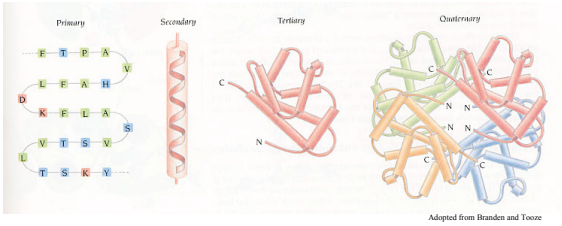
BINF 731

Protein Structure Analysis

Iosif Vaisman

2004

Protein Structure Hierarchy



- Primary - the sequence of amino acid residues
- Secondary - ordered regions of primary sequence (helices, beta-sheets, turns)
- Tertiary - the three-dimensional fold of a protein subunit
- Quaternary - the arrangement of subunits in oligomers.

Anfinsen's Dogma

Three-dimensional structure of a protein is determined solely by its amino-acid sequence.

Native conformation of the protein is the global-minimum free energy conformation.

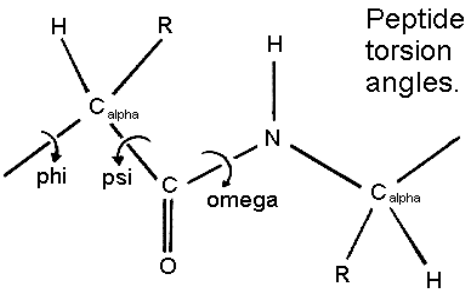
Complexity of protein structure (Levinthal paradox)

100 residue protein
3 conformations per residue

number of distinct conformations:
$3^{100} \cong 10^{48}$

sampling time $\cong 10^{30}$ years

Levinthal paradox



3 conformations per residue is a very conservative estimate

Complexity

P (Polynomial)
complexity class of decision problems for which execution time of a computation is no more than a polynomial function of the problem size

NP (Nondeterministic Polynomial)
complexity class of decision problems for which answers can be checked by an algorithm whose run time is polynomial in the size of the input

Protein Folding Problem

Given: sequence
Find: structure

The problem is NP-complete

Protein Folding Problem

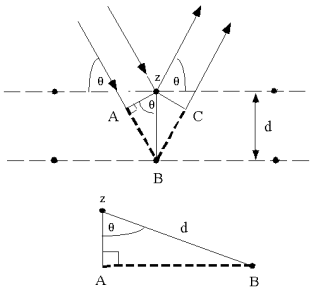
Problem for us, not for proteins.
They just fold...

(Ken Dill)

Protein Structure Determination

X-ray crystallography
NMR spectroscopy
Neutron diffraction
Electron microscopy
Atomic force microscopy

X-ray crystallography

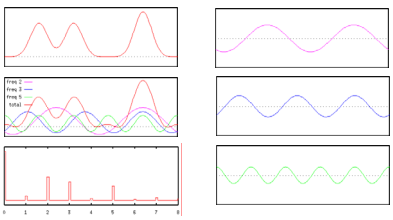


Bragg's Law

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

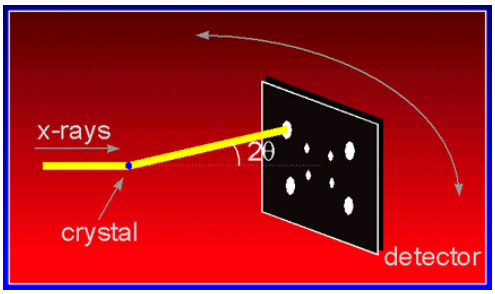
X-ray crystallography

Phase determination: MIR and MAD
(Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction)

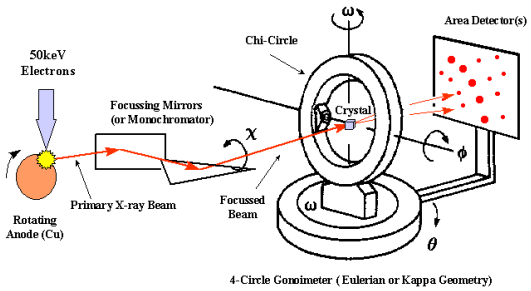


Fourier Transforms

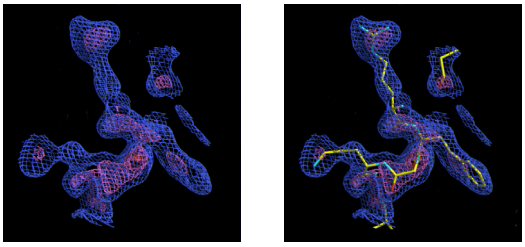
X-ray crystallography



X-ray crystallography

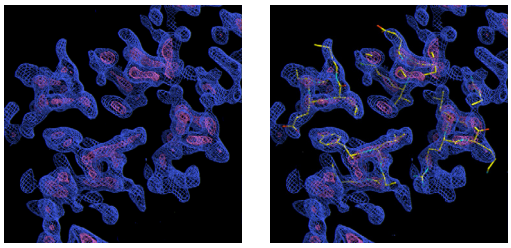


X-ray crystallography



Electron density map created from multi-wavelength data (Arg)

X-ray crystallography



Experimental electron density map and model fitting
(apoE four helix bundle)

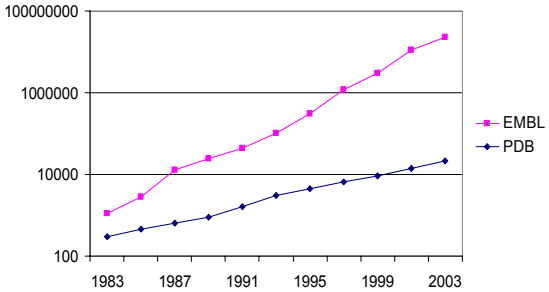
X-ray crystallography

Confidence in structural features of proteins determined by X-ray crystallography

(These are *rough* estimates, and depend strongly on the quality of the data.)

Structural feature	Resolution				
	5 Å	3 Å	2.5 Å	2.0 Å	1.5 Å
Chain tracing	—	Fair	Good	Good	Good
Secondary structure	Helices fair	Fair	Good	Good	Good
Sidechain conformations	—	—	Fair	Good	Good
Orientation of peptide planes	—	—	Fair	Good	Good
Protein hydrogen atoms visible	—	—	—	—	Good

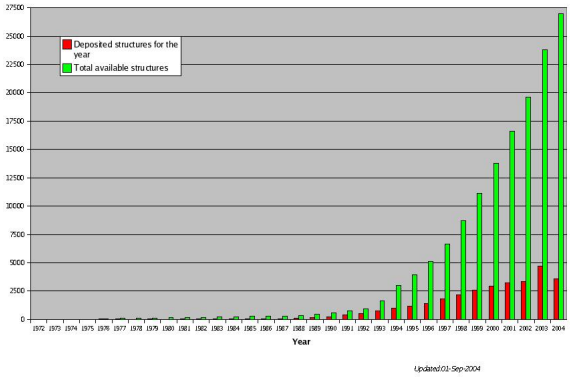
Dynamics of Database Growth



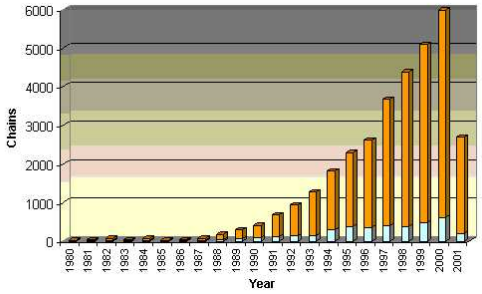
PDB Holdings List: 28-Sep-2004

	Prot	Pr/DNA	DNA	Carb	total
X-ray	21606	1067	748	14	23435
NMR	3276	103	610	4	3993
Total	24882	1170	1358	18	27428

PDB Growth

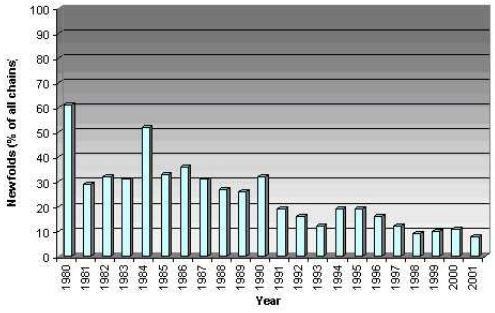


Growth of New Folds in PDB

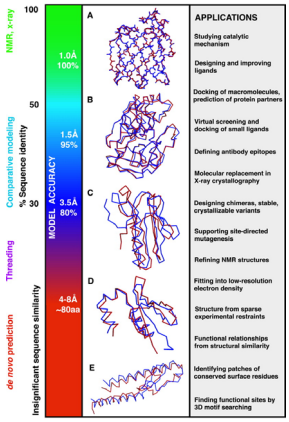


"new folds" (blue) and "old folds" (orange)

Growth of New Folds in PDB



Protein structure models



Adapted from D.Baker and A.Sali, Science, 2001, 294, 93-96