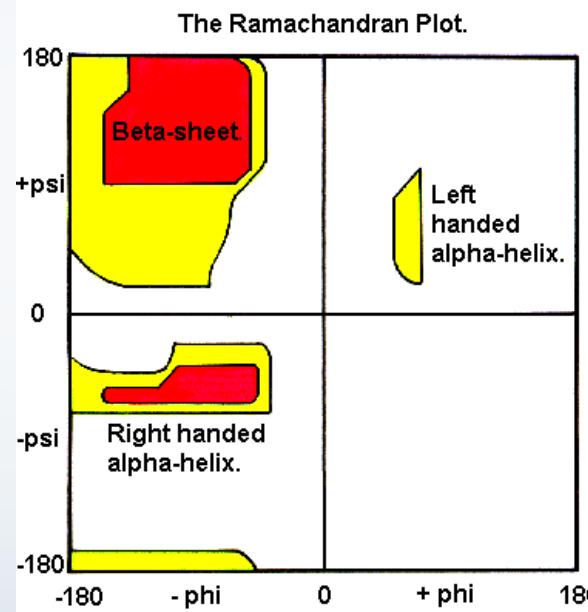
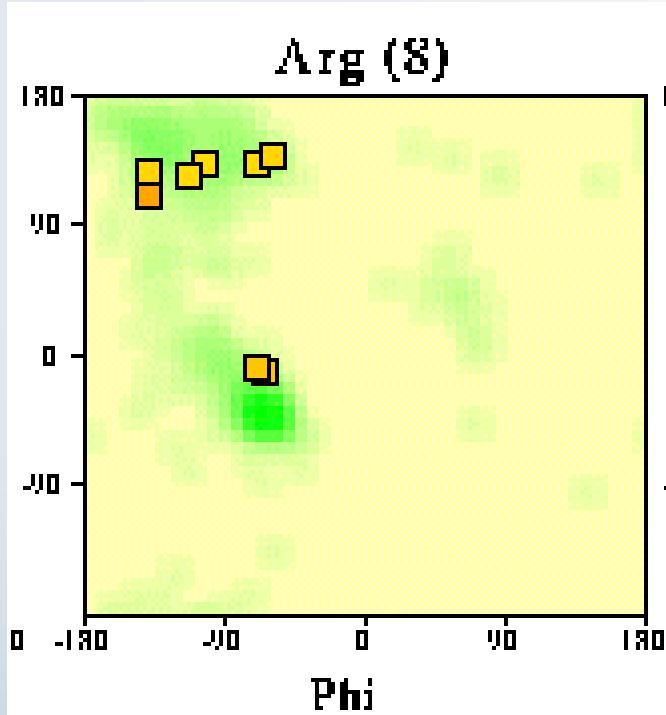
Ramachandran Plot

- White = sterically disallowed conformations (atoms in the polypeptide come closer than the sum of their van der Waals radii)
- Red = sterically allowed regions (namely right-handed alpha helix and beta sheet)
- Yellow = sterically allowed if shorter radii are used (i.e. atoms allowed closer together; brings out left-handed helix)

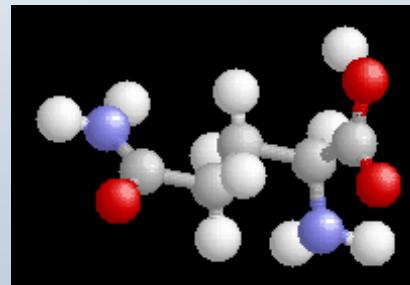
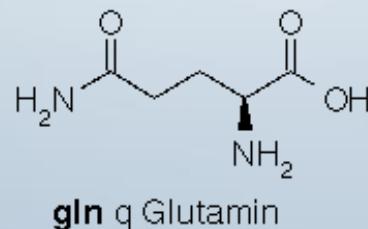
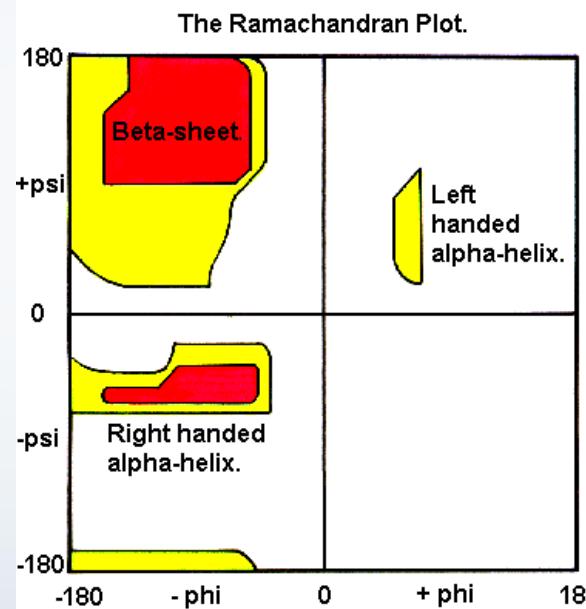
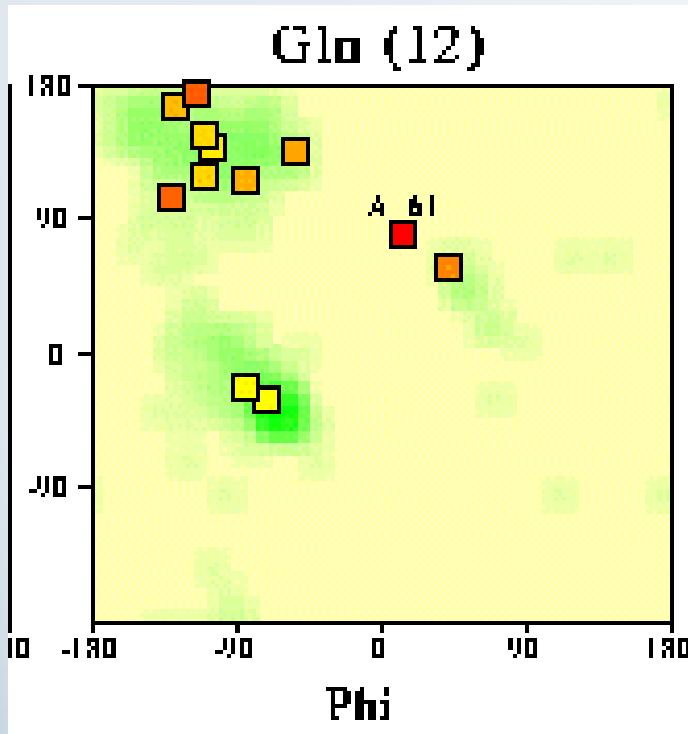
Alanine Ramachandran Plot



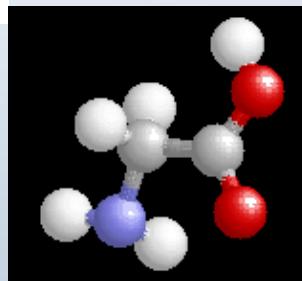
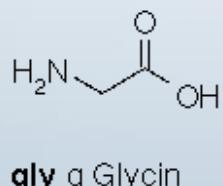
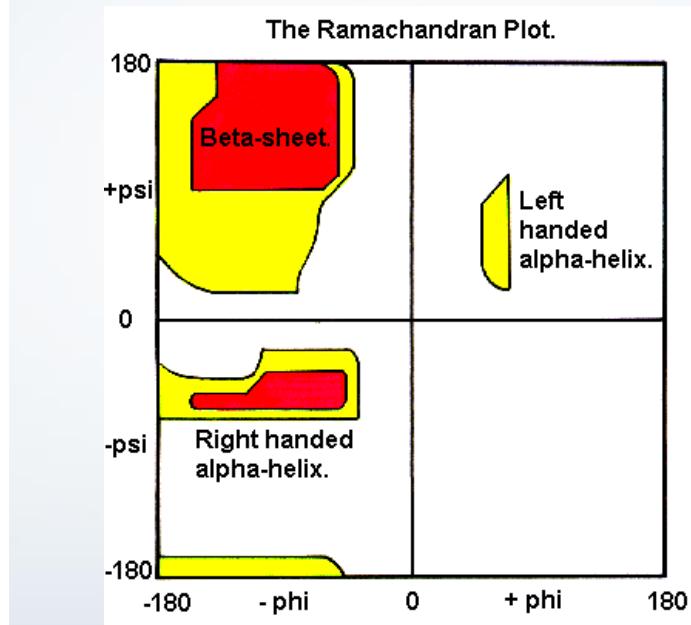
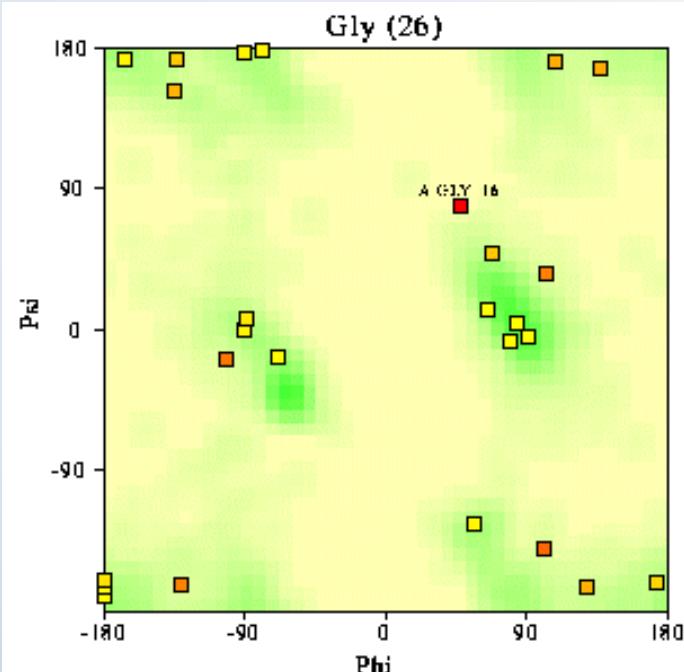
Arginine Ramachandran Plot



Glutamine Ramachandran Plot

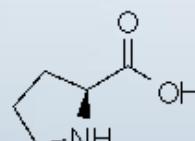
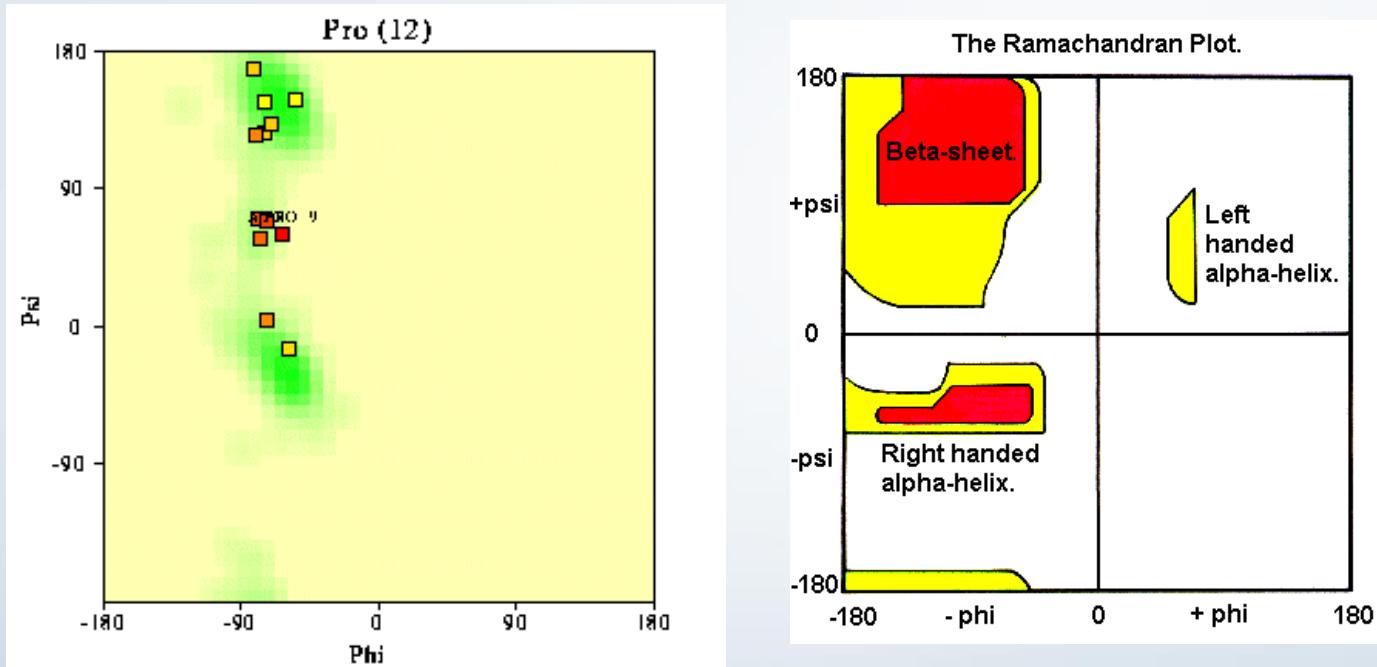


Glycine Ramachandran Plot

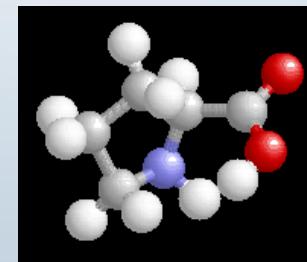


Note more allowed regions due to less steric hindrance

Proline Ramachandran Plot



pro p Prolin

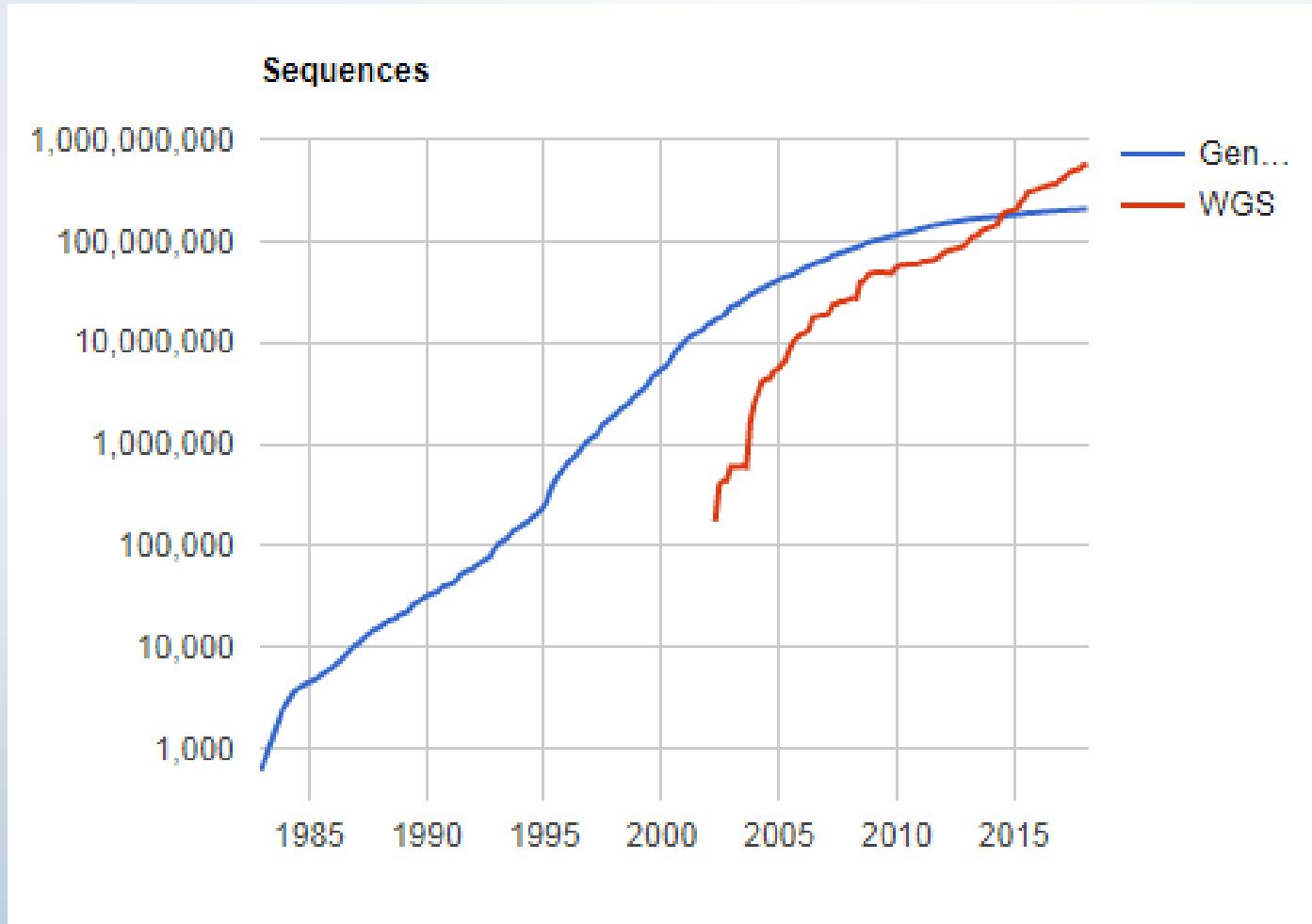


Note less allowed regions due to structure



Macromolecular X-Ray Crystallography

GenBank and WGS Statistics

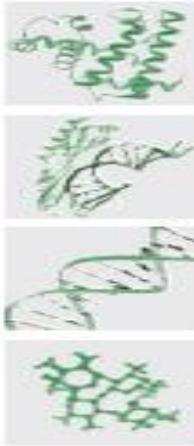


Protein Data Bank: Archive Contents

PDB ARCHIVE CONTENTS ON JULY 1ST, 2016

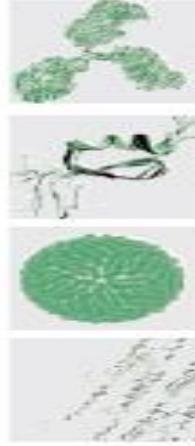
120,057 Released atomic coordinate entries

MOLECULE TYPE



111,440
Proteins, peptides,
and viruses
5,658
Protein/nucleic
acid complexes
2,933
Nucleic acids
26
Other

EXPERIMENTAL METHOD



107,264
X-ray crystallography
11,435
Nuclear Magnetic
Resonance
1,065
3D Electron
Microscopy (3DEM)
197
Other
96
Hybrid

RELATED EXPERIMENTAL DATA FILES

91,864
Structure factors
8,539
NMR restraints
2,297
Chemical shifts
905
3DEM maps

ACCESS



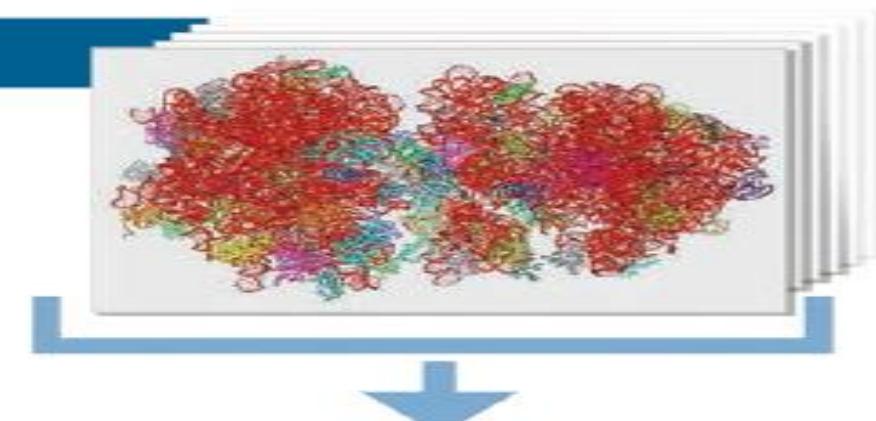
Each month in 2015, rcsb.org
was visited **741,000**
times on average by
315,000 unique visitors



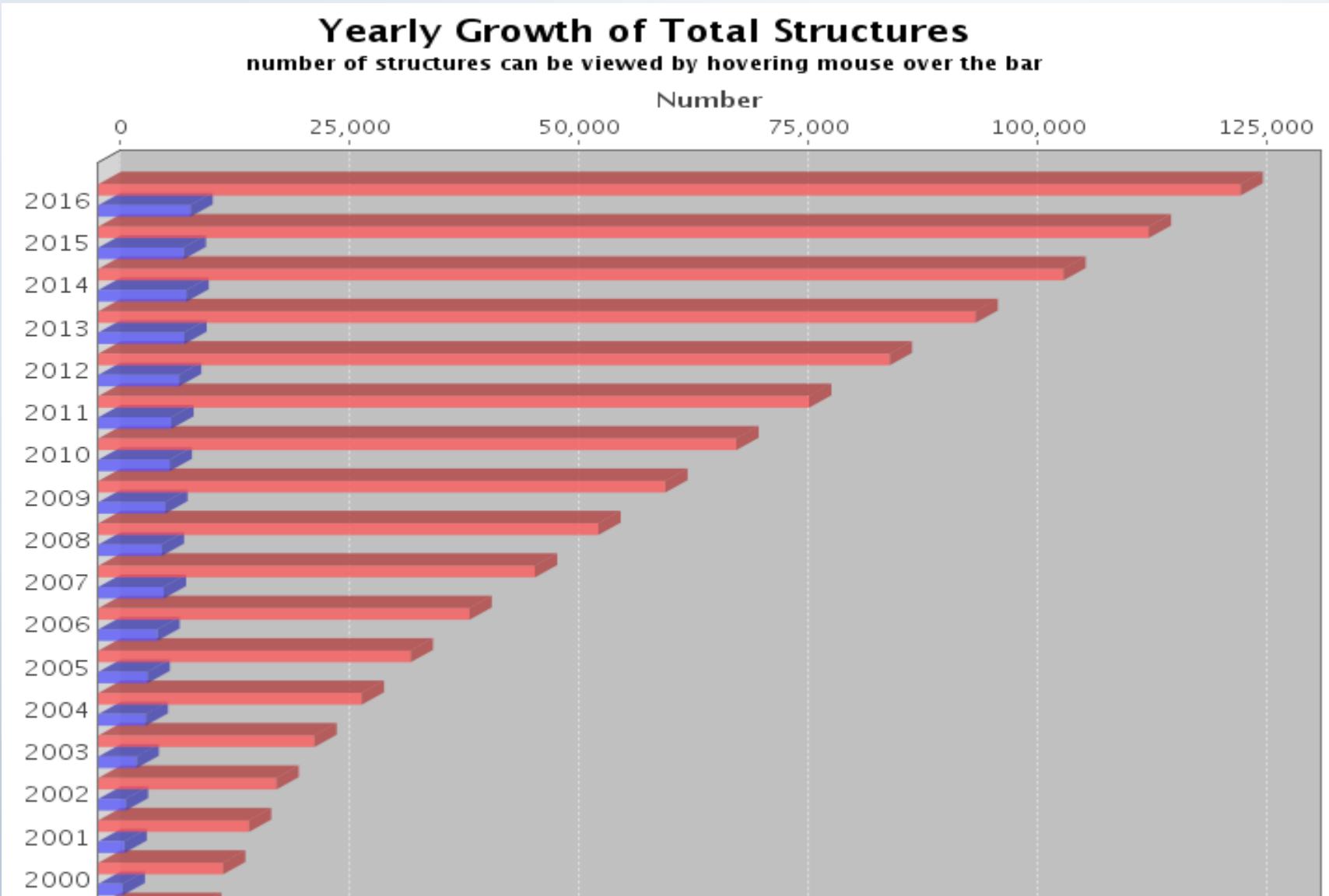
A total of **35,260 GB**
of data were accessed

PDB data are accessed
from **192** of the 195

More than **1.5 million PDB structures**
are downloaded **EVERY day of the year.**



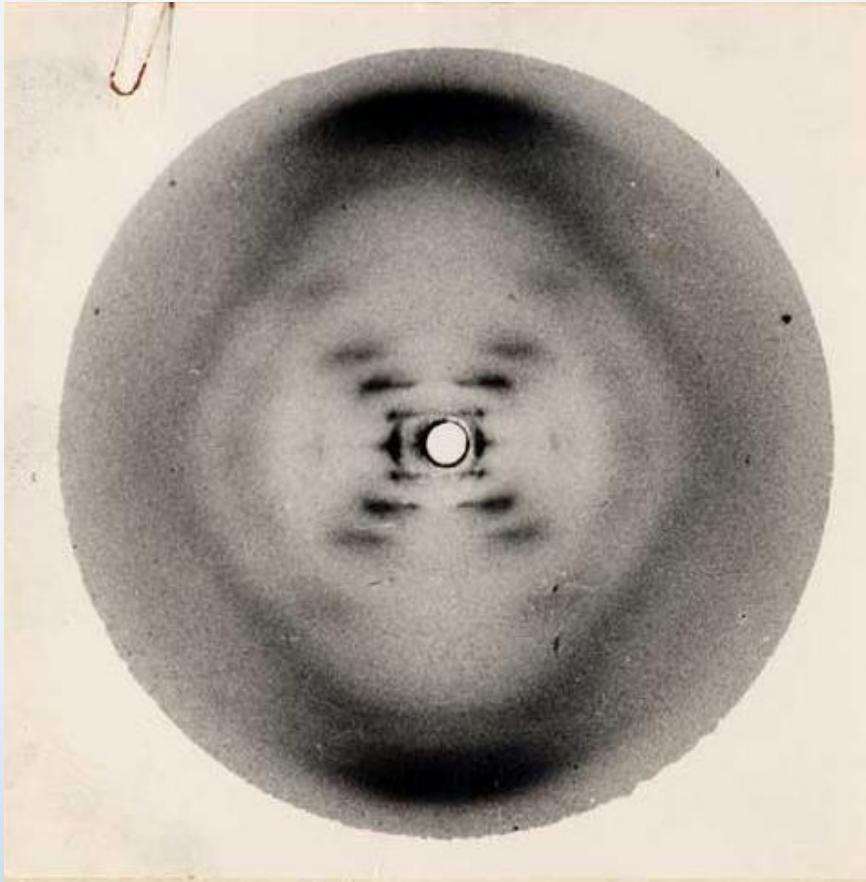
Protein Data Bank: Yearly Growth





Macromolecular X-Ray Crystallography

Macromolecular X-ray Crystallography



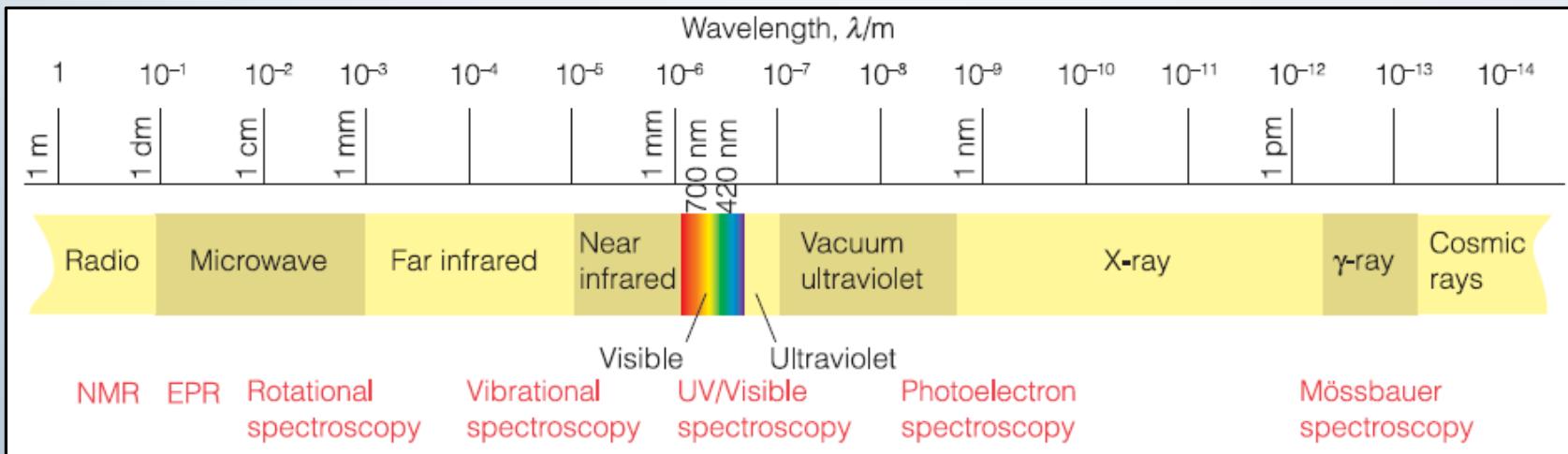
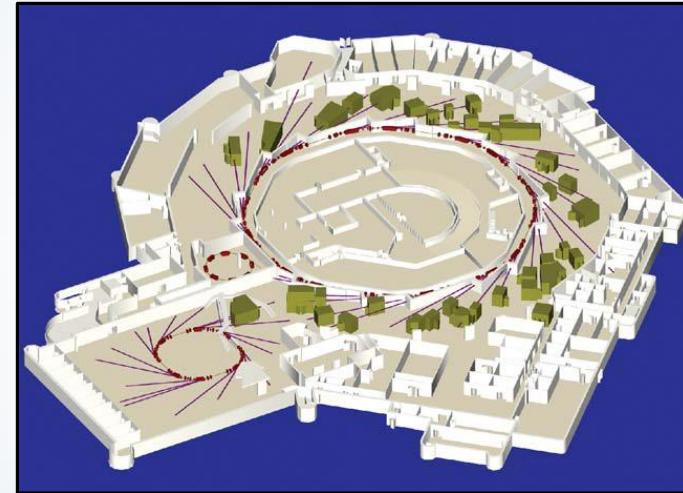
Why X-rays?

to have a wavelength comparable to smallest features to be resolved.

emitted from copper targets bombarded with high energy electrons.

a typical wavelength of 1.5418 Å is very similar to C-C distance.

Synchrotron Light Source



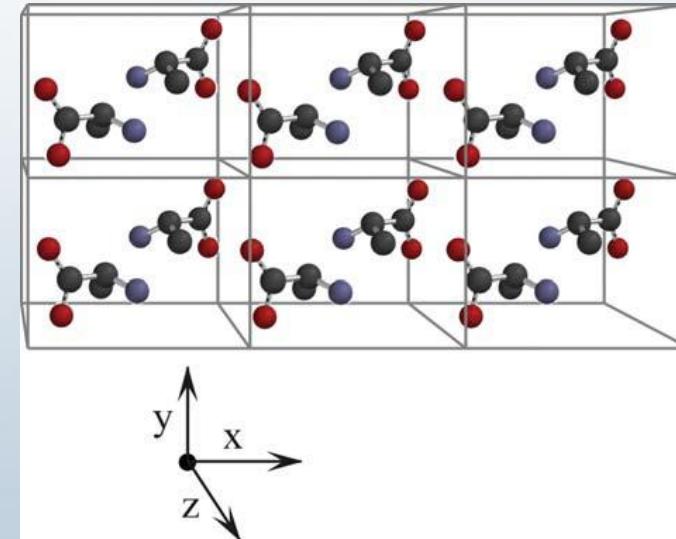
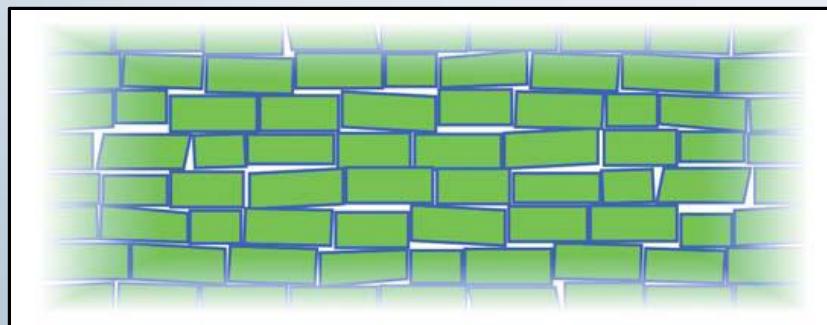
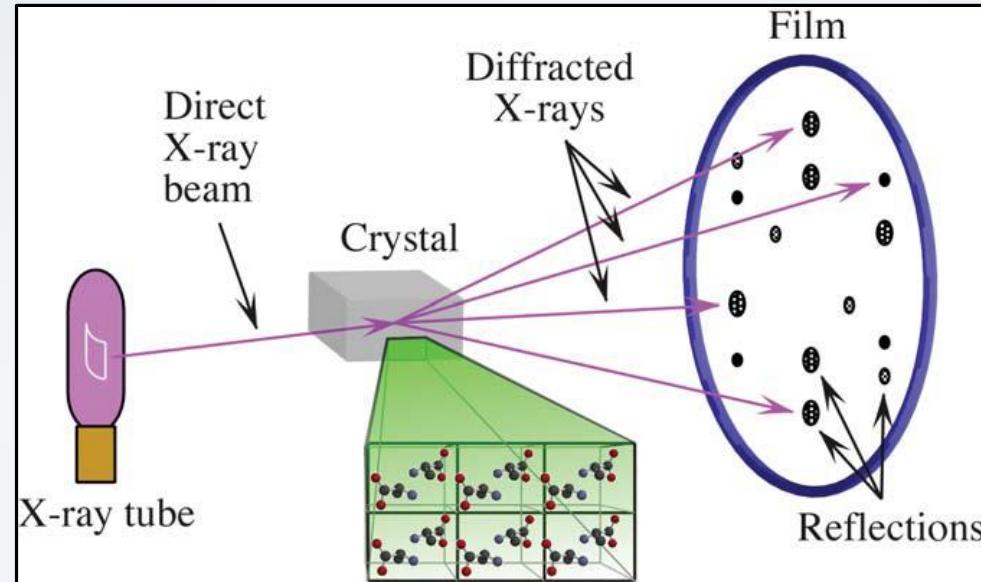
Why crystals?

Scattering from a single molecule:
weak and difficult to detect above
the noise level.

Crystal:
huge numbers of molecules arranged
in the same orientation.

scattered waves **add up in phase**
and raise signal to a measurable
level.

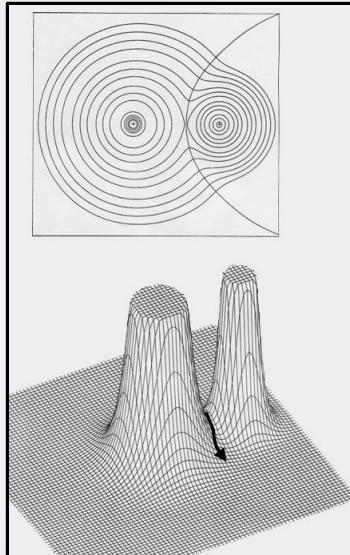
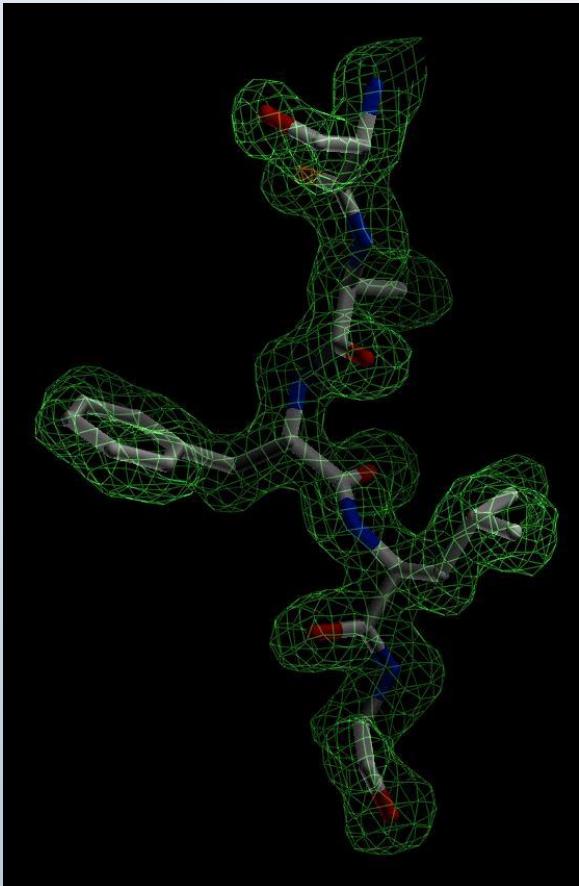
crystal acts as an amplifier



Why electron density?

Intensity of scattered radiation is proportional to the square of the charge/mass ratio.

Crystallography obtains a map of distribution of electrons.

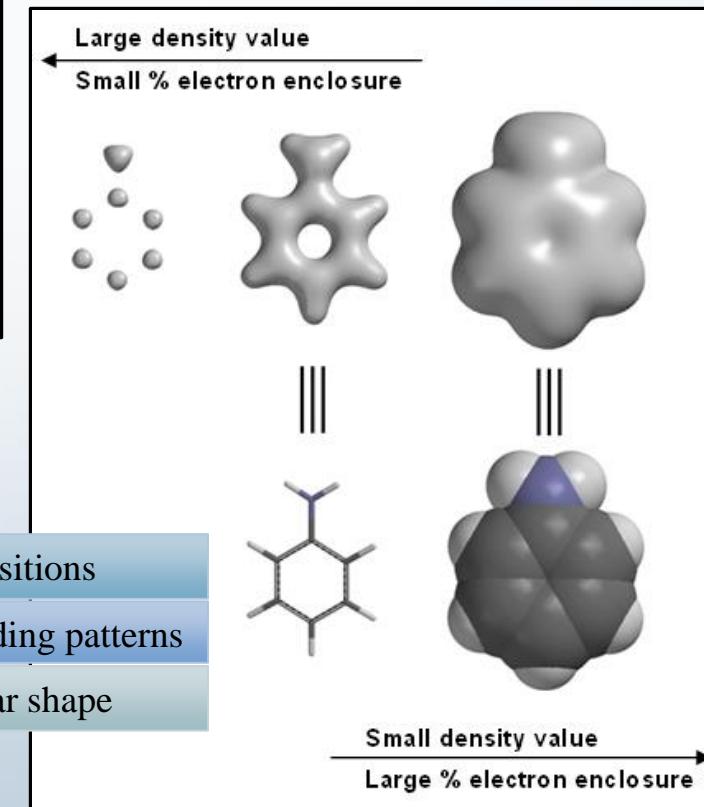


electrons are mostly tightly localized around nuclei.

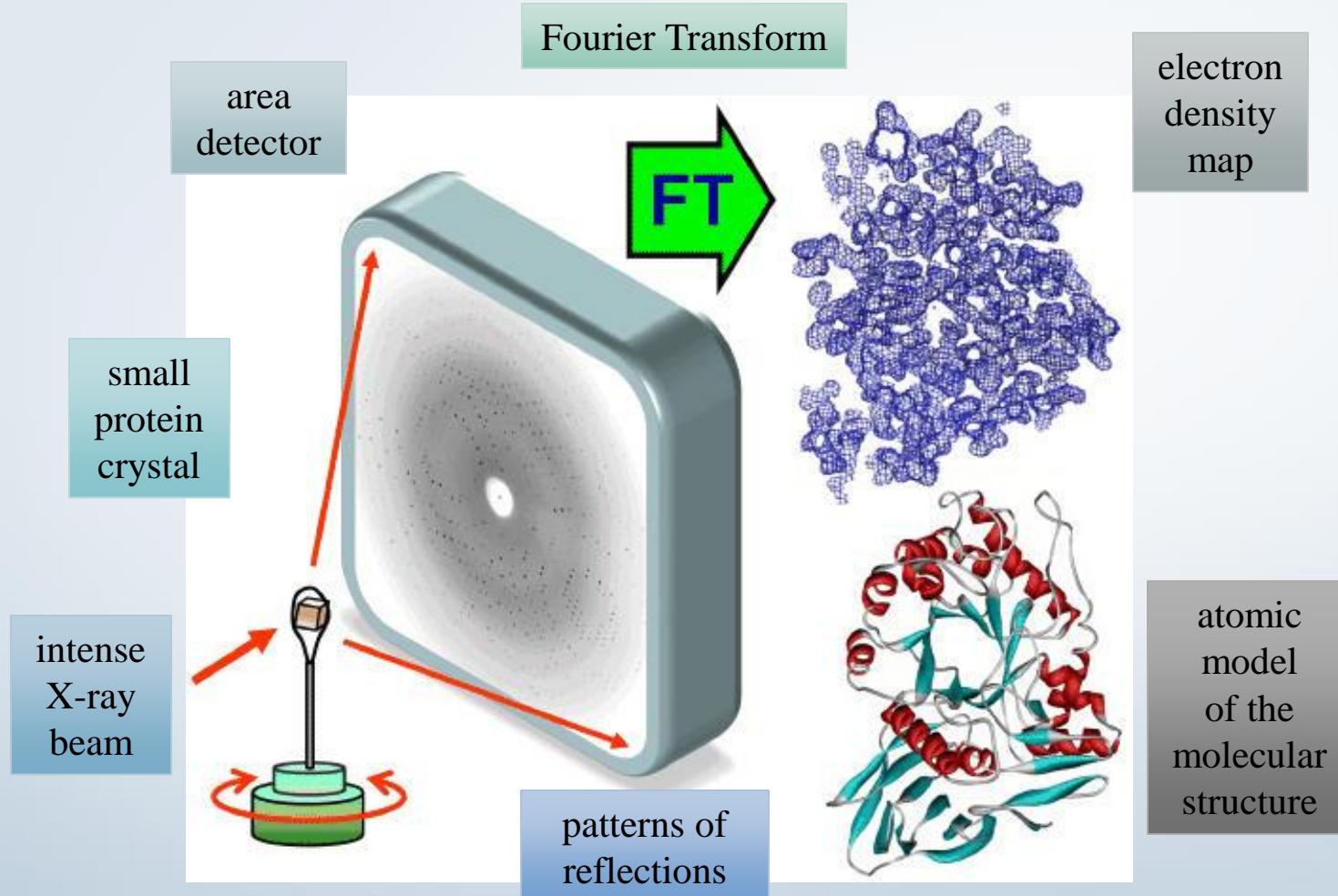
High values: atom positions

Medium values: bonding patterns

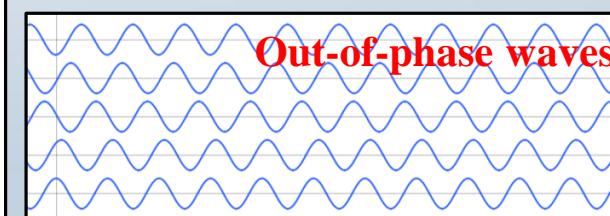
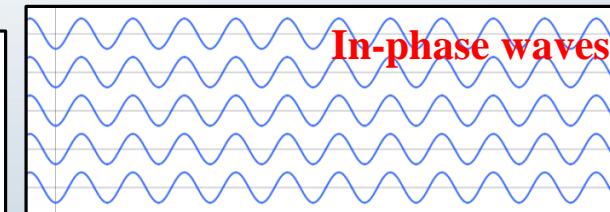
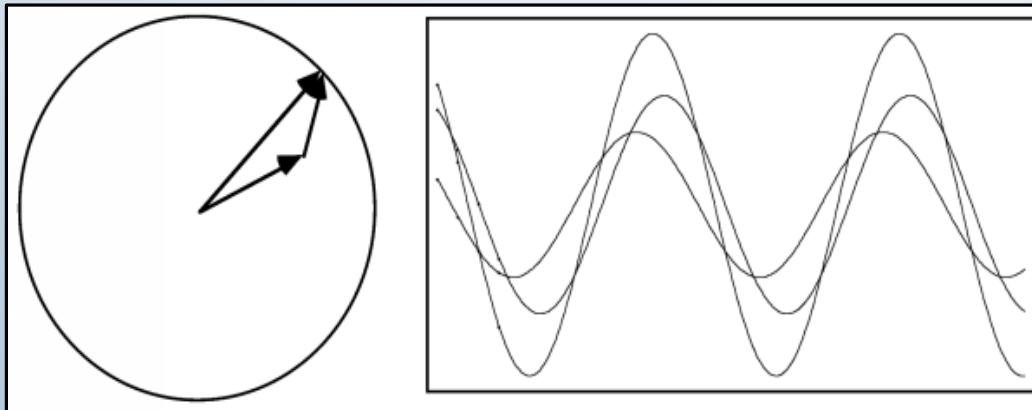
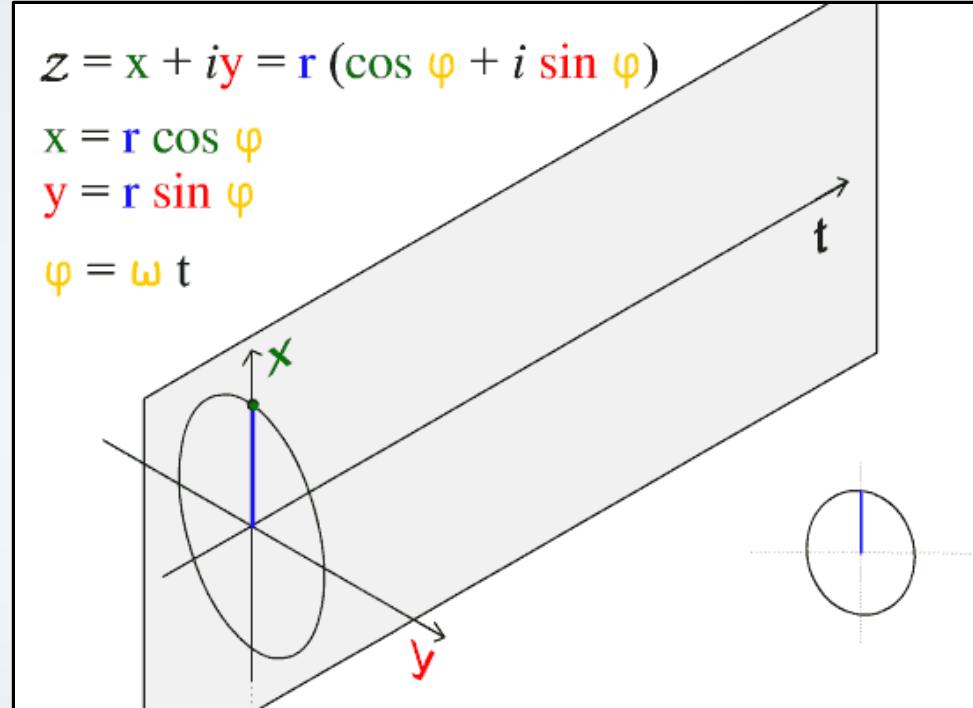
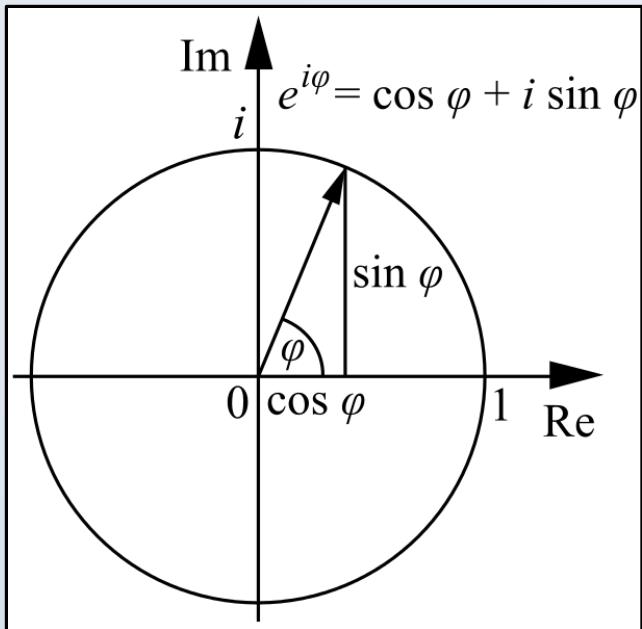
Low values: molecular shape



Overview of the Method

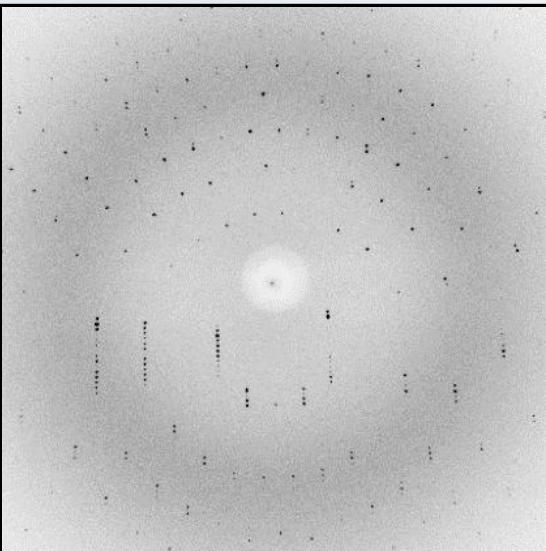


Waves as vectors: Amplitude and Phase

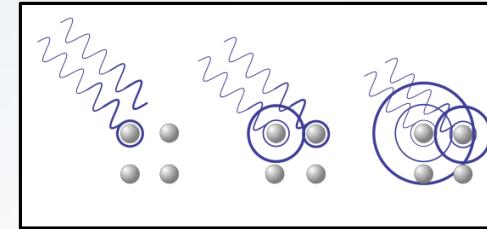


Bragg's Law

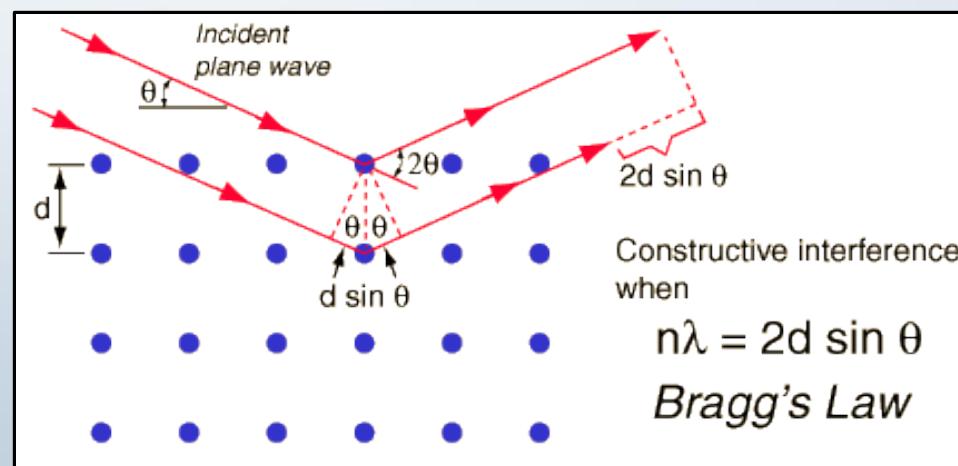
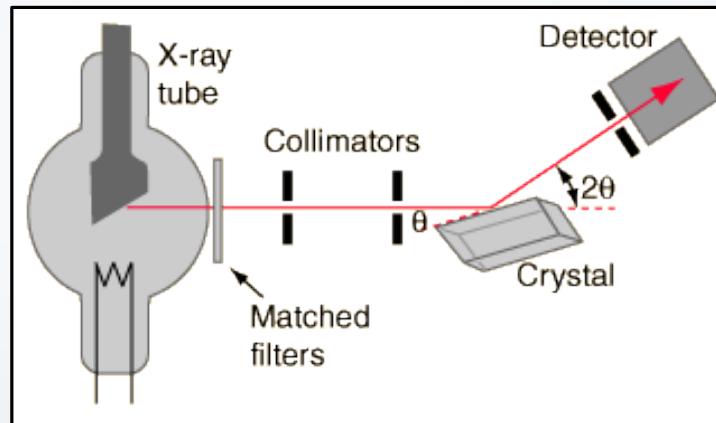
The diffraction pattern from a crystal is an array of spots.



The waves add up in phase in some directions and cancel out in a lot of other directions.



William Lawrence Bragg and his father William Henry Bragg (Nobel prize in 1915)



Bragg's Law

When do waves scatter in phase?

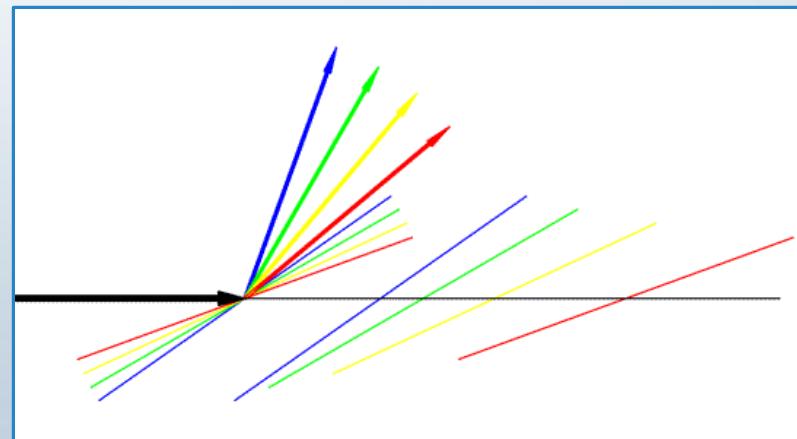
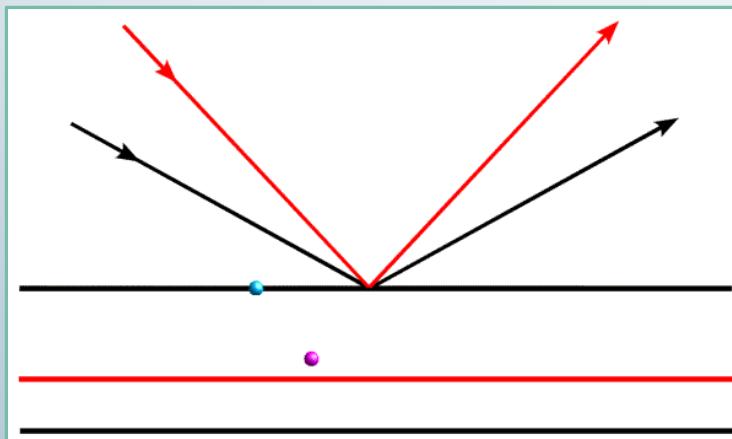
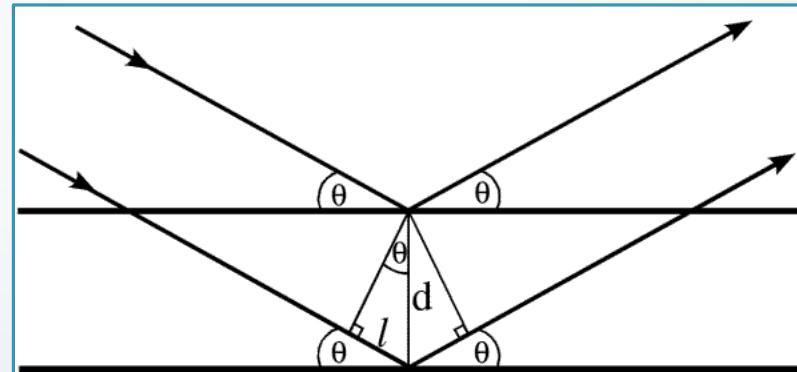
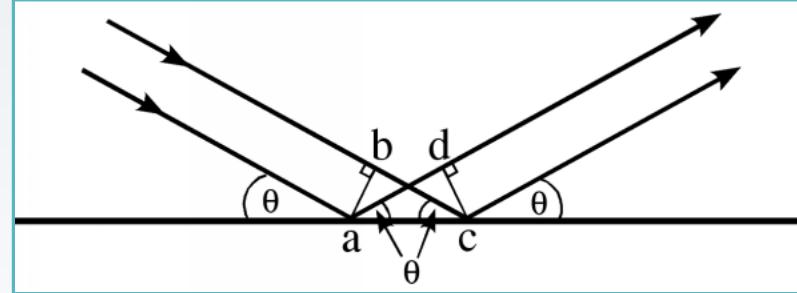
When they have exactly
the same pathlength

When their pathlengths differ by a
multiple of the wavelength

$$\lambda = 2 d \sin\theta$$

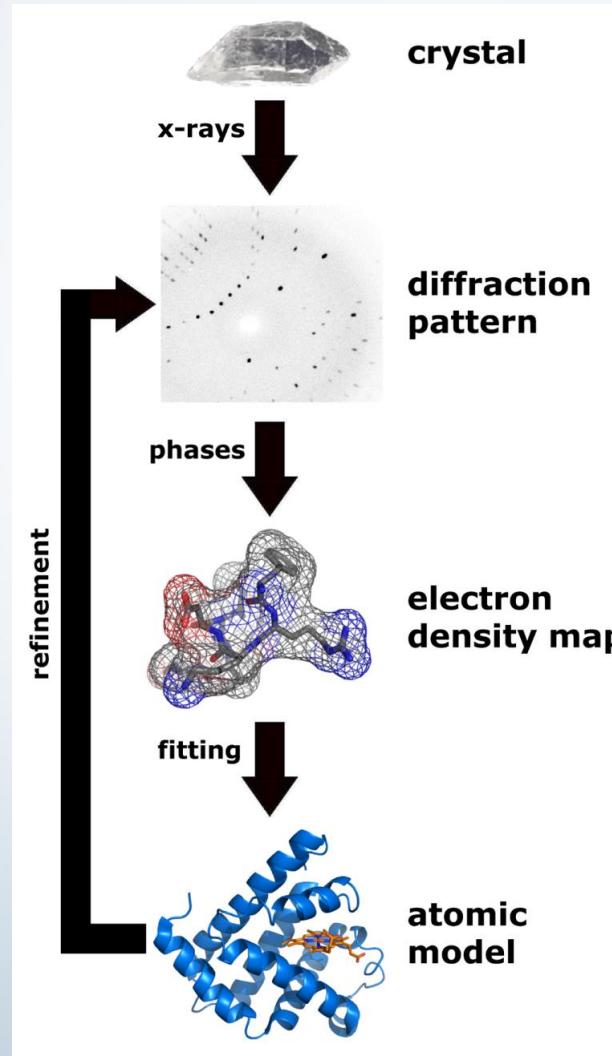
$$\sin\theta/\lambda = 1/(2 d)$$

$$d = \lambda/(2 \sin\theta)$$

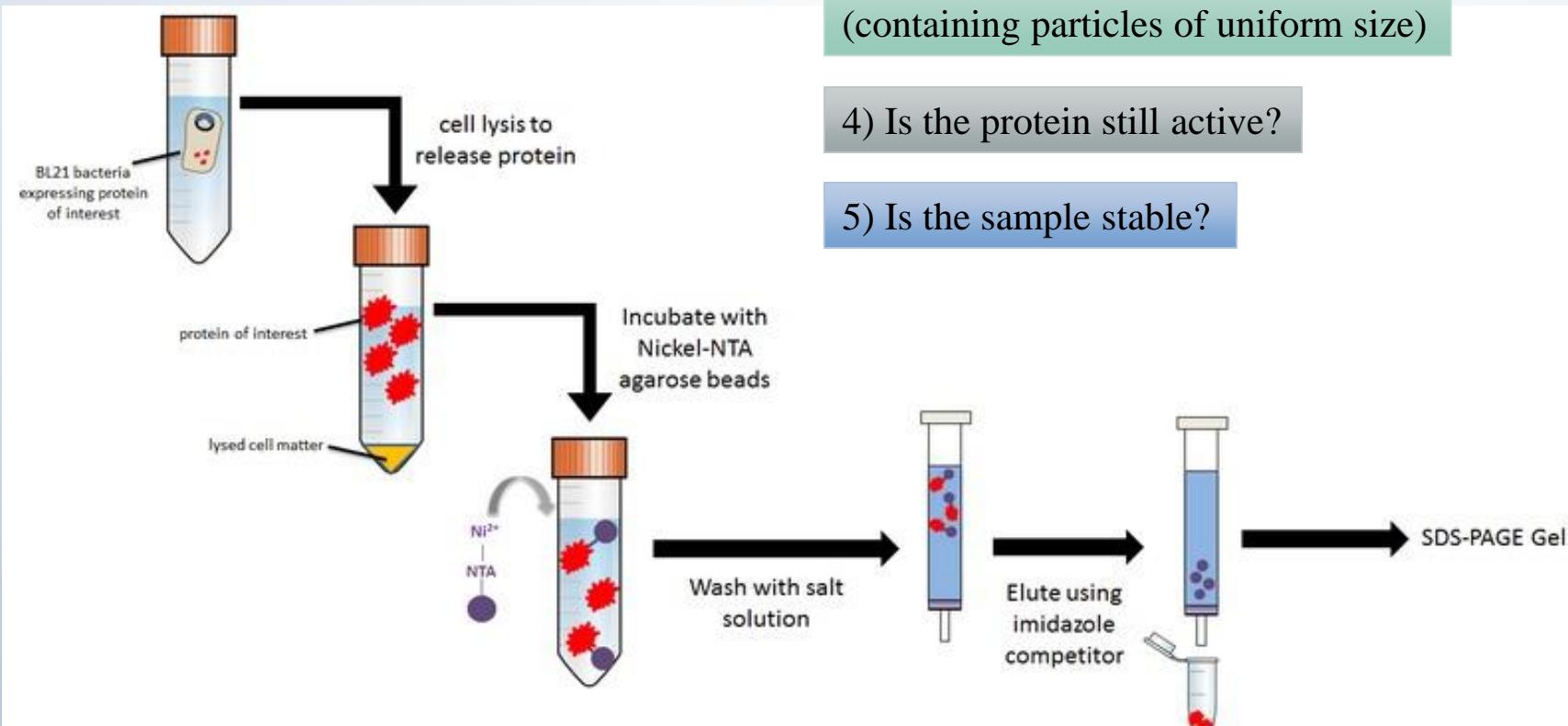
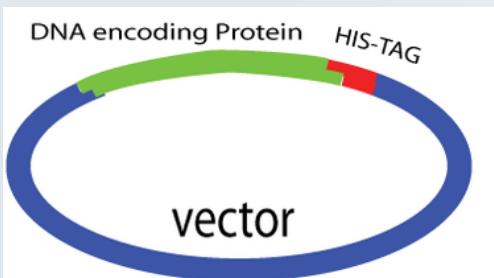


What is involved in a crystal structure determination?

- 1) Protein preparation
- 2) Crystallization
- 3) Testing crystals
- 4) X-ray data collection
- 5) Structure solution
- 6) Model building
- 7) Refinement
- 8) Validation
- 9) Deposition



1) Protein preparation



1) Is it pure and homogeneous?

2) Is the protein soluble and folded?

3) Is the sample monodisperse?
(containing particles of uniform size)

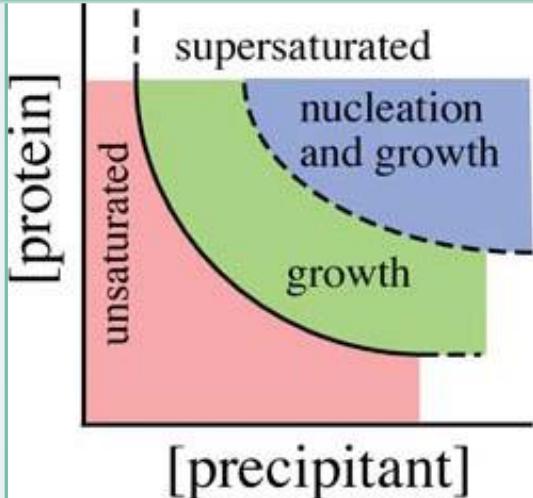
4) Is the protein still active?

5) Is the sample stable?

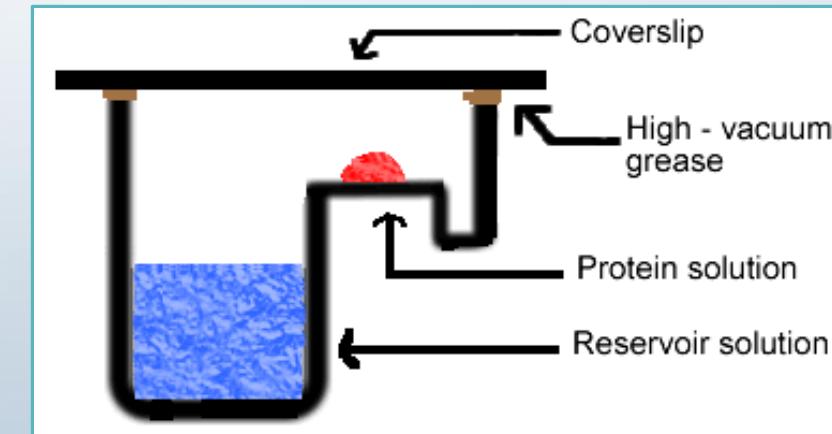
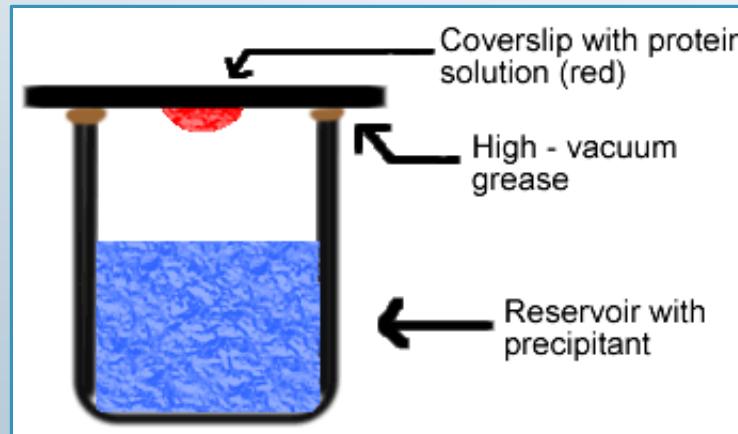
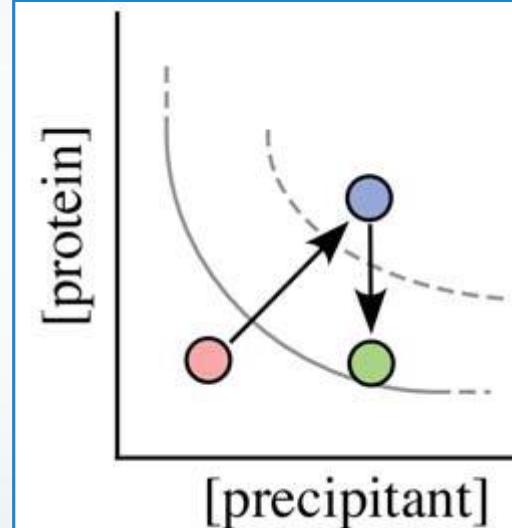
2) Crystallization 3) Testing crystals

Vapor Diffusion Methods:
1) Hanging drop
2) Sitting drop

Phase diagram for crystallization mediated by a precipitant



An ideal strategy for growing large crystals



2) Crystallization



Lysozyme crystals are growing
from an aqueous media

3) Testing crystals

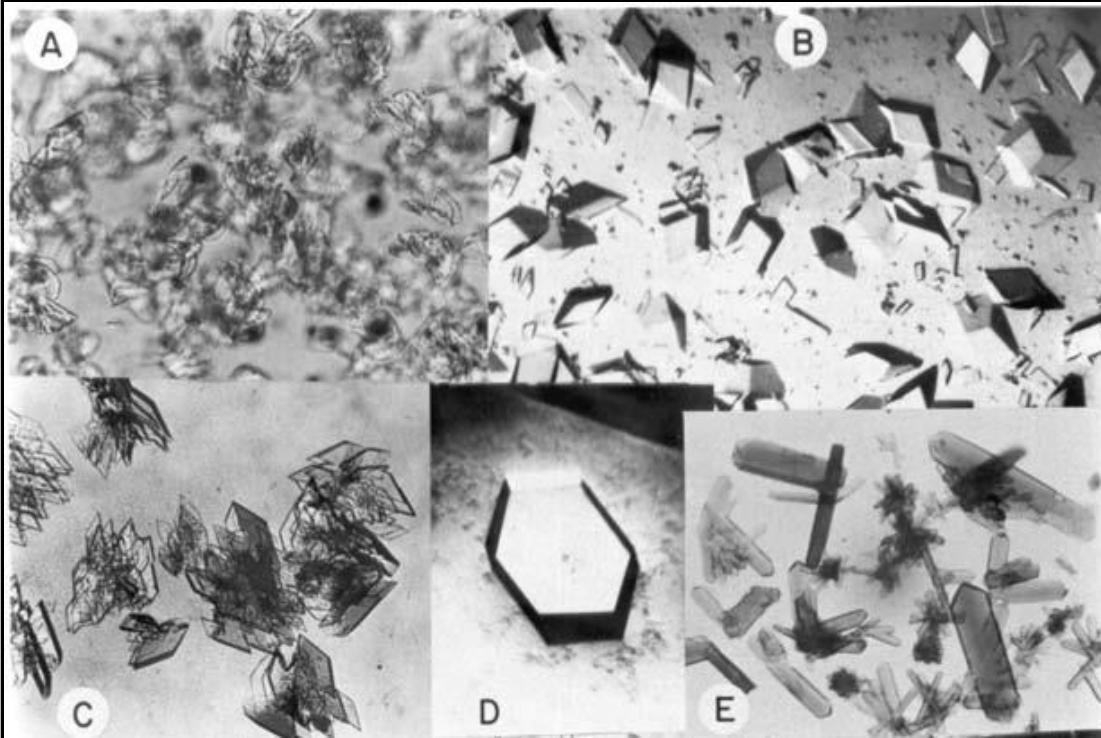
sharp diffraction patterns with
clear reflections at large angles
from the X-ray beam

optical clarity, smooth faces
and sharp edges

Ideally crystal should diffract
to better than 4 Å resolution.



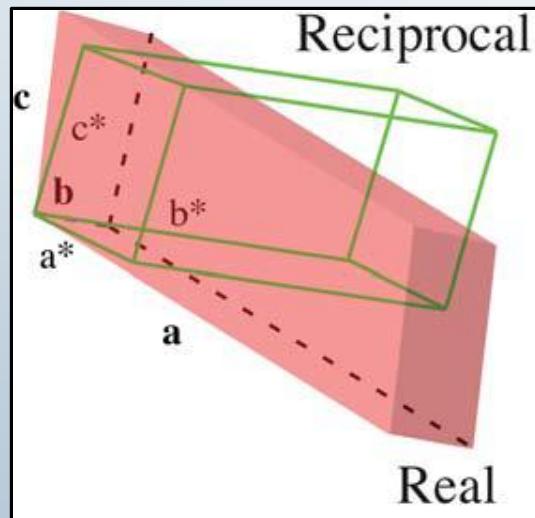
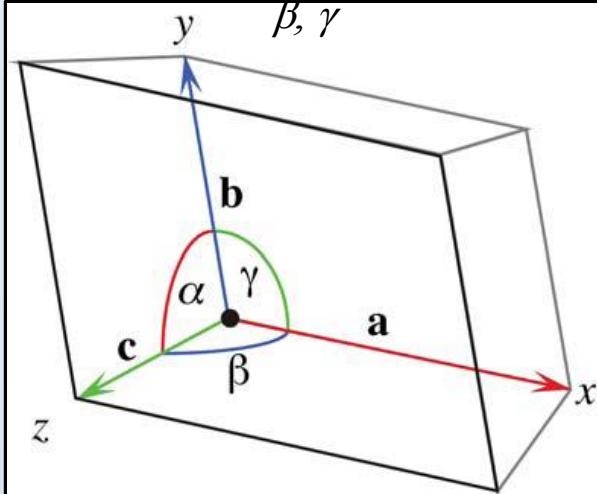
2) Crystallization 3) Testing crystals



- (a) deer catalase
- (b) trigonal form of fructose-1,6-diphosphatase from chicken liver
- (c) cortisol binding protein from guinea pig Sera
- (d) concanavalin B from jack beans
- (e) beef liver catalase

4) X-ray data collection

General (triclinic) unit cell,
with edges **a**, **b**, **c** and angles α ,



unit cell parameters

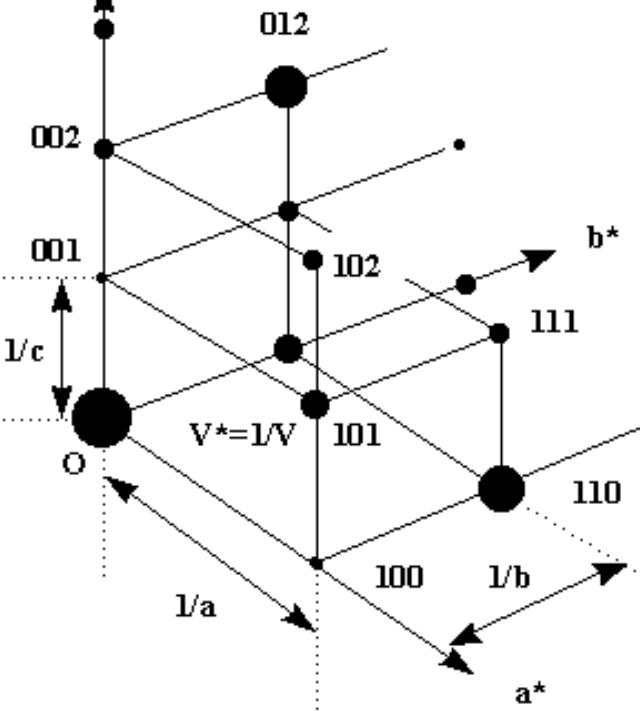
resolution limit

crystal symmetry

crystal orientation

Diffraction Geometry: Reciprocal Lattice

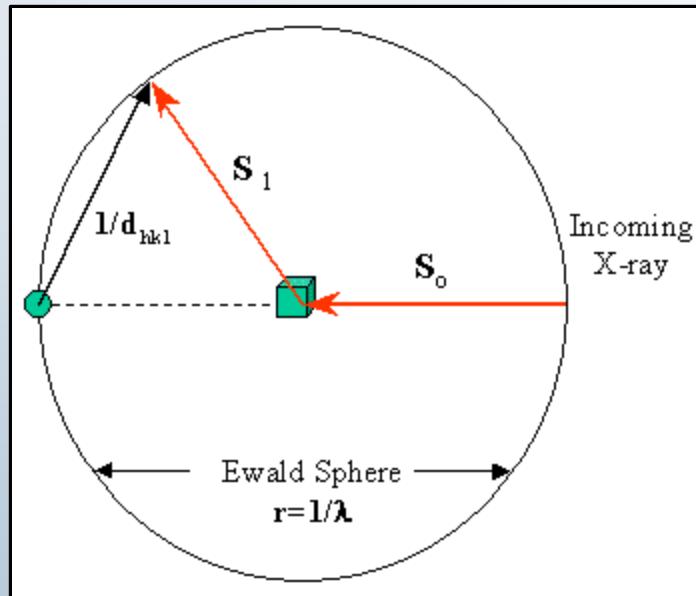
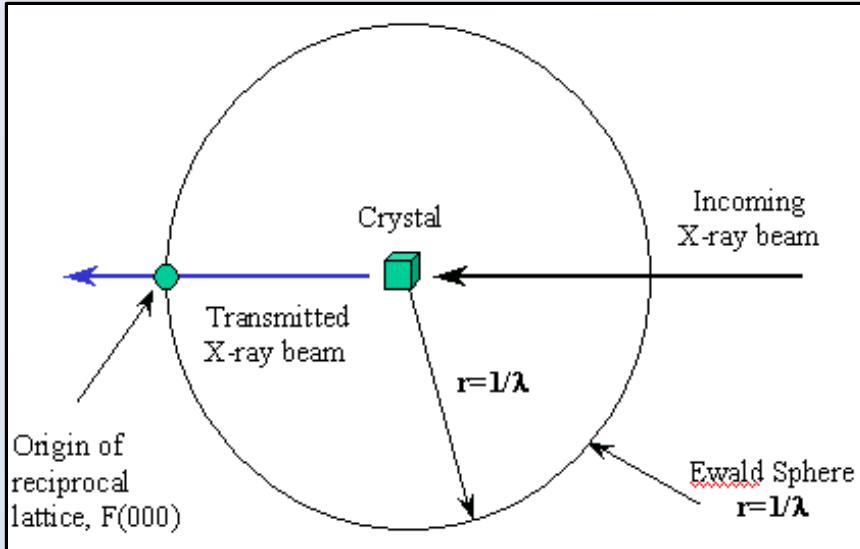
$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{b} \cdot \mathbf{c} \times \mathbf{a}} \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{c} \cdot \mathbf{a} \times \mathbf{b}}$$



reciprocal
lattice is
valuable in
visualizing
diffraction
geometry

coordinates of
reciprocal
lattice points
are **hkl** where
h, **k** and **l** are
integers

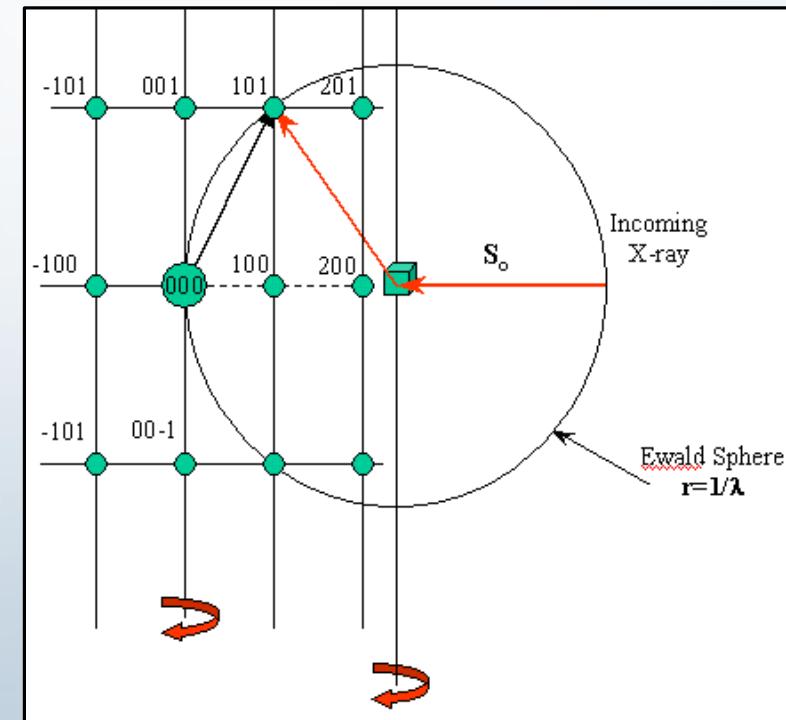
4) X-ray data collection



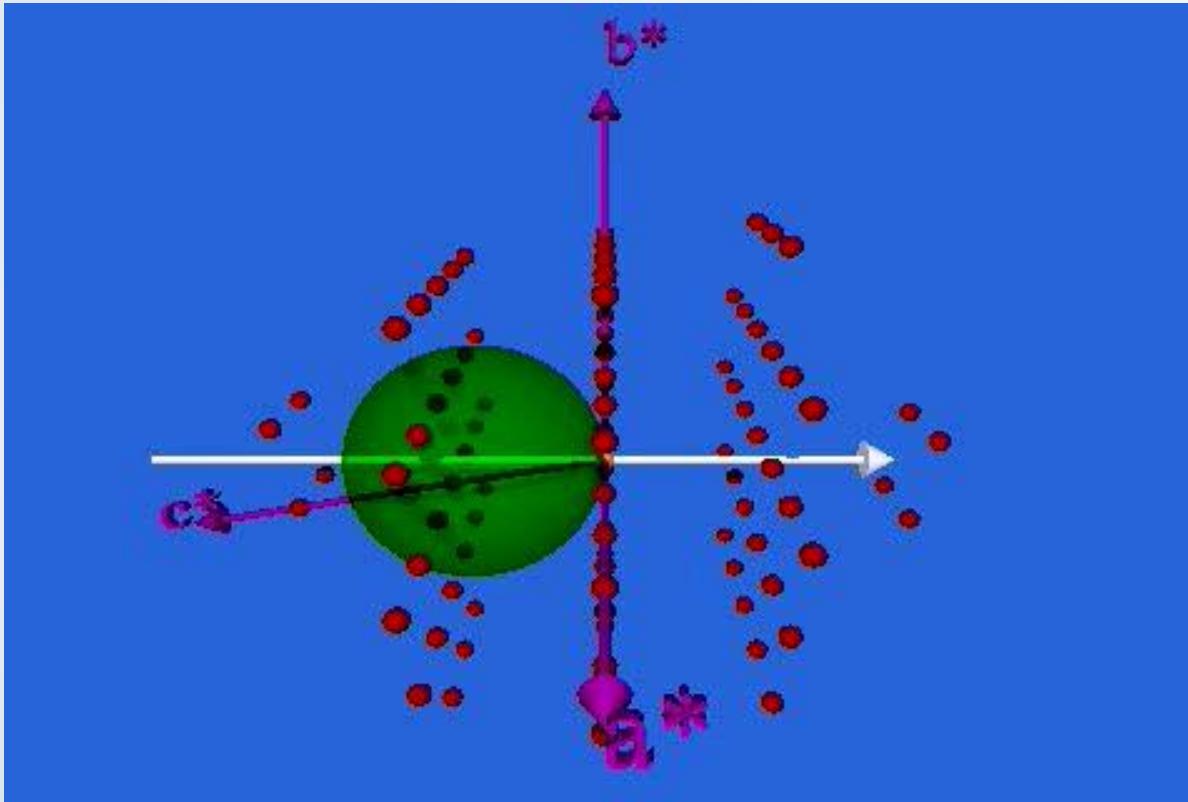
The Ewald Sphere

diffraction maxima (reflections) occur only when Bragg equation is satisfied.

This condition occurs whenever a reciprocal lattice point lies exactly on the Ewald sphere



4) X-ray data collection



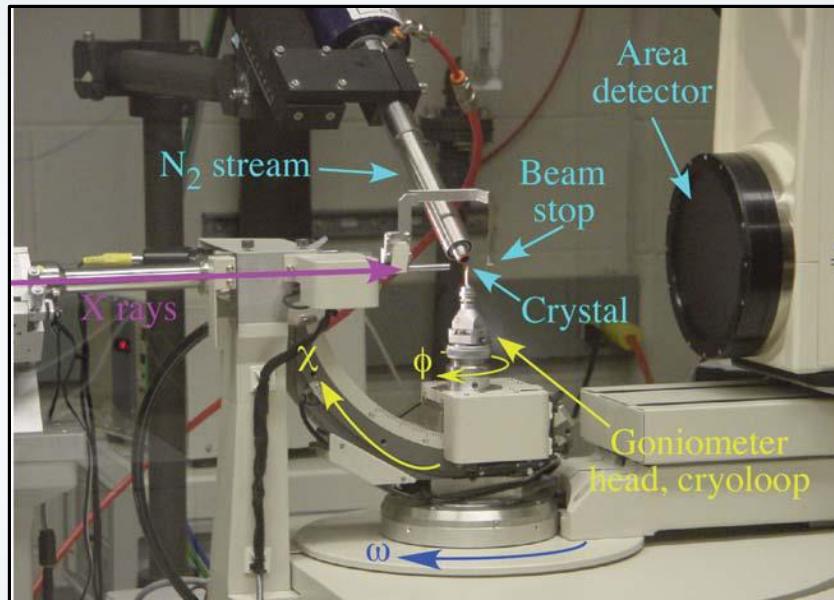
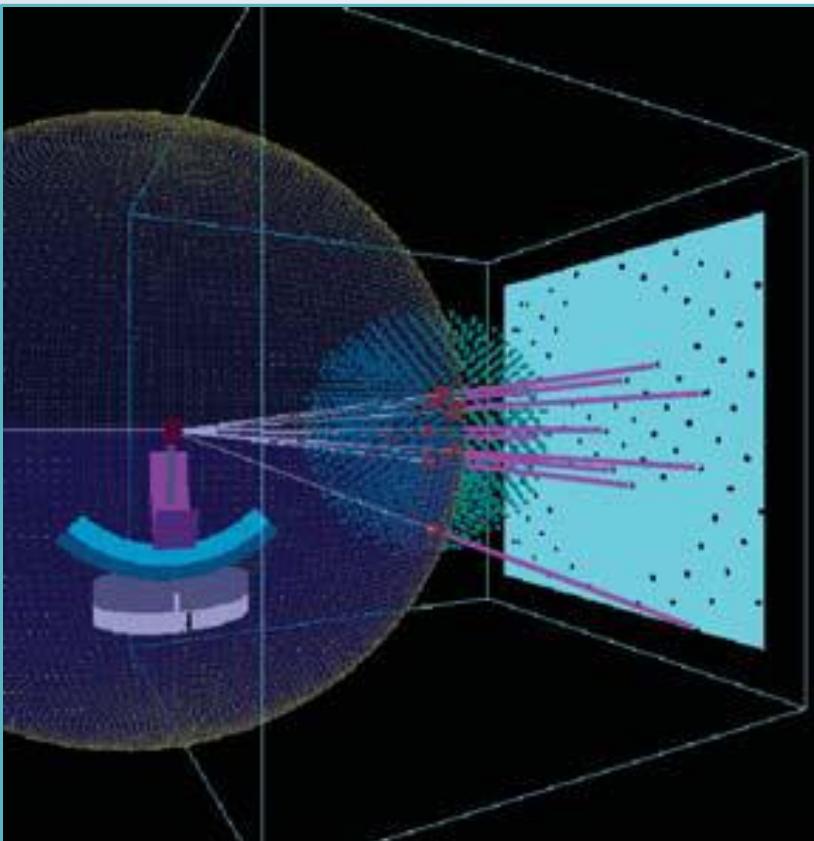
Diffraction from a set of planes with Miller indices \mathbf{hkl} will occur when the corresponding reciprocal lattice point (\mathbf{hkl}) lies exactly on the Ewald sphere

because a real crystal is made up of many small mosaic blocks with a small spread in orientations, a reciprocal lattice point for the crystal will not be a true point, but a small spherical cap

because of wavelength dispersion and beam divergence, Ewald sphere has a finite thickness

4) X-ray data collection

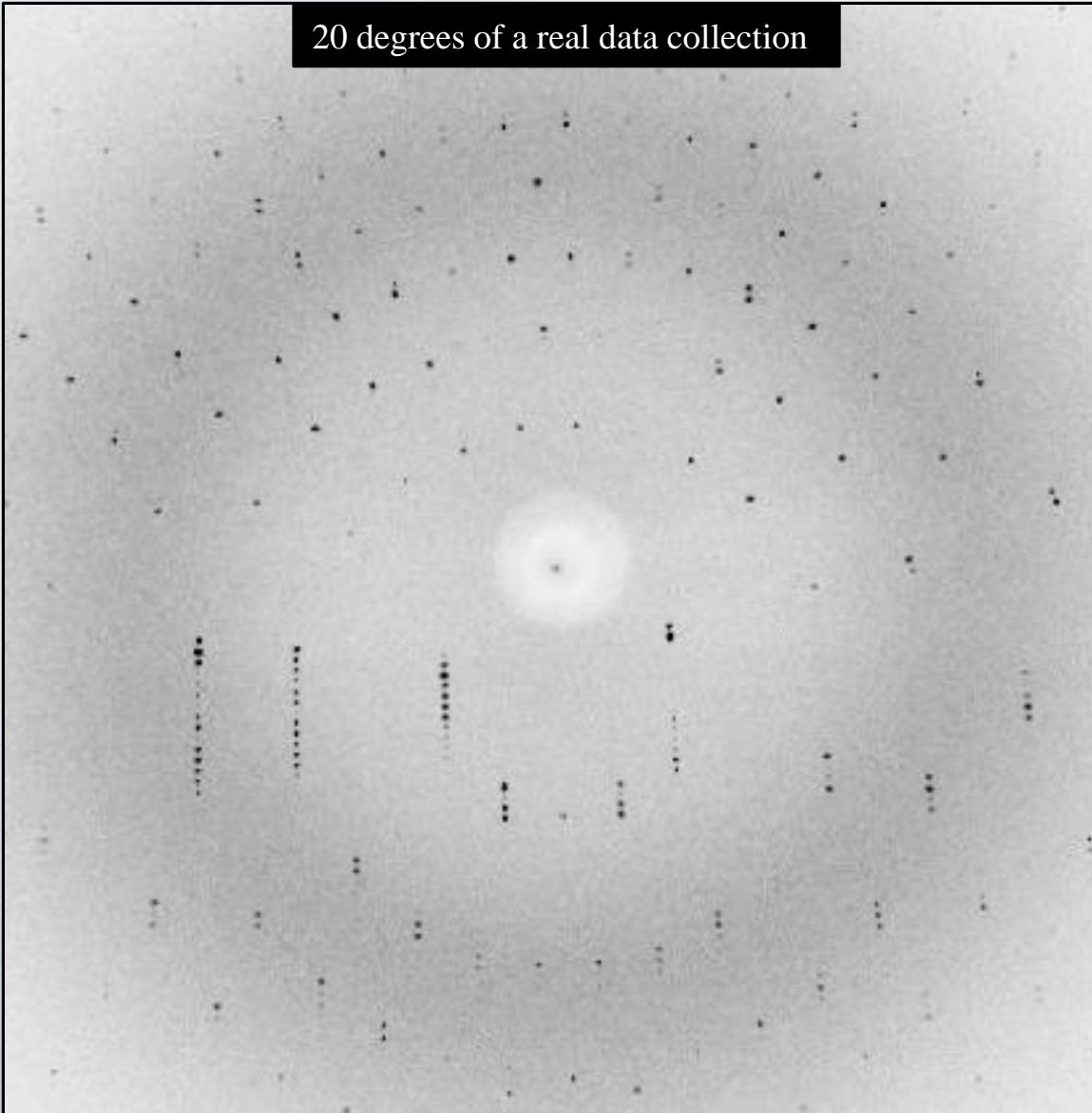
data collection strategy :
maximize resolution and
completeness of data set



rotate crystal by a small angle step by step
and record the X-ray diffraction pattern

4) X-ray data collection

20 degrees of a real data collection

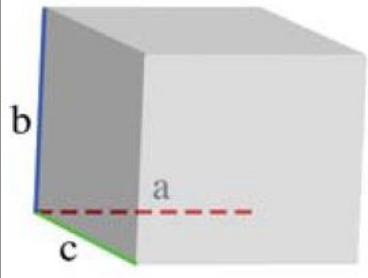


4) X-ray data collection

Most Symmetric

Cubic

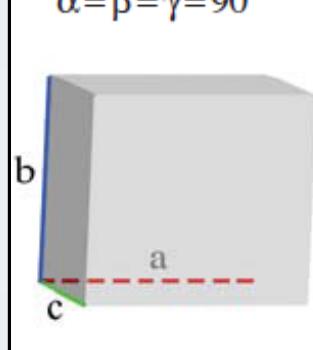
$$a=b=c, \alpha=\beta=\gamma=90^\circ$$



The lower the symmetry,
more data are required

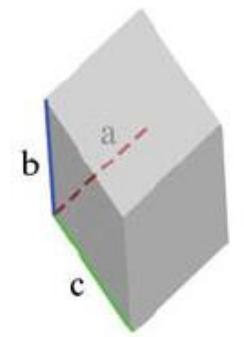
Tetragonal

$$a=b\neq c, \alpha=\beta=\gamma=90^\circ$$



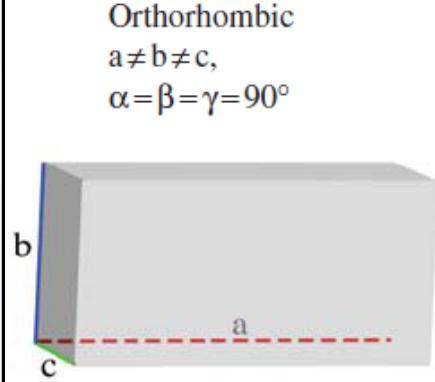
Rhombohedral

$$a=b=c, \alpha=\beta=\gamma\neq 90^\circ$$



Orthorhombic

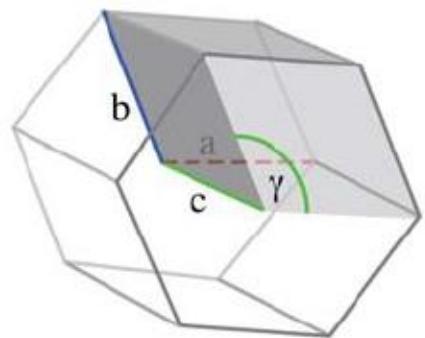
$$a\neq b\neq c, \alpha=\beta=\gamma=90^\circ$$



Least Symmetric

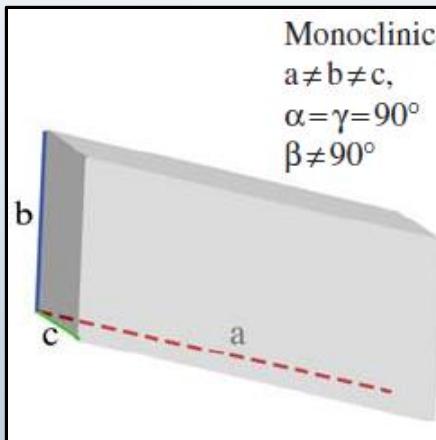
Hexagonal

$$a=b=c, \alpha=\beta=90^\circ, \gamma=120^\circ$$



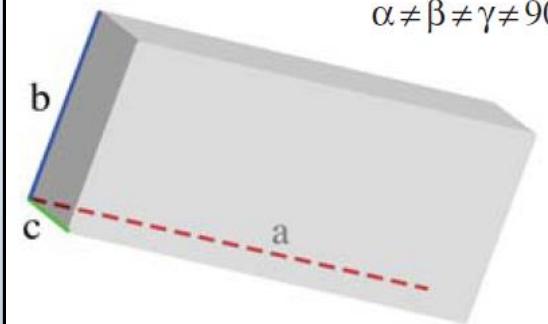
Monoclinic

$$a\neq b\neq c, \alpha=\gamma=90^\circ, \beta\neq 90^\circ$$



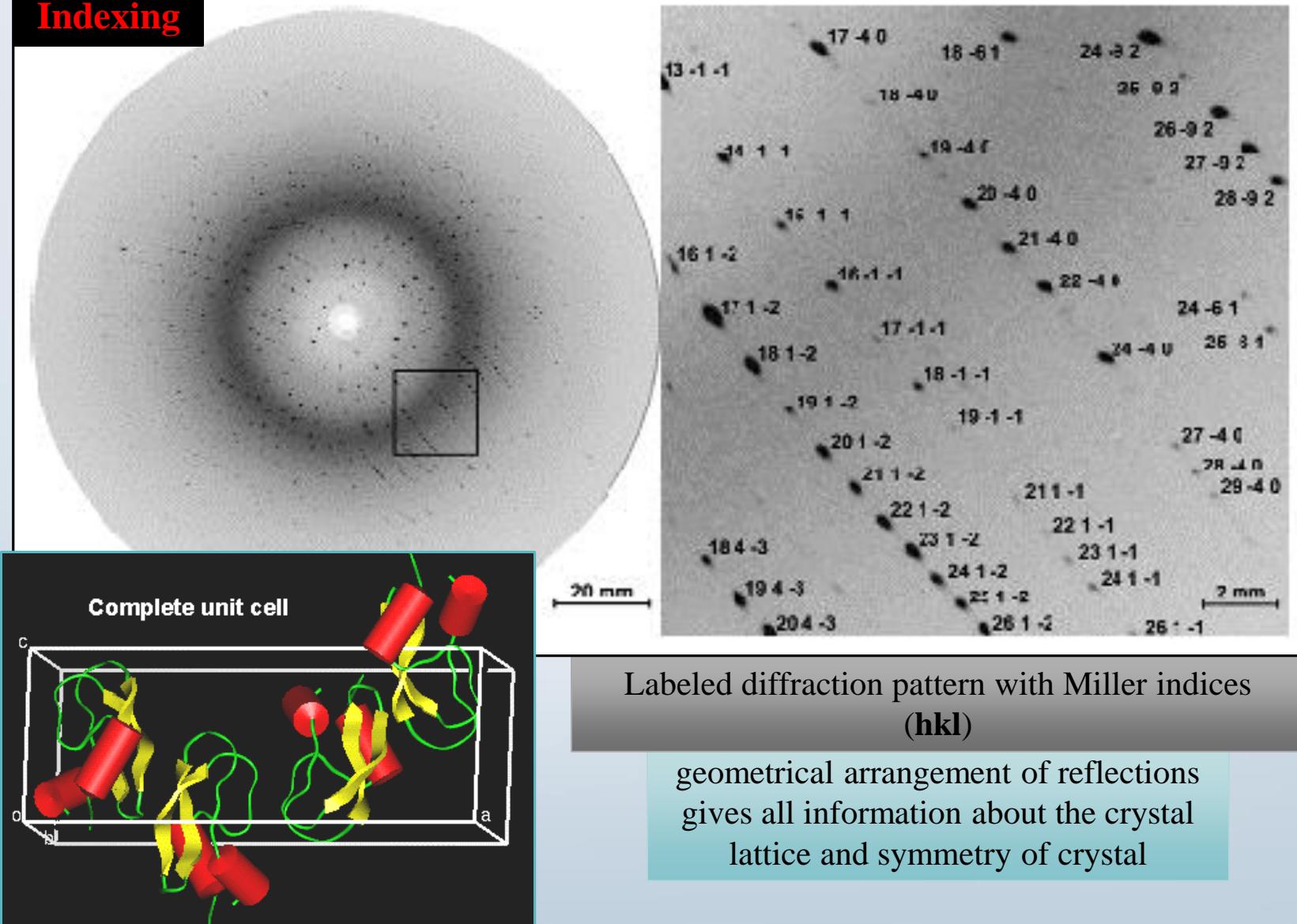
Triclinic

$$a\neq b\neq c, \alpha\neq\beta\neq\gamma\neq 90^\circ$$



4) X-ray data collection

Indexing



4) X-ray data collection

Scaling and Merging

placing all data on a common scale

merging multiple observations to give a unique dataset

rejecting outliers

reducing intensities to amplitudes

$$I_{hkl} \propto |F(hkl)|^2$$

Multiple observations are reduced to a weighted mean intensity and standard deviation

Data quality indicators:

How similar are the measured intensities of equivalent reflections?

Measure of internal agreement of measurements in a crystallographic data set.

R-merge:

$$R = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

$\langle I_{hkl} \rangle$ is the average of symmetry related observations

R-meas:

Redundancy-independant version of R-merge.

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{j=1}^n |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

4) X-ray data collection

Here a typical reflection data listing. The first three rows are the indices h, k, l , of the reflection, followed by $|F|^{\star 2}$ (the intensity) and $\sigma(|F|^{\star 2})$ as a statistical measure.

6	0	0	295587.	3564.27
8	0	0	476981.	3687.43
10	0	0	77658.04	1728.93
12	0	0	76207.25	1882.20
14	0	0	54967.11	1862.80
16	0	0	661011.	1368.47
18	0	0	28076.70	3147.65
20	0	0	129816.	3182.27
22	0	0	87852.19	2770.71
24	0	0	165364.	3898.54
26	0	0	42694.30	2164.06
28	0	0	6260.15	1092.76
30	0	0	2112.36	924.42
32	0	0	3942.09	1122.29
34	0	0	43610.59	2857.37

After some data processing
list of indexed reflections
and their intensities

5) Structure solution

The Fourier transform

The Fourier transform operation is reversible.

Units of the variable \mathbf{h} are reciprocals of the units of \mathbf{x}

This highly general mathematical form is naturally adapted for relating real and reciprocal space

$$F(\mathbf{h}) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i (\mathbf{h} \cdot \mathbf{x})} dx$$

$$f(x) = \int_{-\infty}^{+\infty} F(\mathbf{h}) e^{-2\pi i (\mathbf{h} \cdot \mathbf{x})} dh$$

$$F(\mathbf{h}, k, l) = \int_x \int_y \int_z f(x, y, z) e^{2\pi i (\mathbf{h} \cdot \mathbf{x} + k y + l z)} dx dy dz$$

$$f(x, y, z) = \int_h \int_k \int_l F(\mathbf{h}, k, l) e^{-2\pi i (\mathbf{h} \cdot \mathbf{x} + k y + l z)} dh dk dl$$

Fourier transform is a precise mathematical description of diffraction

Fourier Sum
a sum of simple wave equations or periodic functions that describes or approximates a complicated periodic function.

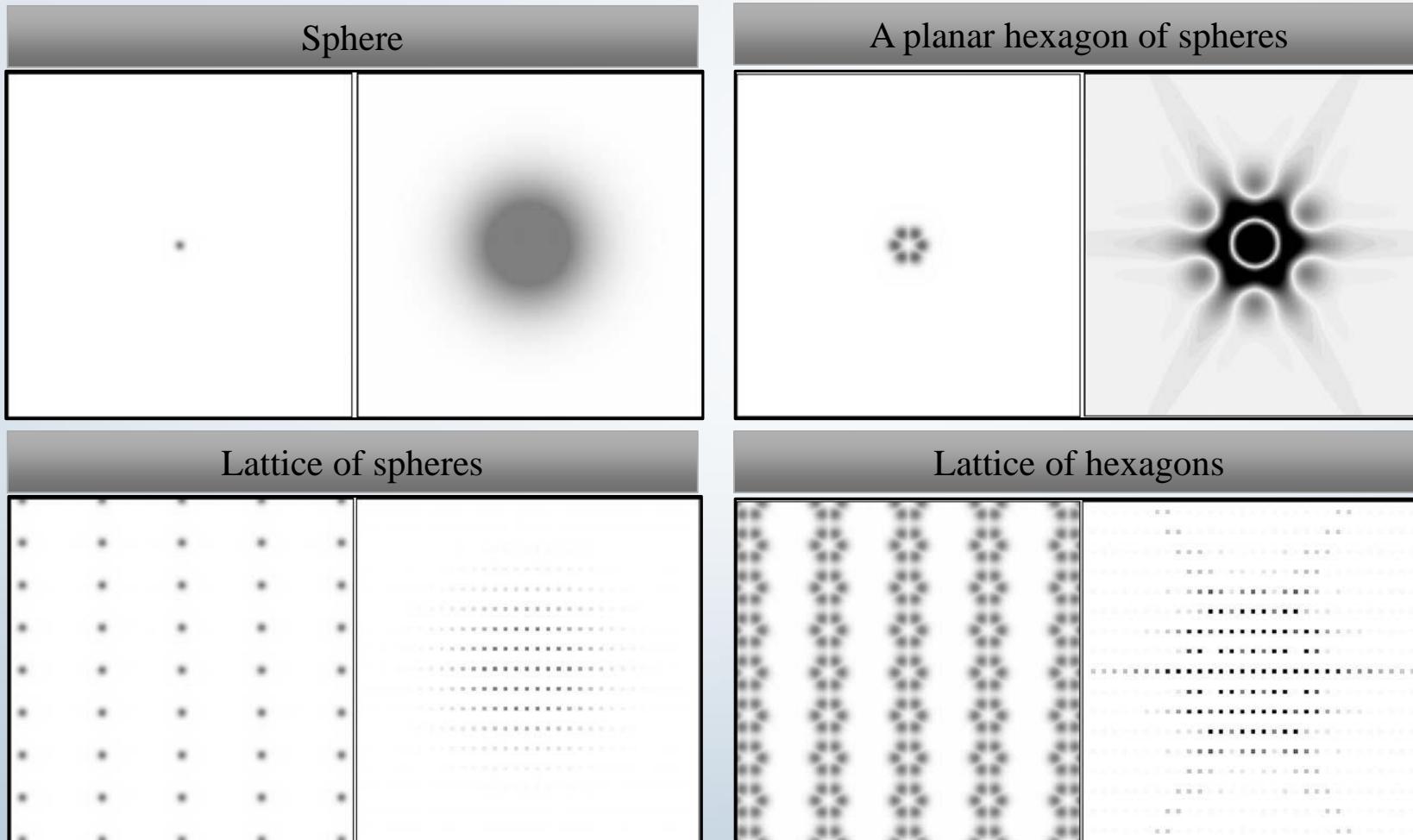
Fourier Synthesis
constructing a Fourier sum to approximate a specific function

Fourier analysis
decomposing a complicated function into its components

Fourier Transform
an operation that transforms a function containing variables of one type into a function whose variables are reciprocals of original type

5) Structure solution

The Fourier transform



Note on difference of real space and reciprocal space: **spacing** and **shape**

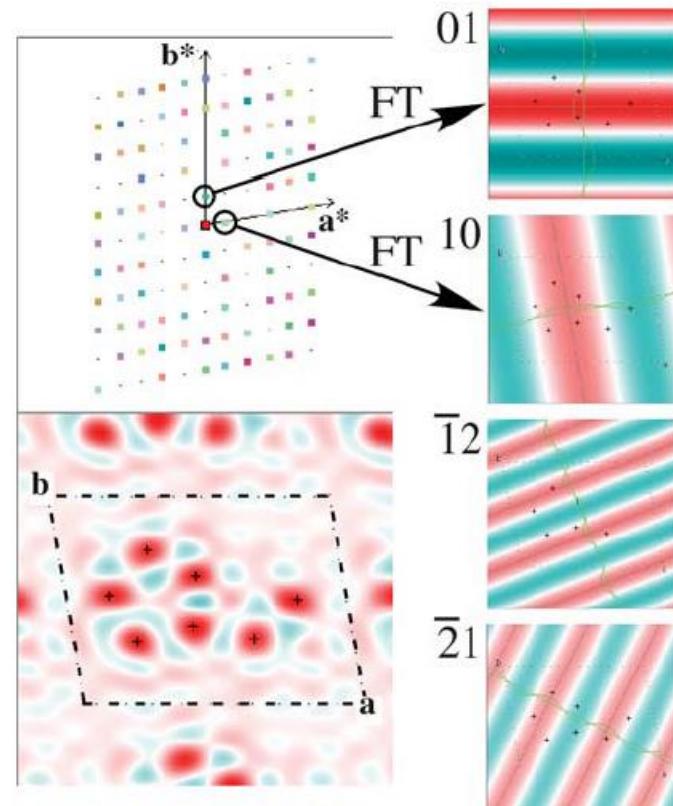
5) Structure solution

The Fourier transform

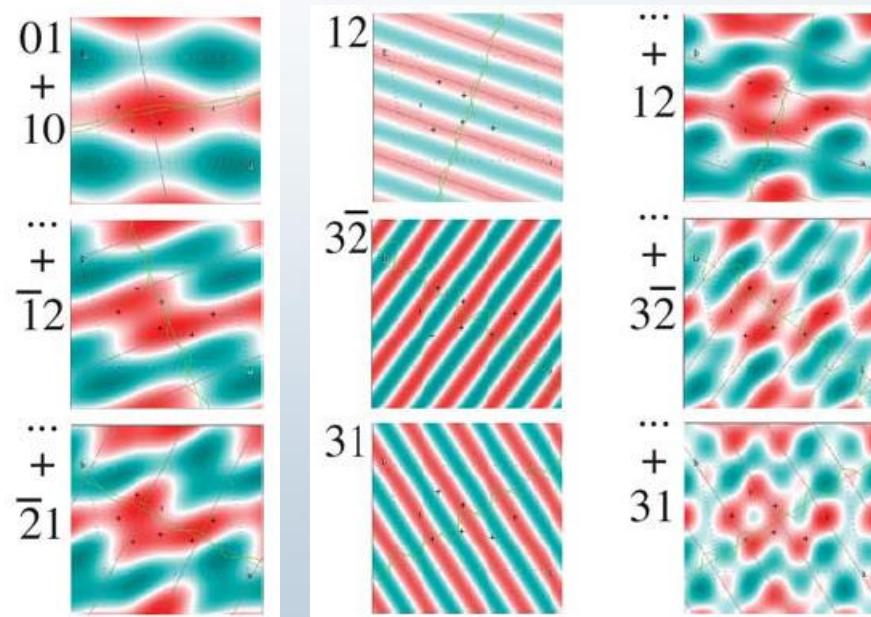
If a crystal does not produce diffracted rays at large angles from direct beam resolution of resulting image is poor

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Comparison of **low-angle** and **high-angle** reflections



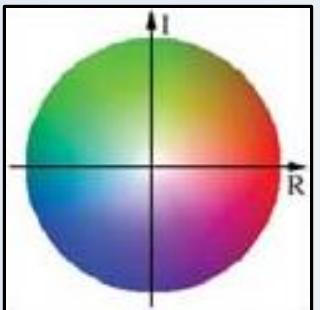
By addition of each FT term far from center of diffraction, we go from an **overall** to a **detailed** shape of electron density



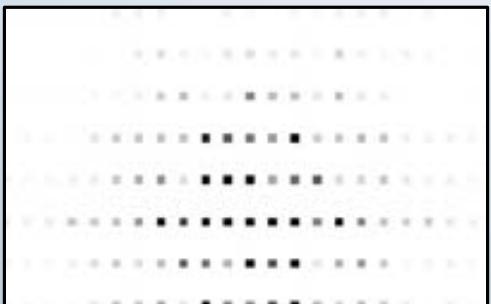
5) Structure solution

The Phase Problem

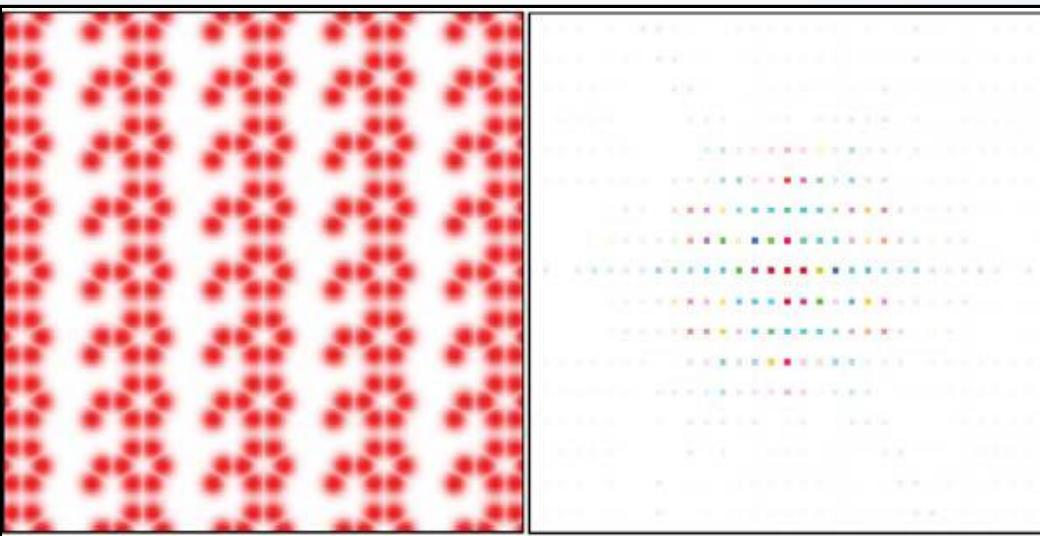
Phase angle of a reflection corresponds to angle of its color on color wheel



Darkness of color indicates intensity of a reflection



Experimental diffraction patterns do not contain phase information



5) Structure solution

The Phase Problem

Atomic structure factor
contribution of the single
atom j to reflection hkl

scattering factor of atom j

$$f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

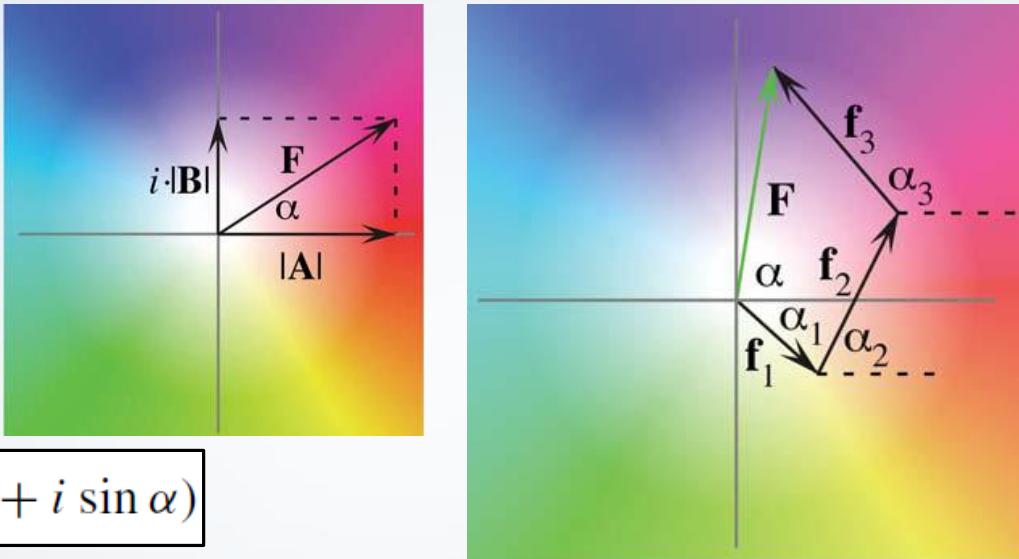
$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dV$$

if we know the structure, either as atoms or as electron density, we can calculate
the diffraction pattern, **including the phases of all reflections**

5) Structure solution

The Phase Problem

$$\mathbf{F} = |\mathbf{F}| \cdot e^{i\alpha}$$



$$\mathbf{F} = |\mathbf{A}| + i |\mathbf{B}| = |\mathbf{F}| \cdot (\cos \alpha + i \sin \alpha)$$

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{i\alpha_{hkl}} e^{-2\pi i (hx + ky + lz)}$$

$$\alpha = 2\pi\alpha'$$

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{2\pi i \alpha'_{hkl}} e^{-2\pi i (hx + ky + lz)}$$

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i (hx + ky + lz - \alpha'_{hkl})}$$

5) Structure solution

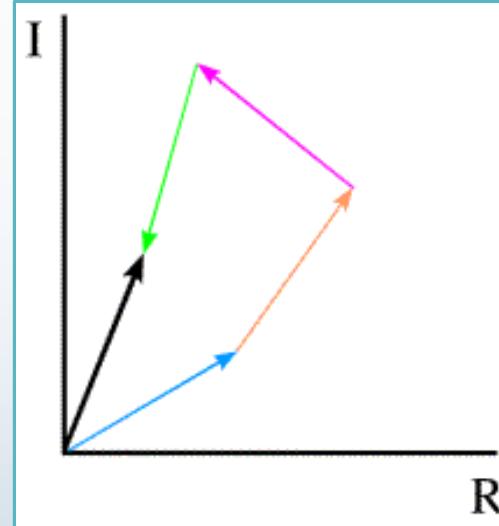
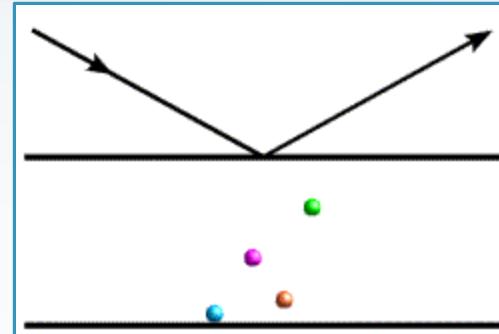
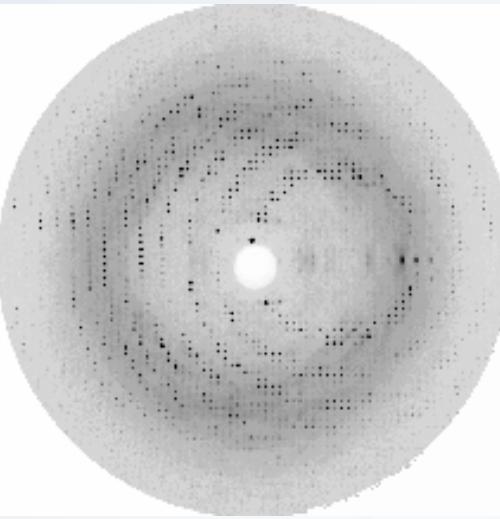
The Phase Problem

Solving a structure means solving "the phase problem".

Each reflection contains information on all atoms.
Each atom contributes to the intensity of each reflection.

To recombine a diffraction pattern, **both amplitude and phase** are required for each reflection.

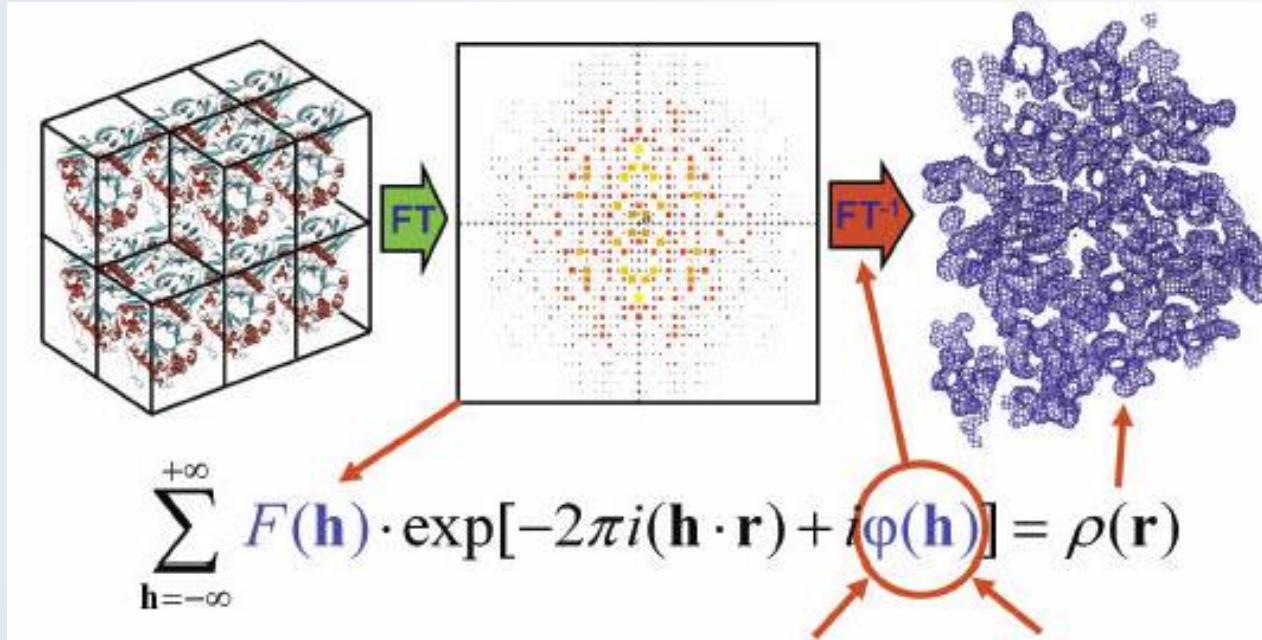
Only amplitudes can be recorded experimentally and **all phase information is lost**.



Structure Factor (F):
The vector representing the overall scattering from a particular set of Bragg planes.

Structure amplitude:
The magnitude of the structure factor.

5) Structure solution



Tackling the Phase Problem

1) Molecular Replacement (MR)

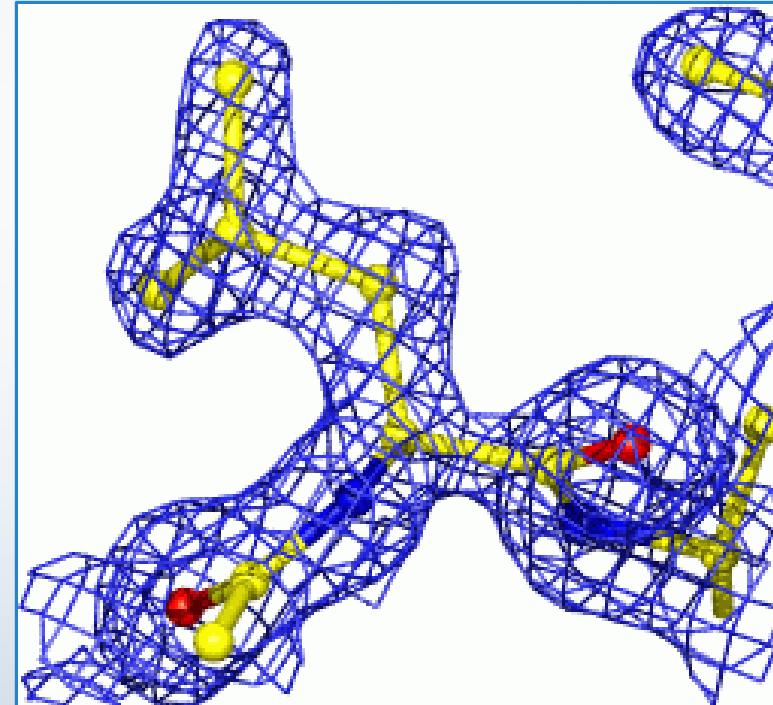
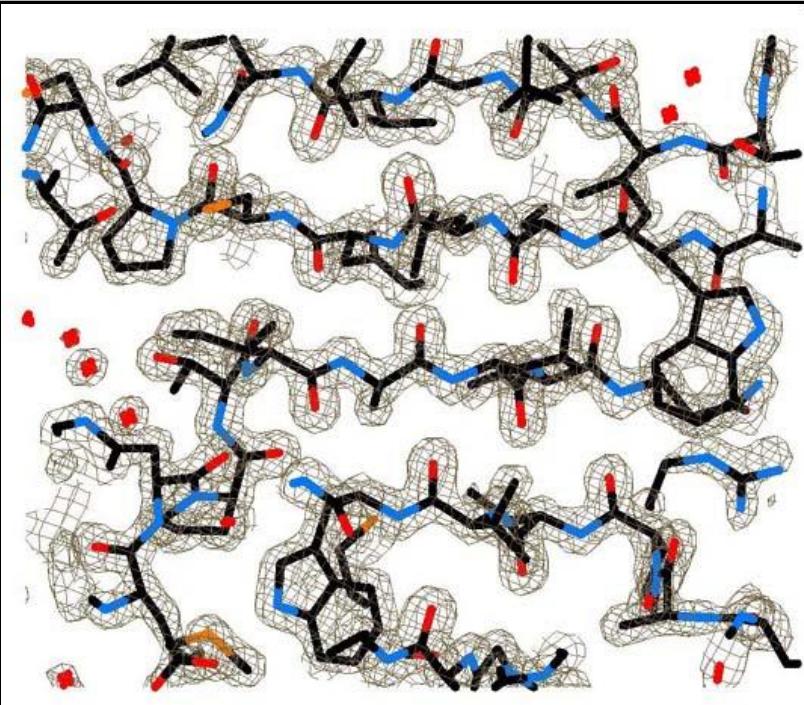
2) Multiple Isomorphous Replacement (MIR)

3) Multiwavelength Anomalous Dispersion (MAD)

6) Model building

The process where the electron density map is interpreted in terms of a set of atomic coordinates

contour lines show the electron density contoured at its root-mean-square (RMS) level.



The amount of detail that is visible is dependent on the resolution and the quality of the phases.

Resolution is reasonably high (1.8 \AA) and it is fairly easy to see that this blob of density is likely to be due to a leucine.

6) Model building

Unfortunately, the **impact of phases** on the appearance of the electron density is **much greater** than that of the **intensities**

With **perfect phases**, even at **around 4 Å resolution**, map interpretation would not be a major problem (this situation often arises in the study of viruses)

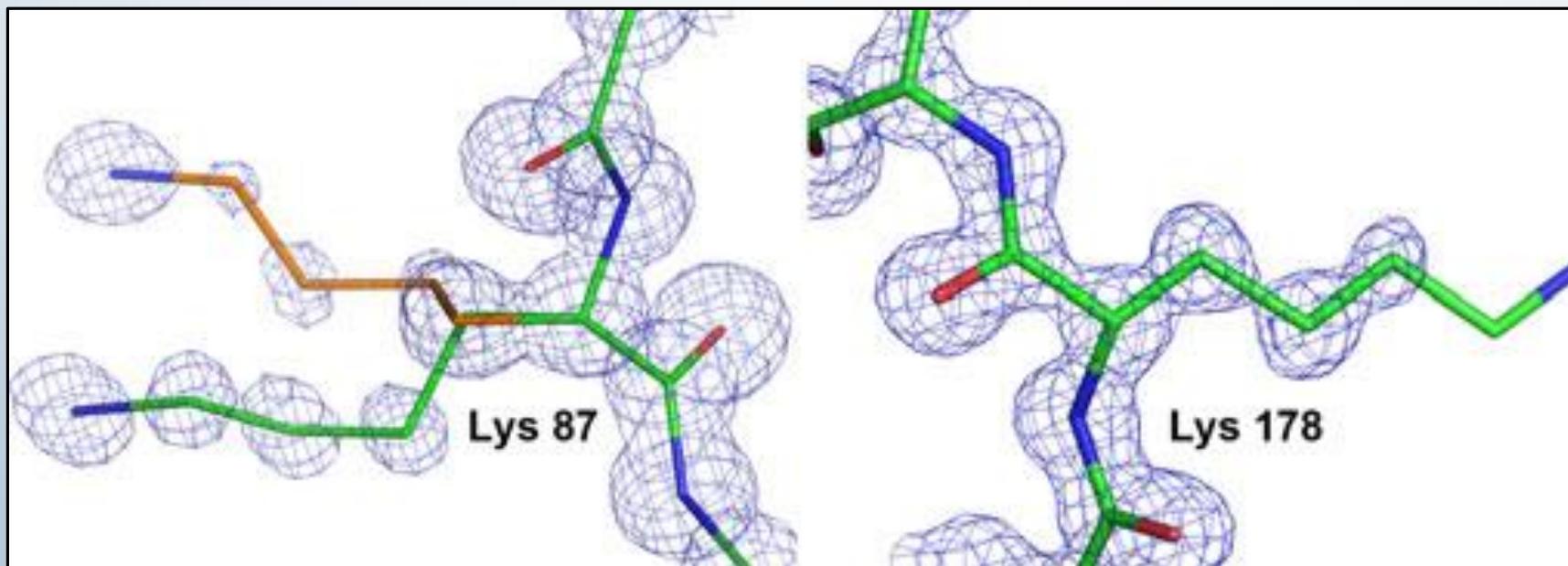
Regions of high flexibility are not visible at all due to:

1) static disorder:

Structure varies from one molecule to the next within crystal

2) Dynamic disorder:

The region is mobile within the crystal
(eliminated in cryogenic data collection)



7) Refinement

Adjustment of parameters of the model **to improve the fit** to the data

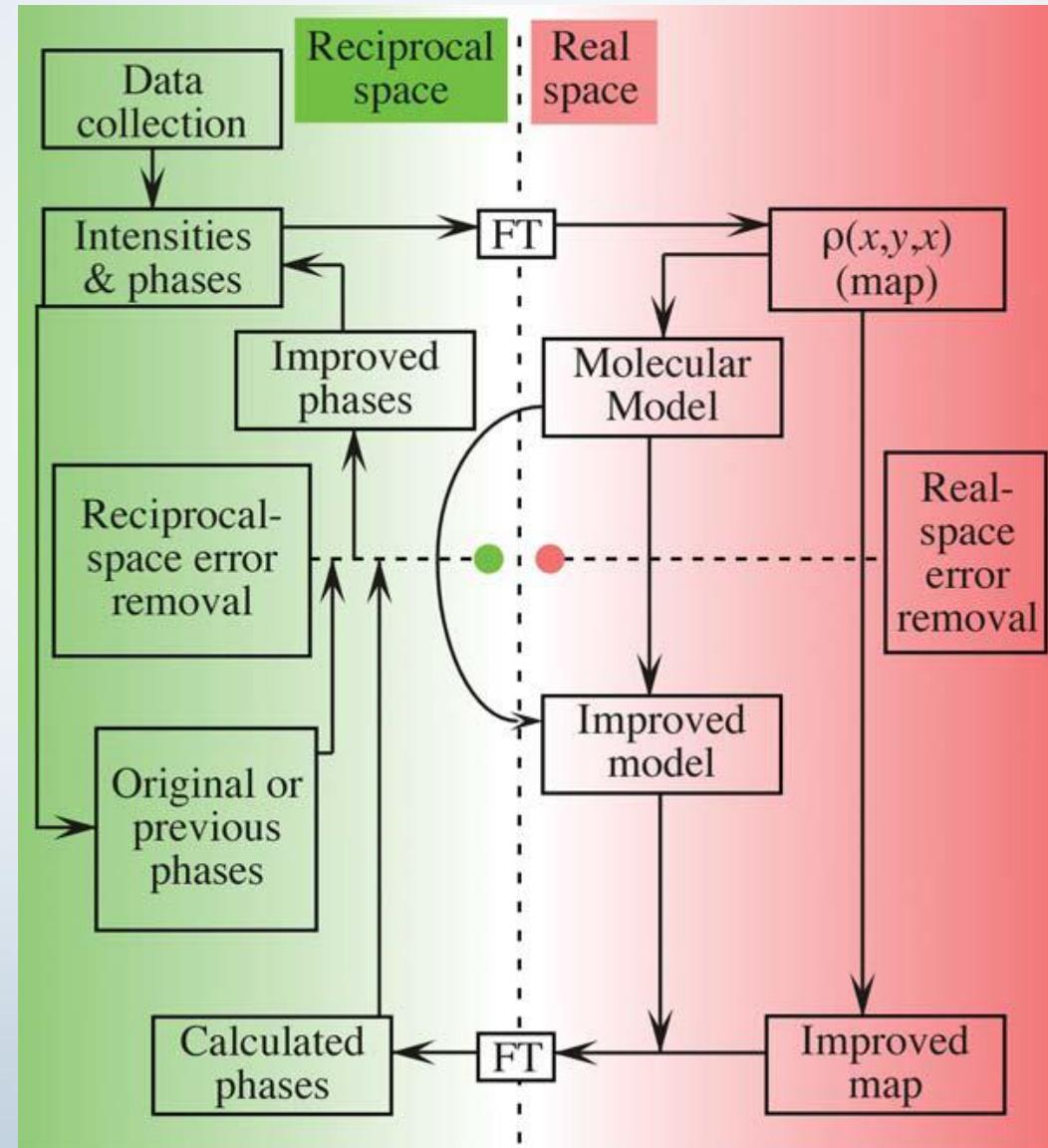
improving the phases which results in **clearer maps** and therefore better models.

go round this cycle **several times** until get little or no further improvements.

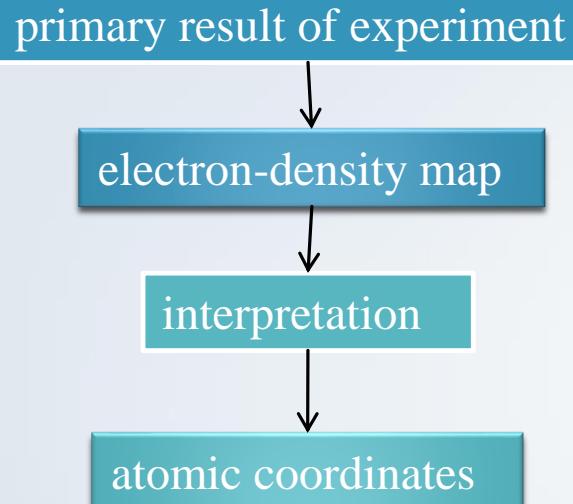
At this stage we would expect a crystallographic **R-factor of below 25%**

final model must make **chemical sense** and there must be no large regions of electron density unaccounted for.

Apply chemical common sense in the form of stereochemical restraints.



8) Validation

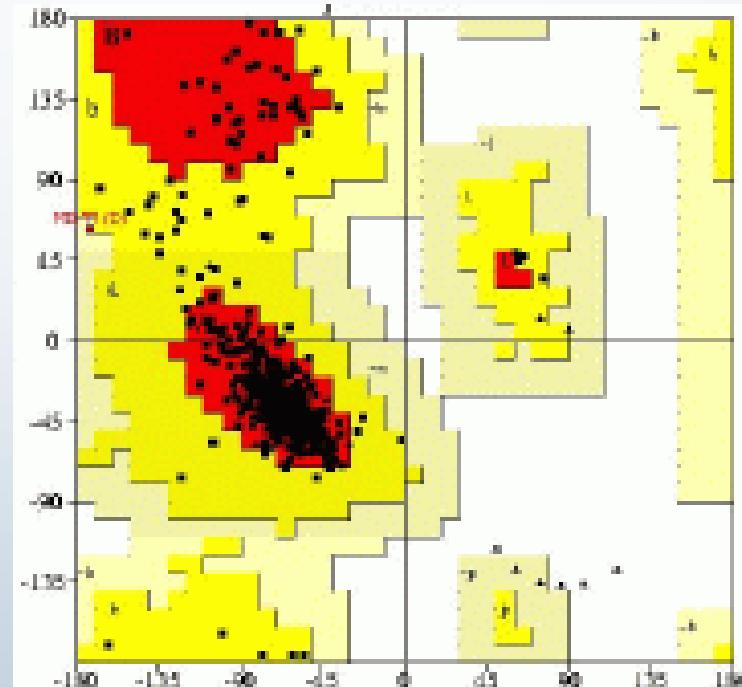


A thorough check to see if the model looks more and more like a “**real protein**”.

Check the distribution of the main-chain torsion angles **phi and psi**

For most protein structures, there is a **very poor observation-to-parameter ratio**

With enough parameters you can **fit an elephant**



9) Deposition

1) Cambridge Structural Database (CSD)

small molecules (<http://www.ccdc.cam.ac.uk>)

2) Inorganic Crystal Structure Database (ICSD)

inorganic compounds (<http://www.fiz-karlsruhe.de/icsd.html>)

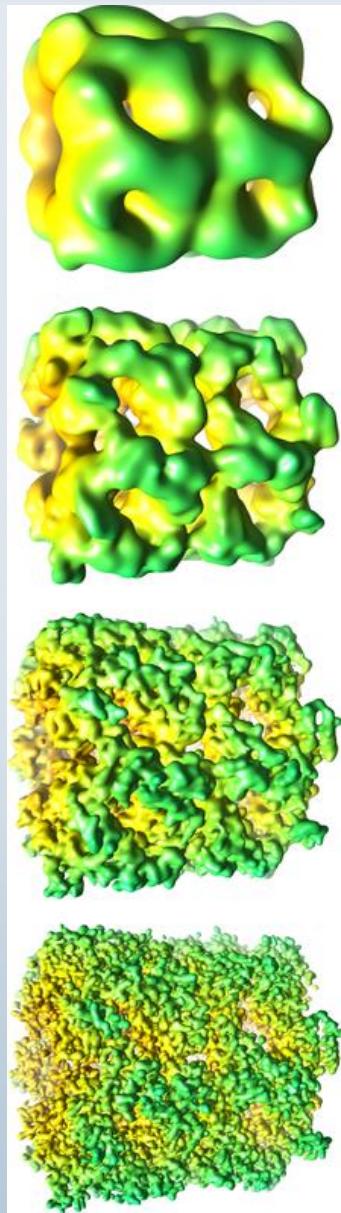
3) Protein Data Bank (PDB)

biomolecular structures (www.wwpdb.org)

The screenshot shows the homepage of the Inorganic Crystal Structure Database (ICSD). At the top, there is a navigation bar with links for 'COMMUNITY', 'RESEARCH & CONSULTANCY', 'SOLUTIONS', 'NEWS & EVENTS', 'SUPPORT & RESOURCES', and 'THE CCDC'. A 'Website Feedback' input field is also present. Below the navigation bar, a large banner features the text 'Knowledge and discovery from crystal structure data' next to a 3D molecular model. The bottom of the page has a footer with the FIZ Karlsruhe logo and links for 'Home', 'FIZ Karlsruhe Home', 'Press Room', 'Sitemap', and 'Print page'.

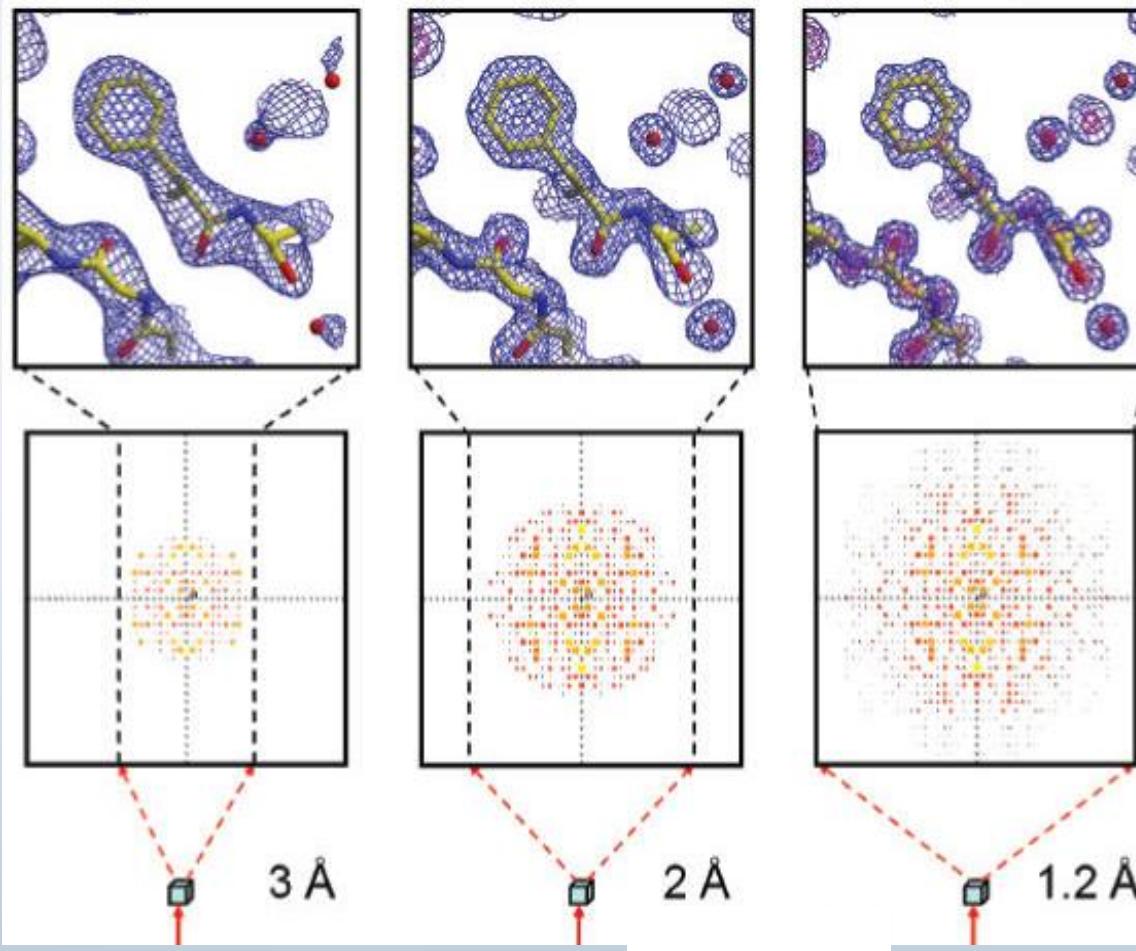
The screenshot shows the homepage of the Worldwide Protein Data Bank (wwPDB). It features the 'wwPDB' logo with a green circular icon containing a white 'w'. The text 'WORLDWIDE PROTEIN DATA BANK' is displayed above the logo. To the right, there is a stylized graphic of a protein structure against a blue background. Below the logo, the text 'Welcome to the Worldwide Protein Data Bank' is visible. At the bottom, a navigation bar includes links for 'Home', 'wwPDB Agreement', 'Statistics', 'FAQ', 'News', 'About Us', and social media icons for Facebook and RSS feed.

Resolution



Most protein crystals diffract between 1.8 and 3 Å.

To increase resolution grow better crystals and cryo-cool the crystals to near liquid nitrogen temperature.



A rough guide to the resolution of protein structures

> 4 Å

Individual coordinates meaningless.
Secondary structure elements can be determined.

3 – 4 Å

Fold possibly correct, but errors are very likely.
Many sidechains placed with wrong rotamer.

2.5 – 3 Å

Fold likely correct except that some surface loops might be mismodelled.
Several sidechains (lys, glu, gln, ser, val, thr) likely to have wrong rotamers.

2 – 2.5 Å

Number of sidechains in wrong rotamer is considerably less.
Fold normally correct and number of errors in surface loops is small.
Water molecules and small ligands become visible.

1.5 – 2 Å

Few residues have wrong rotamer.
Folds are rarely incorrect, even in surface loops.

0.5 – 1.5 Å

Structures have almost no errors.
Individual atoms in a structure can be resolved.
Rotamer libraries and geometry studies are made from these structures.

R-factors

How well the refined structure
predicts the observed data?

a measure of the agreement between the crystallographic
model and the experimental X-ray diffraction data

$$R = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}} \quad I_{hkl} \propto |F(hkl)|^2$$

R-factor, real space, or RSR

Measure of the similarity between an electron-density map calculated
directly from the model and one calculated from experimental data

$$\text{RSR} = \Sigma |\rho_{\text{obs}} - \rho_{\text{calc}}| / \Sigma |\rho_{\text{obs}} + \rho_{\text{calc}}|$$

Both Rmerge and R-factor are global indicators, showing the overall
agreement, between equivalent intensities or observed and calculated
amplitudes, and cannot be used to pinpoint individual poorly measured
reflections or local incorrectly modeled structural features

B – Factor

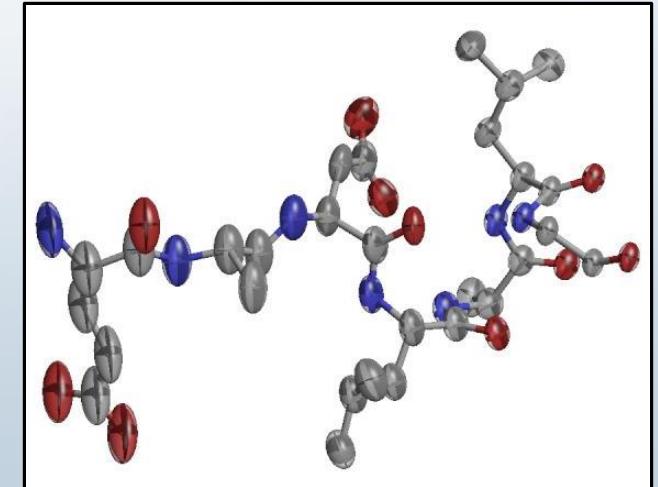
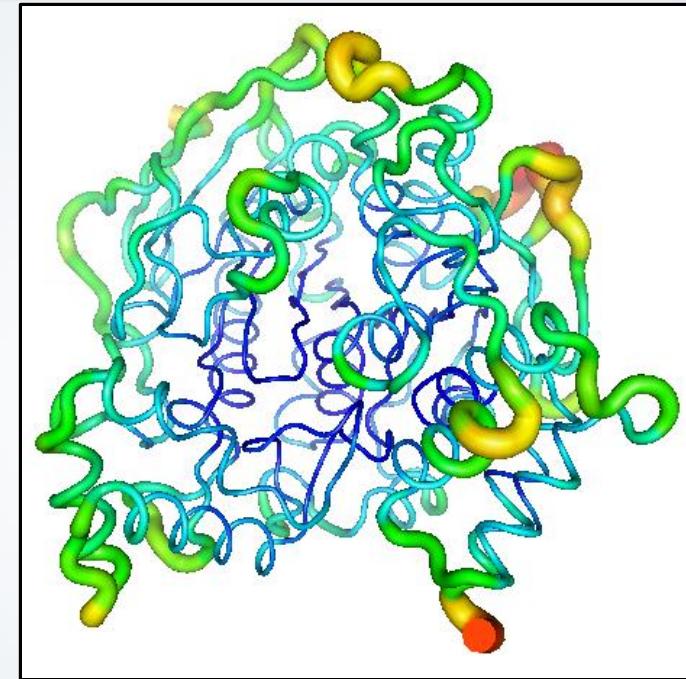
mean squared displacement is the most common measure of the spatial extent of random motion

$$B = 8\pi^2 \langle u^2 \rangle$$

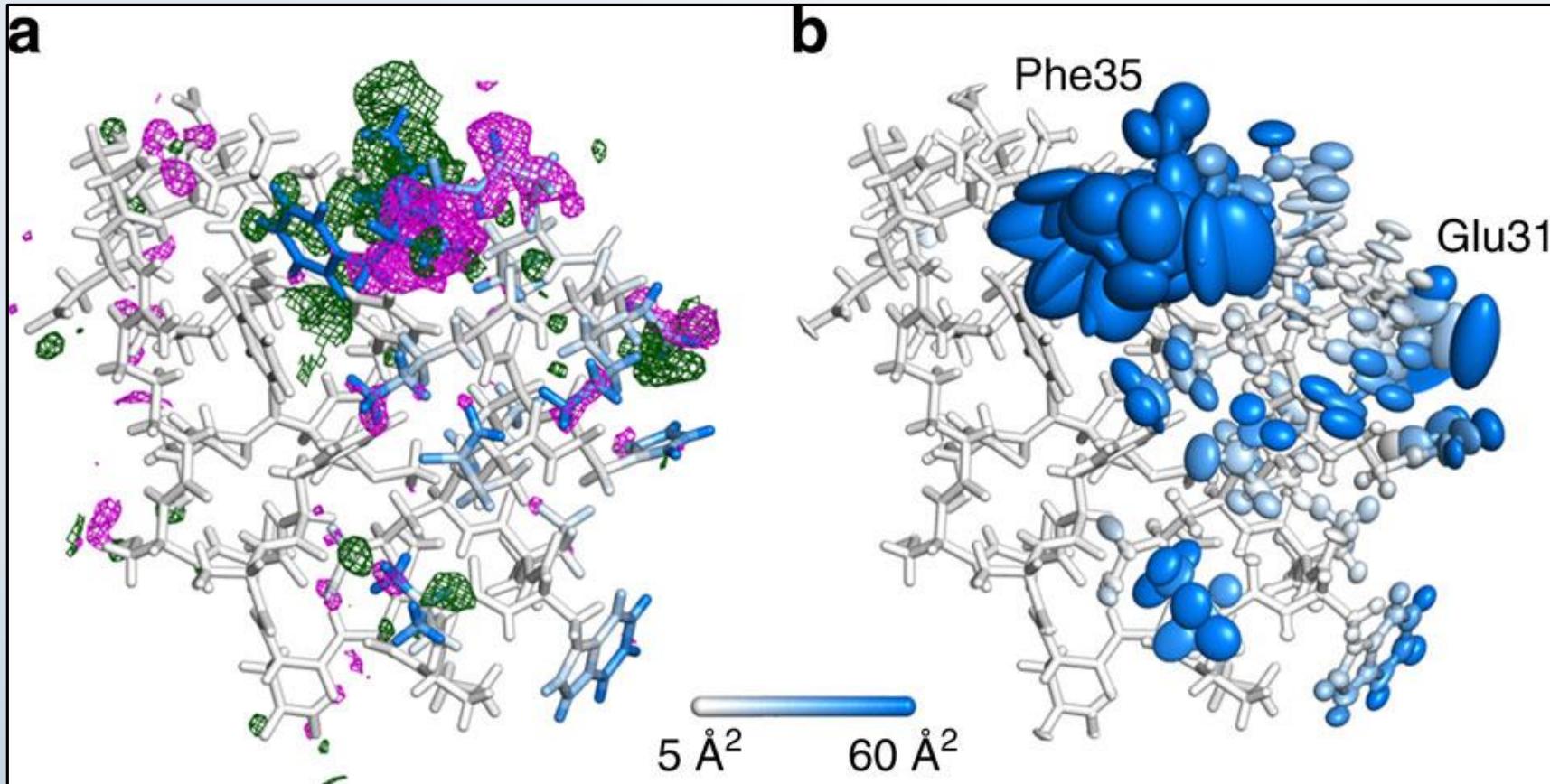
Atoms with low B-factors belong to a part of the structure that is well-ordered.

Atoms with large B-factors generally belong to part of the structure that is very flexible.

The B-factor model used is usually isotropic, i.e. describes only the amplitude of displacement, but more elaborate models describe the individual anisotropic displacement of each atom.

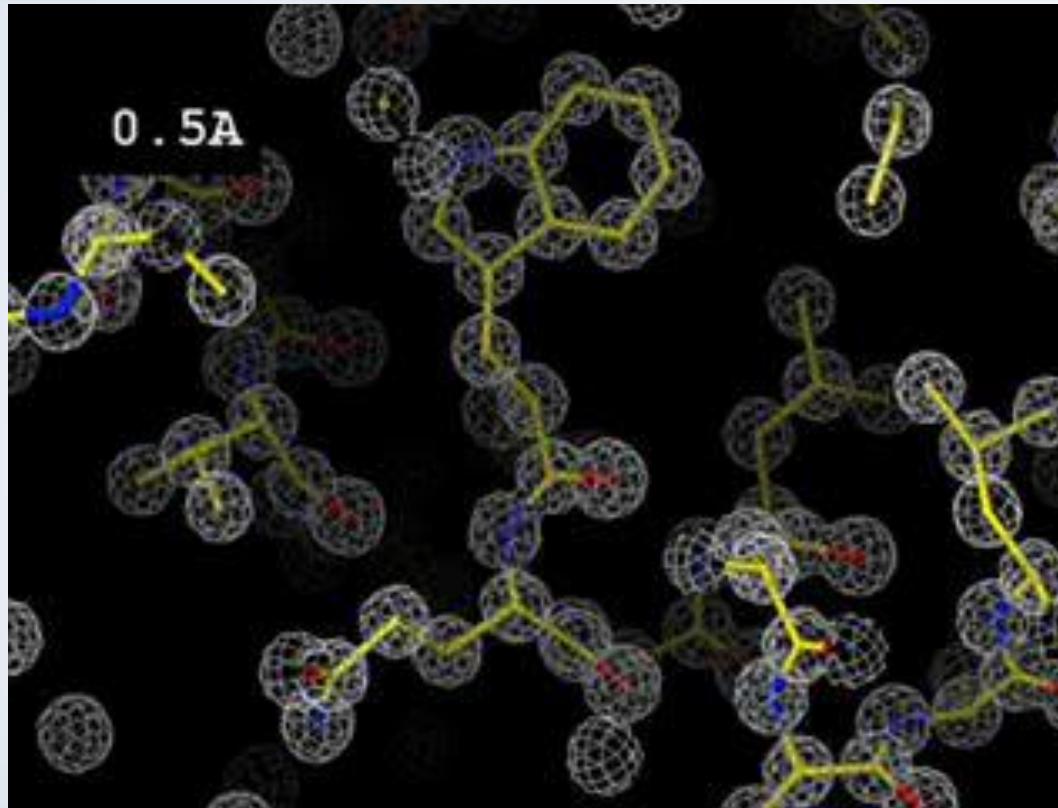


B – Factor



Refined electron density difference map ($F_o - F_c$)
and its accompanying structural model and the
same final model with its anisotropic ellipsoids.

The effect of resolution

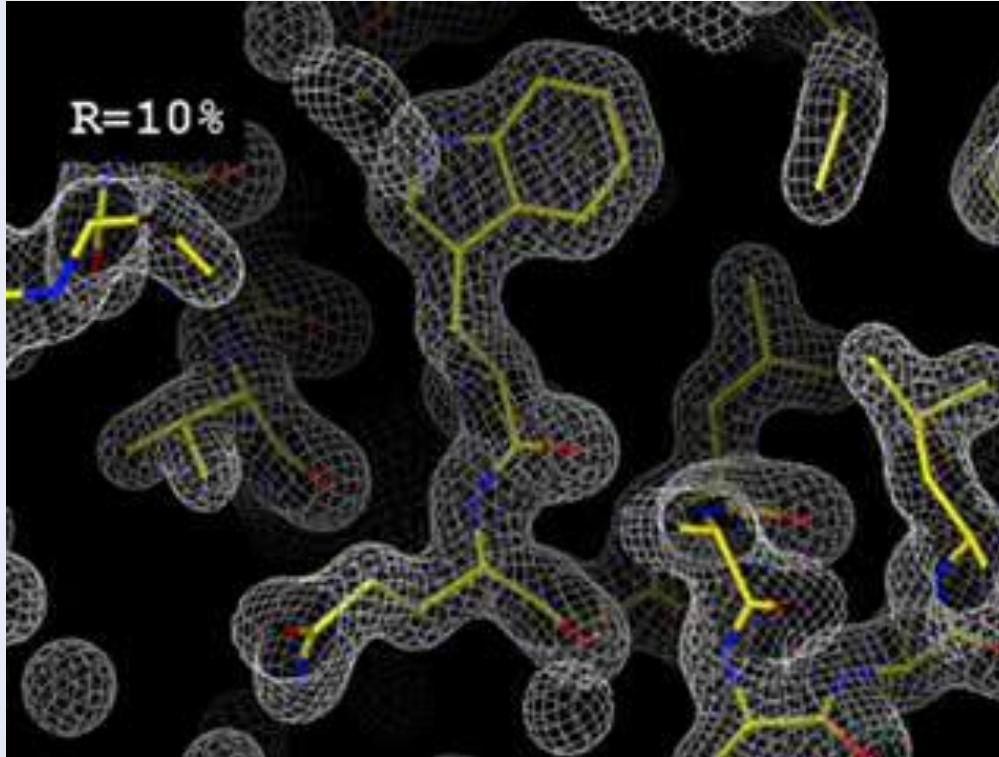


Calculated electron density map, contoured at 1 sigma, as the resolution limit is adjusted slowly from 0.5 Å to 6 Å.

The phases are perfect, and so are the amplitudes (R-factor = 0.0%) for all the resolutions displayed.

even for a perfect map, you expect side chains to poke out of density at 3.5 Å

The importance of amplitudes (intensities)



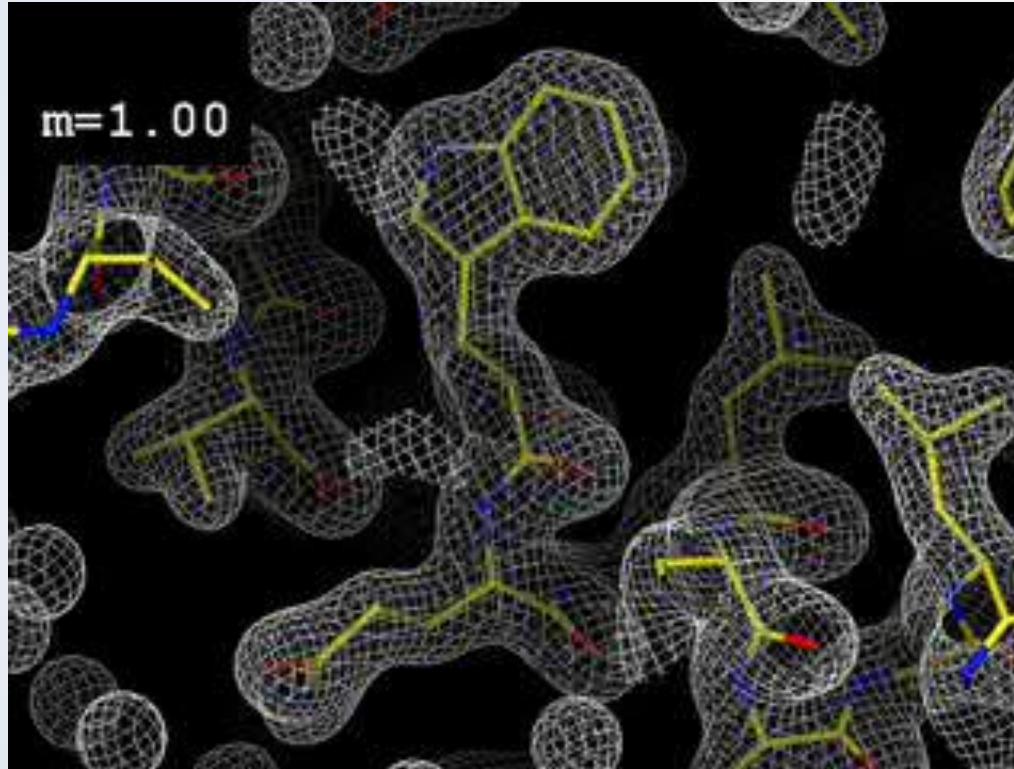
The slow changing of all the amplitudes to a different set of randomly selected values while holding the phases constant.

It is interesting to note that the map hardly changes at all until the R-factor gets higher than 30%.

The maximum R-factor you can get for two random data sets is 75%, which is end of movie.

The resolution here is 1.5 Å, and the phases are always perfect.

The importance of phases



Merging a perfect calculated map with another map, calculated with the same amplitudes, but with phases obtained from a model with randomly positioned atoms.

Always preserves the amplitudes, but changes the phases slowly to a new set of values.

The resolution here is 1.5 Å, and the R-factor is always 0.0 %.

Protein File Formats

Yazdan Asgari

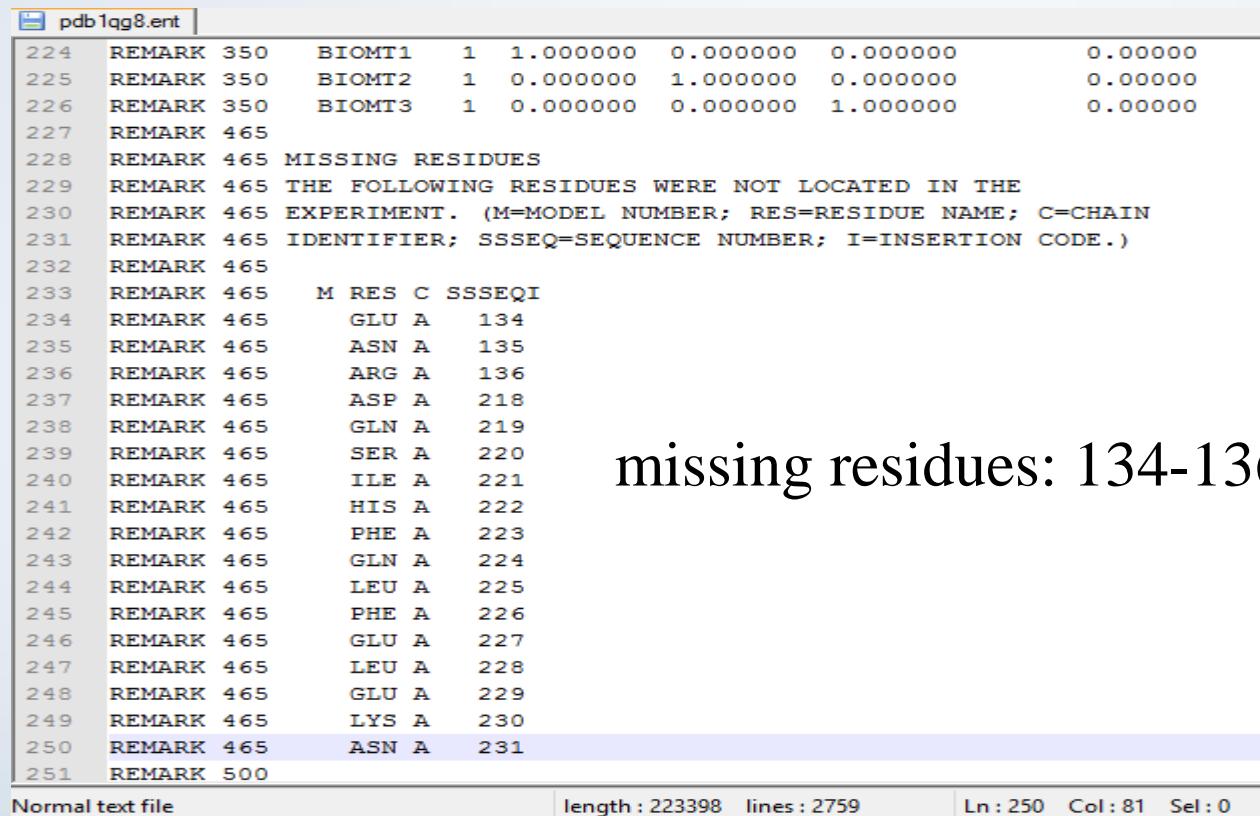
2019

PDB Format

Missing Residues

- Often, you will encounter files in the PDB which have **missing residues**.
- Special care must be taken in this case, as some softwares such as MODELLER **only** reads the ATOM and HETATM records, **not** the SEQRES records, and so will not handle missing residues automatically.
- Unfortunately PDB is **not reliable enough** to be able to automatically rely on SEQRES.

Missing Residues – Example: 1qg8



```
pdb1qg8.ent
224 REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.000000
225 REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.000000
226 REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.000000
227 REMARK 465
228 REMARK 465 MISSING RESIDUES
229 REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
230 REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
231 REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
232 REMARK 465
233 REMARK 465 M RES C SSSEQI
234 REMARK 465 GLU A 134
235 REMARK 465 ASN A 135
236 REMARK 465 ARG A 136
237 REMARK 465 ASP A 218
238 REMARK 465 GLN A 219
239 REMARK 465 SER A 220
240 REMARK 465 ILE A 221
241 REMARK 465 HIS A 222
242 REMARK 465 PHE A 223
243 REMARK 465 GLN A 224
244 REMARK 465 LEU A 225
245 REMARK 465 PHE A 226
246 REMARK 465 GLU A 227
247 REMARK 465 LEU A 228
248 REMARK 465 GLU A 229
249 REMARK 465 LYS A 230
250 REMARK 465 ASN A 231
251 REMARK 500

Normal text file length : 223398 lines : 2759 Ln : 250 Col : 81 Sel : 0
```

missing residues: 134-136 and 218-231

You can use softwares such as Modeller to fill in these missing residues by treating the original structure (without the missing residues) as a template, and building a comparative model using the full sequence.

Adding Residue to a PDB file

simple script example for Modeller

The screenshot shows a code editor window with two tabs at the top: "pdb1qg8.ent" and "missing_residues.py". The "missing_residues.py" tab is active, displaying the following Python script:

```
1  from modeller import *
2  from modeller.automodel import *      # Load the automodel class
3
4  log.verbose()
5  env = environ()
6
7  # directories for input atom files
8  env.io.atom_files_directory = ['.', '../atom_files']
9
10 class MyModel(automodel):
11     def select_atoms(self):
12         return selection(self.residue_range('1', '5'),
13                           self.residue_range('86', '100'),
14                           self.residue_range('291', '291'),
15                           self.residue_range('401', '404'),
16                           self.residue_range('440', '453'),
17                           self.residue_range('565', '582'))
18
19 a = MyModel(env, alnfile = '2e4u.ali',
20             knowns = '2E4U', sequence = '2e4u',
21             assess_methods=(assess.DOPE, assess.normalized_dope, assess.GA341))
22 a.starting_model= 1
23 a.ending_model = 2
24
25 a.make()
```

At the bottom of the editor, there is a status bar with the following information:

Python file | length : 851 lines : 25 | Ln : 25 Col : 9 Sel : 0

mmCIF Format

mmCIF Format

- It has been developed under the auspices of the International Union of Crystallography (IUCr), is to [extend the Crystallographic Information File \(CIF\)](#) data representation used for describing small molecule structures and associated diffraction experiments.
- It is referred to as the [macromolecular](#) Crystallographic Information File (mmCIF).
- The result of this effort was a core dictionary of data items sufficient for archiving the small molecule crystallographic experiment and its results.
- The format of the small molecule CIF dictionary and the data files based upon that dictionary conform to a restricted version of the [Self Defining Text Archive and Retrieval](#) (STAR) representation developed by Hall. STAR permits a data organization that may be understood by analogy with a spoken language.

How Does an mmCIF Data File Differ from a PDB File?

The mmCIF differs from the existing PDB file in the following important ways:

- **Level of detail**
 - ❑ Features either not present (e.g. description of the biological active molecule) or not accessible to the computer (e.g. experimental details found in REMARK records) are recorded in the mmCIF.
- **Level of description**
 - ❑ Each item of data in a mmCIF matches an entry in the mmCIF dictionary. The PDB Guide to Authors is the closest analogy to the mmCIF dictionary. However, the mmCIF dictionary is just another file with the same format as the mmCIF data file and is completely computer readable. This facilitates the validation of data and leads to a more consistent representation of structures.
- **Format**
 - ❑ This is the most difficult aspect of the transition for a macromolecular crystallographer since the format of data represented by an mmCIF **is significantly different** from the PDB format. It is this aspect of mmCIF that we shall look at first.



Validation and Deposition of Structural Data

Protein Data Bank

In 1971, the PDB archive was established with **seven** structures.
Scientists would submit their coordinate data to the PDB,
who would then **mail them by post** to interested users.

The wwPDB member organizations are:

- 1) the RCSB PDB (US)
- 2) Protein Data Bank Europe (PDBe, UK)
- 3) Protein Data Bank Japan (PDBj)
- 4) the BioMagResBank (BMRB, US).

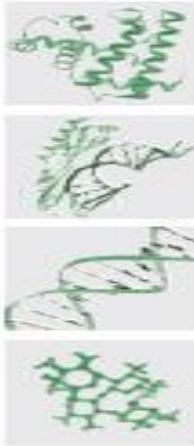


Protein Data Bank: Archive Contents

PDB ARCHIVE CONTENTS ON JULY 1ST, 2016

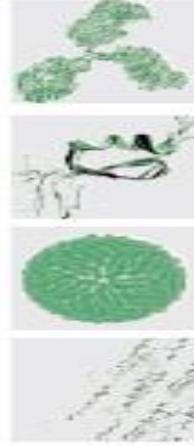
120,057 Released atomic coordinate entries

MOLECULE TYPE



111,440
Proteins, peptides,
and viruses
5,658
Protein/nucleic
acid complexes
2,933
Nucleic acids
26
Other

EXPERIMENTAL METHOD



107,264
X-ray crystallography
11,435
Nuclear Magnetic
Resonance
1,065
3D Electron
Microscopy (3DEM)
197
Other
96
Hybrid

RELATED EXPERIMENTAL DATA FILES

91,864
Structure factors
8,539
NMR restraints
2,297
Chemical shifts
905
3DEM maps

ACCESS



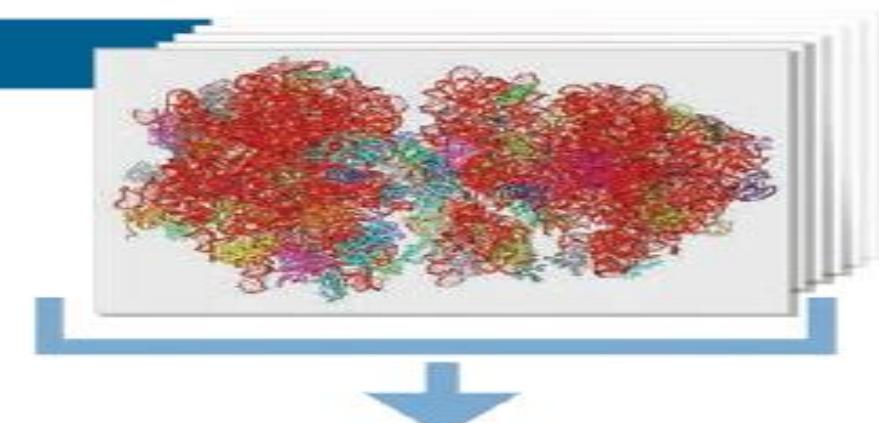
Each month in 2015, rcsb.org
was visited **741,000**
times on average by
315,000 unique visitors



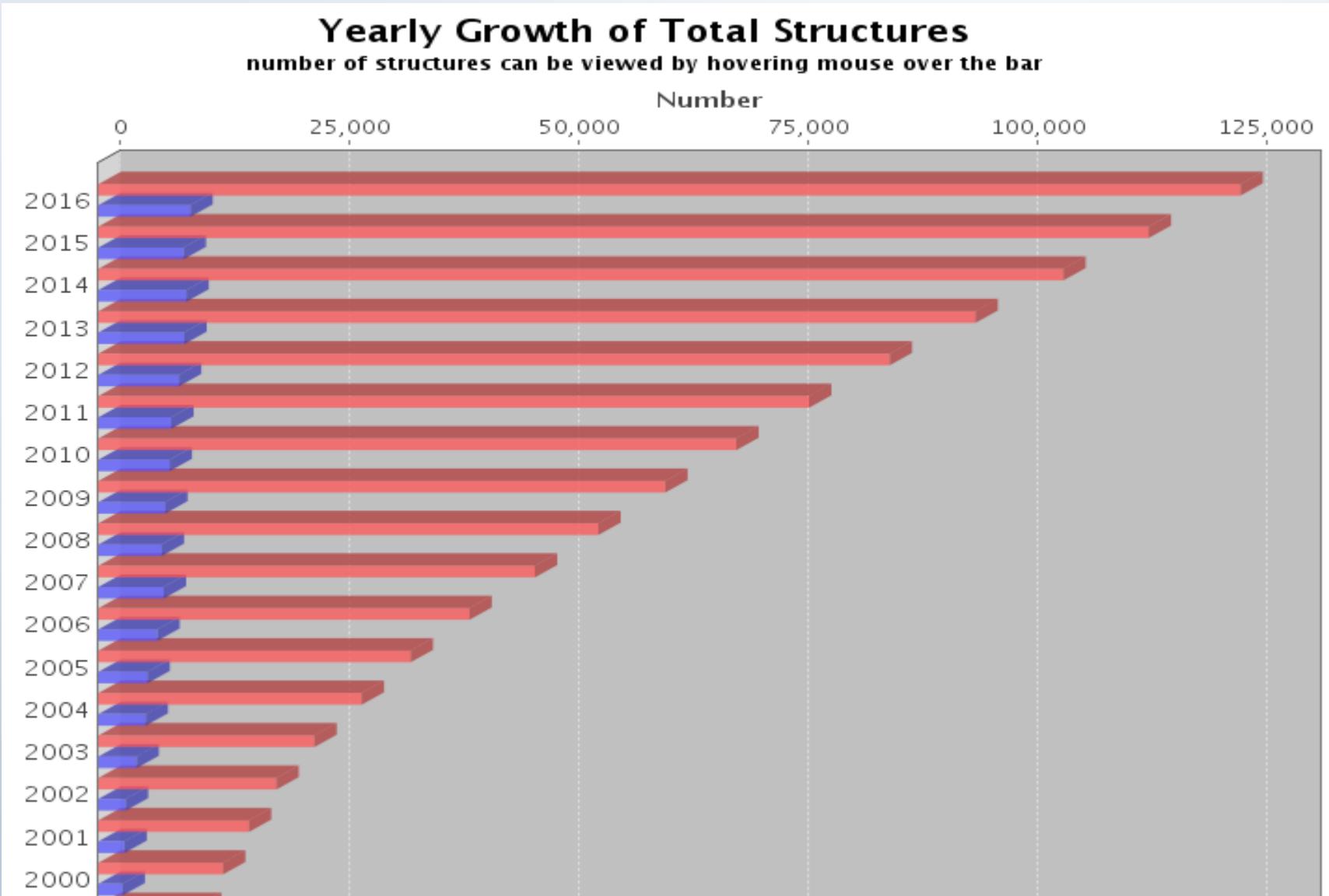
A total of **35,260 GB**
of data were accessed

PDB data are accessed
from **192** of the 195

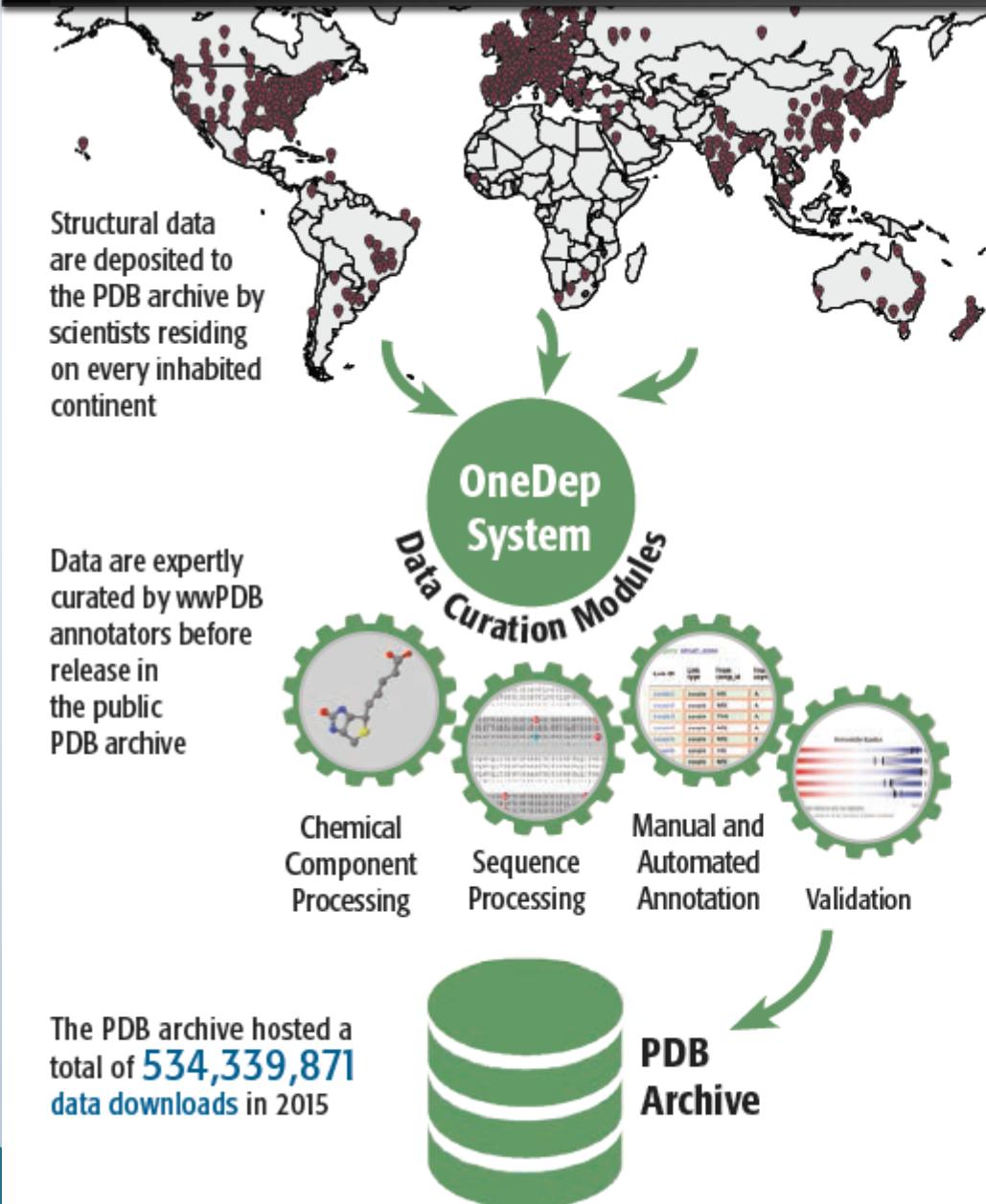
More than **1.5 million PDB structures**
are downloaded **EVERY day of the year.**



Protein Data Bank: Yearly Growth



Protein Data Bank: Deposition and Annotation



recommendations from wwPDB expert task forces.

Global adoption of the OneDep system has already resulted in the improved representation of data in the archive.

In 2015, the wwPDB curated **10,956** deposited structures. The PDB archive is on track to receive **~12,000** new structures in 2016

Best Practices in Data Representation

The wwPDB brings together community experts to provide guidance on best practices. In 2015, the wwPDB/CCDC/D3R Ligand Validation Workshop brought co-crystal structure determination experts from academe and industry together with X-ray Crystallography and Computational Chemistry software developers to discuss and develop best practices for validation of co-crystal structures; editorial/refereeing standards for publishing co-crystal structures; and recommendations for ligand representation across the archive.

These recommendations have been published in *Structure*¹, providing guidance for data archiving, annotation, and validation practices of co-crystal structural data.

How to Assess the Quality

FEBS Journal 275 (2008) 1–21



REVIEW ARTICLE

Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures

Alexander Wlodawer¹, Wladek Minor^{2,3}, Zbigniew Dauter⁴ and Mariusz Jaskolski^{5,6}



FEBS Journal 280 (2013) 5705–5736

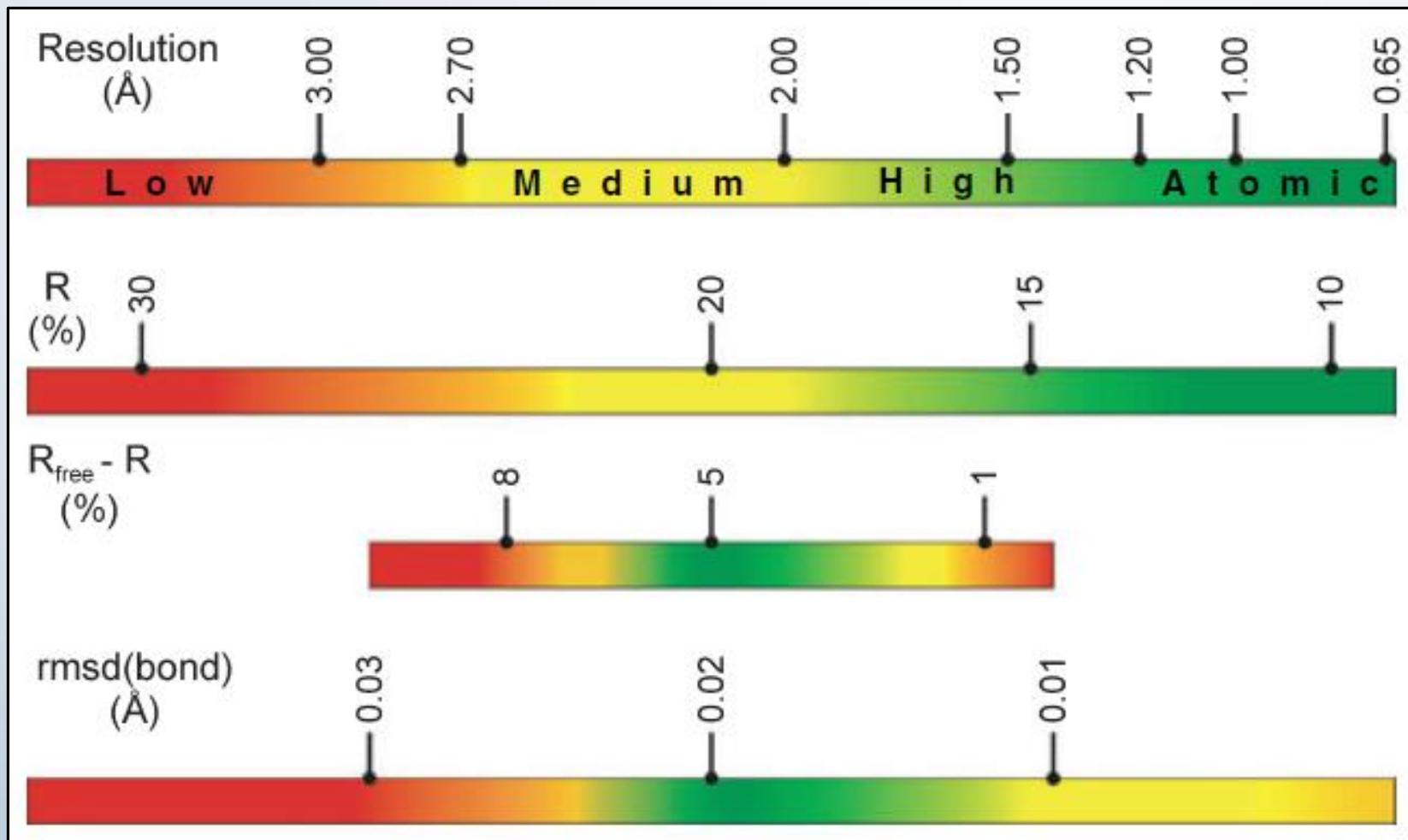


REVIEW ARTICLE

Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination

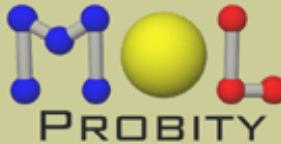
Alexander Wlodawer¹, Wladek Minor^{2,3,4,5}, Zbigniew Dauter⁶ and Mariusz Jaskolski⁷

How to Assess the Quality



Acta Cryst. (2010). D66, 12–21

MolProbity: all-atom structure validation for macromolecular crystallography



Main page

FILE UPLOAD/RETRIEVAL (MORE OPTIONS)

PDB/NDB code: type:

type:

MolProbity4 structure validation now provides many of its validation metrics through CCTBX, the open-source component of the Phenix crystallographic package. CCTBX allows for consistent validation results with Phenix, as well as added functionality, such as geometry regularization of NQH flips. Read more about this change [here](#).

We have updated Reduce to add hydrogens at a length more consistent with electron-cloud positions, and accordingly adjusted the Van der Waals radii in Probe to compensate for the change. This will affect comparison of results calculated with older versions of MolProbity.

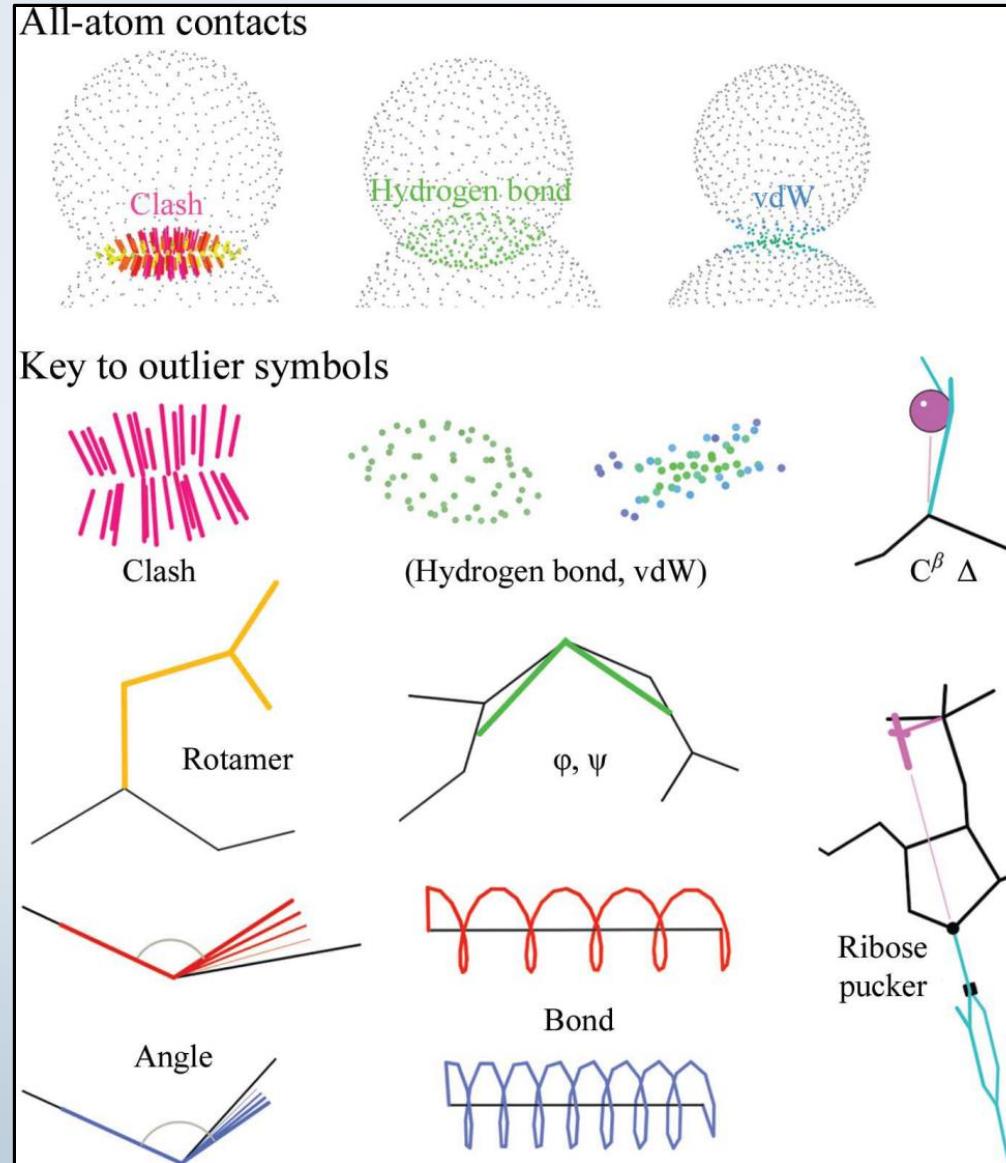
Main page
About hydrogens
Evaluate X-ray
Evaluate NMR
Fix up structure
Work with kins

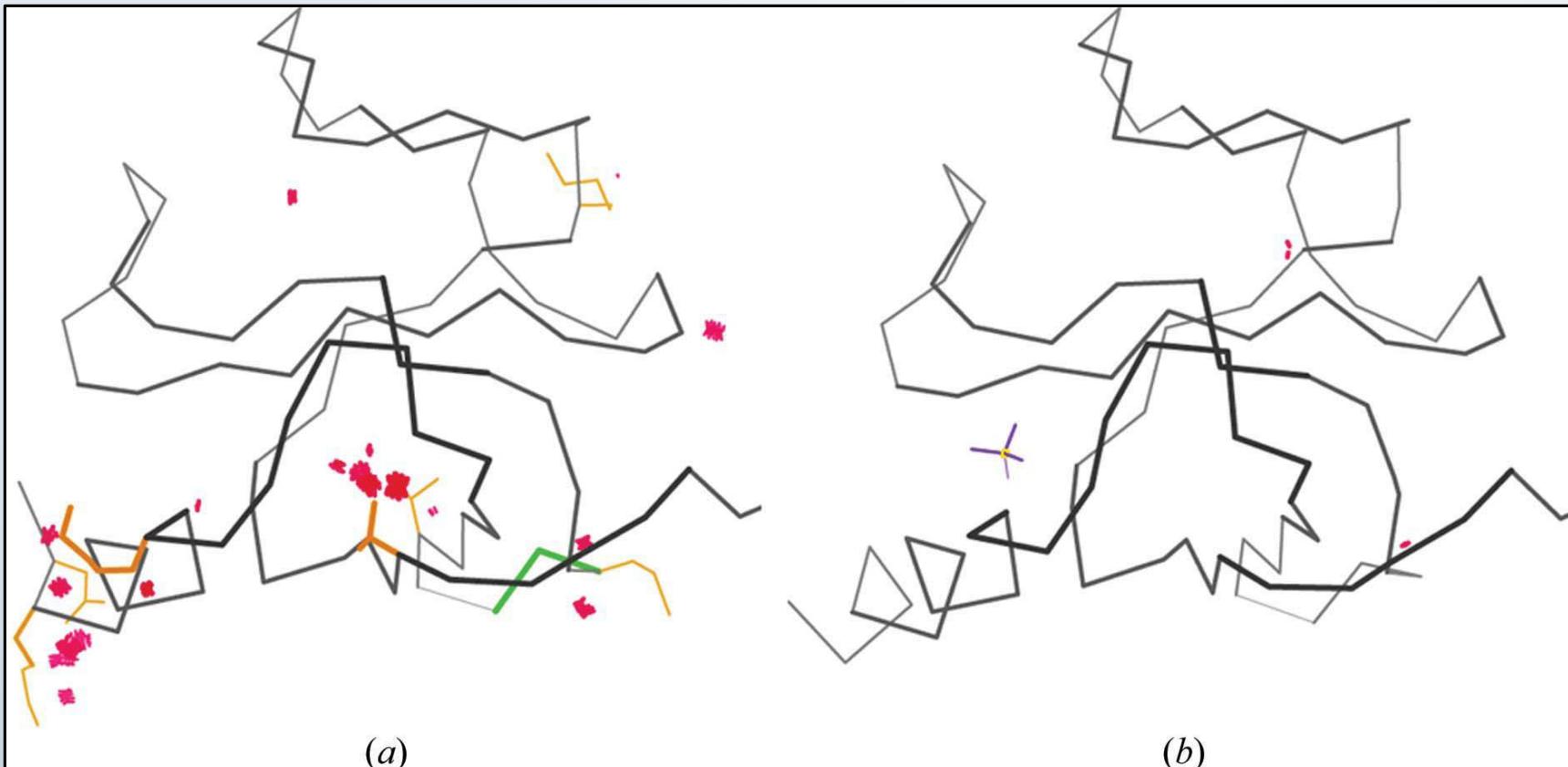
View & download files
Lab notebook
Feedback & bugs
Site map

Save session
Log out

You are using 0% of your 200 Mb of disk space.

MolProbity

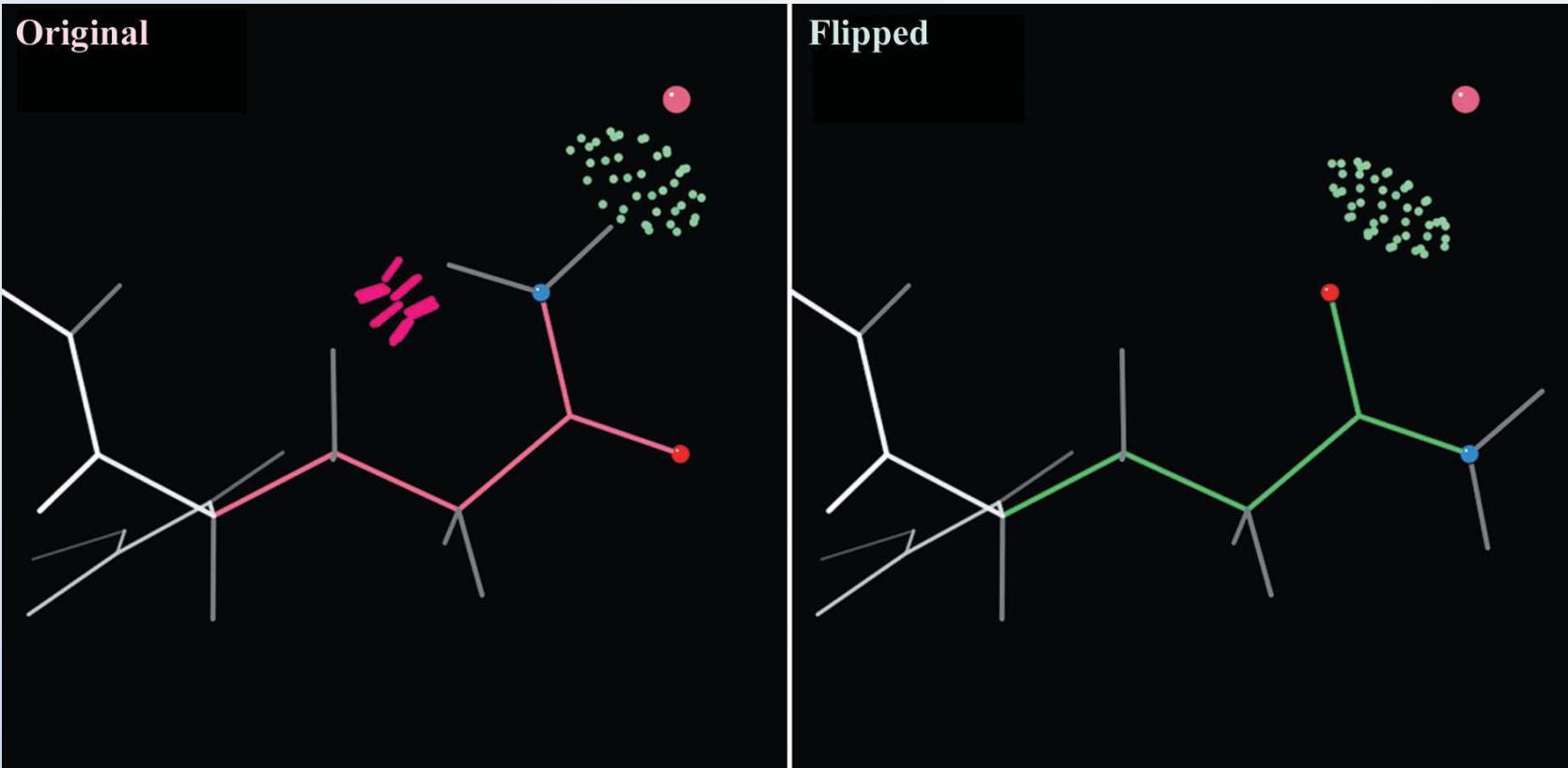




three major clusters of clash, rotamer
and Ramachandran problems plus a few
isolated outliers

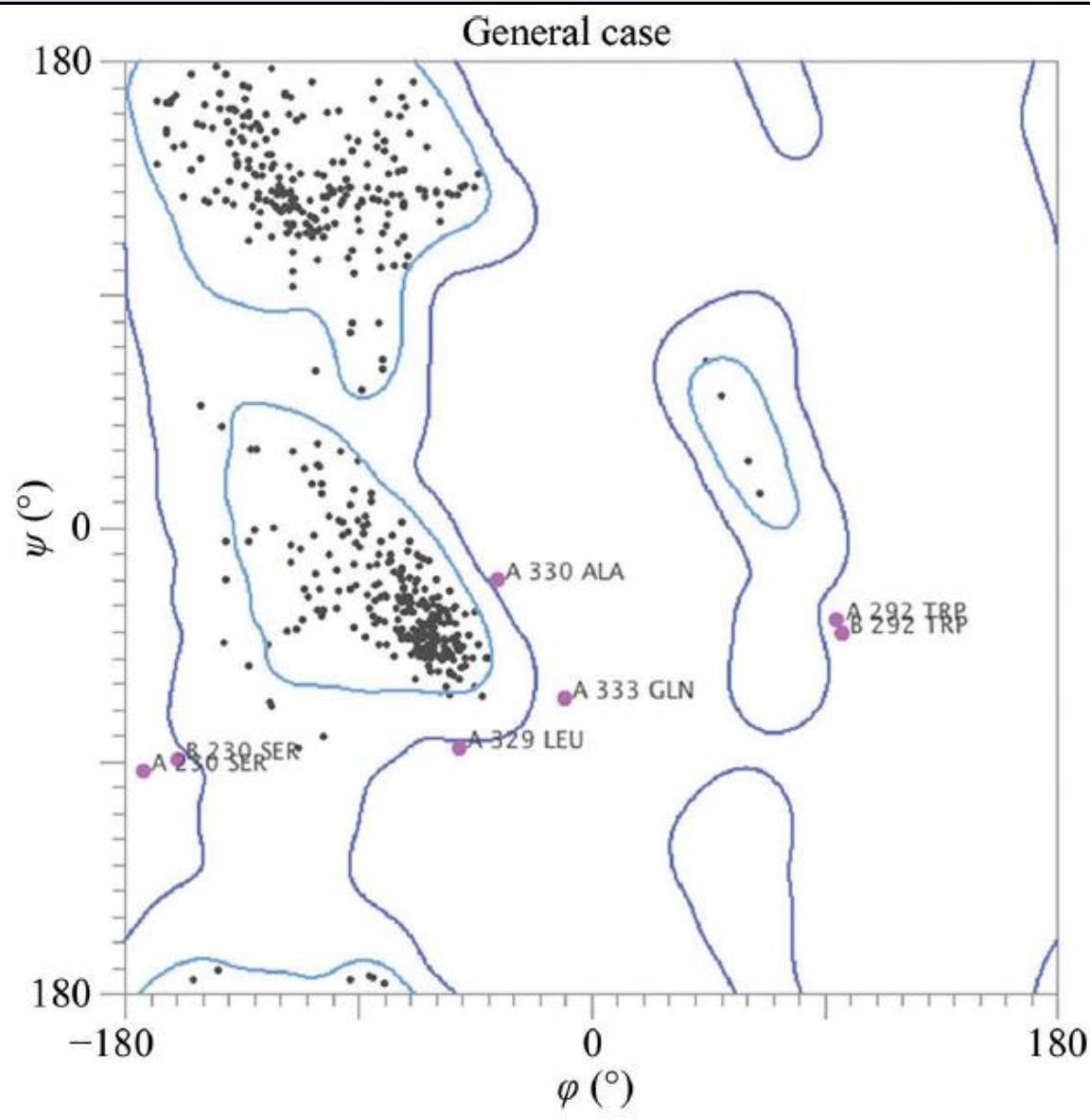
essentially no outliers, a 4% lower R_{free} , a
bound sulfate and an additional turn of
helix at the N-terminus

MolProbit



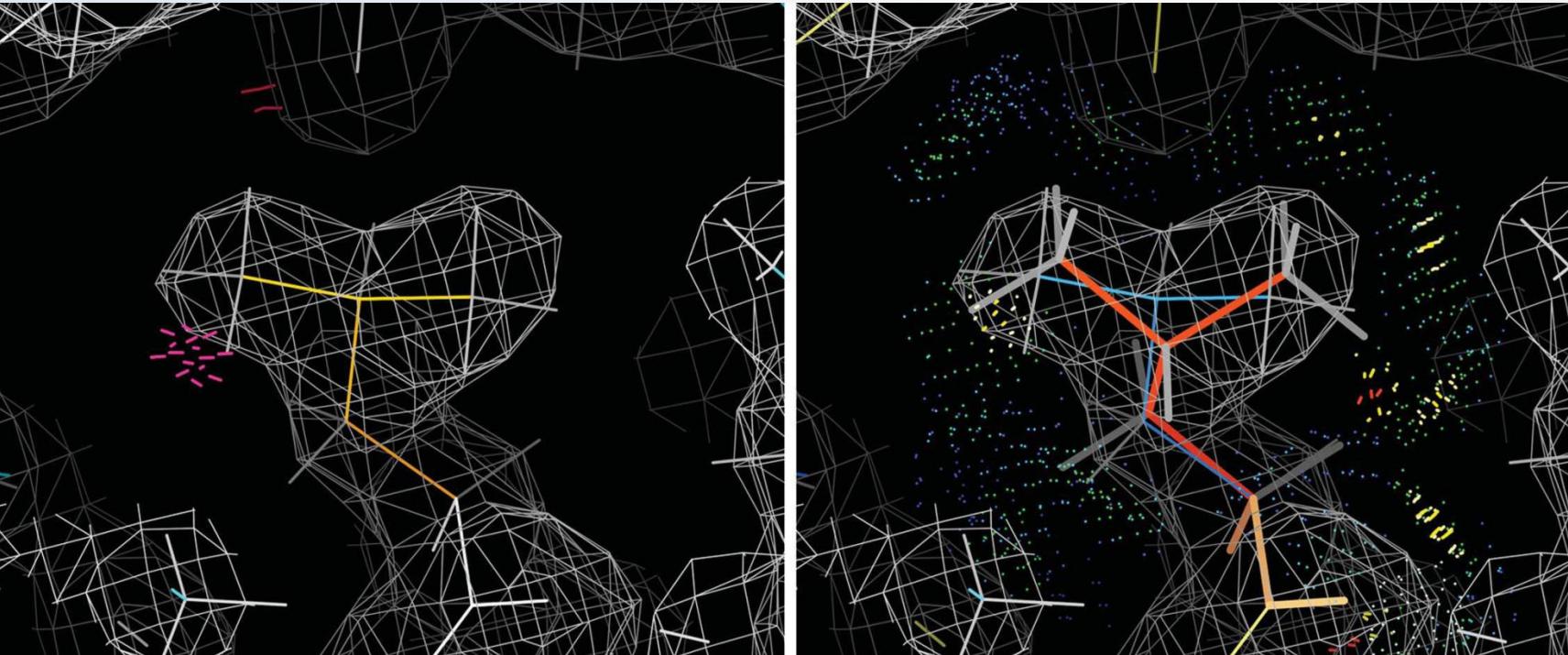
Both orientations make a hydrogen bond to the crystallographic water, but the original has a serious internal clash of the NH₂ group with its own CB hydrogen

MolProbit



MolProbity

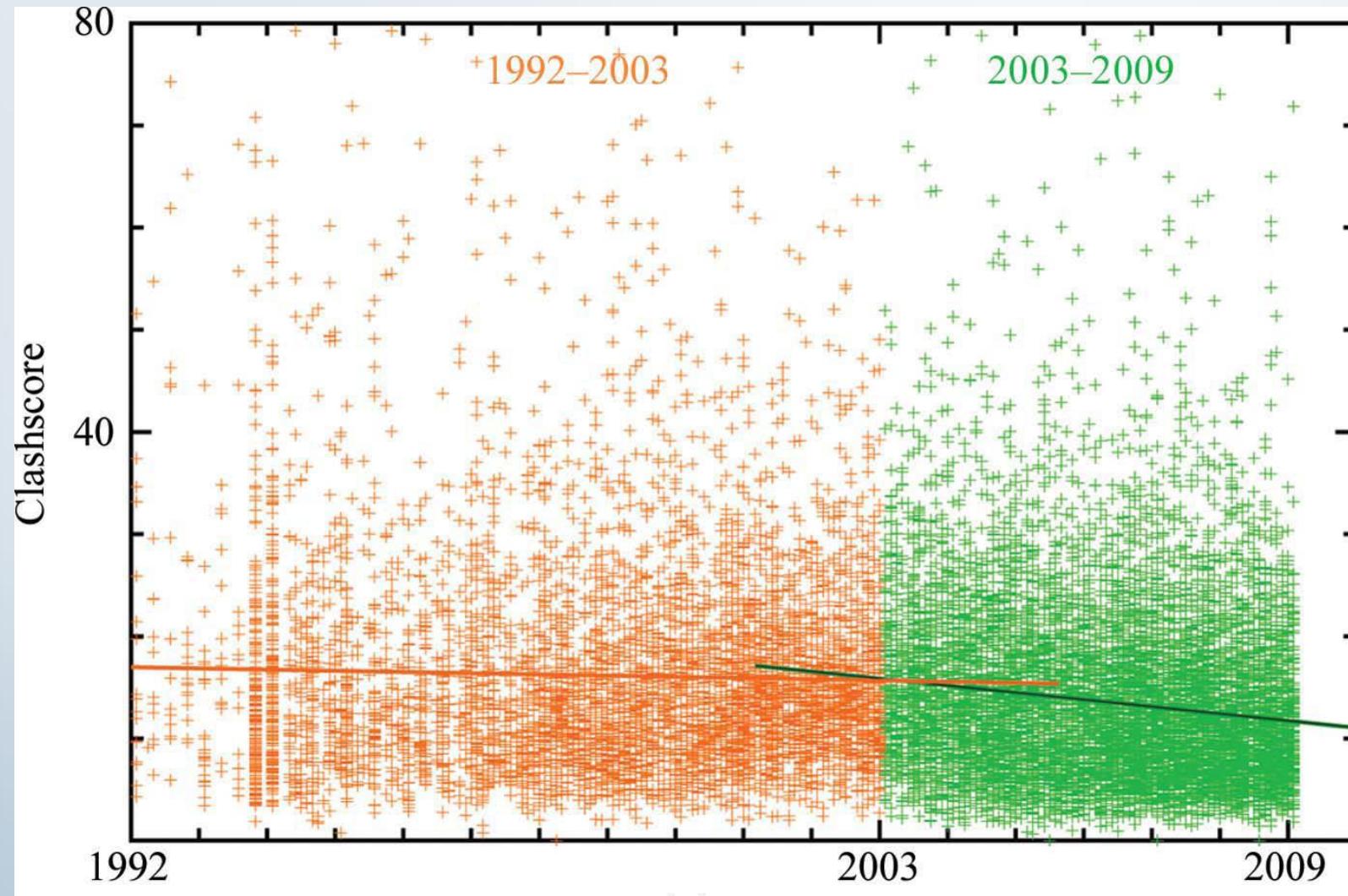
Rebuilding of a backward-fitted Leu side chain (1.7 Å resolution)



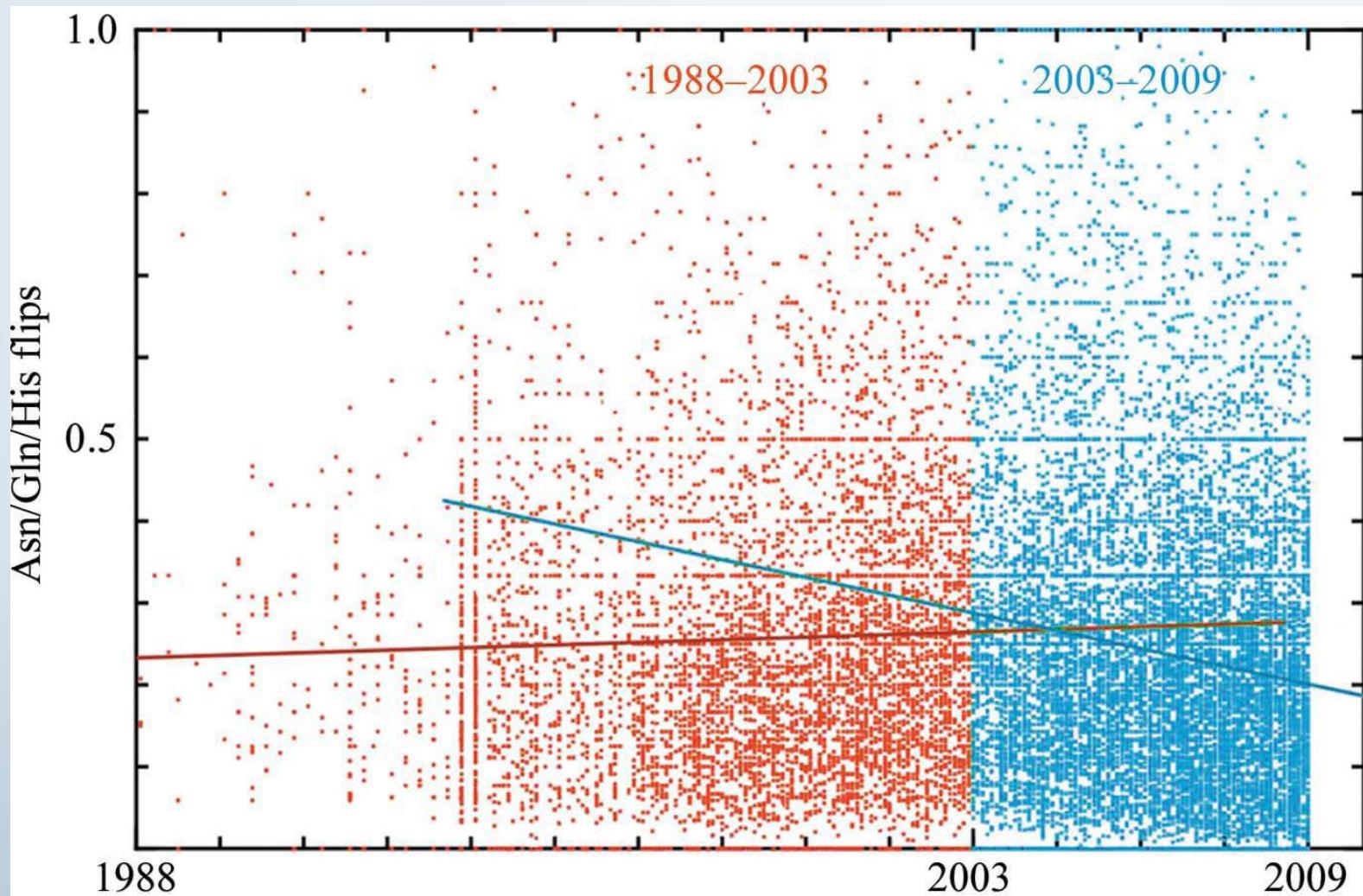
fits the density fairly well but is
a rotamer outlier with a clash

also fits the density well
with good all-atom packing

MolProbit



MolProbit



PDB Validation Report: Announcement

02-August-2013

Summarises the **quality** of the structure and highlights specific **concerns**

New wwPDB X-ray structure validation reports support depositors, journal editors and referees.

The wwPDB partners are pleased to announce that new, considerably more informative X-ray structure validation reports are now being provided to depositors as part of the structure annotation process. The reports can be used by depositors to assess the quality of their structures, and by journal editors and referees as a useful tool in the manuscript-review process.

The new reports implement recommendations of a large group of community experts on validation [1,2]. They were originally scheduled to be introduced as part of the **new wwPDB Deposition & Annotation system**, which will come online in 2014. However, initial feedback on the reports has been very positive and therefore the reports are provided for all new X-ray crystal structures deposited PDBe, PDBj, and RCSB PDB as of August 1, 2013.

"Validation 'at the gate' is crucial to improving the quality and consistency of the structural archive," said Gerard Kleywegt, Head of PDBe. "The new wwPDB validation reports for X-ray crystal structures draw on a wealth of community experience in the validation of models, experimental data and the fit of the model to these data. The new style reports are a huge improvement compared to the reports previously provided by the wwPDB partners. Initial feedback from depositors, annotation staff and users alike has been very positive and this has in fact motivated us to start delivering the reports much earlier than we had planned. We strongly encourage journal editors and referees to request these reports from depositors as part of the manuscript submission and review process."

PDB Validation Report: Criteria

Structure

Ways & Means



Structure 19, 1395–1412, October 12, 2011

A New Generation of Crystallographic Validation Tools for the Protein Data Bank

Randy J. Read,^{1,*} Paul D. Adams,² W. Bryan Arendall, III,³ Axel T. Brunger,⁴ Paul Emsley,⁵ Robbie P. Joosten,^{6,7} Gerard J. Kleywegt,^{8,9} Eugene B. Krissinel,^{9,10} Thomas Lütteke,^{6,11} Zbyszek Otwinowski,¹² Anastassis Perrakis,⁷ Jane S. Richardson,³ William H. Sheffler,¹³ Janet L. Smith,¹⁴ Ian J. Tickle,¹⁵ Gert Vriend,⁶ and Peter H. Zwart²

Conclusions of
X-ray
Validation Task
Force of
wwPDB

Validation
Criteria

Primary validation criteria were chosen to cover complementary aspects of experimental data:

- 1) model-to-data match**
- 2) geometry**
- 3) Conformation**
- 4) packing**

Most measures of model accuracy correlate strongly with resolution

Geometric and Conformational

Atomic and Molecular Interactions

Structure-Factor and Electron-Density

Non-Protein Components

Early depositions (before 1990):

< 1% of the entries

but

> 30% of the worst outliers on each criterion

PDB Validation Report: Scores

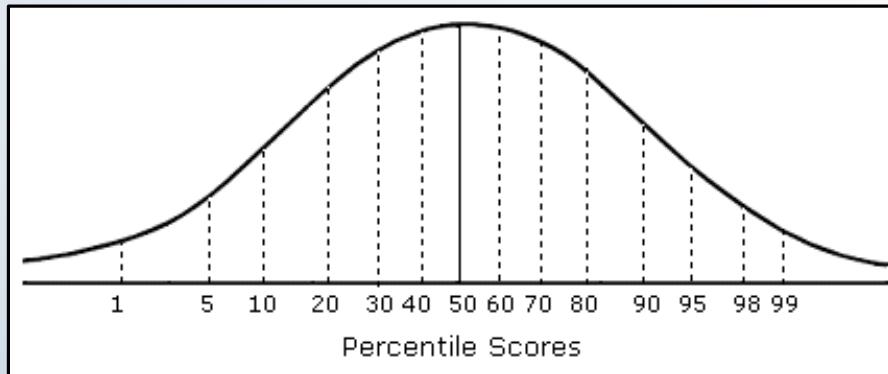
Scores relative to distributions

Percentiles:

Percentage of structures in the PDB that are worse



$$PR = \left(\frac{f_b + \frac{1}{2} f_w}{N} \right) * 100$$



Advantage:

Place different criteria on a **common scale** and no need to remember target values for individual criteria

Disconcerting Effect:

As the average quality of structures increase percentile scores **change over time**

Z-score:

Deviation from mean value, divided by estimated standard deviation

$$Z = \frac{x - \langle x \rangle}{\sigma(x)}$$

RMS-Z scores:

Root-Mean-Square value of the Z-scores for a criterion

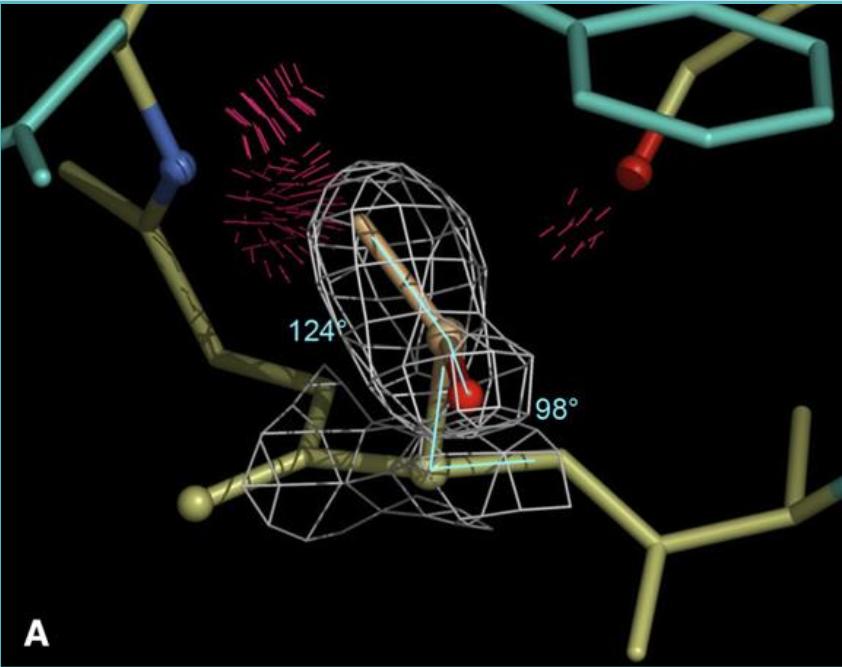
$$\text{RMS-Z} = \sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}$$

Amount of variation expected in each validation measure.

Populations:

- 1) Entire PDB structures
- 2) Structures at similar resolution
- 3) Set of small molecule structures

PDB Validation Report: Bonds, Angles, Planes, Chirality



A

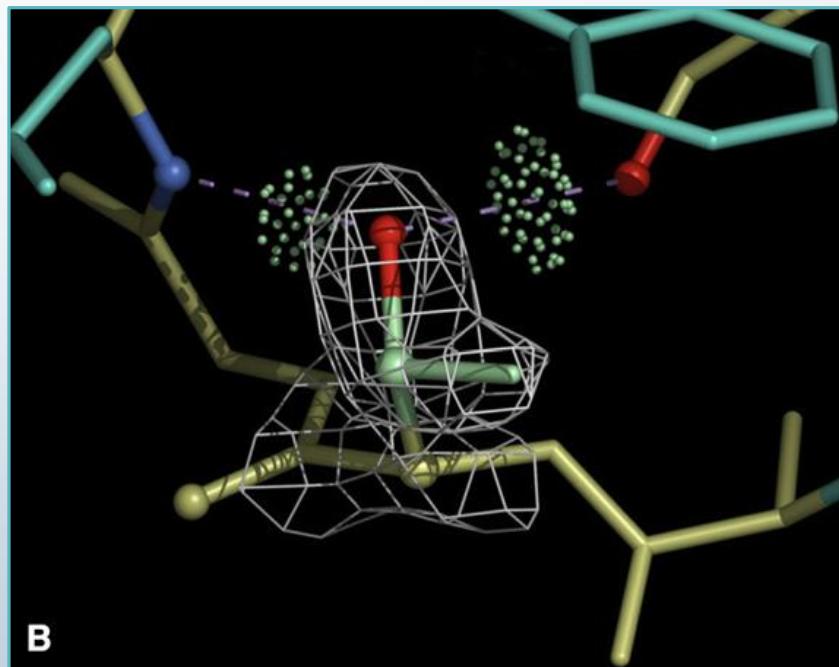
Thr32 in 1SBP at 1.7 Å Resolution

two bond-angle outliers at 5σ and 7σ

steric clashes, poor rotamer, no H-bond

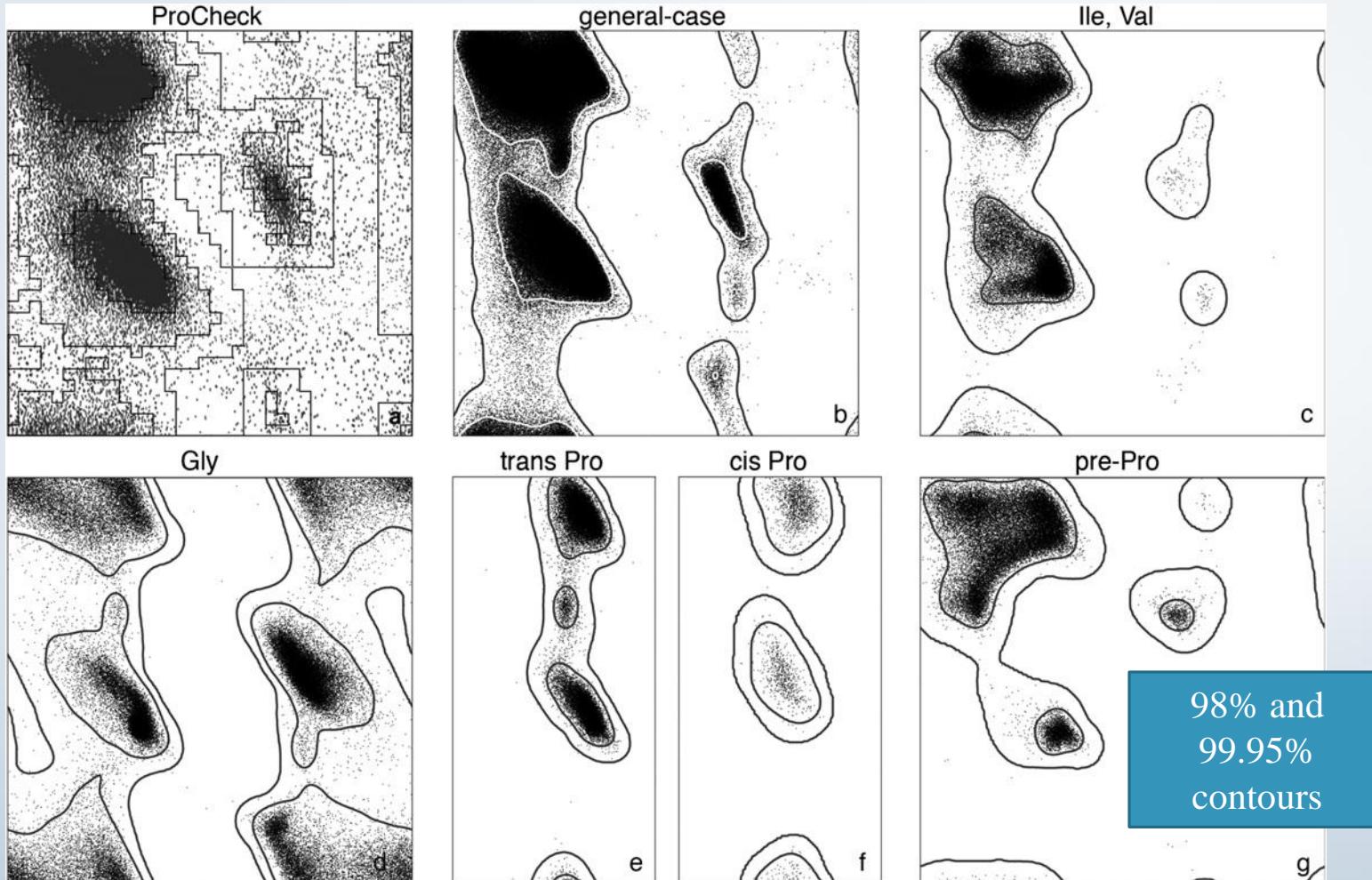
Side-chain has been turned 180°

Target values for means and standard deviations of bond lengths, angles, and planes were obtained by analyzing the high-resolution, small-molecule structures in the **Cambridge Structural Database**



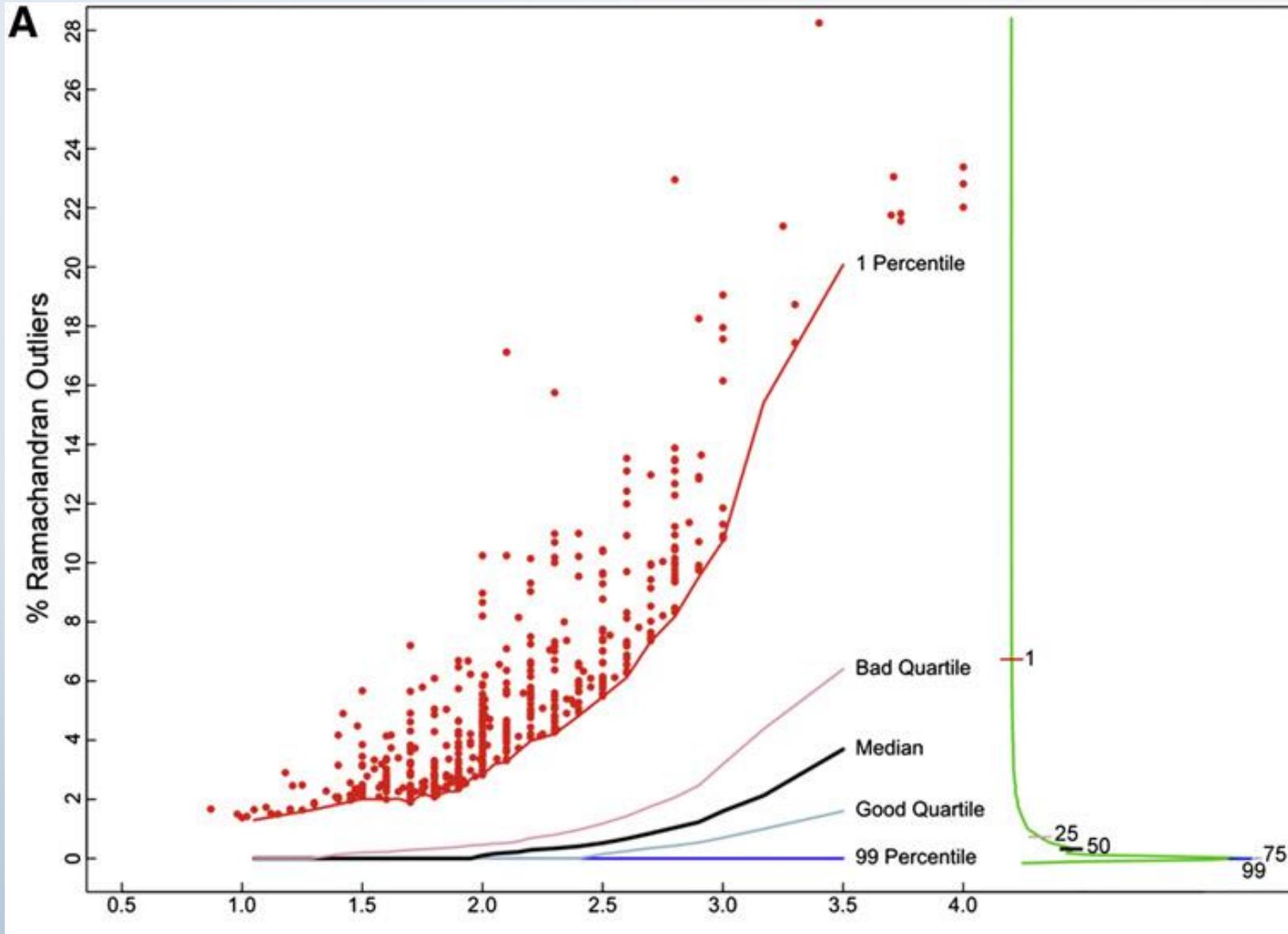
B

PDB Validation Report: Protein Backbone Conformation



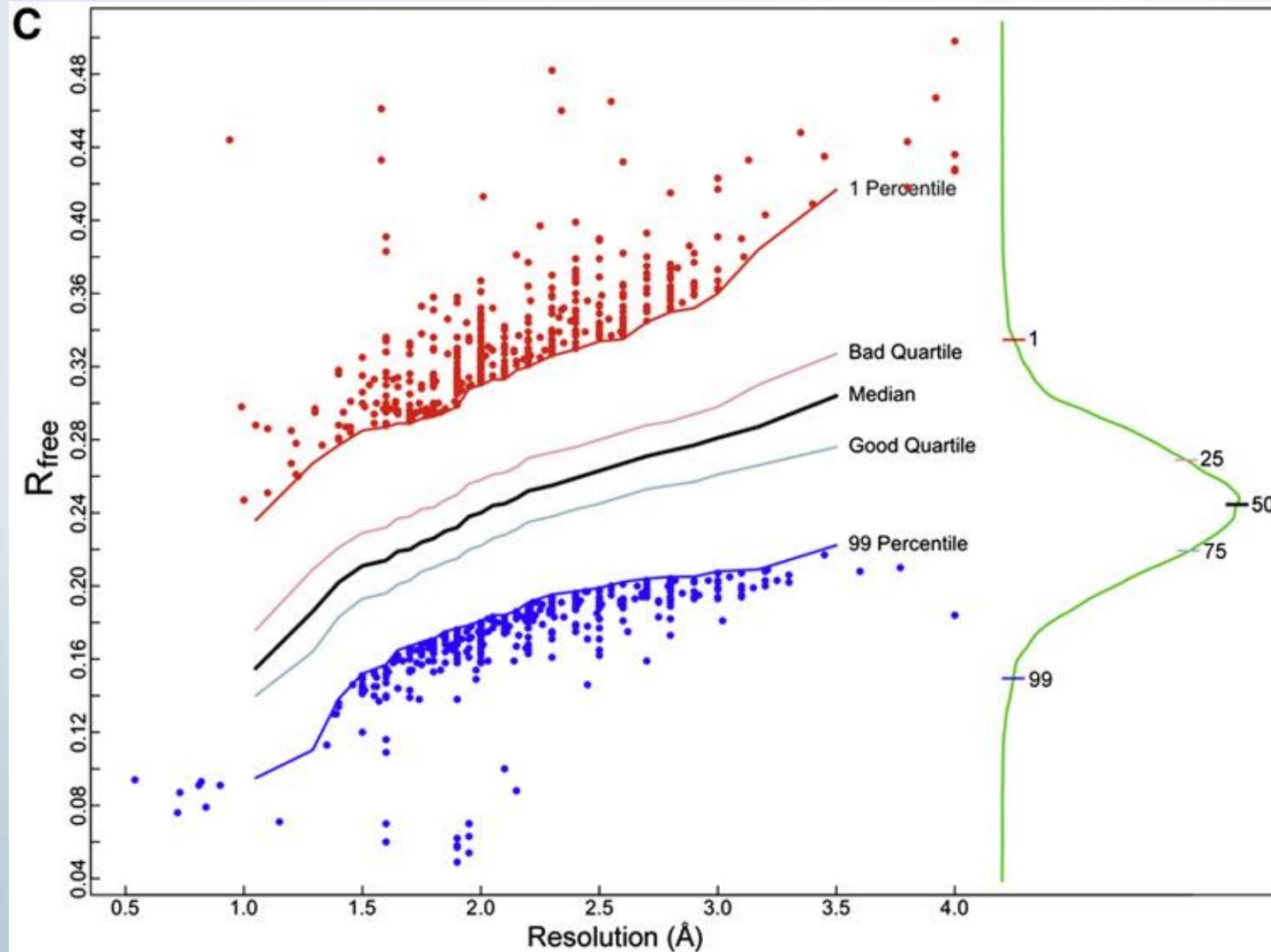
distributions for 6 amino-acid categories, from 825,000 residues after quality-filtering
by resolution ($<2\text{\AA}^\circ$), alternate conformations, and backbone B-factor ($<30\text{\AA}^\circ 2$)

PDB Validation Report: Protein Backbone Conformation



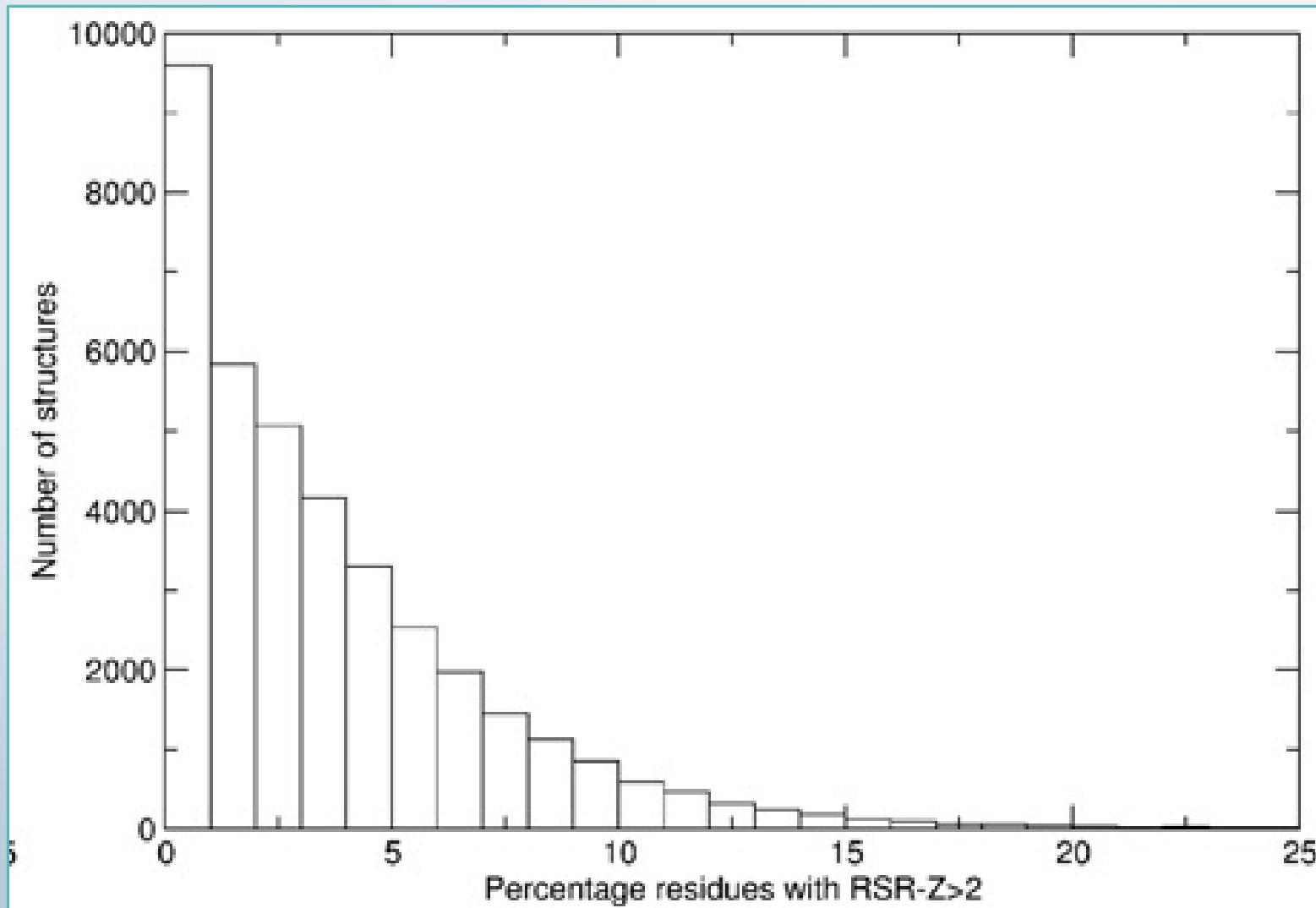
PDB Validation Report: model-to-data agreement

1) global quality of crystal structures: R_{free} 2) local quality of fit to the electron density: RSR-Z



PDB Validation Report: model-to-data agreement

1) global quality of crystal structures: R_{free} 2) local quality of fit to the electron density: RSR-Z



PDB Validation Report: Full Report

RCSB PDB
PROTEIN DATA BANK An Information Portal to
105383 Biological
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands **Go**

Advanced Search | Browse by Annotations

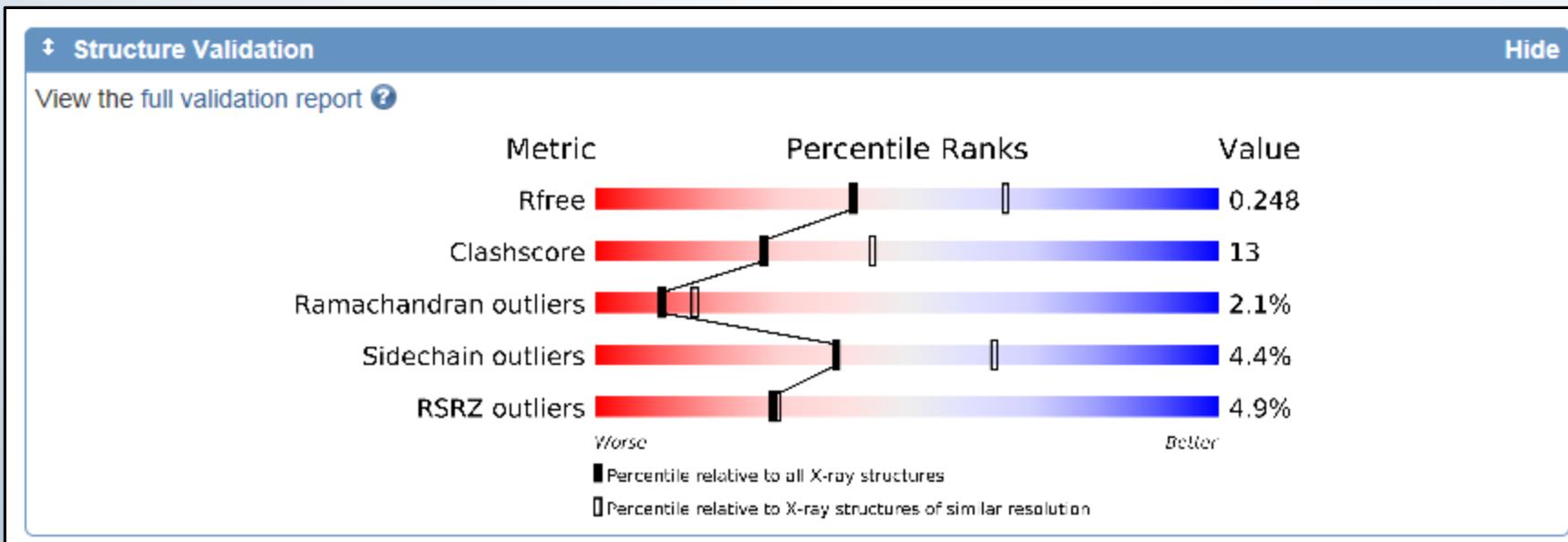
PDB-101 Worldwide PDB EMDataBank

Summary 3D View Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Links

CRYSTAL STRUCTURE OF HUMAN SERUM ALBUMIN 1AO6

DOI:10.2210/pdb1ao6/pdb

Display Files ▾
Download Files ▾
Download Citation ▾



PDB Validation Report: Good, Bad, Ugly

The screenshot shows the homepage of the Worldwide Protein Data Bank (wwPDB). At the top left is the wwpdb logo with the text "WORLDWIDE PROTEIN DATA BANK". To the right is a banner with a stylized green and blue protein structure and the text "Welcome to the Worldwide Protein Data Bank". Below the banner is a navigation bar with links: Home, wwpdb Agreement, Statistics, FAQ, News, About Us, a Facebook icon, and a RSS feed icon.

Access the PDB FTP:

- RCSB PDB | PDBe | PDBj
- Archive Download
- Chemical Component Dictionary
- Biologically Interesting Molecule Reference Dictionary (BIRD)

New Deposition and Annotation System

- Tutorial
- System Information
- FAQ

Validation Reports

- Reports

Server Information

Deposit Data to the PDB:

- RCSB PDB | PDBe
- PDBj | BMRB

18-September-2014

Inclusion of Large Structures in the Main PDB Archive

The wwpdb recently combined entries that represent large structures (such as ribosomes) across multiple PDB files (SPLIT entries) into single files. These combined structures have been issued new PDB IDs and are represented in the archive in both PDBx/mmCIF and PDBML formats.

Useful links and resources

- Validation Report announcement
- Validation Report User Guide
- Sample validation report PDFs for:
 - 1CBS, a 1.8Å structure of a small protein and a ligand, an entry with better overall quality relative to all X-ray structures
 - 1FCC, a 3.2Å structure with worse overall quality relative to all X-ray structures
 - 1EG1, a 3.6Å structure with possible ligand geometry and fit to density issues
- Deposit a crystal structure now at PDBe, PDBj or RCSB PDB

PDB Validation Report: 1CSB vs. 1FCC

PDB ID : 1CBS

Title : CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID

Authors : Kleywelt, G.J.; Bergfors, T.; Jones, T.A.

Deposited on : 1994-09-28

Resolution : 1.80 Å (reported)

PDB ID : 1FCC

Title : CRYSTAL STRUCTURE OF THE C2 FRAGMENT OF STREPTOCOCAL PROTEIN G IN COMPLEX WITH THE FC DOMAIN OF HUMAN IGG

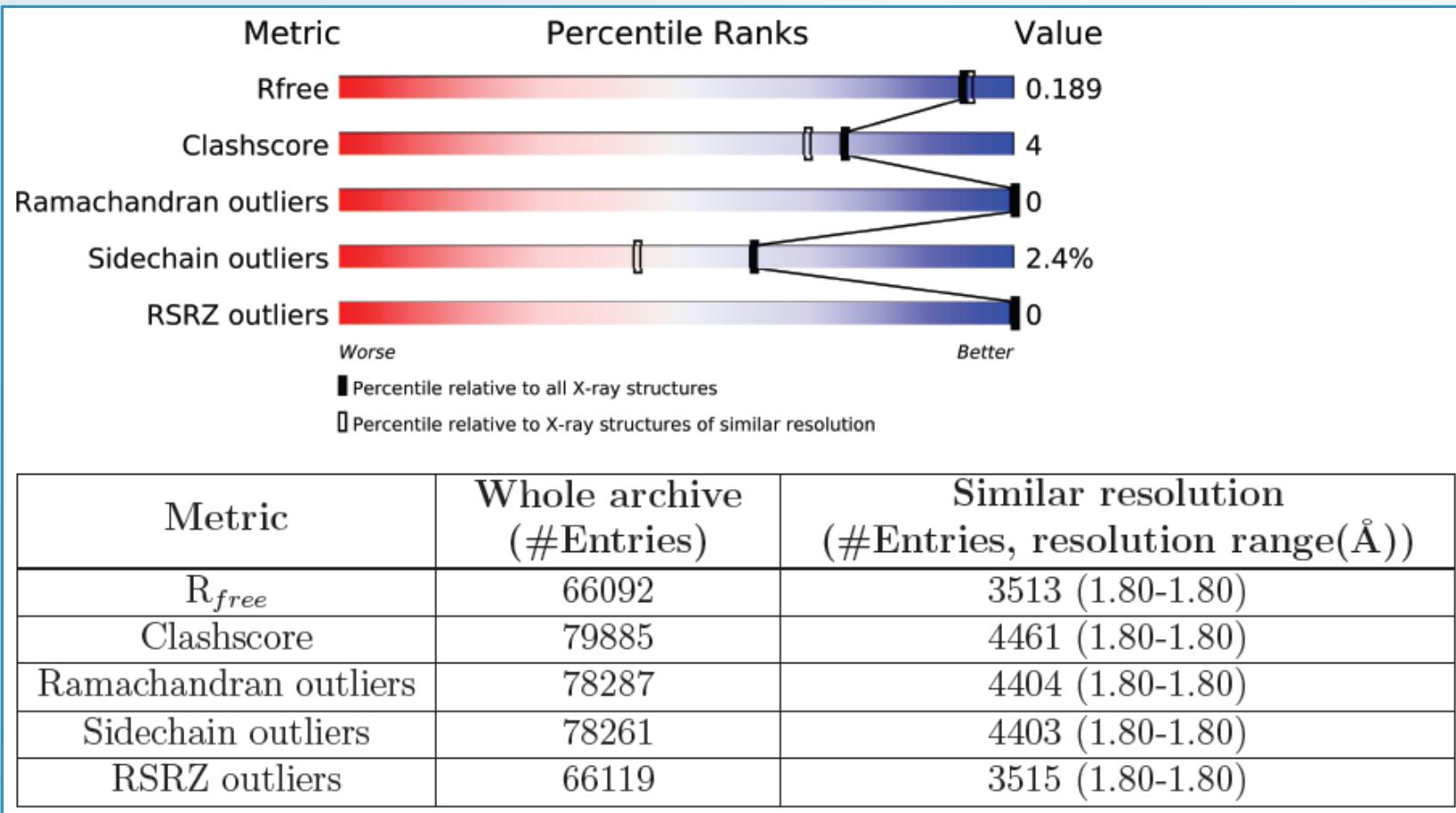
Authors : Sauer-Eriksson, A.E.; Kleywelt, G.J.; Uhlen, M.; Jones, T.A.

Deposited on : 1995-01-17

Resolution : 3.20 Å (reported)

PDB Validation Report: 1CSB vs. 1FCC

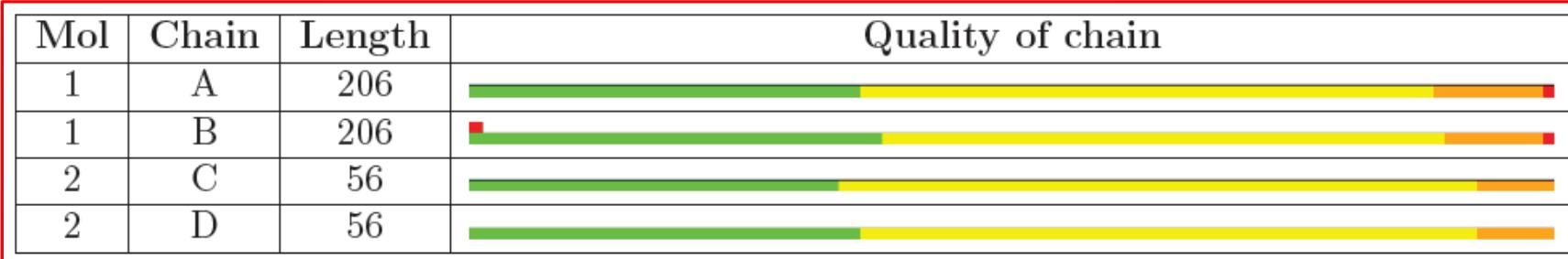
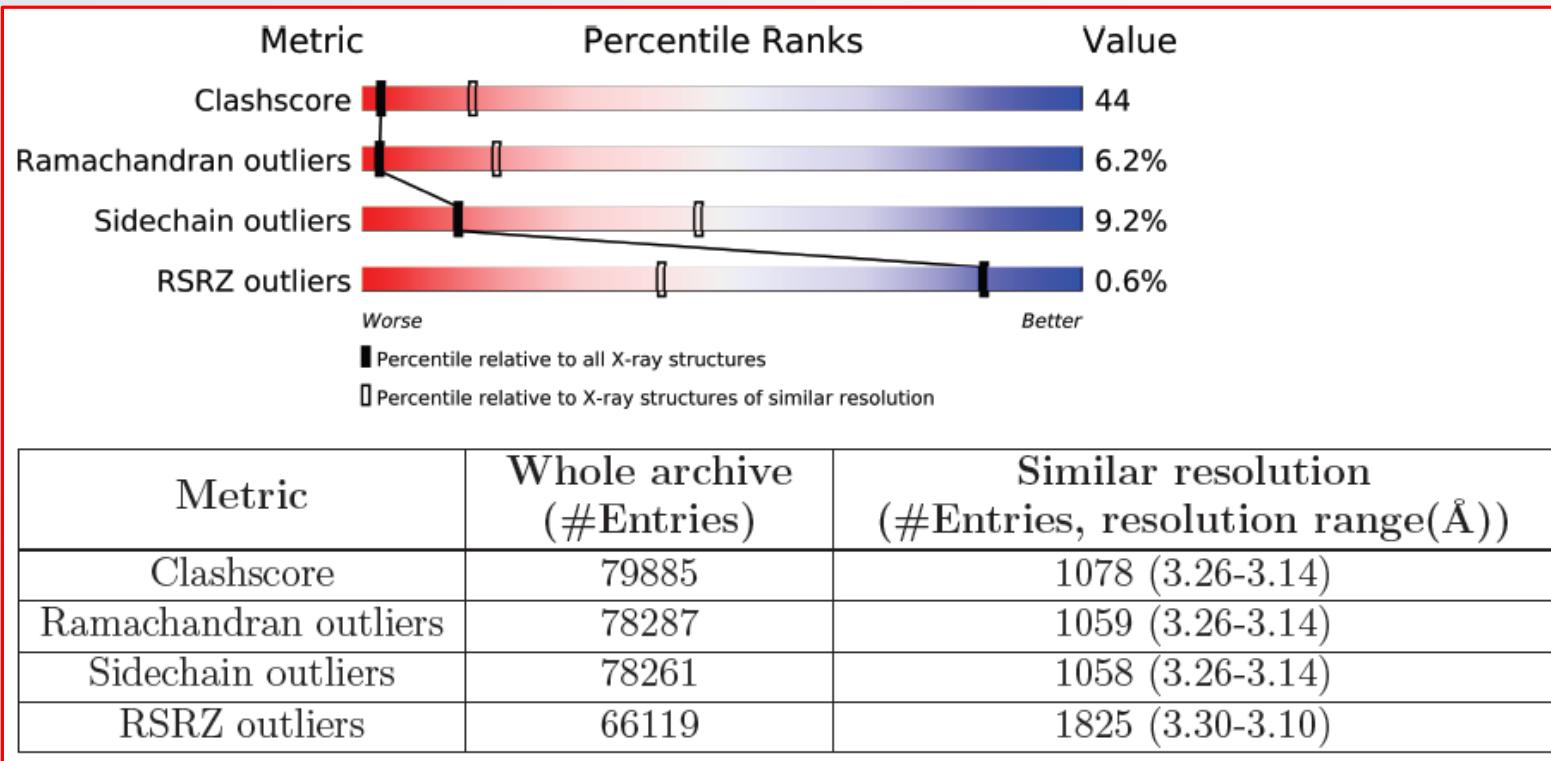
Overall quality at a glance



Mol	Chain	Length	Quality of chain
1	A	137	

PDB Validation Report: 1CSB vs. 1FCC

Overall quality at a glance



PDB Validation Report: 1CSB vs. 1FCC

Entry composition

There are 14 discrepancies between the modeled and reference sequences

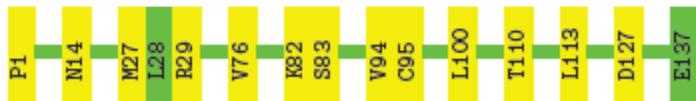
Chain	Residue	Modelled	Actual	Comment	Reference
A	272	GLN	GLU	CONFLICT	UNP P01857
A	283	GLN	GLU	CONFLICT	UNP P01857
A	294	GLN	GLU	CONFLICT	UNP P01857
A	312	ASN	ASP	CONFLICT	UNP P01857
A	315	ASP	ASN	CONFLICT	UNP P01857
A	356	GLU	ASP	CONFLICT	UNP P01857
A	358	MET	LEU	CONFLICT	UNP P01857
B	272	GLN	GLU	CONFLICT	UNP P01857
B	283	GLN	GLU	CONFLICT	UNP P01857
B	294	GLN	GLU	CONFLICT	UNP P01857
B	312	ASN	ASP	CONFLICT	UNP P01857
B	315	ASP	ASN	CONFLICT	UNP P01857
B	356	GLU	ASP	CONFLICT	UNP P01857
B	358	MET	LEU	CONFLICT	UNP P01857

PDB Validation Report: 1CSB vs. 1FCC

Residue-property plots

- Molecule 1: CELLULAR RETINOIC ACID BINDING PROTEIN TYPE II

Chain A:



Chain B:



PDB Validation Report: 1CSB vs. 1FCC

Data and refinement statistics

Property	Value		
Space group	P 21 21 21		
Cell constants	45.65Å	47.56Å	77.61Å
a, b, c, α , β , γ	90.00°	90.00°	90.00°
Resolution (Å)	8.00 – 1.80 14.93 – 1.80		
% Data completeness (in resolution range)	90.3 (8.00-1.80) 90.5 (14.93-1.80)		
R_{merge}	(Not available)		
R_{sym}	(Not available)		
$\langle I/\sigma(I) \rangle^{\textcolor{red}{1}}$	3.77 (at 1.79Å)		
Refinement program	X-PLOR		
R , R_{free}	0.200	,	0.237
	0.184	,	0.189
R_{free} test set	1479 reflections (10.17%)		
Wilson B-factor (Å ²)	14.8		
Anisotropy	0.434		
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 58.9		

PDB Validation Report: 1CSB vs. 1FCC

Data and refinement statistics

Property	Value		
Space group	P 43 21 2		
Cell constants	110.60Å	110.60Å	160.30Å
a, b, c, α , β , γ	90.00°	90.00°	90.00°
Resolution (Å)	8.00	–	3.20
	18.35	–	3.20
% Data completeness (in resolution range)	72.0 (8.00-3.20) 72.7 (18.35-3.20)		
R_{merge}	(Not available)		
R_{sym}	(Not available)		
$\langle I/\sigma(I) \rangle$	-		
Refinement program	X-PLOR		
R , R_{free}	0.289	,	0.357
	0.294	,	(Not available)
R_{free} test set	No test flags present.		
Wilson B-factor (Å ²)	87.0		
Anisotropy	0.065		
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.20 , 21.9		

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Standard geometry

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.47	0/1107	0.71	0/1491

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.80	1/1702 (0.1%)	0.94	1/2316 (0.0%)
1	B	0.81	1/1702 (0.1%)	0.94	1/2316 (0.0%)
2	C	0.73	0/440	0.92	0/597
2	D	0.73	0/440	0.92	0/597
All	All	0.79	2/4284 (0.0%)	0.94	2/5826 (0.0%)

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Standard geometry

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	293	GLU	CG-CD	5.22	1.59	1.51
1	A	293	GLU	CG-CD	5.20	1.59	1.51

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	288	LYS	N-CA-C	-8.45	88.18	111.00
1	A	288	LYS	N-CA-C	-8.44	88.20	111.00

There are no chirality outliers.

There are no planarity outliers.

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Close contacts

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1091	0	1106	7	0
2	A	22	0	27	2	0
3	A	100	0	0	2	0
All	All	1213	0	1133	9	0

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1656	0	1630	161	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1656	0	1630	159	0
2	C	434	0	415	39	1
2	D	434	0	415	40	0
All	All	4180	0	4090	368	2

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Close contacts

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:82:LYS:HG2	1:A:100:LEU:HD21	1.56	0.87

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:A:200:REA:H8	2:A:200:REA:H181	1.81	0.61
1:A:82:LYS:HG2	1:A:100:LEU:CD2	2.32	0.59
2:A:200:REA:C8	2:A:200:REA:H181	2.41	0.51
1:A:1:PRO:HB2	1:A:113:LEU:HD12	1.97	0.47
1:A:76:VAL:HG23	3:A:310:HOH:O	2.16	0.46
1:A:29:ARG:NH2	3:A:387:HOH:O	2.48	0.45
1:A:83:SER:HB3	1:A:95:CYS:SG	2.57	0.44
1:A:94:VAL:HG22	1:A:110:THR:HG22	2.03	0.41

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Close contacts

All (368) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:357:GLU:HG3	1:B:349:TYR:CZ	1.85	1.10
1:A:371:GLY:HA2	1:A:403:SER:OG	1.73	0.89
1:A:357:GLU:HG3	1:B:349:TYR:OH	1.72	0.89
1:B:371:GLY:HA2	1:B:403:SER:OG	1.73	0.87
1:B:375:SER:HB3	1:B:404:PHE:CZ	2.10	0.86
1:A:375:SER:HB3	1:A:404:PHE:CZ	2.10	0.86
1:A:238:PRO:HA	1:A:265:ASP:HB2	1.57	0.86
1:B:238:PRO:HA	1:B:265:ASP:HB2	1.57	0.85
1:B:338:LYS:HD2	1:B:339:ALA:O	1.77	0.85
1:B:346:PRO:HD3	1:B:429:HIS:HD2	1.42	0.84
1:A:357:GLU:CG	1:B:349:TYR:CZ	2.59	0.84
1:A:238:LYS:HD2	1:A:239:ALA:O	1.77	0.84

PDB Validation Report: 1CSB vs. 1FCC

[Model quality](#)

[Torsion angles](#)

[Protein backbone](#)

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	135/137 (98%)	132 (98%)	3 (2%)	0	100 100

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	204/206 (99%)	171 (84%)	19 (9%)	14 (7%)	2 14
1	B	204/206 (99%)	171 (84%)	19 (9%)	14 (7%)	2 14
2	C	54/56 (96%)	45 (83%)	7 (13%)	2 (4%)	5 34
2	D	54/56 (96%)	45 (83%)	7 (13%)	2 (4%)	5 34
All	All	516/524 (98%)	432 (84%)	52 (10%)	32 (6%)	2 19

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Torsion angles

Protein backbone

Mol	Chain	Res	Type
1	A	288	LYS
1	A	291	PRO
1	A	292	ARG
1	A	298	SER
2	C	46	ASP
1	B	288	LYS
1	B	291	PRO
1	B	292	ARG
1	B	298	SER
2	D	46	ASP
1	A	285	HIS
1	A	339	ALA
1	A	385	GLY
1	A	390	ASN
1	A	441	LEU
2	C	21	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	285	HIS
1	B	339	ALA
1	B	385	GLY
1	B	390	ASN
1	B	441	LEU
2	D	21	VAL
1	A	289	THR
1	B	289	THR
1	A	297	ASN
1	A	438	GLN
1	B	297	ASN
1	B	438	GLN
1	A	287	ALA
1	B	287	ALA
1	A	247	PRO
1	B	247	PRO

PDB Validation Report: 1CSB vs. 1FCC

Model quality

Torsion angles

Protein sidechains

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	123/123 (100%)	120 (98%)	3 (2%)	61 44

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	193/193 (100%)	175 (91%)	18 (9%)	13 46
1	B	193/193 (100%)	175 (91%)	18 (9%)	13 46
2	C	46/46 (100%)	42 (91%)	4 (9%)	15 51
2	D	46/46 (100%)	42 (91%)	4 (9%)	15 51
All	All	478/478 (100%)	434 (91%)	44 (9%)	13 47

PDB Validation Report: 1CSB vs. 1FCC

Fit of model and data

Mol	Chain	Analysed	$\langle \text{RSRZ} \rangle$	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	137/137 (100%)	-0.33	0 100 100	6, 13, 27, 43	0

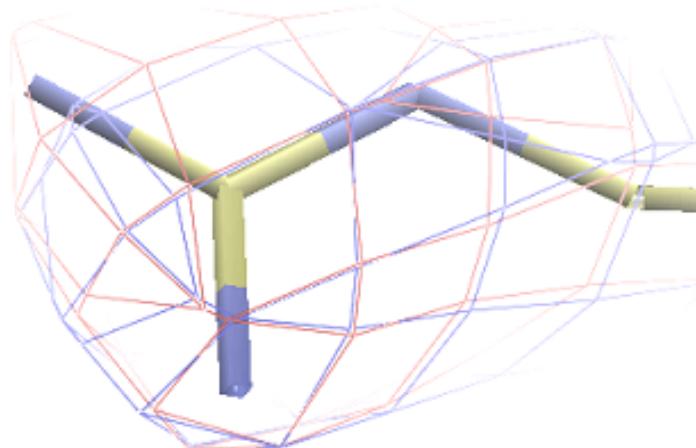
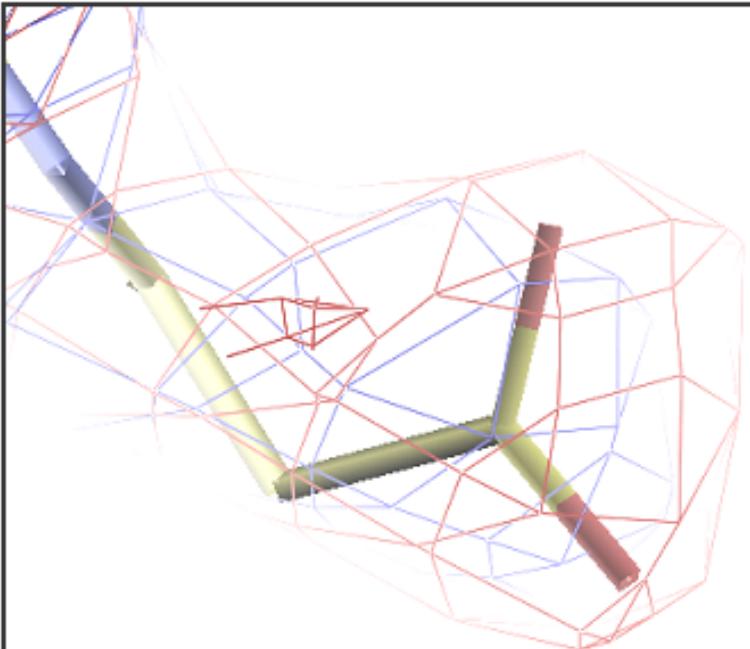
Mol	Chain	Analysed	$\langle \text{RSRZ} \rangle$	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	206/206 (100%)	-0.16	0 100 100	2, 37, 83, 125	0
1	B	206/206 (100%)	0.08	3 (1%) 70 21	2, 37, 83, 125	0
2	C	56/56 (100%)	-0.05	0 100 100	2, 39, 73, 86	0
2	D	56/56 (100%)	-0.07	0 100 100	2, 39, 73, 86	0
All	All	524/524 (100%)	-0.05	3 (0%) 86 41	2, 38, 83, 125	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	298	SER	2.7
1	B	238	PRO	2.2
1	B	239	SER	2.0

PDB_REDO

A databank of updated and optimised X-ray structure models



PDB_REDO

Enter a PDB code (4 characters):

Submit

PDB_RED0: 1CSB vs. 1FCC

Warnings!

- Used extra refinement runs to compensate for possible R-free bias

Structure

Spacegroup

P 21 21 21

Cell dimensions

a: 45.650 Å

b: 47.560 Å

c: 77.610 Å

α: 90.00°

β: 90.00°

γ: 90.00°

Resolution

1.80 Å

Experimental data

Reflections

All: 14678

Test set: 1496 (10.2%)

Resolution range

14.93 Å

1.80 Å

Warnings!

- A new R-free set (0.0813 fraction of all reflections) was created for this optimisation

Structure

Spacegroup

P 43 21 2

Cell dimensions

a: 110.600 Å

b: 110.600 Å

c: 160.300 Å

α: 90.00°

β: 90.00°

γ: 90.00°

Resolution

3.20 Å

Experimental data

Reflections

All: 12297

Test set: 0 (0.0%)

Resolution range

18.35 Å

3.20 Å

PDB_RED0: 1CSB vs. 1FCC

R-values etc.

	From PDB header	Calculated from data	After conservative optimisation	After full optimisation
R	0.2000	0.1785	0.1626	0.1610
R-free	0.2370	0.1833	0.2010	0.1962
σ R-free		0.0047	0.0052	0.0051
Z(R-free)		6.40	-1.27	-0.73

R-values etc.

	From PDB header	Calculated from data	After conservative optimisation	After full optimisation
R	0.2890	0.3155	0.2354	0.2218
R-free	0.3570	0.3028	0.2695	0.2679
σ R-free		0.0067	0.0059	0.0059
R-free Z-score		11.31	2.20	-0.29

PDB_REDO: 1CSB vs. 1FCC

WHAT_CHECK validation

	Original PDB entry	Conservatively optimised	Fully optimised
1st generation packing quality ¹	1.331	1.339	1.319
2nd generation packing quality ¹	0.076	0.248	0.471
Ramachandran plot appearance ¹	1.724	2.025	2.217
Chi-1/Chi-2 rotamer normality ¹	-1.120	0.448	0.736
Backbone conformation ¹	1.207	1.335	1.324
Bond length RMS Z-score ²	0.433	0.467	0.449
Bond angle RMS Z-score ²	0.676	0.686	0.664
Total number of bumps ³	3	3	0
Unsatisfied H-bond donors/acceptors ³	3	3	2

WHAT_CHECK validation

	Original PDB entry	Conservatively optimised	Fully optimised
1st generation packing quality ¹	-1.344	-0.348	-0.613
2nd generation packing quality ¹	-1.605	-0.353	-0.764
Ramachandran plot appearance ¹	-4.557	-2.886	-2.763
Chi-1/Chi-2 rotamer normality ¹	-5.175	-3.195	-4.027
Backbone conformation ¹	-0.393	0.143	0.068
Bond length RMS Z-score ²	0.712	0.336	0.510
Bond angle RMS Z-score ²	0.959	0.548	0.752
Total number of bumps ³	193	48	97
Unsatisfied H-bond donors/acceptors ³	50	46	40

PDB_RED0: 1CSB vs. 1FCC

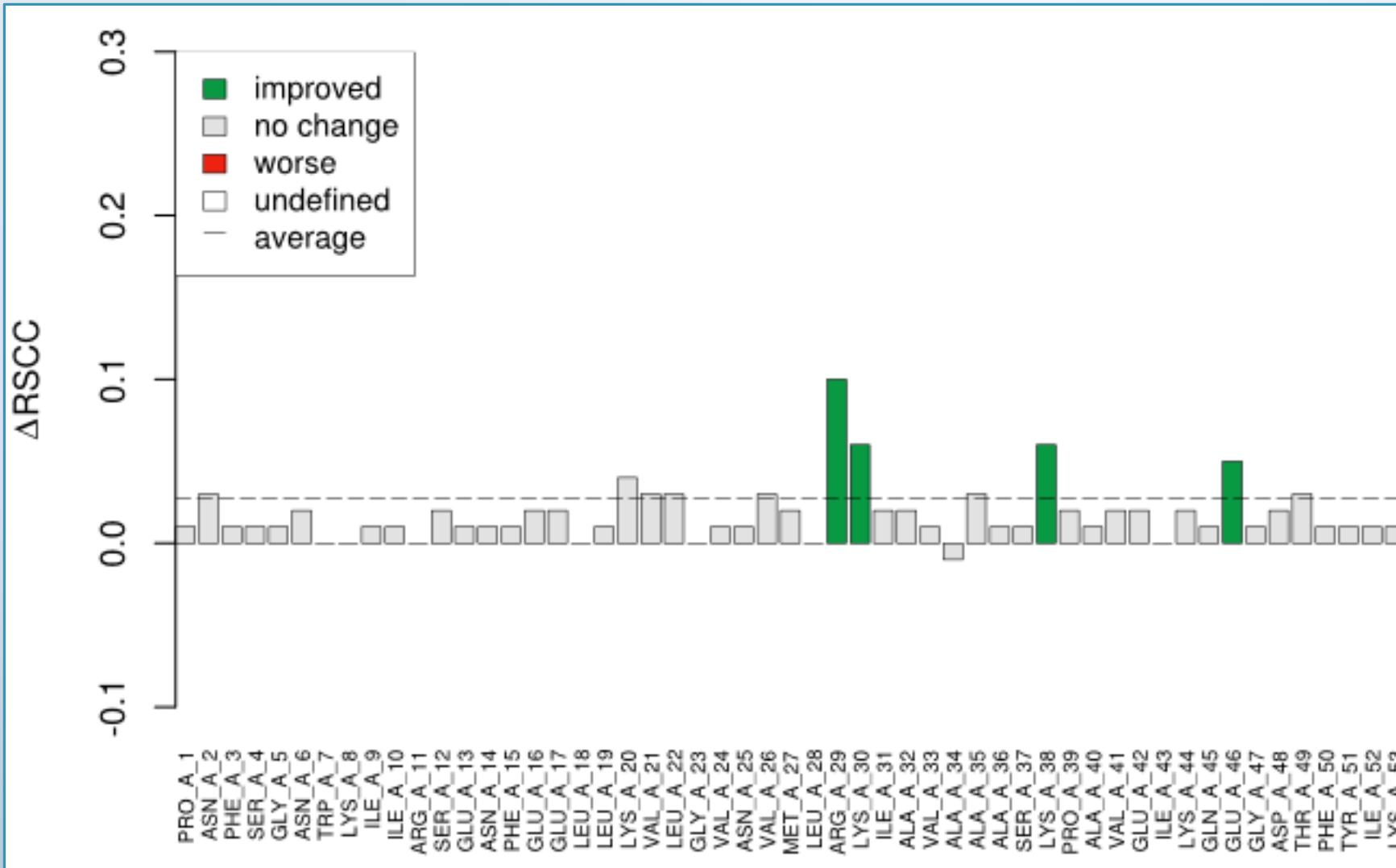
Density map fit (real-space correlation coefficient)

Significantly improved residues	11
Significantly deteriorated residues	0
Changes of all residues	Plot

Electron density map validation

	Real-space correlation coefficient	Real-space R-factor
Significantly improved residues	262	100
Significantly deteriorated residues	0	0
Changes of all residues	Plot	Plot

PDB_REDO: 1CSB vs. 1FCC



PDB_REDO: 1CSB vs. 1FCC

