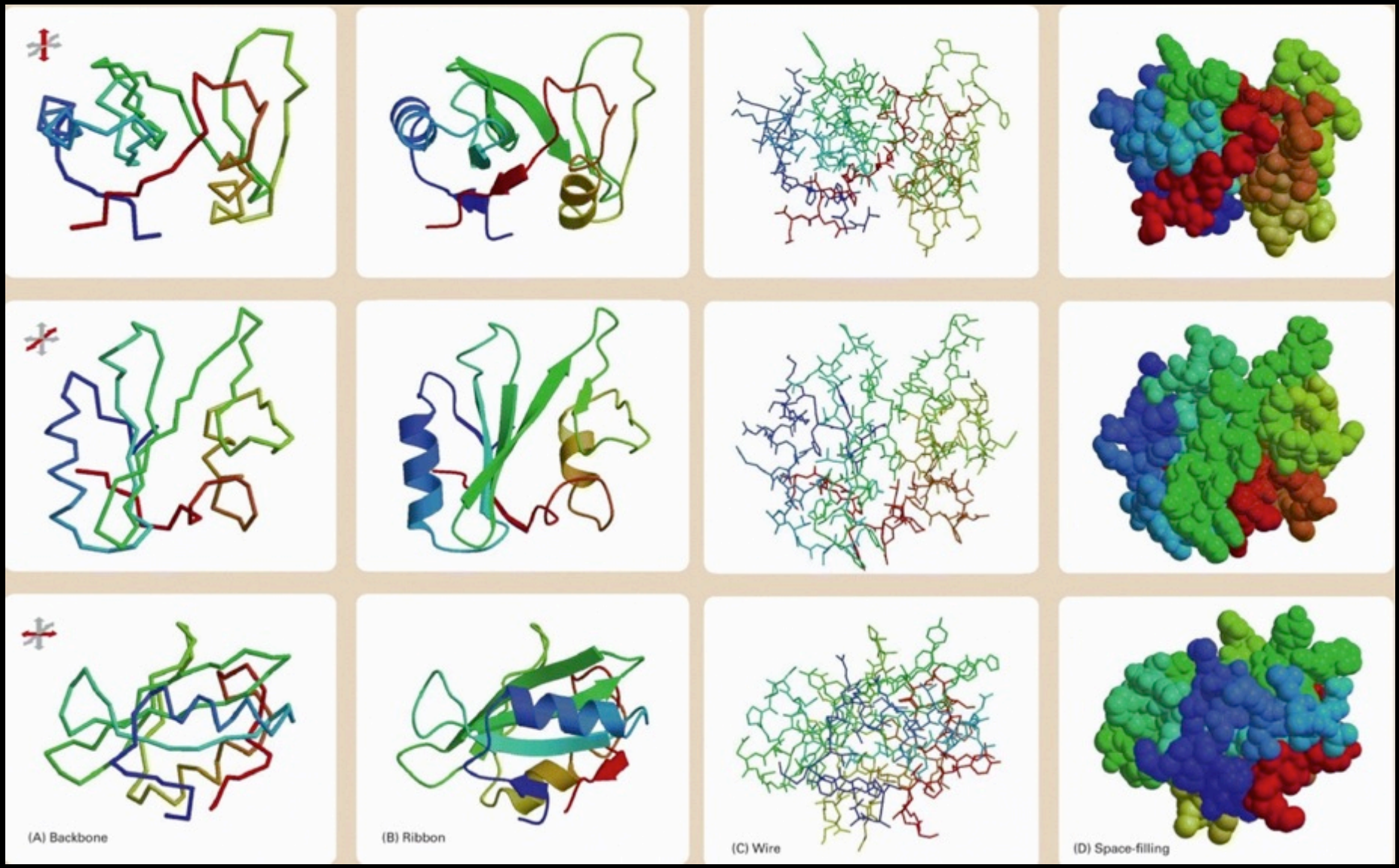
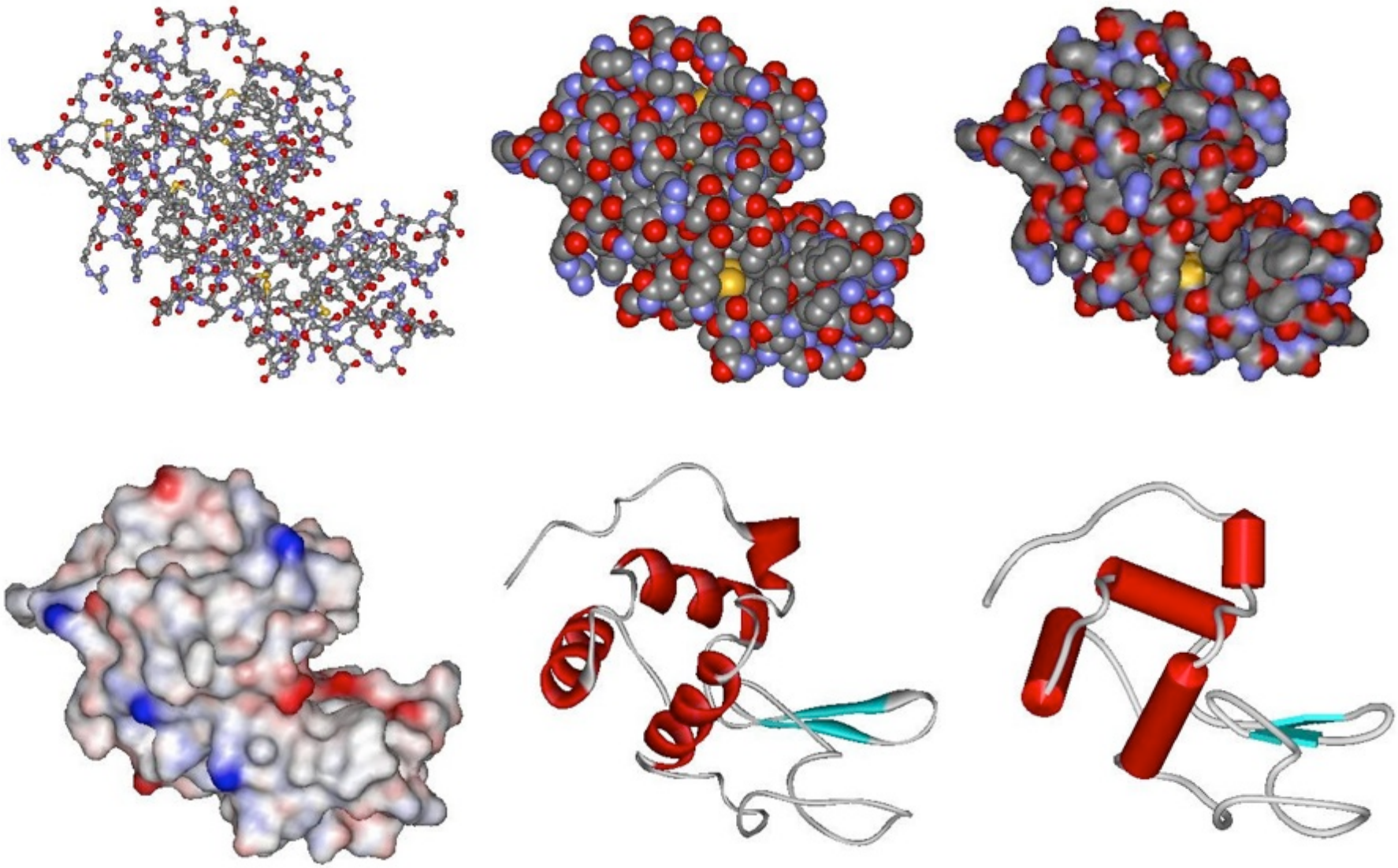
Different ways of illustrating biomacromolecular 3-D structure



The 100 aa SH2 (Src Homology 2) domain

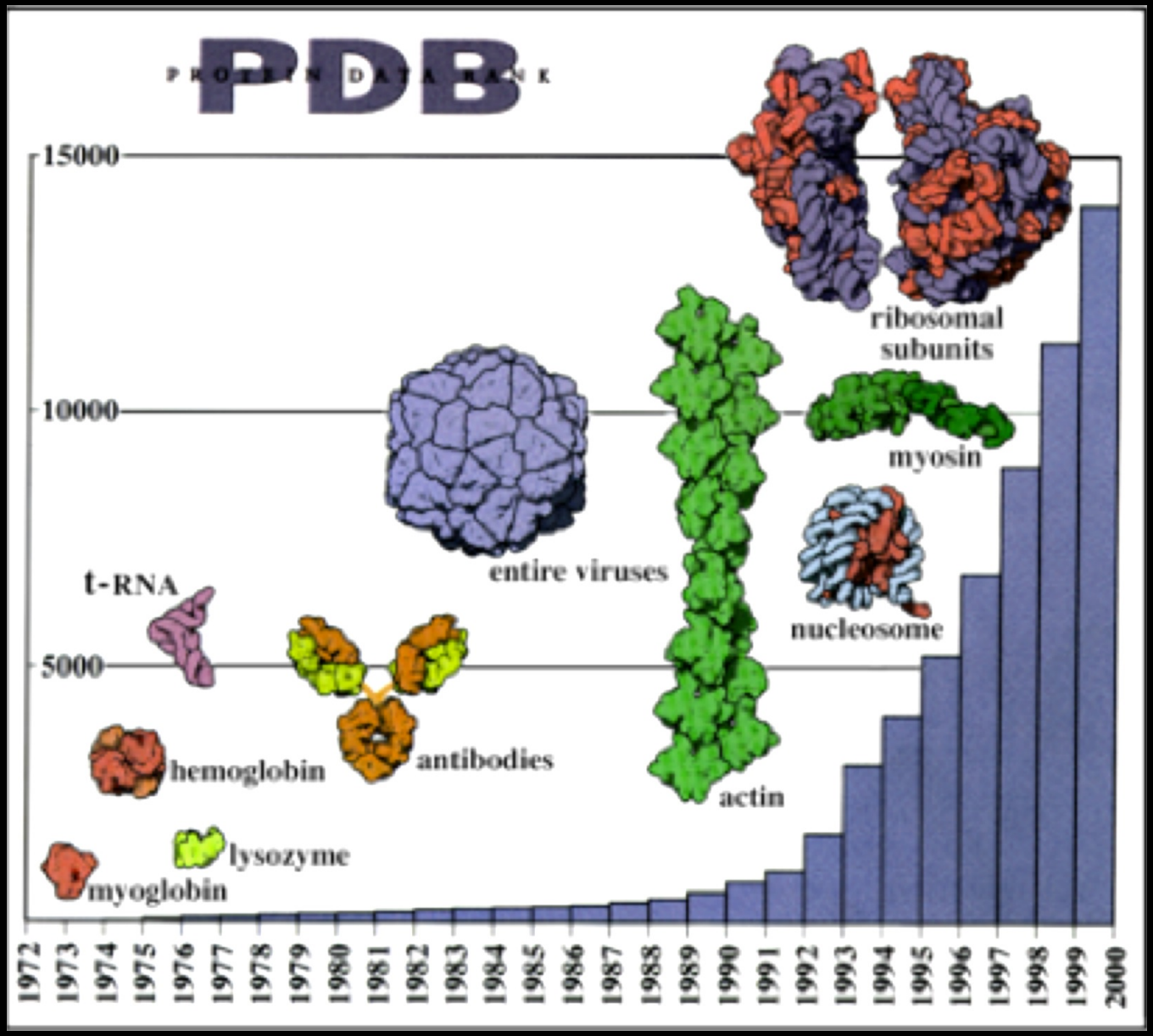
Alberts *et al.*, (2002). *Molecular Biology of the Cell*, 4th ed., pp. 138-139. Garland Science.

HEW lysozyme (129 aa, M_r 14.9 kDa)



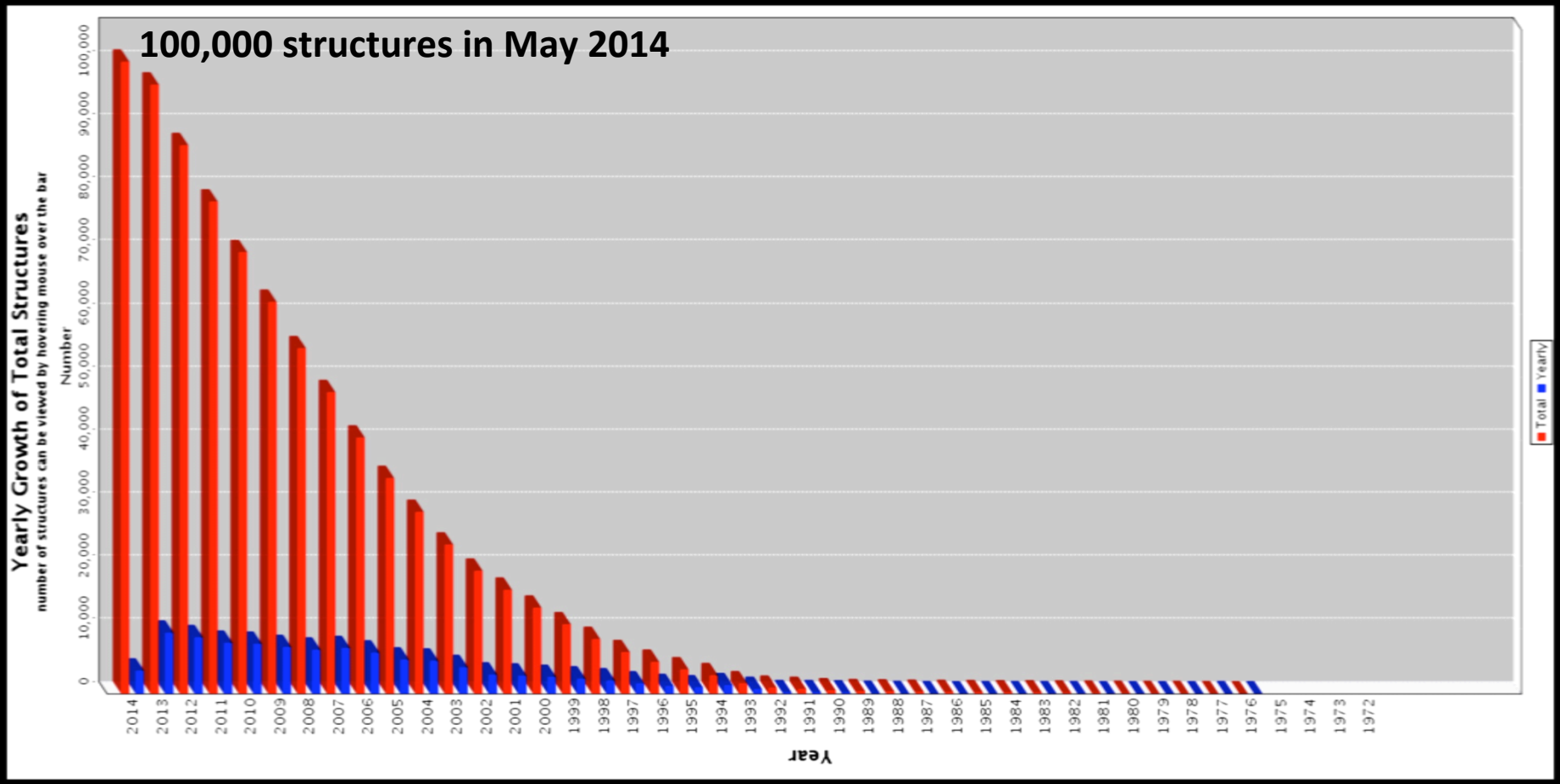
http://lysozyme.co.uk/lysozyme-structure.php

THE PROTEIN DATA BANK
at the turn of the Millennium



http://upload.wikimedia.org/wikipedia/commons/9/99/TriosePhosphateIsomerase_Ribbon_pastel_photo_mat.png

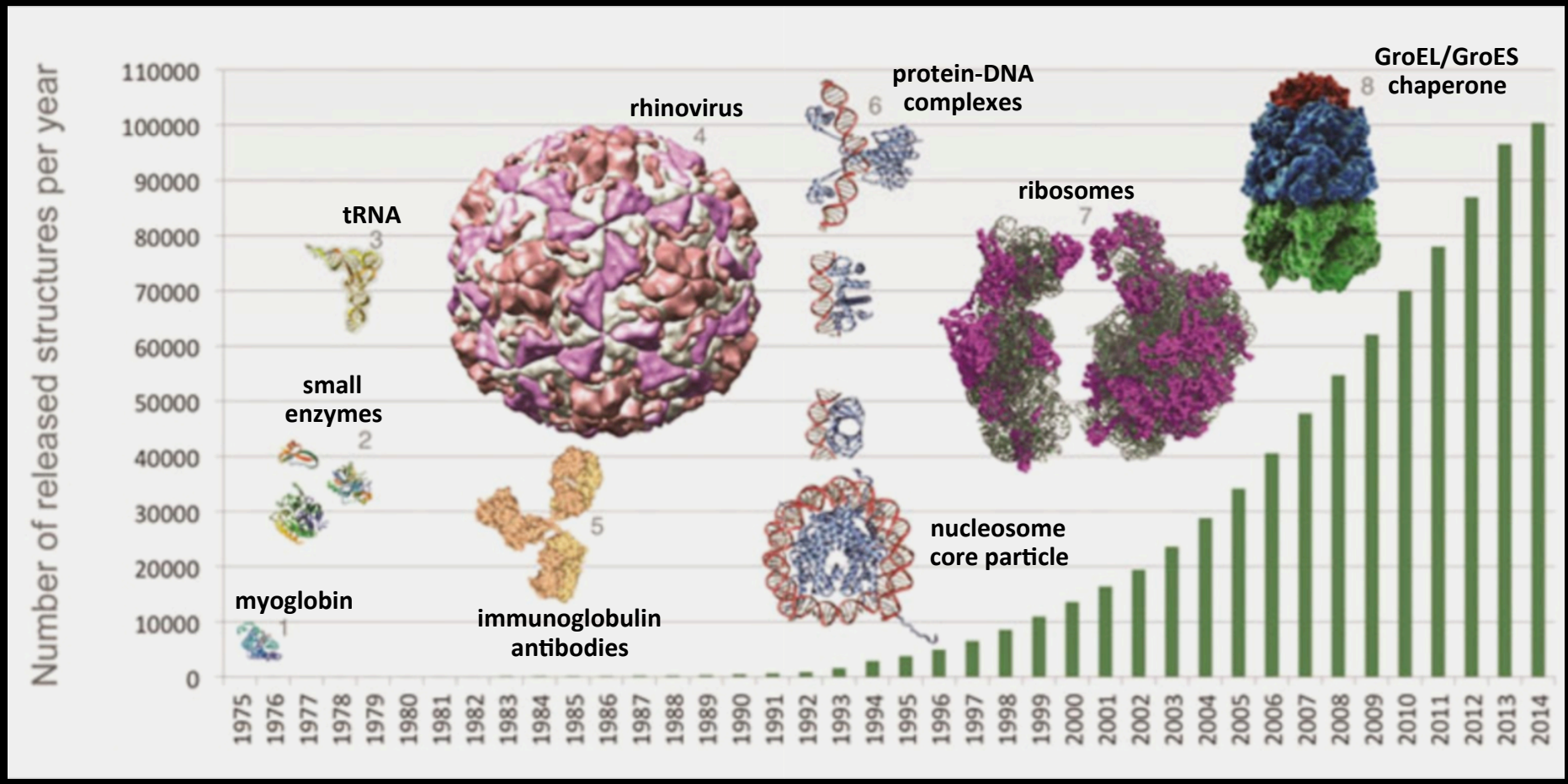
PDB growth curve through mid 2014



Doubling time ~ 4 years

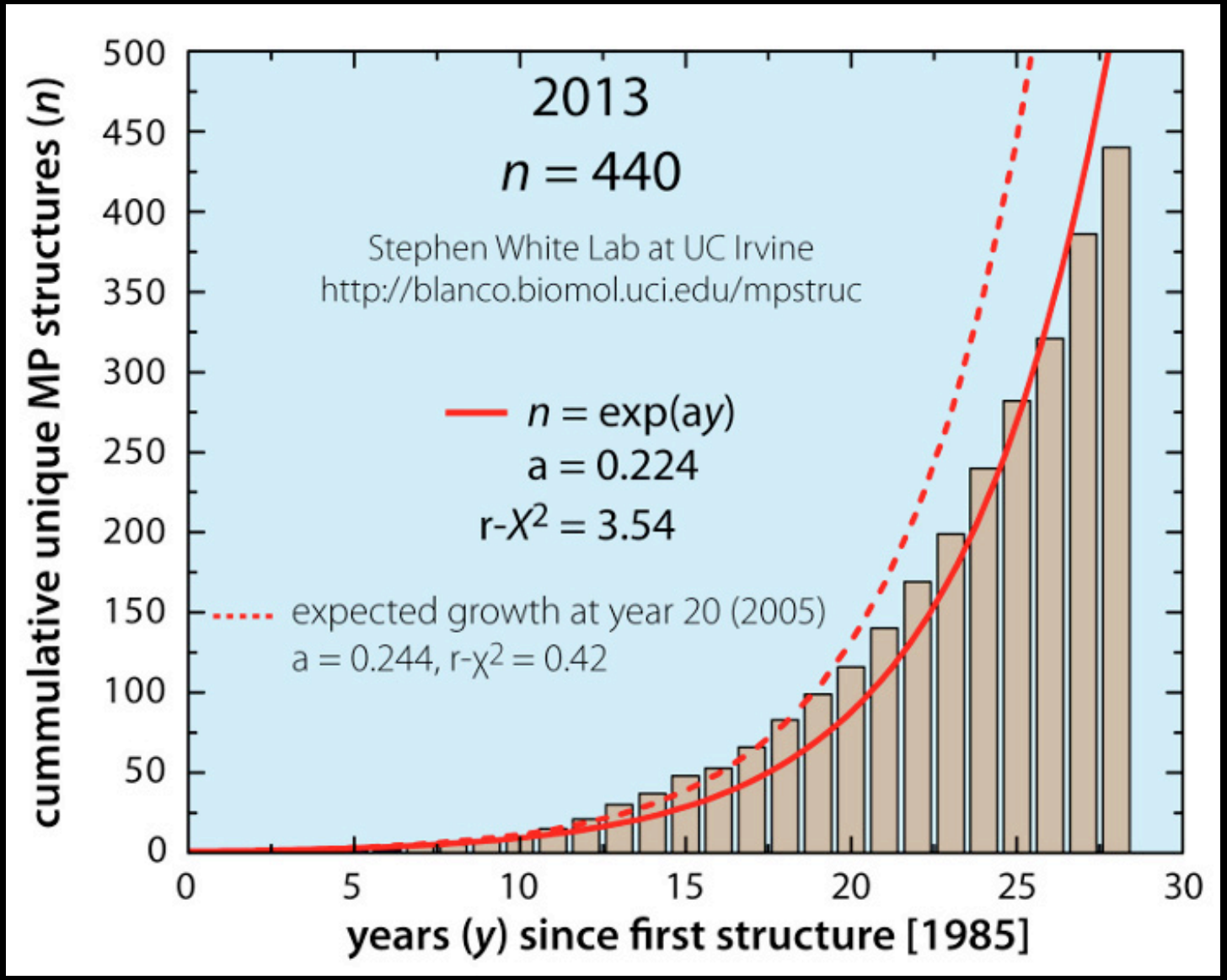
http://www.pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html

PDB grows to 100,000 structures in mid 2014



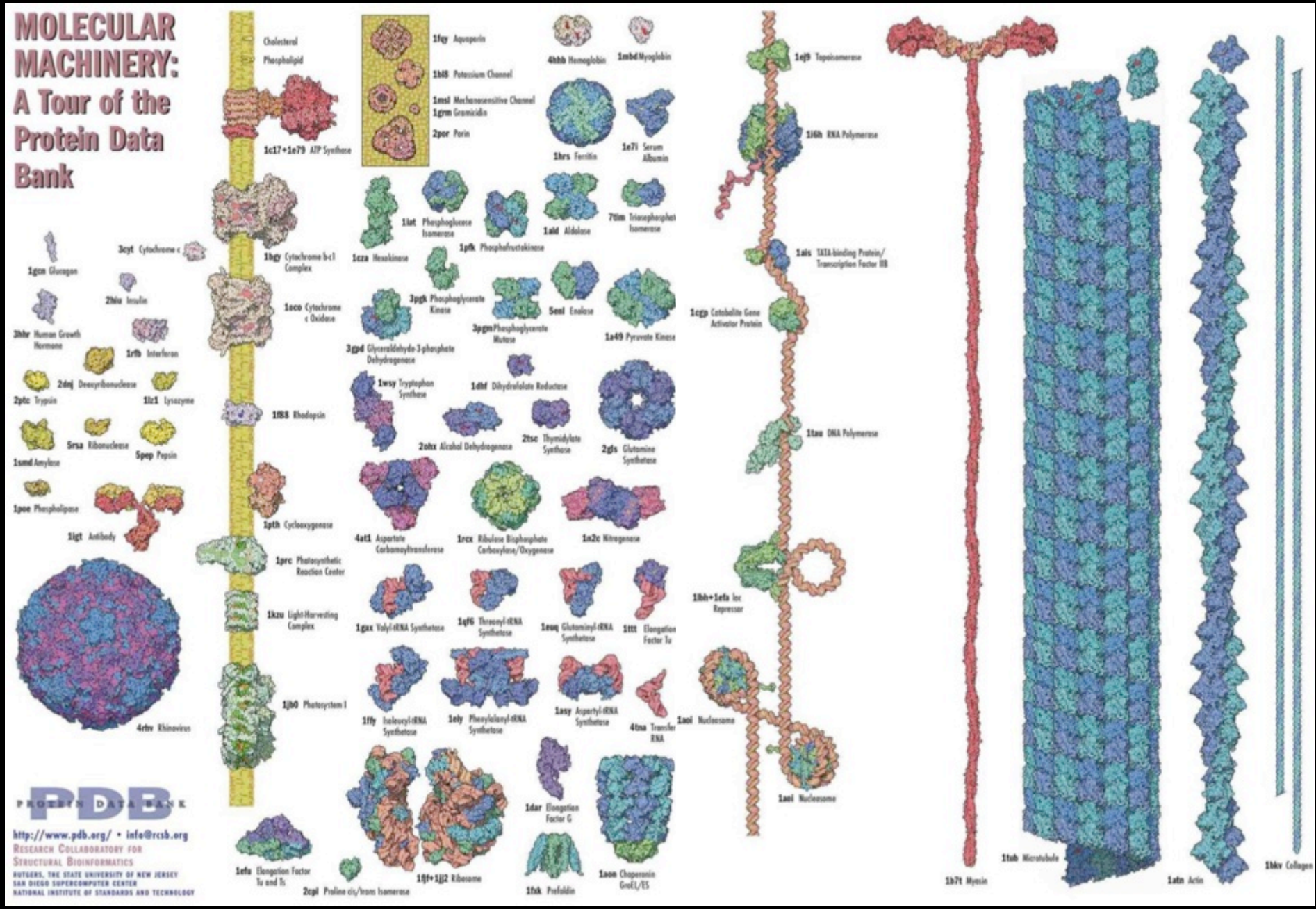
http://loop.nigms.nih.gov/wp-content/uploads/2014/05/PDBstructures_bg.jpg
http://www.eurekalert.org/multimedia/pub/73206.php?from=267554

Growth Curve for Membrane Protein Crystal Structures



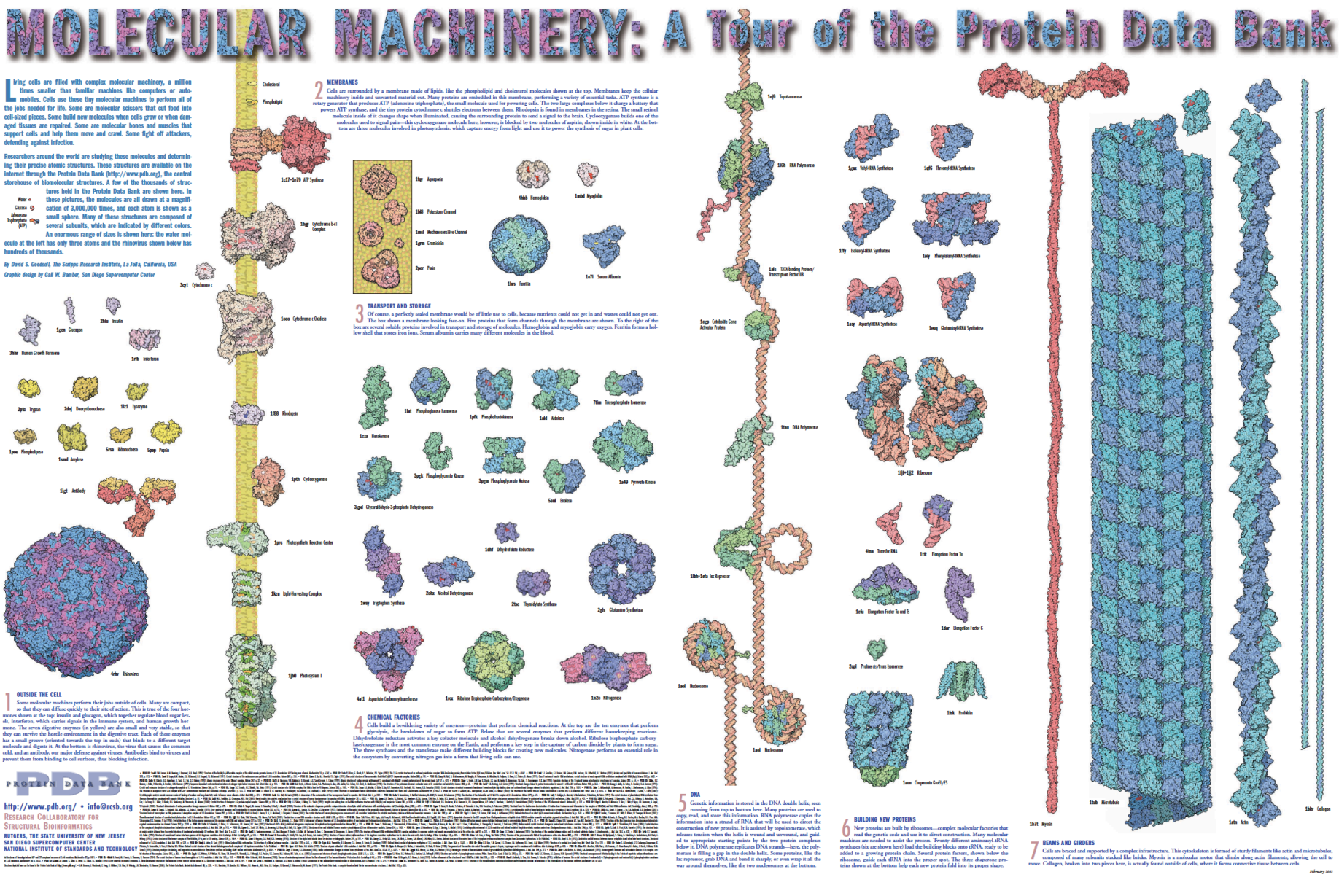
Membrane proteins represent ~30% of known genomes but only ~2% of known structures.
Some 60% of pharmaceuticals target membrane proteins

Architectural elements of structural molecular biology



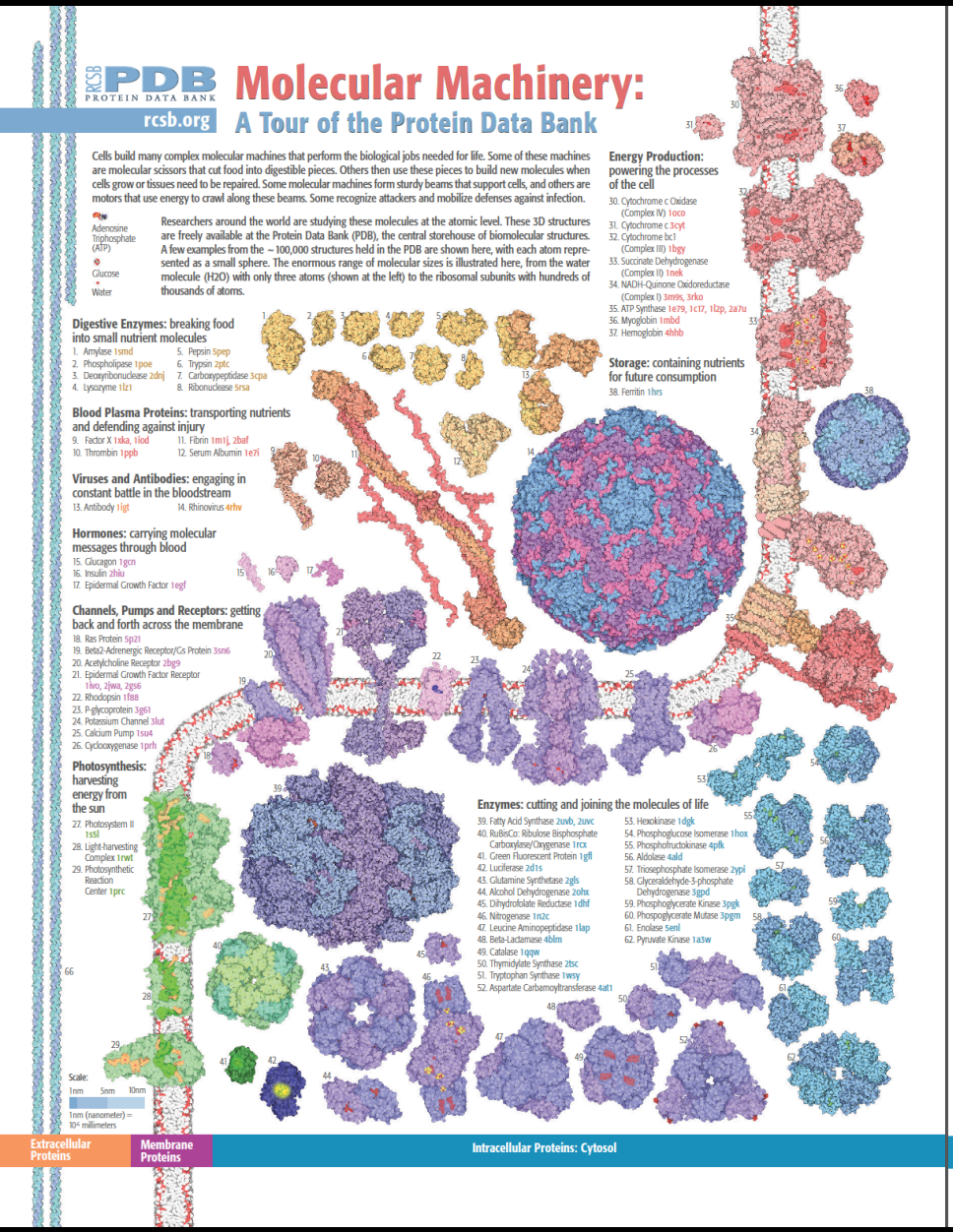
http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/educational_resources/index.html

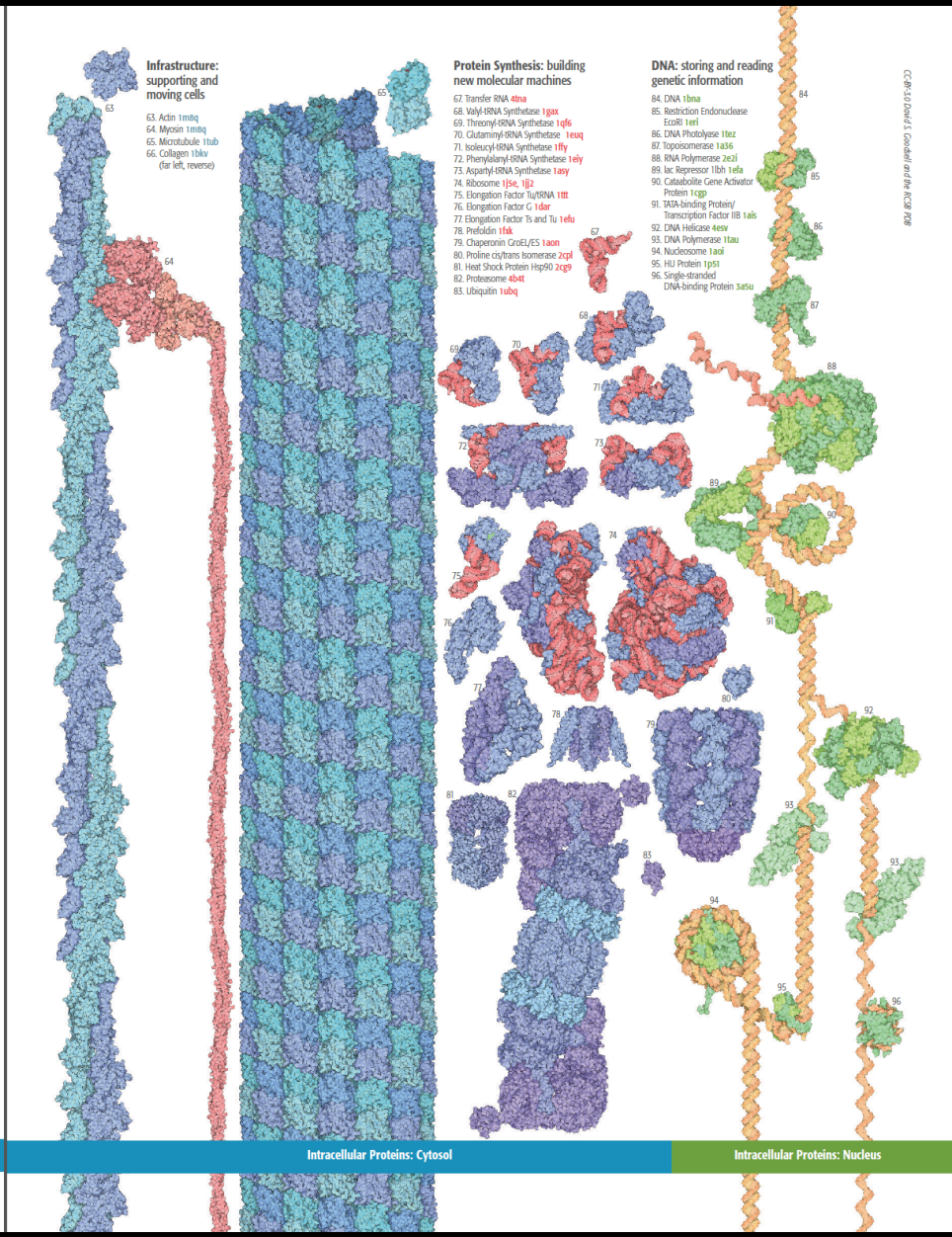
Architectural elements of structural molecular biology



http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/educational_resources/index.html

Architectural elements of structural molecular biology





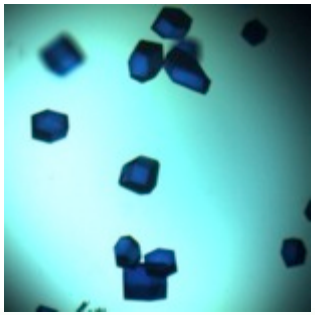
http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/educational_resources/index.html

The protein crystallography challenges

There are four key challenges in crystallography:

1 No material – no crystal - no crystallography

Not all materials are sufficiently soluble and not all lend themselves to self-assembly into periodic structures – that is, protein crystals.

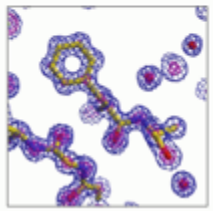


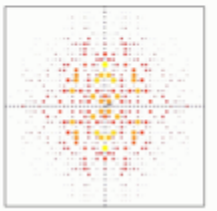
2 The Phase Problem

The major component for the Fourier transformation of the data into electron density is missing - needs strategy.

$$F_{hkl} = \sum$$

$$\rho_{(x,y,z)} = \frac{1}{V}\sum_{-h}^{h}\sum_{-k}^{k}\sum_{-l}^{l}|F_{hkl}|\cdot\exp[-2\pi i(hx+ky+lz-\alpha_{hkl})]$$





The Crystallographic Phase Problem

3 Nonlinearity in Refinement

Each and every atom in the whole structure contributes to each and every data point (measured reflections or structure factor amplitudes)

$$F_{(hkl)} = \sum_{n=1}^{atoms}\rho_{(xyz)}e^{2\pi i(hx+ky+lz)}$$

$$\rho_{(xyz)} = V^{-1}\sum_{-hkl}^{+hkl}F_{(hkl)}e^{-2\pi i(hx+ky+lz)}$$

4 Low determinacy of Refinement

The data to parameter ratio is often poor – stereochemical restraints required that bias the model towards prior expectations.

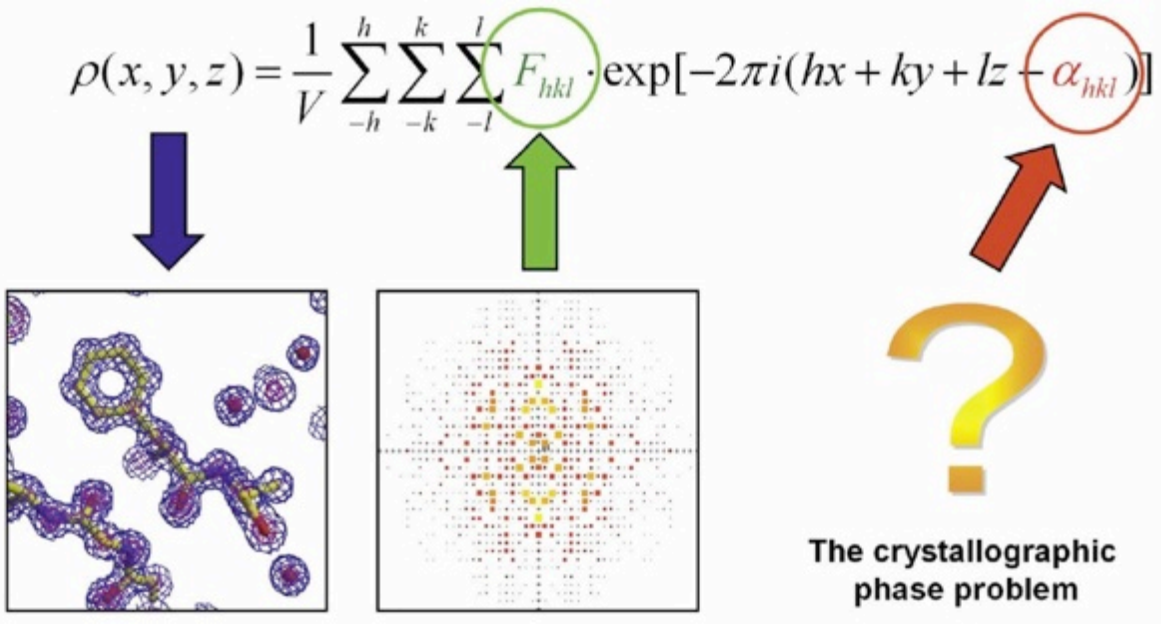
$$F_{hkl} = |F_{hkl}|e^{i\alpha_{hkl}} \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} \rho(x,y,z)$$

The crystallographic phase problem

Diffraction spot intensity alone does not suffice to reconstruct the electron density of the molecule

Figure 1-7 The crystallographic phase problem. In order to reconstruct the electron density of the molecule, two quantities need to be provided for each reflection (data point): the structure factor amplitude, F_{hkl}, which is directly obtained through the experiment and is proportional to the square root of the measured intensity of the diffraction spot or reflection; and the phase angle of each reflection, α_{hkl}, which is not directly observable and must be supplied by additional phasing experiments. The methods and mathematics of electron density reconstruction by Fourier methods are extensively treated in Chapter 9.

Biomolecular Crystallography B.Rupp (2010)



The exponential and dominating term in the Fourier reconstruction of the electron density – the phases of each reflection – must be provided by additional phasing experiments

The "fundamental theorem" of structural crystallography

The fundamental theorem of arithmetic

Every nonzero integer has a unique expression as a product of primes.

The fundamental theorem of algebra

Every univariate polynomial of degree *n* has exactly *n* zeros.

The fundamental theorem of the calculus

If the derivative of $f(x)$ is $g(x)$, then the integral of $g(x)$ is $f(x)$.

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = g(x) \quad \Rightarrow \quad \int_a^b g(x)\,\mathrm{d}x = f(x)\Big|_a^b = f(b) - f(a) \quad \Rightarrow \quad \int g(x)\,\mathrm{d}x = f(x) + C \,.$$

The "fundamental theorem" of structural crystallography

The crystal structure factors F_{hkl} in diffraction or reciprocal *hkl* space and the unit-cell scattering density distribution $\rho(x,y,z)$ in crystal or direct *xyz* space are related by Fourier transformation,

$$F_{hkl} = \left|F_{hkl}\right| \mathrm{e}^{i\varphi_{hkl}} \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} \rho(x,y,z) \qquad \begin{cases} \mathcal{F}\left[F_{hkl}\right] = \rho(x,y,z) & \text{Fourier synthesis} \\ \mathcal{F}^{-1}\left[\rho(x,y,z)\right] = F_{hkl} & \text{Fourier analysis} \end{cases}$$

where the $\left|F_{hkl}\right|$ and φ_{hkl} are, respectively, the amplitudes and phases of the Laue-Bragg scattered beams of radiation diffracted by a crystal.

The "fundamental theorem" of structural crystallography

$$F_{hkl} = |F_{hkl}| e^{i\alpha_{hkl}} \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftarrows}} \rho(x,y,z)$$

$$\begin{cases} \mathcal{F}\left[F_{hkl}\right] = \rho(x,y,z) & \text{Fourier synthesis} \\ \mathcal{F}^{-1}\left[\rho(x,y,z)\right] = F_{hkl} & \text{Fourier analysis} \end{cases}$$

$$\begin{cases} \rho(x,y,z) = \dfrac{1}{V_{\text{cell}}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}| \exp\left[i\alpha_{hkl} - 2\pi i(hx+ky+lz)\right] \\ F_{hkl} = V_{\text{cell}} \int_0^1 \int_0^1 \int_0^1 \rho(x,y,z) \exp\left[+2\pi i(hx+ky+lz)\right] \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \end{cases}$$

$$\begin{cases} F_{hkl} = \sum_{a=1}^{N} f_a(S_{hkl}) \exp\left[2\pi i(hx_a+ky_a+lz_a)\right] = |F_{hkl}| e^{i\alpha_{hkl}} \\ \rho(x,y,z) = \dfrac{1}{V_{\text{cell}}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}| \cos\left[\alpha_{hkl} - 2\pi(hx+ky+lz)\right] \end{cases}$$

$$S_{hkl} = \frac{1}{d_{hkl}} = 2\left(\frac{\sin\theta_{hkl}}{\lambda}\right) \qquad \text{and} \qquad \begin{cases} |F_{\bar{h}\bar{k}\bar{l}}| = |F_{hkl}| \\ \alpha_{\bar{h}\bar{k}\bar{l}} = -\alpha_{hkl} \end{cases}$$

Structure determination involves a large array of different techniques

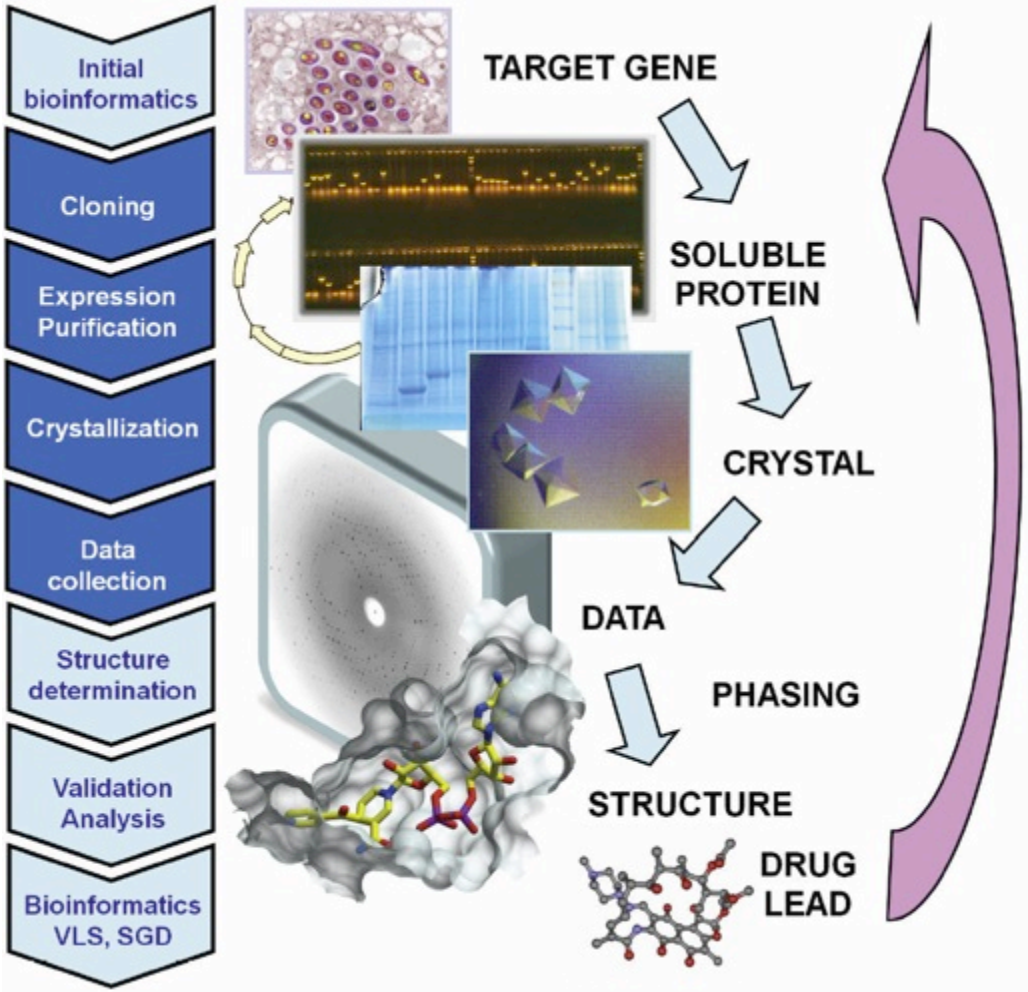


Figure 1-4 Overview of protein structure determination. The bar on the left side of the figure lists major stages of a crystal structure determinaton project. The dark blue shading indicates experimental procedures while the light shading indicates work performed in-silico on computers. The results of the structure analysis frequently feed back into the design of a refined study, particularly in structure guided drug discovery. VLS: virtual ligand screening; SGD: structure guided drug discovery. Consult Figure 1-8 for a more detailed diagram of key steps in structure determination and the corresponding Chapters in this book.

Biomolecular Crystallography B.Rupp (2010)

Major steps in course to be covered

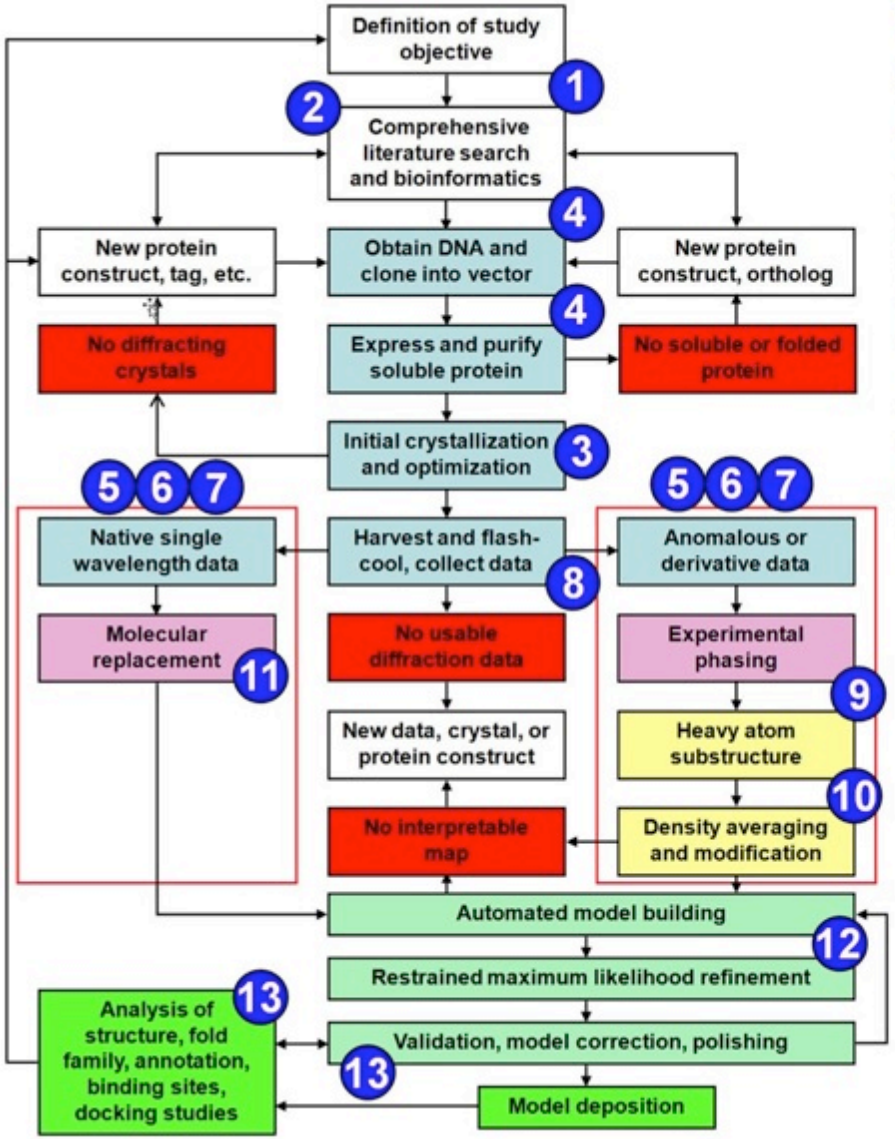
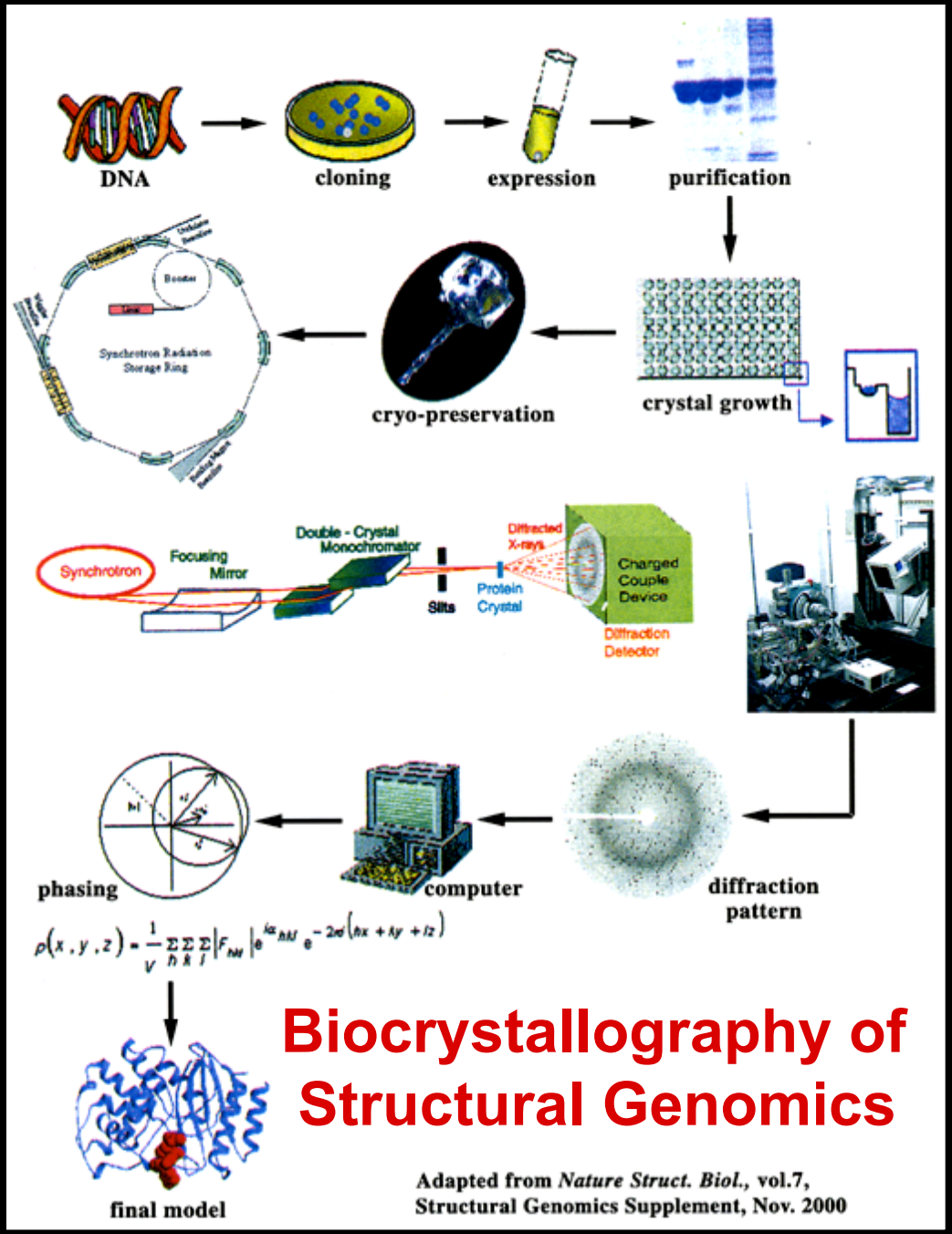
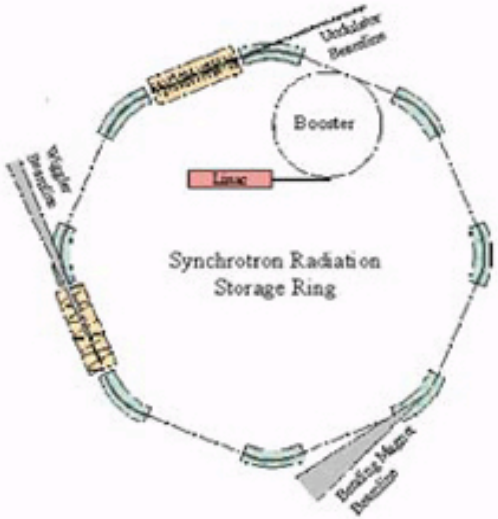
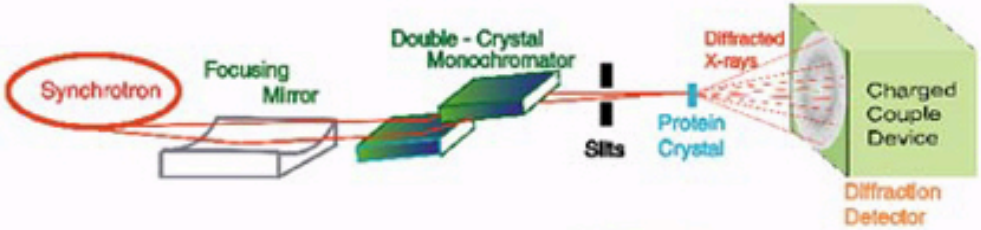


Figure 1-8 Key stages in the structure determination process. The flow diagram provides an overview of the major steps in a structure determination project, labeled with the chapter numbers treating the subject or related general fundamentals. Blue shaded boxes indicate experimental laboratory work, while all steps past data collection are conducted *in silico*. Protein production is discussed in Chapter 4, once we understand the process of crystallization and the requirements a protein must meet to be able to crystallize (Chapter 3).

Biomolecular Crystallography B.Rupp (2010)


DNA
cloning
expression
purification
crystal growth
cryo-preservation
Synchrotron Radiation Storage Ring
Booster
Linac
Synchrotron
Focusing Mirror
Double - Crystal Monochromator
Slits
Protein Crystal
Diffracted X-rays
Charged Couple Device
Diffraction Detector
diffraction pattern
computer
phasing
final model
Biocrystallography of Structural Genomics
Adapted from Nature Struct. Biol., vol.7, Structural Genomics Supplement, Nov. 2000

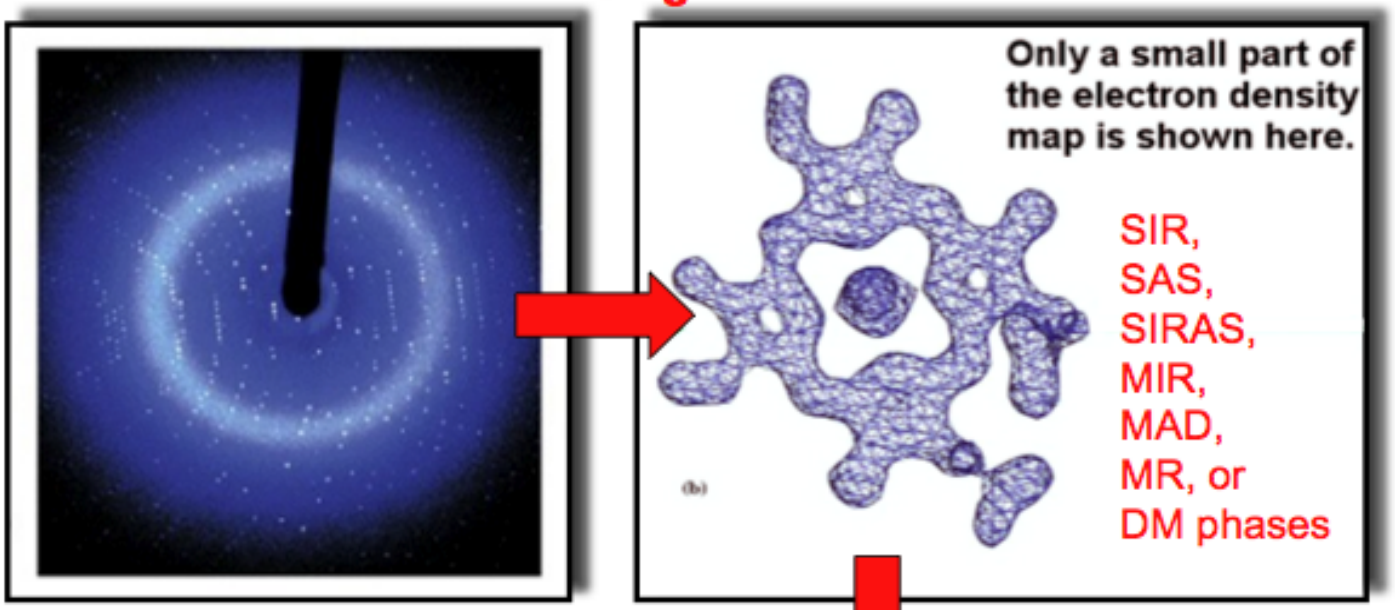

Booster
Linac
Synchrotron Radiation
Storage Ring


Synchrotron
Focusing
Mirror
Double - Crystal
Monochromator
Slits
Protein
Crystal
Diffracted
X-rays
Charged
Couple
Device
Diffraction
Detector

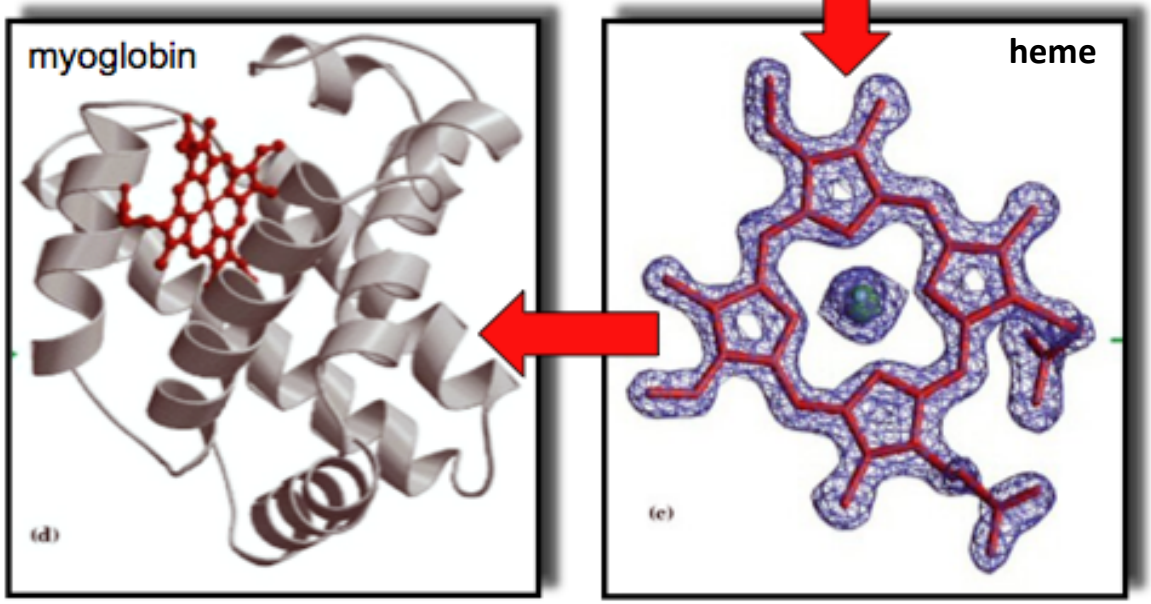
Biomacromolecular Crystal Structure Analysis

Phasing

Diffraction Measurements



Model Fitting



Structure-Function Analysis

Model Refinement

Hierarchy and organization of protein structure

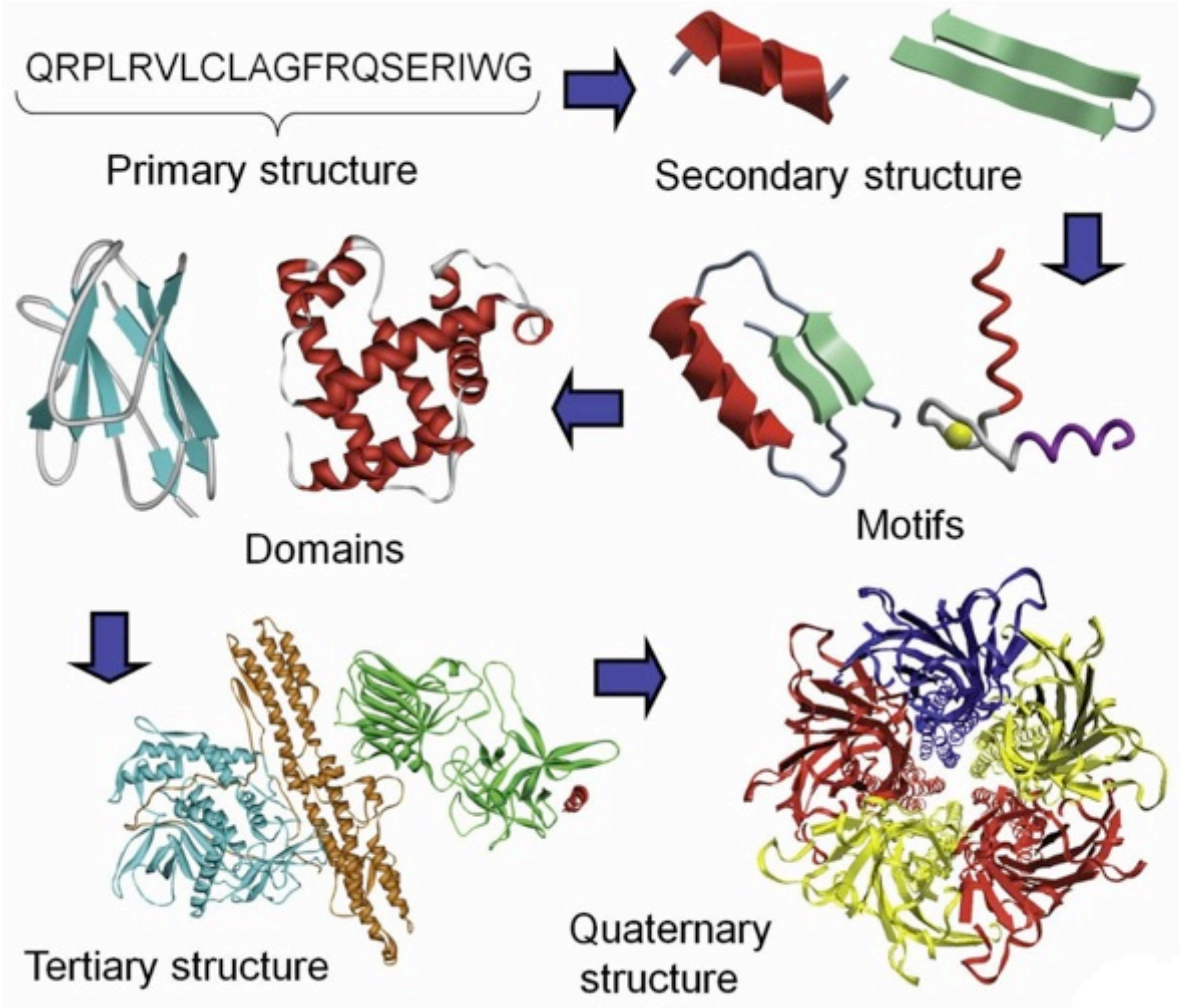


Figure 2-3 The hierarchy of protein structure organization. From a chain of defined amino acid sequence or primary structure, secondary structures form through distinct backbone interactions. Several secondary structure elements combine through side chain interactions into smaller structural motifs, which together with other secondary structure elements form distinct, independently folding structural domains. From the same protein chain, several domains with different functions can fold, assembling into the complete tertiary structure. Finally, several protein chains can form functionally important homo- or hetero-oligomeric assemblies, defining the quaternary structure.

1° aa sequence

2° α, β, loop fragments and motifs

3° domain folds

4° domains assembly

Molecular geometry is determined by distances, angles, and torsions

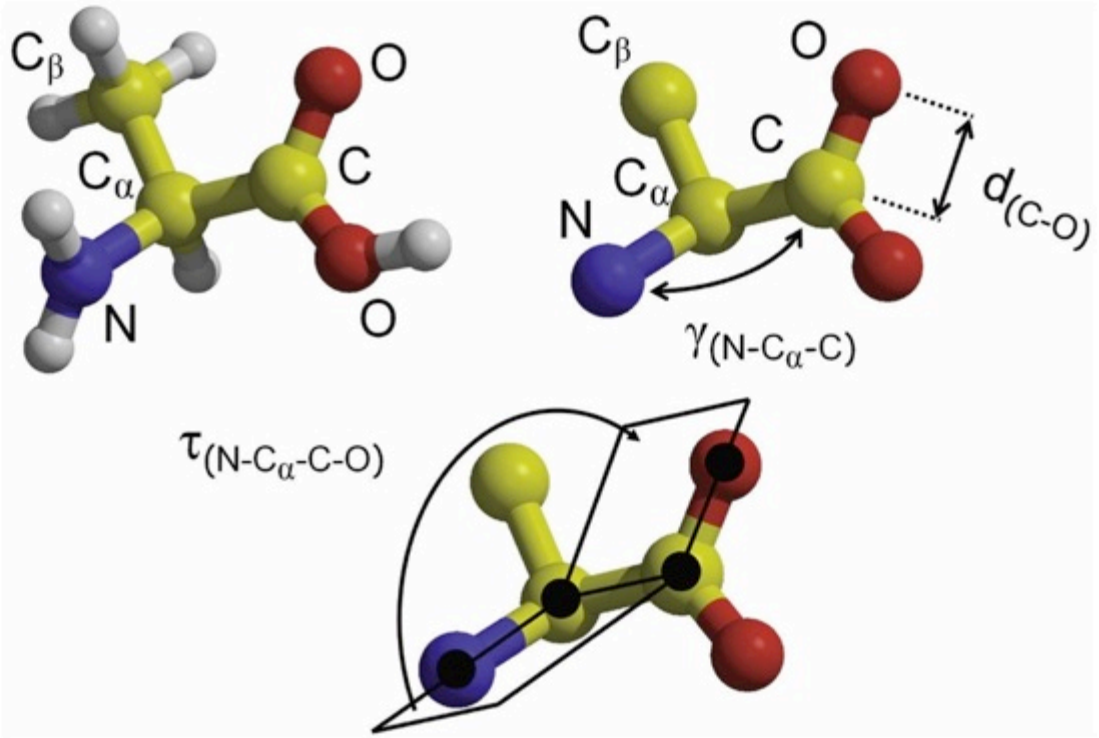


Figure 2-4 Definition of bond length, bond angle, and torsion angle. A ball and stick representation of the amino acid alanine is labeled with atom names and shows the definition of bond lengths, bond angles, and torsion angles. The color scheme applied to distinguish atoms is a common default scheme used by protein model display programs (carbon atoms yellow, oxygen red, nitrogen blue). Hydrogen atoms in riding positions are explicitly shown only in the top left panel (gray atoms) and are normally omitted. Because the torsion angle can be interpreted as the angle between two planes, torsion angles are also called dihedral angles.

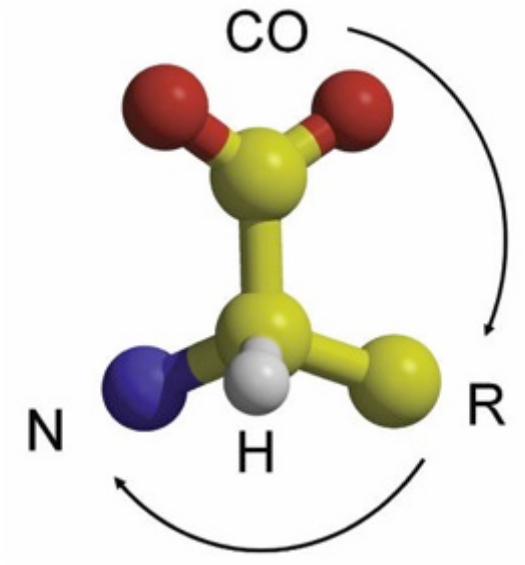
Biomolecular Crystallography B.Rupp (2010)

Peptide bond	Average length (Å)	Single bond	Average length (Å)	Hydrogen bond	Average length (Å)
Cα–C	1.525 ± 0.021	C–C	1.540 ± 0.027	O–H - - - O–H	2.8 ± 0.2
C–N	1.329 ± 0.014	C–N	1.489 ± 0.030	N–H - - - O=C	2.9 ± 0.2
N–C	1.458 ± 0.019	C–O	1.420 ± 0.020	O–H - - - O=C	2.8 ± 0.2

Table 2-1 Common bond lengths. Mean values and standard deviations of some covalent bond lengths and hydrogen bonds commonly occurring in protein structures. The average or mean values are derived from accurate small molecule X-ray structures and high resolution protein structures. Because they also serve as the basis for stereochemical restraints in macromolecular refinement (Chapter 12), the mean values are also known as restraint target values.[11]

Cα -chirality - the CO-R-N rule

Sidebar 2-2 Hydrogen atoms in protein structure models. In protein crystallographic work, the hydrogen atoms are generally only visible in electron density at ultra high resolution (about 1.2 Å or better). Because their so-called *riding* positions are known and can be calculated when needed, they are normally omitted in crystallographic models (the presence of hydrogen atoms is, however, implicitly and fully accounted for in crystallographic distance and energy restraints). The remaining panels of Figure 2-4 therefore show alanine in a representation typical for a crystallographic model, *sans* hydrogens. We will generally omit hydrogen atoms in riding positions in protein structure models, unless they are needed to emphasize a structural or stereochemical feature.



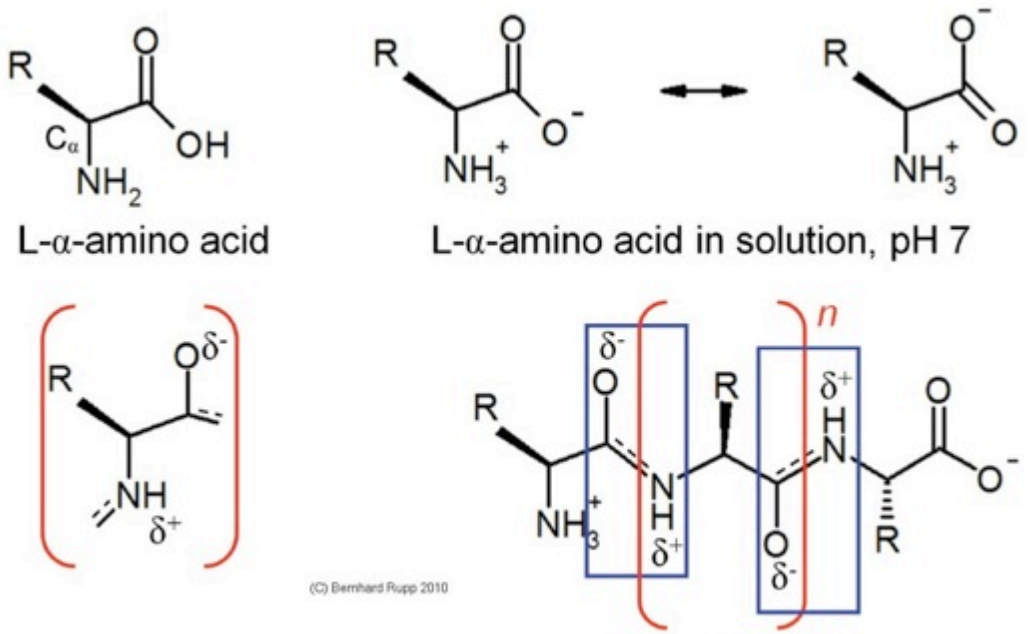
Biomolecular Crystallography B.Rupp (2010)

Figure 2-5 The CORN rule. The CORN rule is a practical aid for determining the configuration of chiral Cα centers of amino acids. When the central Cα-atom is viewed with the H atom pointing toward the observer (or out of the paper plane), the ligand sequence reads "CO-R-N" in clockwise rotation for a L-amino acid.

With exception of non-chiral glycine, all proteinogenic amino acids are chiral with 2S configuration – L- amino acids

Primary structure, sequence, peptide bond

Figure 2-19 Definition of L-α-amino acid, residue, and polypeptide chain. Hydrogen atoms at the chiral Cα atom are omitted. Top left: L-α-amino acid in uncharged state. In solution, the functional terminal groups are charged, and because of resonance the two oxygen atoms are equivalent. Bottom row: a residue of an L-α-amino acid. A residue lacks the second carboxyl oxygen and the second hydrogen atom on the amide group, which are lost during formation of the peptide bonds. The bottom right panel shows the formal composition of a polypeptide chain of *n* residues, with the blue boxes containing the planar peptide bonds.



99.9% non-Pro trans

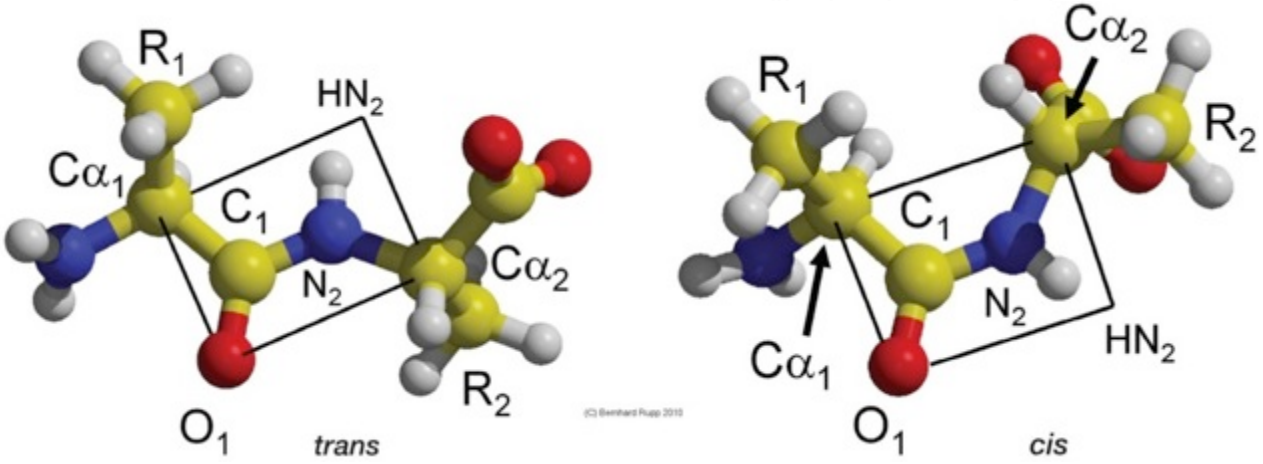


Figure 2-8 *Trans*- and *cis*-peptide bonds. The *trans*-peptide bond, recognizable by the two sequential Cα atoms opposed, has an ω-angle of 180° while the *cis*-peptide bond with an ω-angle of 0° can be easily recognized by the Cα atoms located on the same side of the peptide bond. In both cases, the peptide bond is planar and contains the same atoms, but in *cis* conformation (right panel) the torsional freedom is much more limited because of side chain collisions. The Cα–Cα distance in a *cis*-peptide is ~2.9 Å, while the Cα–Cα distance across a *trans*-peptide bond is ~3.8 Å.

Protein backbones prefer *trans* geometry

Figure 2-8 *Trans*- and *cis*-peptide bonds. The *trans*-peptide bond, recognizable by the two sequential Cα atoms opposed, has an ω-angle of 180° while the *cis*-peptide bond with an ω-angle of 0° can be easily recognized by the Cα atoms located on the same side of the peptide bond. In both cases, the peptide bond is planar and contains the same atoms, but in *cis* conformation (right panel) the torsional freedom is much more limited because of side chain collisions. The Cα–Cα distance in a *cis*-peptide is ~2.9 Å, while the Cα–Cα distance across a *trans*-peptide bond is ~3.8 Å.

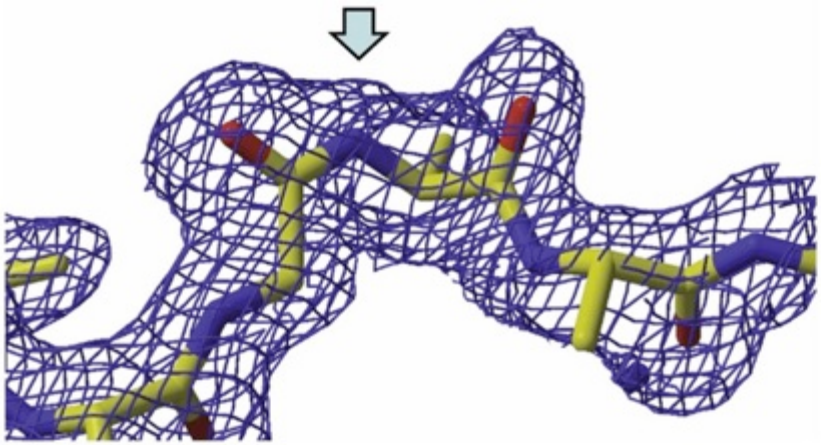
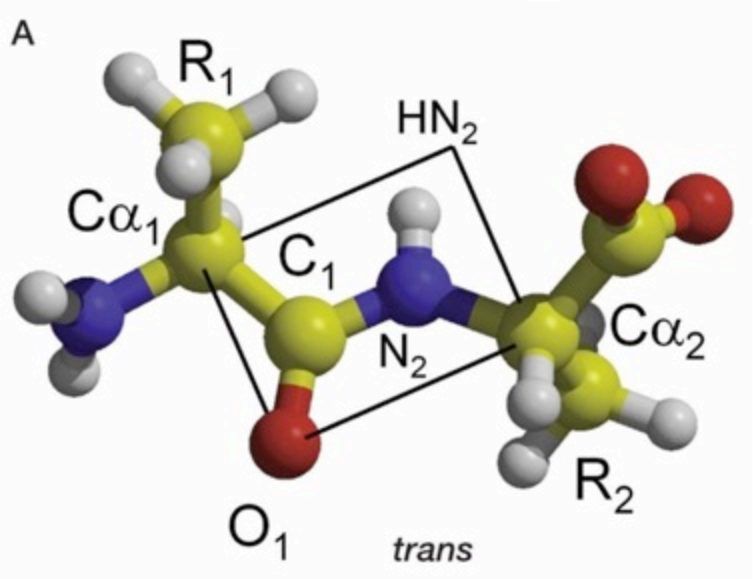
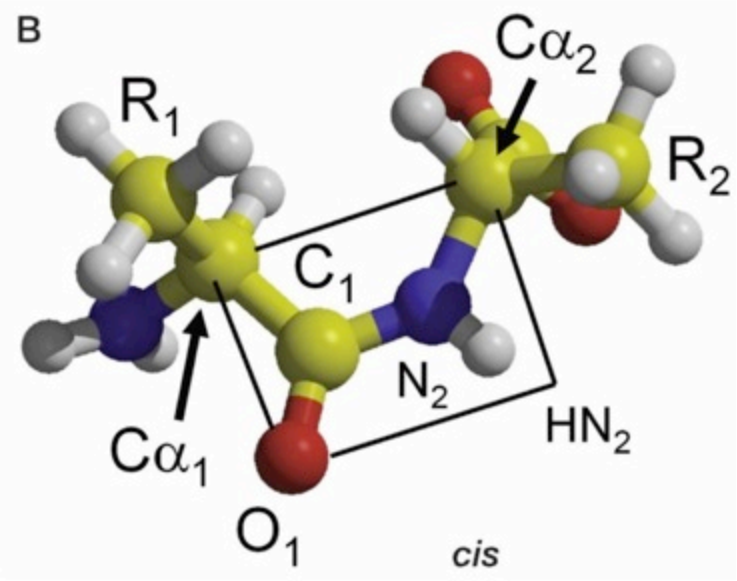


Figure 2-9 *Cis*-peptide bond and its electron density. A *cis*-peptide bond between Gly and Ala in a connecting loop is shown in experimental electron density (blue grid). Both the clear electron density and the fact that the *cis* conformation is structurally conserved in this particular fold family allow making this particular cis assignment with confidence. PDB entry 1upi.[13]





Secondary structure is defined by van der Waals repulsion and backbone torsion – not side chains!

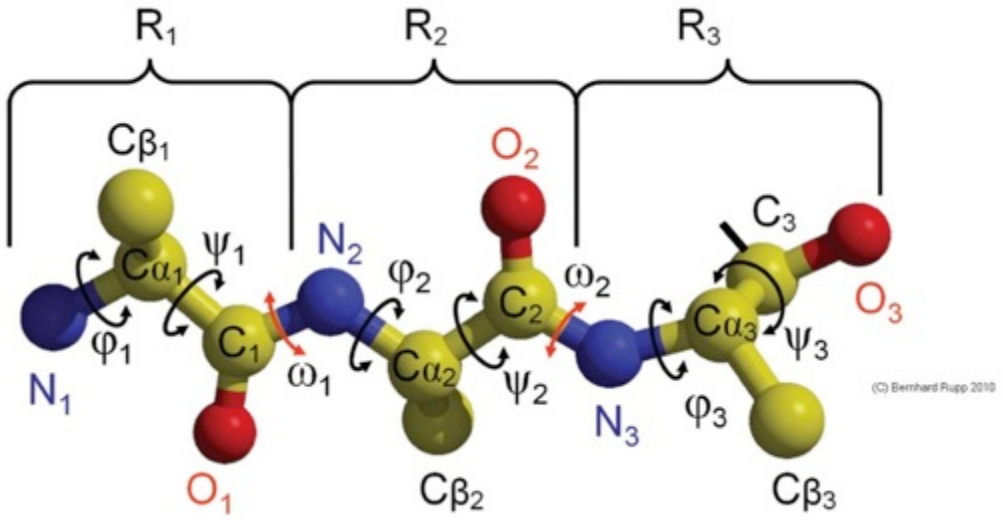


Figure 2-7 Backbone torsion angles. The N-terminal 3-residue stretch of a peptide Ala-Ala-Ala- containing three *trans*-peptide bonds is shown. Three torsion angles for each residue, φ (phi), ψ (psi), and ω (omega), define the conformation of the peptide backbone. While combinations of torsion angles φ and ψ are only restrained by van der Waals repulsion and fall into several allowed, energetically favored regions, the *trans* omega-torsion around the partially delocalized, planar peptide bond (short red arrow) is highly restrained to 180°.

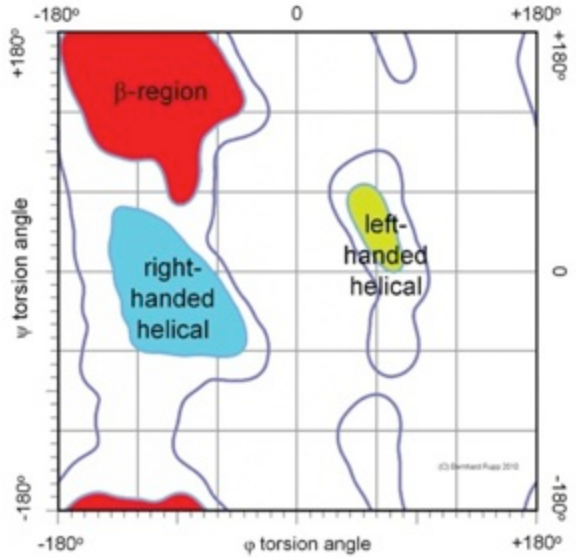


Figure 2-10 Backbone torsion angle distribution. The protein backbone torsion angles are represented in the form of an empirical Ramachandran plot.[15] The colored regions are energetically most preferred, with the surrounding regions having higher torsional energies. The φ–ψ combinations outside the blue contoured regions are so unfavorable that they are usually not observed. The torsion angles of various secondary structure elements cluster in typical regions. The right-handed helical region largely contains residues in right-handed α-helices, and some 3_{10}- and π-helices. The β-region contains β-strands, β-proline helices, and β-turns, while the left-handed helical region contains torsion angle pairs as observed in left-handed helices. Note that the left-handed helical region has opposite signs in the φ–ψ distribution compared with the regular, right-handed helical region. Left-handed helix conformations tend to be short and rare, because a clash between backbone carbonyl oxygen and Cβ atoms makes them energetically less favorable.

- Ramachandran energy surface
- outliers indicate points of high local energy
- ~2-5% high energy conformations are normal
- not restrained in refinement – will reliably reveal errors

Protein secondary structure

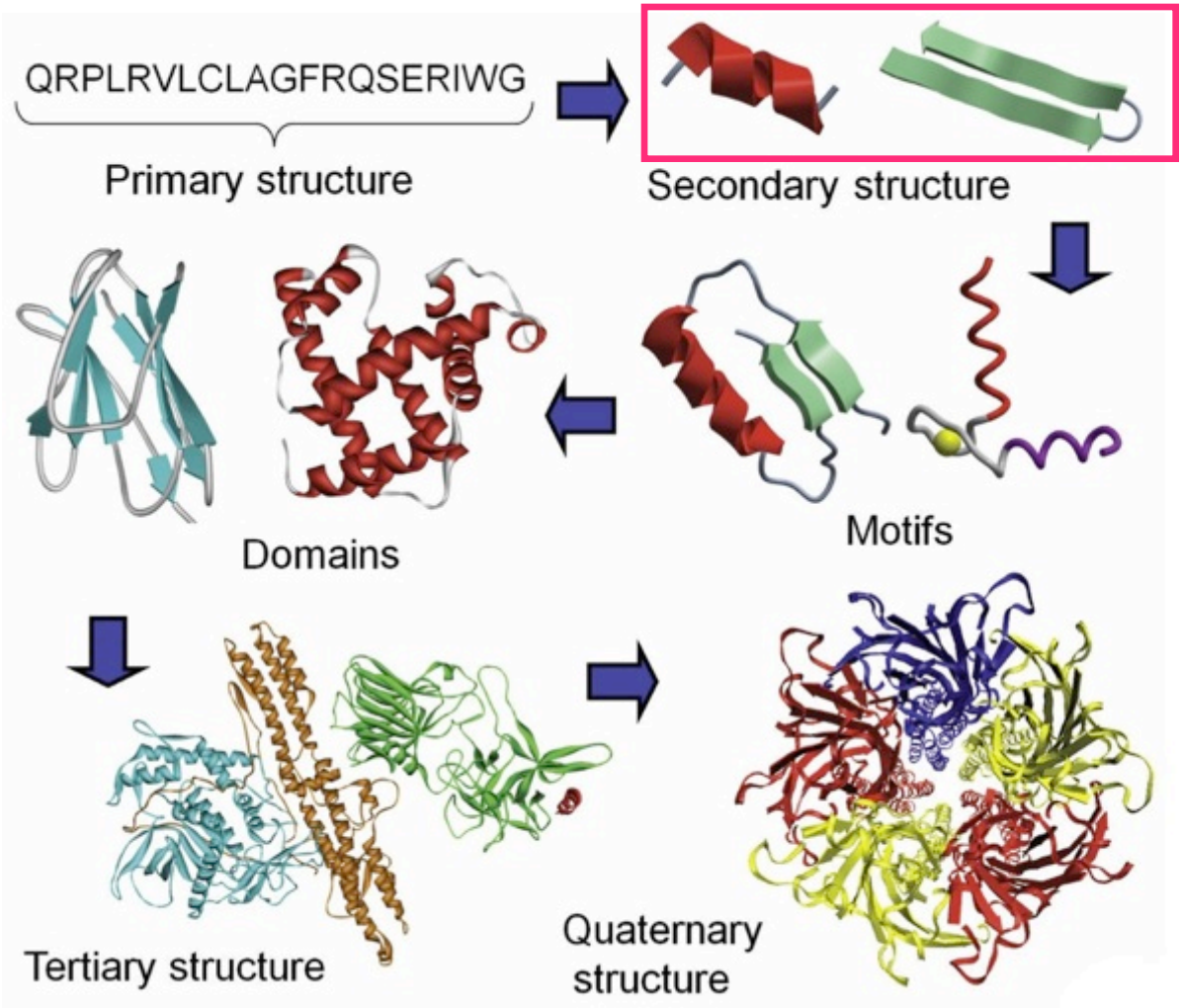


Figure 2-3 The hierarchy of protein structure organization. From a chain of defined amino acid sequence or primary structure, secondary structures form through distinct backbone interactions. Several secondary structure elements combine through side chain interactions into smaller structural motifs, which together with other secondary structure elements form distinct, independently folding structural domains. From the same protein chain, several domains with different functions can fold, assembling into the complete tertiary structure. Finally, several protein chains can form functionally important homo- or hetero-oligomeric assemblies, defining the quaternary structure.

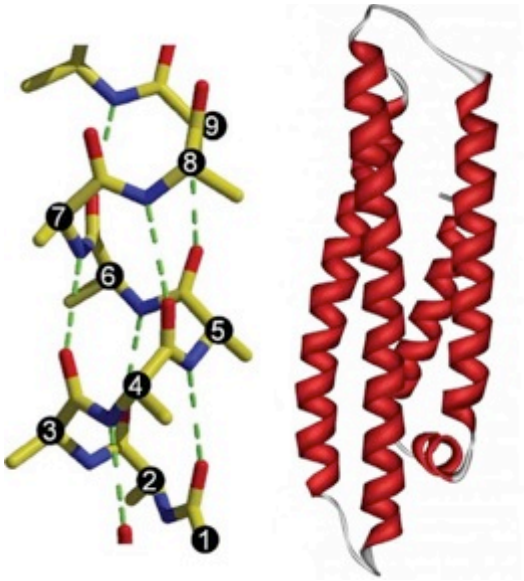
1° aa sequence

2° α, β, loop fragments and motifs

3° domain folds

4° domains assembly

Secondary structure elements – helices, sheets, turns and loops



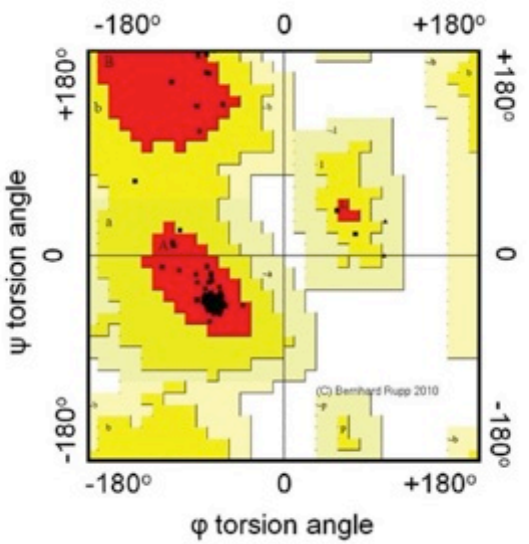
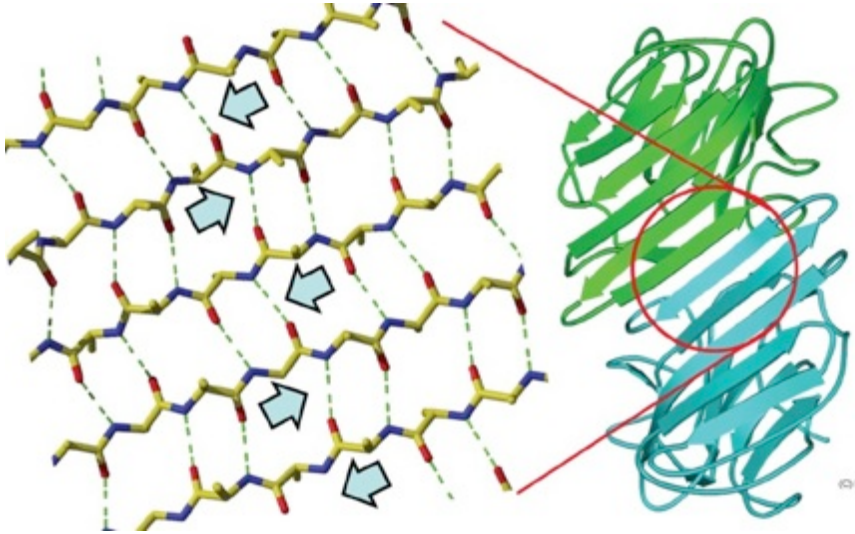


Figure 2-11 Protein α-helices. Left panel shows a polyalanine stretch in α-helical conformation. Note the hydrogen bonds (green dotted lines) from the backbone carbonyl oxygen of residue *n* to the backbone nitrogen of residue *n*+4. The residues are labeled at the Cα carbon atom. Center panel: A typical 4-helix bundle is the 22 kDa fragment of apolipoprotein E4, an allelic isoform of apo-E implicated in late onset familial Alzheimer's disease.[23] Note that in the Ramachandran plot[24] (right panel) practically all residues except those few located in the connecting loops and turns are sharply clustered in the right-handed α-helical region. PDB entry 1b68.[25]



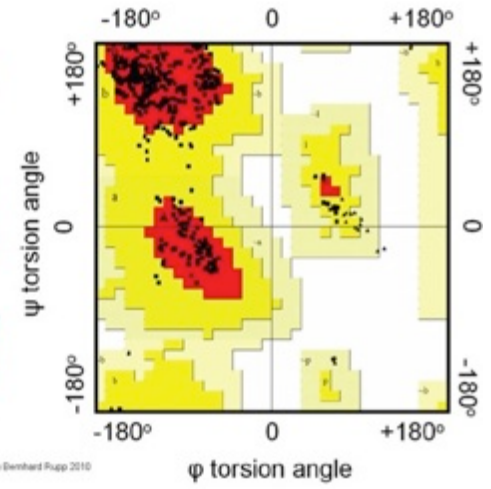


Figure 2-15 The β-sheet sandwich structure of concanavalin A. Left panel: a magnification of the hydrogen bond pattern in the large antiparallel sheet continuing across the dimer interface in the ribbon diagram. The dimer interface between the green and cyan monomers is formed through intermolecular continuation of the antiparallel β-sheet between the two molecules (two of the depicted dimers finally form an active tetramer). In the Ramachandran plot (right panel) the large majority of β-sheet torsion angle pairs are clustered in the left side of the β-region. The smaller, second cluster in the right part of the β-region results from the significant number of β-turns present in this structure. The remaining turns and loops contain helical torsions (but do not form complete helices), and a few torsions fall into the left-handed helical region. PDB entry 1gkb.[32]

About helix handedness and screw axes

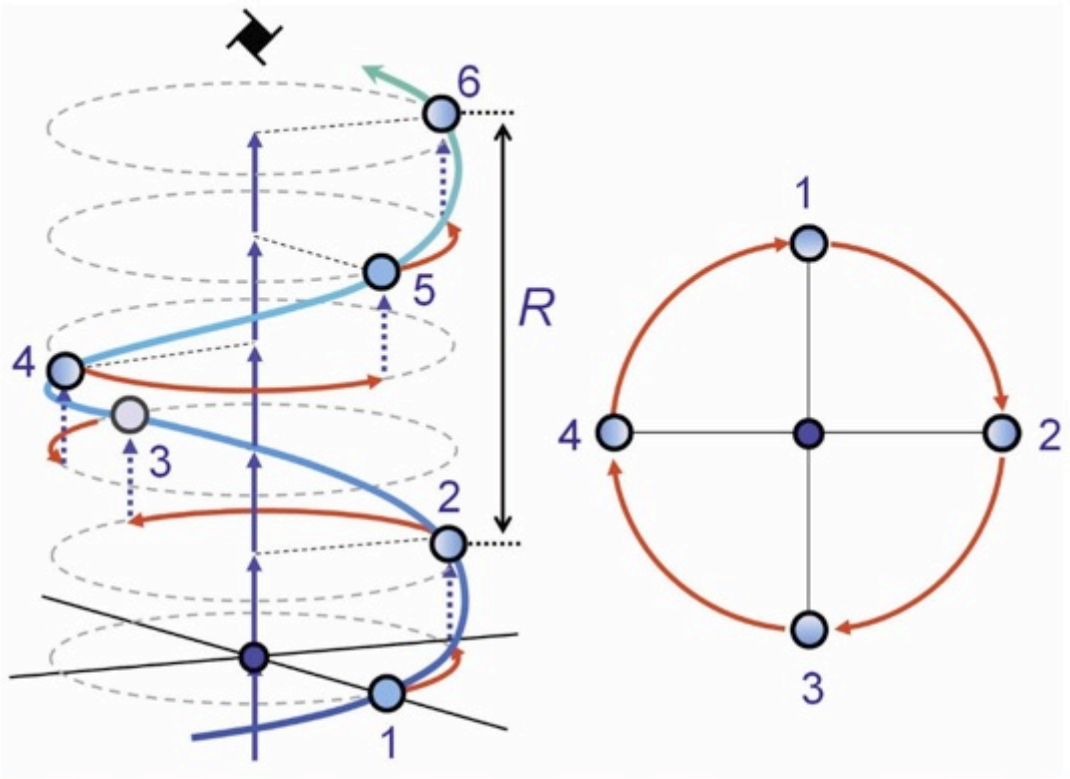


Figure 2-12 Helix orientation. A succession of counter clockwise rotations by 90° in a projection with the helical axis pointing upwards plus translation along the helical axis leads to a right-handed helix (left drawing). The sense or handedness of the helix is determined by looking along the helical axis (right panel) and following the sequence of the points (it does not matter in which direction one looks along the helix- or screw axis, the handedness remains the same). In crystallographic plain axes as well as screw axes, the rotation is exactly 360/*n* degrees, with *n* = 2, 3, 4, and 6. The depicted helix is compatible with a crystallographic 4-fold (4_1) screw operation, indicated by the symbol above it. The right-handed protein α-helix in contrast has a non-integer number of ~3.6 residues per turn, corresponding to ~100° counter clockwise rotation between residues. Screw operations are further explained in Chapter 5.

Other helices: the 3_{10} helix and the PP-II helix

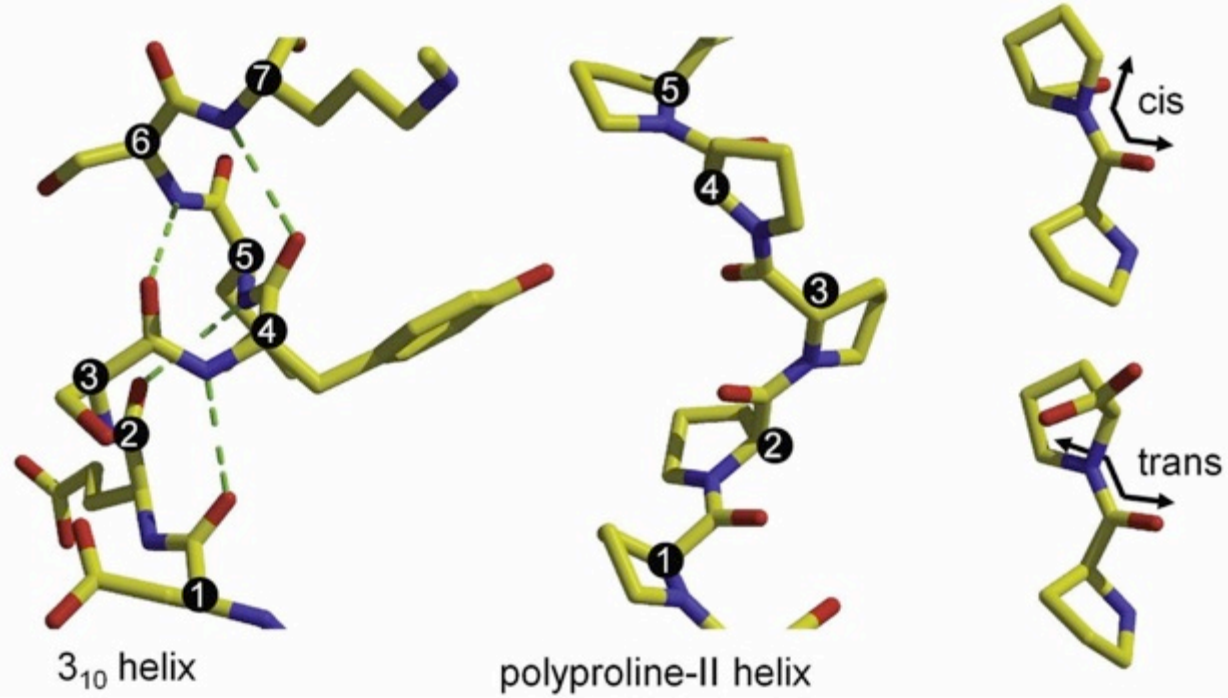


Figure 2-13 The right-handed 3_{10} helix and the left-handed trans-polyproline-II helix. The residues of the helices are numbered at their Cα positions. Note the *n*+3 backbone hydrogen bonds in the 3_{10} helix. No hydrogen network can exist in the proline helix because of the absence of a hydrogen atom on the proline nitrogen. The right panel emphasizes the *cis* and *trans* conformation of a proline dipeptide. Note the similar conformation of the proline rings in both *cis* and *trans* conformation. PDB entries 2d2e[30] and 1vzj.[31]

Biomolecular Crystallography B.Rupp (2010)

Strictly speaking the PP-II helices are not secondary structure elements because they lack the structure-stabilizing hydrogen-bonded backbone interactions

Pleated β-sheet structures are mostly parallel
Or anti-parallel, sometimes irregular

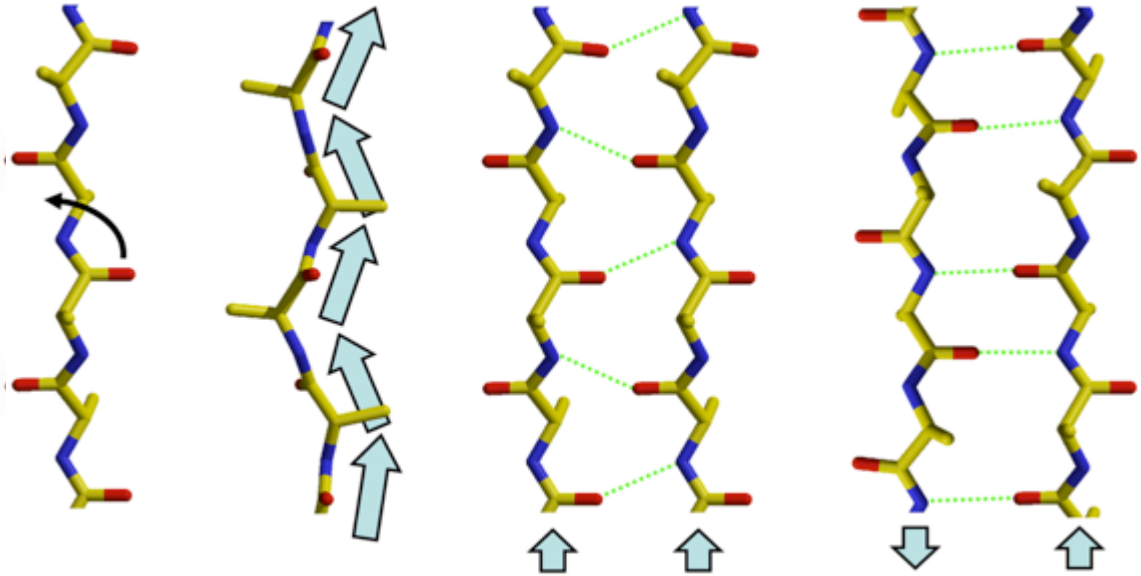


Figure 2-14 Formation of β-sheets from individual β-strands. Left pair: the extended conformation of the β-strand viewed in the sheet plane, and viewed across the sheet plane. Note that the carboxyl groups alternate in extending *in* the plane of the sheet, while the residues protrude alternating *above* and *below* each strand. Also visible is the typical pleat of the sheet in cross section. Right two pairs: inter-strand hydrogen bond arrangement in parallel and antiparallel β-sheets.

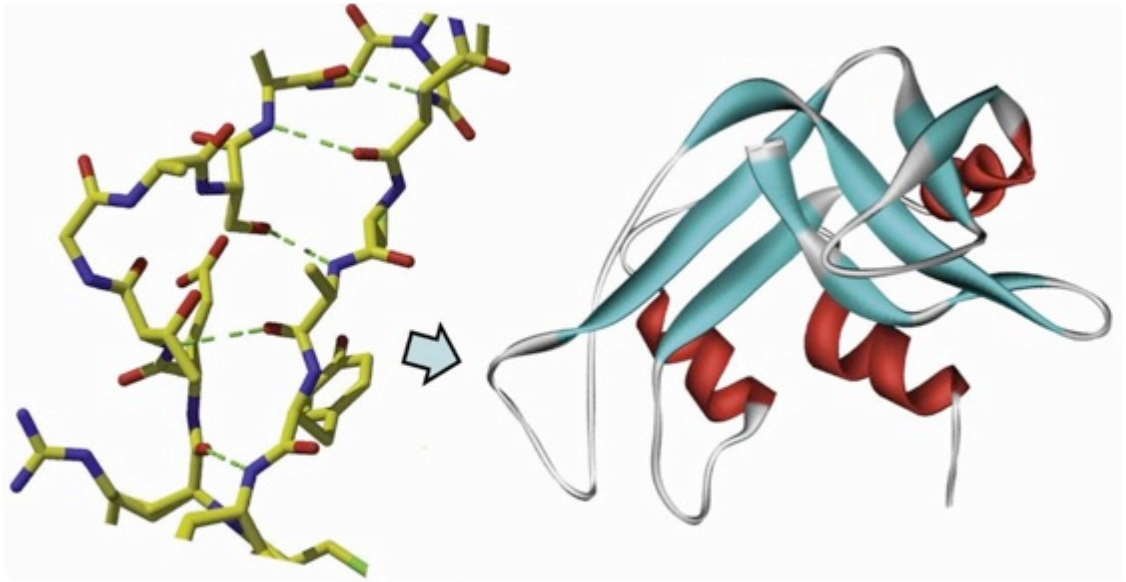


Figure 2-16 Ribonuclease A contains irregular β-sheets and a β-bulge. The left panel shows the interruption of the antiparallel β-sheet through an insertion of an additional residue in one strand. In contrast to the near-perfect regularity of the sheets in concanavalin A (Figure 2-15), the β-strands are irregular and show crossovers and a bifurcation. PDB entry 7rsa.[34]

Secondary structure elements – helices, sheets, turns and loops

Figure 2-17 Types I′ and II′ are the most frequent β-turns. In β-turns, two residues separate the hydrogen bonded residues, while in the α-turn three residues are located between the hydrogen bonded residues. Note that the β-sheet with the α-turn in the right panel has also a bulge after residue number five. PDB entries 1gkb[32] and 1upi.[13]

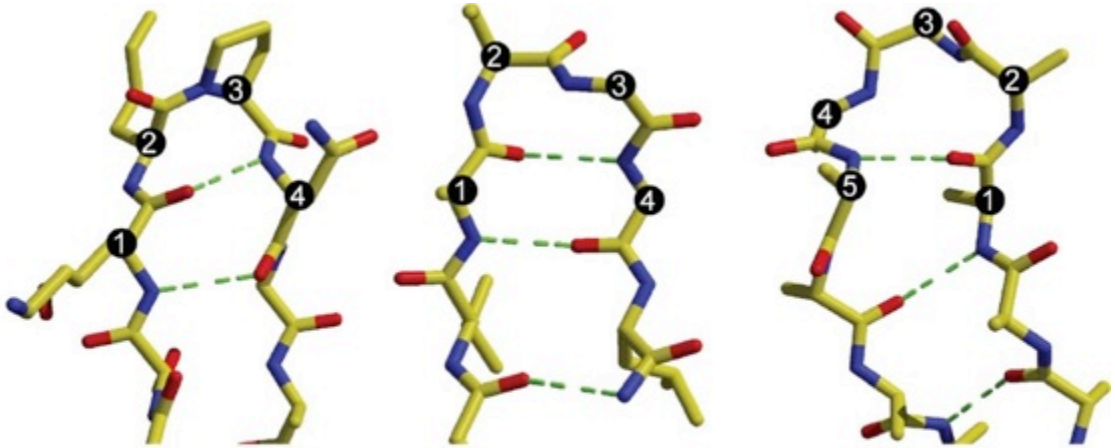
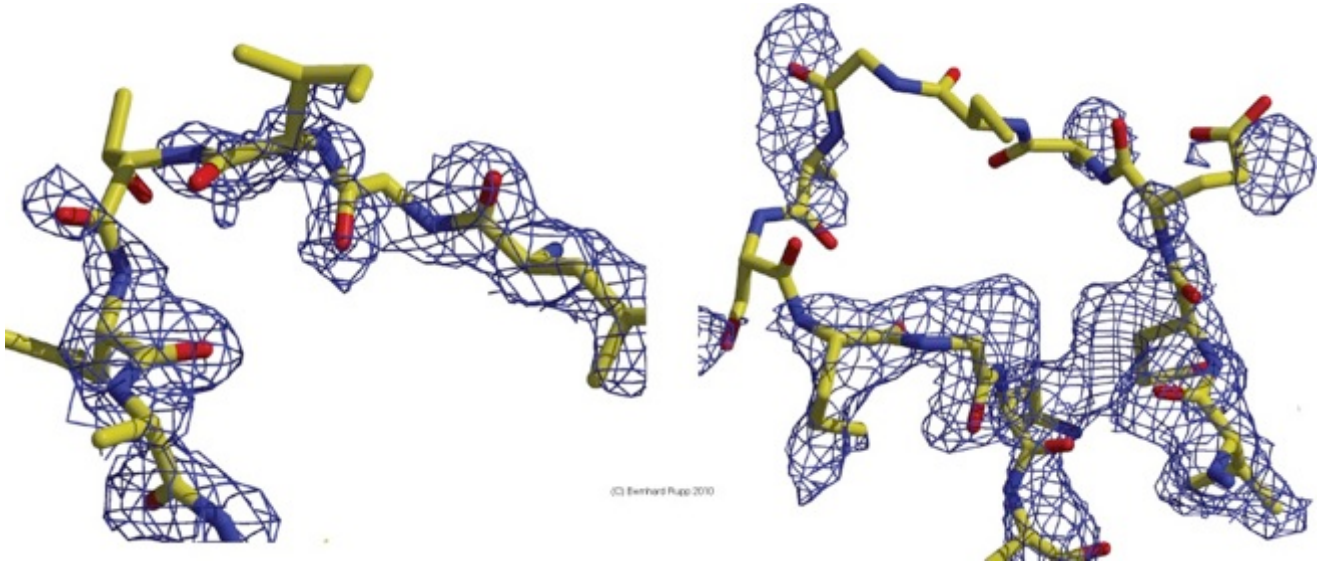


Figure 2-18 Solvent-exposed loops. In the left panel, the backbone conformation of the loop can be reasonably well traced, and the general location of the loop is still reasonably well defined. In the right panel, the situation is much worse, and the tracing of the entire loop is uncertain.



Protein structure hierarchy

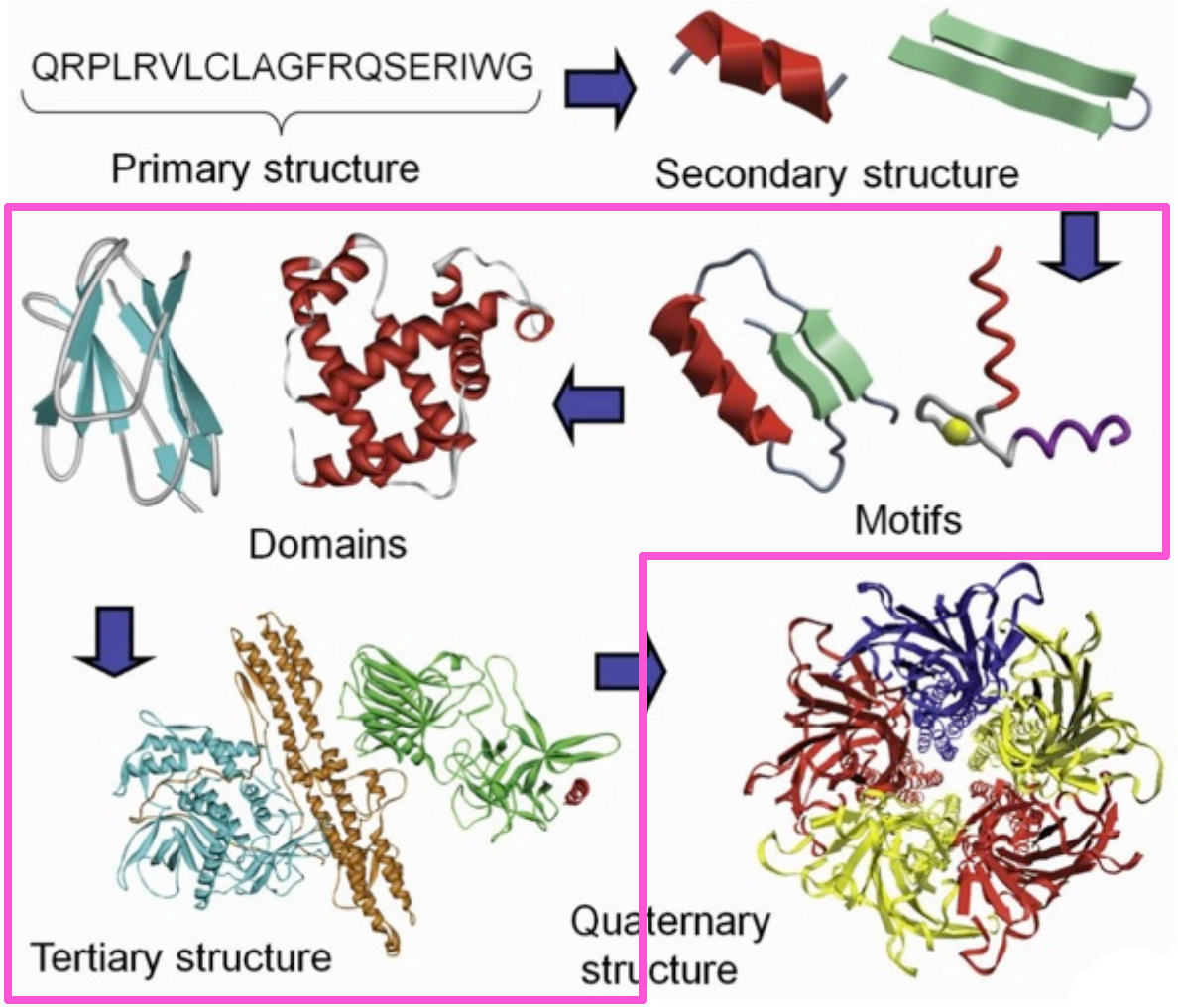


Figure 2-3 The hierarchy of protein structure organization. From a chain of defined amino acid sequence or primary structure, secondary structures form through distinct backbone interactions. Several secondary structure elements combine through side chain interactions into smaller structural motifs, which together with other secondary structure elements form distinct, independently folding structural domains. From the same protein chain, several domains with different functions can fold, assembling into the complete tertiary structure. Finally, several protein chains can form functionally important homo- or hetero-oligomeric assemblies, defining the quaternary structure.

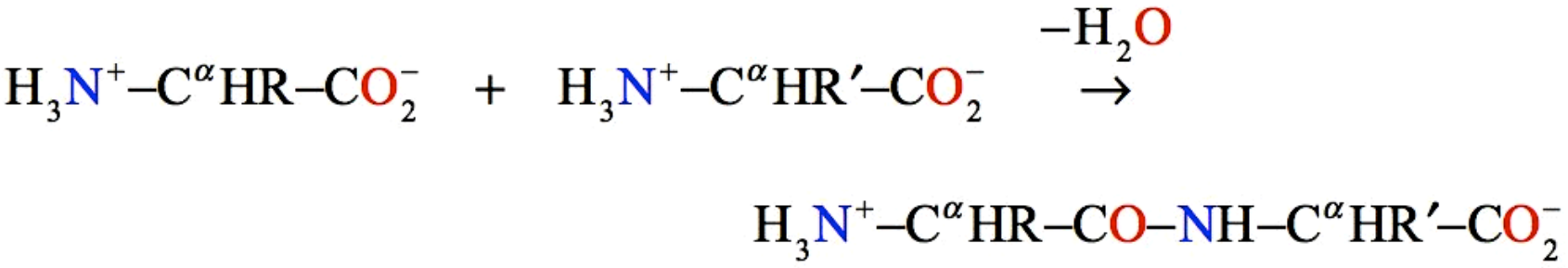
1° aa sequence

2° α, β, loop fragments and motifs

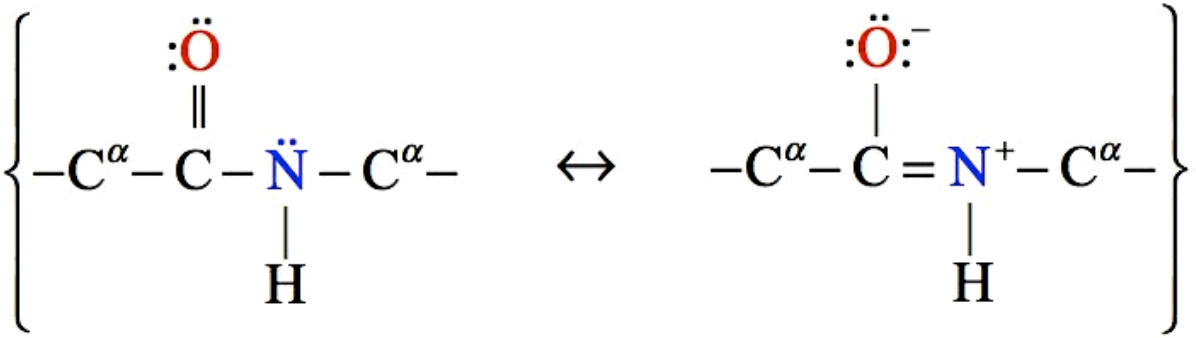
3° domain folds

4° domains assembly

Peptide formation by amino acid condensation

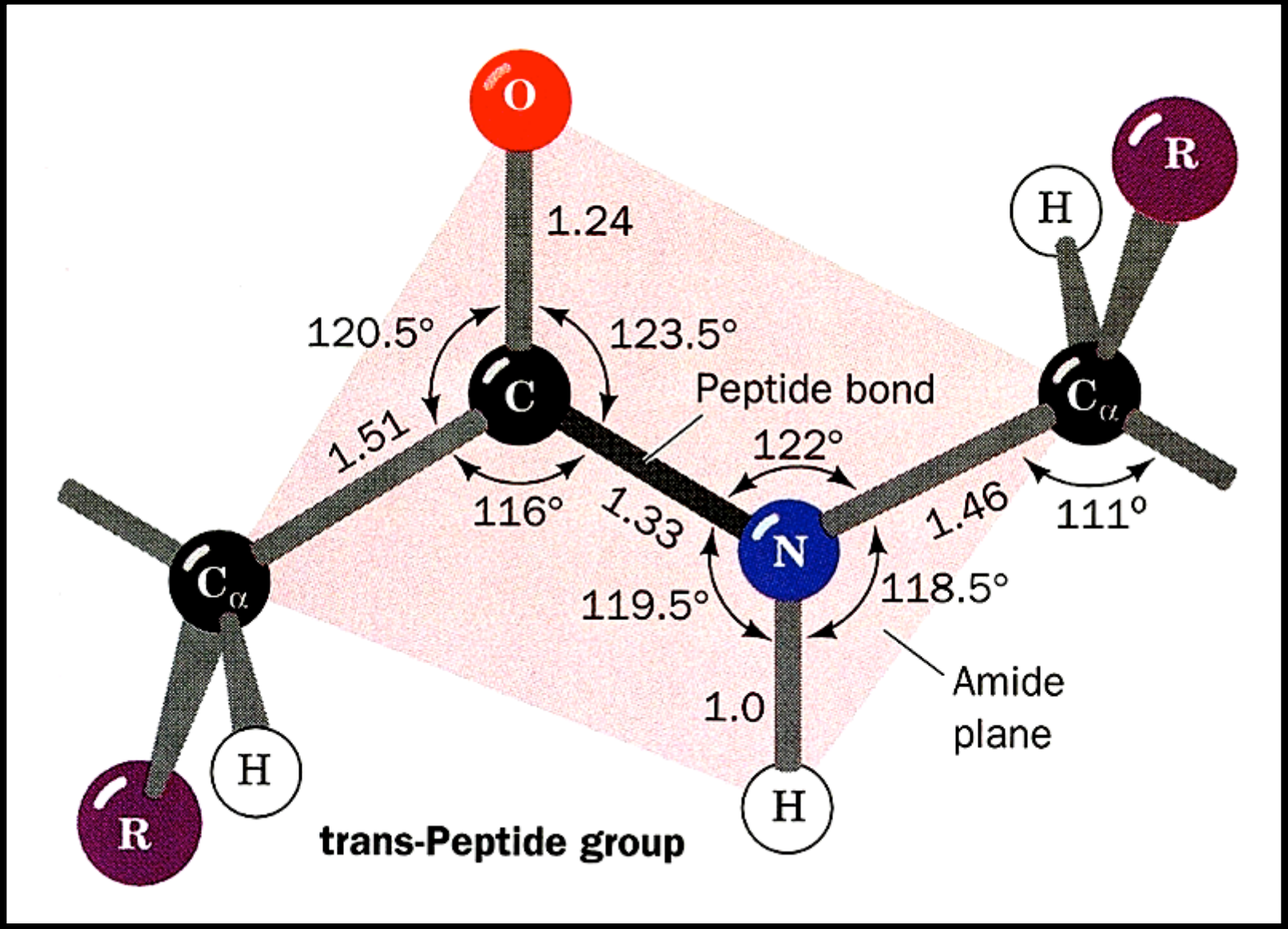


Trans, planar –CO–NH– peptide link



Canonical resonance structures

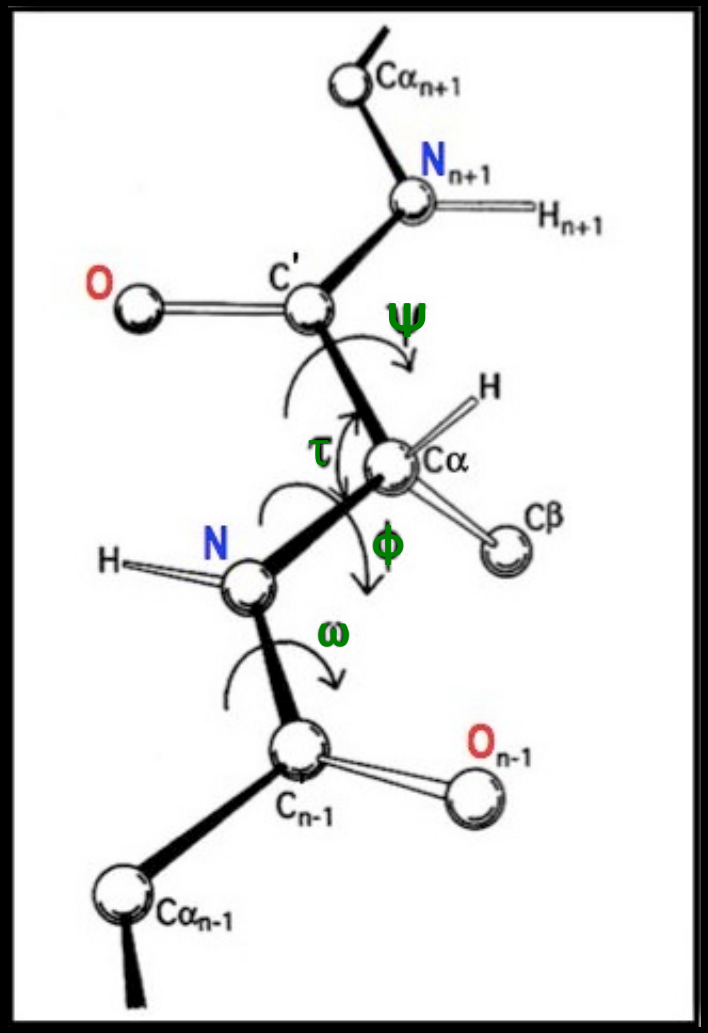
Valence geometry in the *trans*, planar peptide link



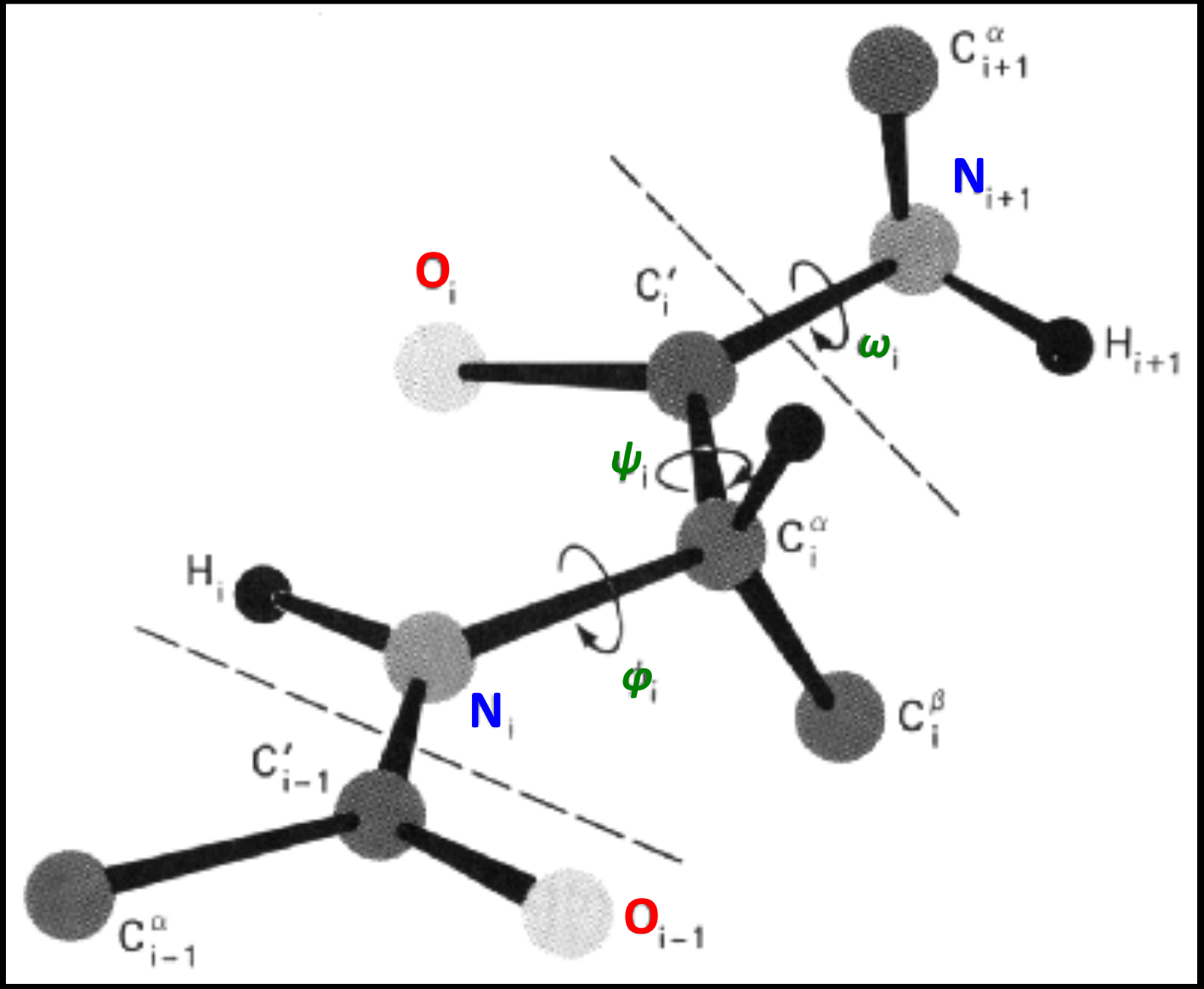
Polypeptide backbone

conformation angles ϕ and ψ, and valence angle τ, at Cα

trans, planar –CO–NH– peptide links with conformation angle $\omega \approx 180°$



Jane S. Richardson, http://kinemage.biochem.duke.edu/teaching/anatax/html/anatax.1b.html


C^{α}_{i+1}
N_{i+1}
O_i
C'_i
ω_i
H_{i+1}
ψ_i
C^{α}_i
H_i
φ_i
N_i
C^{β}_i
C'_{i-1}
C^{α}_{i-1}
O_{i-1}

What is a crystal – formal view

Definition: A 3-dimensional translationally periodic stacking of unit cells

A very useful mathematical concept for computation but it ignores the minutiae of protein self assembly and details of crystal contacts.

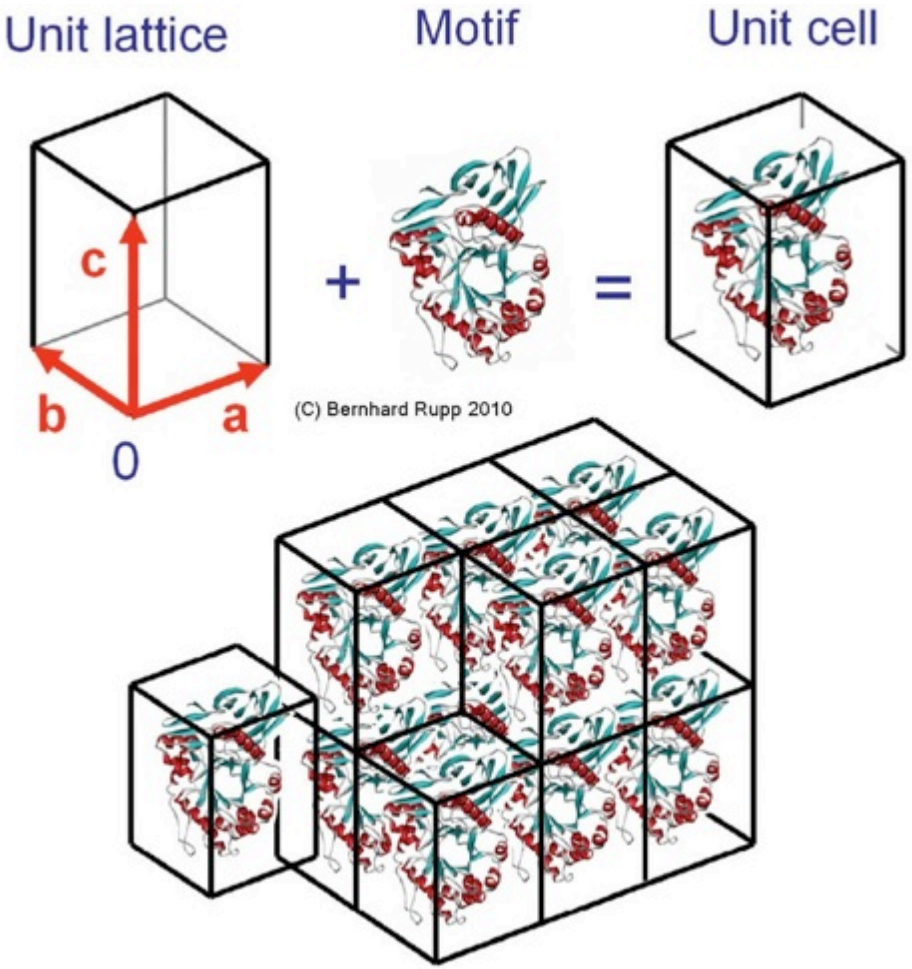


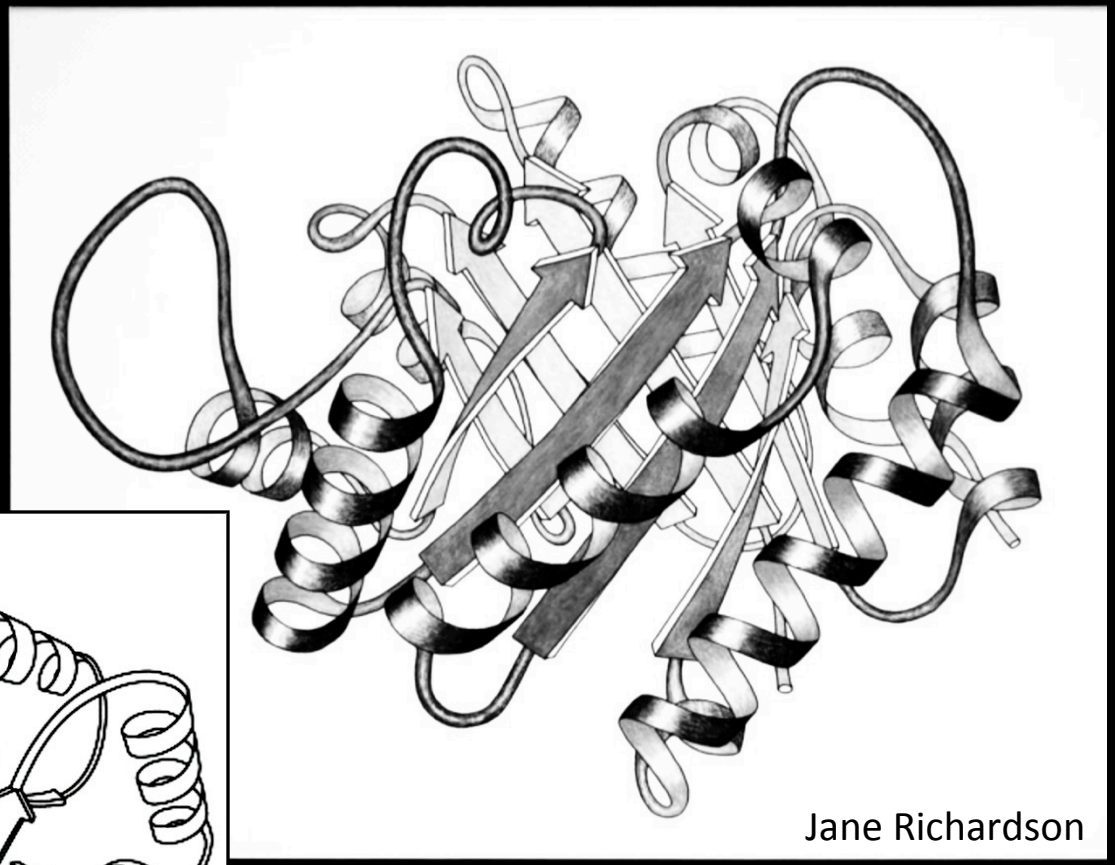
Figure 5-24 Assembly of a primitive triclinic 3-dimensional crystal from unit cells. In analogy to the 2-dimensional case, the unit lattice is filled with a motif, and the crystal is built from translationally stacked unit cells. The basis vectors form a right-handed system [0, **a**, **b**, **c**].

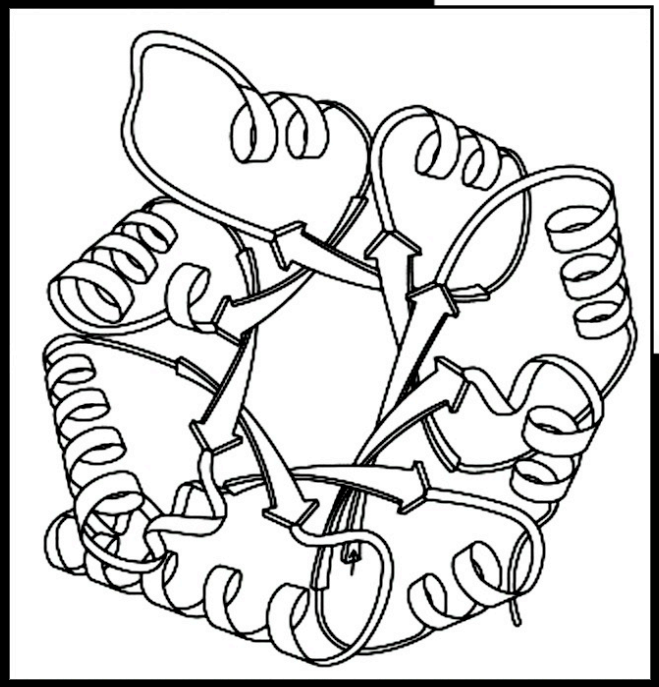
Bernhard Rupp's generic protein molecular structure motif used in his various crystal structure illustrations

a TIM barrel structure

α helix β strand β sheet turn loop

Triosephosphate isomerase (TIM barrel)
Ribbon cartoon as hand-drawn by Jane Richardson (ca. 1980)

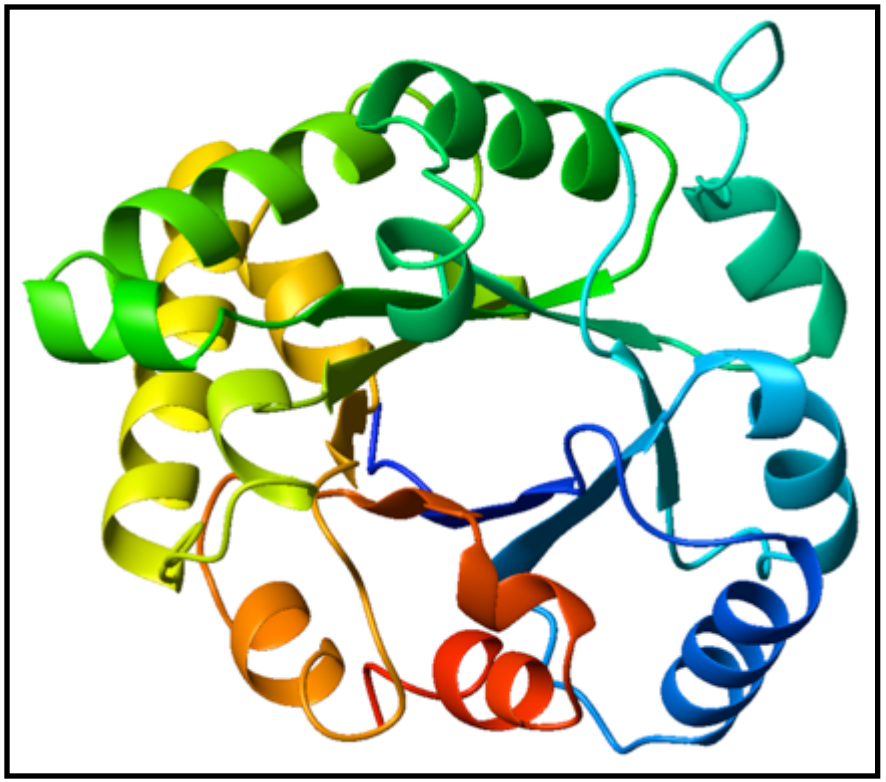


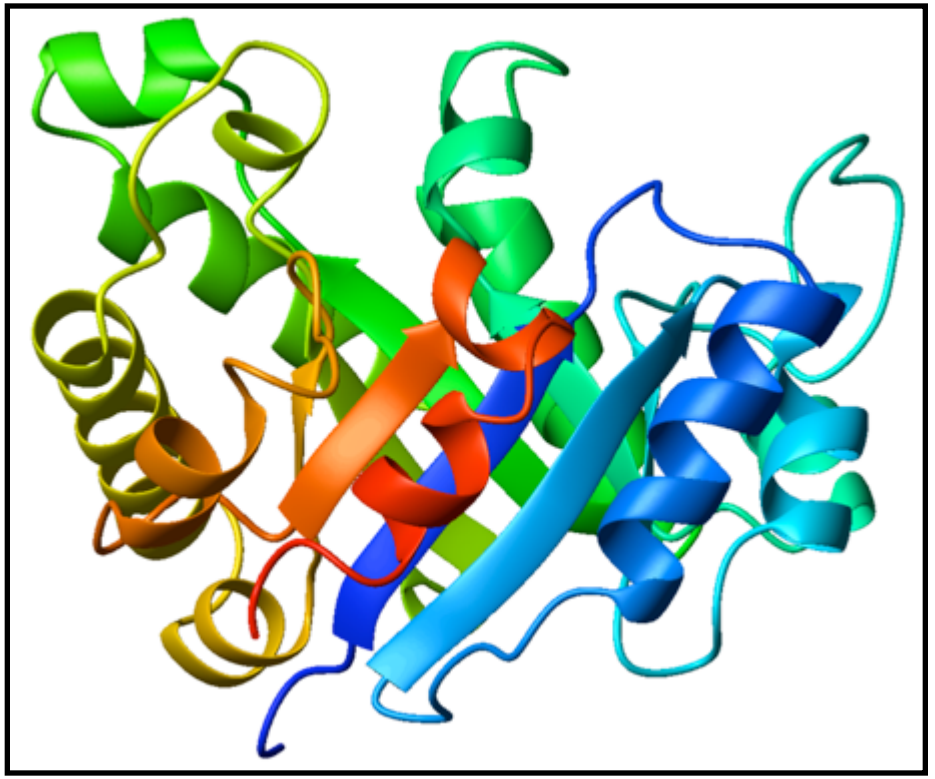


http://en.wikipedia.org/wiki/Ribbon_diagram

Triosephosphate Isomerase Monomer, TIM Barrel

Ribbon cartoon colored by N → C sequence direction



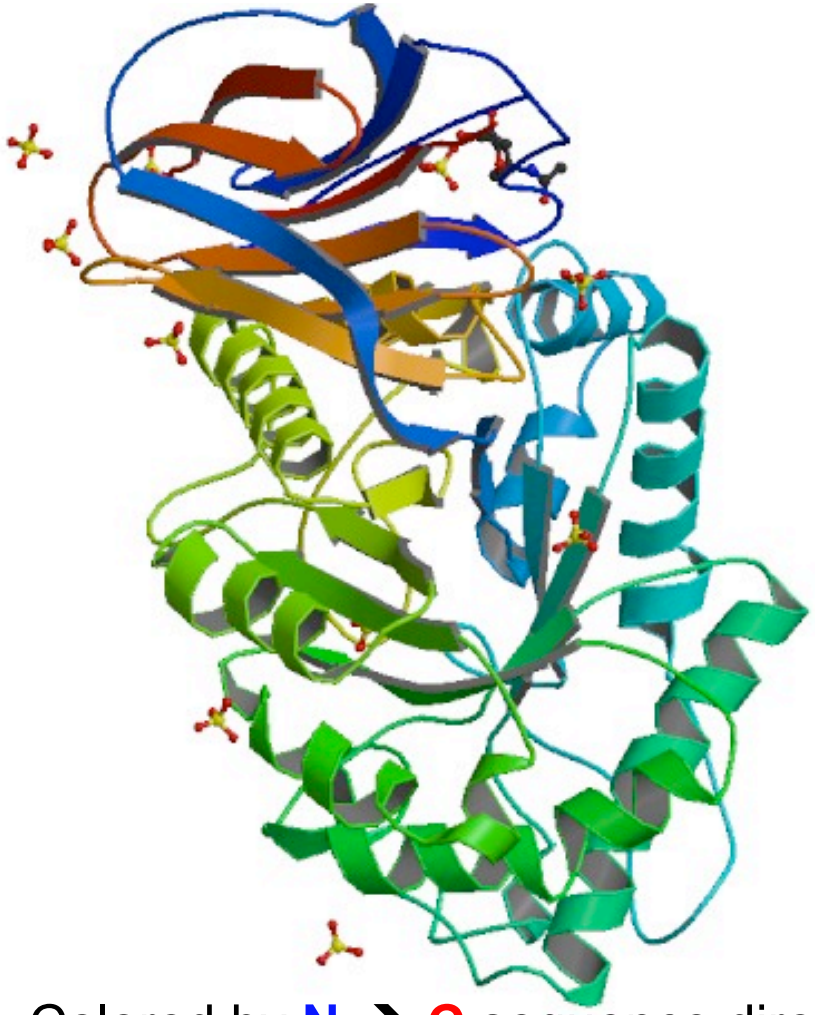


blue → green → yellow → orange → red
N-terminus **C-terminus**

http://en.wikipedia.org/wiki/Triose_phosphate_isomerase

Human acid-β-glucosidase

Bernhard Rupp's generic protein molecular structure in his various crystal structure illustrations



Colored by **N** ➔ **C** sequence direction

PDB ID **1ogs**

$C\ 2\ 2\ 2_1$

a = 107.74 Å α = 90°

b = 285.23 β = 90

c = 91.68 γ = 90

Z = 16

~ 500 aa

~ 4000 protein atoms

$M_r \approx 56$ kDa

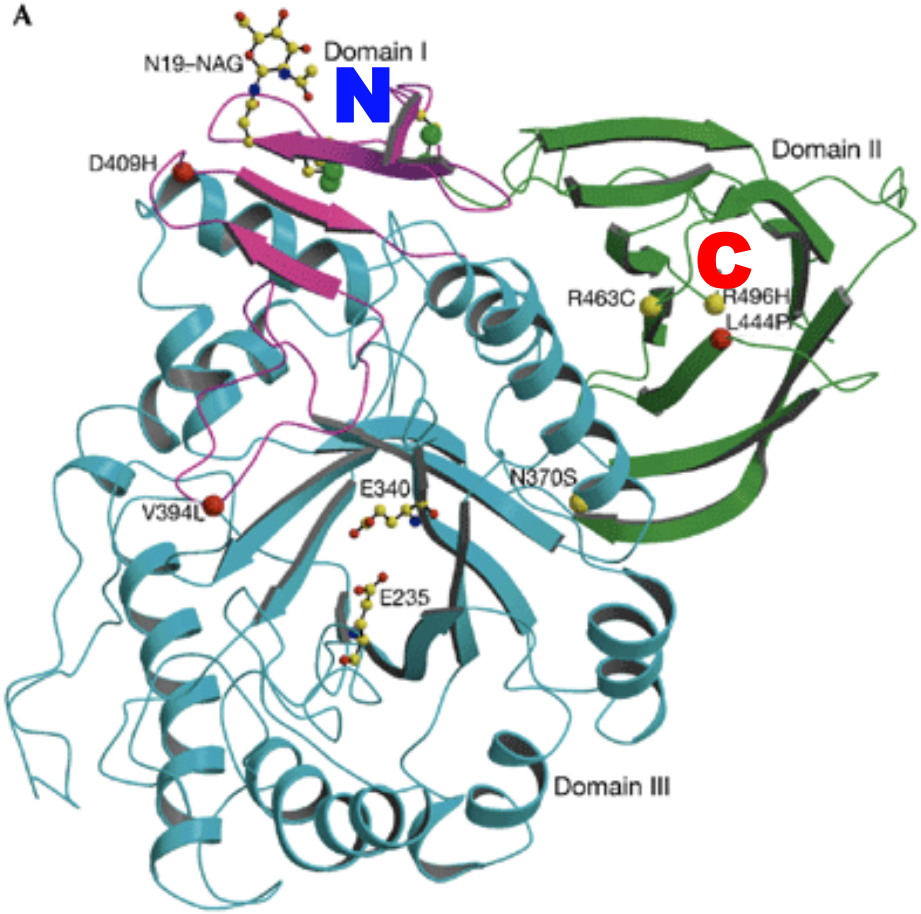
$V_M = 2.9$ Å^3 Da^{-1}

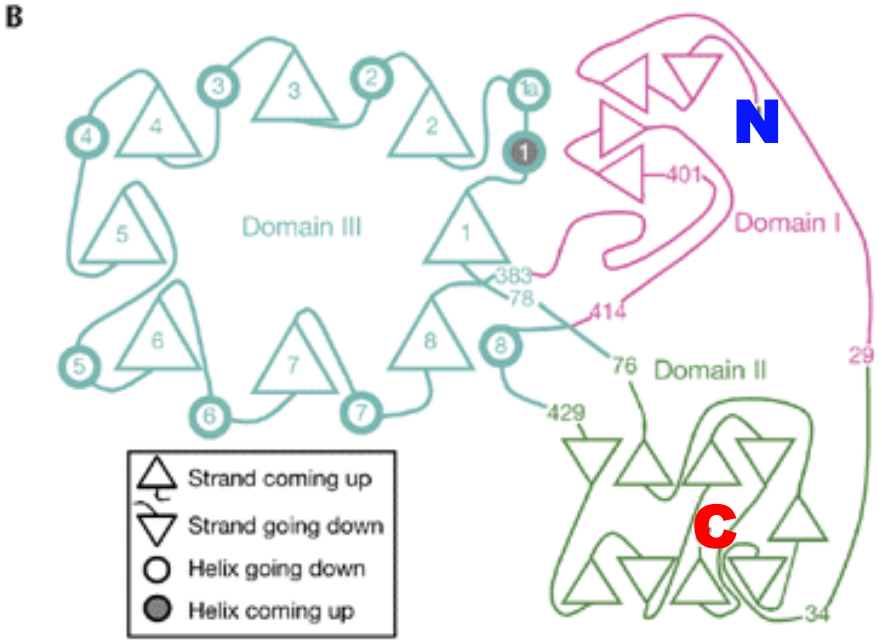
$V_s = 0.58$

Gaucher's disease results from accumulations of glucosylceramide due to inherited acid-β-glucosidase mutations.

Human acid-β-glucosidase (PDB ID 1ogs)

Bernhard Rupp's generic protein molecular structure in his various crystal structure illustrations





Colored by domains I, II, III

Domain III is a TIM barrel structure

http://embor.embopress.org/content/4/7/704

What is a crystal – formal view

Definition: A 3-dimensional translationally periodic stacking of unit cells

A very useful mathematical concept for computation but it ignores the minutiae of protein self assembly and details of crystal contacts.

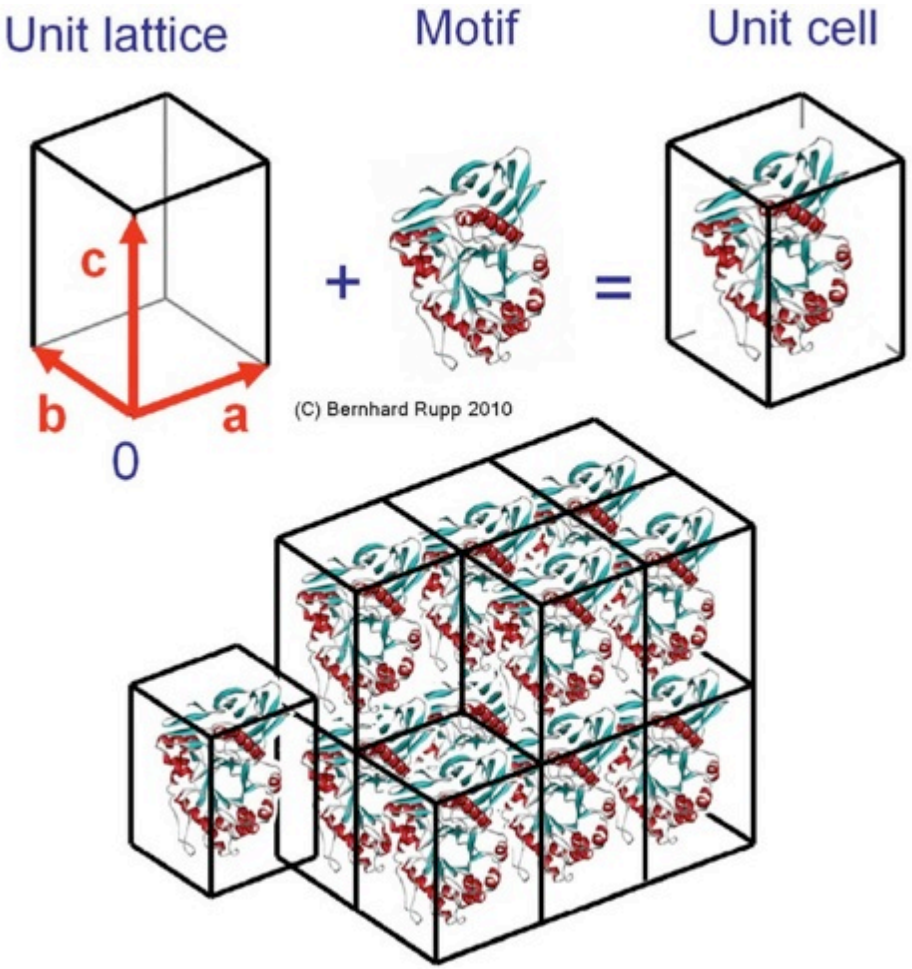


Figure 5-24 Assembly of a primitive triclinic 3-dimensional crystal from unit cells. In analogy to the 2-dimensional case, the unit lattice is filled with a motif, and the crystal is built from translationally stacked unit cells. The basis vectors form a right-handed system [0, **a**, **b**, **c**].

What is protein crystal – biocrystallization view

Protein-protein contacts are mediated by weak and sparse, non-bonded intermolecular interactions

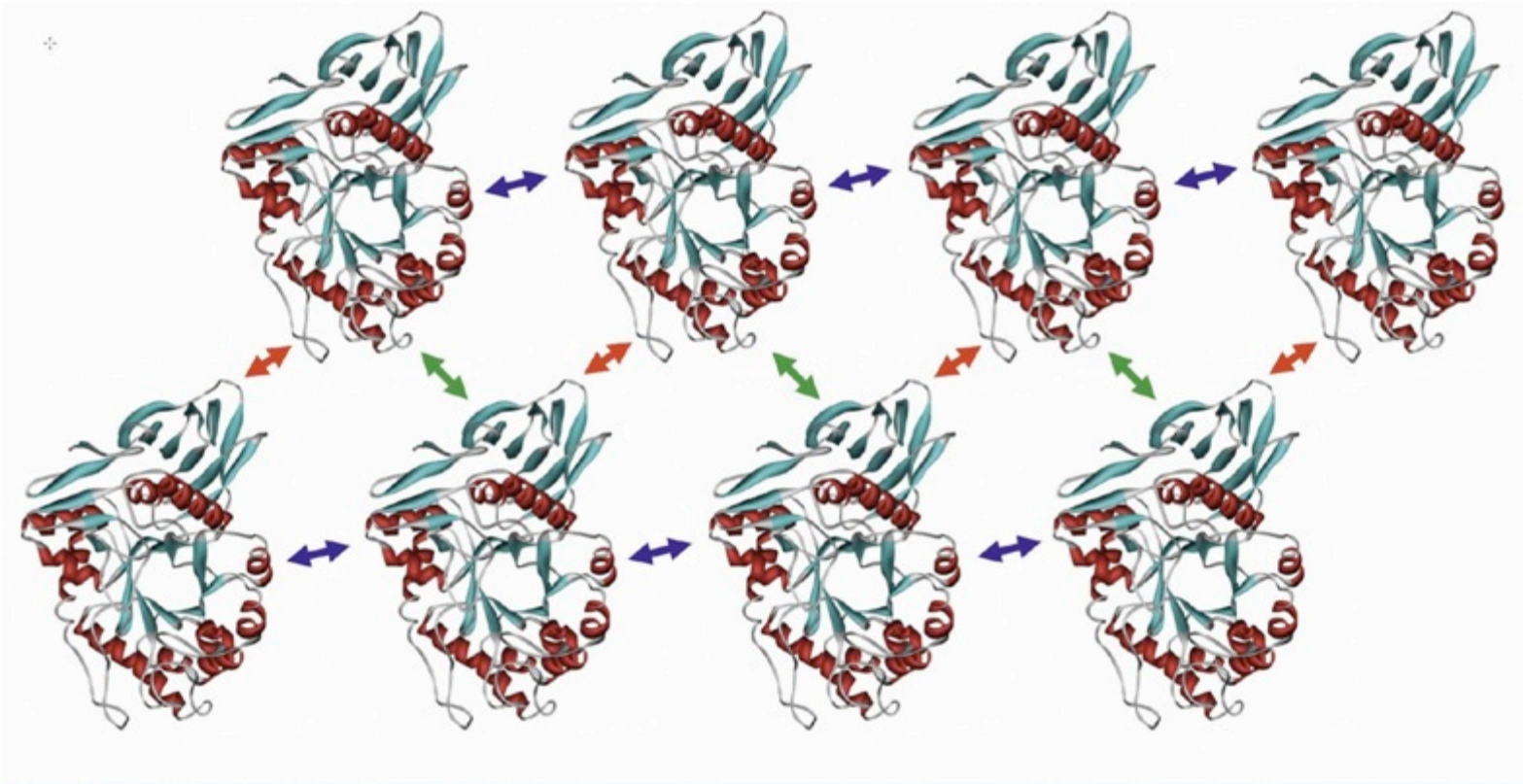


Figure 3-5 Protein crystals are formed by a sparse network of weak intermolecular interactions. The example shows protein molecules assembled into a primitive 2-dimensional lattice, connected by three different types (red, green, blue) of periodically repeating intermolecular interactions. The interactions are both sparse and weak, and as a consequence protein crystals are fragile and sensitive to mechanical stress and environmental changes.

Crystal as network of protein molecules

Note that also all of the three different intermolecular interactions repeat periodically – a few wrong molecules can disrupt the lattice

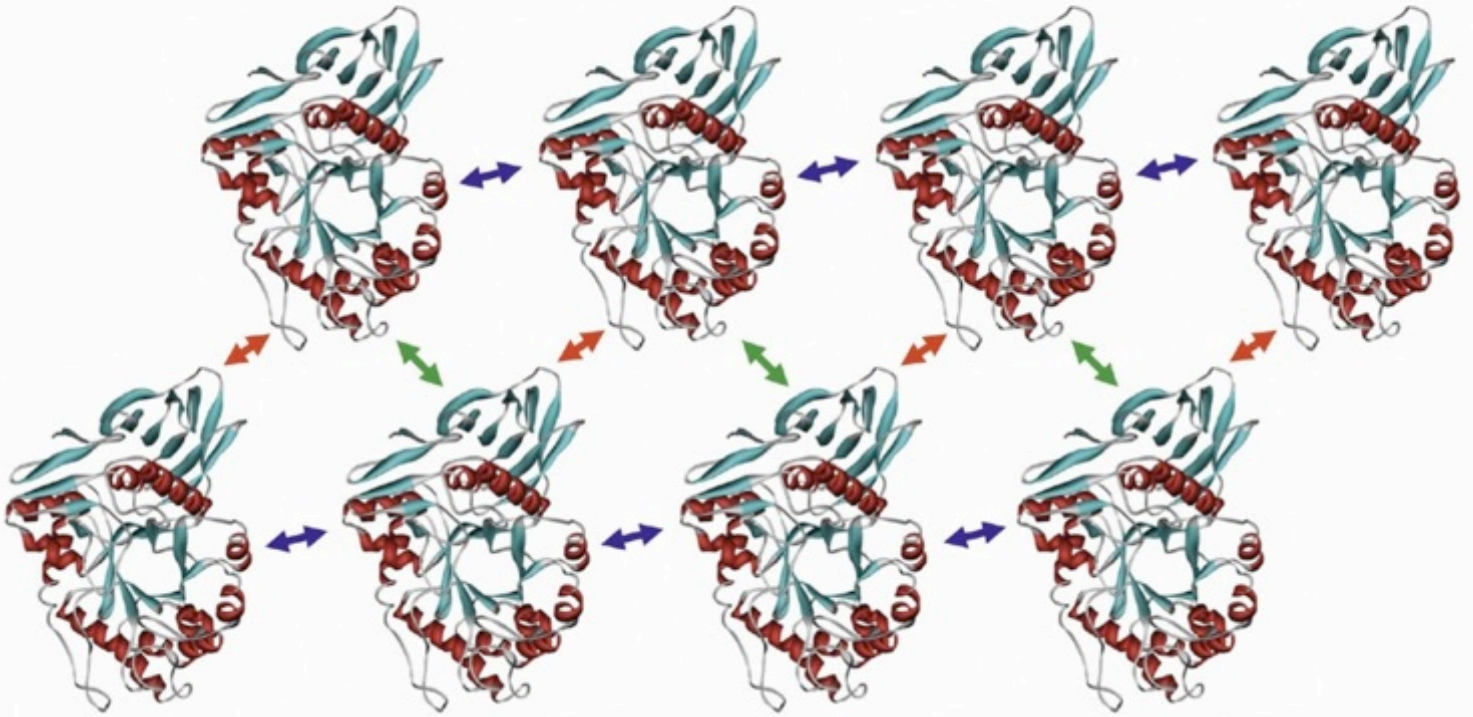
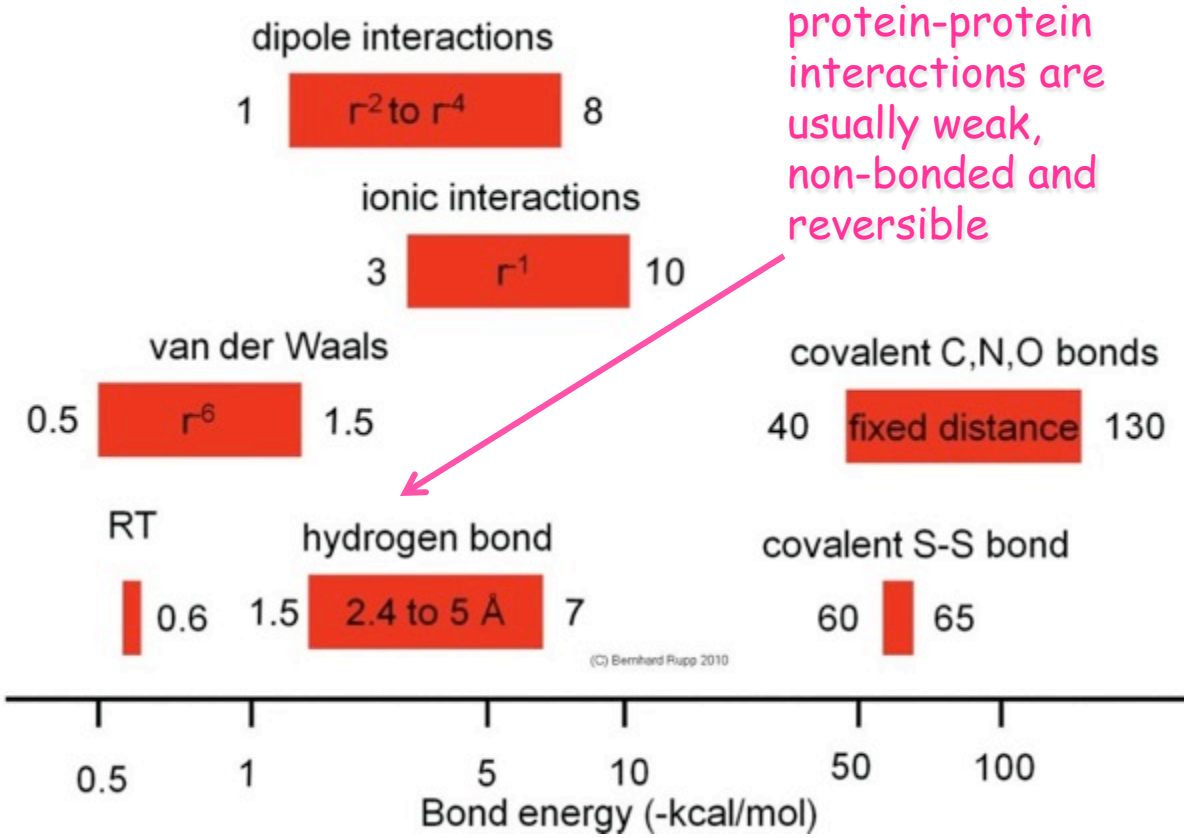


Figure 5-5 Two-dimensional "crystal" in primitive plane group *p*1. Three different types of periodically repeating intermolecular interactions form the crystal contacts. Note that only a few contacts connect our molecules, explaining the general fragility of protein crystals.

Energy range of non-bonded interactions

Figure 2-29 Typical ranges for bond energies of side chain interactions. RT denotes the thermal energy at room temperature (293 K). Note the logarithmic energy scale. The numbers in the red boxes give the approximate radial distance dependence for charged and polar interactions, and an approximate interaction range for the directionally dependent hydrogen bonds. The numbers left and right of the red boxes flank the approximate bond energy range. In contrast to the weak non-covalent interactions, covalent bonds have specific and discrete bond distances and bond angles.



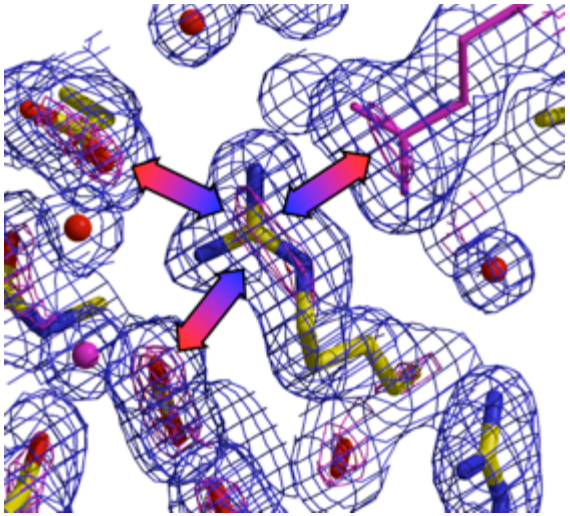
protein-protein interactions are usually weak, non-bonded and reversible

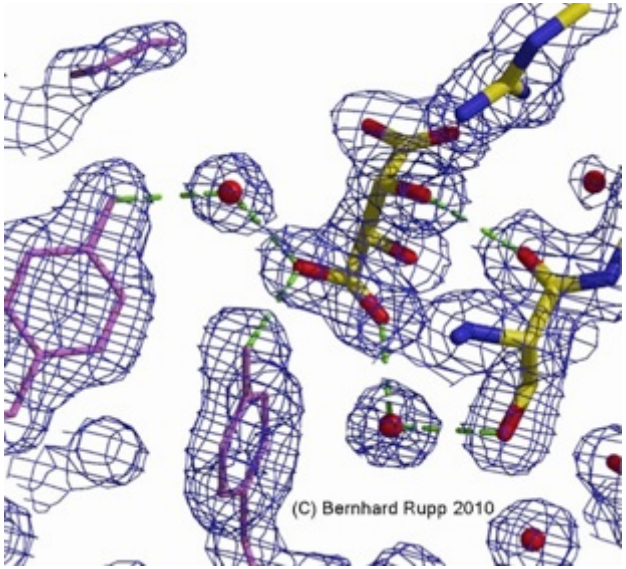
Conclusion: Small changes in environment can have significant impact on protein crystallization

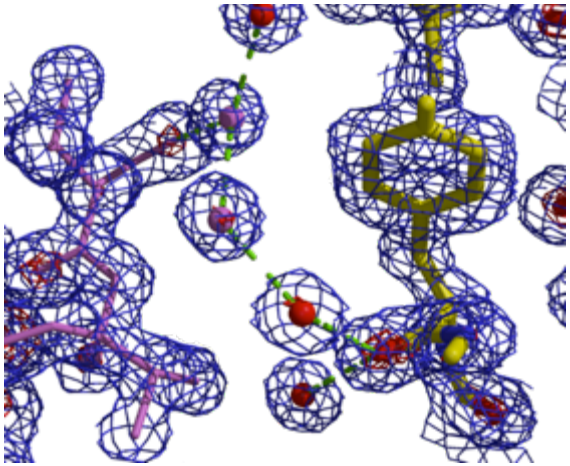
Interactions between molecules

Preciously few weak interactions that must be in the right place for self-assembly into in a fragile protein crystal

- Hydrogen bonds
- Salt bridges (charged interactions)
- Hydrophobic contacts
- VdW contacts
- Solvent mediated







What is a crystal – formal view

Definition: A 3-dimensional translationally periodic stacking of unit cells

A very useful mathematical concept for computation but it ignores the minutiae of protein self assembly and details of crystal contacts.

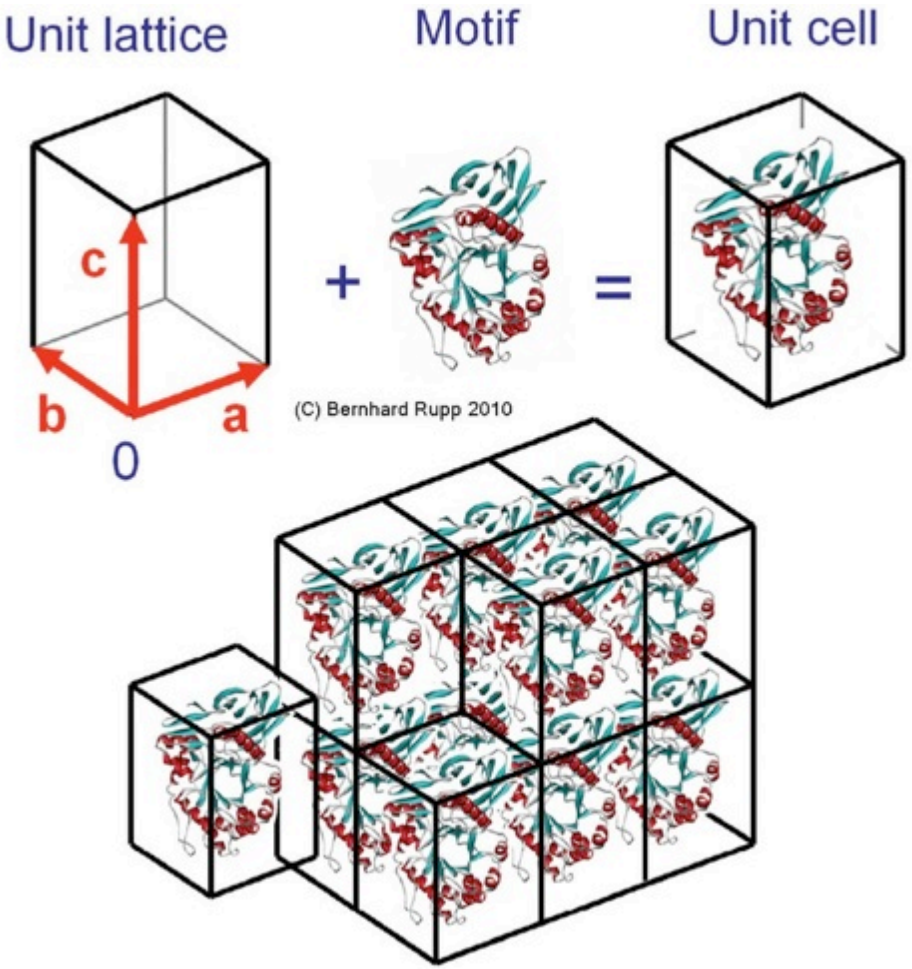


Figure 5-24 Assembly of a primitive triclinic 3-dimensional crystal from unit cells. In analogy to the 2-dimensional case, the unit lattice is filled with a motif, and the crystal is built from translationally stacked unit cells. The basis vectors form a right-handed system [0, **a**, **b**, **c**].

Plane lattices, motifs, and unit cells

Figure 5-3 Assignment and nomenclature of plane lattice vectors. A plane square unit lattice exemplifies the assignment and nomenclature of plane unit lattice vectors **a** and **b**, the corresponding scalar lattice parameters *a*, *b*, and the enclosed angle γ. Table 5-1 lists the remaining plane lattice types.

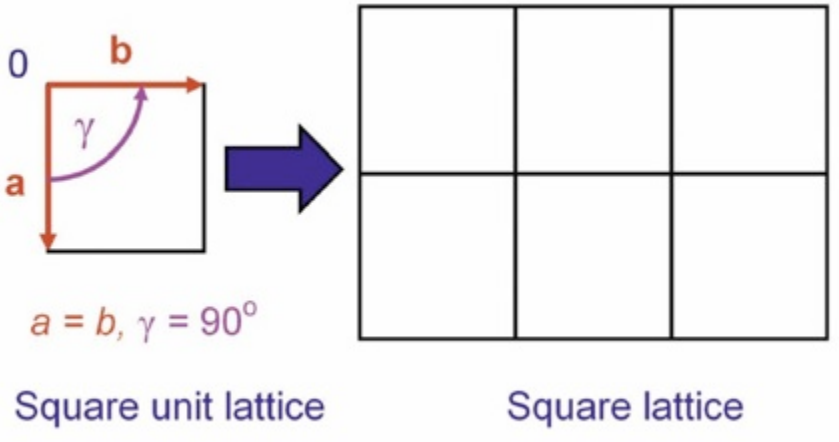
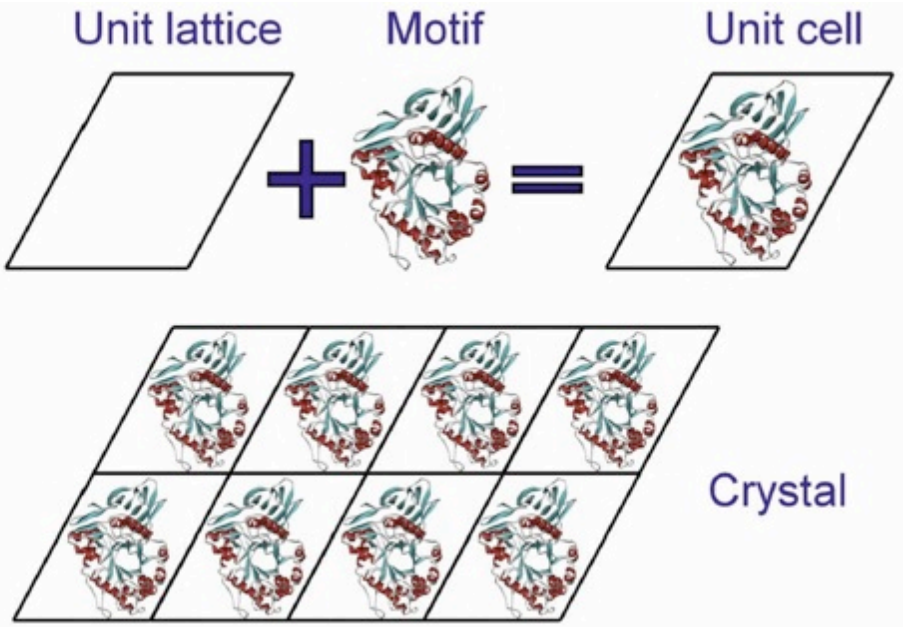


Figure 5-4 Crystals as translationally periodic arrangements of unit cells. Filling the oblique unit lattice with a motif creates an oblique unit cell. The unit cells can be stacked to form an extended, translationally periodic arrangement of unit cells—the actual crystal.



Different unit cell origins

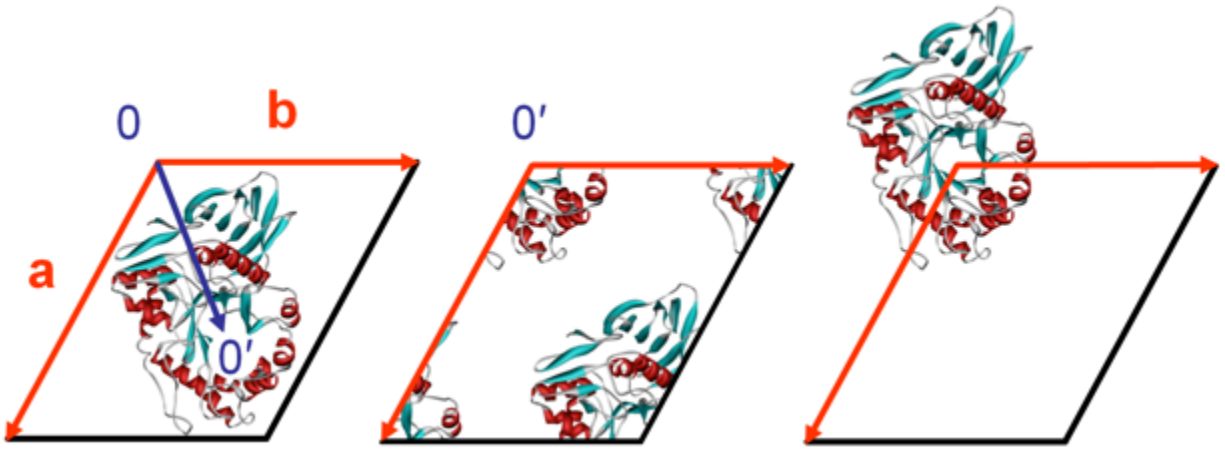
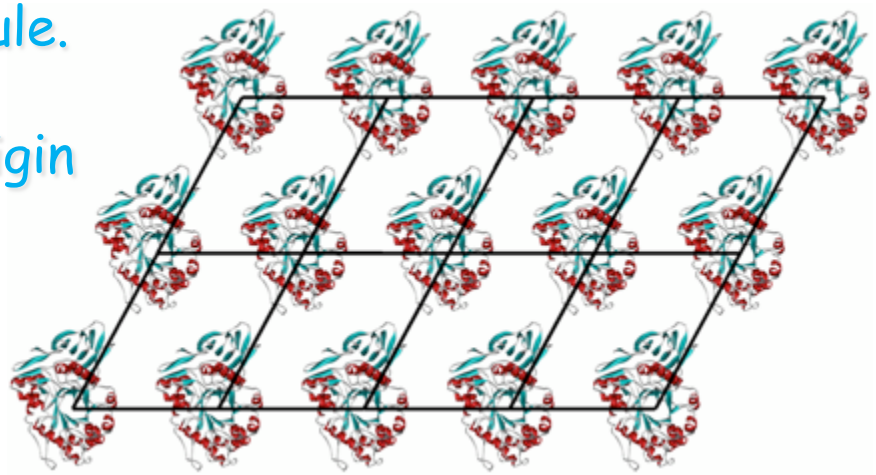
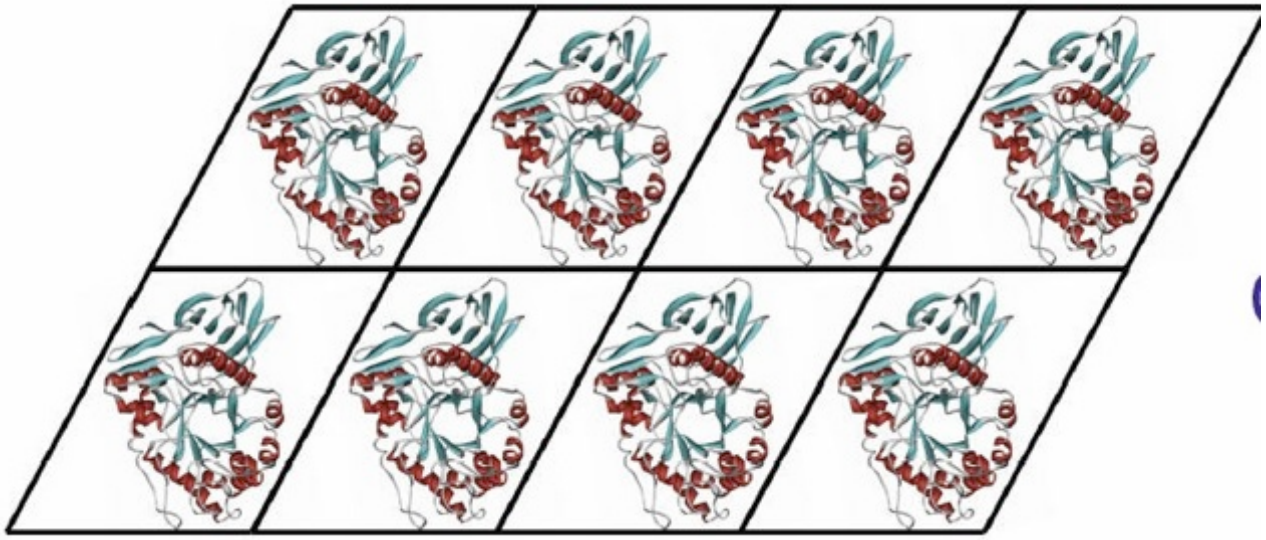


Figure 5-7 Different unit cell origins. Left: the unit cell chosen with the origin so that the whole molecule happens to fit completely within the until cell boundaries, which is rarely possible in reality. Middle: a different choice of origin, with the molecule displayed in fragments within the unit cell boundaries. The origin shift vector 00′ is indicated in blue in the left panel. Right: the same origin as in the center panel, but this time the intact molecule is displayed, preferably, but not necessarily, close to the unit cell origin. For crystallographic purposes, the three different representations of the molecule are equivalent and contain the exact same information. For ease of visualization, a representation containing an intact molecule is preferred.

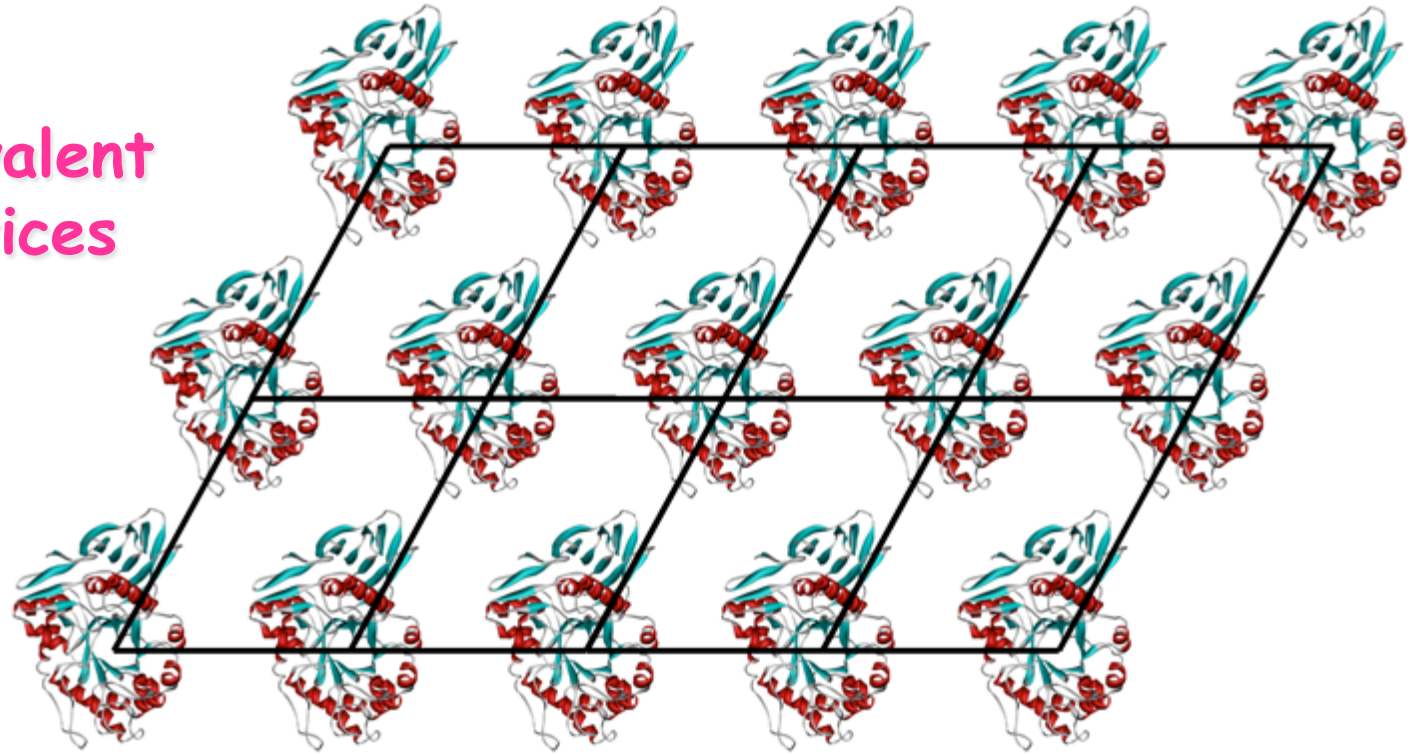
Different origin choices may be possible. Regardless on which origin we place the primitive lattice, a unit cell will always contain the same 'amount' of molecule. Some choices are more handy than others. Only in primitive p1 is the origin choice entirely arbitrary.



Lattice origin shift



Equivalent lattices



Symmetry in the unit cell: 2-fold rotation

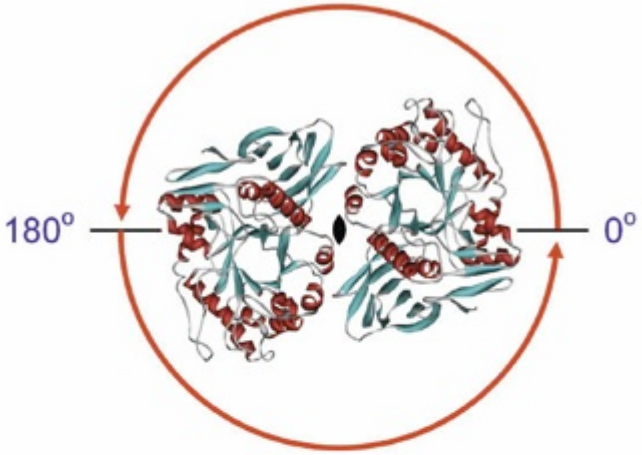


Figure 5-8 Two-fold rotation operation applied to a molecular motif. The 2-fold rotation axis is perpendicular to the paper plane, and its location is depicted by the black dyad symbol (⧫). By definition, rotations are applied counterclockwise.

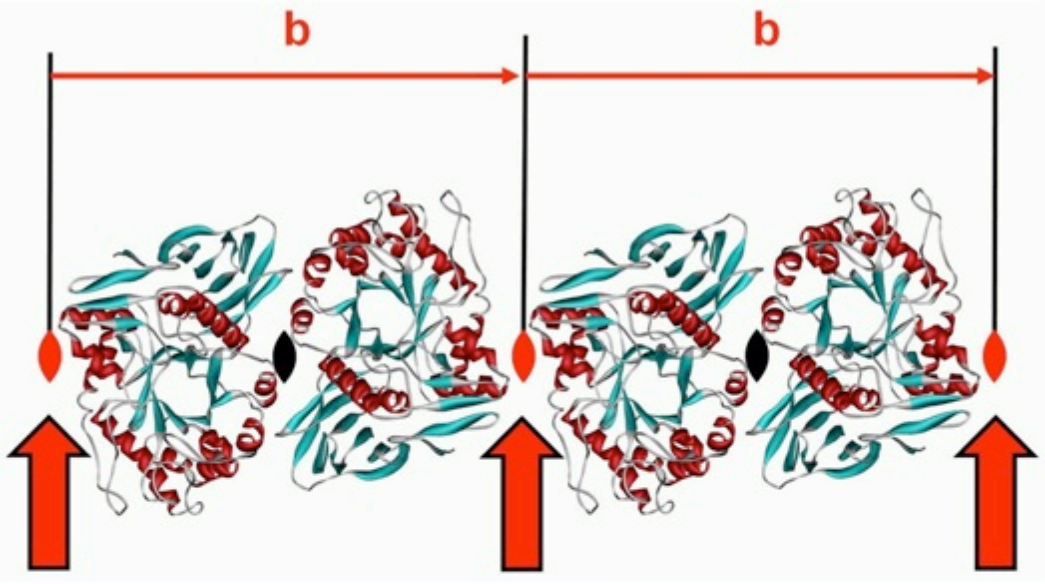


Figure 5-9 Translation of unit cell contents. The translational arrangement of the molecules related by a 2-fold axis perpendicular to the paper plane generates additional 2-fold symmetry axes (depicted by the red dyad symbol ⧫).

Different unit cells and origins of a plane p2 crystal

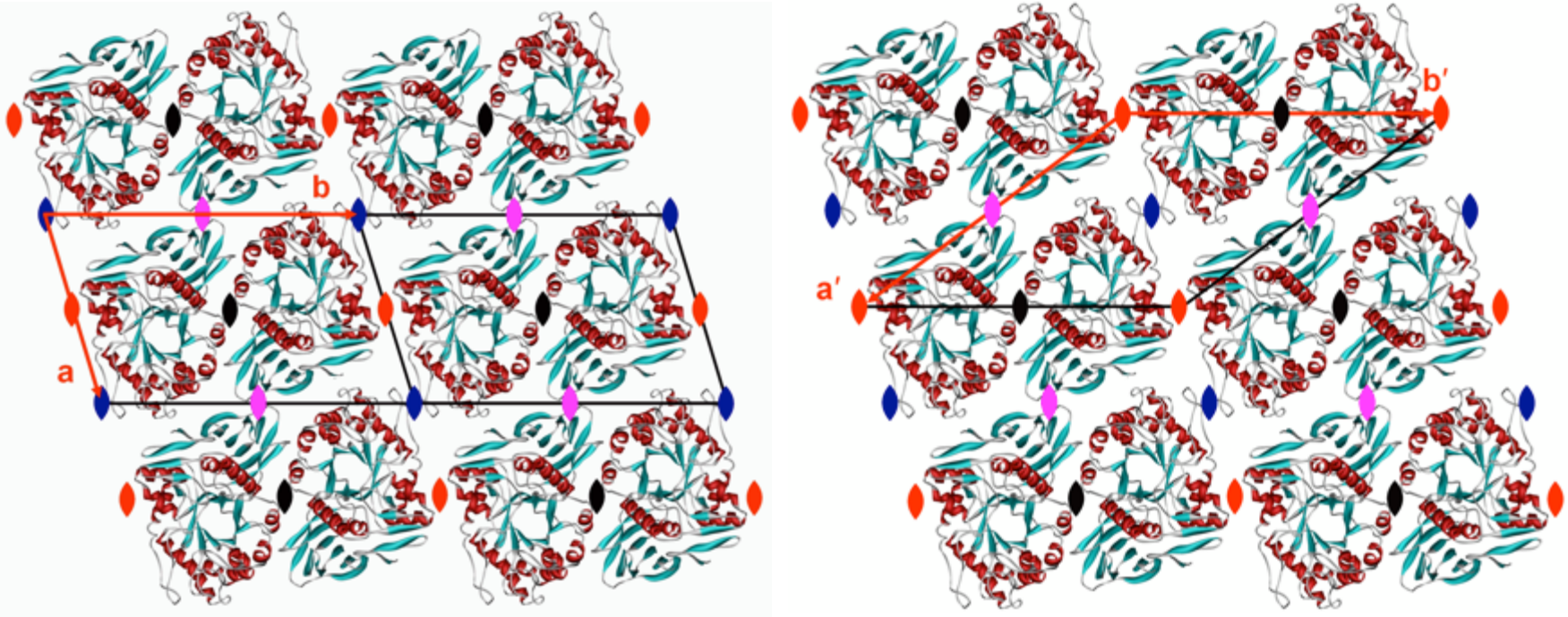


Figure 5-10 Different choices of unit cell and unit cell origins.
A crystal belonging to plane group *p*2 is superimposed with two different unit cells with different origin choices. Note that additional symmetry elements generated by the unit cell translation are located on the cell edges and corners. The crystal packing is very tight, with many intermolecular contacts and narrow solvent channels. Such an unusually well-packed crystal with low solvent content often diffracts well.[3]

The two unit cells have different origins but have the same volume. Each p2 unit cell contains two molecules.

Symmetry in the unit cell: 4-fold rotation

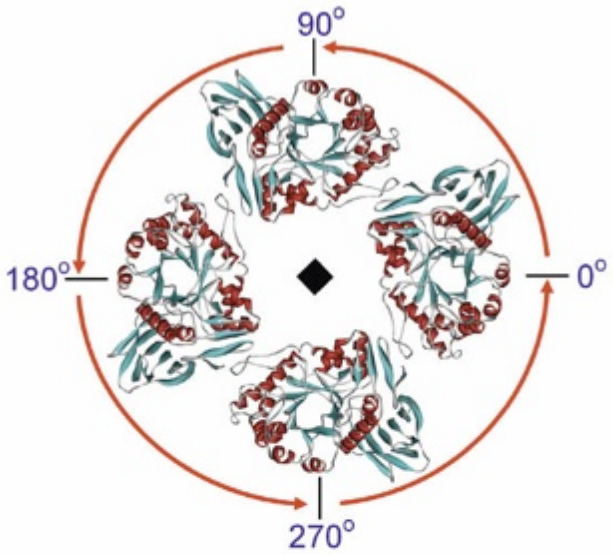


Figure 5-11 Rotation around a 4-fold rotation axis. The tetragonal unit cell is generated by rotation of the molecular motif around a 4-fold axis, depicted by the tetrad symbol (◆). Note that the 4-fold operation has generated a molecular assembly with a distinct solvent channel in the center in the direction of the 4-fold axis.

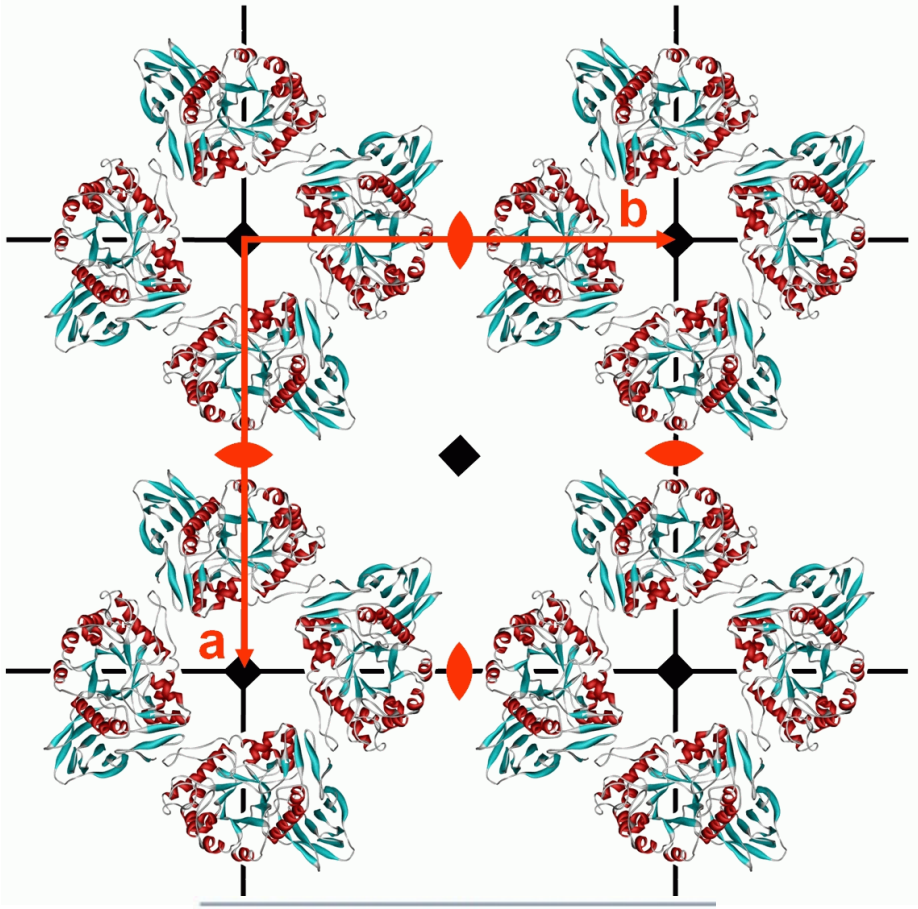


Figure 5-12 A *p*4 plane structure. Additional new 2-fold (⧫) and 4-fold (◆) rotation axes are created by the unit cell translations. The structure has extensive solvent channels, quite typical for crystal structures with high order rotation axes.

The asymmetric unit suffices to reconstruct the unit cell

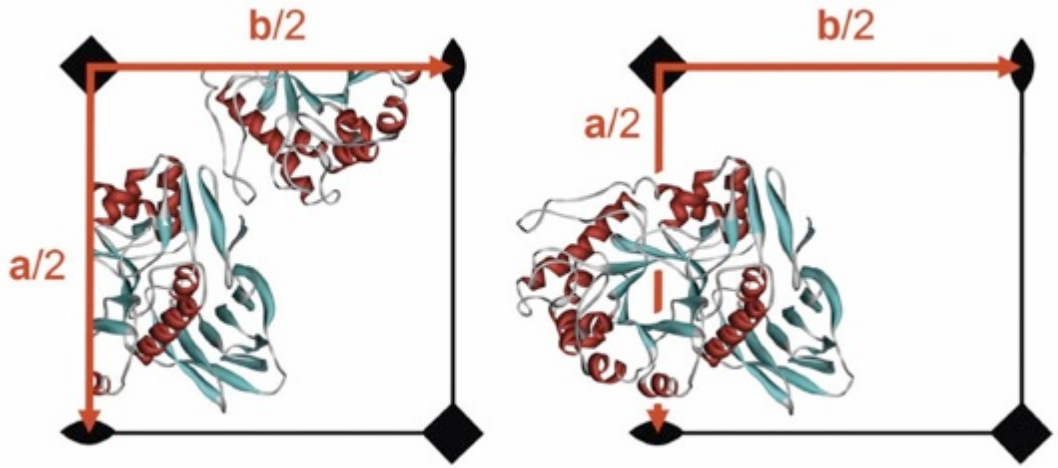
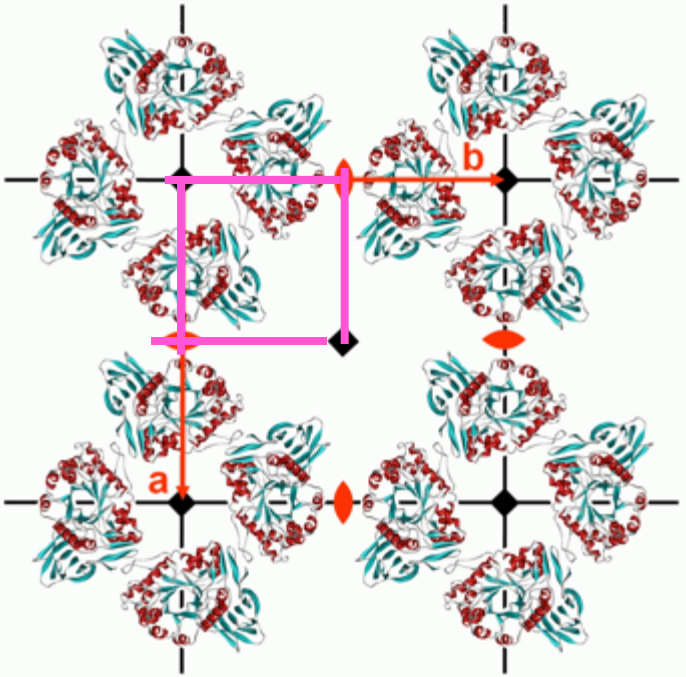


Figure 5-13 Asymmetric unit of the *p*4 structure. The asymmetric unit of *p*4 covers one fourth of the unit cell. The asymmetric unit to the left would be ideal for producing a tile (or in crystallographic computations of the unit cell contents), but the representation to the right is much better suited for displaying the molecule.

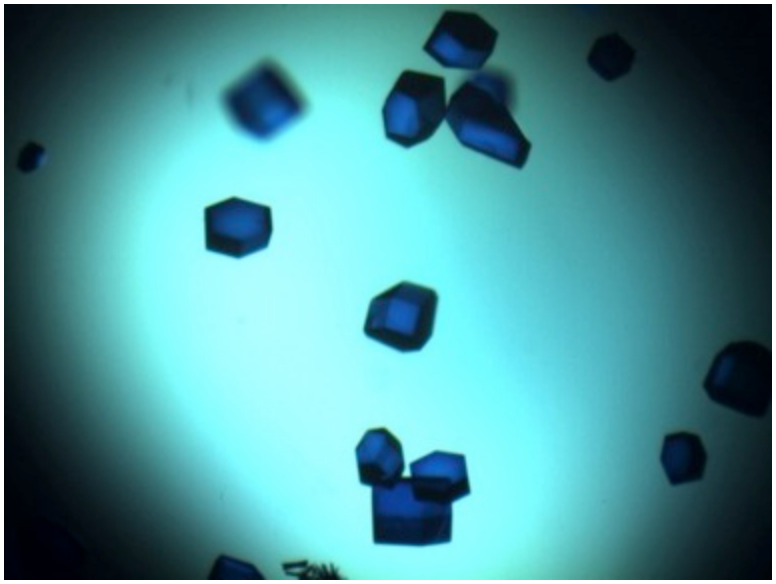
Biomolecular Crystallography B.Rupp (2010)

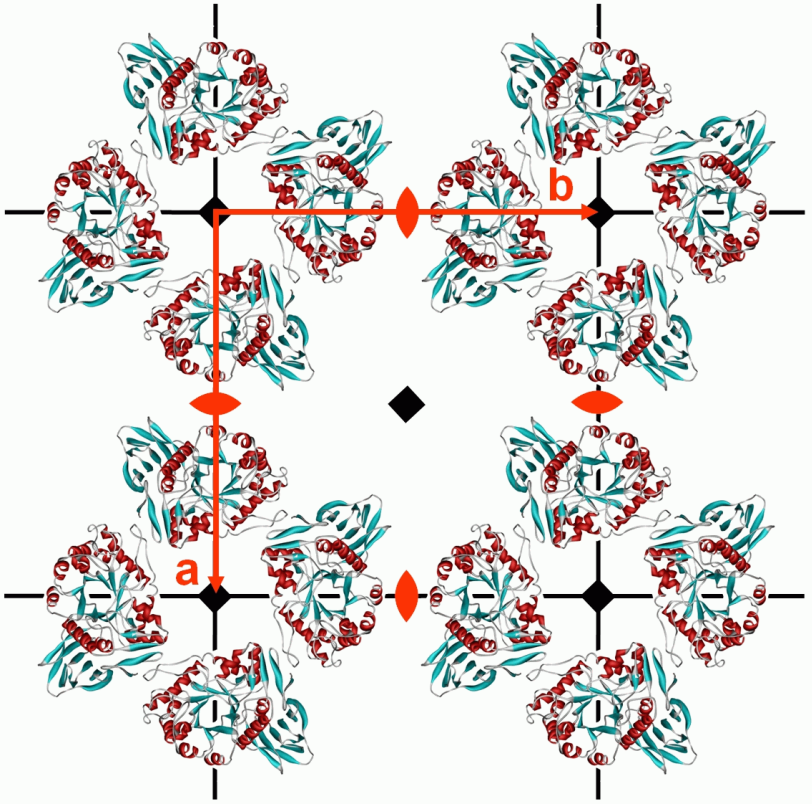


The contents of the asymmetric unit together with the unit cell symmetry allows reconstruction of the entire unit cell (and the entire crystal structure).

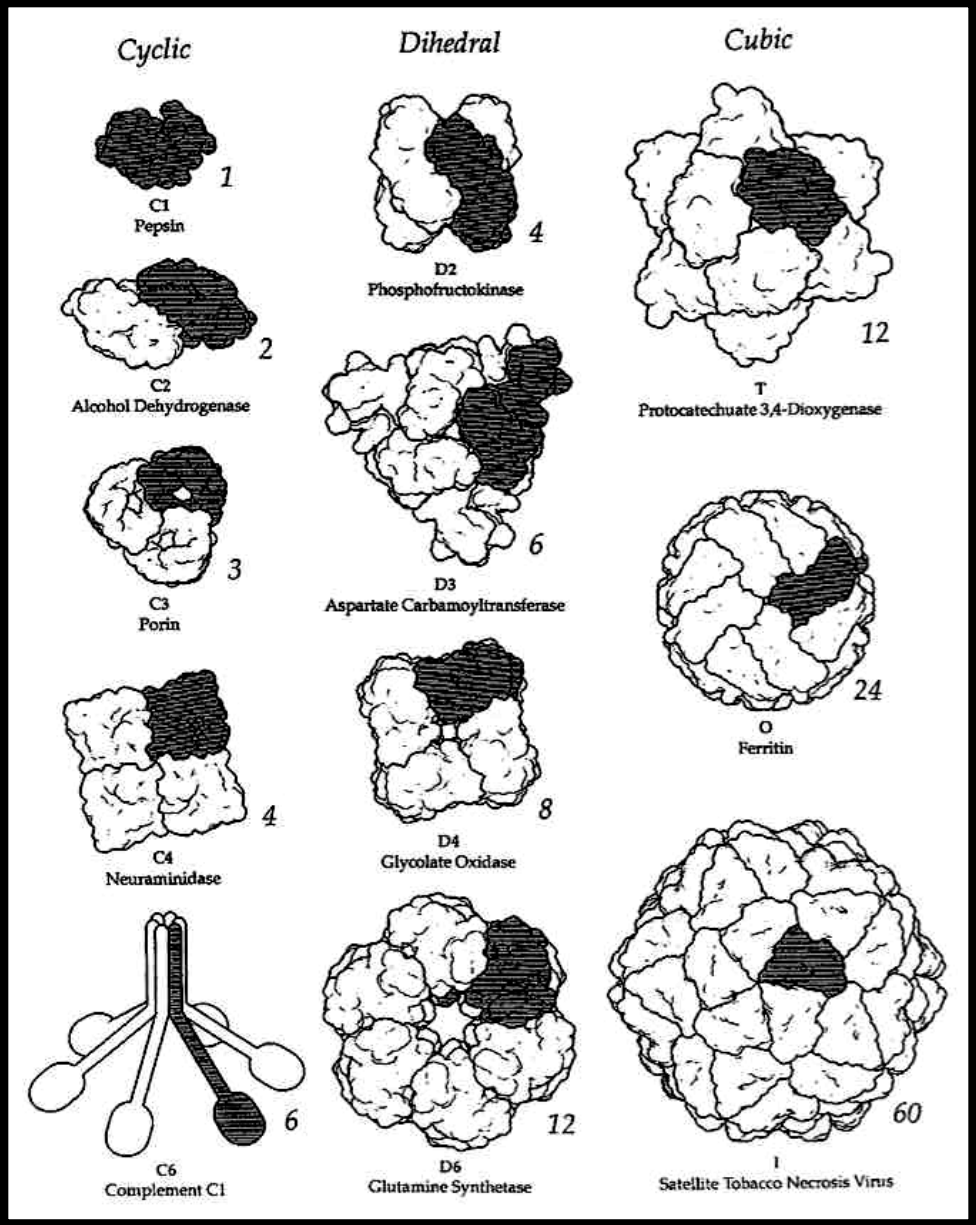
Consequences of a crystal being a network of sparse, weak, and specific interactions

- Sensitive to mechanical stress
- Sensitive to environmental changes - ΔT, ΔpH, Δμ
- Contain large fraction of solvent
- Contain solvent channels
  important for ligand soaking





Crystallographic point groups of protein homomultimers



Goodsell & Olsen (2000). *Ann. Rev. Biophys. Biomol. Struct.* 29, 105.

What is a 'stable' protein crystal ?

A) How many contacts and B) how strong ?

A) ~ 15 contacts/molecule
B) contact surface area 100-500 $Å^2$
obligate dimers: ~ 800 $Å^2$ and up
in between gray area

Free energy of crystallization

$$\Delta G_c = \Delta H_c - T(\Delta S_{protein} + \Delta S_{solvent})$$

Not much — Decisive term

Crystallization is strongly entropy driven !
rationale for surface (entropy) engineering

Protein solubility and solubility diagrams

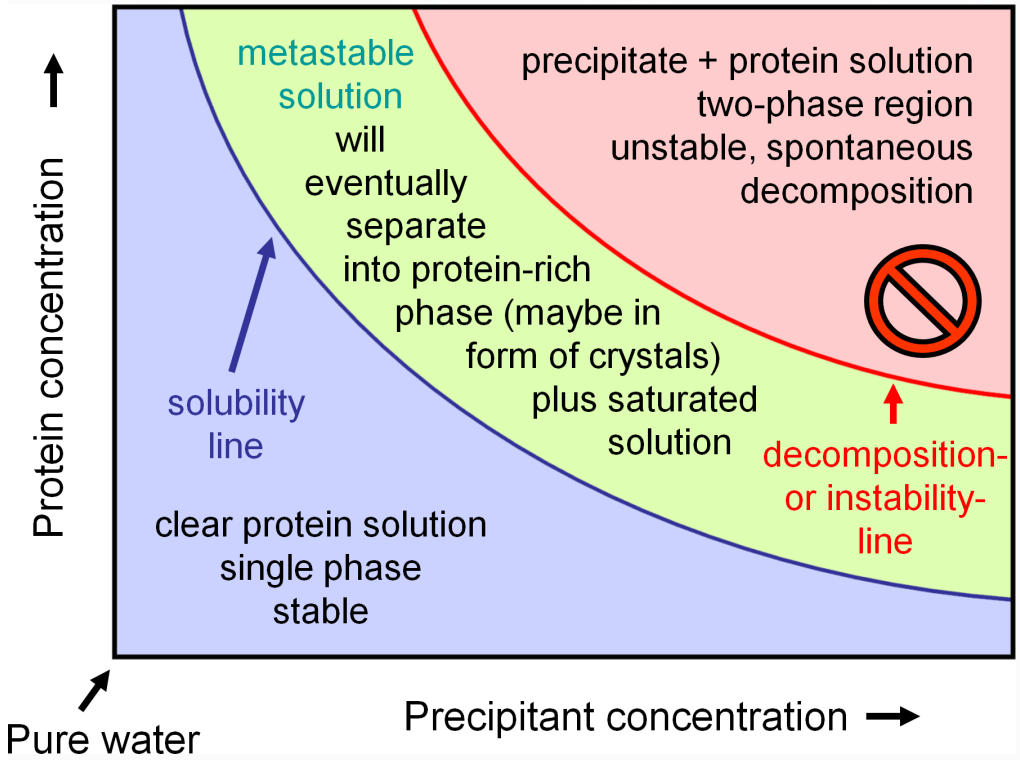


Figure 3-7 A basic solubility phase diagram for a given temperature. The diagram visualizes the general observation that the higher the precipitant concentration in the solution, the lower the maximal achievable protein concentration in the solution and *vice versa*. Between the solubility line and the decomposition line lies the metastable region representing the supersaturated protein solution, which will eventually—given the necessary kinetic nucleation events—equilibrate and separate into a protein-rich phase (often in the form of precipitate or crystals) and saturated protein solution.

Protein crystallization is a special case of phase separation from thermodynamically metastable (supersaturated) solution under the control of kinetic parameters

Thermodynamics determine whether it can happen, kinetics whether it does actually happen

Protein solubility and nucleation regimes

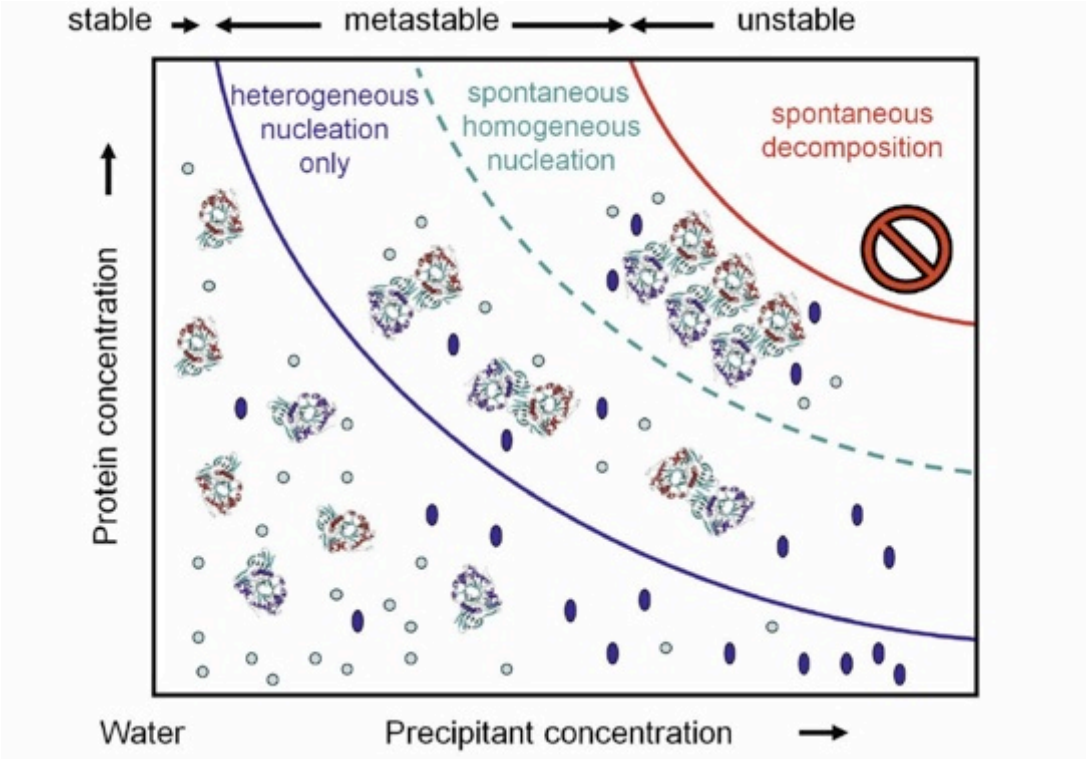


Figure 3-10 The location of nucleation zones in a protein crystallization diagram. Precipitant molecules are represented as dark blue ovals and water molecules as light blue circles. As a rule of thumb, higher supersaturation is necessary for spontaneous formation of stable crystallization nuclei (homogeneous nucleation), while at low supersaturation nucleation requires external seeds in the form of microcrystals or other particulate matter (heterogeneous nucleation). The zone of homogeneous (spontaneous) nucleation is occasionally referred to as the "labile" zone, but this term should be avoided because it leads to confusion with the different concept of labile or neutral equilibrium used in physical chemistry and thermodynamics.

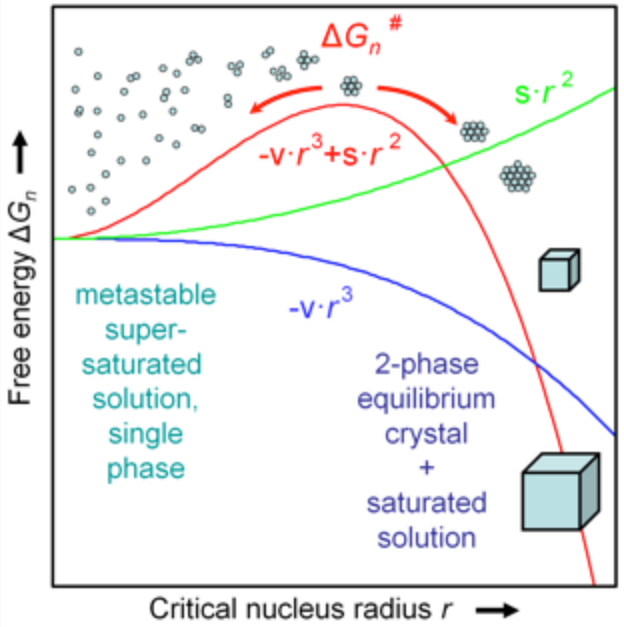


Figure 3-9 Nucleation energy. To achieve crystallization, nucleation must overcome the kinetic barrier that exists for phase separation (crystallization) from the metastable solution (left). At the critical size, it is equally likely for the nucleus (symbolized by the small aggregate) to fall apart again (left red arrow) or to grow into a crystal (right red arrow). Once a nucleus above a critical size defined by the critical free energy of nucleation $\Delta G_n^{\#}$ is formed, additional gain of binding enthalpy overcomes entropic loss during crystal growth, and the system can proceed towards its 2-phase equilibrium state (right side of image).

Protein crystals are not perfect inside

Figure 3-11 Atomic force microscope images of crystal growth. (Panel A) The atomic force microscope images of the 001 surface of glucose isomerase show the two most common growth patterns observed in crystal growth: step growth starting from 2-dimensional nucleation islands (A, left image) and a double-spiral growth pattern (A, right image). Panel B shows formation of supercritical 2-dimensional nuclei on the 001 surface of cytomegalovirus (CMV), a member of the herpes virus family. As indicated by the arrows, in this case only two virions (B, left image) suffice to generate a critical nucleus from which new step growth commences (B, right image). Images courtesy of Alexander McPherson and Aaron Greenwood, University of California, Irvine.

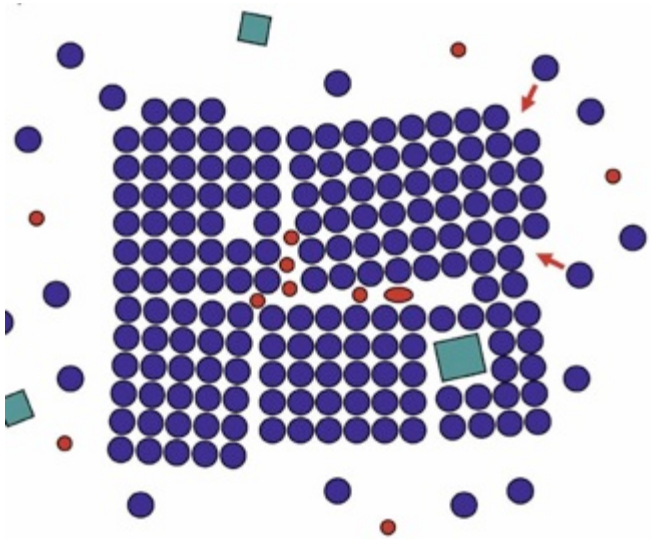


Figure 3-12 Growth of a real mosaic crystal. The schematic drawing shows a crystal growing in a solution of protein molecules (blue spheres). Small impurities (red) and some larger detritus (green squares) are also present in the solution. New molecules attach preferentially to steps and edges (red arrows) and we can recognize a growth defect in the form of a hole; impurities are enclosed at the domain boundaries; and a larger piece of detritus is incorporated at a domain boundary. Individual domains can be substantially misaligned, in this case about 6°; such a highly mosaic crystal would not be useful for diffraction experiments.

Phenomena of mosaicity and twinning complicate data collection