COMPUTATIONAL STRUCTURAL BIOLOGY: BIRTH AND FUTURE

Neve Ilan

15 September 2010

Michael Levitt

Structural Biology & Computer Science

Stanford

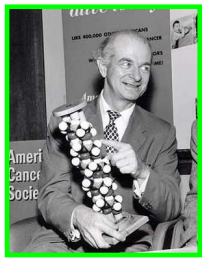
http://csb.stanford.edu/levitt

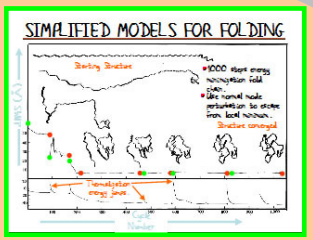
OVERVIEW

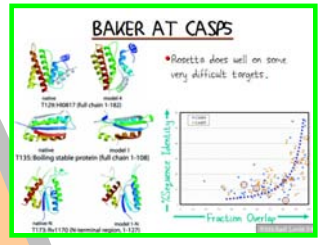
Modeling

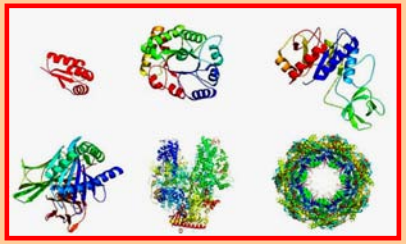
Experiment

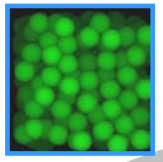
Simulation

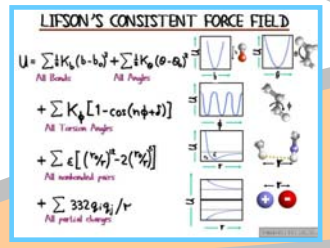


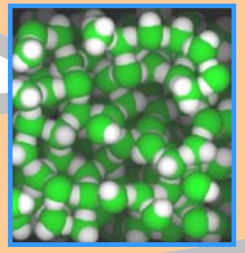


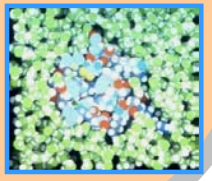


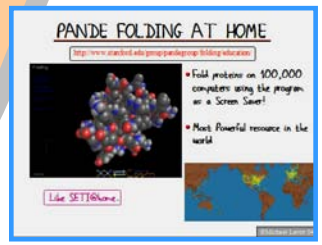












Models for Evolution

Large-Scale Structure

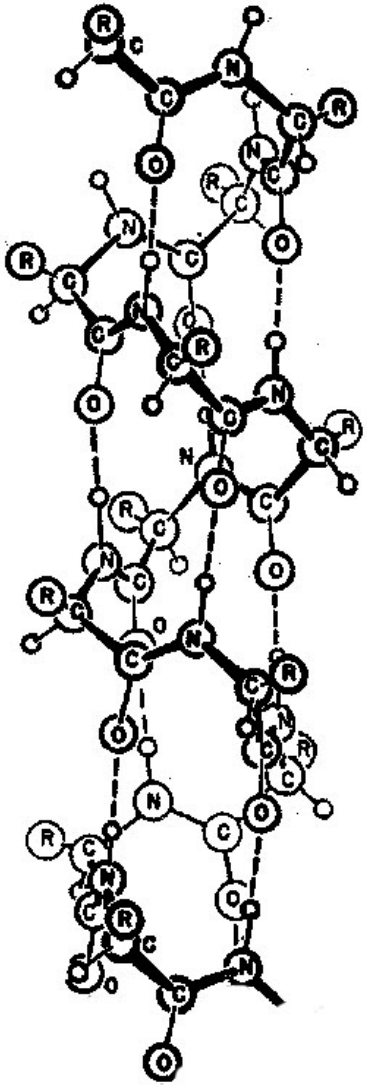
Accurate Simulation

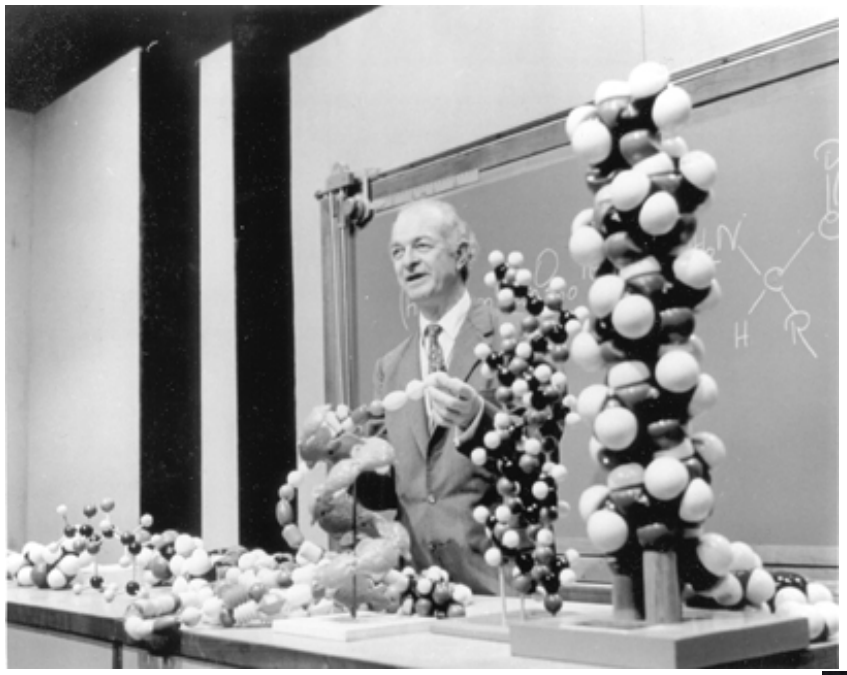
1950 1960 1970 1980 1990 2000 2010

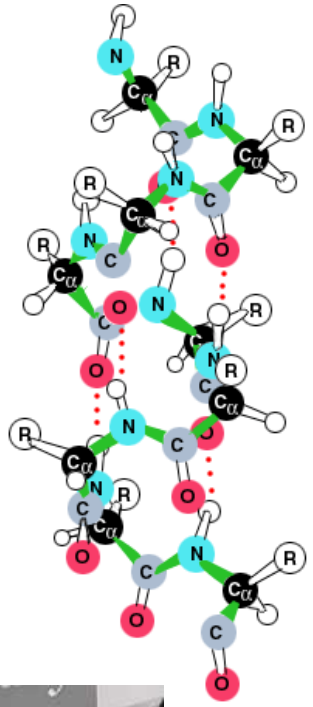
PREHISTORIC TIMES (NON-COMPUTER)

Early Modeling

1950: PAULING THE GREAT CHEMIST

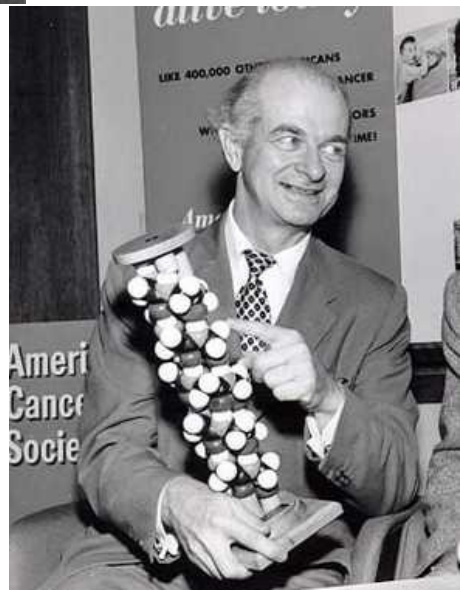






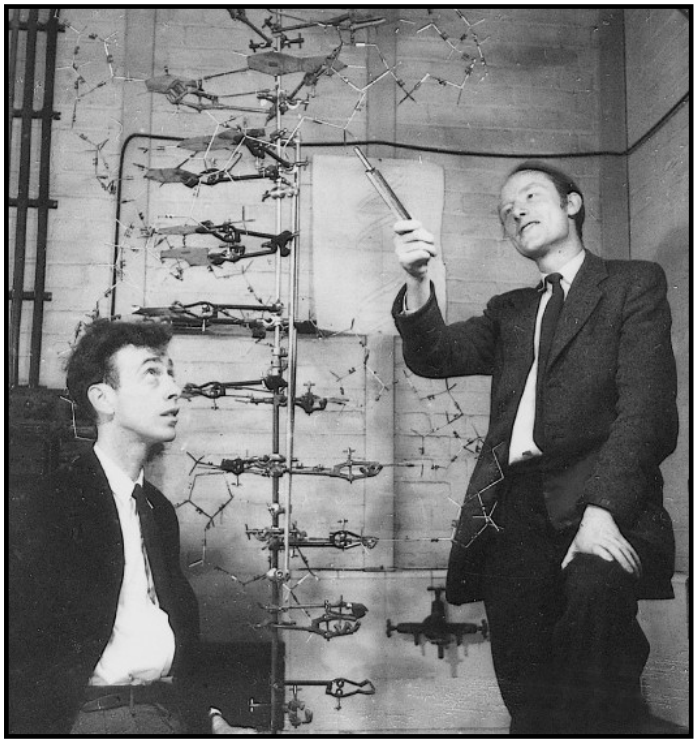
1951

The alpha-helix



1901–1994

1952: WATSON & CRICK MODEL DNA



DNA: Fame to Modelers

Early Experiment

1953: DNA MODEL AND EXPERIMENT

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–8], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[4] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through

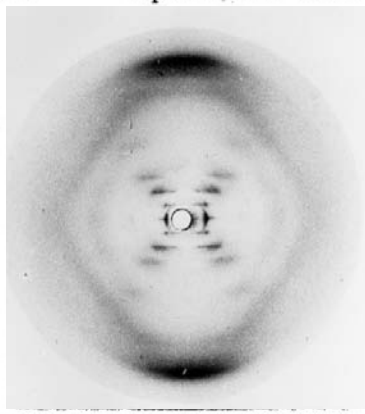


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion (J_n) on the nth layer line. The helical configuration produces side-bands on this fundamental frequency, the effect being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to *C* in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

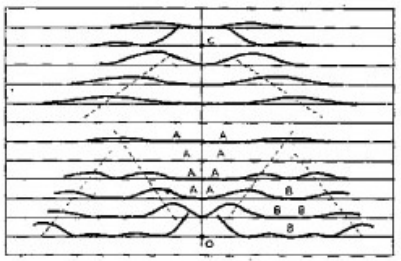
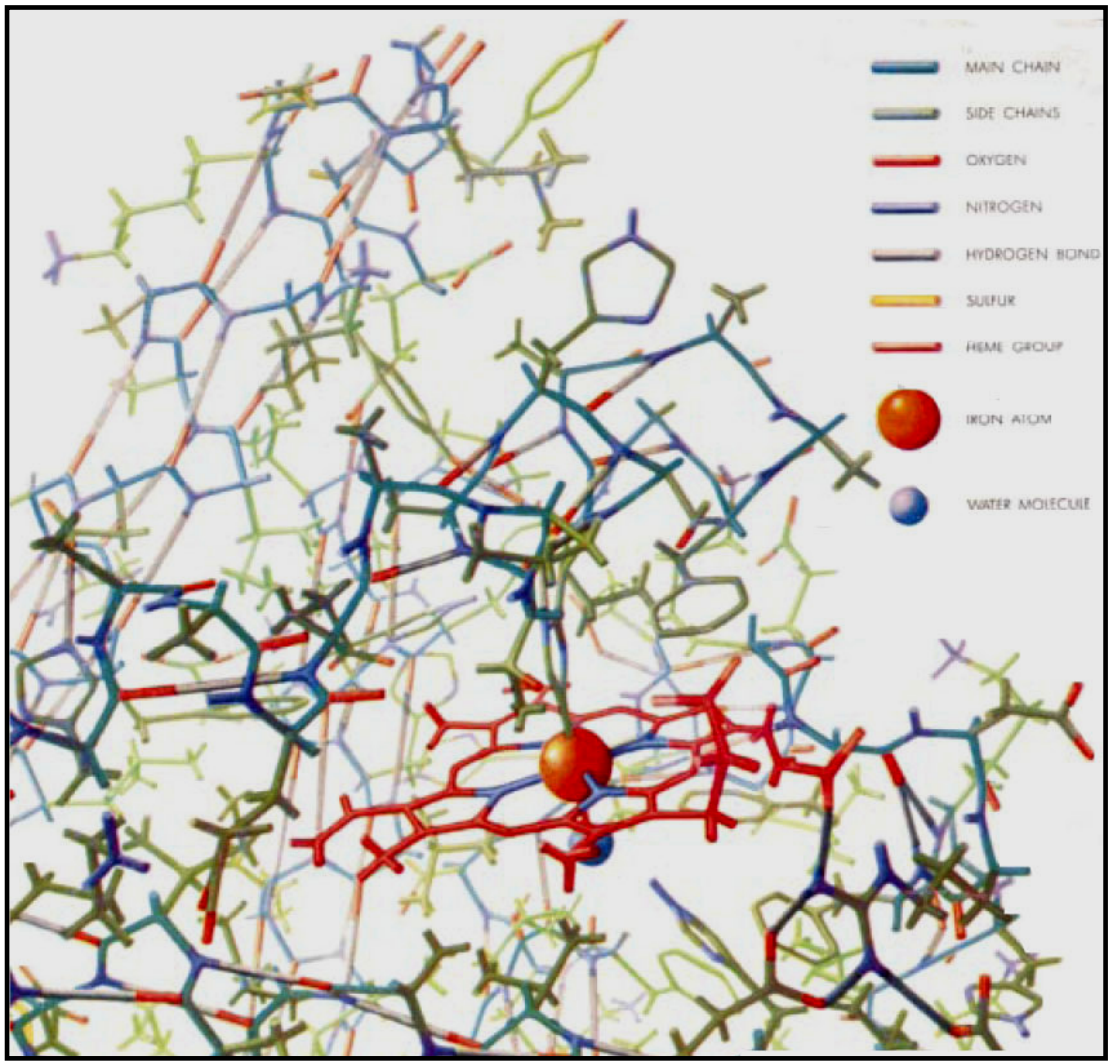


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About *C* on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

1959: KENDREW AND MYOGLOBIN

Scientific American 1961

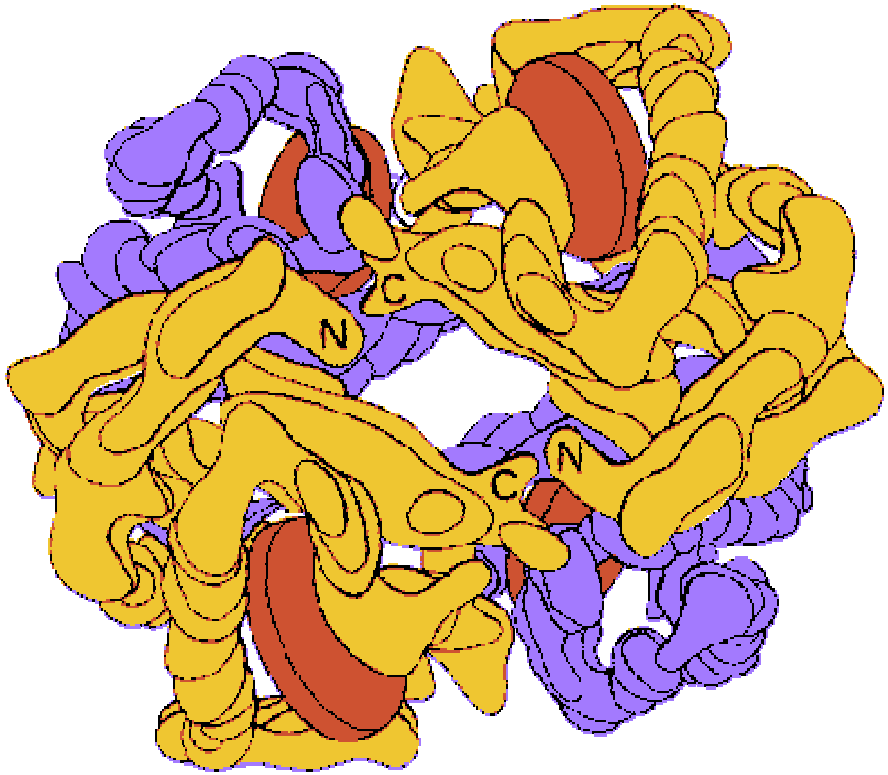


Painted by artist
Irving Geis

1917–1997

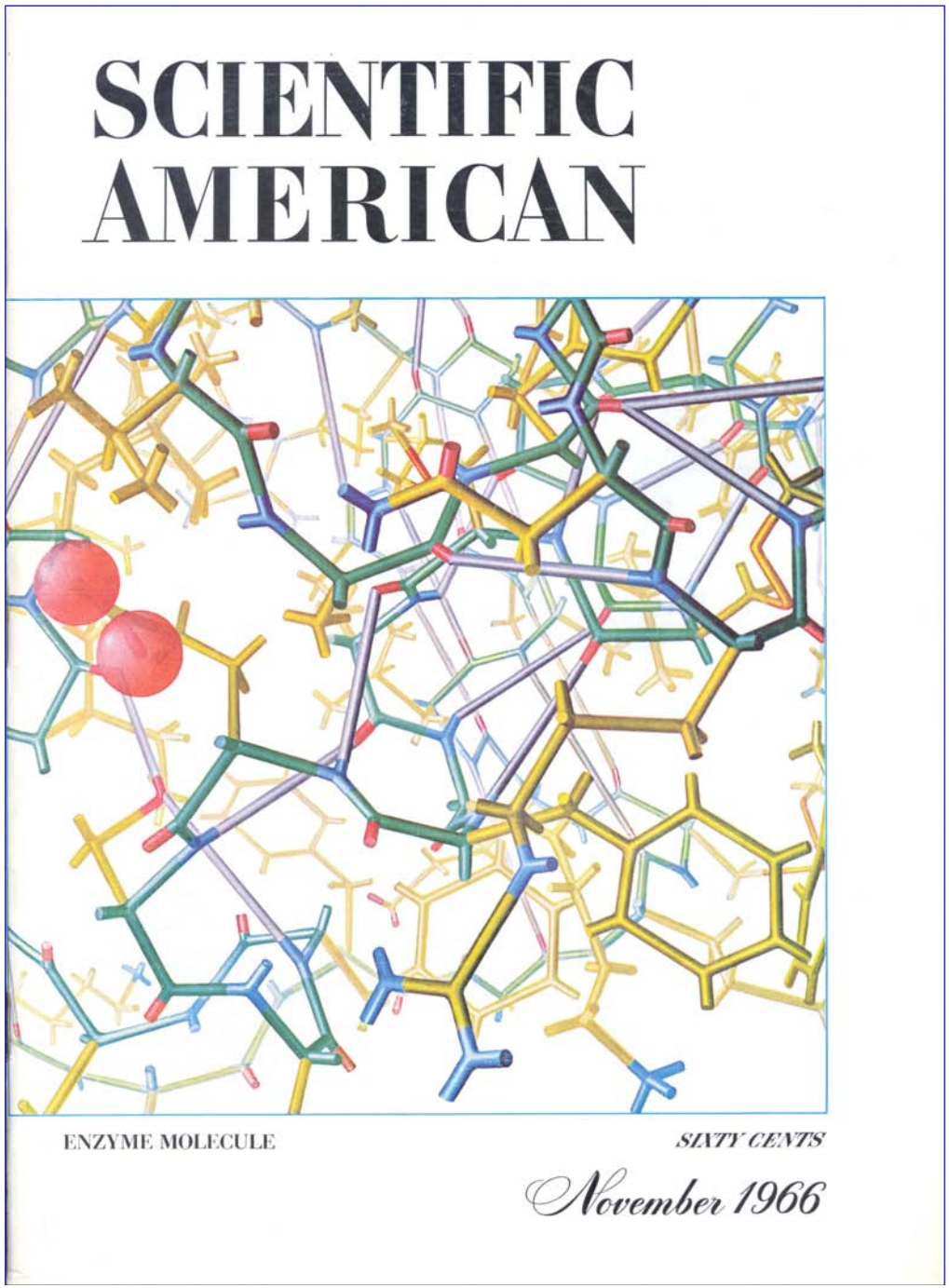
1962: PERUTZ AND HEMOGLOBIN

1914-2002

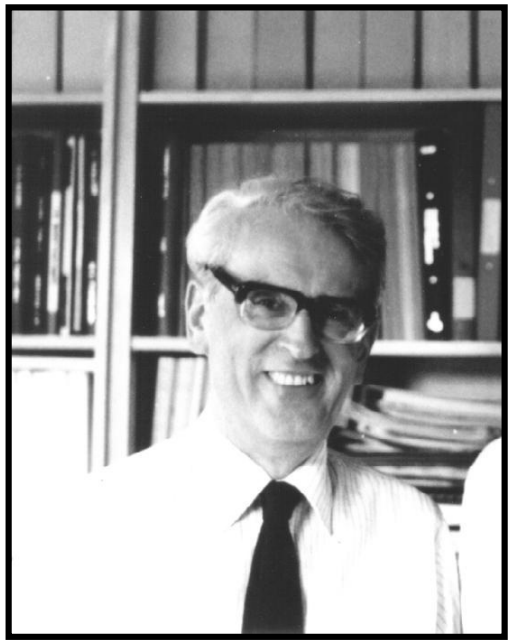


The real hero of the structural biology.

1965: PHILLIPS AND LYSOZYME



1924–1999

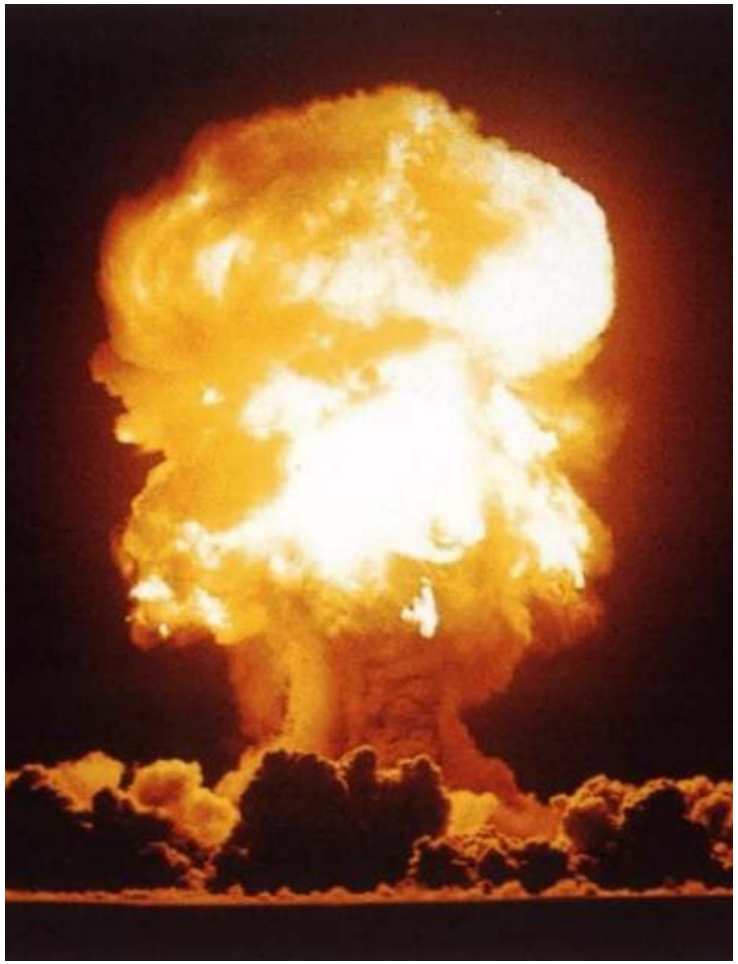


Great supporter of Computational Biology

Kedumin 1985

Early Simulation

1943-1945: LOS ALAMOS



The Birth of the Monte Carlo Method.

When any sufficiently large nuclear explosion occurs within a container, unless the radioactive material is properly contained and the timing of triggering explosions perfect, neutrons stream out of one side of the container. This leak causes an asymmetrical, much weaker, and more unpredictable blast. In order to make the most potent blast possible, a series of complex events must be modeled so that the radioactive material explodes symmetrically. This research appears under the hygienic guise of solving the "neutron diffusion problem." Until 1943, when von Neumann and Stanley Ulam worked on the neutron diffusion problem, there were essentially only two sorts of modeling employed by scientists and mathematicians to describe complex events: deterministic methods (which are essentially applied mathematics) and variations on stochastic techniques (which were known simply as simulation).

To get around the apparently inevitable incorporation of the random, von Neumann devised a third kind of simulation called the "Monte Carlo" in homage to the games of luck he enjoyed in the gambling capital of Europe. He held that random elements in simulations were unacceptable, a form of contamination tantamount to cheating at cards. Indeed, his aversion to stochastic modeling and his appreciation of rule-based games is at the heart of his epistemology. In the Monte Carlo simulation, Von Neumann devised a non-stochastic formula for approximating the stochastic operators in non-trivial simulations. Essentially, he had found a deterministic way to model random events. At the same time, he had rigged the game in the house's favor. When the Monte Carlo simulation worked, it suggested not only that we could describe nature without relying on randomness or chance, but that nature itself was deterministic.

http://trace.ntu.ac.uk/frame2/articles/borg/JvN.html

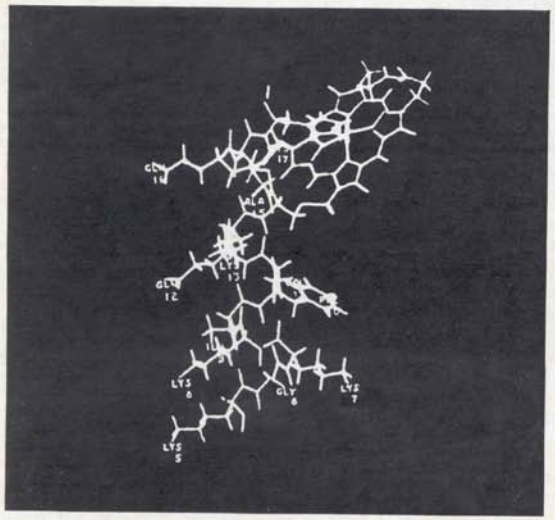
1966: LEVINTHAL

Molecular Model-building by Computer

In which biochemists observe models of giant molecules as they are displayed on a screen by a computer and try to fold them into the shapes that they assume in nature

by Cyrus Levinthal

Many problems of modern biology are concerned with the detailed relation between biological function and molecular structure. Some of the questions currently being asked will be completely answered only when one has an understanding of the structure of all the molecular components of a biological system and a knowledge of how they interact. There are, of course, a large number of problems in biology into which biologists have some insight but concerning which they cannot yet ask suitable questions in terms of molecular structure. As they see such problems more clearly, however, they invariably find an increasing need for structural information. In our laboratory at the Massachusetts Institute of Technology we have recently started using a computer to help gain such information about the structure of large biological molecules.

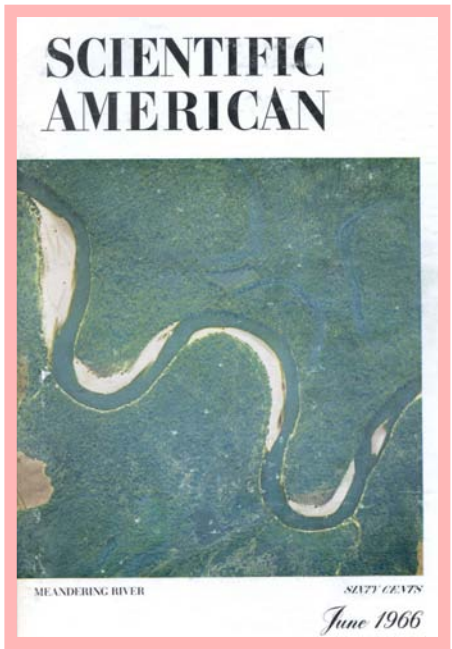


MOLECULAR MODEL of a segment of cytochrome *c*, a protein that plays an important role in cell respiration, is shown as it is displayed on an oscilloscope screen. The protein has 104 amino acid subunits; this segment consists of units 5 through 18 (designated here by their abbreviated names). The heme group, which acts as a carrier of electrons, is known to be attached to amino acids 14 and 17. In the hypothetical structure shown here this stretch of the molecule is assumed to be in the characteristic "alpha helix" configuration.

For the first half of this century the metabolic and structural relations among the small molecules of the living cell were the principal concern of biochemists. The chemical reactions these molecules undergo have been studied intensively. Such reactions are specifically catalyzed by the large protein molecules called enzymes, many of which have now been purified and also studied. It is only within the past few years, however, that X-ray-diffraction techniques have made it possible to determine the molecular structure of such protein molecules. These giant molecules, which contain from a thousand to tens of thousands of atoms, constitute more than half of the dry weight of cells. Protein molecules not only act as enzymes but also provide many of the cell's structural components. Another class of giant molecules, the nucleic acids, determine what kind of protein the cell can produce, but most of the physiological behavior of a cell is determined by the properties of its proteins.

The X-ray-diffraction methods for investigating the three-dimensional structure of protein molecules are difficult and time-consuming. So far the structures of only three proteins have been worked out: myoglobin, hemoglobin and lysozyme [see "The Three-dimensional Structure of a Protein Molecule," by John C. Kendrew, SCIENTIFIC AMERICAN, December, 1961, and "The Hemoglobin Molecule," by M. F. Perutz, November, 1964]. In their studies of the hemoglobin molecule M. F. Perutz and his associates at the Laboratory of Molecular Biology in Cambridge, England, have observed that the structure of the molecule changes slightly when

Earliest molecular graphics.



BIRTH OF COMPUTATIONAL STRUCTURAL BIOLOGY

1967-1976

KENDREW, ME & ISRAEL

amazon.com Prime | Michael's Amazon.com | Books | See all 41 Product Categories | Your Account | Cart | Your Lists

Browse Subjects | Hot New Releases | Bestsellers | The New York Times® Best Sellers | Libros En Español

Books | GO | Gift Certificates

Note: This item is only available from third-party sellers (see all offers.)
Not eligible for Amazon Prime. (Learn more.)

hael Levitt

mage available
ur own customer image

The Thread of Life: an introduction to molecular biology. Based on the series of B.B.C. Television Lectures of the same title (Hardcover)
by John C. Kendrew (Author), b/w photos. Illustrated by Diagrams (Illustrator)

No custome

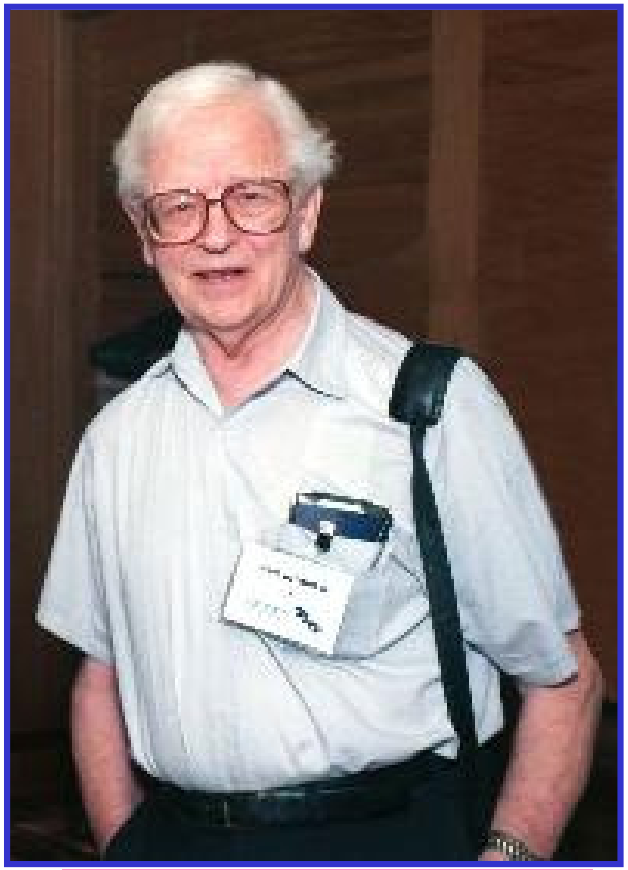
Availability:

1 used & n

how customers can search
side this book.

Episodes

The WAY AHEAD (07/03/1964)
LIVING ARCHITECTURE The VIRUSES (29/02/1964)
The MESSENGER OF THE GENES (15/02/1964)
NUCLEIC ACID The INFORMATION CARRIER (08/02/1964)
REPRODUCTION AND GENETICS (01/02/1964)
PROTEINS IN THREE DIMENSIONS (25/01/1964)
INSIDE THE CELL (11/01/1964)
The REVOLUTION IN BIOLOGY (04/01/1964)
SOLVING THE CODE (22/02/1964)
PROTEINS IN ONE DIMENSION (18/01/1964)



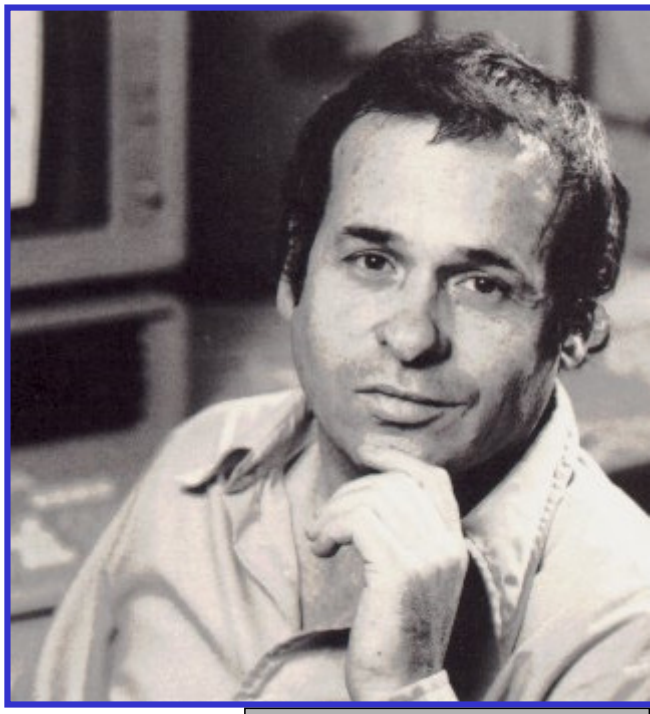
1996 in Israel

But then something changed John's plans: he had become friendly with Suzi Ambache. Susi and her two sisters belonged to a Jewish family from South Africa. John, a slim good-looking young man with a roundish face and dark hair, fell for Susi. John considered matrimony and started learning Hebrew. On 6 August, John wrote once more to his

Suzy Eban

YOM KIPPUR 1967

Arieh Warshel

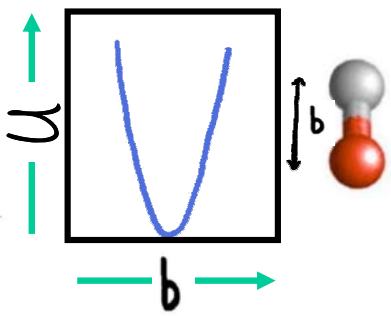


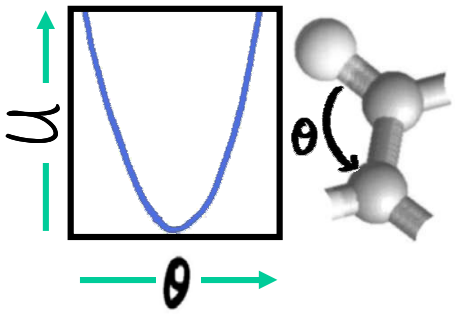
SHNEIOR LIFSON (1914-2001)

LIFSON'S CONSISTENT FORCE FIELD

$$U = \sum \tfrac{1}{2} K_b (b - b_o)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_o)^2$$

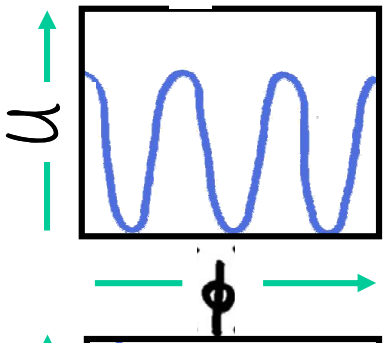
All Bonds All Angles

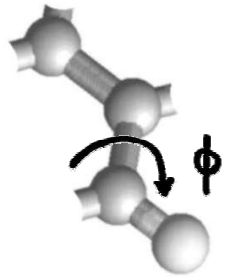




$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

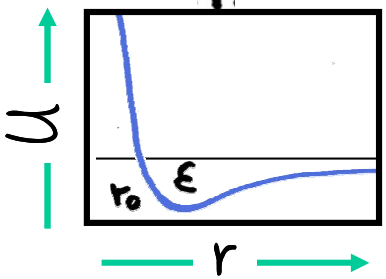
All Torsion Angles

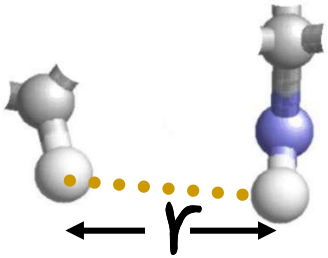




$$+ \sum \varepsilon \left[(r_o/r)^{12} - 2(r_o/r)^6 \right]$$

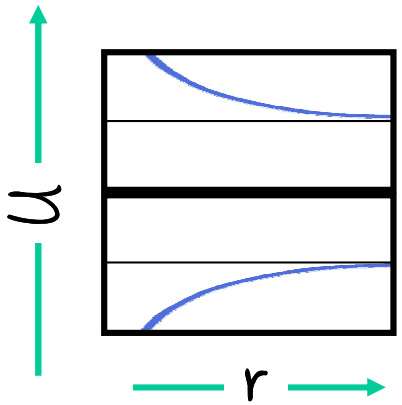
All nonbonded pairs

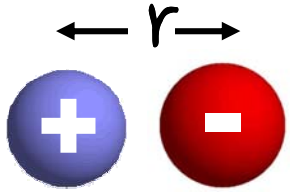




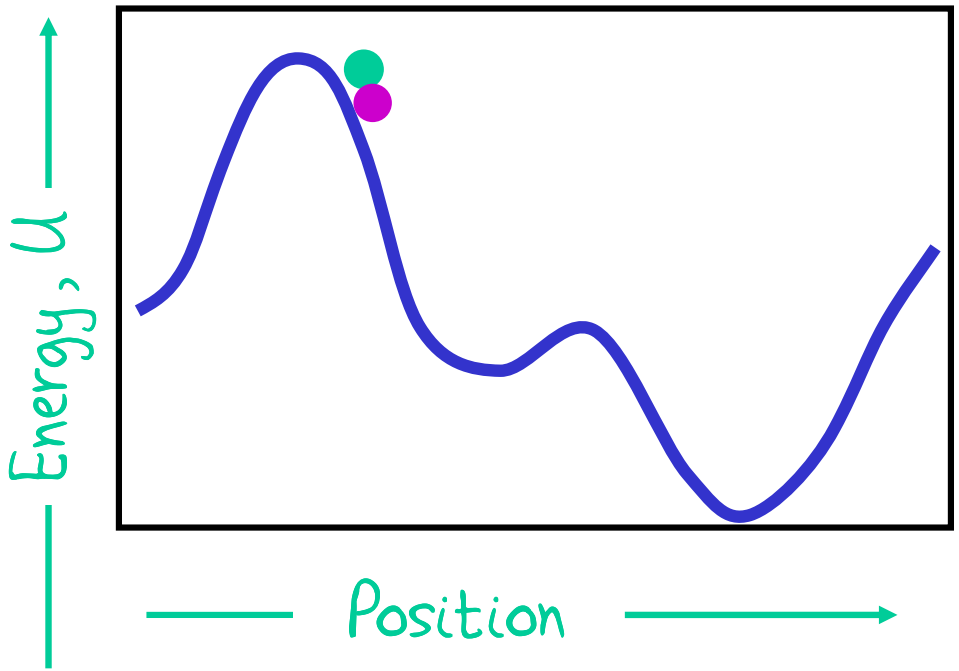
$$+ \sum 332 q_i q_j / r$$

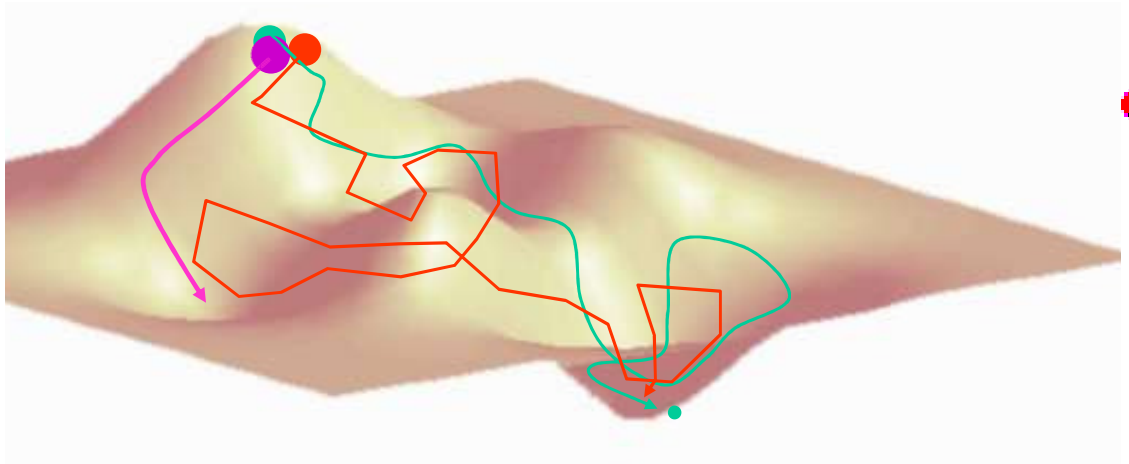
All partial charges





MOVING OVER ENERGY SURFACE





- Energy Minimization drops into local minimum.
- Molecular Dynamics uses thermal energy to move smoothly over surface.
- Monte Carlo Moves are random. Accept with probability $\exp(-\Delta U/kT)$.

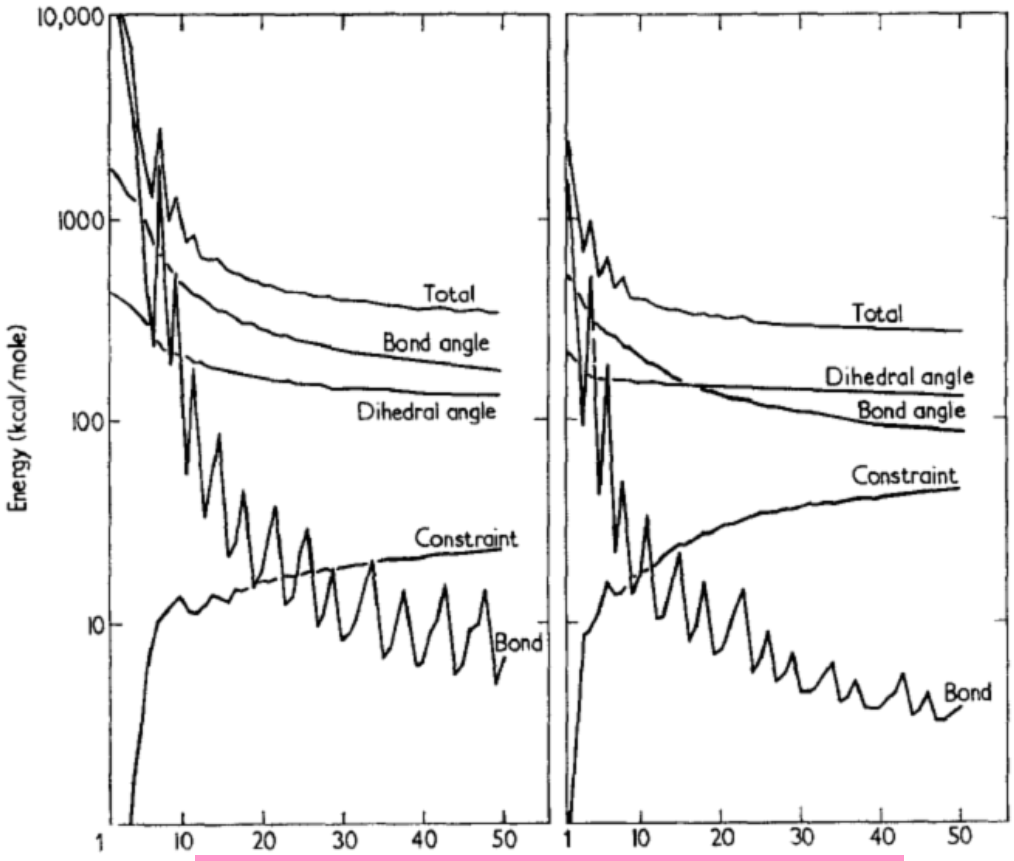
MACROMOLECULAR ENERGY MINIMIZATION

Refinement of Protein Conformations using a Macromolecular Energy Minimization Procedure

MICHAEL LEVITT† AND SHNEIOR LIFSON

Weizmann Institute of Science
Rehovot, Israel

J. Mol. Biol. (1969) **46**, 269–279

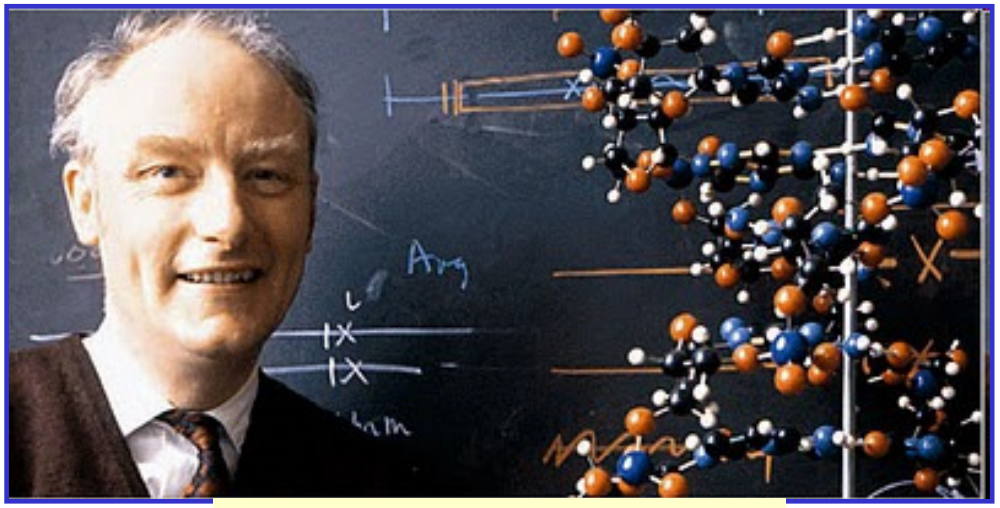


Structure refinement

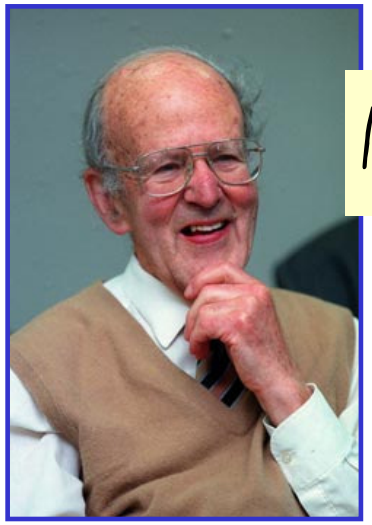
```
ALA     NC(C)AO
ILE     NC(C(C)CC)AO
SER     NC(CO)AO
PRO     *NC(CC*C)AO
MET     NC(CC.SC)AO
ARG     NC(CCCNB(N)N)AO
ASN     NC(CAON)AO
GLN     NC(CCAON)AO
PHE     NC(C*BBBBB*B)AO
```

MENTORS IN CAMBRIDGE

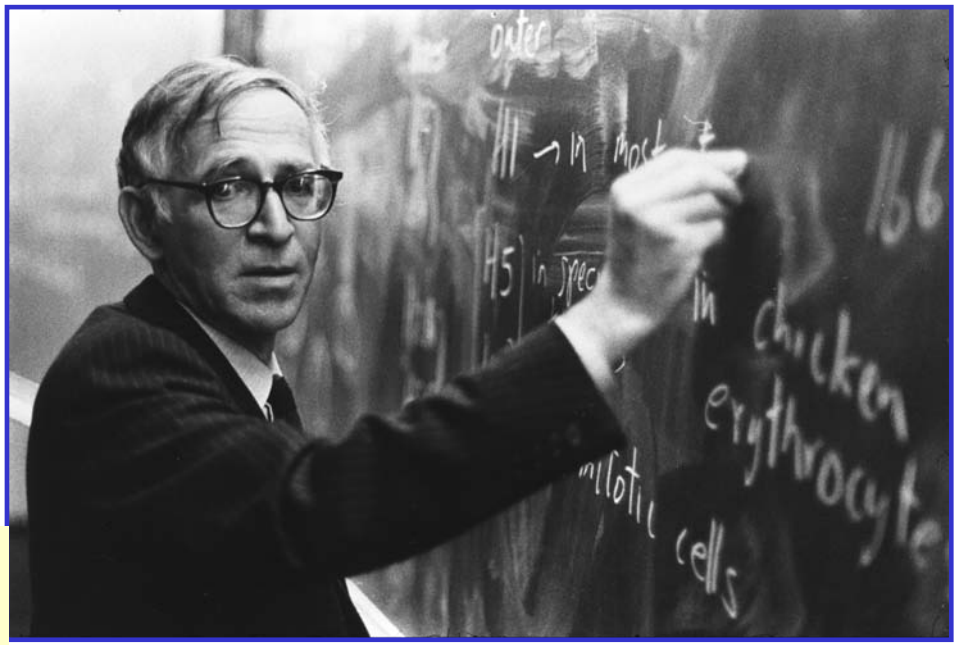
Bob Diamond



Francis Crick



Max Perutz



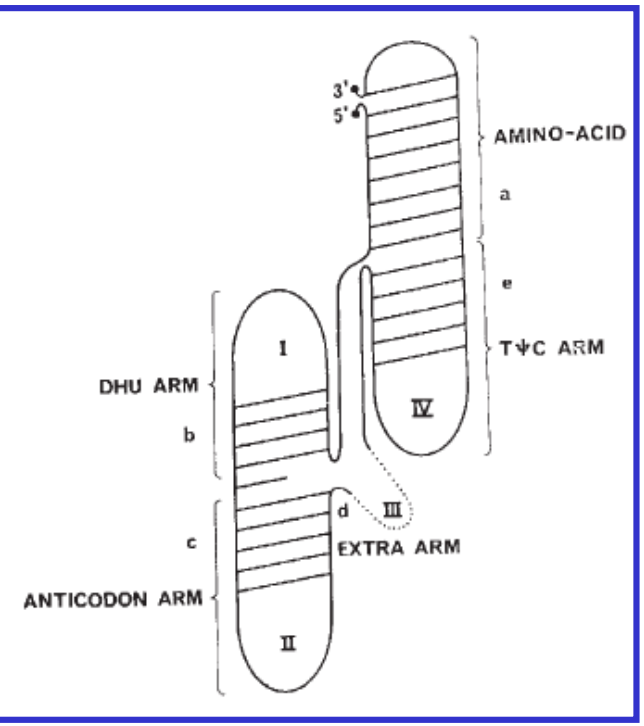
Aaron Klug

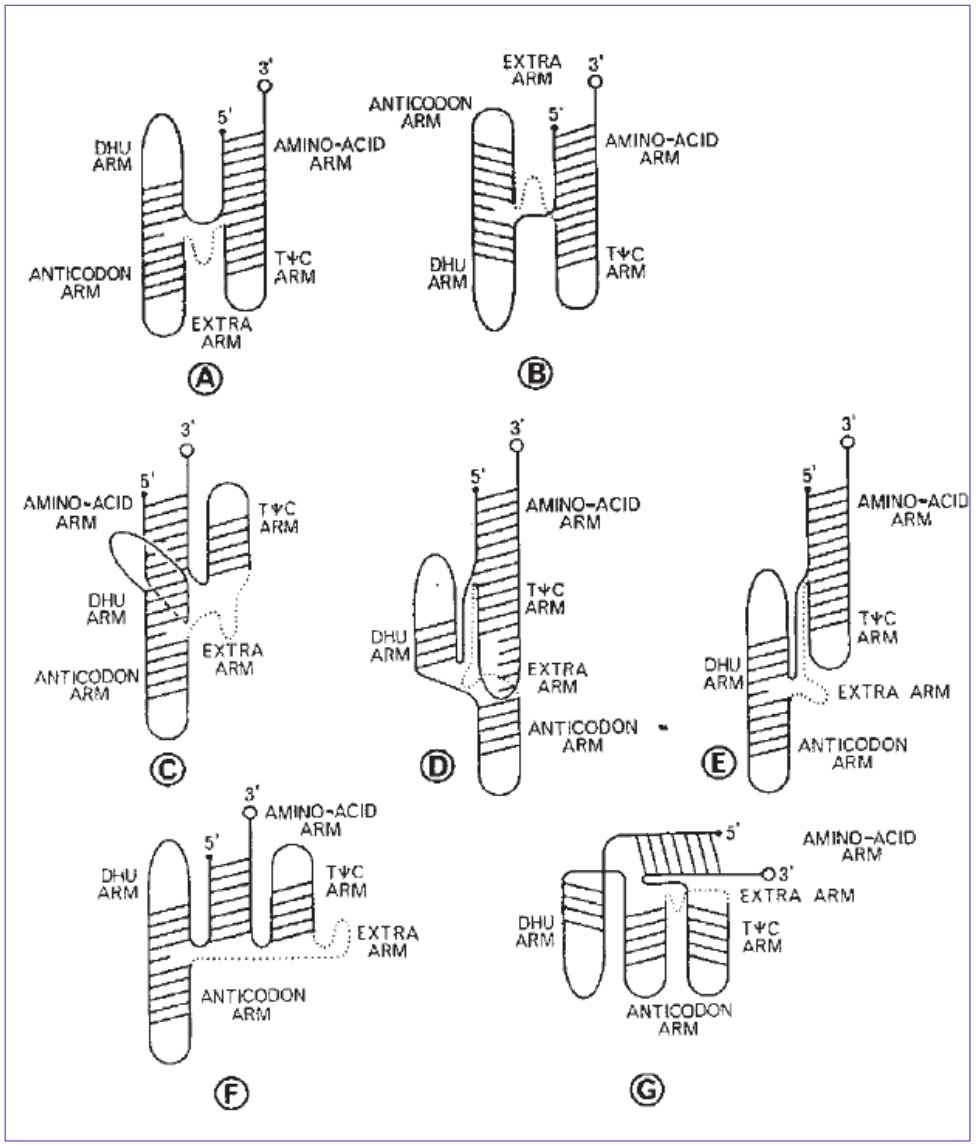
DETAILED MODEL FOR tRNA

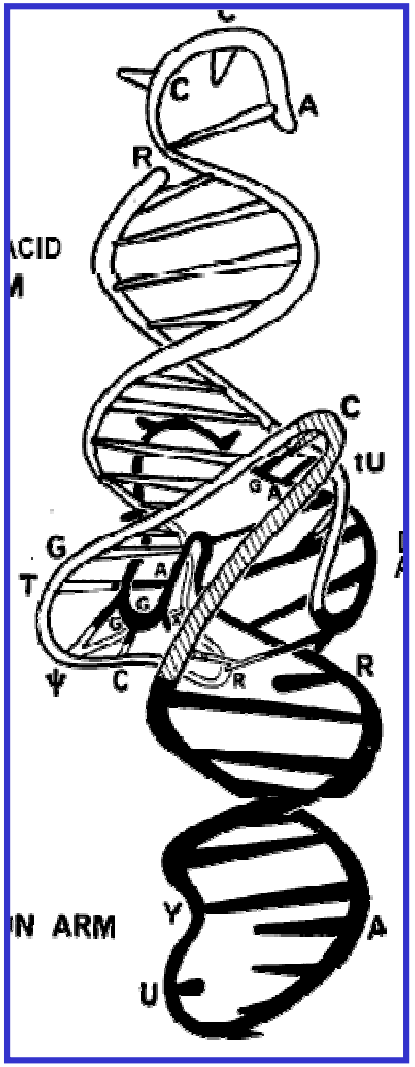
MICHAEL LEVITT

MRC Laboratory of Molecular Biology,
Cambridge

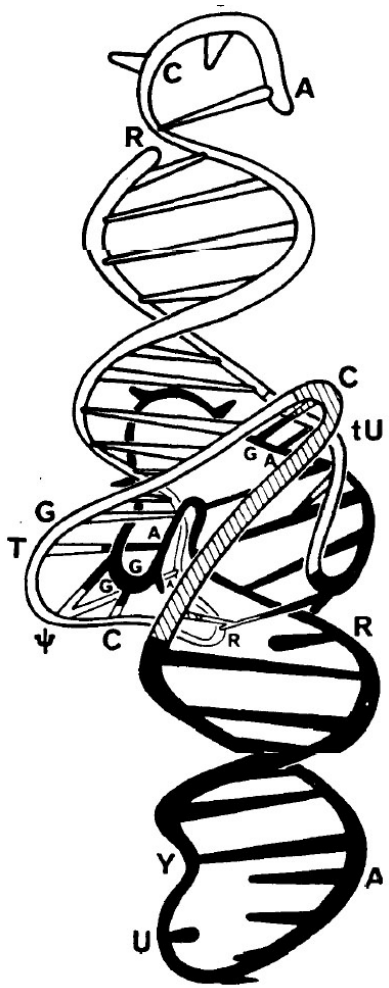
NATURE VOL. 224 NOVEMBER 22 1969

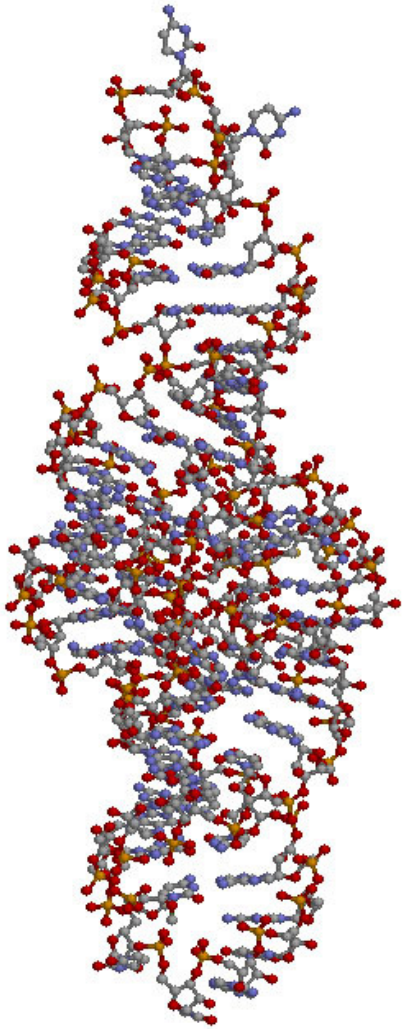


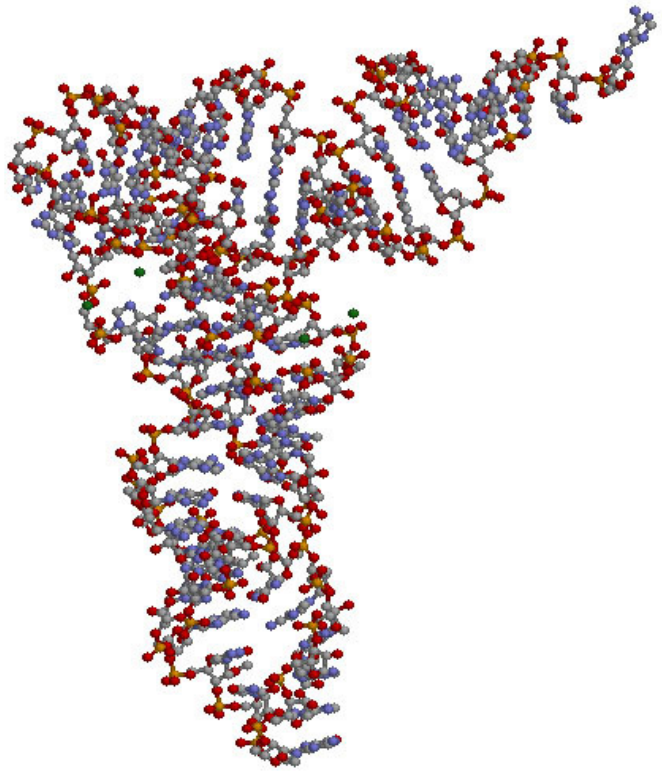




1969: MY tRNA MODEL WAS WRONG







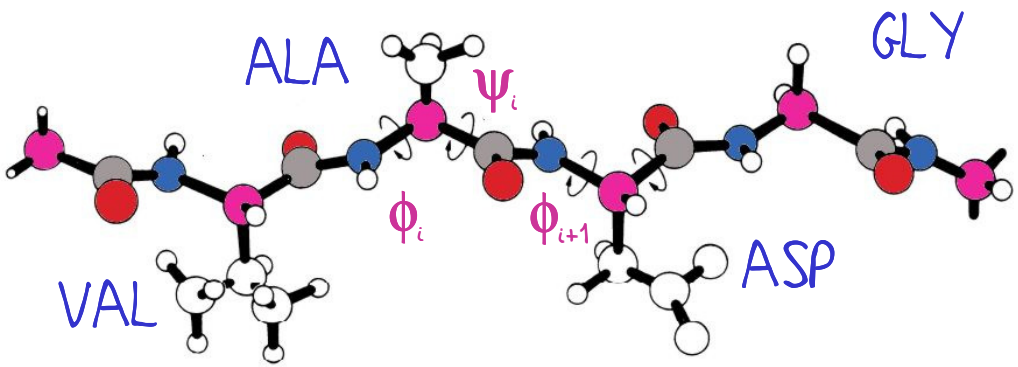
Lessons for future

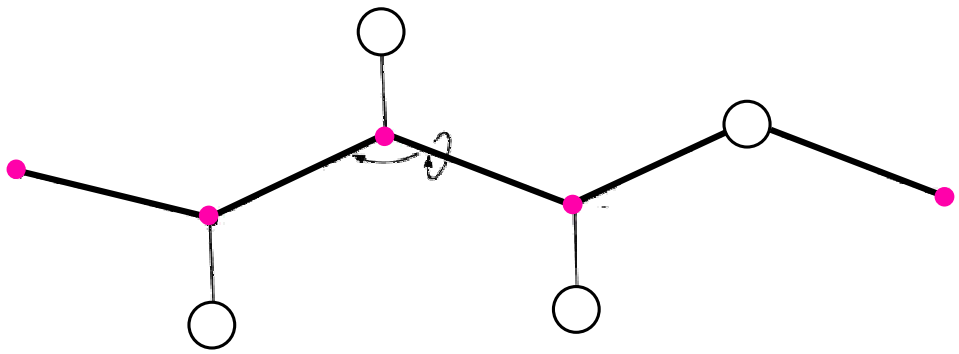
COMPUTER SIMULATION OF PROTEIN FOLDING

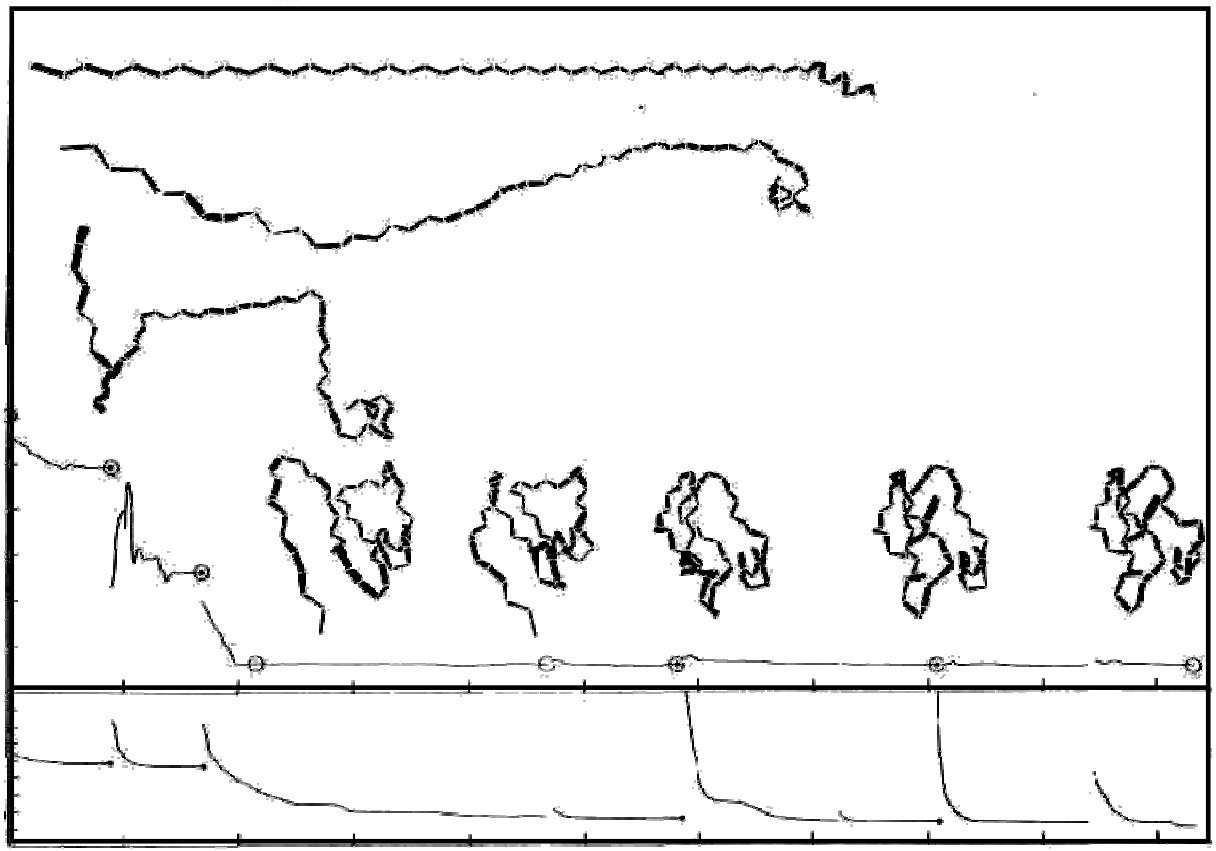
Michael Levitt* & Arieh Warshel*

Department of Chemical Physics, Weizmann Institute of Science, Rehovoth, Israel

Nature Vol. 253 February 27 1975







Reduced models

Fold protein with 1000 steps of minimization.

Escape from local minima with normal modes jumps.

THEORETICAL STUDIES OF ENZYMIC REACTIONS

J. Mol. Biol. (1976) **103**, 227–249

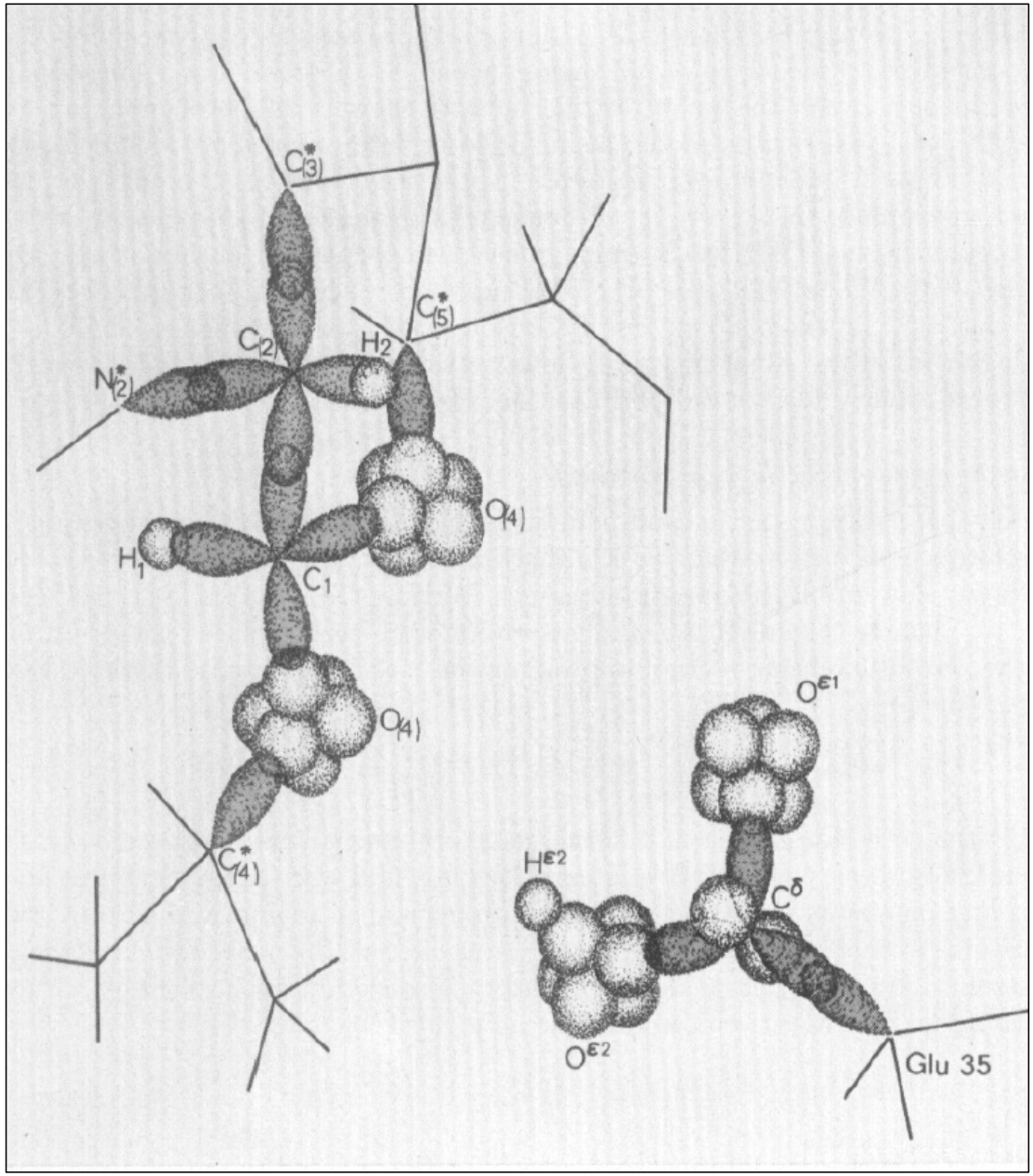
A. WARSHEL AND M. LEVITT

Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

and

Department of Chemical Physics
The Weizmann Institute of Science
Rehovot, Israel

Most cited by far



STRUCTURAL PATTERNS IN PROTEINS

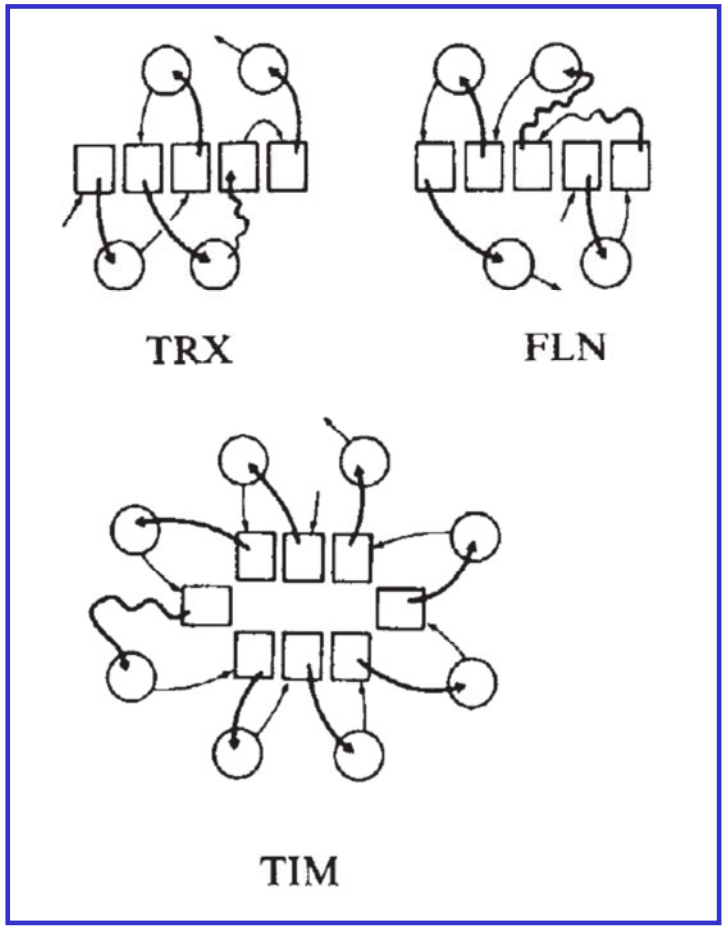
Nature Vol. 261 June 17 1976

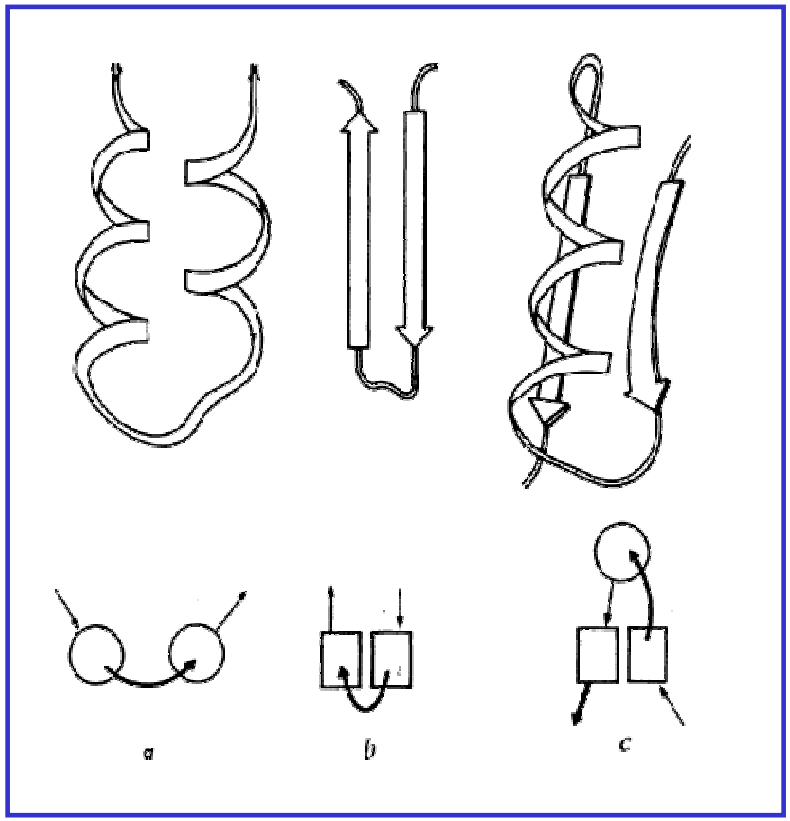
Michael Levitt

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Cyrus Chothia

Service de Biochemie Cellulaire, Institut Pasteur, 75724 Paris, France



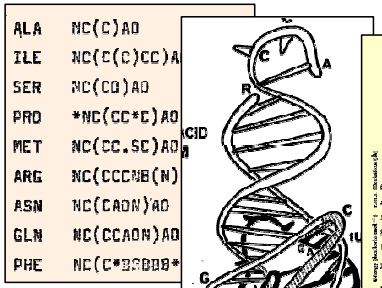


Structure classification

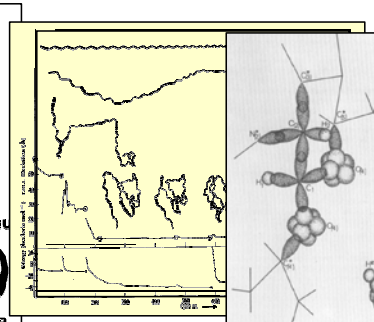
THE GOLDEN YEARS

1977-2007

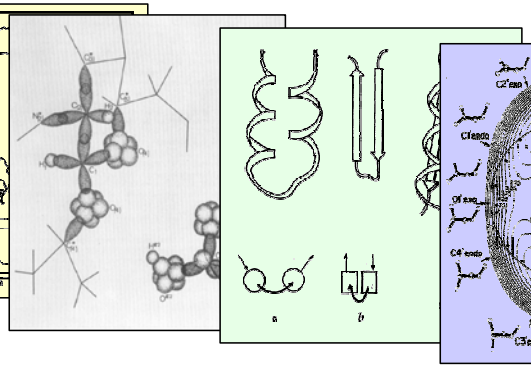
1969



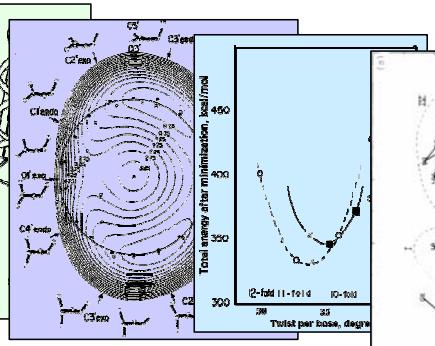
1975



1976

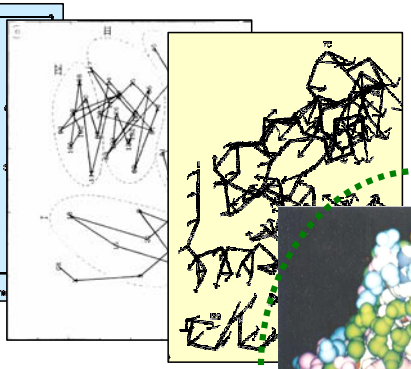


1978

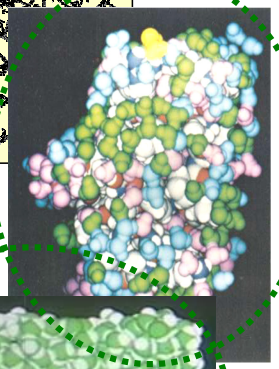


1983

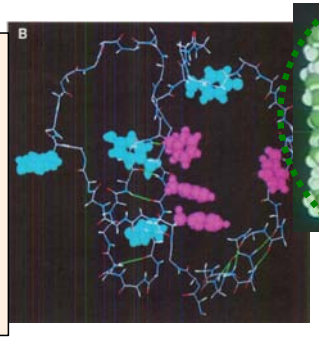
1985

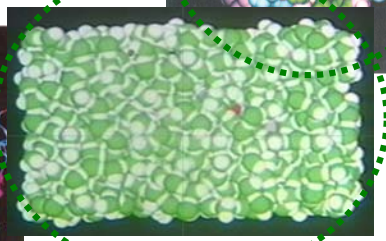


1988



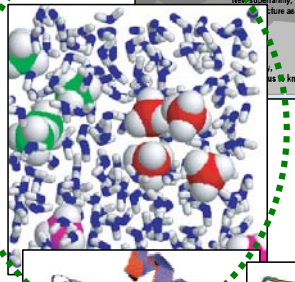
1992



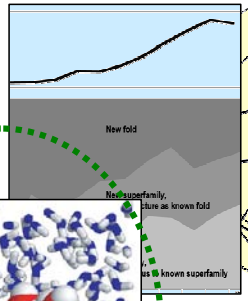


1991

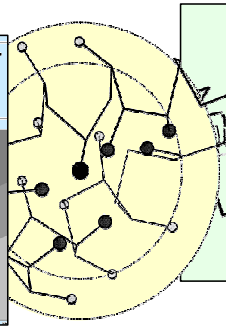
2000



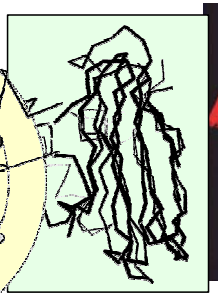
1999



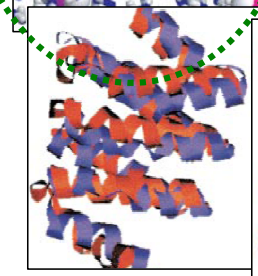
1995



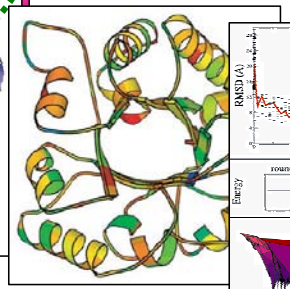
1993



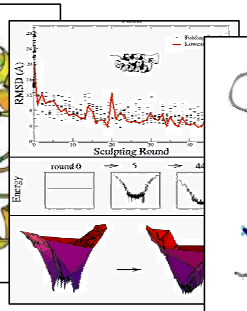
2001



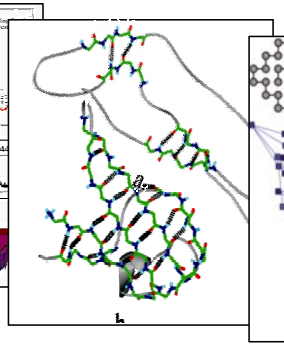
2002



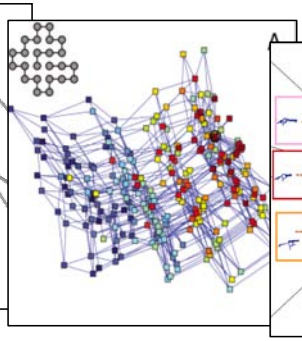
2003



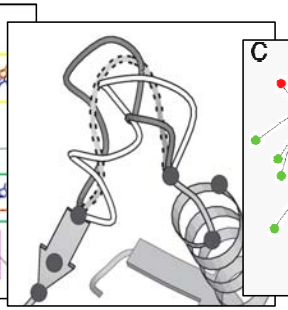
2004

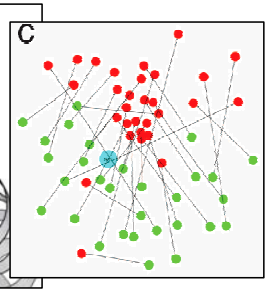


2005



2007



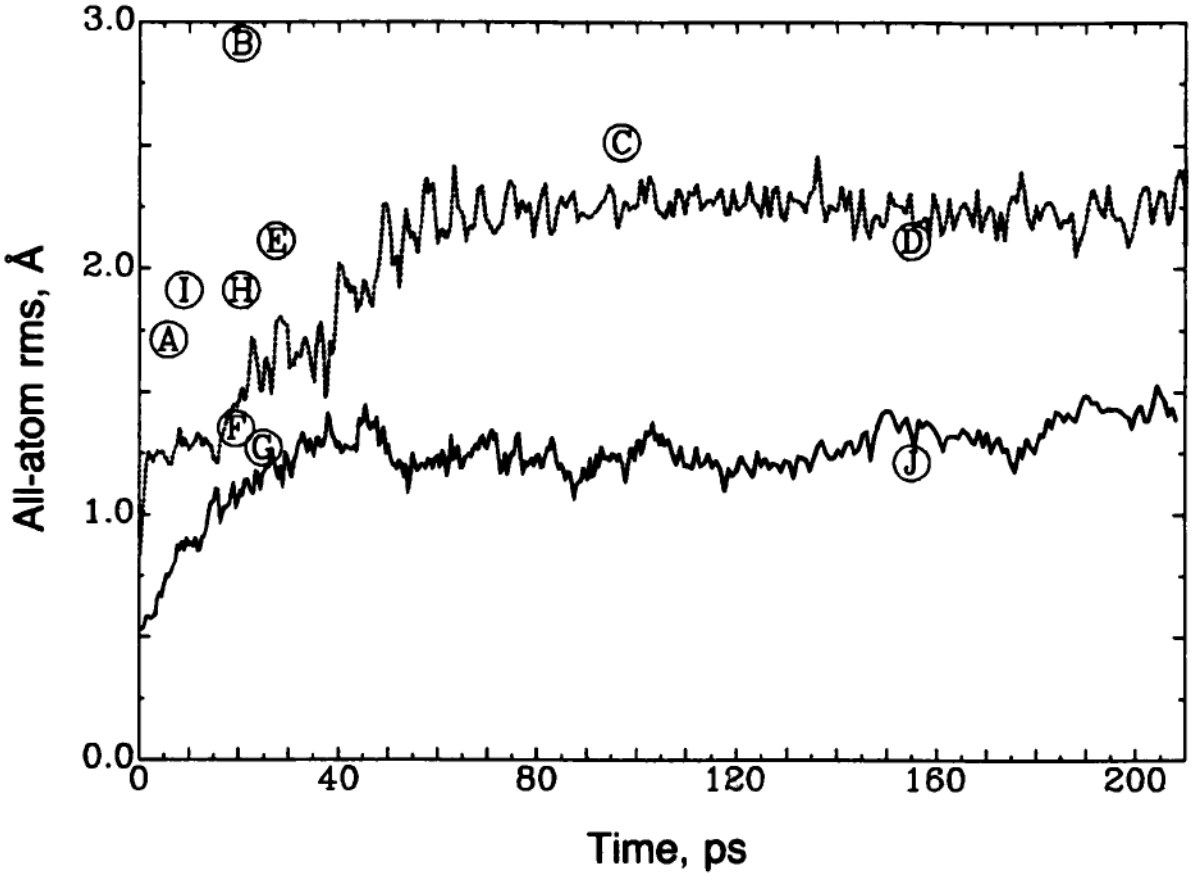


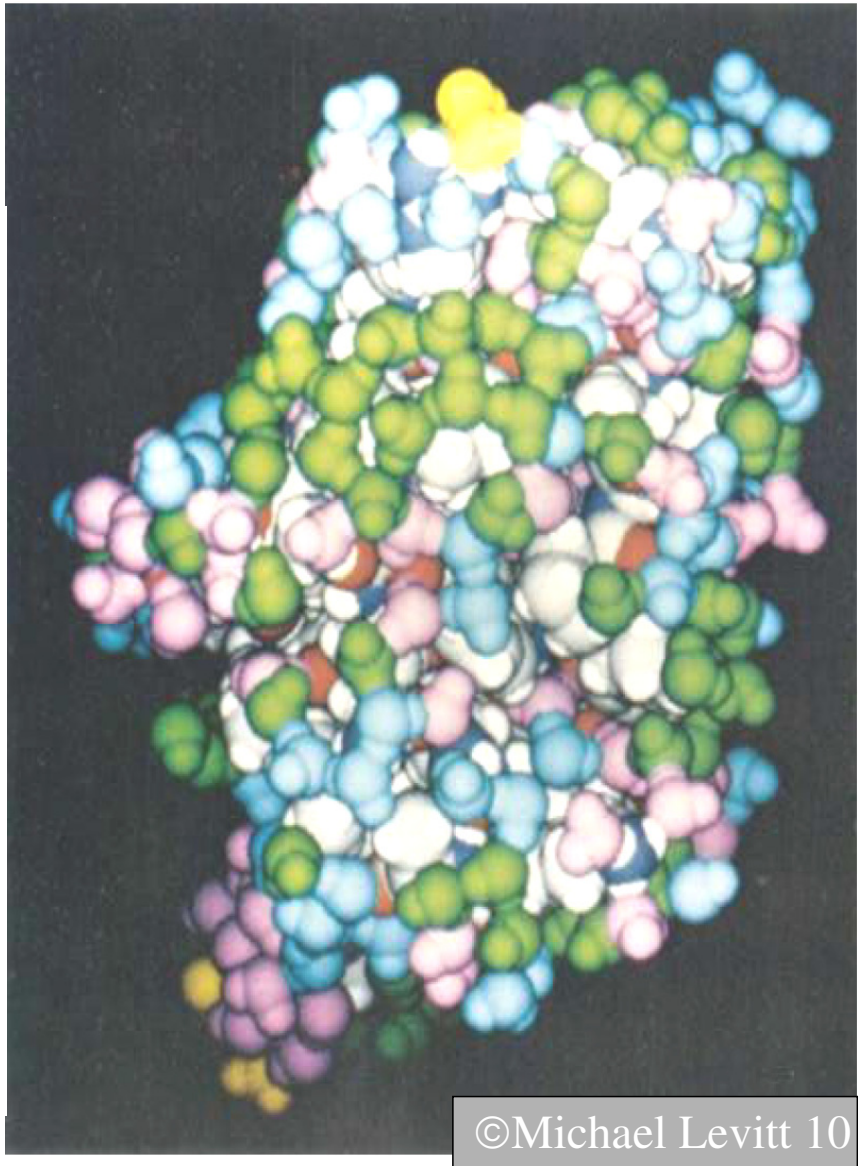
ACCURATE SIMULATION OF PROTEIN DYNAMICS IN SOLUTION

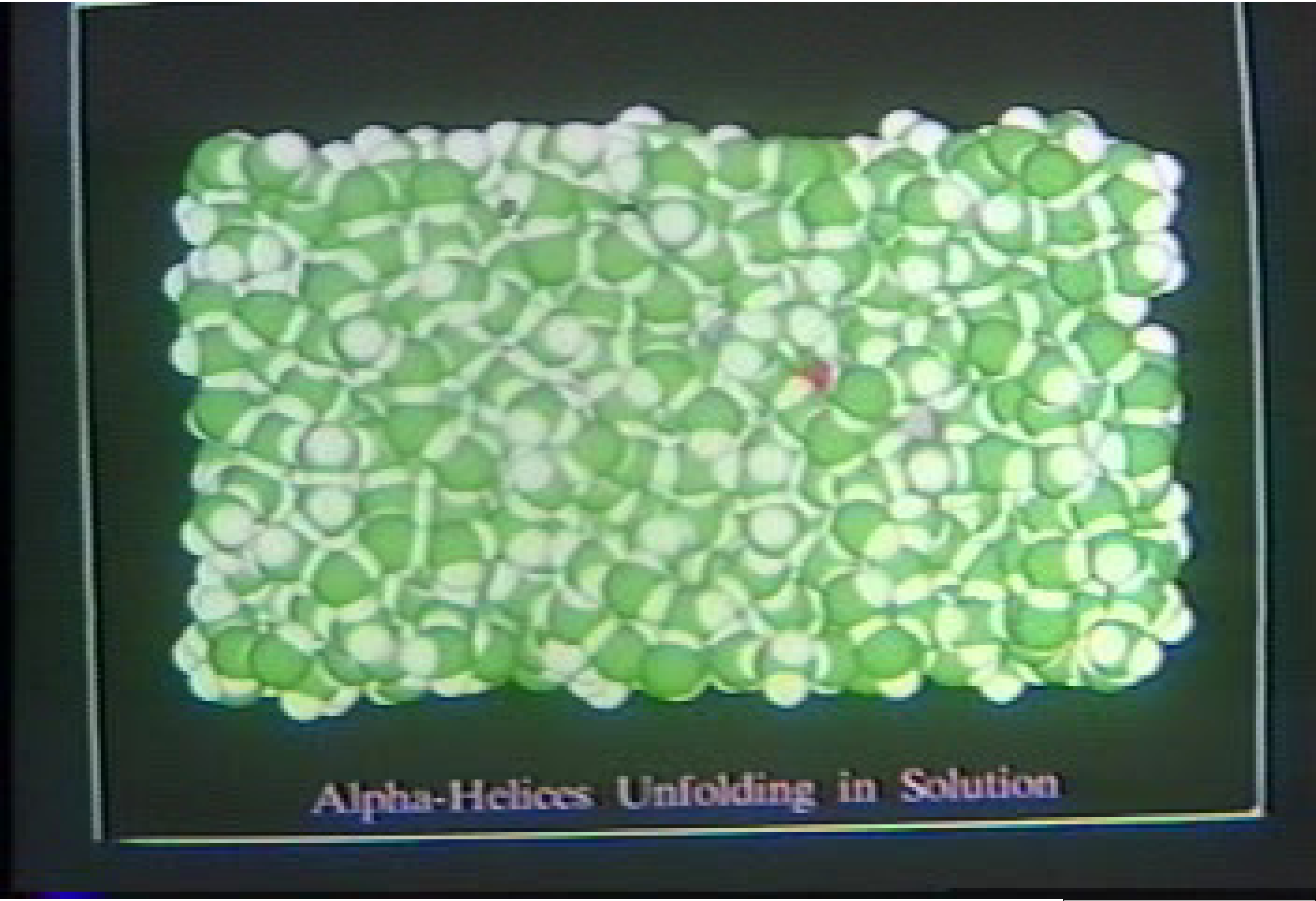
MICHAEL LEVITT* AND RUTH SHARON

Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100

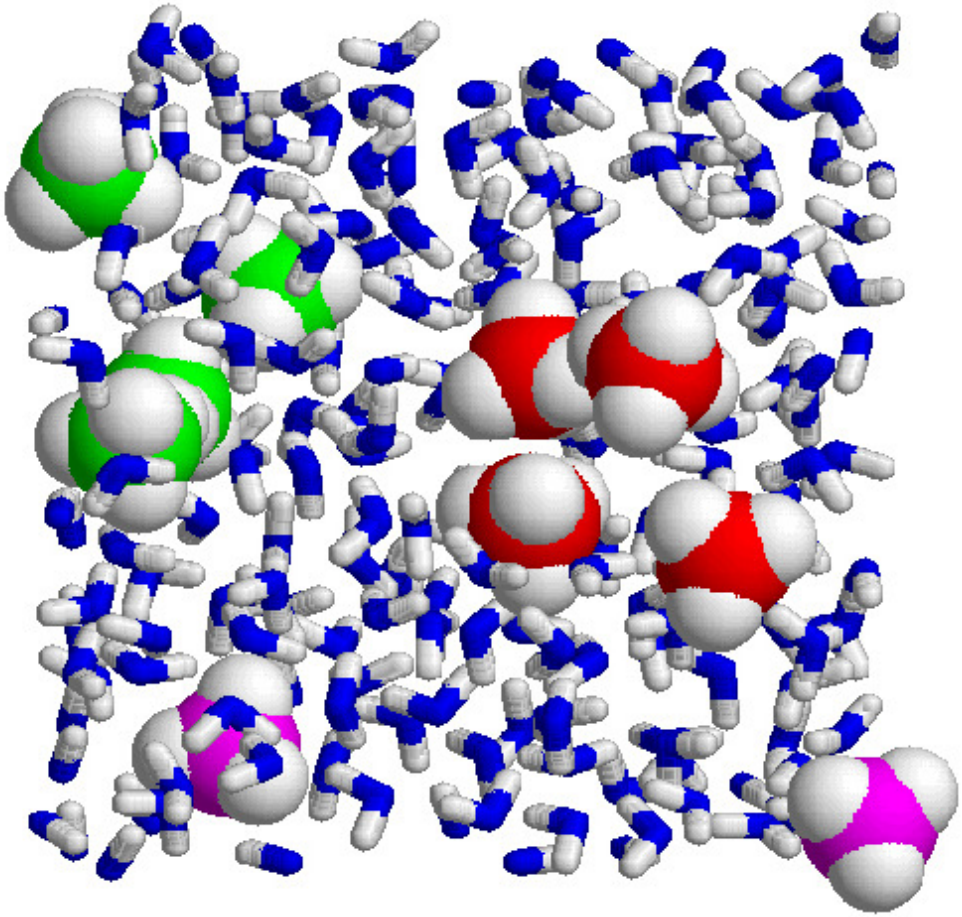
Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7557–7561, October 1988



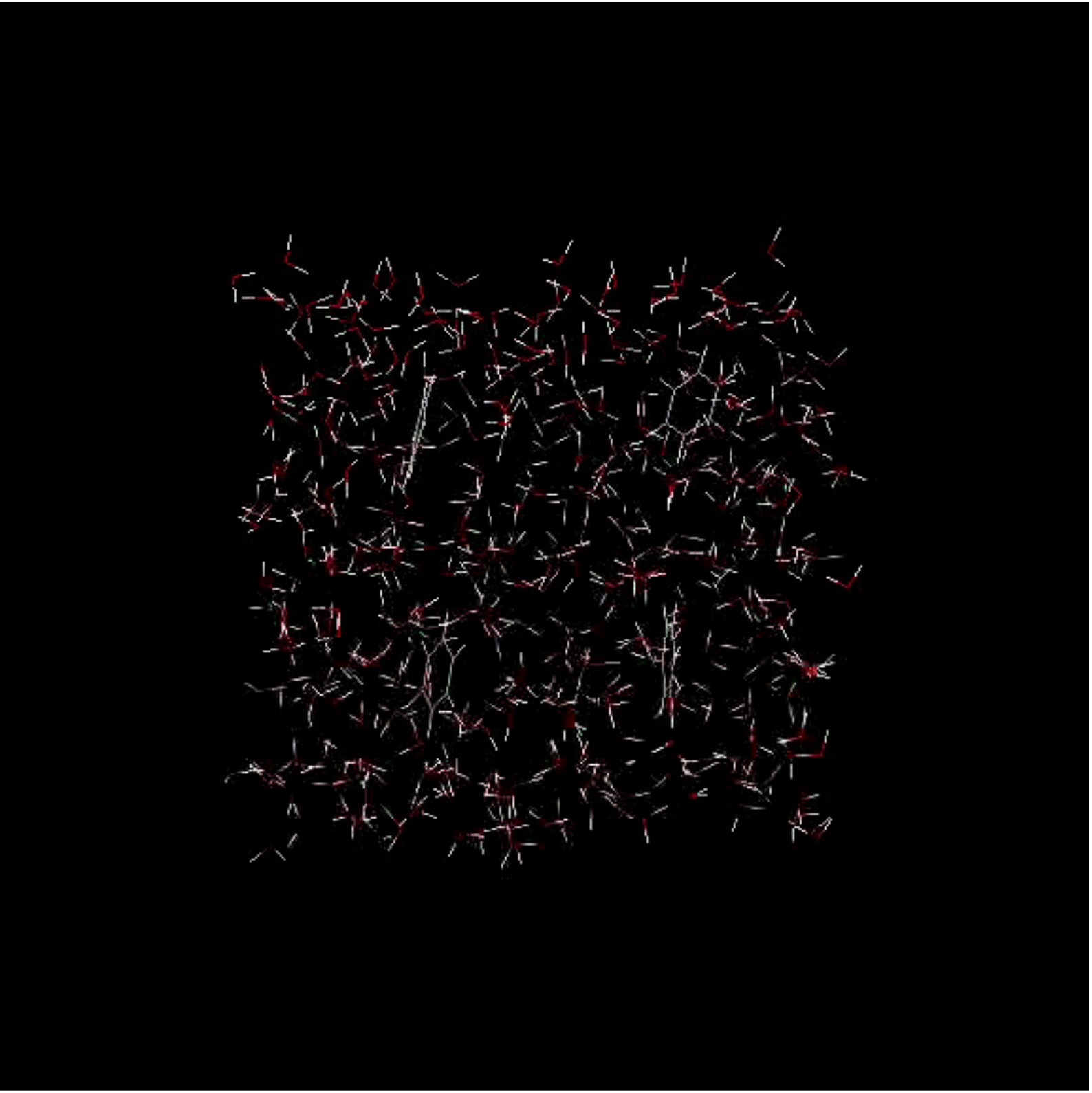




SIMULATING HYDROPHOBIC EFFECT



- 1 nanosecond MD simulations.
- Periodic water boxes.
- 30 mM to 3 Molar concentration solution.



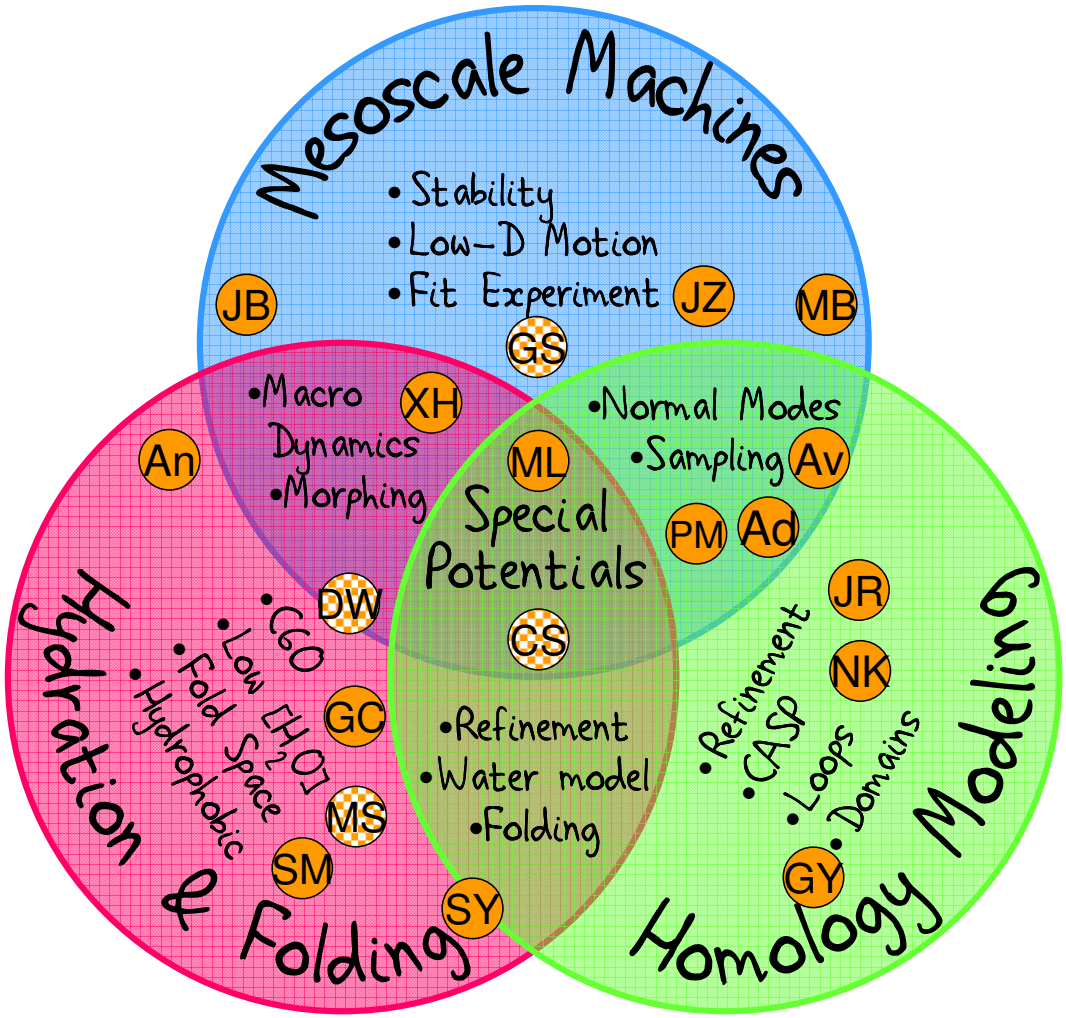
- 4 Benzenes in a periodic box of water.
- Simulate for 1 ns at 298 K.
- See clustering.

THE PRESENT & FUTURE

2008-2038

A LARGE, DIVERSE GROUP

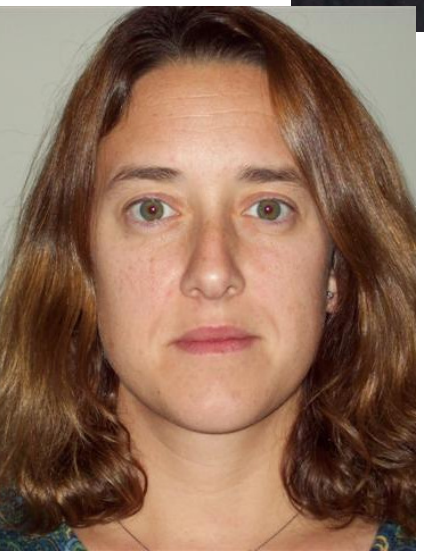
- JB Jenelle Bray
- MB Marie Brut
- GC Gaurav Chopra
- XH Xuhui Huang
- NK Nir Kalisman
- PM Peter Minary
- SM Sergio Moreno
- JR João Rodrigues
- Av Avraham Samson
- An Andrea Scaiewicz
- GS Gunnar Schroeder
- Al Alena Shmygelska
- Ad Adeline Sim
- CS Chris Summa
- MS Michael Sykes
- DW Dahlia Weiss
- SY Simon Ye
- GY Golan Yona
- JZ Junjie Zhang

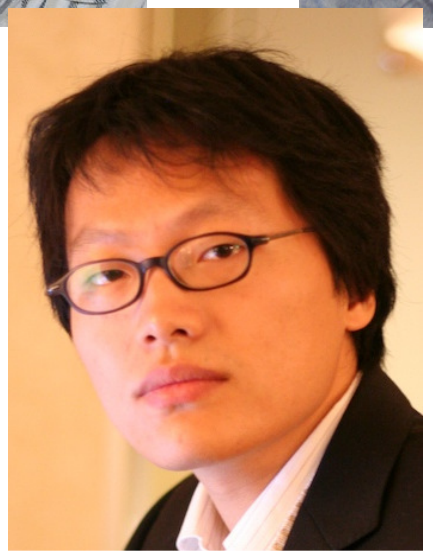








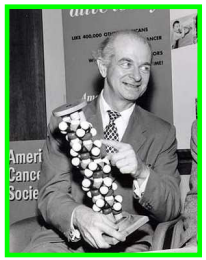


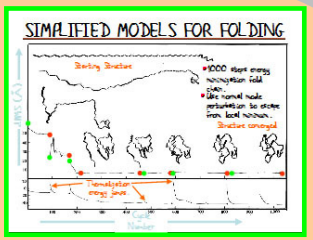

FUTURE

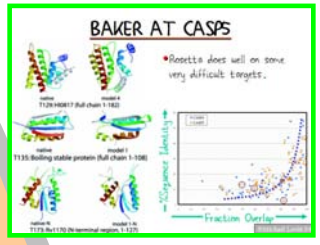
Modeling

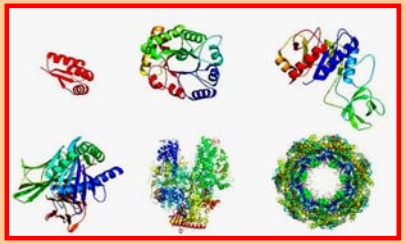
Experiment

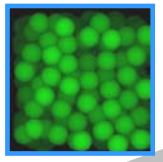
Simulation

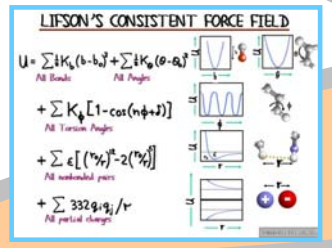


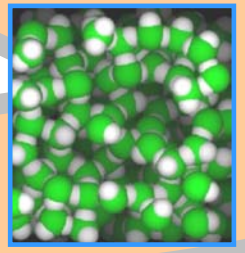


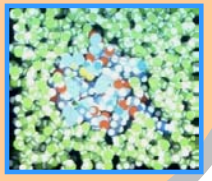


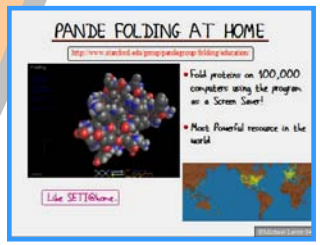












Models for Evolution

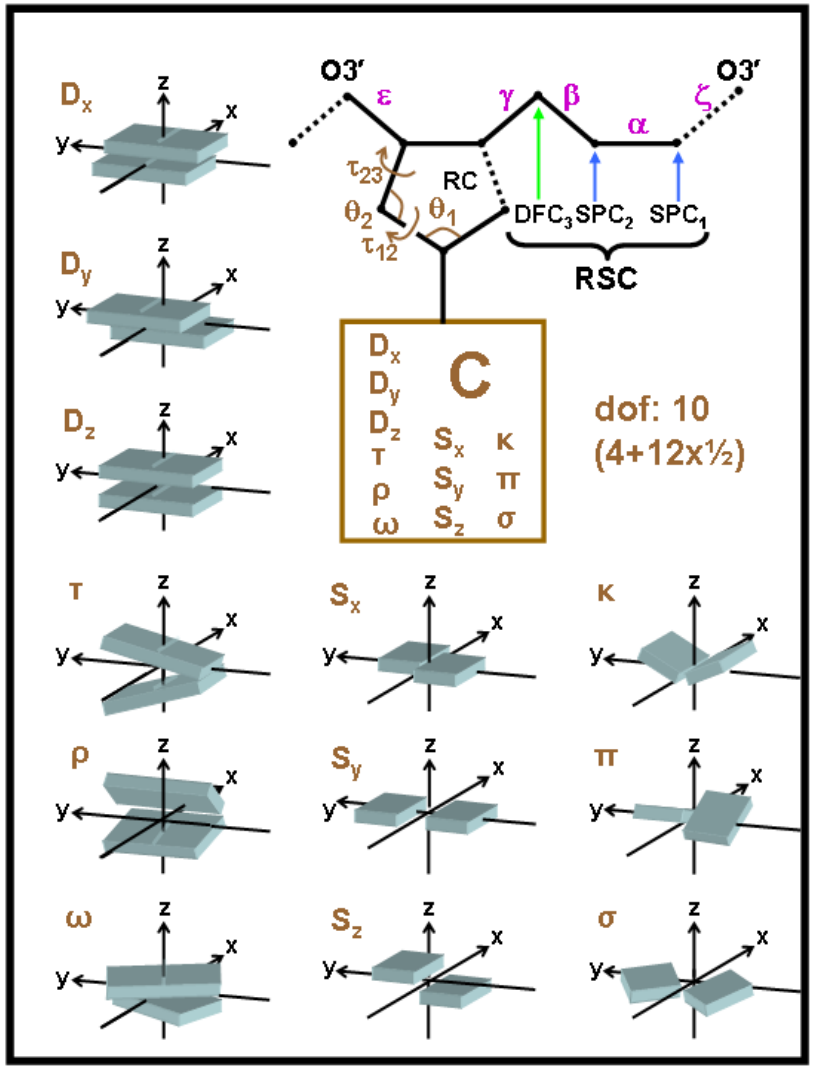
Large-Scale Structure

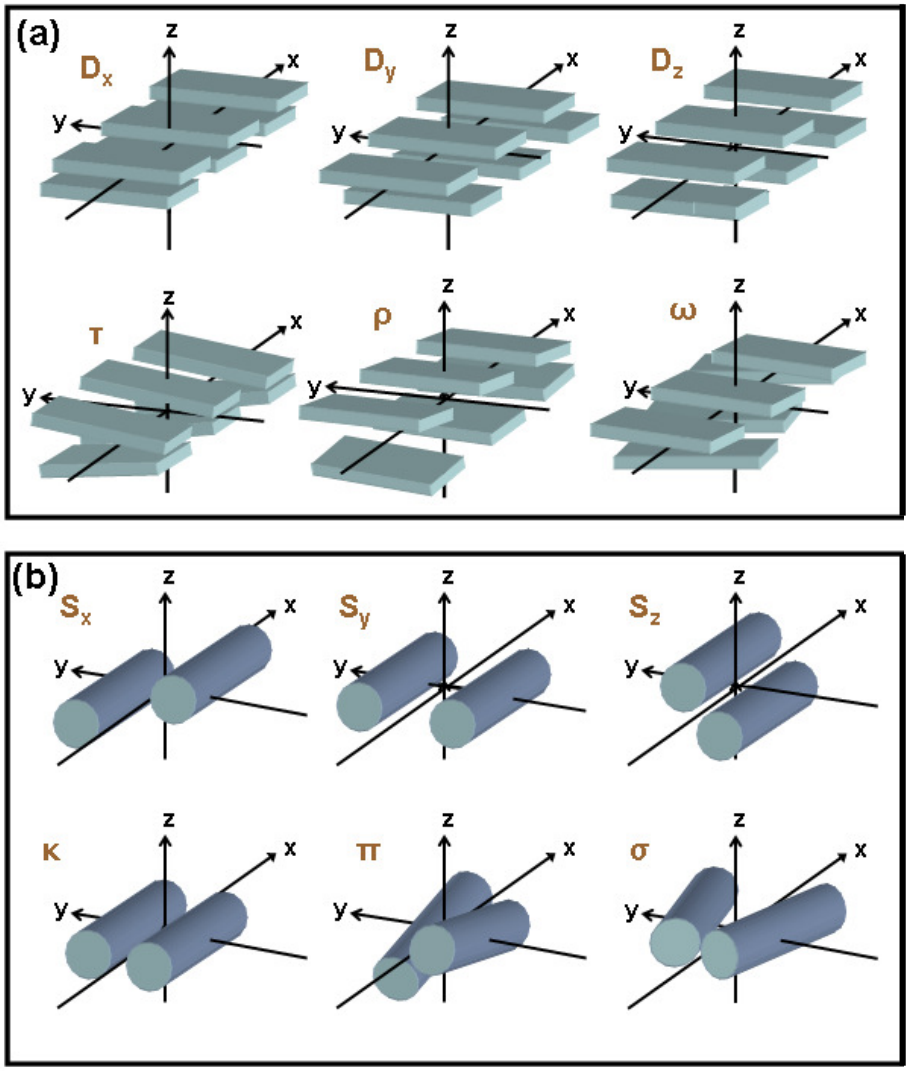
Accurate Simulation

1950 1960 1970 1980 1990 2000 2010

Peter Minary

NATURAL VARIABLES for DNA & PROTEINS





S_x : Shear
S_y : Stretch
S_z : Stagger

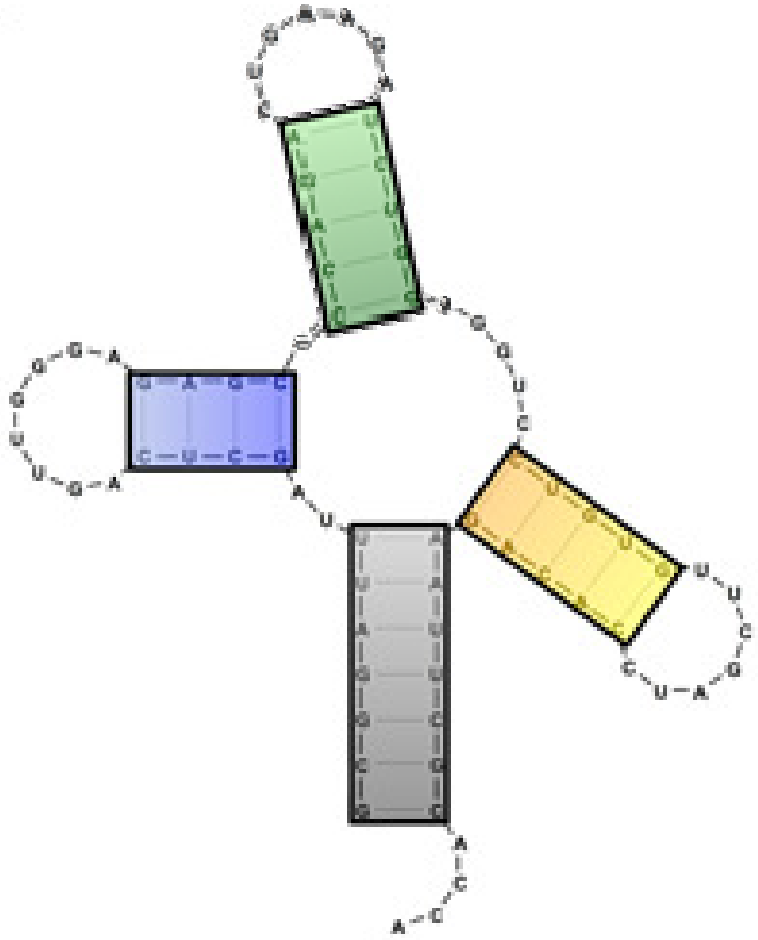
κ : Buckle
π : Propeller
σ : Opening

D_x : Shift
D_y : Slide
D_z : Rise

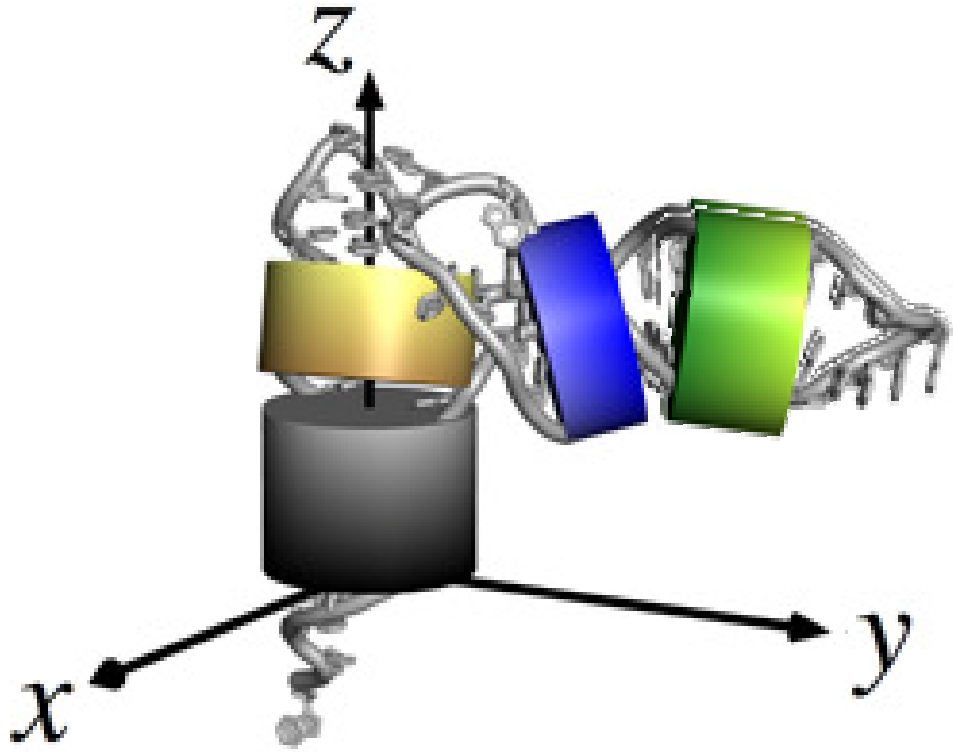
τ : Tilt
ρ : Roll
ω : Twist

NATURAL MOVES FOR RNA MODELING HARD

A

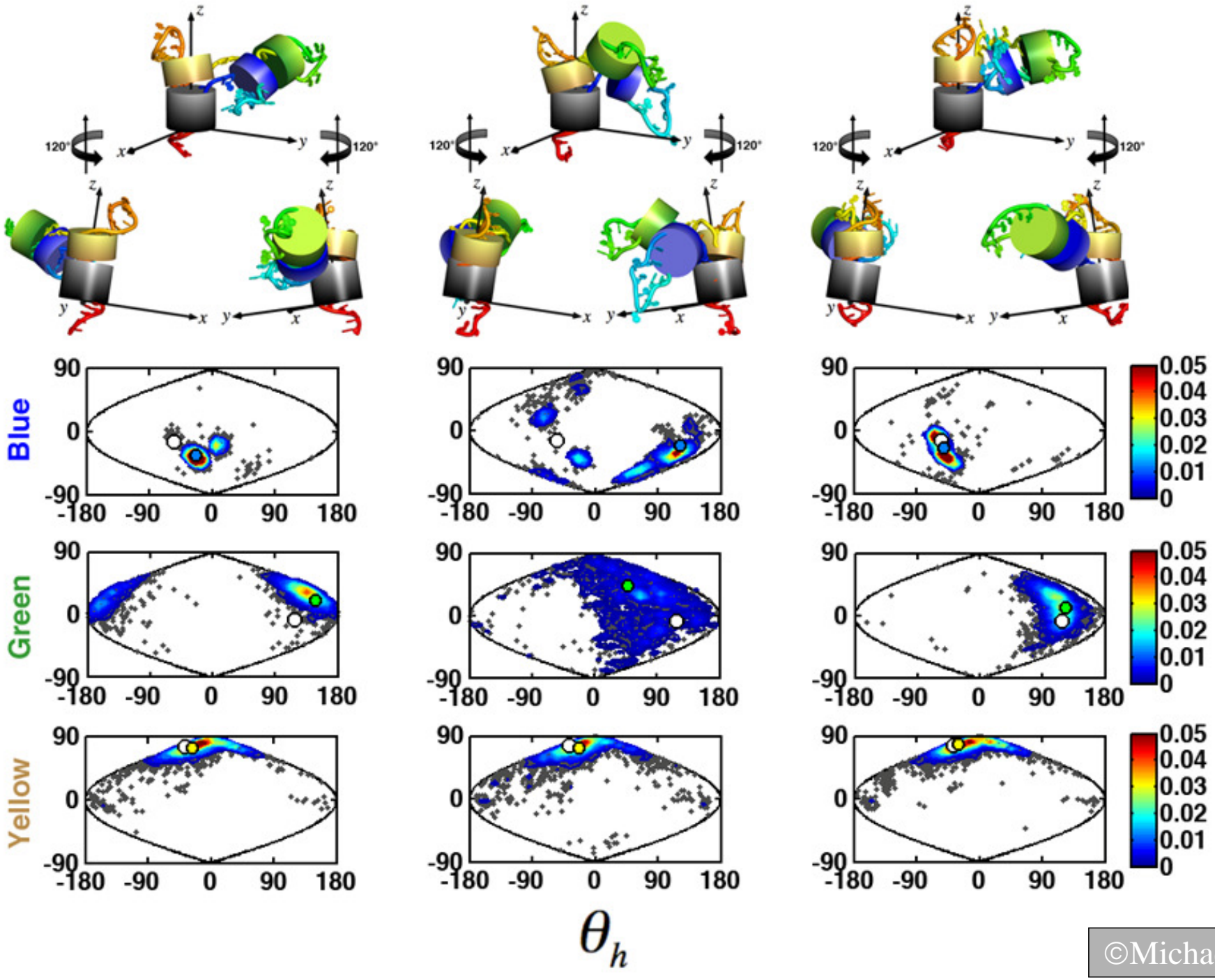


B



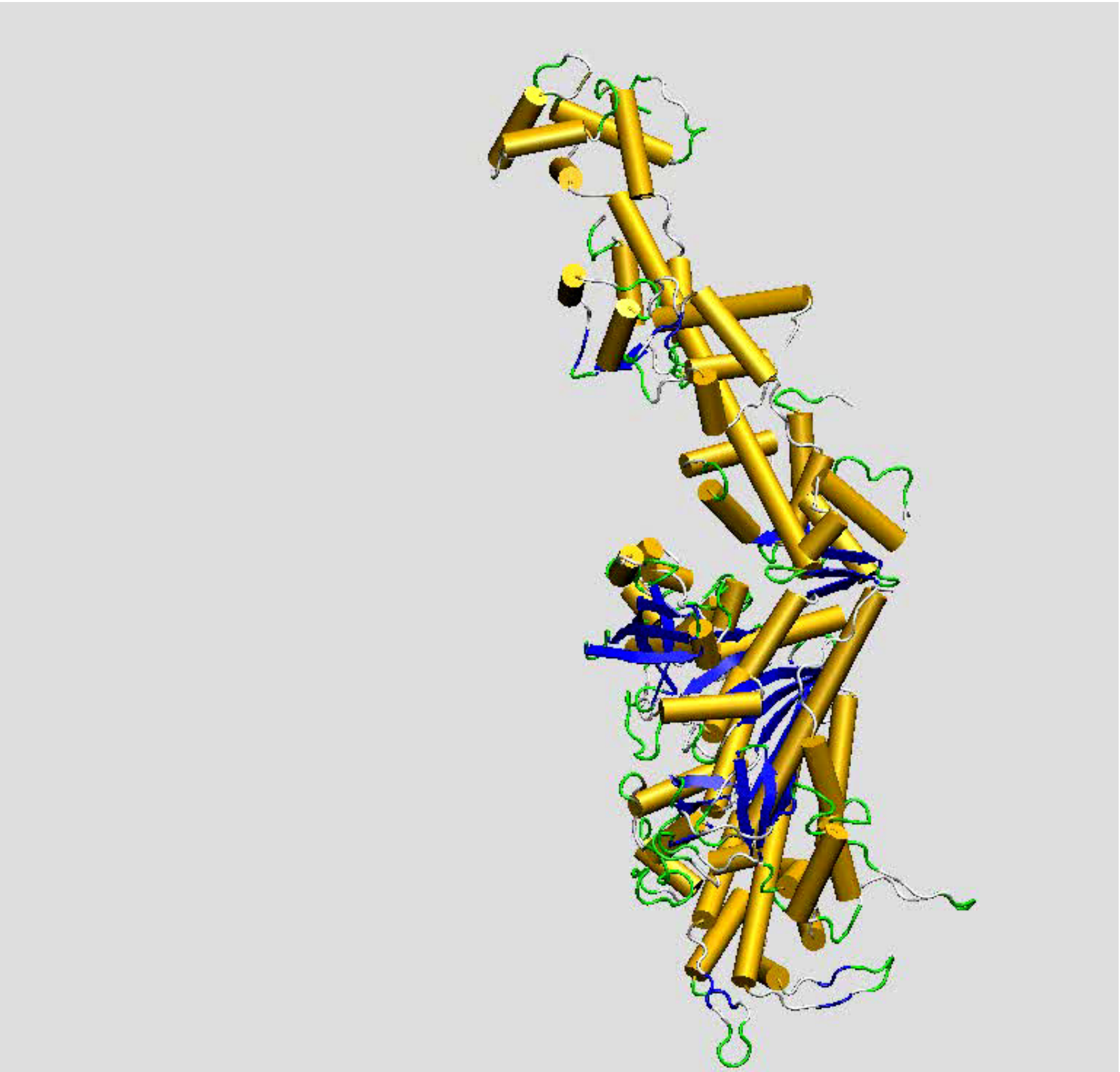
Adelene Sim

SYSTEMATIC SEARCH OF RNA STRUCTURES



Dahlia Weiss

MYOSIN MORPHING



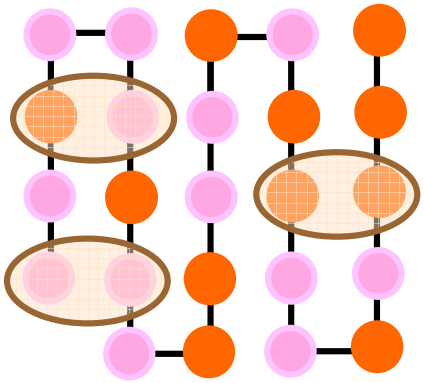
Sergio Moreno

Xia Yu

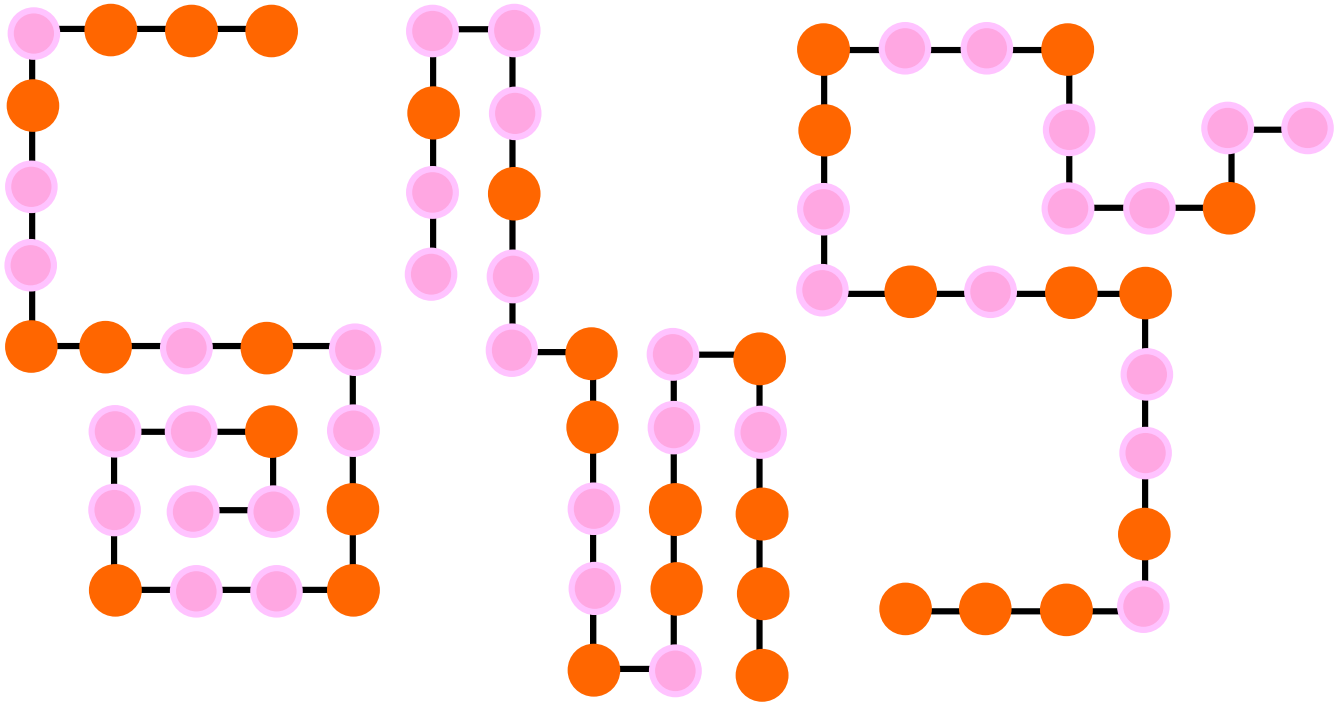
TOY MODELS OF PROTEIN UNIVERSE

- Take one sequence and try each shape.

PPHPPPHPPHHPPHPHHPPHPHHH

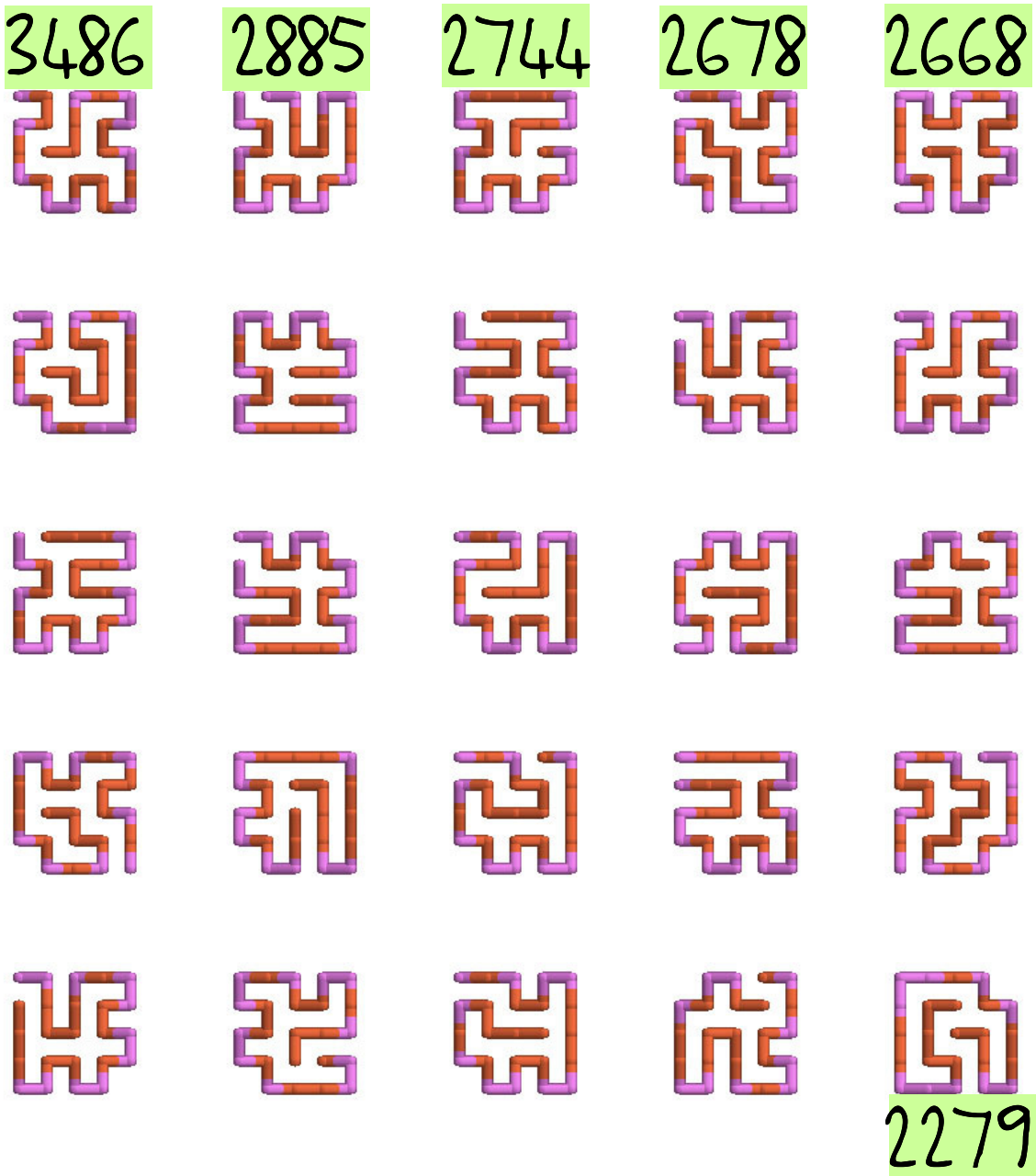


- Calculate the energy of each sequence on each shape: $E_{HH}=-2$, $E_{HP}=E_{PP}=-1$.



- Each sequence selects the shape or shapes that are most stable for it.

WHAT IS SPECIAL ABOUT SELECTED SHAPES?

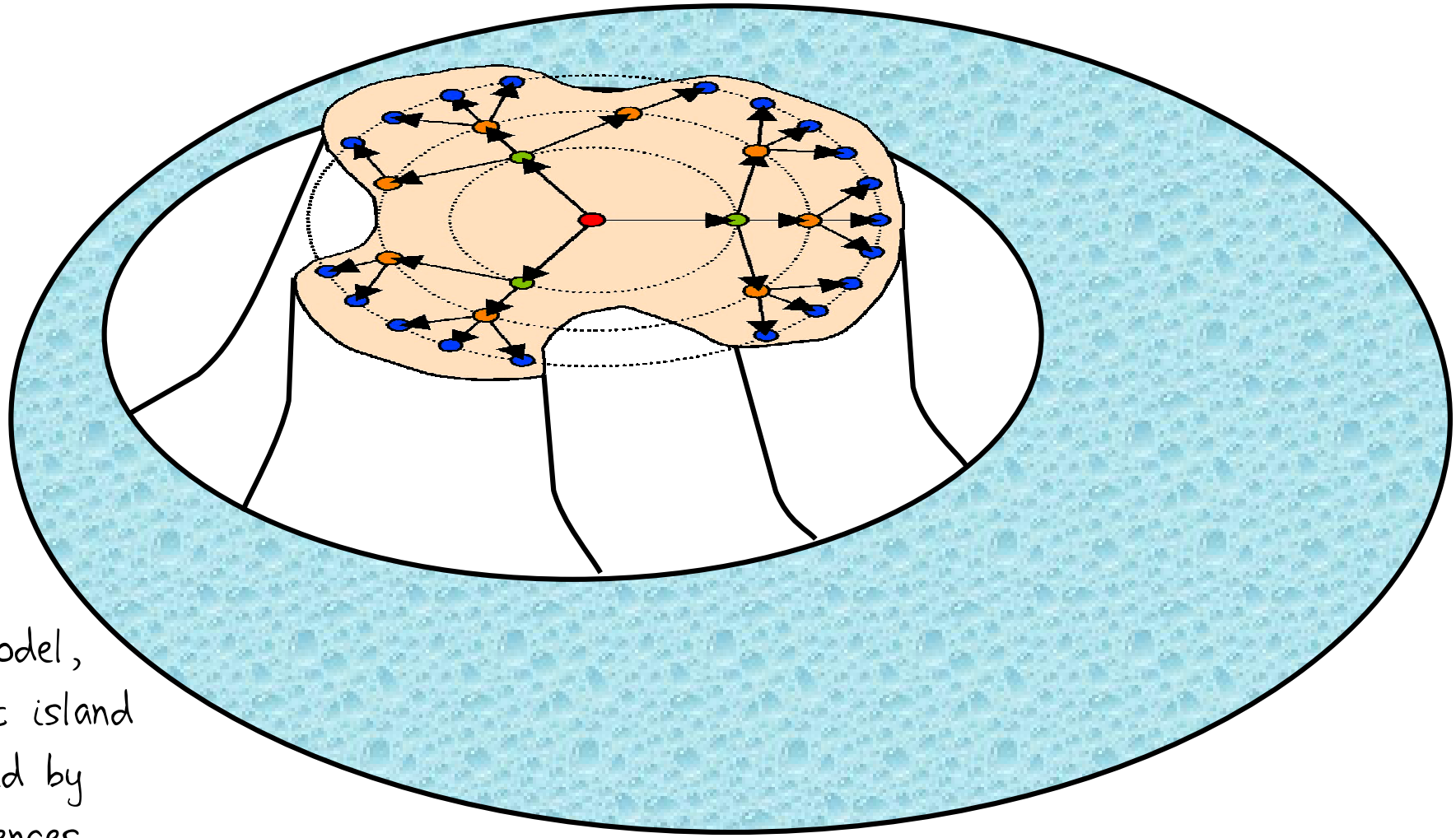


Folds used by many sequences have high designability.

At least 2279 sequences favor each of the folds shown here.

AN ISLAND OF FUNCTION

- Proteins with sequences on the island have the needed function. They can multiply and evolve.

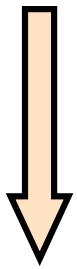


For our model, the biggest island is populated by 3486 sequences.

EVOLVE BY MUTATION

- Change any one amino acid at random

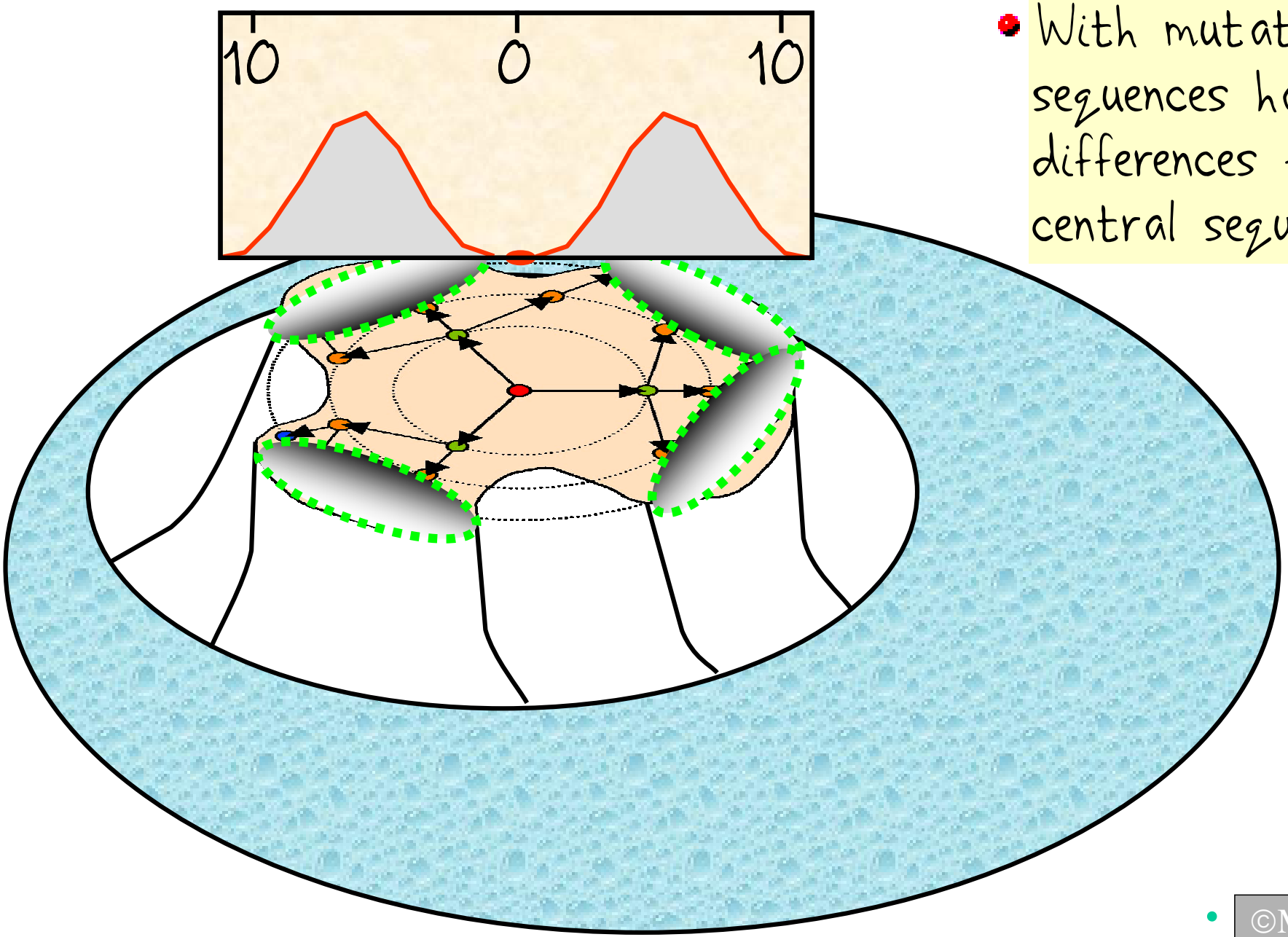
PPHPPPHPPHHPPHPHHPPHPHHH



PPHPHPHPPHHPPHPHHPPHPHHH

Clonal Reproduction

MUTATION POPULATES EDGE OF ISLAND



- With mutation sequences have many differences from central sequence.

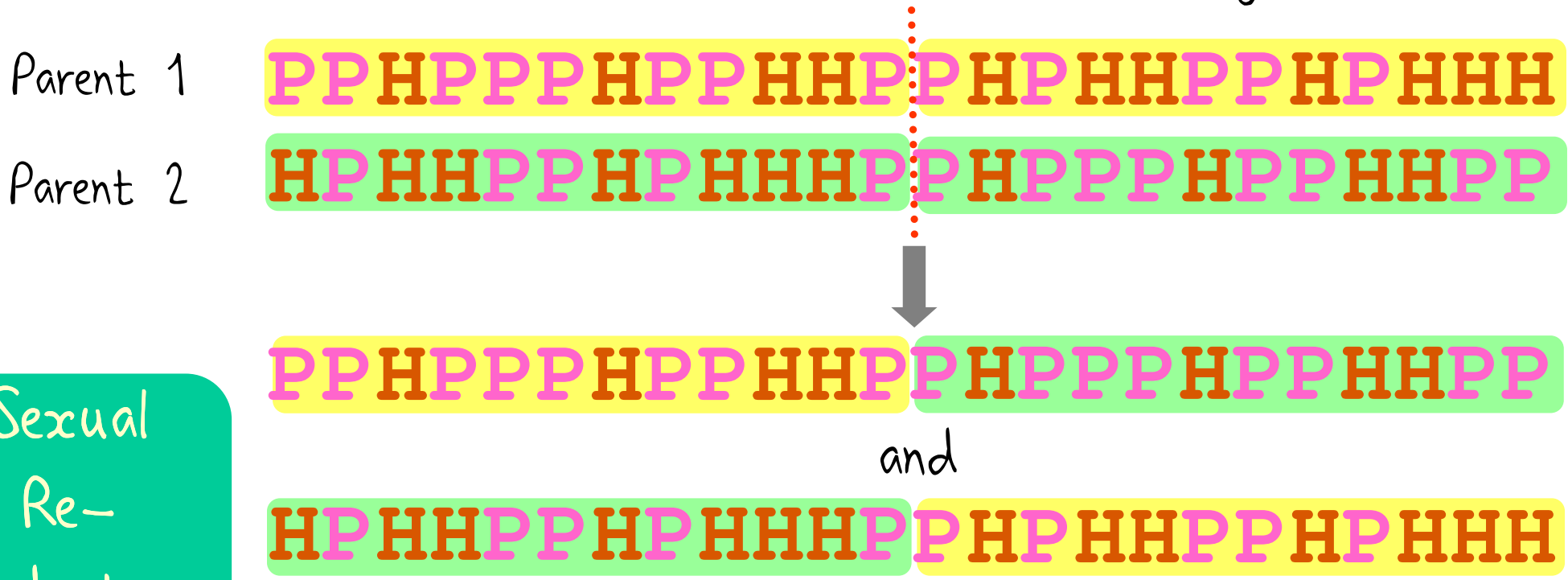
EVOLVE BY RECOMBINATION

- Choose two parents at random from population

PPHPPPHPPHHPPHPHHPPHPHHH

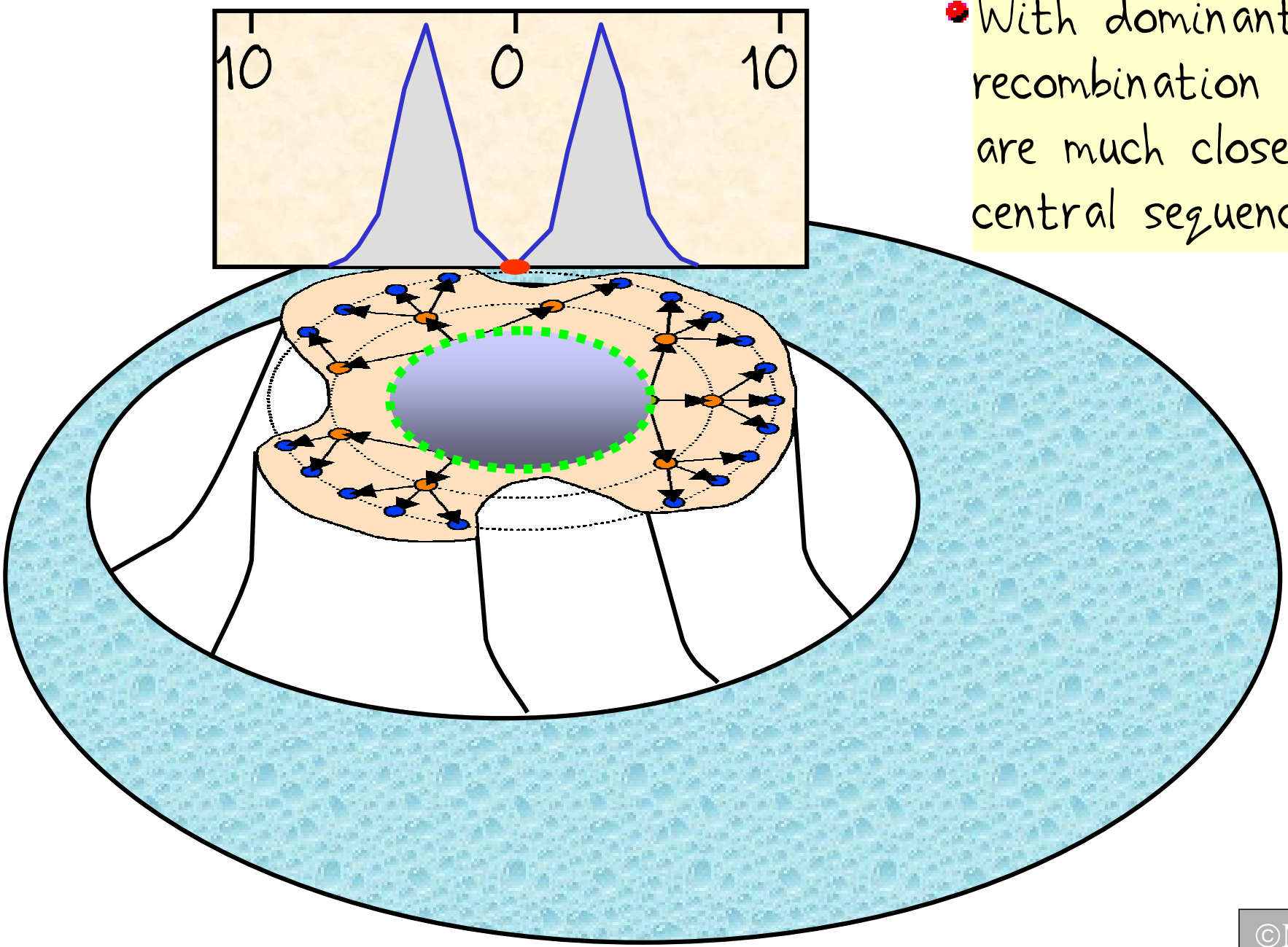
HPHHPPHPHHHPPHPPPHPPHHPP

- Choose a random cross-over point and merge



Sexual Re-production

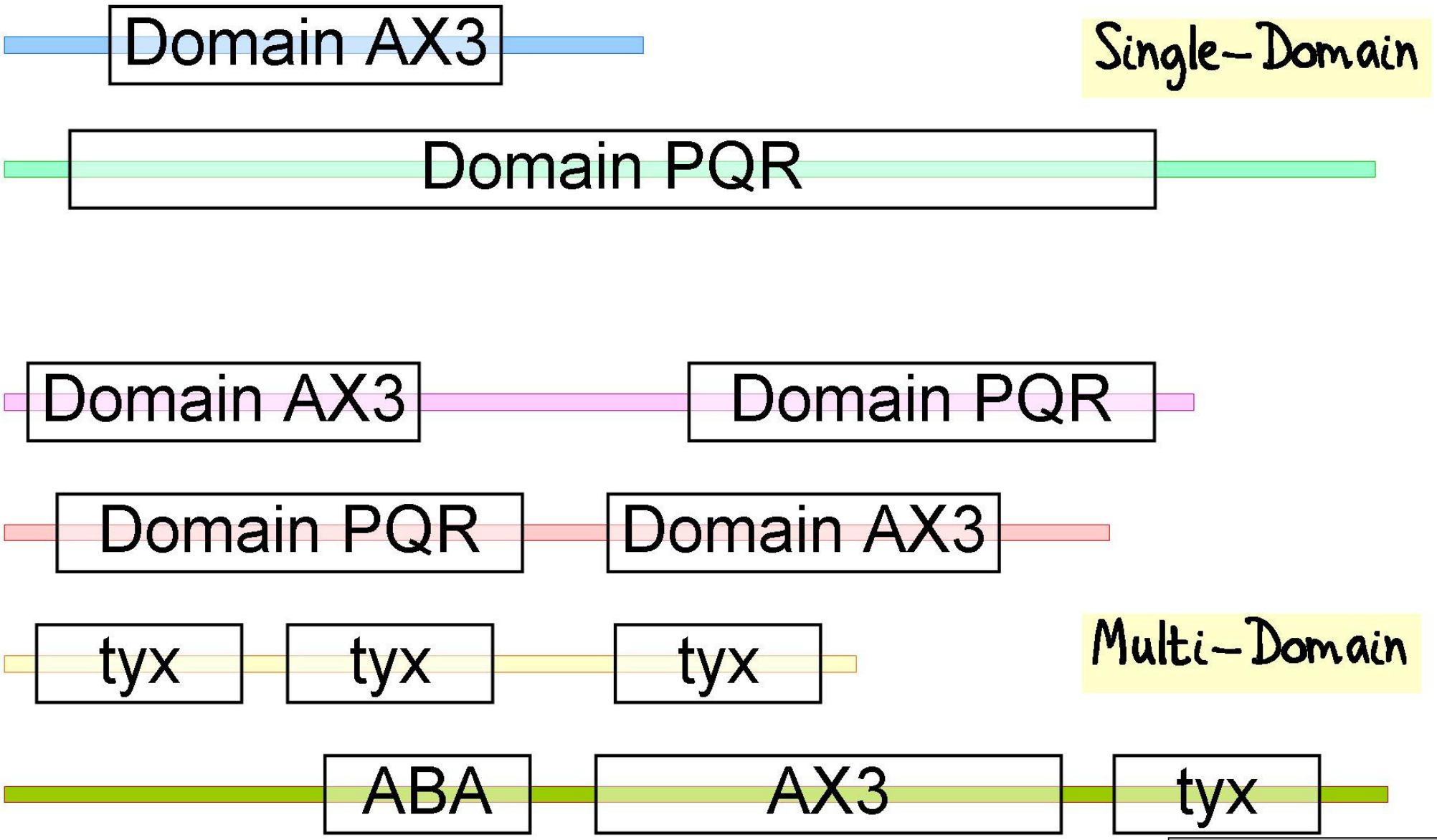
RECOMBINATION POPULATES MIDDLE OF ISLAND



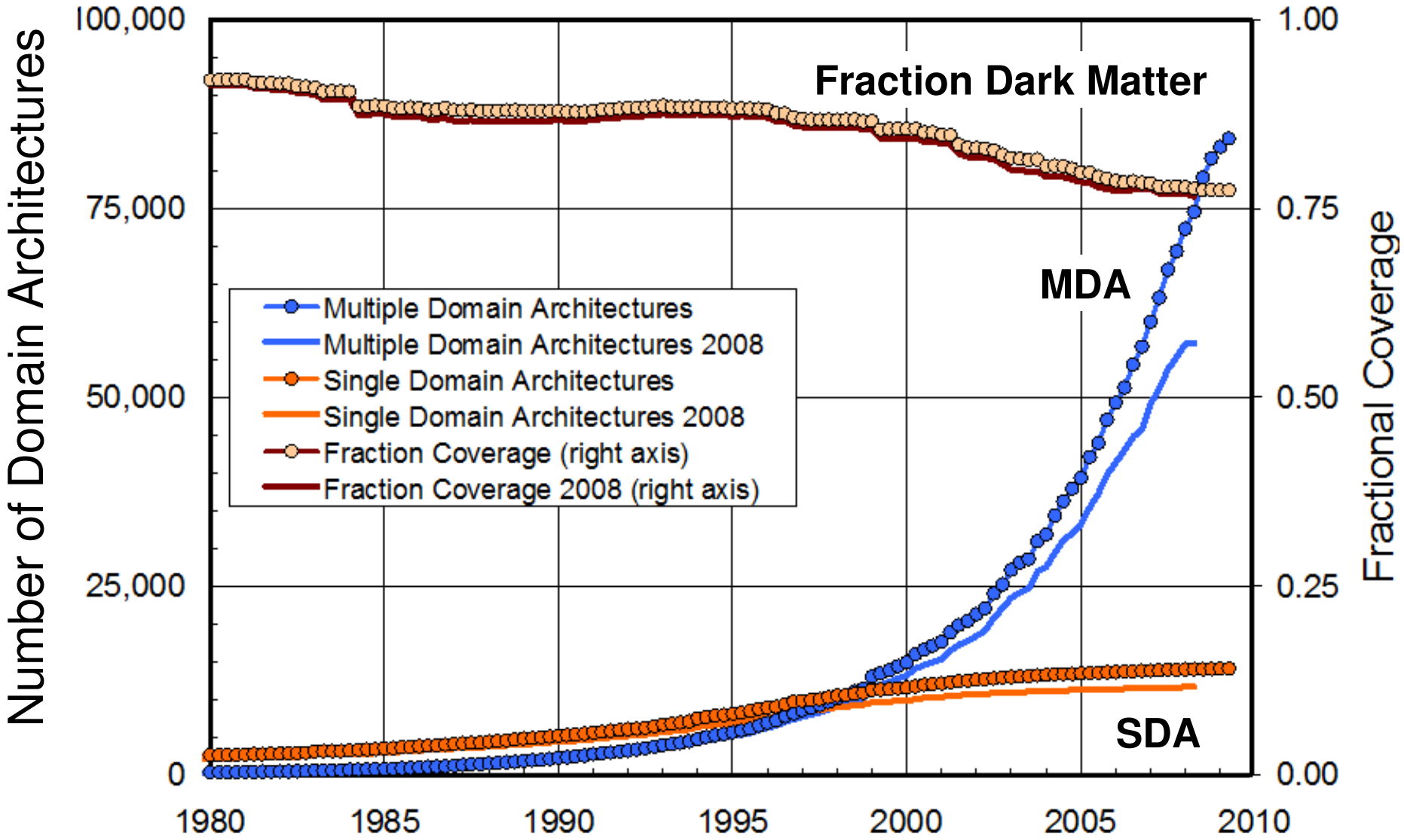
- With dominant recombination sequences are much closer to the central sequence.

My own work...

SINGLE AND MULTI-DOMAIN SEQUENCES



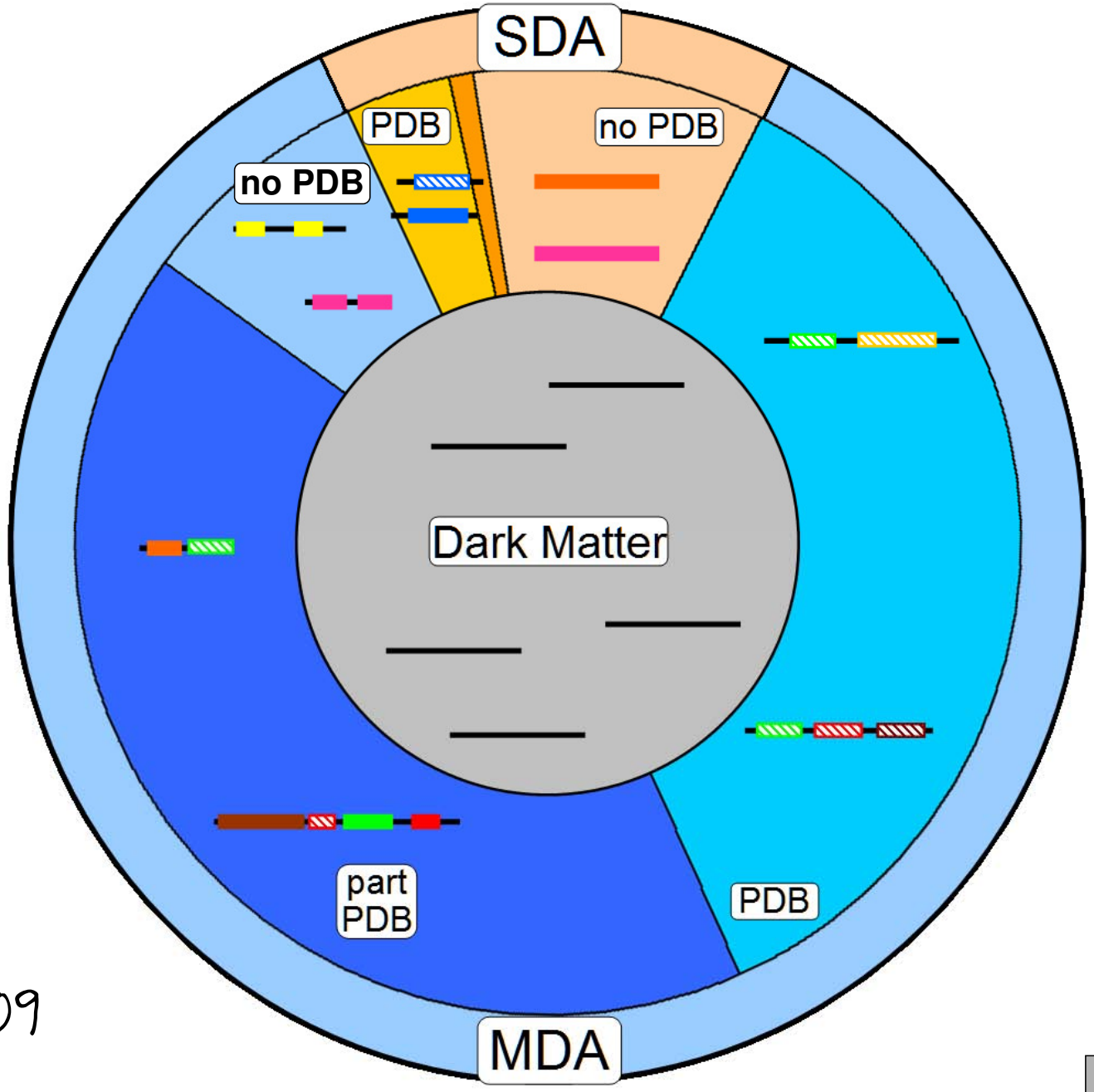
EXPONENTIAL MULTI-DOMAIN GROWTH WITH TIME



MDAs are growing rapidly while SDAs hardly change.

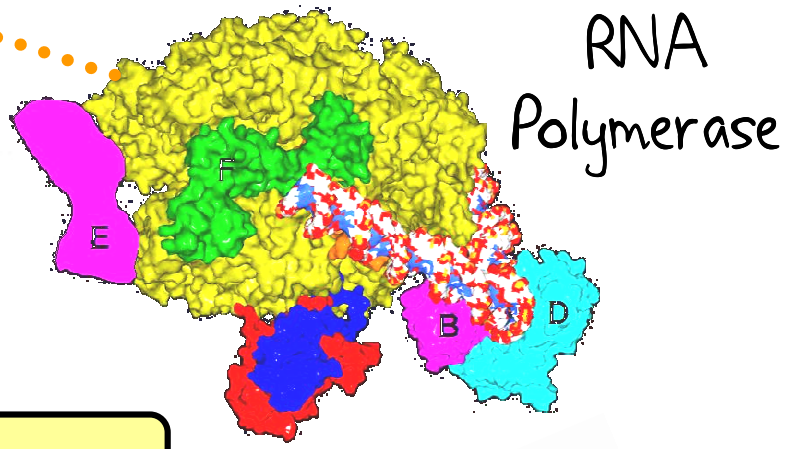
Sequence Coverage is high and increasing.

UNIQUE UNIVERSE



PNAS, 2009

THE CORE MACHINERY OF LIFE



Package DNA

DNA

Polymerase

RNA

Fold Protein

Protein

Nucleotides

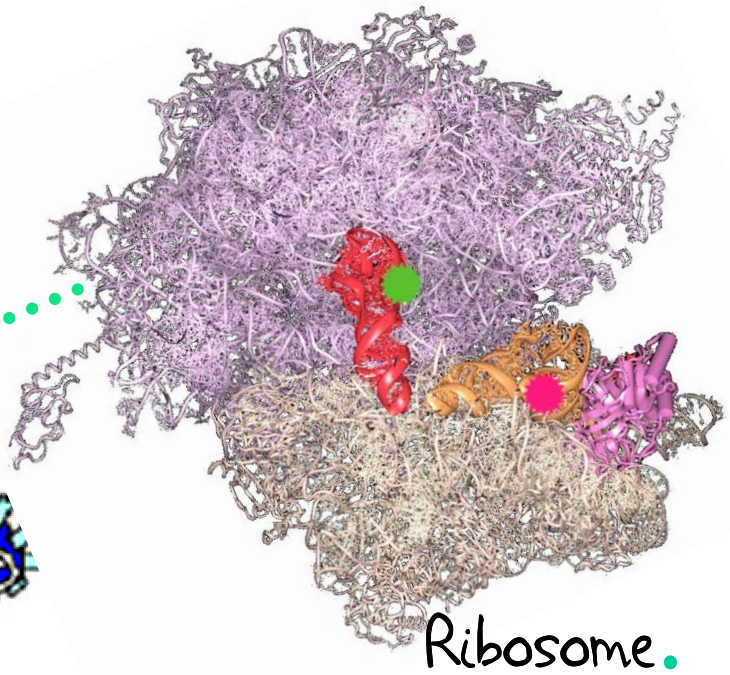
Make RNA

Amino
acid

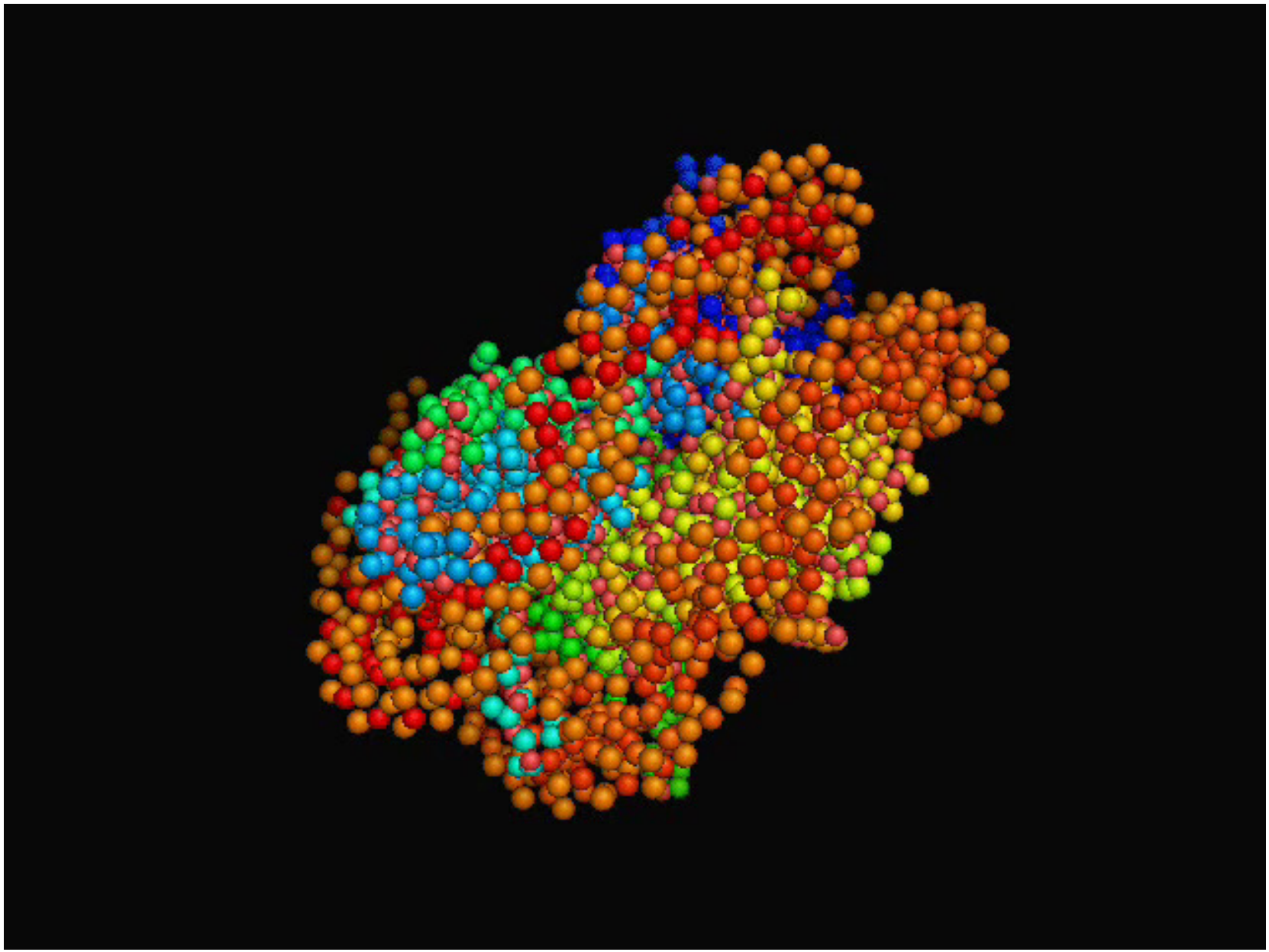
mRNA

Ribosome

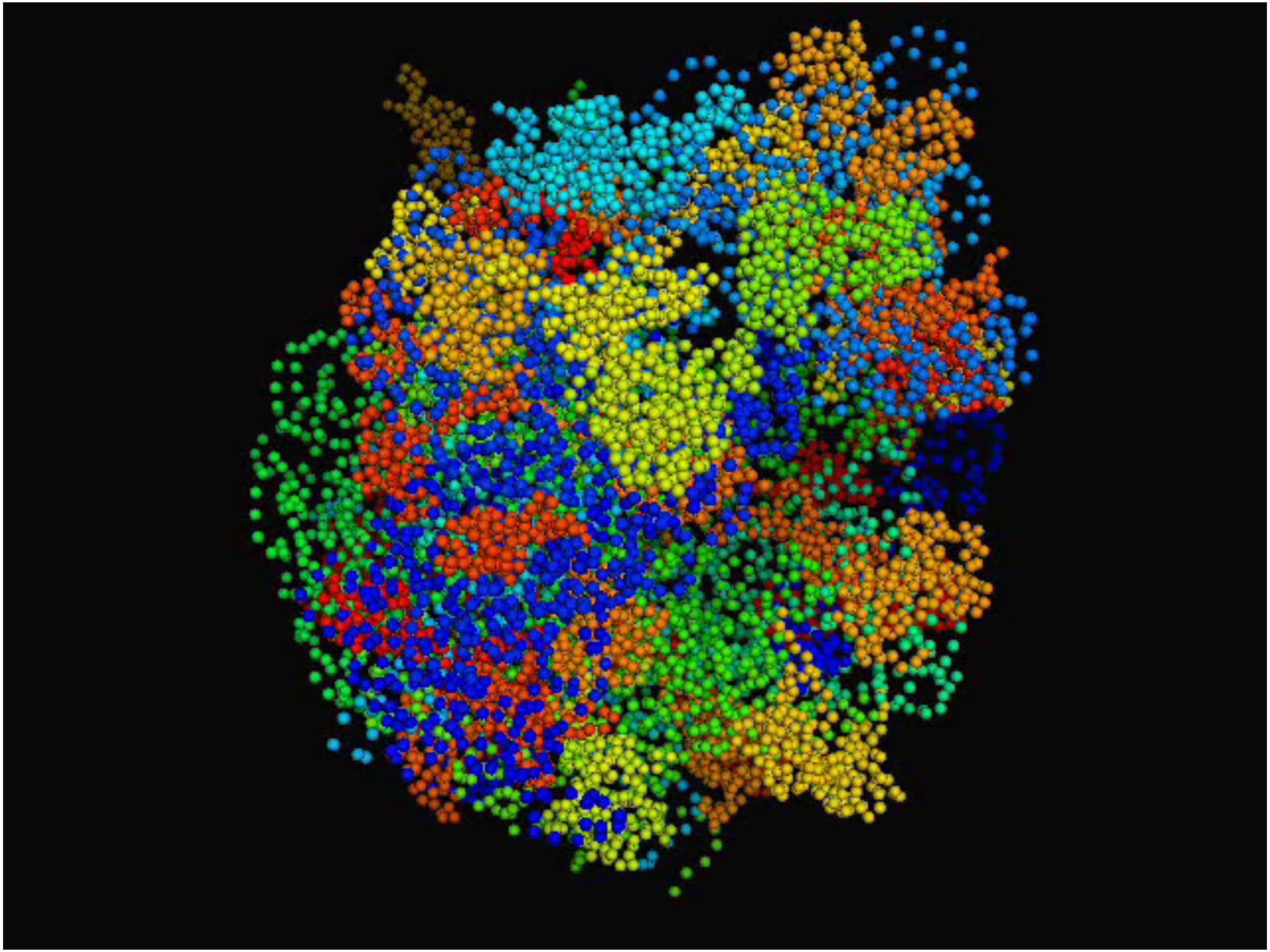
Make Protein



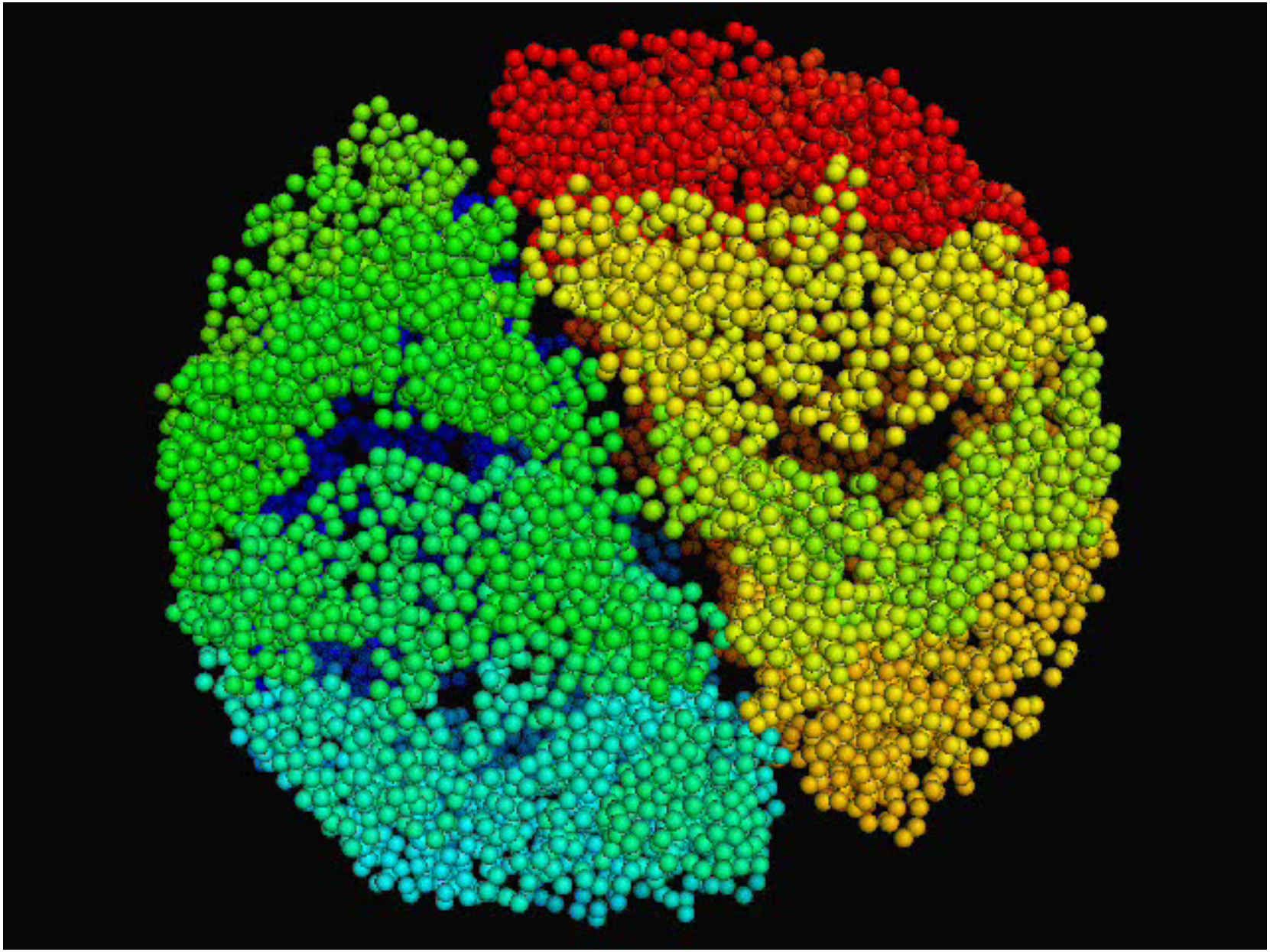
NUCLEOSOME NORMAL MODE 2



70S RIBOSOME NORMAL MODE 2



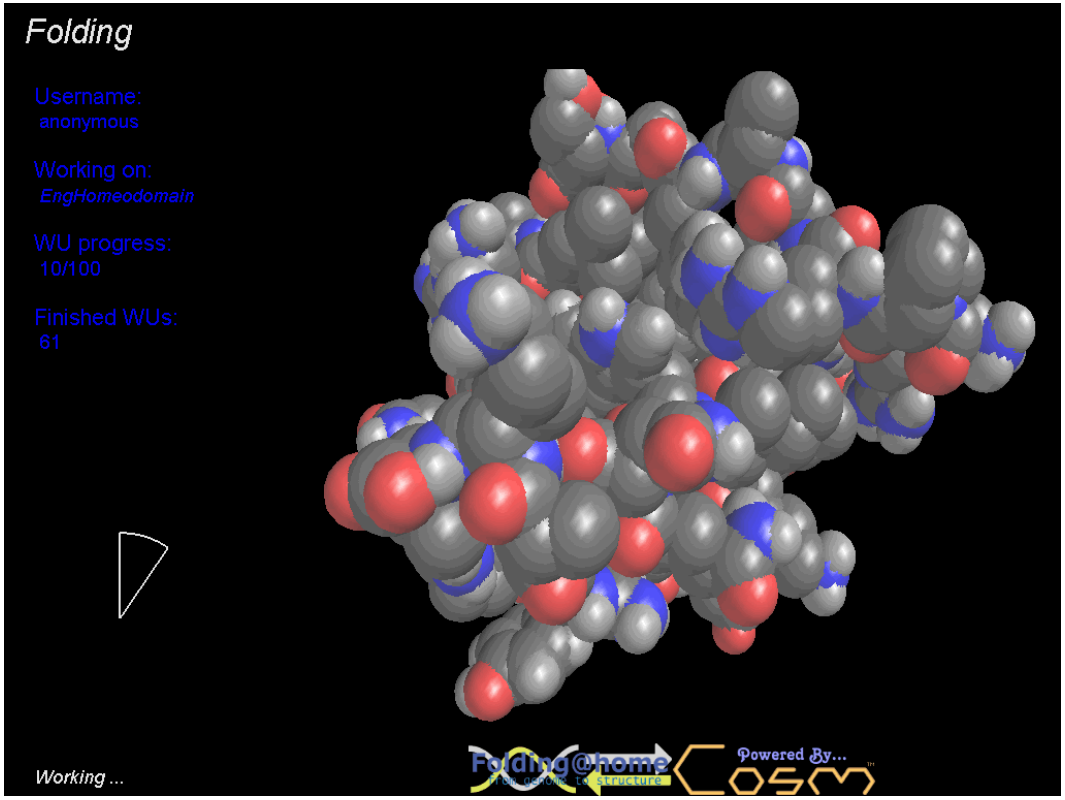
Chaperonin Eu NORMAL MODE 1



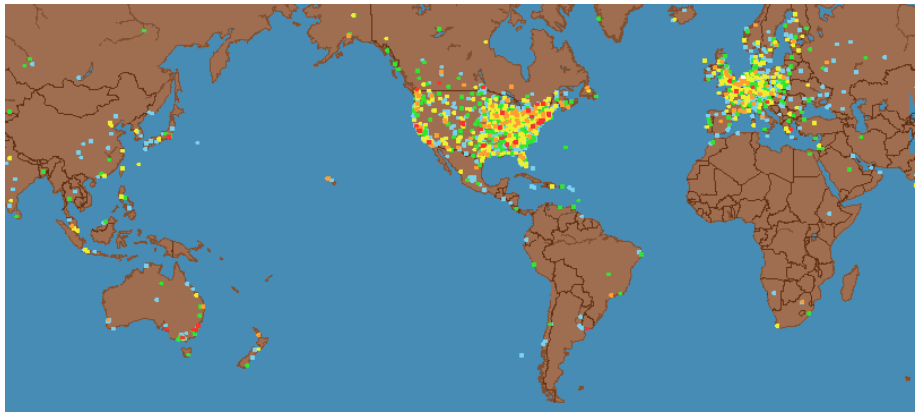
Novel
Computational
Machinary

PANDE FOLDING AT HOME

http://www.stanford.edu/group/pandegroup/folding/education/



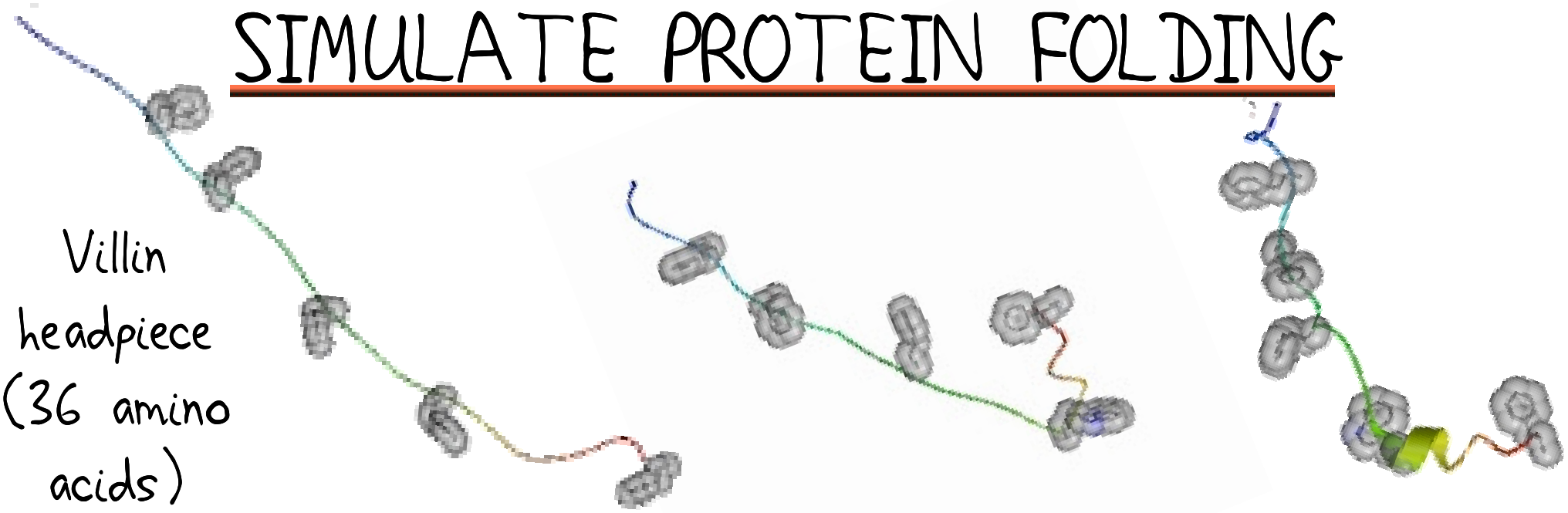
- Fold proteins on 100,000 computers using the program as a Screen Saver!

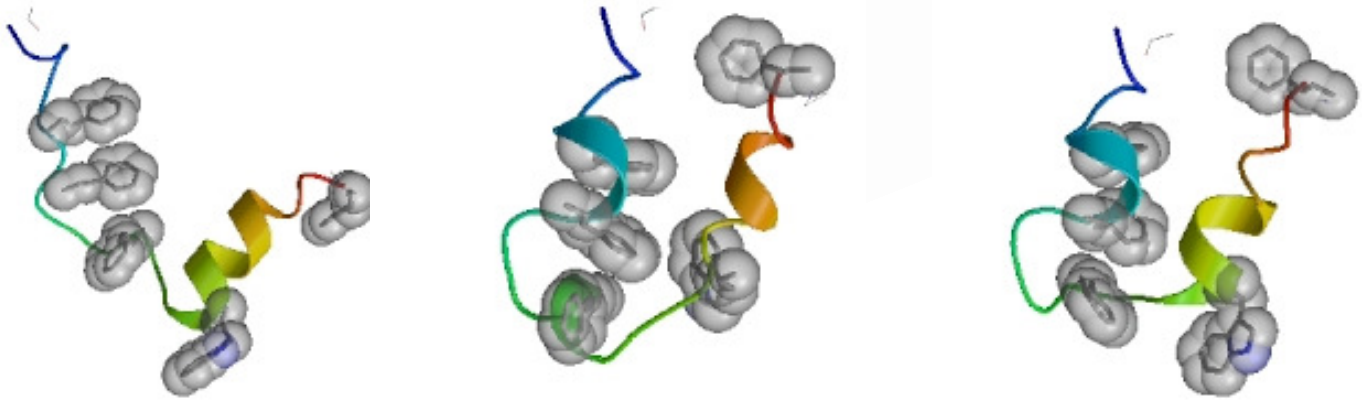


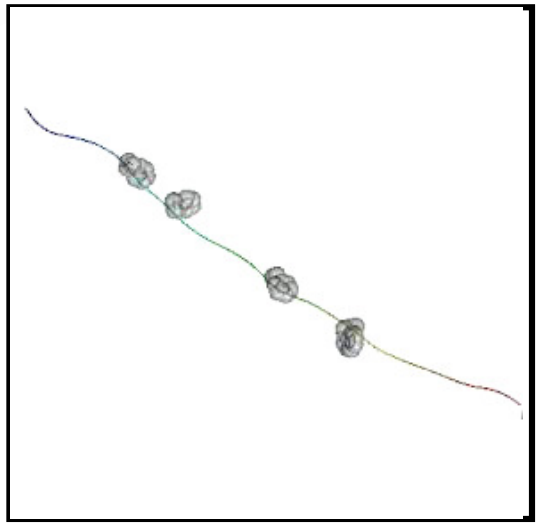
- Most Powerful resource in the world.

MOLECULAR DYNAMICS CAN SIMULATE PROTEIN FOLDING

Villin headpiece (36 amino acids)







For small proteins can fully simulate folding.

Pande Group

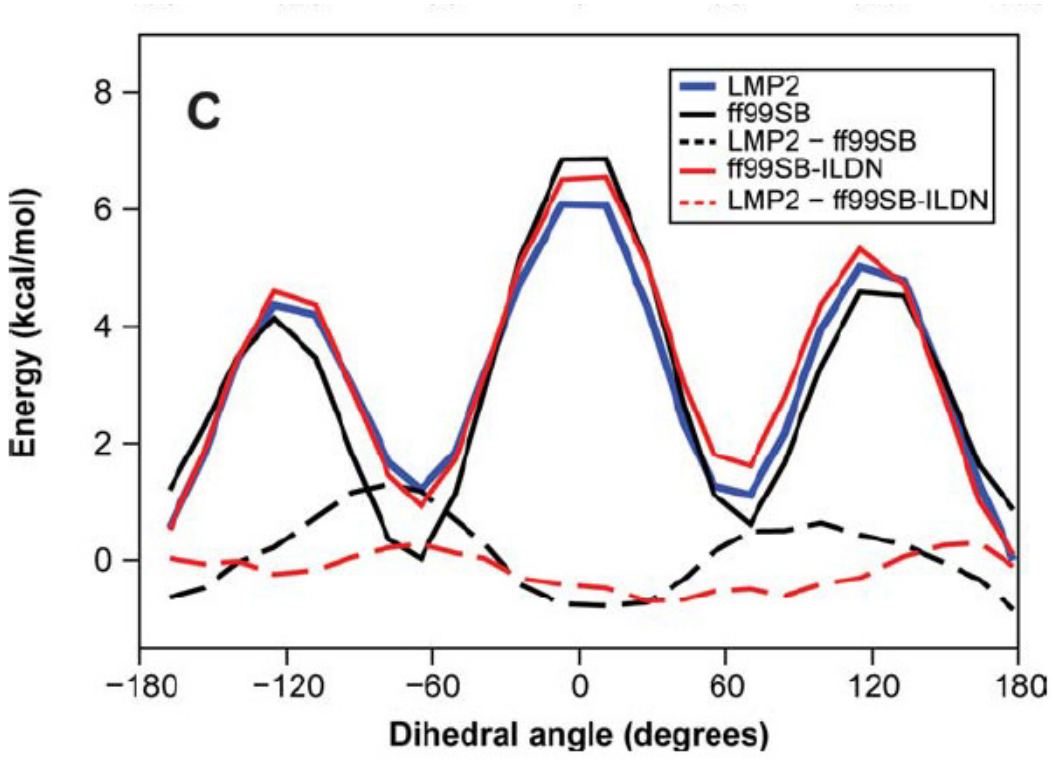
FOLDING PROTEINS WITH ANTON

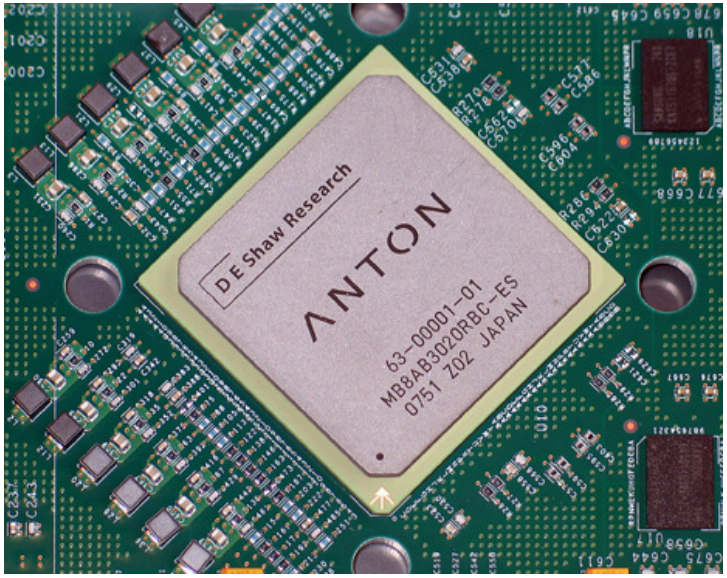
Improved side-chain torsion potentials for the Amber ff99SB protein force field

Kresten Lindorff-Larsen,[1] Stefano Piana,[1] Kim Palmo,[1] Paul Maragakis,[1] John L. Klepeis,[1] Ron O. Dror,[1] and David E. Shaw[1,2*]

[1] D. E. Shaw Research, New York, New York 10036

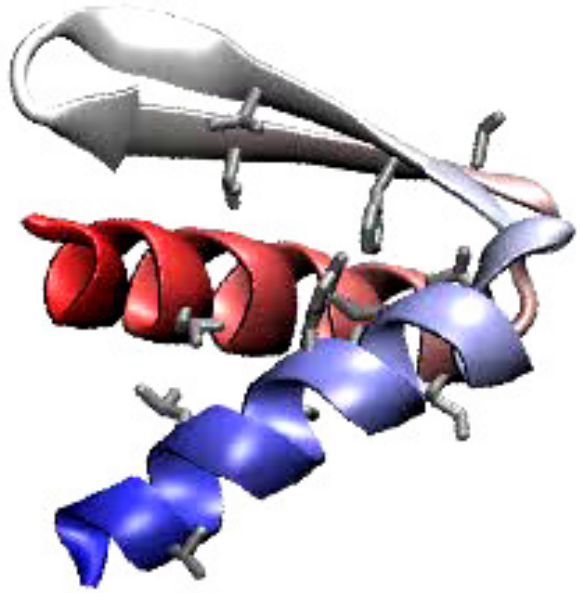
[2] Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032





D. E. Shaw Research
120 W. 45th St., 39th Fl.
New York, NY 10036

GBW UNFOLDS AND FOLDS



Gaurav Chopra

QUANTUM MECHANICS

$$U^{TOTAL} = \min_{\mathbf{t}} \left\{ \sum_{ab} \left(U^{ES}_{ab}(\mathbf{t}_a, \mathbf{t}_b; \mathbf{R}_a, \mathbf{R}_b) + U^{EX}_{ab}(r_{ab}) \right) + \sum_{a} U^{IN}_{a}(\mathbf{t}_a; \mathbf{R}) \right\} + \sum_{ab} U^{DS}_{ab}(R_{ab})$$

where $R_{ab} = \left| \mathbf{R}_a - \mathbf{R}_b \right|$ and $r_{ab} = \left| (\mathbf{R}_a + \mathbf{t}_a) - (\mathbf{R}_b + \mathbf{t}_b) \right|$

Algodign PNAS 102:7829 (2005)

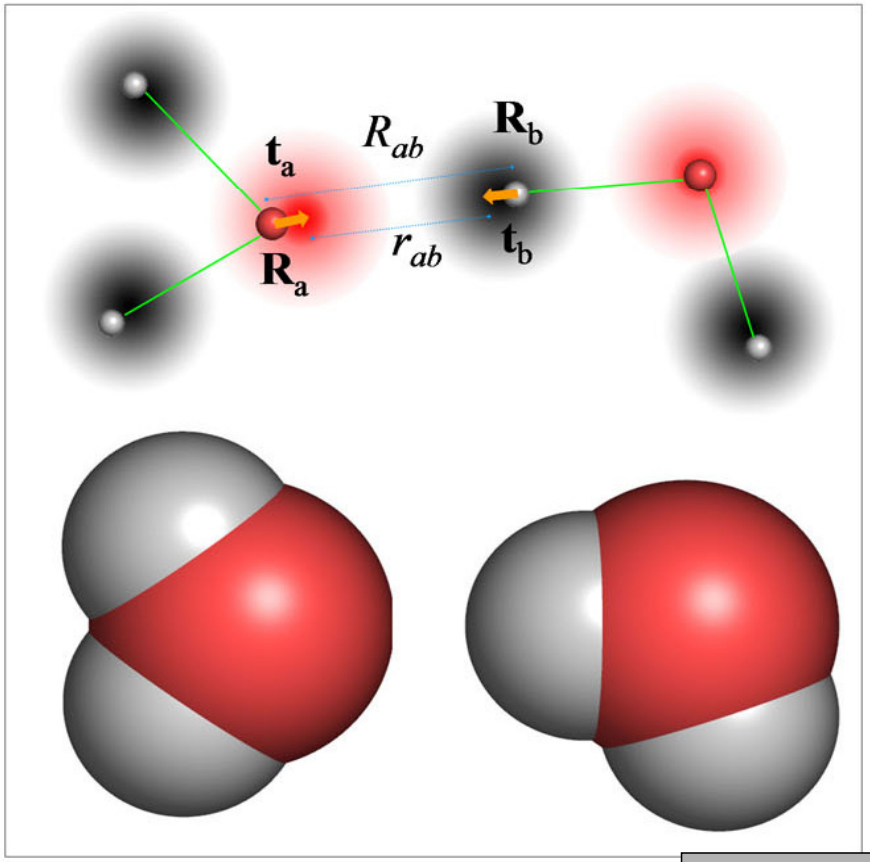
$$U^{ES}_{ab}(\mathbf{t}_a, \mathbf{t}_b; \mathbf{R}_a, \mathbf{R}_b) = \tilde{Z}_a \tilde{Z}_b \varphi(R_{ab}; 0, 0) + \tilde{Q}_a \tilde{Q}_b \varphi(r_{ab}; \tilde{w}_a, \tilde{w}_b)$$
$$+ \tilde{Q}_a \tilde{Z}_b \varphi(\left| \mathbf{R}_a + \mathbf{t}_a - \mathbf{R}_b \right|; \tilde{w}_a, 0)$$
$$+ \tilde{Z}_a \tilde{Q}_b \varphi(\left| \mathbf{R}_b + \mathbf{t}_b - \mathbf{R}_a \right|; 0, \tilde{w}_b)$$

$$U^{EX}_{ab}(r_{ab}) = \tilde{C}_a \tilde{C}_b \left(1 + \left(\frac{2 r_{ab}}{\tilde{w}_a + \tilde{w}_b} \right)^2 \right) \exp\left(- \frac{2 r_{ab}}{\tilde{w}_a + \tilde{w}_b} \right)$$

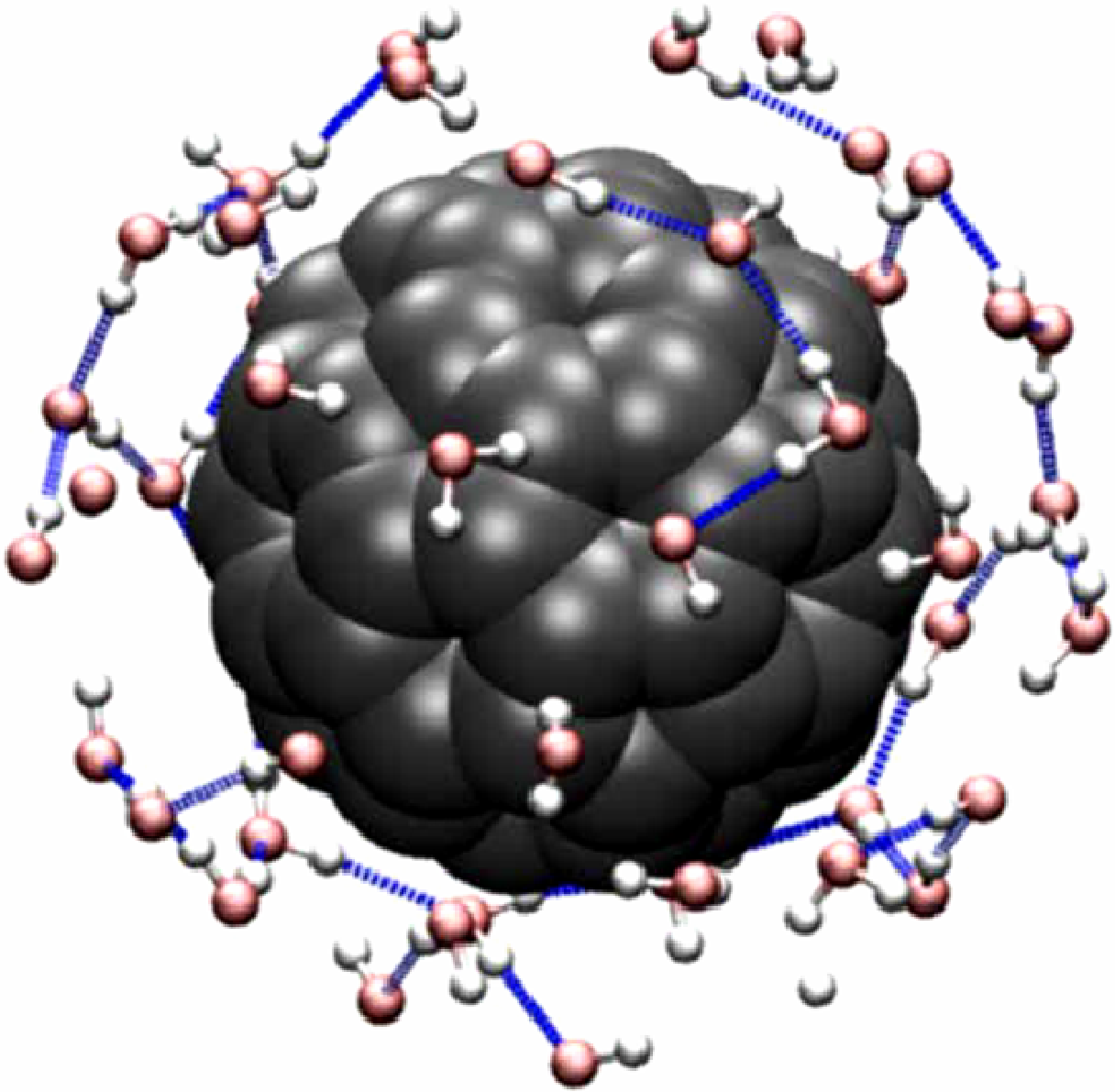
$$U^{IN}_{a}(\mathbf{t}_a; \mathbf{R}) = \frac{\tilde{Q}_a^2 \tilde{t}_a^2}{\tilde{\alpha}_a} \left(1 - \sqrt{1 - \frac{1}{\tilde{t}_a^2} \left(\mathbf{t}_a - \mathbf{t}_a^0 \right)^2} \right)$$

where $\mathbf{t}_a^0 = - \sum_b \tilde{t}_{ab} \mathbf{n}_{ab} / \tilde{w}_0$

$$U^{DS}_{ab}(R_{ab}) = - \frac{2 \tilde{E}_a \tilde{E}_b}{\tilde{E}_a + \tilde{E}_b} \left(\frac{\tilde{R}_0}{R_{ab} + \min(\tilde{R}_a, \tilde{R}_b)} \right)^6$$



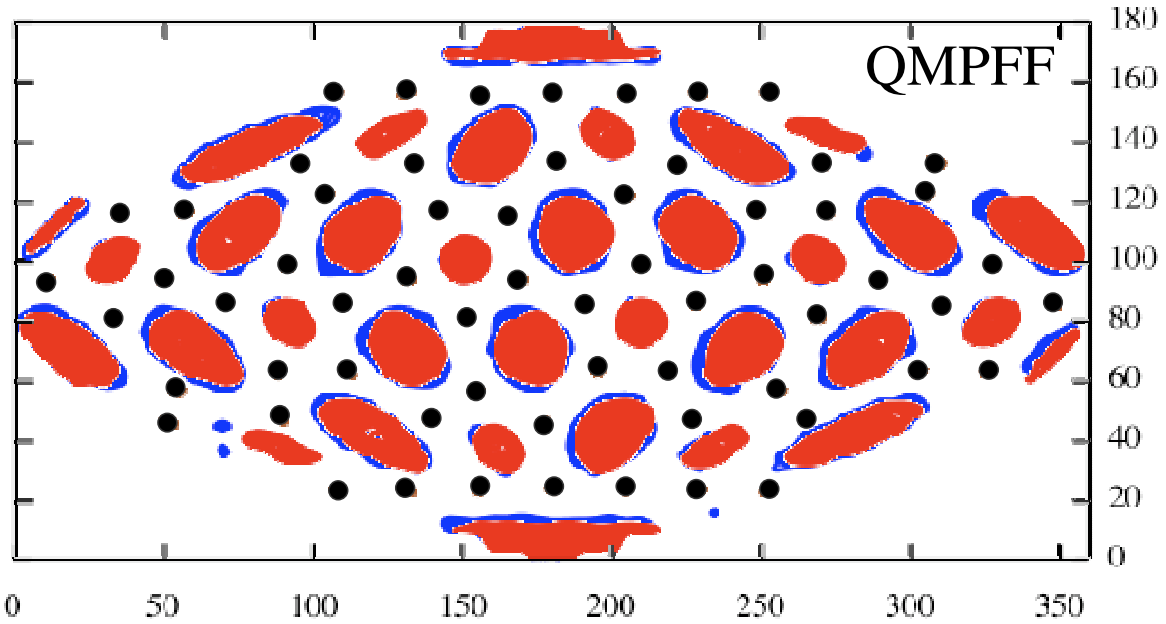
WATER AROUND C60



Oxygen

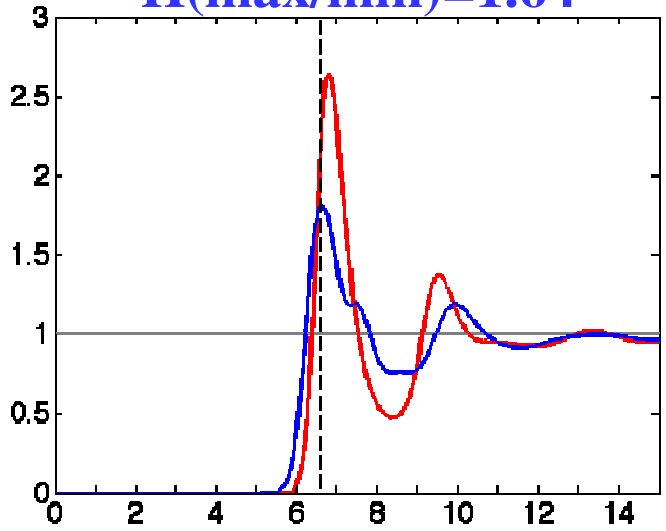
Hydrogen

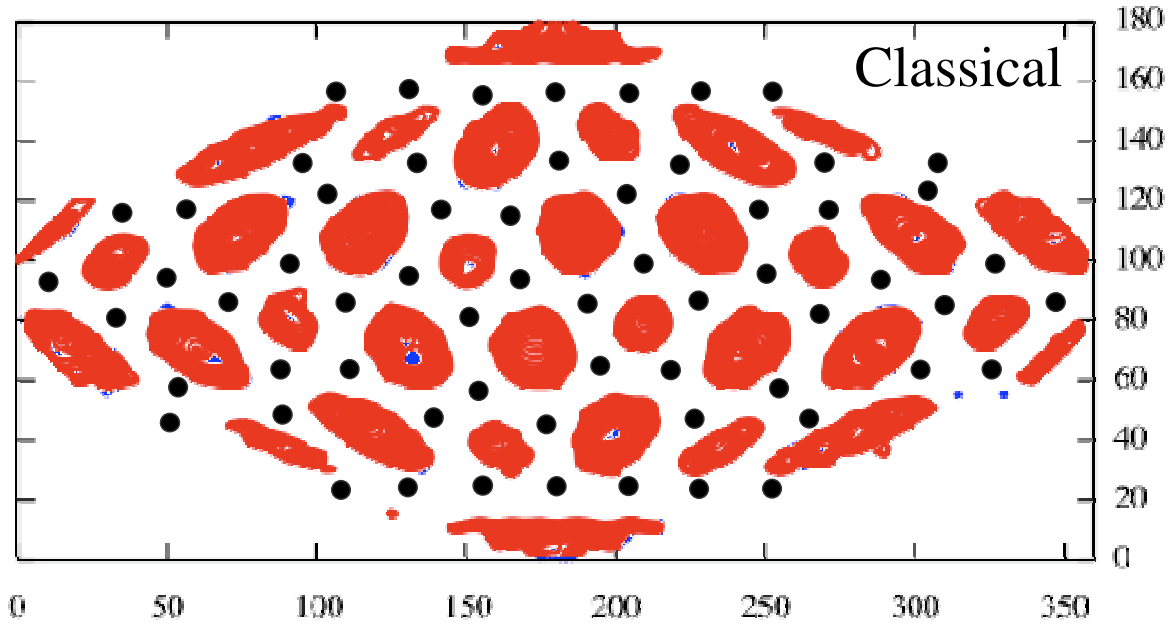
O AND H DENSITY AT R = 6.6 Å



O(max/min)=1.97

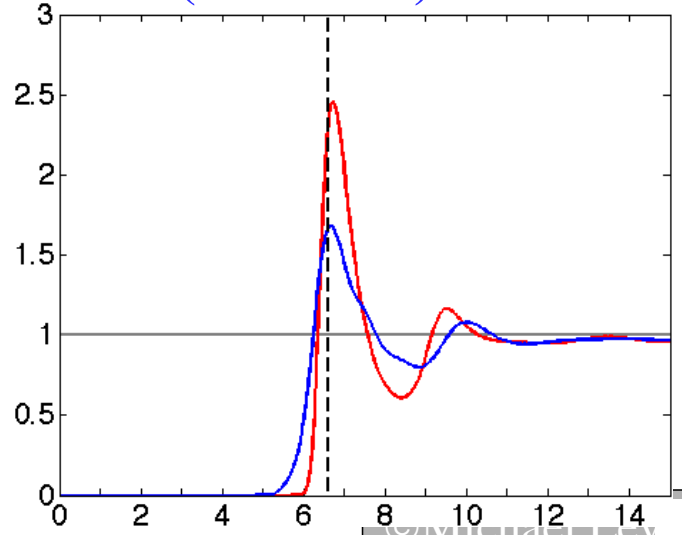
H(max/min)=1.64





O(max/min)=1.37

H(max/min)=1.38



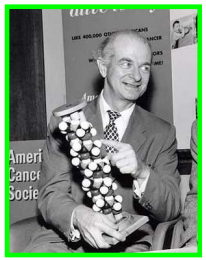
Final Words

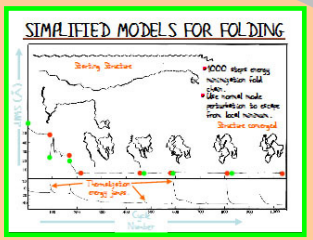
SUMMARY

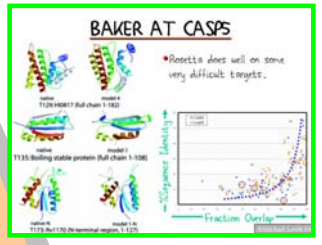
Modeling

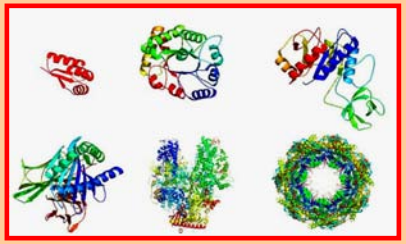
Experiment

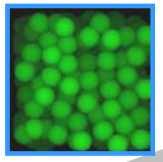
Simulation

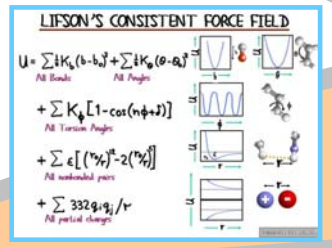


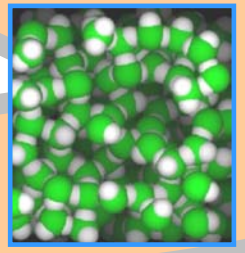


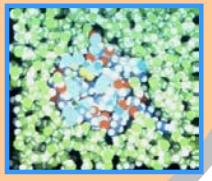


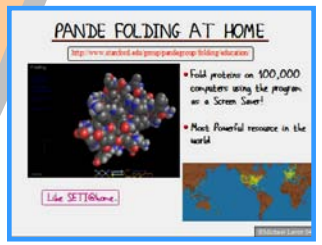








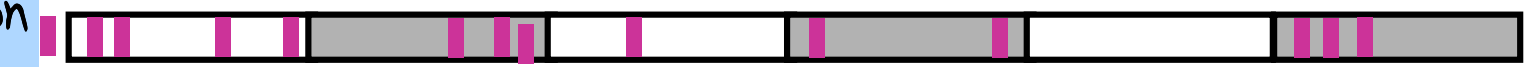




Models for Evolution

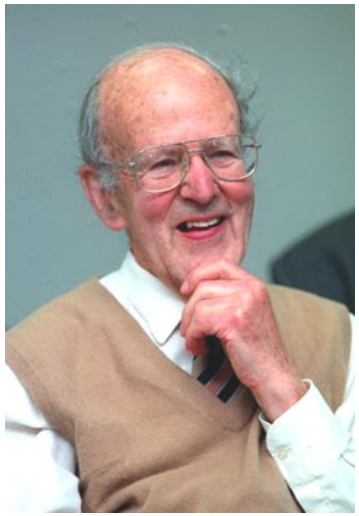
Large-Scale Structure

Accurate Simulation



1950 1960 1970 1980 1990 2000 2010

PERUTZ: MORE SCIENCE AFTER NOBEL PRIZE



1914–2002

Nobel Prize in 1962

66 Year Career

Royal Medal August 1971

Klosters 1977

FAMILY SUPPORT

THE END