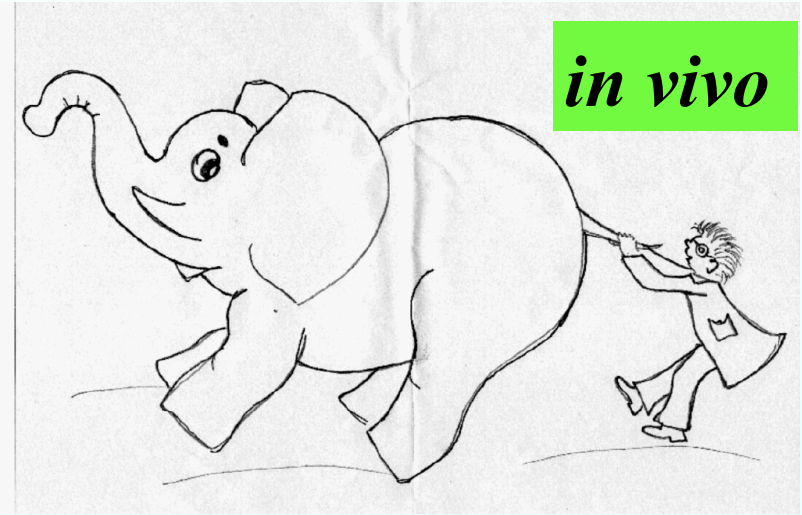
Research in Structural Bioinformatics and Molecular Biophysics

OUTLINE:

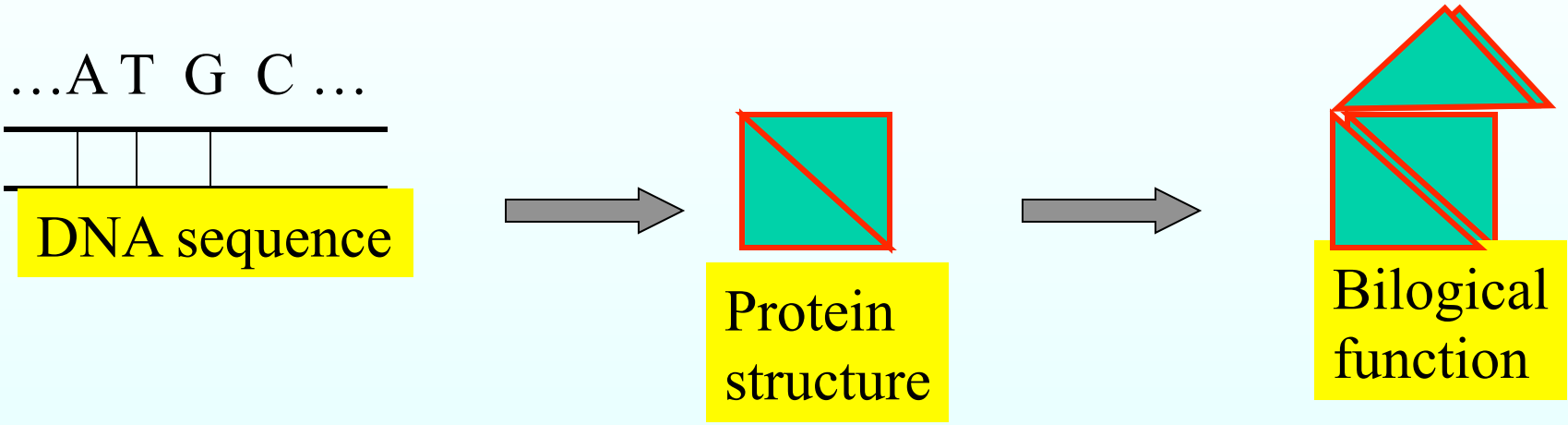
- What is it and why is it useful?
- EXAMPLES:
- a. Biomolecular surface story.
- b. Improving enzyme's function.
- c. Folding proteins.

Alexey Onufriev, Dept of Computer Science and Physics, Virginia Tech

The emergence of "*in virtuo*" Science.



Biological function = f(3D molecular structure)



Key challenges:

Biomolecular structures are complex (e.g. compared to crystal solids).

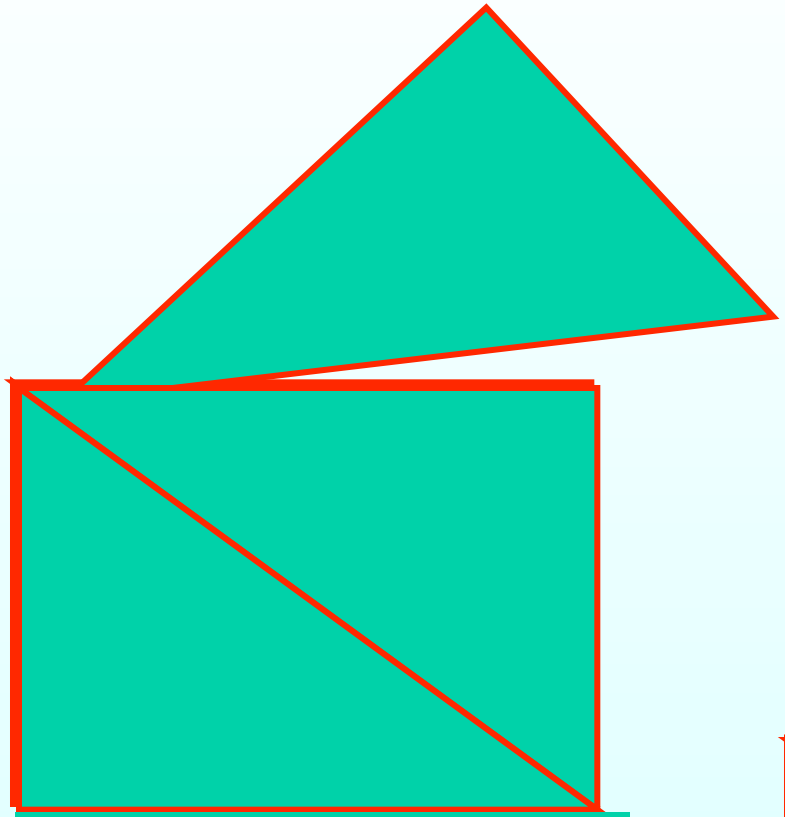
Biology works on many time scales.

Experiments can only go so far.

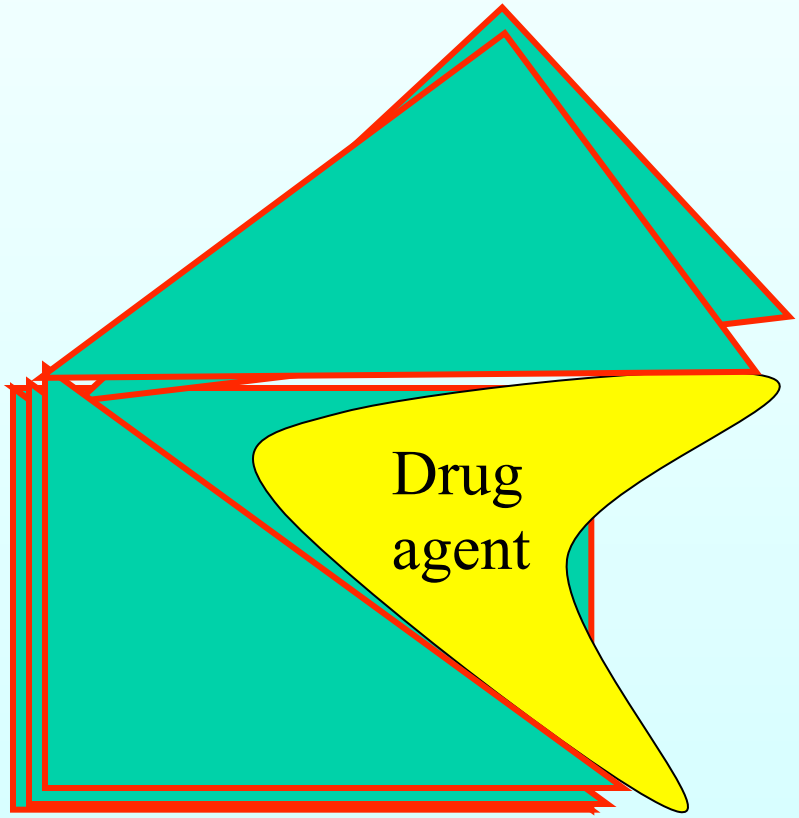
A solution: Computational methods.

Why bother? ***Example: rational drug design.***

If you block the enzymes function – you kill the virus.

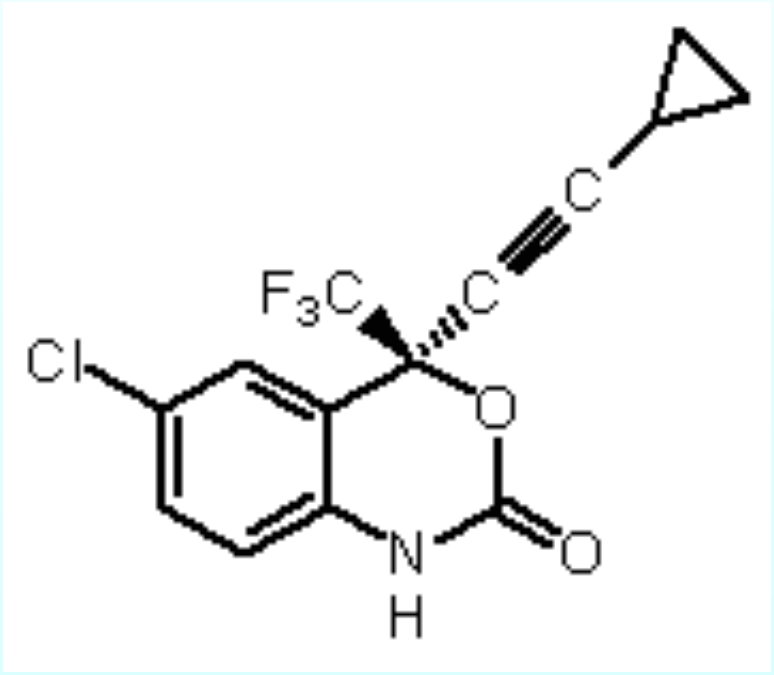


e.g.:viral endonuclease (cuts DNA, RNA)



Example of successful computer-aided (rational) drug design:
One of the drugs that helped slow down the AIDS epidemic

(part of anti-retro viral cocktail).



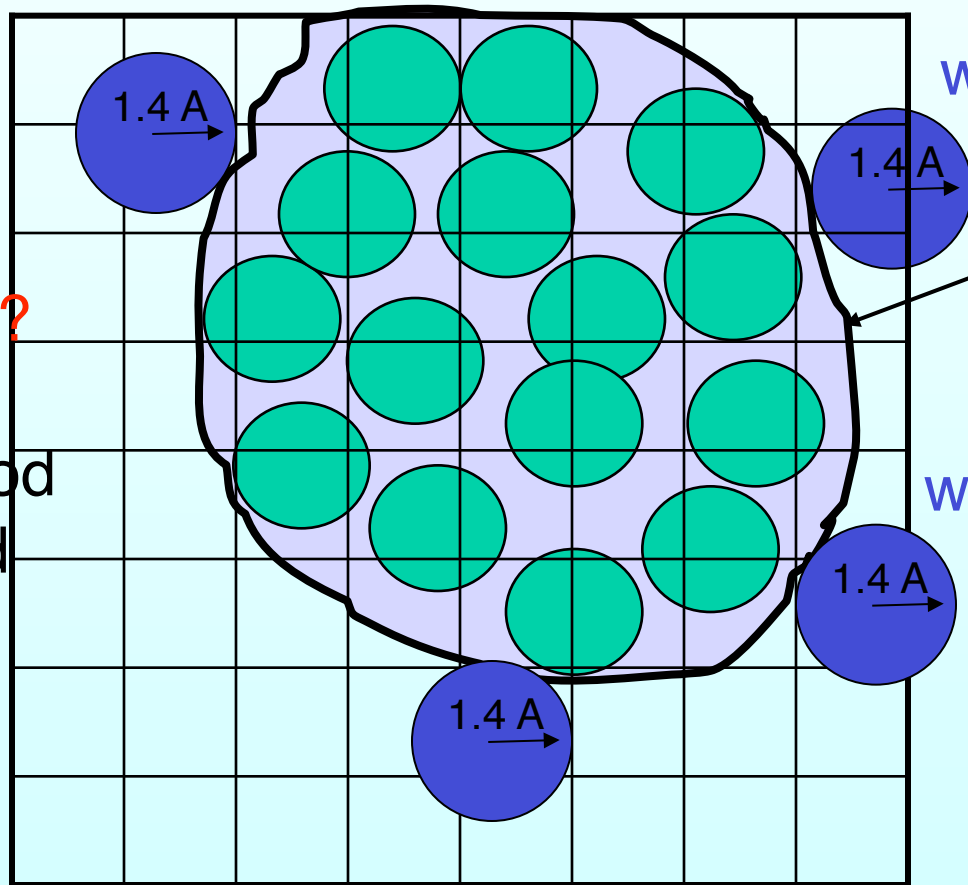
The drug blocks the function of a key viral protein. To design the drug, one needs a precise 3D structure of that protein.

Molecular shape DOES matter.

One can learn a lot from appropriate shape analysis.

Example of a computer-science challenge: molecular surface and volume

Need a SIMPLE, EFFICIENT approximation for volume and surface:



Molecular surface => no water within.

Grid computation?
A possibility, but not a good idea if speed is a factor.

A typical PDB entry (header)

myoglobin

```
HEADER    OXYGEN TRANSPORT                        13-DEC-97   101M
TITLE     SPERM WHALE MYOGLOBIN F46V N-BUTYL
ISOCYANIDE AT PH 9.0                              COMPND      MOL_ID: 1;
COMPND    2 MOLECULE: MYOGLOBIN;
COMPND    3 CHAIN: NULL;
COMPND    4 ENGINEERED: SYNTHETIC GENE;
COMPND    5 MUTATION: INS(M0), F46V, D122N
SOURCE    MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: PHYSETER CATODON;
SOURCE    3 ORGANISM_COMMON: SPERM WHALE;
SOURCE    4 TISSUE: SKELETAL MUSCLE;
SOURCE    5 CELLULAR_LOCATION: CYTOPLASM;
SOURCE    6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE    7 EXPRESSION_SYSTEM_STRAIN: PHAGE RESISTANT
SOURCE    8 EXPRESSION_SYSTEM_CELLULAR_LOCATION:
SOURCE    9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE    10 EXPRESSION_SYSTEM_PLASMID: PEMBL 19+
KEYWDS    LIGAND BINDING, OXYGEN STORAGE, OXYGEN
BINDING, HEME,                    KEYWDS   2 OXYGEN TRANSPORT
EXPDTA    X-RAY DIFFRACTION
AUTHOR    R.D.SMITH,J.S.OLSON,G.N.PHILLIPS JUNIOR
```

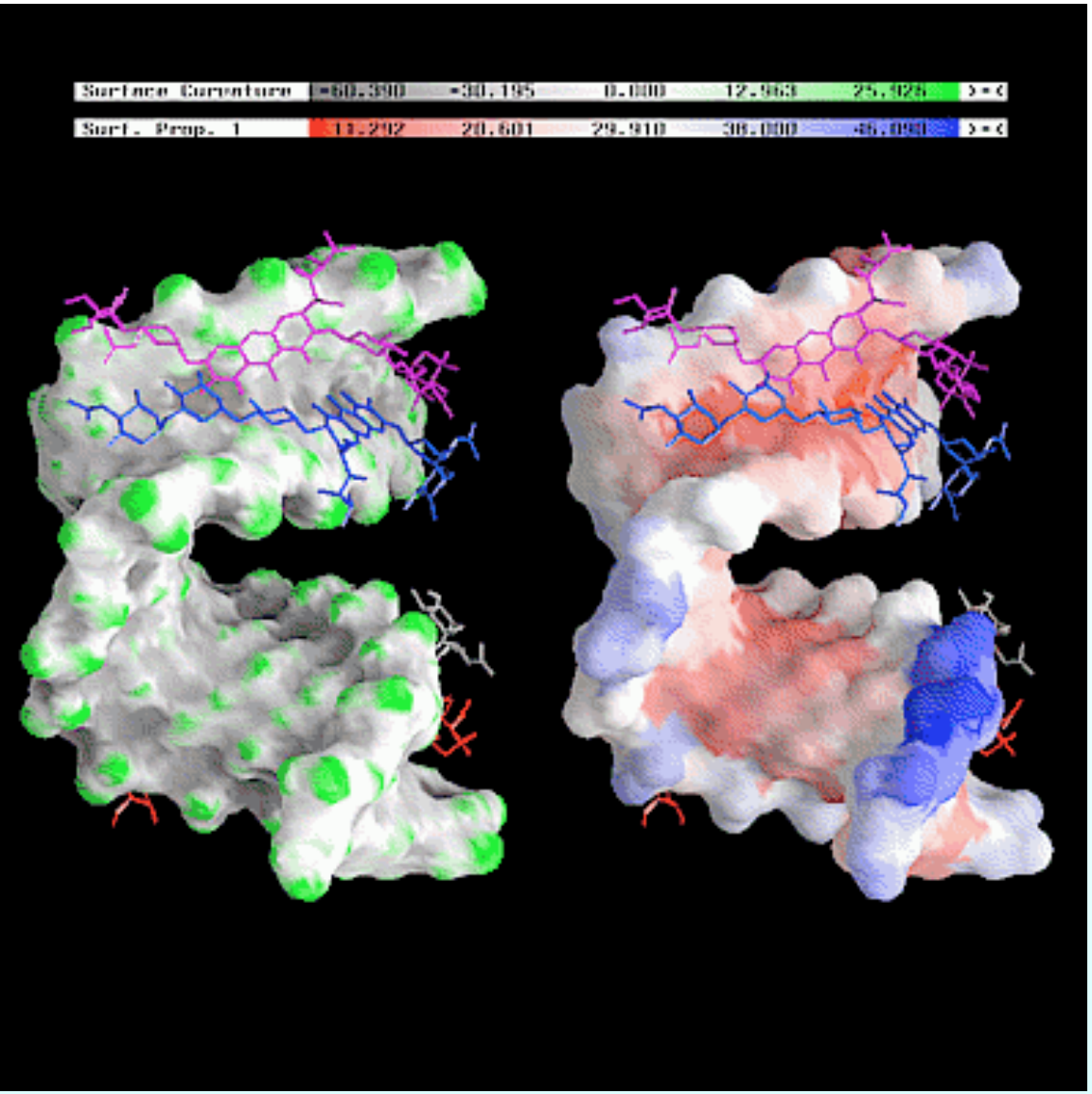
Key Part: atomic coordiantes (x,y,z)

```
                                X   Y    Z
ATOM 1 N       MET 0 24.277 8.374 -9.854 1.00 38.41 N
ATOM 2 CA     MET 0 24.404 9.859 -9.939 1.00 37.90 C
 ATOM 3 C       MET 0 25.814 10.249 -10.359 1.00 36.65 C
ATOM 4 O       MET 0 26.748 9.469 -10.197 1.00 37.13 O
 ATOM 5 CB    MET 0 24.070 10.495 -8.596 1.00 39.58 C
ATOM 6 CG     MET 0 24.880 9.939 -7.442 1.00 41.49 C
ATOM 7 SD     MET 0 24.262 10.555 -5.873 1.00 44.70 S
ATOM 8 CE     MET 0 24.822 12.266 -5.967 1.00 41.59 C
ATOM 9 N       VAL 1 25.964 11.453 -10.903 1.00 34.54 N
ATOM 10 CA    VAL 1 27.263 11.924 -11.359 1.00 32.46 C
 ATOM 11 C      VAL 1 27.392 13.428 -11.115 1.00 30.70 C
ATOM 12 O      VAL 1 26.443 14.184 -11.327 1.00 31.42 O
 ATOM 13 CB     VAL 1 27.455 11.631 -12.878 1.00 32.95 C
ATOM 14 CG1    VAL 1 28.756 12.209 -13.382 1.00 32.87 C
 ATOM 15 CG2    VAL 1 27.432 10.131 -13.140 1.00 33.54 C
 ATOM 16 N      LEU 2 28.555 13.855 -10.636 1.00 27.76 N
 ATOM 17 CA     LEU 2 28.797 15.269 -10.390 1.00 25.21 C
ATOM 18 C      LEU 2 29.492 15.903 -11.585 1.00 24.21 C
 ATOM 19 O      LEU 2 30.250 15.240 -12.306 1.00 23.80 O
ATOM 20 CB     LEU 2 29.688 15.470 -9.152 1.00 24.30 C
 ATOM 21 CG       LEU 2 29.084 15.416 -7.751 1.00 22.96 C
 ATOM 22 CD1     LEU 2 28.730 13.988 -7.390 1.00 22.03 C
```

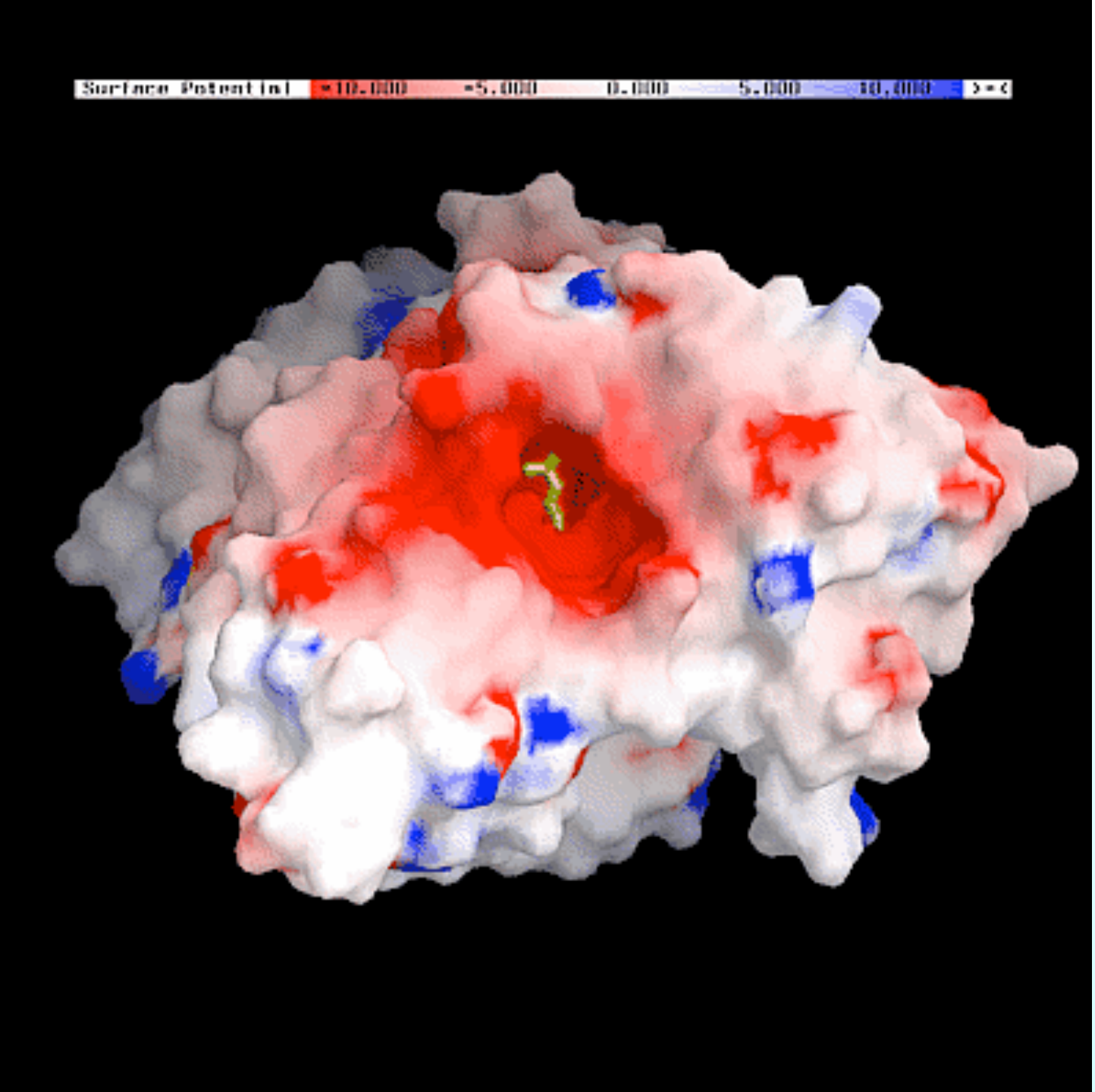
How to infer something meaningful from this?

Meaningful visualization helps.

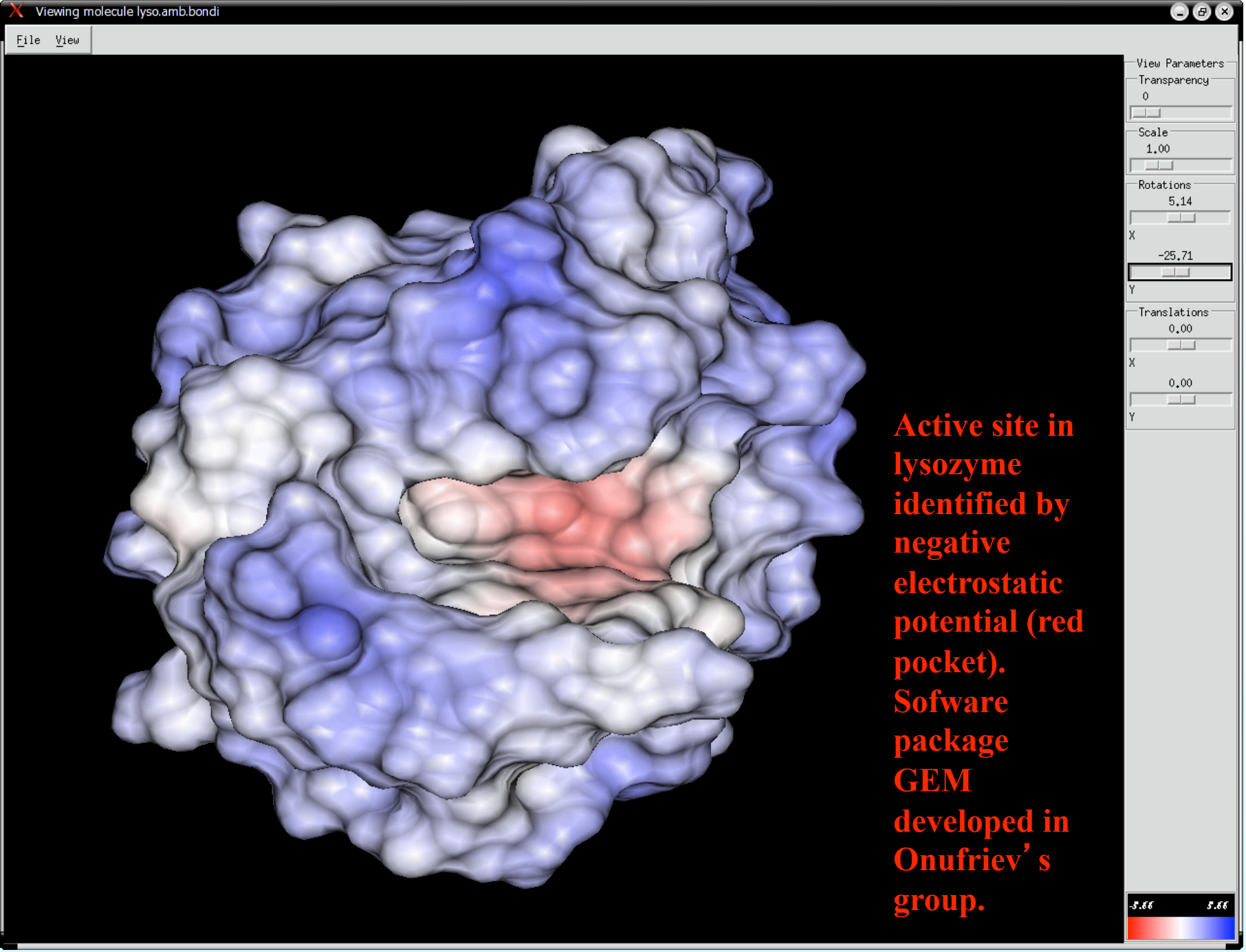
Examples.



The surface of a short DNA fragment which binds to a drug dimer (chromomyosin) is shown color coded on the left by curvature and on the right by *B value* (structural flexibility). The latter are propagated to the surface from the B values of the atoms below. The drug molecule is represented in stick mode. Note that where the drug binds the DNA has significantly lower B values, indicating it is less mobile. Also note from the left hand surface that the effect of binding the drug is to cause the surface of the major groove to "flex" outward, while the minor groove widens.



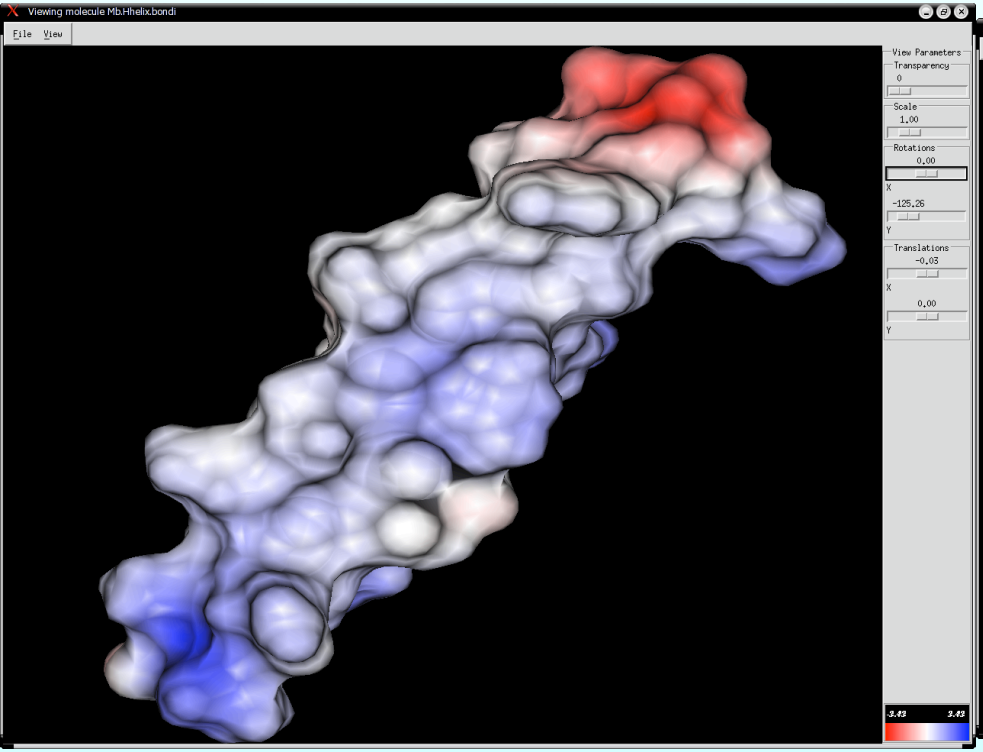
Molecular surface of acetyl choline esterase molecule (structure by Sussman et al.) color coded by electrostatic potential. The view is directly into the active site and acetyl choline is present in a bond representation. Note the depth of the pocket, its negative nature corresponding to the postive charge on the acetyl choline *(small worm-like thing*


Viewing molecule lyso.amb.bondi
File View
View Parameters
Transparency
0
Scale
1.00
Rotations
5.14
X
-25.71
Y
Translations
0.00
X
0.00
Y
-5.66
5.66
Active site in lysozyme identified by negative electrostatic potential (red pocket). Sofware package GEM developed in Onufriev's group.

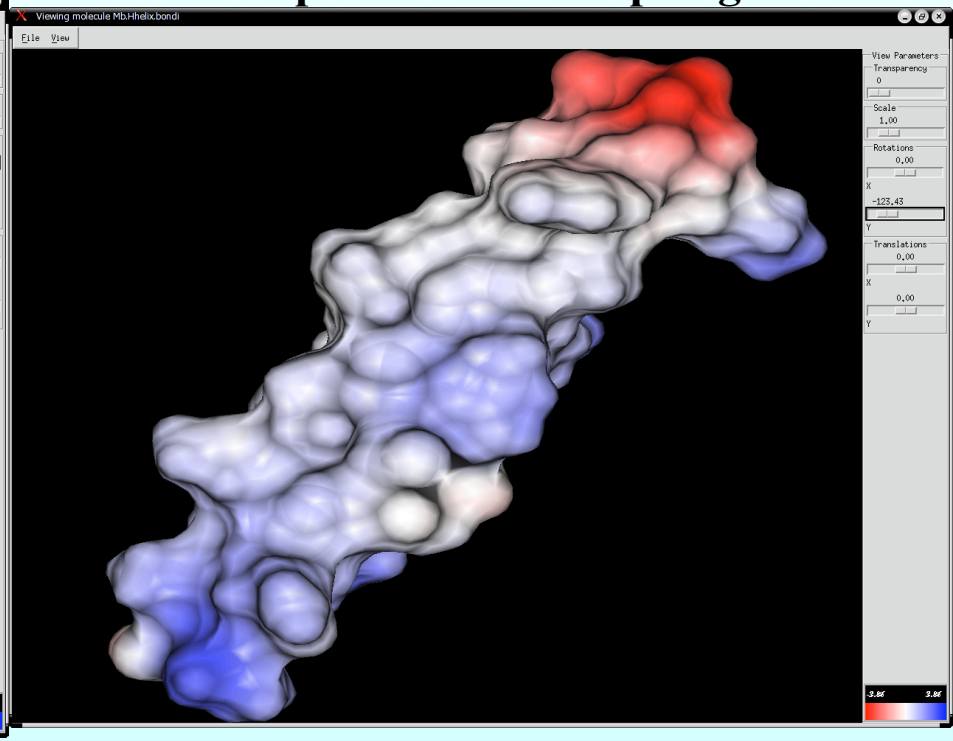
We can do the same thing, but much much faster, based on the "virtual water" ideas.

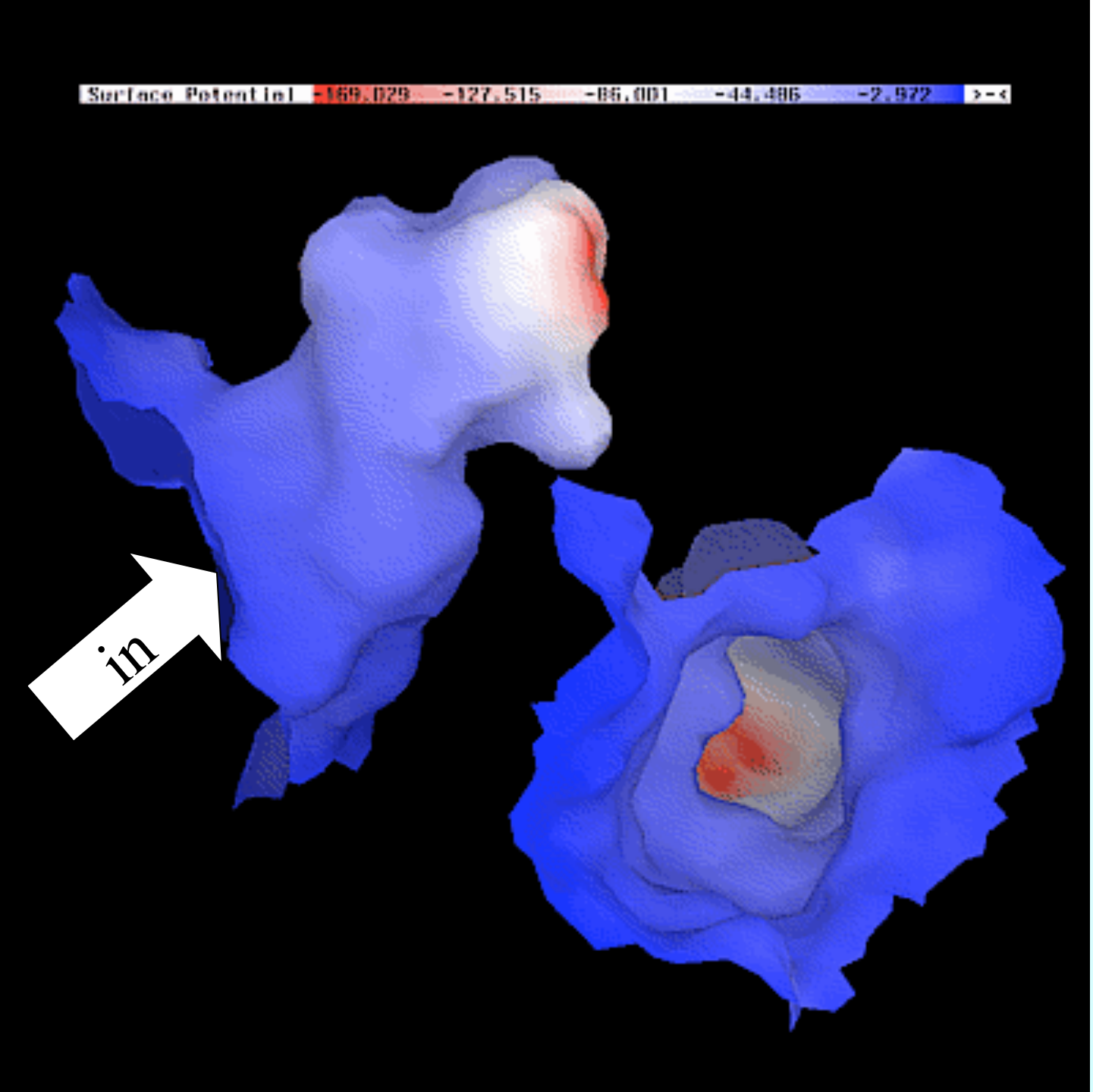
Example: potential of α-helix dipole.

DelPhi **(grid-based traditional method)**

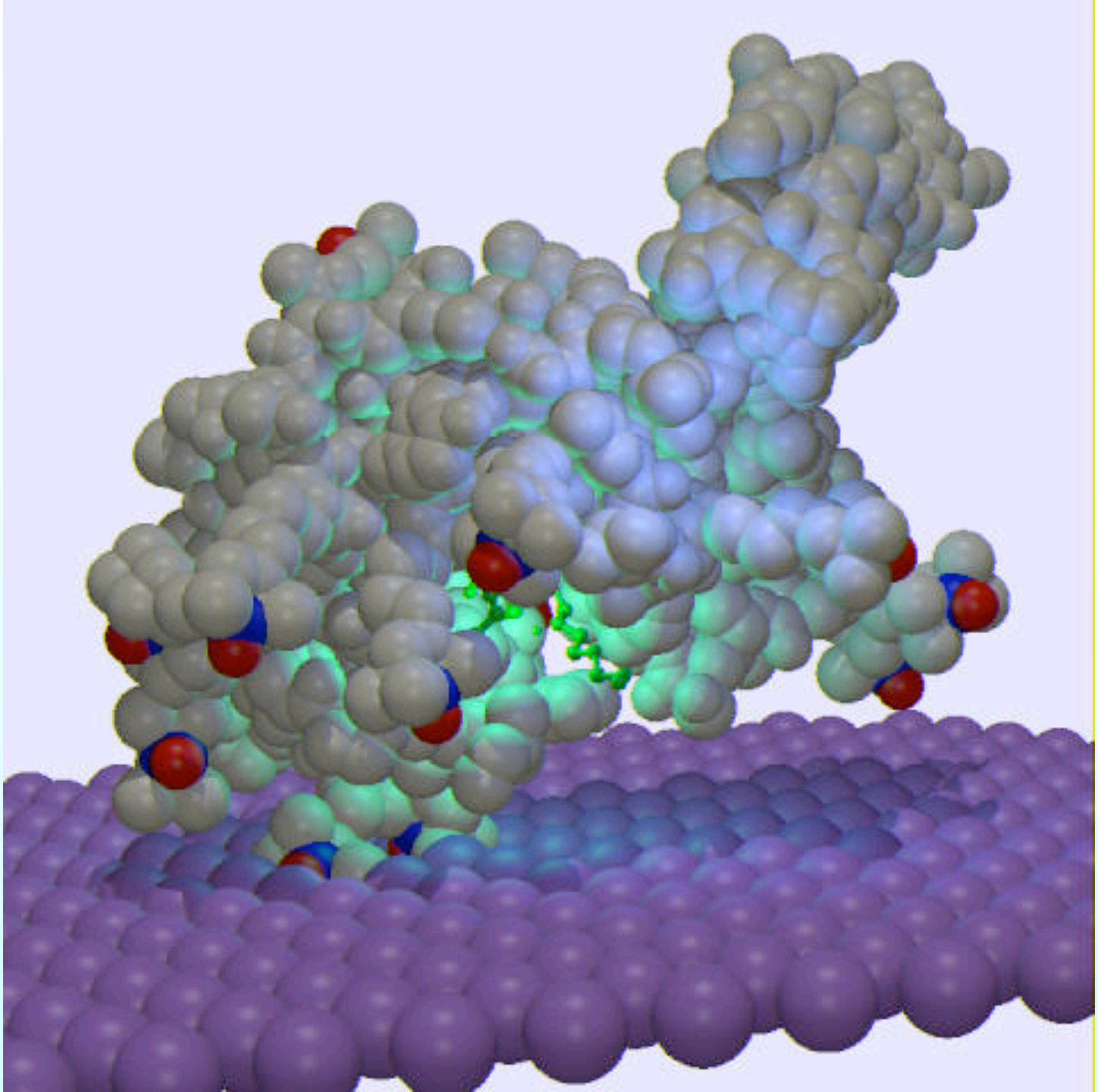


GEM **(our analytical method)**
Developed in CS6104 Spring 04





The surface of the active site of acetly choline esterase seen from two different angles, color coded by electrostatic potential. Note the potential gets more negative the deeper in one goes. Also note that one view of the surface is lit from the inside, the other from the outside, i.e the latter is the former "inverted"



Yet another cool picture …

As if this this was not already complex enough…

the molecules are ALIVE (i.e. they move).

Everything that living things do …
…can be explained by the wiggling and jiggling of atoms.
R. Feynman

Suggests the approach: model what nature does, *i.e.* let the molecule evolve with time according to underlying physics laws.

"Everything that living things do…

can be reduced to wiggling and jiggling of atoms"

R. Feynmann

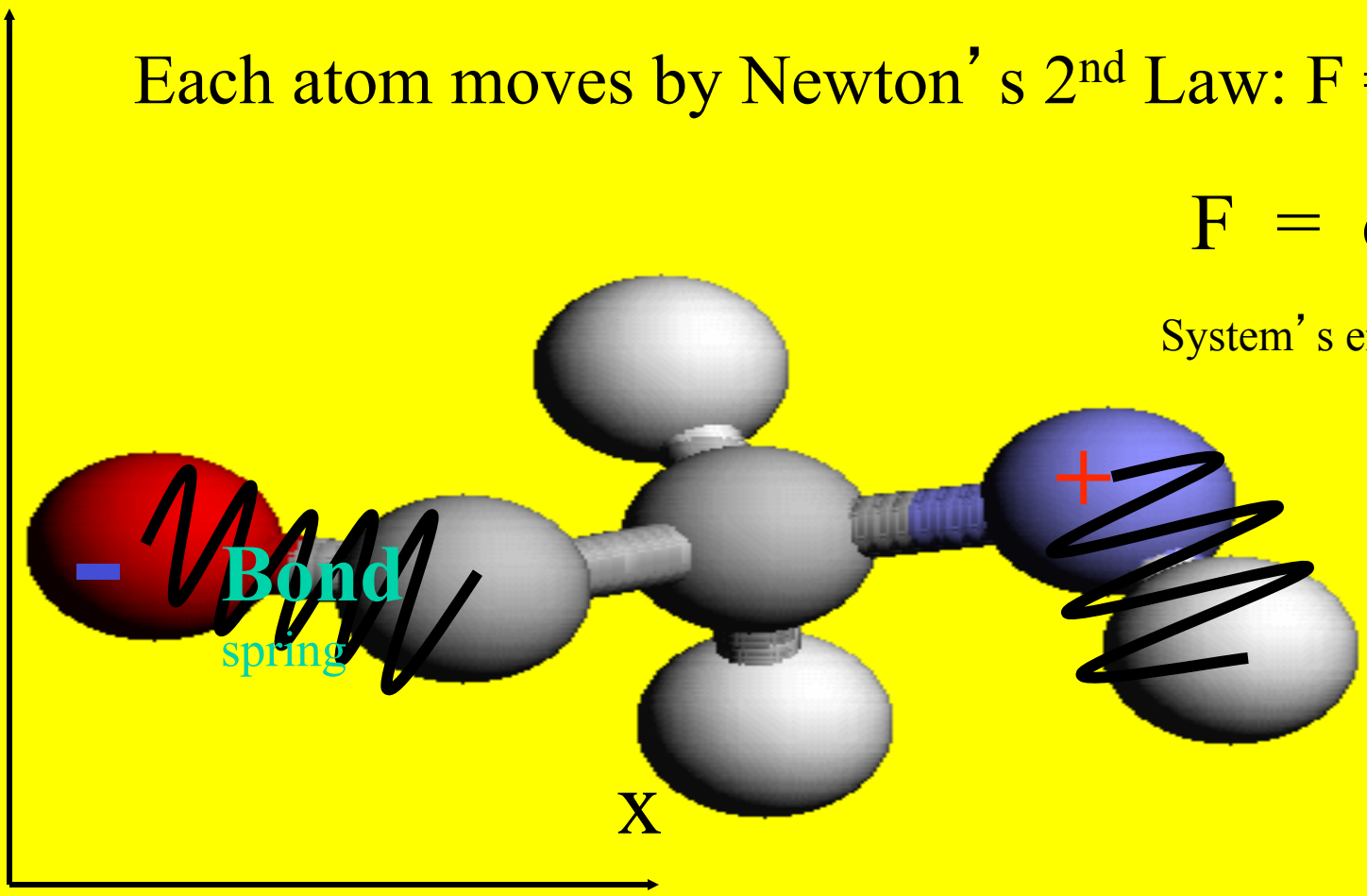
Suggests the approach: model what nature does, *i.e.* let the molecule evolve with time according to underlying physics laws.

Principles of Molecular Dynamics (MD):

Each atom moves by Newton's 2nd Law: $F = ma$

$$F = dE/dr$$

System's energy



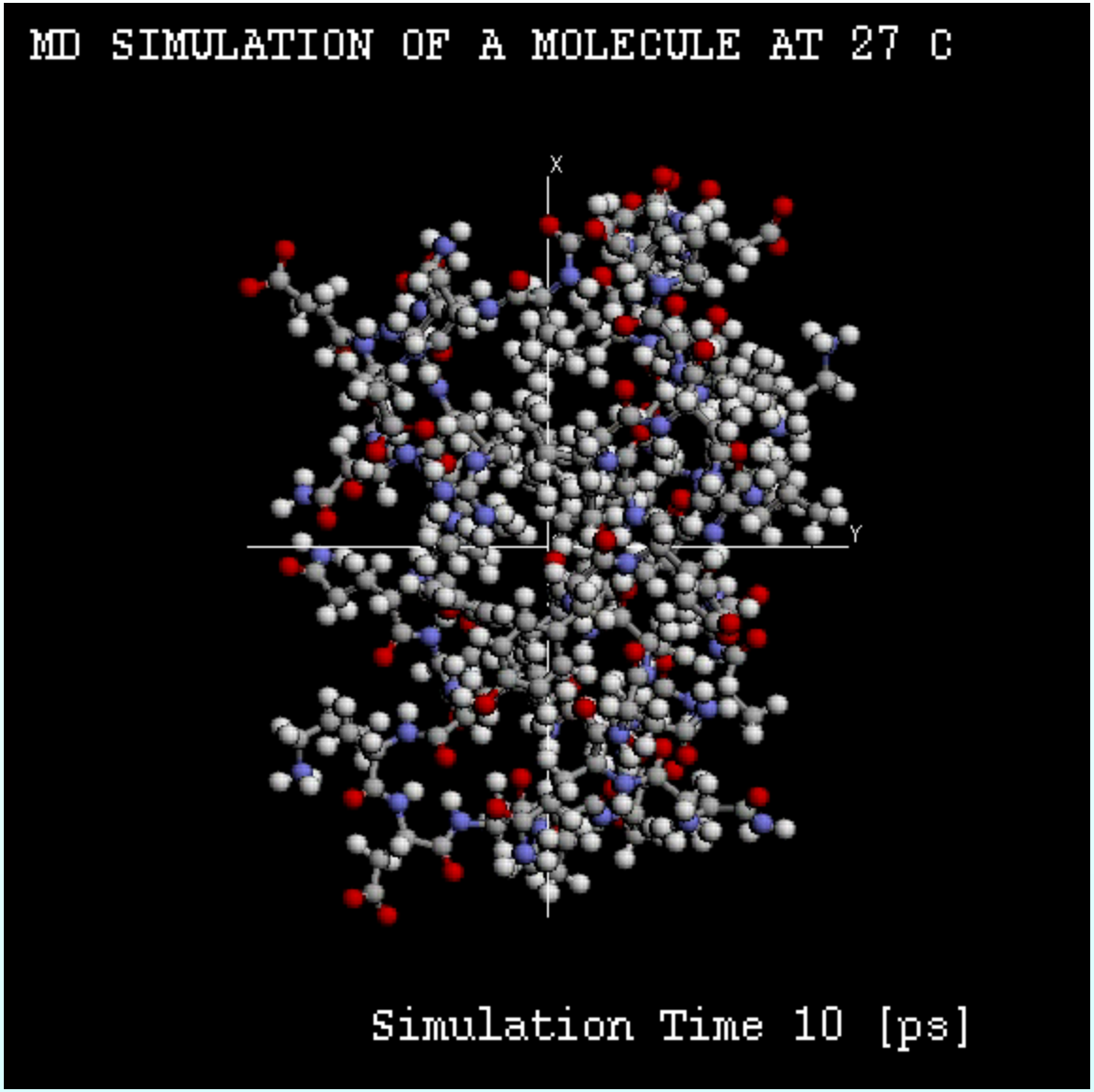
$$E = Kr^2 + A/r^{12} - B/r^6 + Q_1Q_2/r + \ldots$$

Bond stretching

VDW interaction

Electrostatic forces

Now we have positions of all atoms as a function of time.

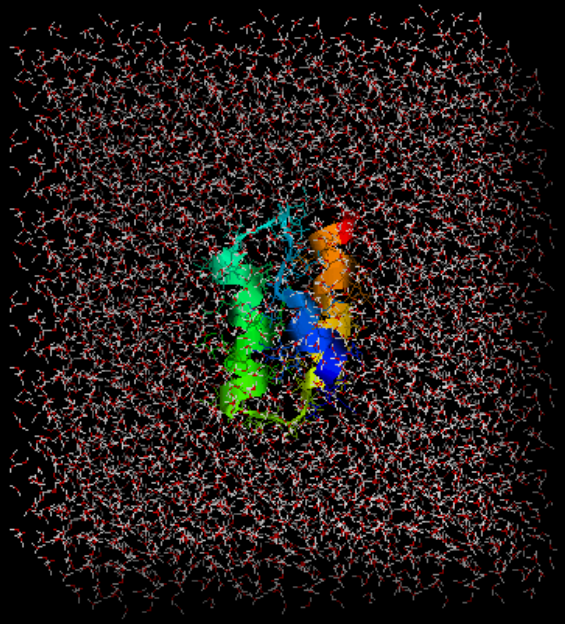


Can compute statistical averages, fluctuations; Analyze side chain movements, Cavity dynamics, Domain motion, Etc.

Computational advantages of representing water implicitly, via the ``virtual water" model

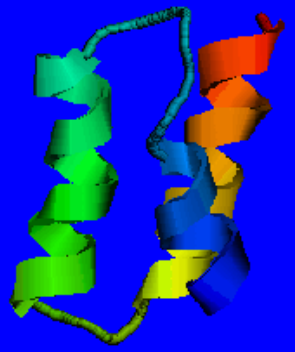
(currently being developed in my group at VT)

Explicit water (traditional)



Large computational cost. Slow dynamics.

Implicit water as dielectric continuum



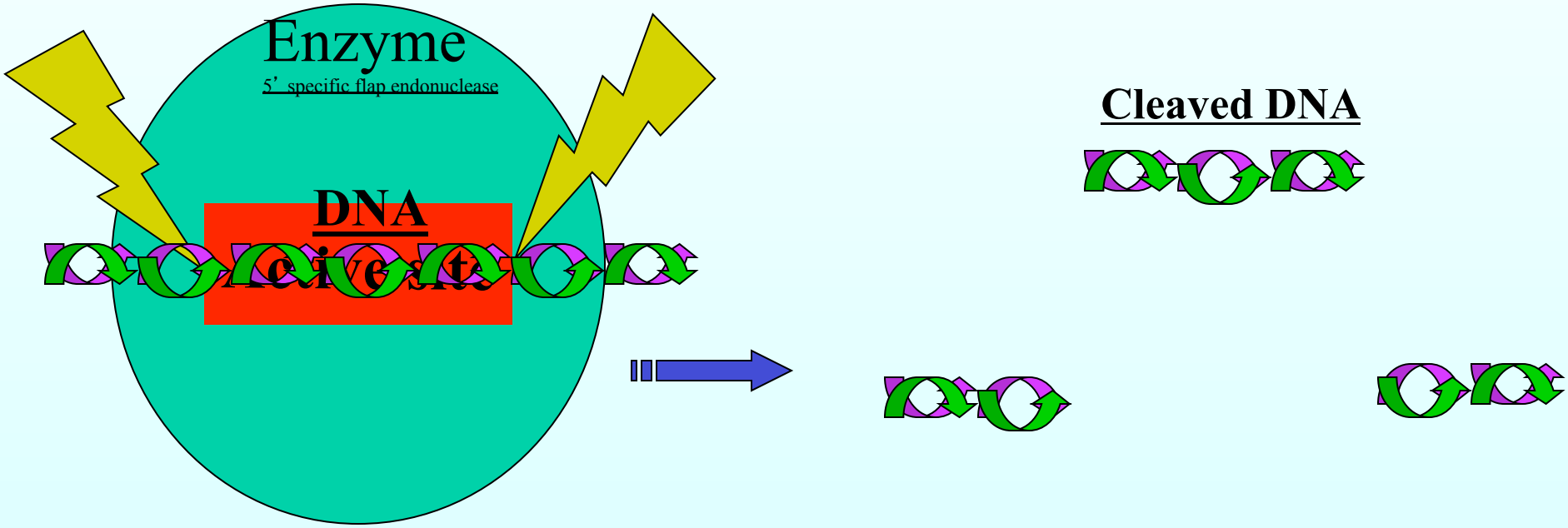
Low computational cost. Fast dynamics.

No need to track individual water molecules

No drag of viscosity

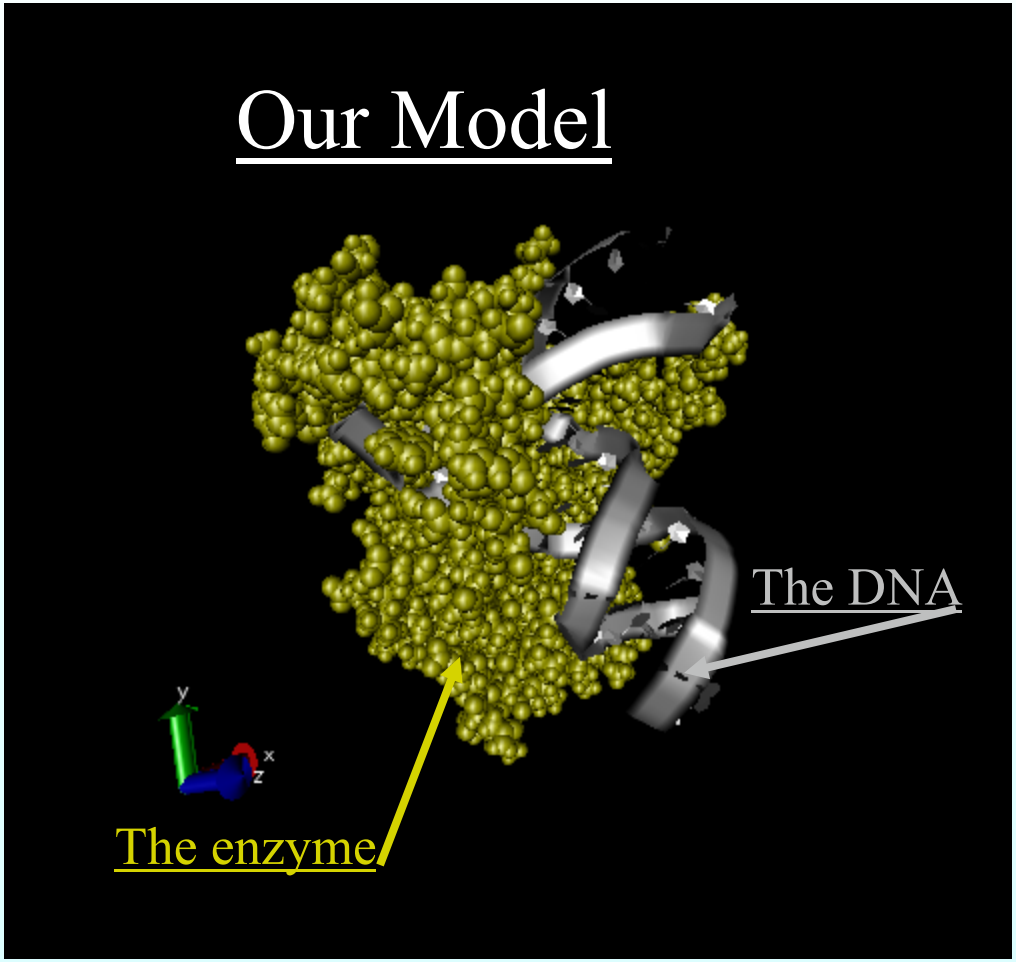
An industrial application: improving the function of a commercial enzyme.

Collaboration with the Third Wave Technologies, Inc. Madison, WI



Problem: to understand the mechanism, need structure of the enzyme-DNA complex (unavailable from experiment).

Solution: model the structure using molecular dynamics (and other) computational techniques



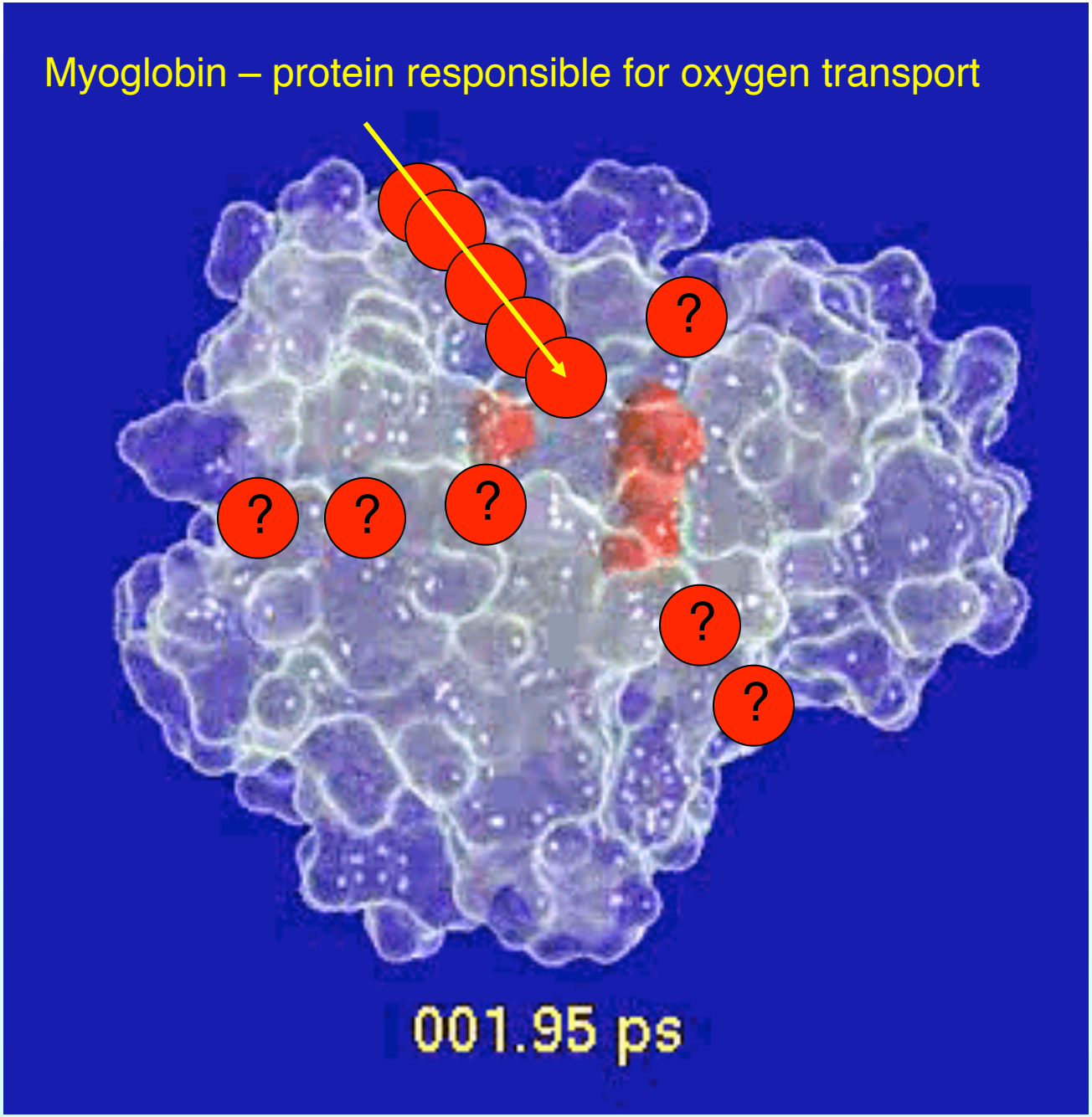
Result: On the basis of the model, mutations were introduced that improved the enzyme's function.

So, molecular volume changes with time.

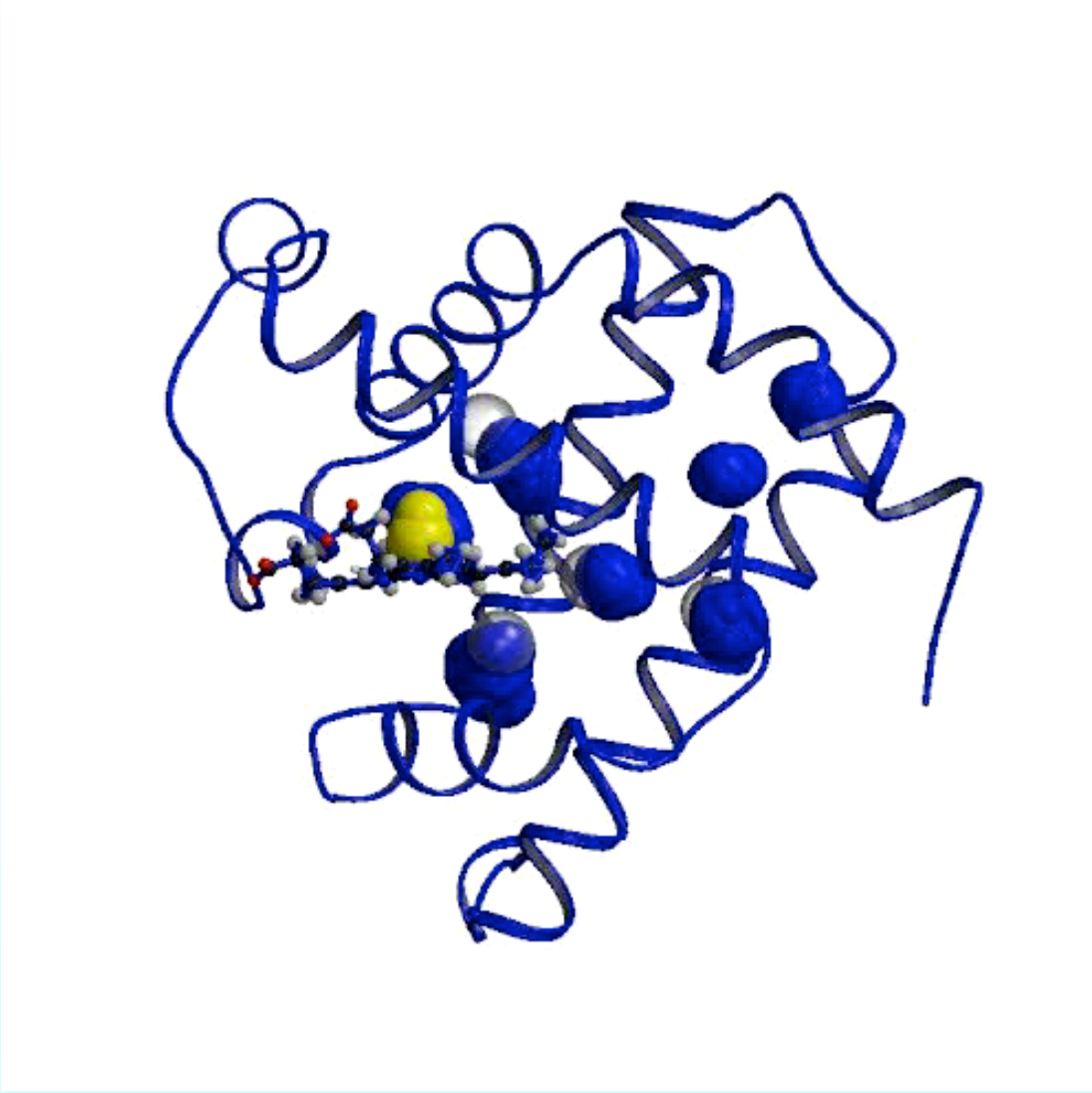
How does that help?

Example:
Resolves the problem with oxygen uptake by myoglobin.

How oxygen gets inside myoglobin? Single vs. multiple channels.

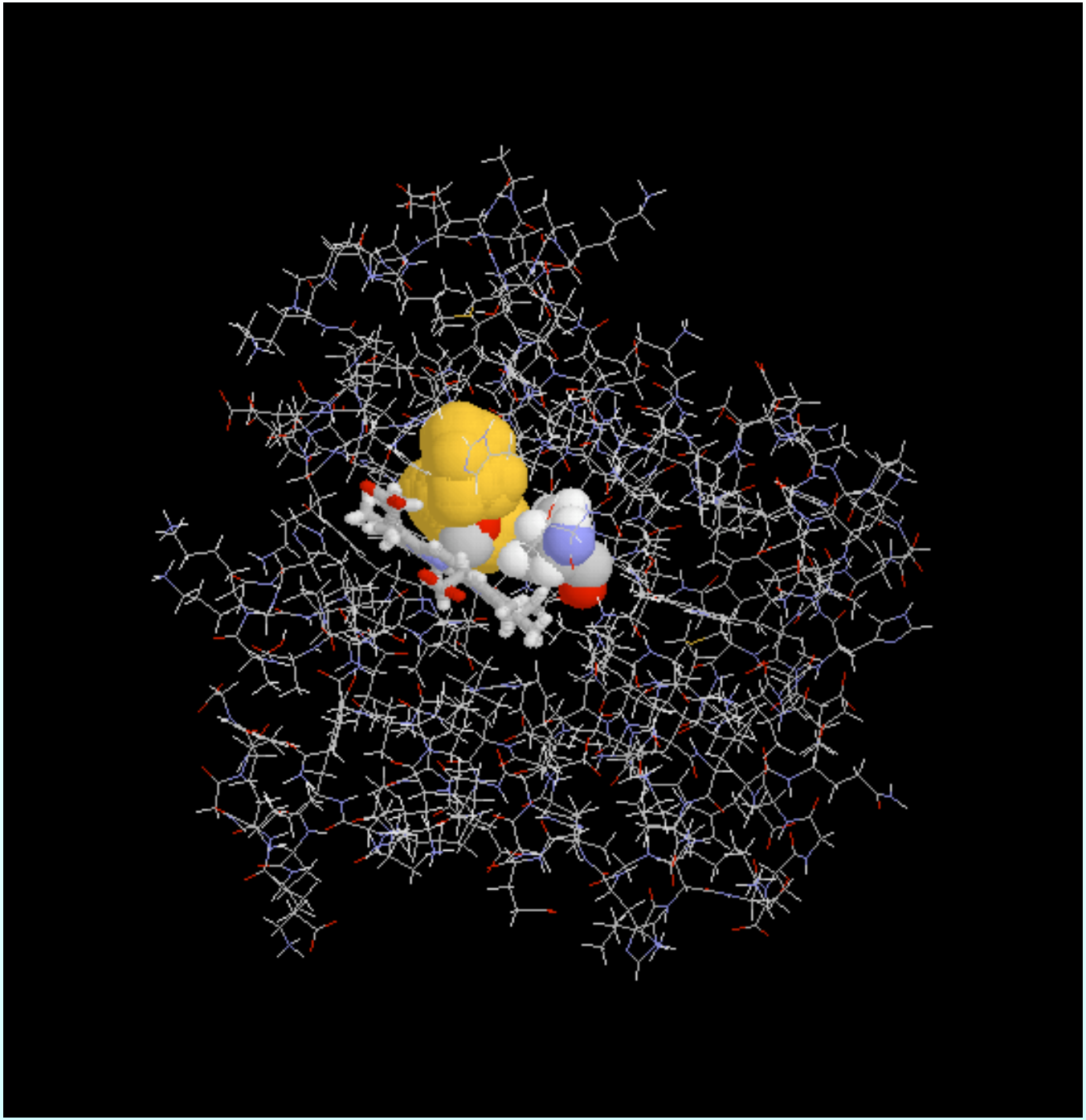


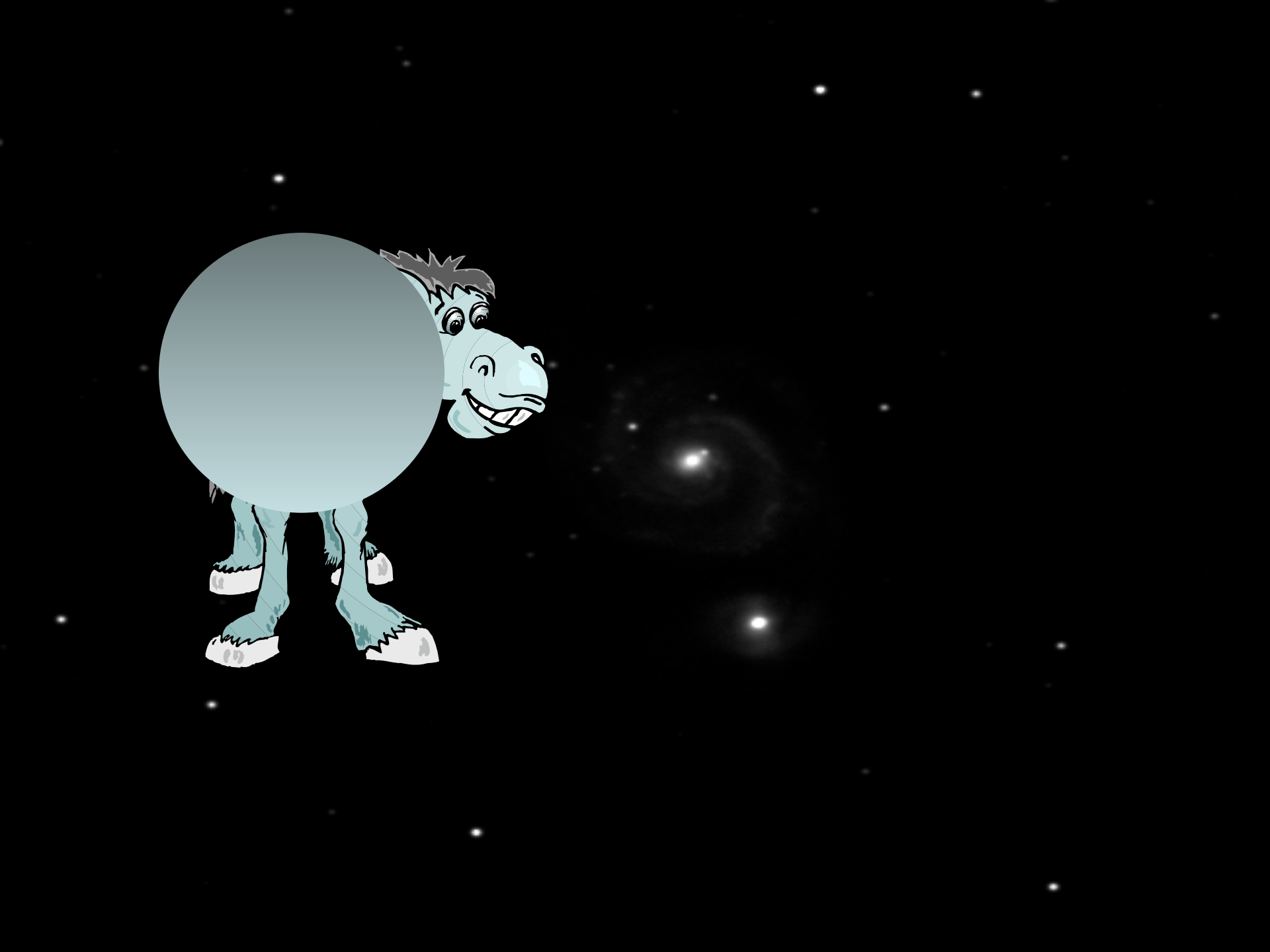
Holes in the protein as a function of time



How do we explain the specific location of the pathways?

Dynamic pathways occur in the ¨loose˝ space in-between the helices and in the loop regions.



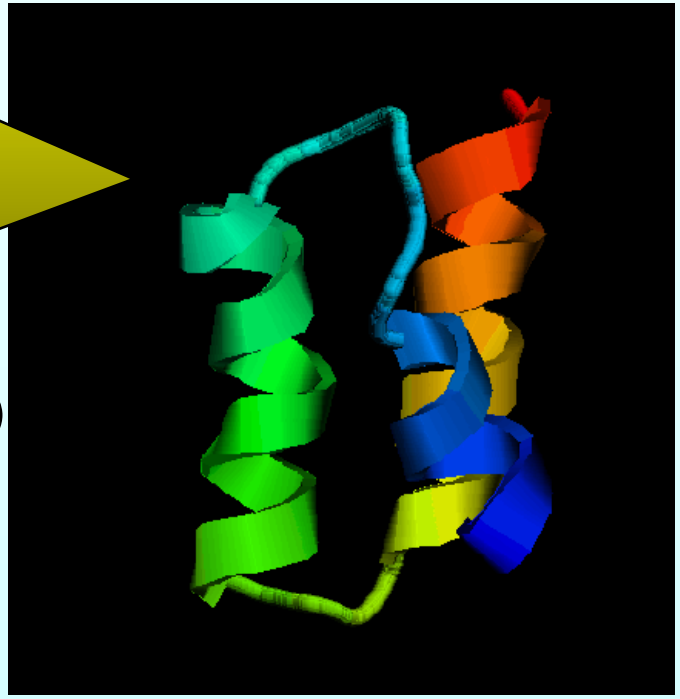


THEME I. *Protein folding.*

Amino-acid sequence – translated genetic code.

MET—ALA—ALA—ASP—GLU—GLU--….

How?

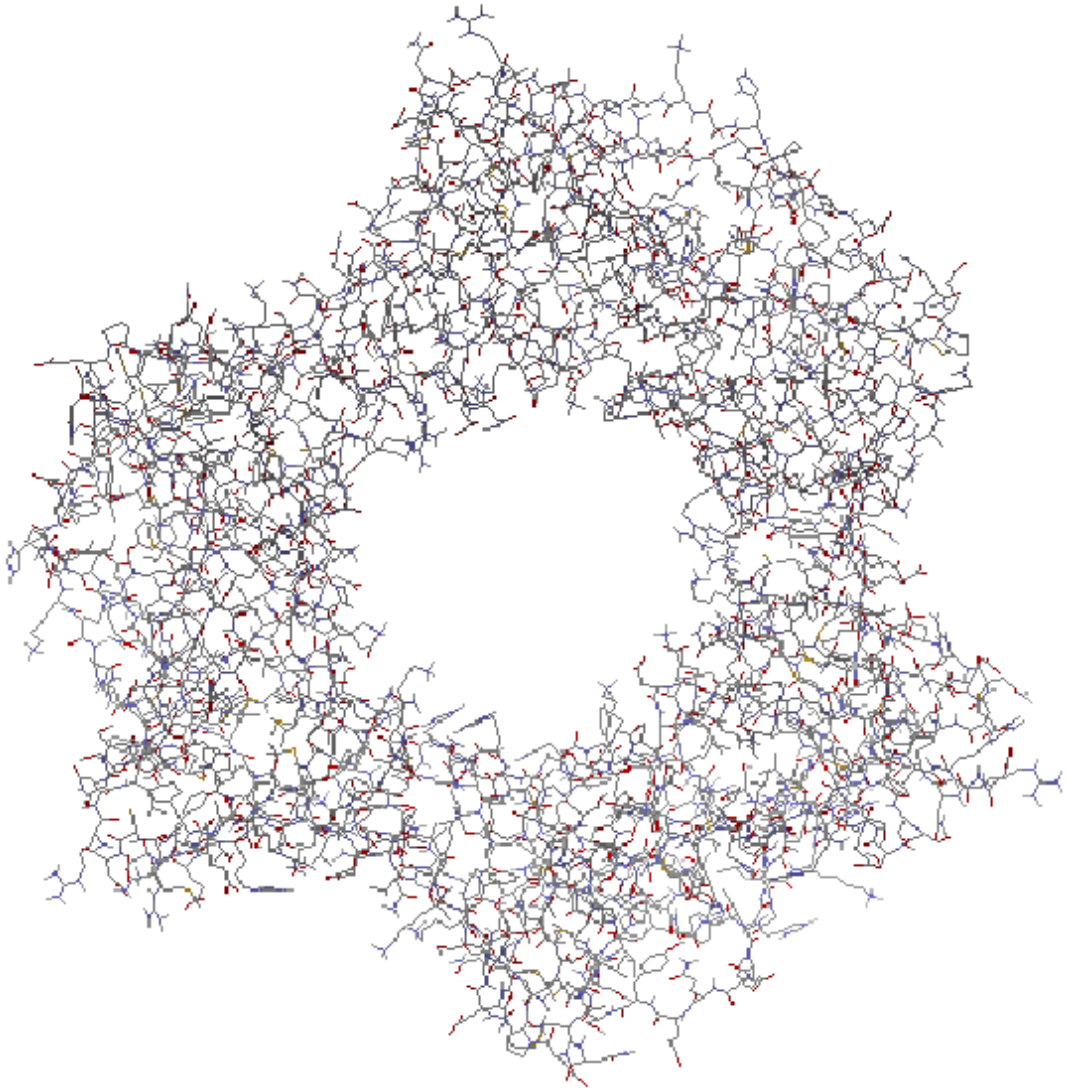


Experiment: amino acid sequence uniquely determines protein's 3D shape (ground state)

Nature does it all the time. Can we?

Complexity of protein design

Example: PCNA – a human DNA-binding protein.



Single amino-acid
(phenilalanin)

Drawn to scale

The magnitude of the protein folding challenge:

A small protein is a chain of ~ 50 mino acids (more for most).

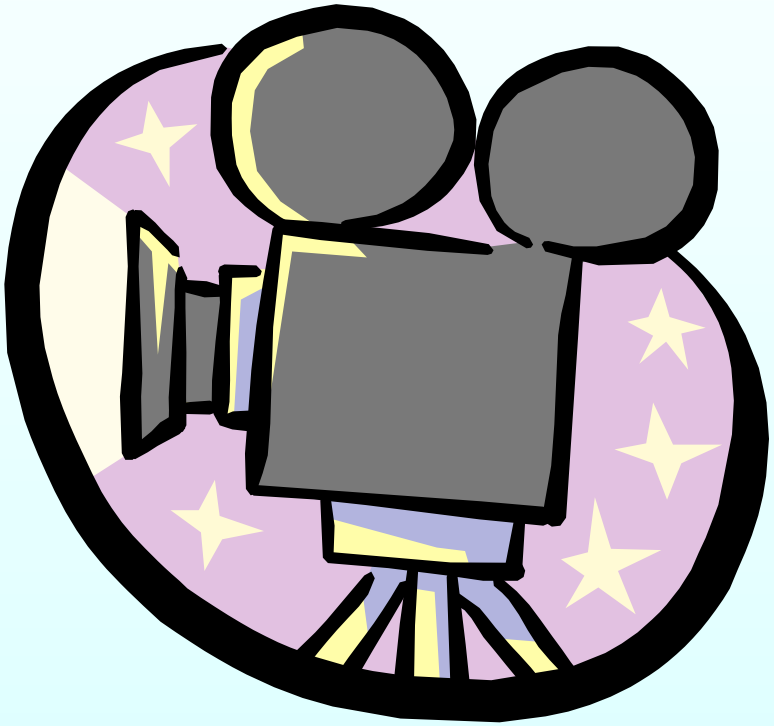
Assume that each amino acid has only 10 conformations (vast underestimation)

Total number of possible conformations: 10^{50}

Say, you make one MC step per femtosecond.

Exhaustive search for the ground state will take 10^{27} years.

Why bother: protein's shape determines its biological function.



Research in Structural Bioinformatics:

SUMMARY:

Through a combination of novel computational approaches we can gain insights into aspects of molecular function inaccessible to experiment and "traditional" (sequence) bioinformatics, and make contributions to both the applied and fundamental science.