ArPS Summer School Biophysics: From Molecules to Organisms

Part 1: Molecular Biophysics

24-26 August 2024

Molecular Biophysics

- What is Life? An Historical Prelude
- Central Dogma(s) of Molecular Biology
- The Storage of Information
- The Protein Folding Problem
- Basics of Biomolecular Structure
- Biological Function of Macromolecules
- The Ligand-Perturbed Free Energy Landscape
- Graphic Visualizaton of Biomolecules

Physics in Biology: What's the Difference?

- All of the components of the cell arrange themselves to make up the cell
- There is no "guide", "life force", or "volition". The only "decisions" obey the laws of physics

Schrodinger 1942: What is Life?

- Living things are much bigger than atoms
- Living things must be ordered
- Statistical Physics assumes order arises from large numbers of small molecules
- Example: magnetism-nuclear magnetization vs. temperature

Schrodinger: What is Life?

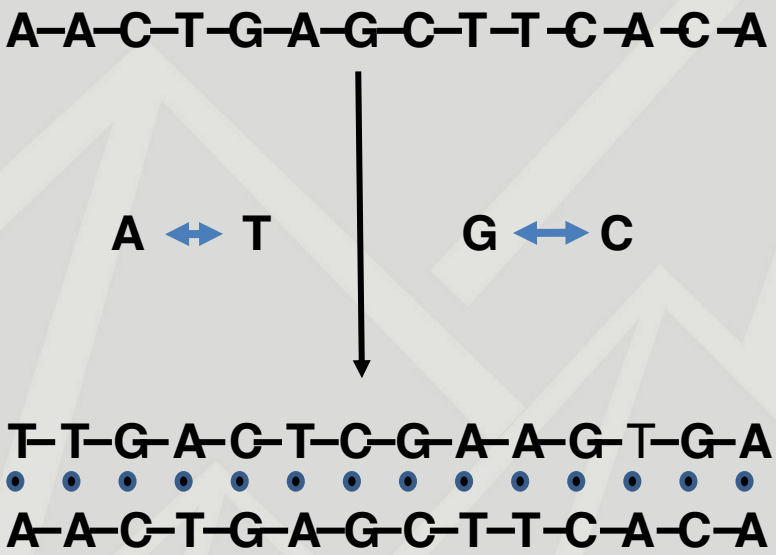
- Living things are much bigger than atoms
- Living things must be ordered
- Statistical Physics assumes order arises from large numbers of small molecules
- Genetic instructions for the cell on 2-48 chromosomes

Schrodinger: What is Life?

the most essential part of a living cell-the chromosome fibre may suitably be called an aperiodic crystal. In physics we have dealt hitherto only with periodic crystals. To a humble physicist's mind, these are very interesting and complicated objects; they constitute one of the most fascinating and complex material structures by which inanimate nature puzzles his wits. Yet, compared with the aperiodic crystal, they are rather plain and dull. The difference in structure is of the same kind as that between an ordinary wallpaper in which the same pattern is repeated again and again in regular periodicity and a masterpiece of embroidery, say a Raphael tapestry, which shows no dull repetition, but an elaborate, coherent, meaningful design traced by the great master

The Classical Central Dogma

The Mechanism of Information Transmission in Biology



RNA, a Passive “Messenger”?

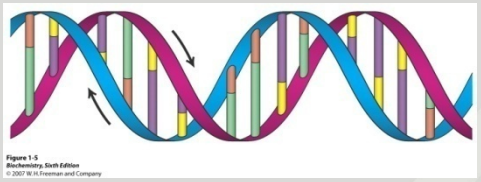
DNA

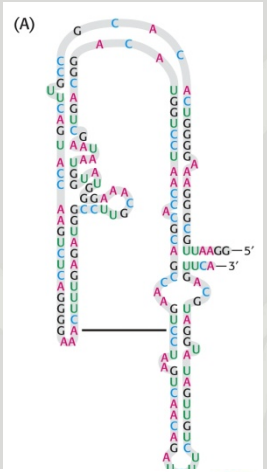
↓

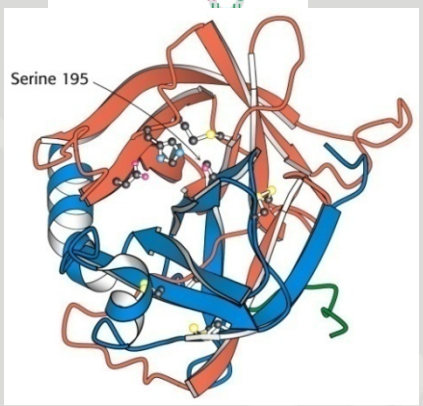
RNA

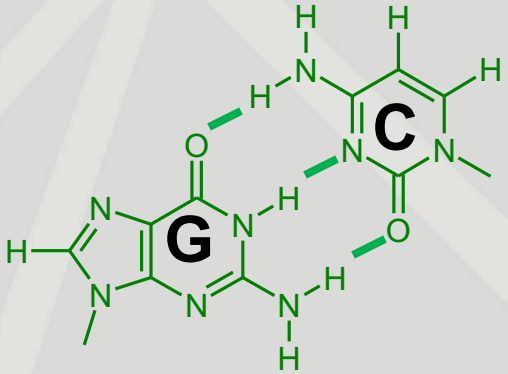
↓

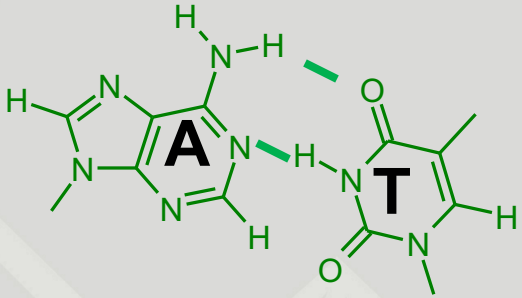
Protein

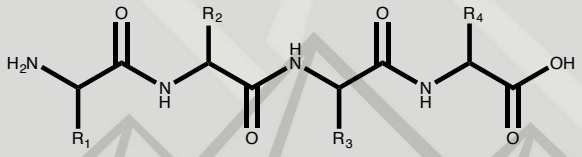




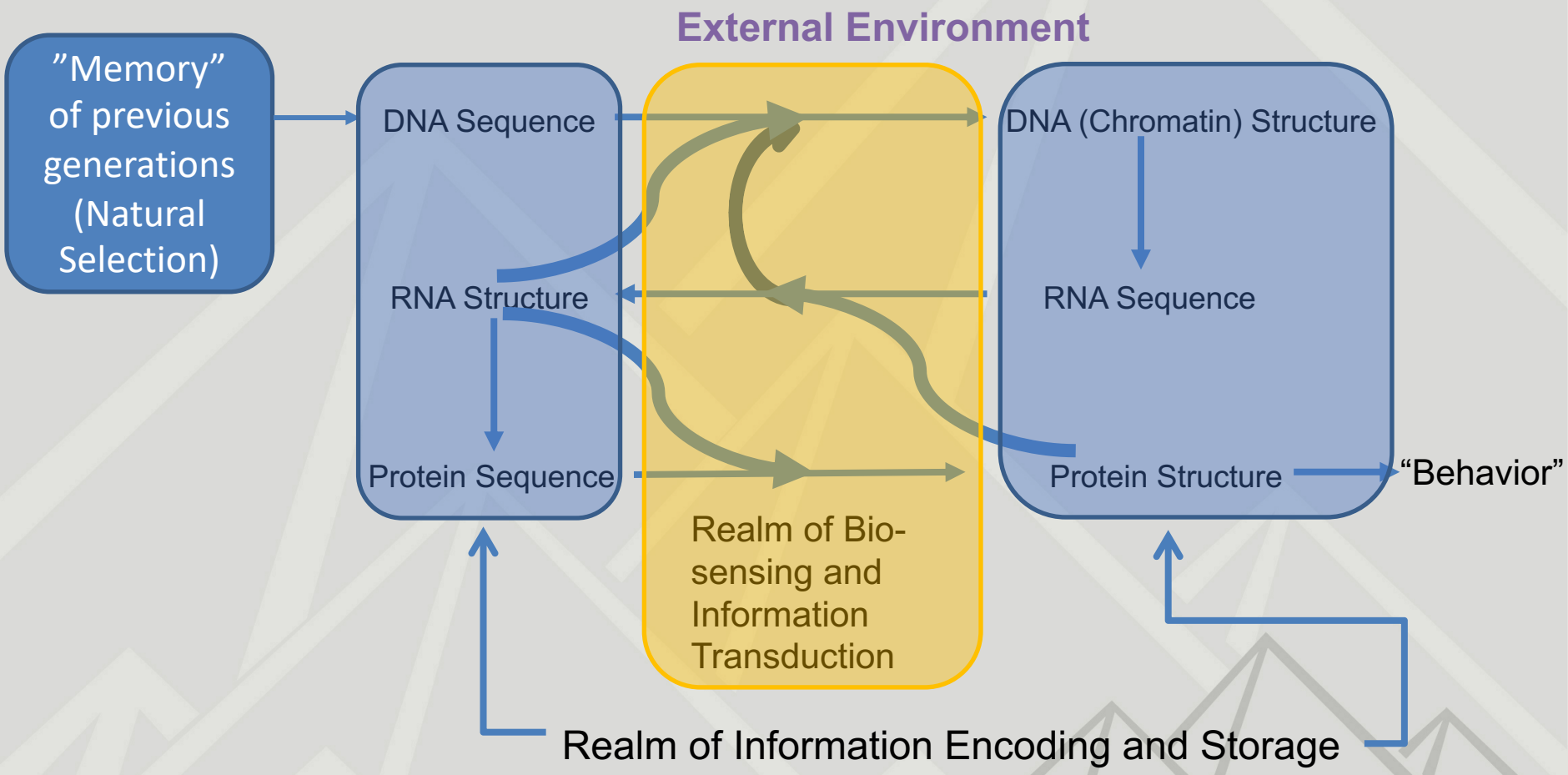




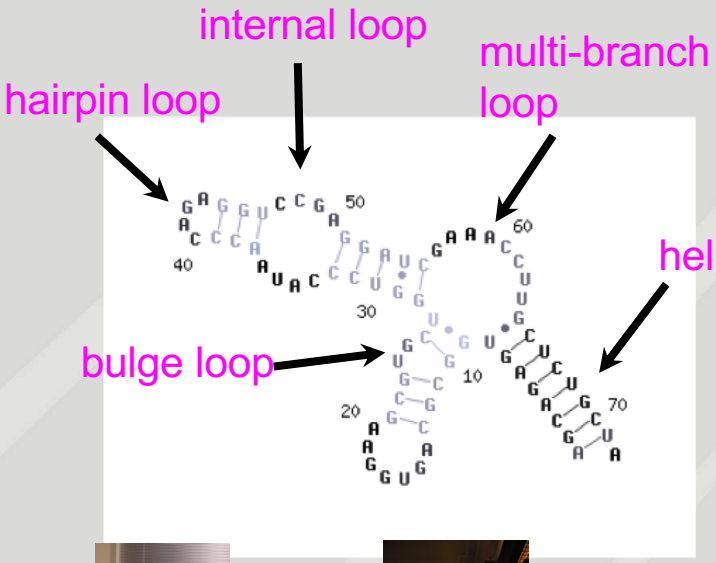




Modern Central Dogma?



RNA Folds as a Single Strand

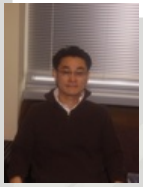


A partition function can be used to determine the base pair probability (BPP)

$$P_{i,j} = \sum_k \frac{e^{-\Delta G(k)/RT}}{Q} = \left(\frac{1}{Q}\right)\sum_k e^{-\Delta G(k)/RT} = \frac{Q_{i \text{ paired to } j}}{Q}$$

Base pair probability could be more informative than a single minimal free energy (MFE)structure

McCaskill, J. S. (1990). Biopolymers **29**(6-7): 1105-1119.



Joohyun Kim (CCT) Wei Huang

Predicting Protein Folding

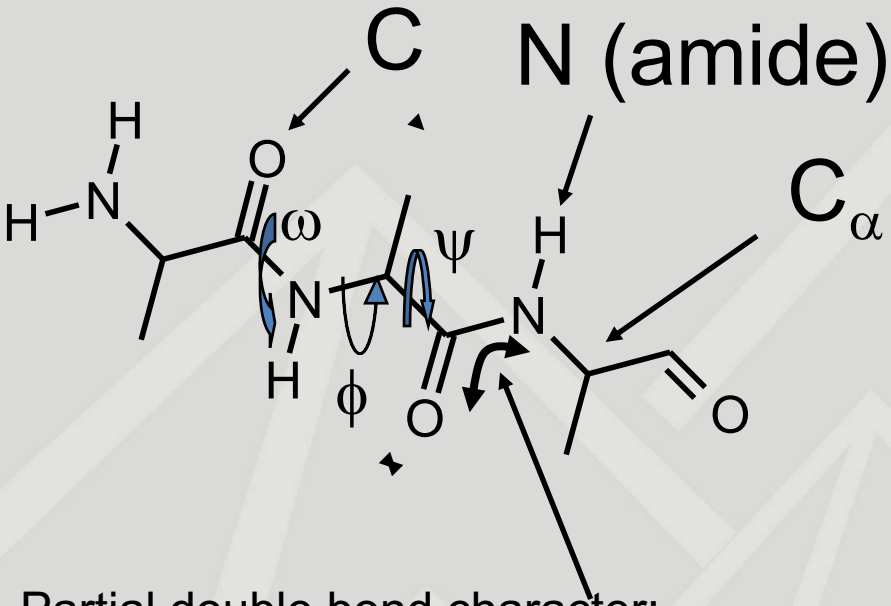
- Evolution selects for function, which is linked to structure
- Mutation can occur more rapidly at RNA level (splicing, recombination, etc.)
- Thermodynamics links sequence to structure, and often but not always to function

Forces driving protein folding

- Conformational entropy: favors conformations which contain dynamic elements
- Hydrogen bonds (intra and with water)
- vdW (hydrophobic interactions)
- Electrostatic forces
- Topological stress due to disulfide bridges

Predicting Protein Structure from First Principles was long the Holy Grail of Biophysics: Protein Structure may be Predicted based on Physics or Homology

Basic components of proteins: the Backbone

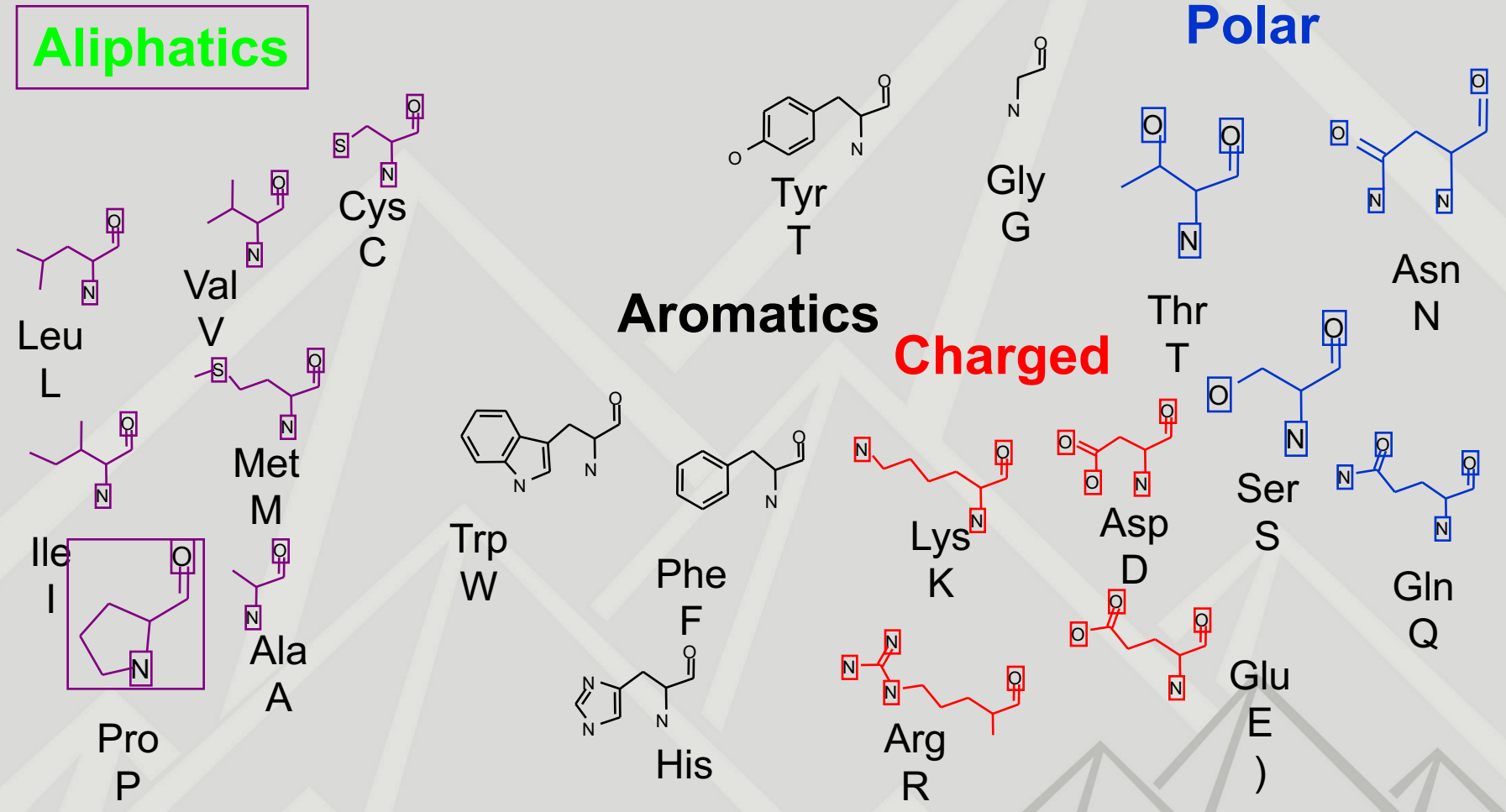


Partial double bond character:
leads to restricted rotations

ϕ and ψ are restricted by steric factors

Residue numbering from amino terminus to $COOH^-$

Protein components: the side chains



The Accessable Conformational Space

- Level 1: Rotation about ϕ and ψ are restricted by steric blocks: leads next figure (for poly-A)
- Level 2: Add potential functions: leads to following figure for poly-A
- Adding side chains further limits accessible angles
- Map of database torsions and accessible torsions are in agreement: restrictions on ϕ and ψ reflect steric constraints

This method then provides us with allowed structures

Ramachandran Maps

position of the CONH planes in successive pleats.

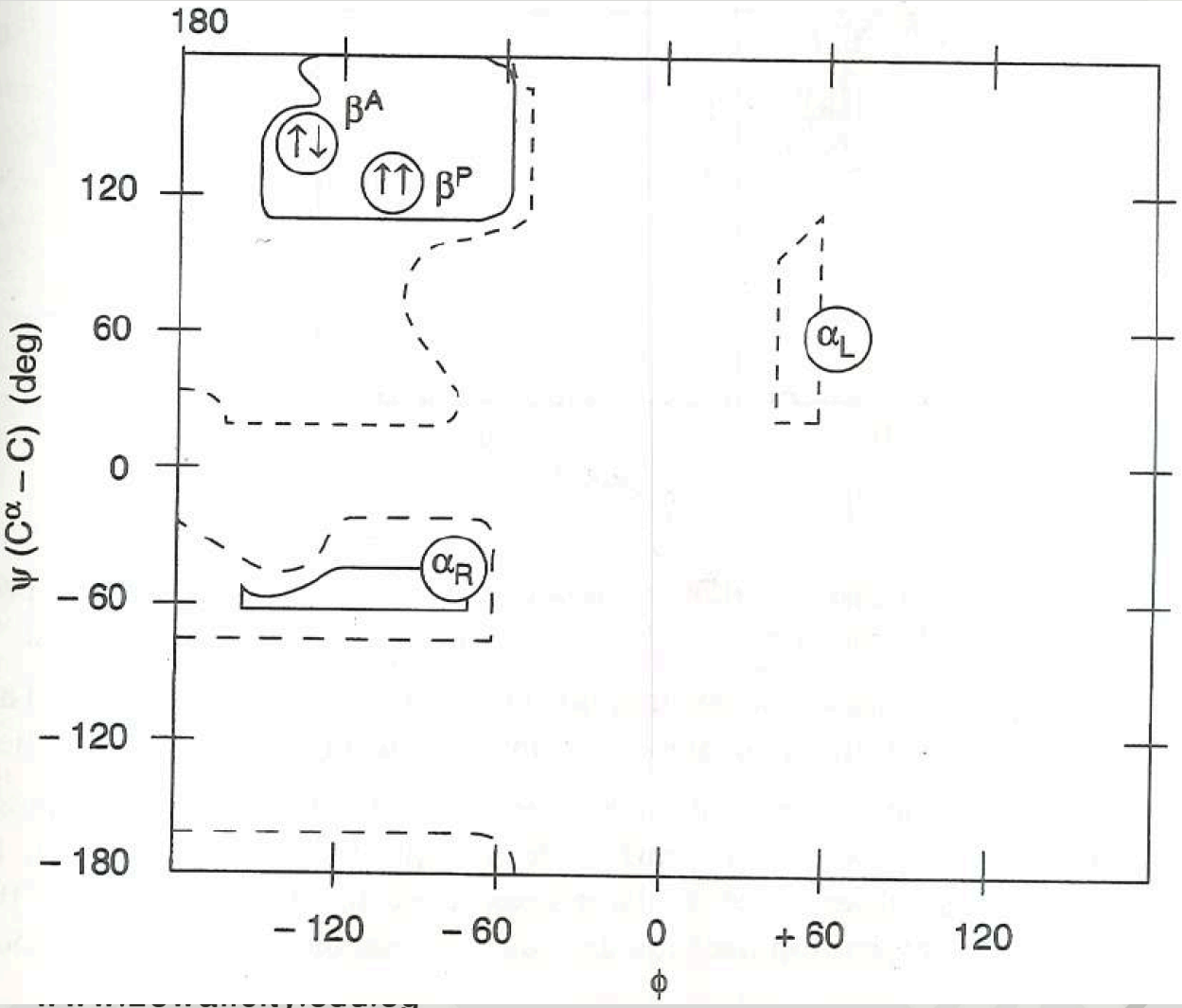


Figure 5.12 A $\Psi-\phi$ map for poly-L-alanine in Ramachandran's method. The contours drawn with solid lines give the allowed distances in Table 5.4. The contours in broken lines correspond to the minimum distances in Table 5.4 (values in parentheses). Note the symmetry about the diagonal of the contours corresponding to α_R and α_L. (From Flory, 1969. Reprinted from *Statistical Mechanics of Chain Molecules* with the permission of Carl Hanser Verlag, Munich.)

Calculating Potentials

Equation 3.4 shows the form used for a potential in molecular mechanics/dynamics calculations:

$$E_p = \Sigma K_1(r - r_o)^2 + \Sigma K_2(\theta - \theta_o)^2 + \Sigma (V/2)[1 - \cos(n\phi + \gamma)] + \text{Electrostatics} + \text{H-bond} + \text{L-J}$$

This expression is a summation over all of the atoms, bond, bond angles, charges, and vdw and dispersion forces

Ramachandran Maps

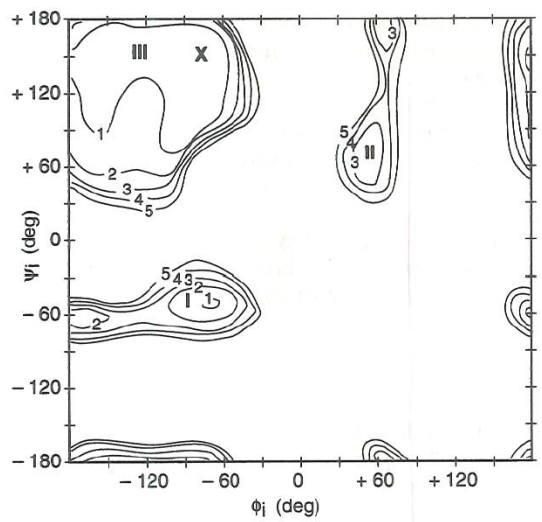


Figure 5.13 A $\phi-\Psi$ map for poly-L-alanine using an interaction potential. The contours are drawn for equal values of potential energy expressed in kcal mol^{-1}, with the zero taken arbitrarily at the cross in region III. The other potential minima in I and II define stable structures for a given (Ψ, ϕ) pair. (From Flory, 1969. Reprinted from *Statistical Mechanics of Chain Molecules* with the permission of Carl Hanser Verlag, Munich.)

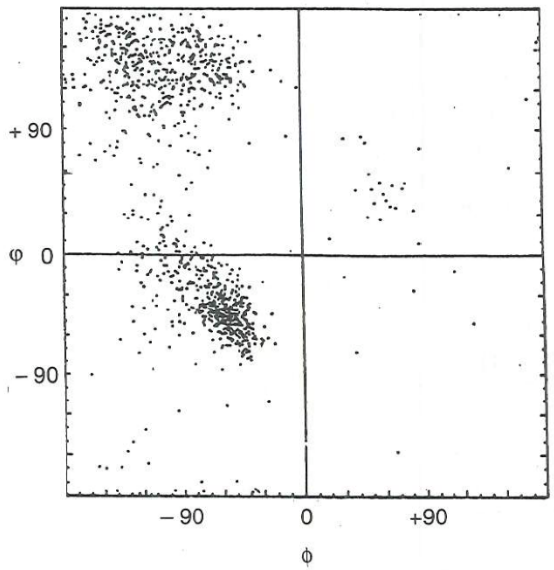
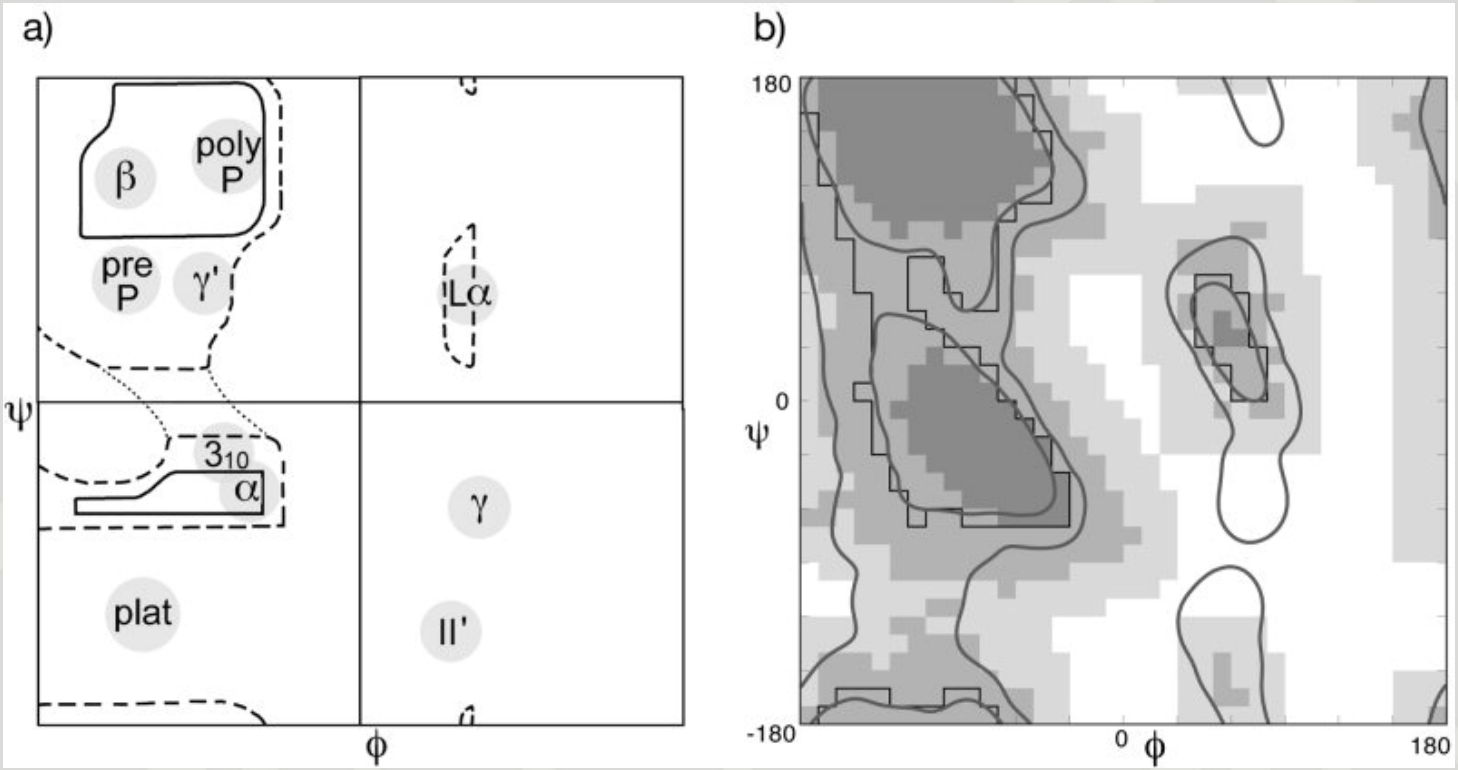


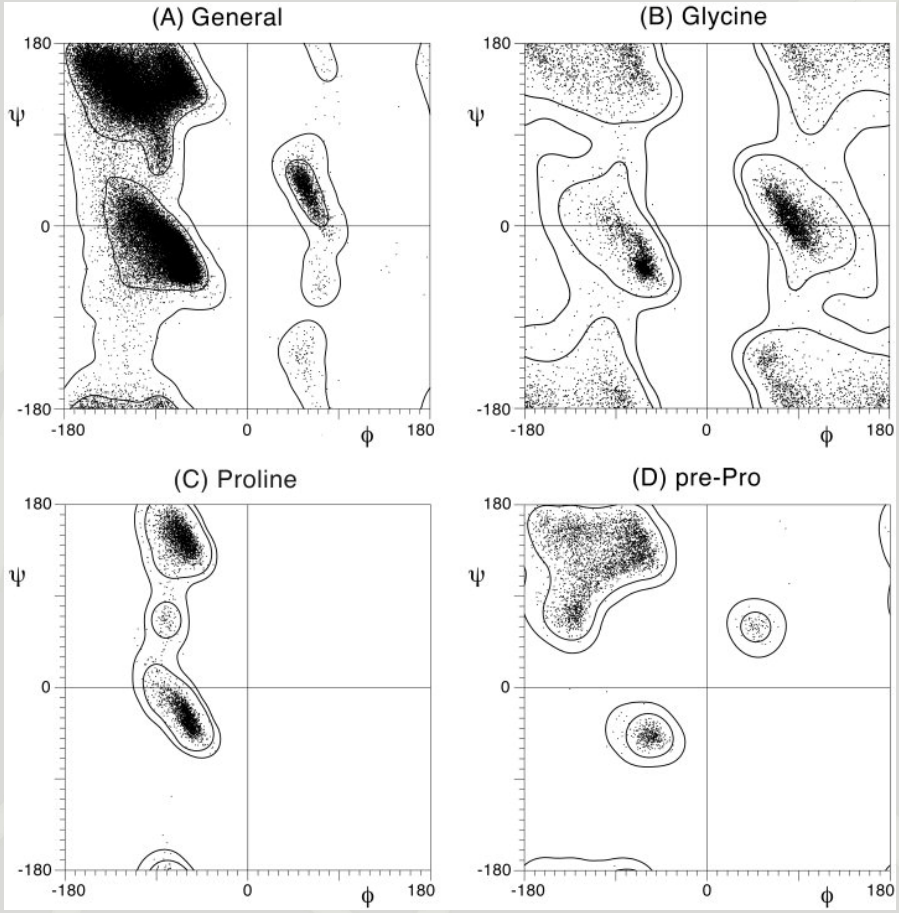
Figure 5.14 A $\phi-\Psi$ map determined from 1000 non-glycine residues in eight proteins whose structures have been refined at high resolution. (From Richardson, 1981. Reprinted from *Advances in Protein Chemistry* with the permission of Academic Press and the author.)

Structure validation by Cα geometry: φ,ψ and Cβ deviation



Proteins: Structure, Function, and Bioinformatics
Volume 50, Issue 3, pages 437–450, 8 JAN 2003 DOI: 10.1002/prot.10286
http://onlinelibrary.wiley.com/doi/10.1002/prot.10286/full#fig1

Structure validation by Cα geometry: φ,ψ and Cβ deviation

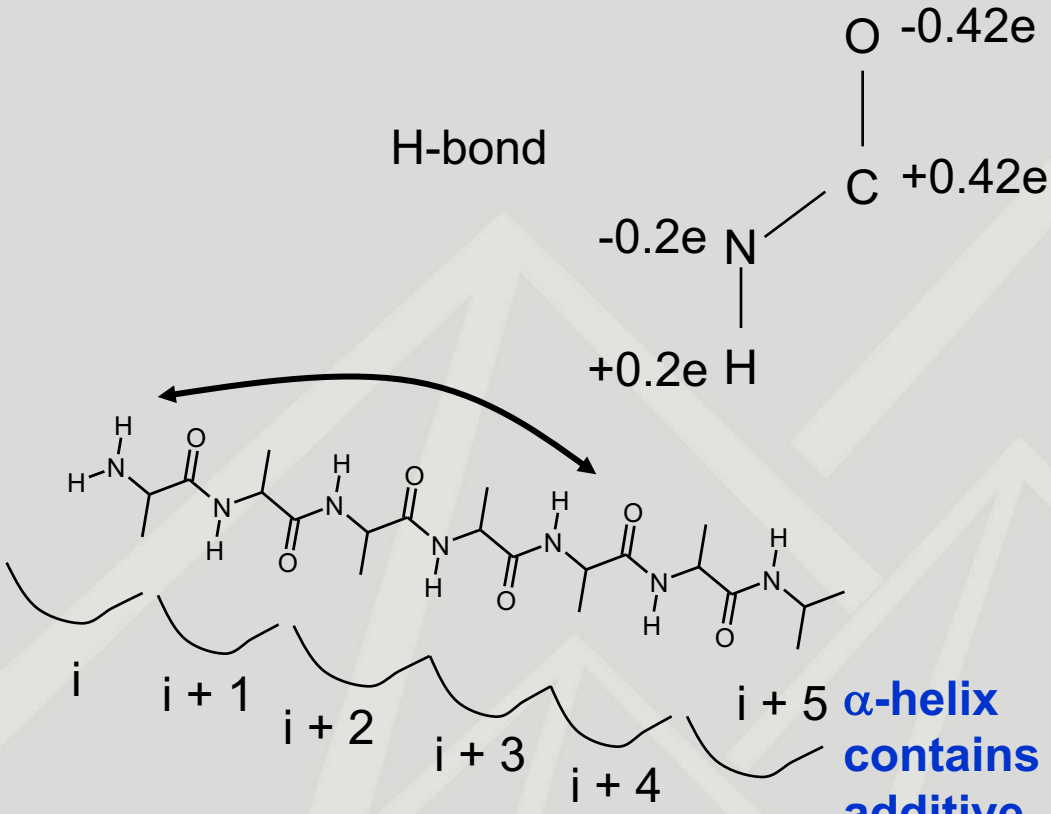


Proteins: Structure, Function, and Bioinformatics
Volume 50, Issue 3, pages 437–450, 8 JAN 2003 DOI: 10.1002/prot.10286
http://onlinelibrary.wiley.com/doi/10.1002/prot.10286/full#fig4

Protein Structures are built from Modular Elements

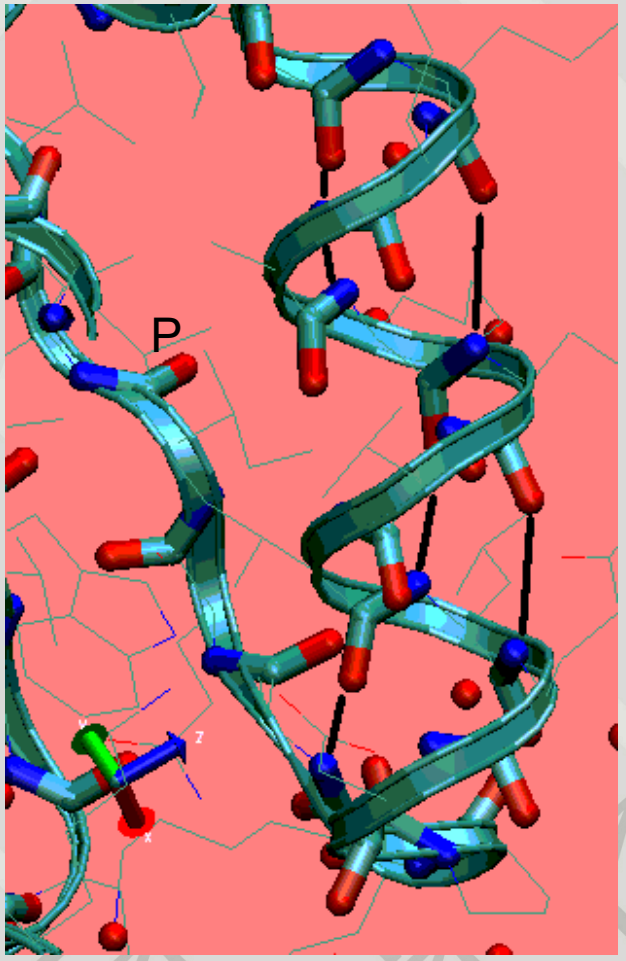
- α-helix
- 3_{10} helix and β-turns
- Polyproline helices
- β sheets

α–Helices Contain a Dipole Moment

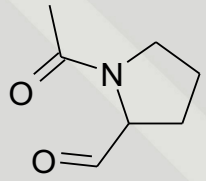


ith carbonyl H-bonded to i + 3th amide

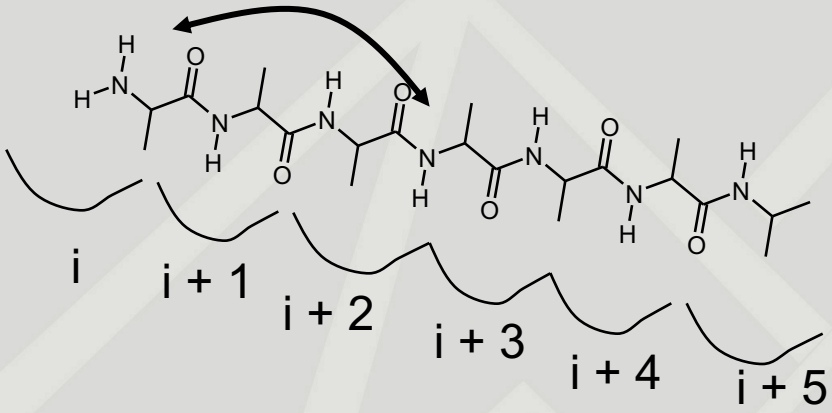
α-helix contains additive dipole moment



3_{10} Helix and β-Turn



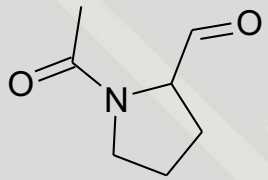
Bonds Occur between i and i + 2



H-bonds (dipoles) not well aligned as in regular α helix

Most common aas in β-Turns:

1. N, C, D, H
2. P, S
3. N, G, D (always G in type II)
4. W, G



Cis-Proline
(form 1)

Trans-Proline
Most common in β turns

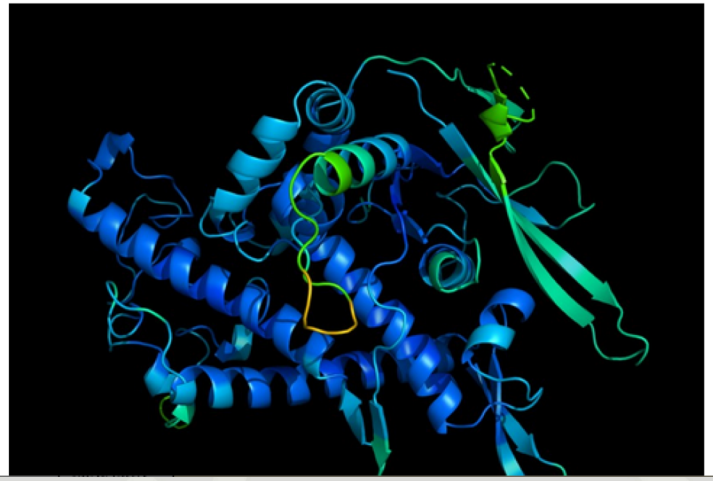
Protein Structure Prediction

NEWS · 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

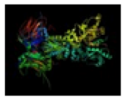
Ewen Callaway



PDF version

RELATED ARTICLES

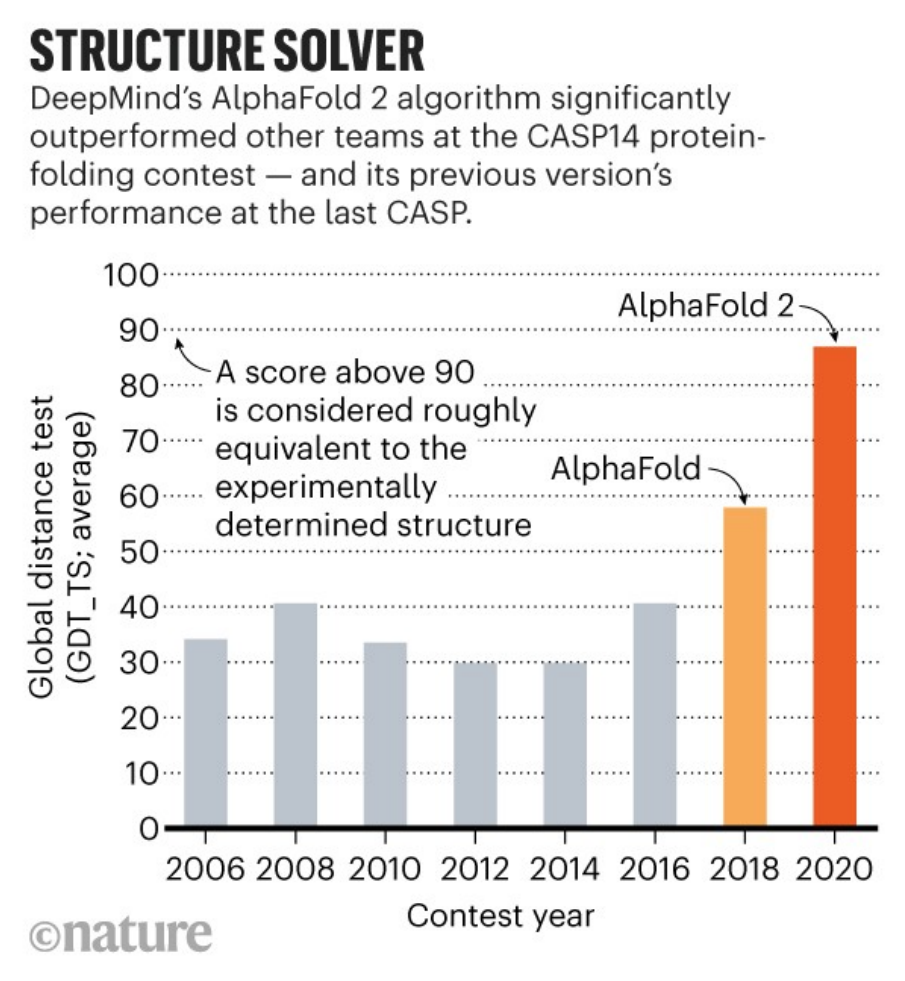
AI protein-folding algorithms solve structures faster than ever



The revolution will not be crystallized: a new method sweeps through structural biology

The computational protein designers

Protein Structure Prediction

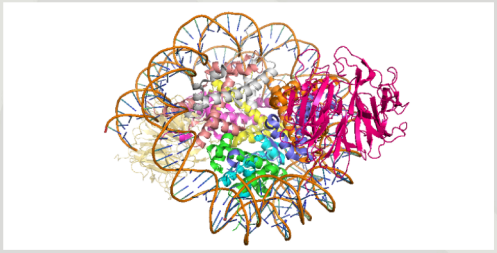


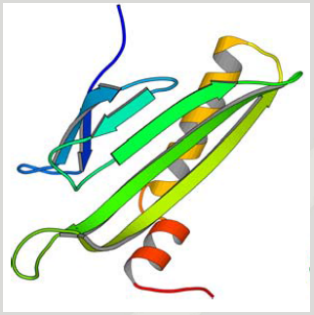
- Deep Mind Won 2020 CASP Competition
- Deep Mind Used >19,000 Structures (protein databank) in the "Learning Set"
- Algorithm Assumes Static Structure
- Still Remains to Predict Change in Fold on Binding to a Partner

https://predictioncenter.org/casp14/

The Cell is Made up of Dynamically Interacting Components

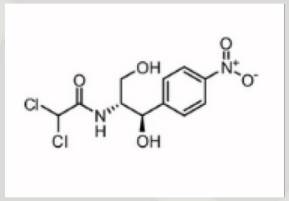
Chromatin 3mvd

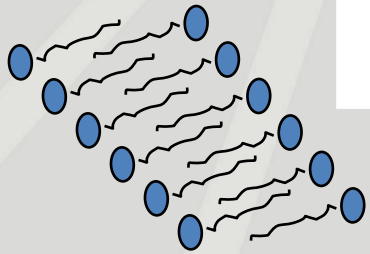




(PDB: 2MS2)

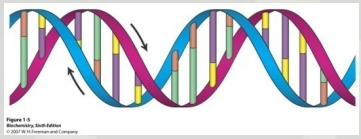
Drugs



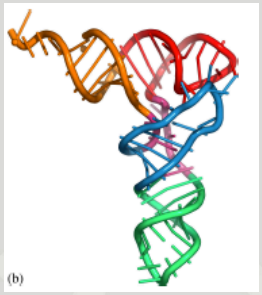


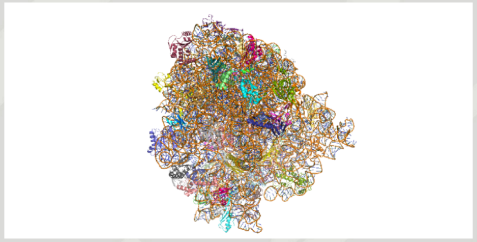
Lipids/membranes

DNA

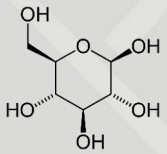


RNA

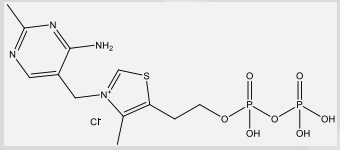




Ribosome 4v49



Metabolites



Vitamins

Ligand Binding and Biomolecular Flexibility

Specificity

- Enzymes
- Ligand/Drug Design
- Synthetic Biology
- Regulation of Gene Expression

Emil Fischer: The Lock and Key "Model" (1894)

"The restricted action of the enzymes on glucosides may therefore be explained by the assumption that only in the case of similar geometrical structure can the molecules so closely approach each other as to initiate a chemical action. To use a picture I would like to say that enzyme and glucoside have to fit together like lock and key in order to exert a chemical effect on each other. The finding that the activity of enzymes is limited by molecular geometry to so marked a degree, should be of some use in physiological research. Still more important though appears to me the proof, that the previously assumed difference between the chemical activity of a cell and the mode of action of chemical reagents is, factually, non-existent."[2]

Angew. Chem. Int. Ed. Engl. **1994**, *33*, 2364–2374

3. Emil Fischer (1852–1919) in [18].

Louis Pasteur
Biography.com

Dan Koshland: The Induced Fit "Model" (1958)

Dan Koshland

The explanation that we suggest to explain these phenomena is as follows: (*a*) a precise orientation of catalytic groups is required for enzyme action; (*b*) the substrate may cause an appreciable change in the three-dimensional relationship of the amino acids at the active site; and (*c*) the changes in protein structure caused by a substrate will bring the catalytic groups into the proper orientation for reaction, whereas a non-substrate will not. This set of postulates has been called "the induced fit" theory for brevity and to emphasize that, while the idea of a fit is retained from the key-lock theory, the fit in this case occurs *only after* the changes induced by the substrate itself.

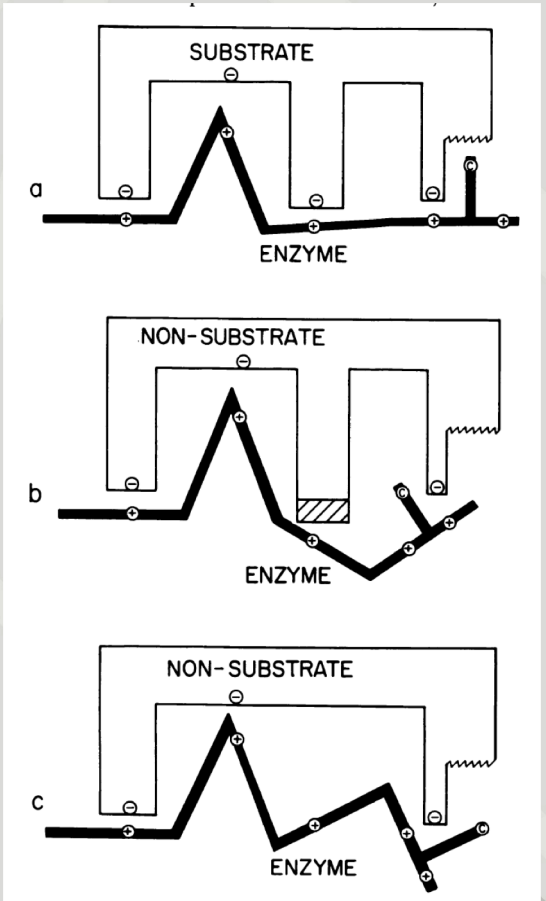
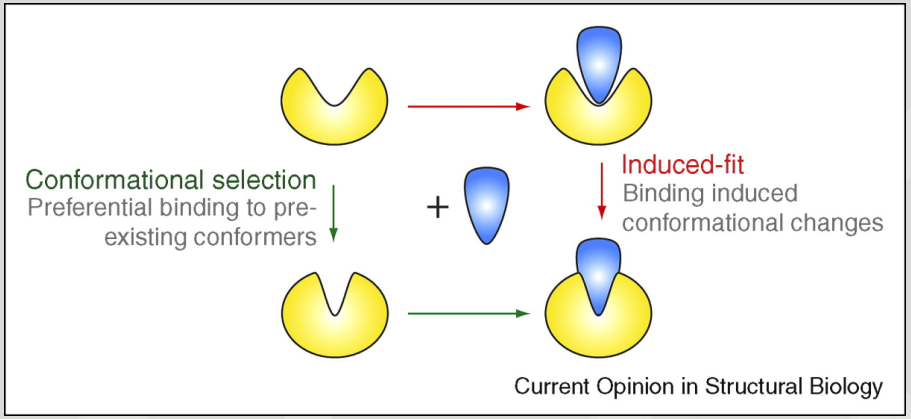


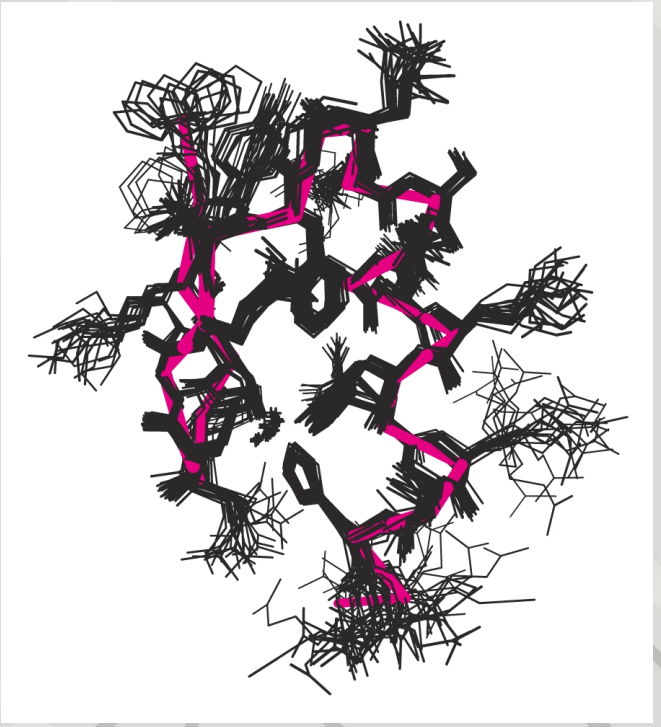
FIG. 1.—Interaction of enzyme with (*a*) substrate, (*b*) compound too large to be substrate, and (*c*) compound too small to be substrate. Circled pluses and minuses indicate any mutually attractive groups on enzyme and substrate. Circled *C* stands for catalytic group, and jagged line for bond to be broken.

1990s: Conformational Capture Model



$$P_i \propto e^{\frac{-\Delta G_i}{kT}}$$

Protein Structure from Nuclear Magnetic Resonance



References

- Molecular Biophysics
  - Biological Physics:Energy, Information, and Life by Philip Nelson
  - Principles of Physical Biochemistry by van Holde, Johnson and Ho
  - Physical Biology of the Cell by Rob Phillips
- Systems Biology
  - Mathematical Modeling in Systems Biology by Brian Ingalls
  - Physical Modeling of Living Systems, by Philip Nelson

Textbook in Progress

Biophysics of Nucleic Acids

- The Free Energy Landscape of DNA and RNA and its Perturbation by Ligands

Biophysics of Nucleic Acids: Proposed Topics

- Introduction
- Nucleic acid composition
- Folding on the Free Energy Landscape
- Experimentally Charting the FEL
- The FEL of Nucleic Acids
  - Nucleotides
  - Dinucleotide stacking
  - Hydrogen bonding and water
- Nucleic Acid Helical Polymorphism
- Cooperative Structural Transitions
- Topology and DNA Supercoiling
- Three Dimensional RNA Folding
- Ions and Solvation
- Ligand binding: Linked Equilibria
- Small molecule Interactions
- Nucleic acid-Nucleic Acid Interactions
- Protein-Nucleic Acid Interactions
- Cryo-EM and Complex Assemblies
- Nucleic Acid Biophysics in Systems Biology