PROTEIN PHYSICS

LECTURES 22-23

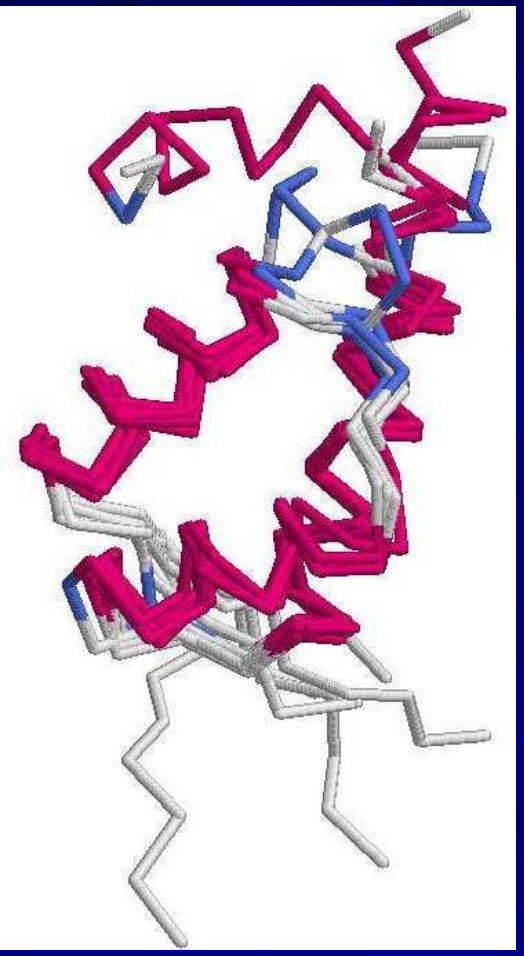
PROTEIN STRUCTURE:

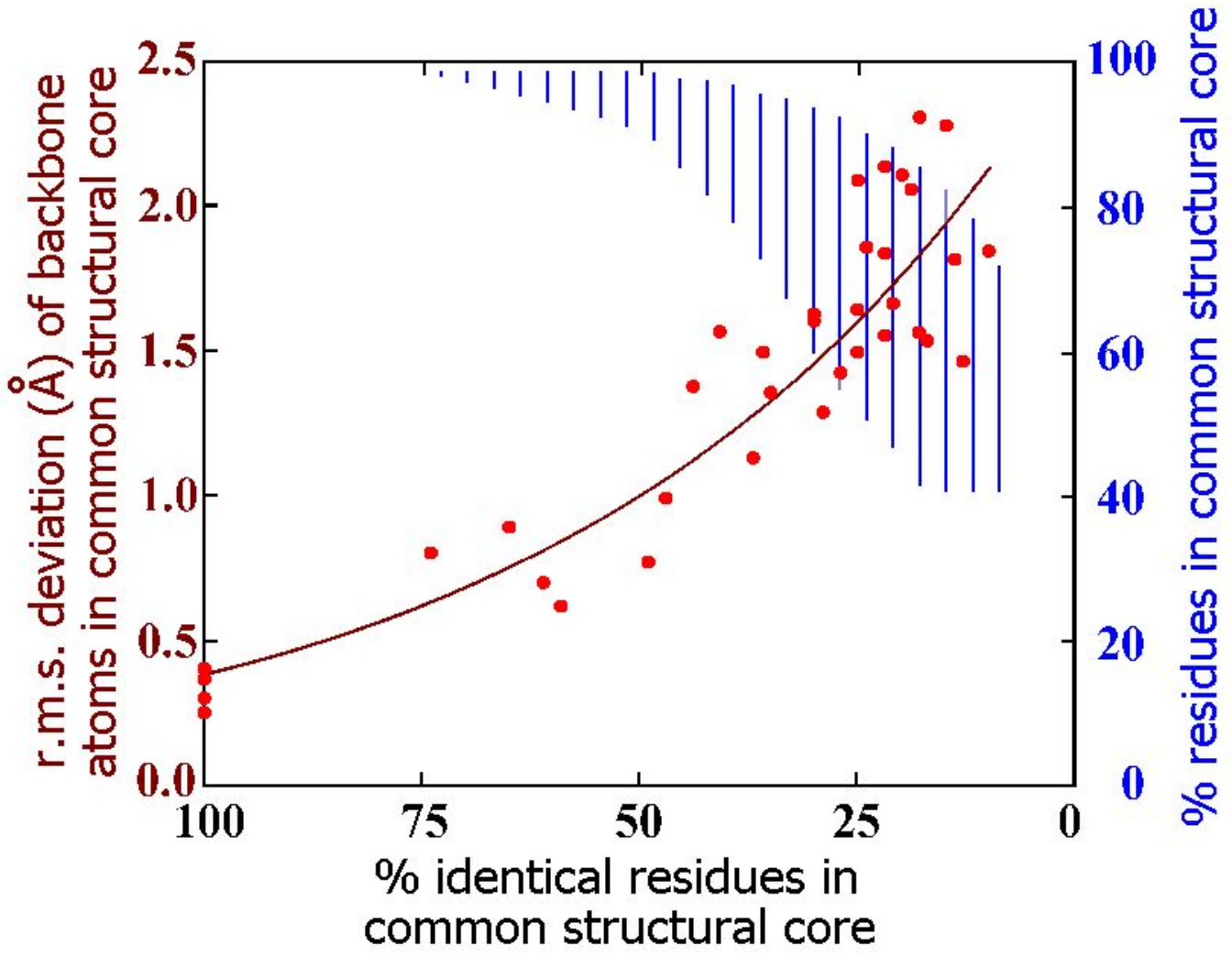
PREDICTION
ENGINEERING
DESIGN

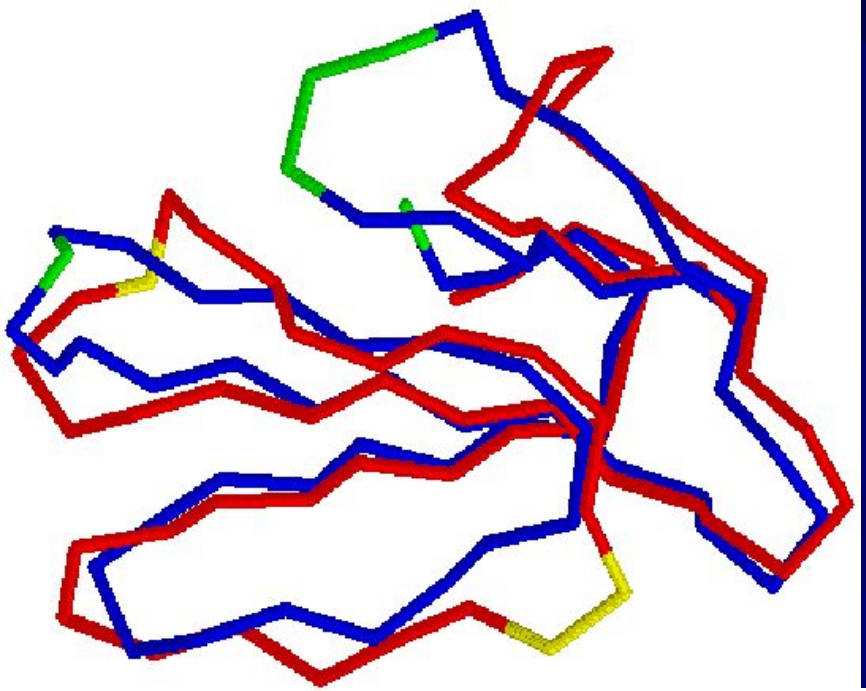
Homology

	1	10
Human	**GD**VE**KG**KK**IF**...	
Rust fungus	GFED**GD**AK**KG**AR**IF**...	

Sequence identity: 60%







PREDICTION FROM HOMOLOGY

SIMILAR SEQUENCES

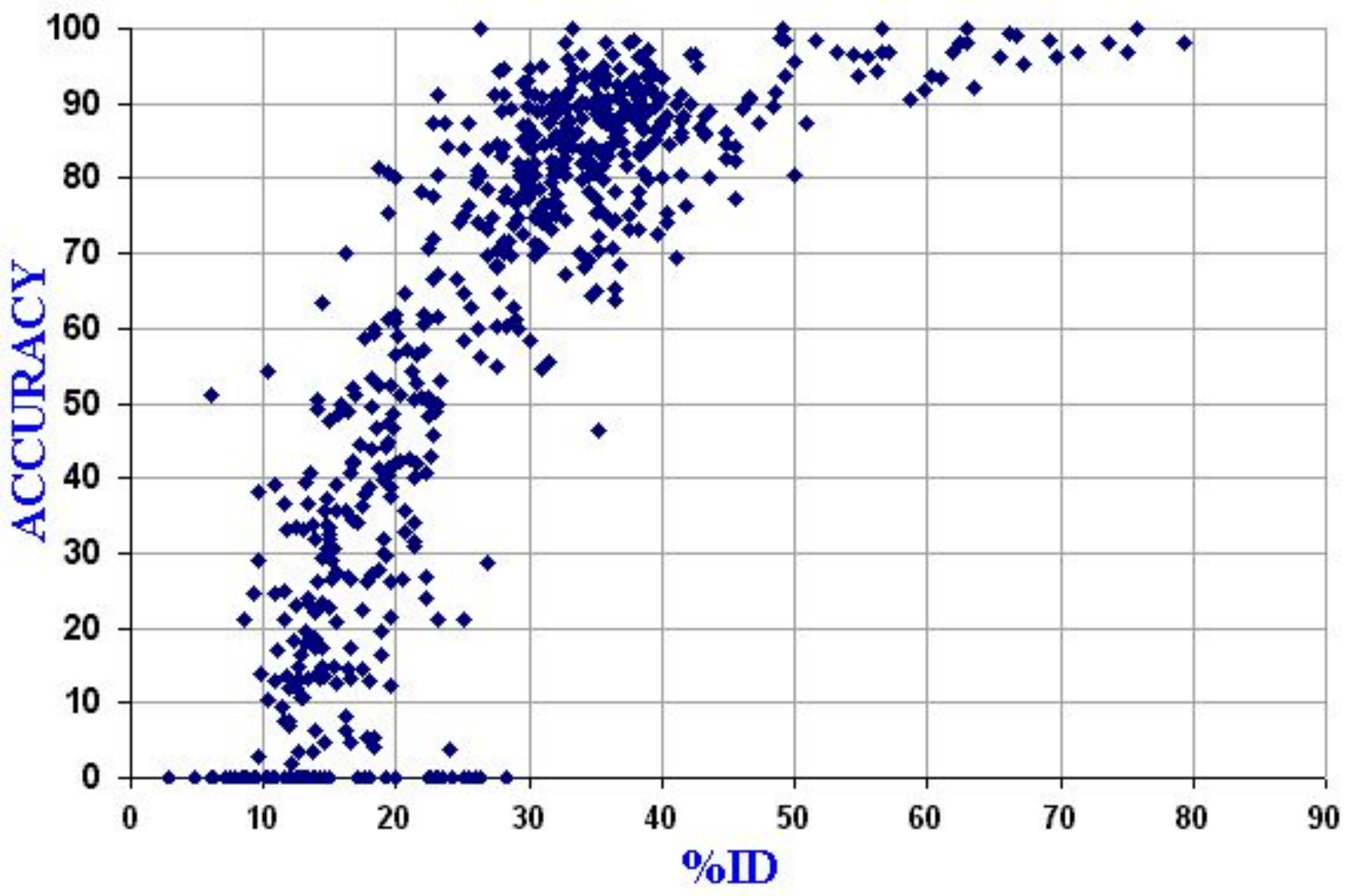
SIMILAR FOLDS

STRUCTURAL ALIGNMENT. SEQ. ID. = 32% ↓

```
1          lkCnqri---PpfwkTCpkGknlCYkmtmraap-mvpvkRGC...
2          riCfnhqssqPqttkTCspGessCYhkqwsdf-rgtiieRGC...
                     |||||||||||||||||||||| ////////
2     -----RIcfnhqssqPqttKTCspGessCYhkqwsdfrgtiieRGC...
1     lkcnqRI--------PpfwKTCpkGknlCYkmtmraapmvpvkRGC...
```

SEQUENCE ALIGNMENT ↑

SEQUENCE ALIGNMENT: **BIOINFORMATICS**

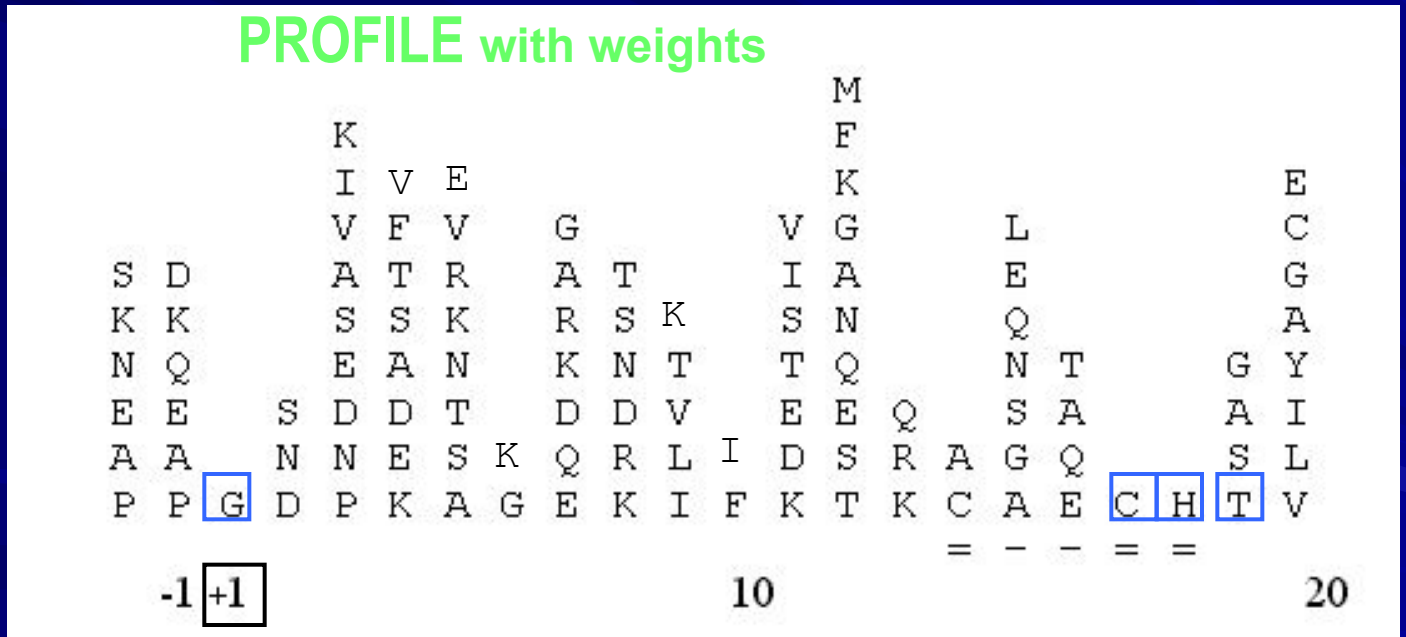


N0 □TWILIGHT□ ======= GOOD PREDICTION =======

Multiple homology

```
                                    1         10        20
Human, chimpanzee                   GDVEKGKKIFIMKCSQCHTV...
Pig, bovine, sheep                  GDVEKGKKIFVQKCAQCHTV...
Chicken, turkey                     GDIVEKGKKIVQKCSQCHTV...
Puget sound dogfish                 GDVEKGKKVFVQKCAQCHTV...
Screw-worm fly                GVPAGDVEKGKKIFVQRCAQCHTV...
Rust fungus                   GFEDGDAKKGARIFKTRCAQCHTL...
Rape, cauliflower       ASFDEAPPGNSKAGEKIFKTKCAQCHTV...
```

PROFILE with weights

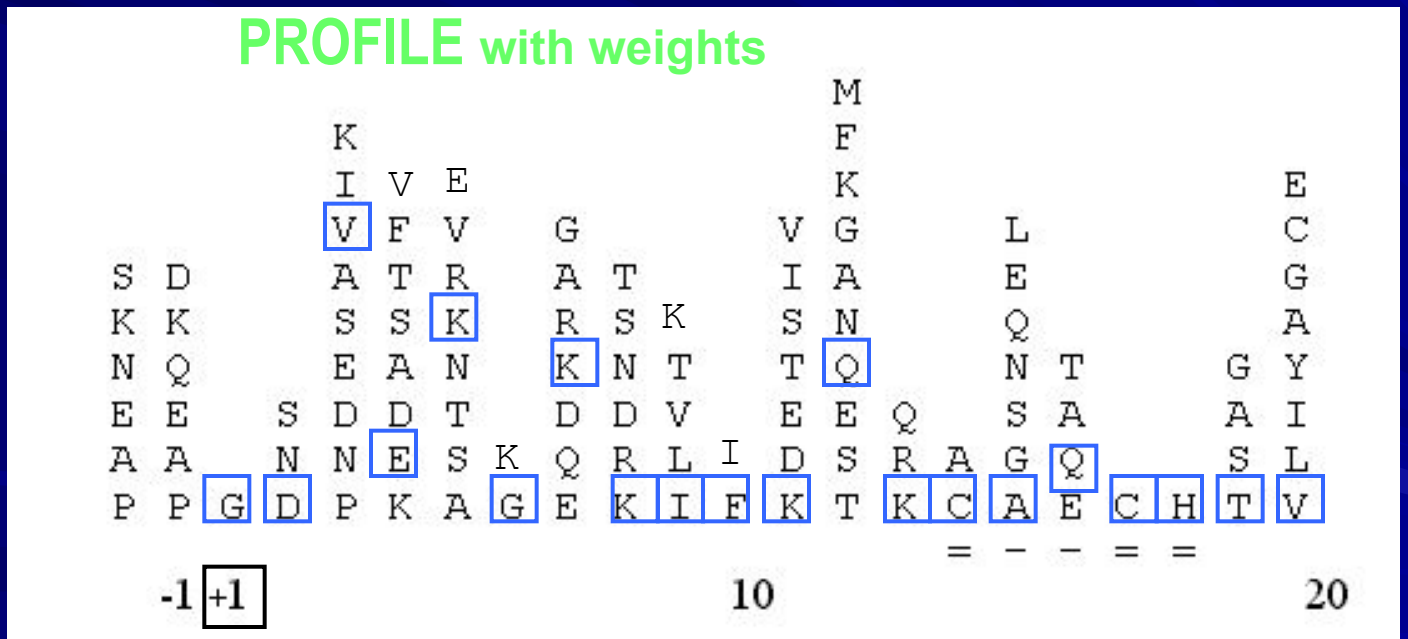


TARGET SEQUENCE . . . A P G D E F G – – H I K K L M A A T C H A L . . .

Multiple homology

```
                                    1         10        20
Human, chimpanzee                   GDVEKGKKIFIMKCSQCHTV...
Pig, bovine, sheep                  GDVEKGKKIFVQKCAQCHTV...
Chicken, turkey                     GDIVEKGKKIVQKCSQCHTV...
Puget sound dogfish                 GDVEKGKKVFVQKCAQCHTV...
Screw-worm fly                GVPAGDVEKGKKIFVQRCAQCHTV...
Rust fungus                   GFEDGDAKKGARIFKTRCAQCHTL...
Rape, cauliflower       ASFDEAPPGNSKAGEKIFKTKCAQCHTV...
```

PROFILE with weights

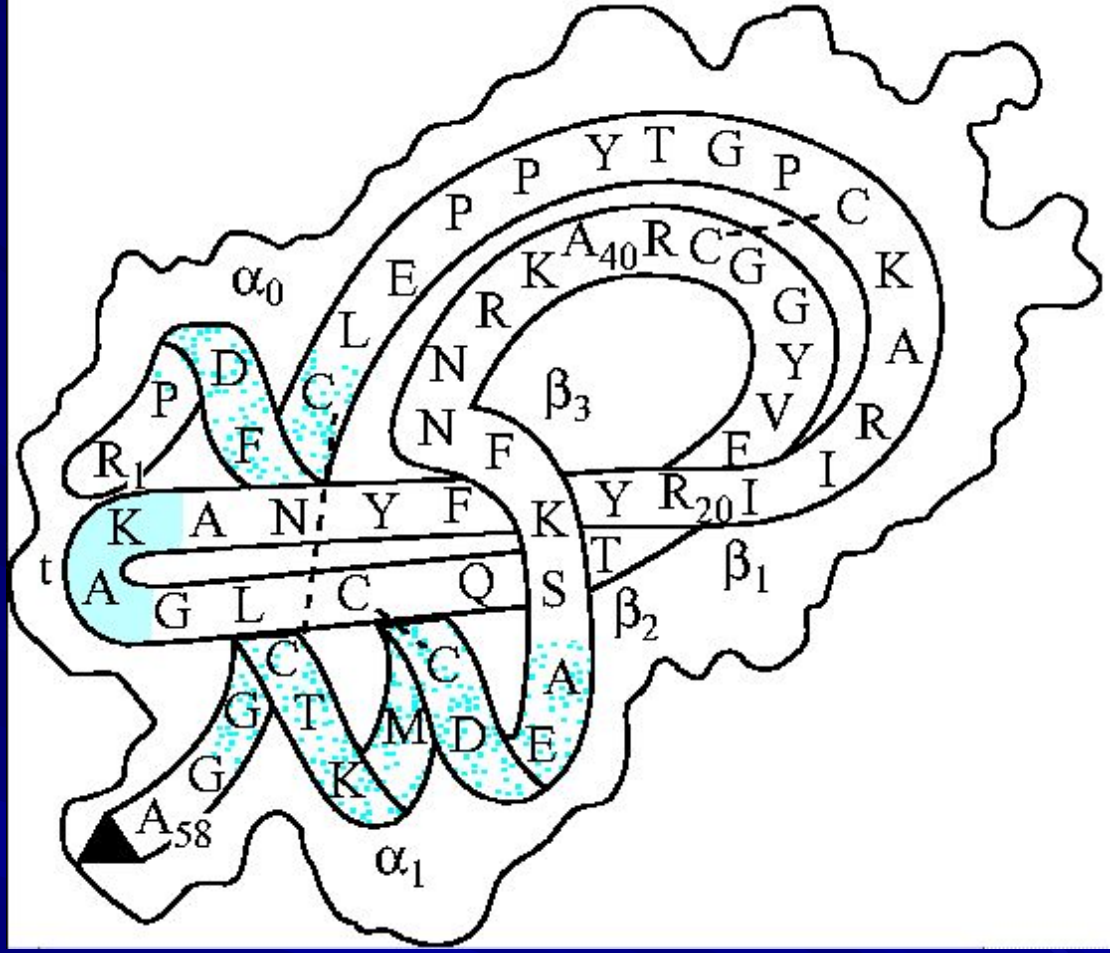


TARGET SEQUENCE . . . A P G D E F G – – H I K K L M A A T C H A L . . .

PREDICTION FROM PHYSICS:

PROTEIN CHAIN FOLDS SPONTANEOUSLY

- SEQUENCE HAS ALL INFO TO PREDICT:



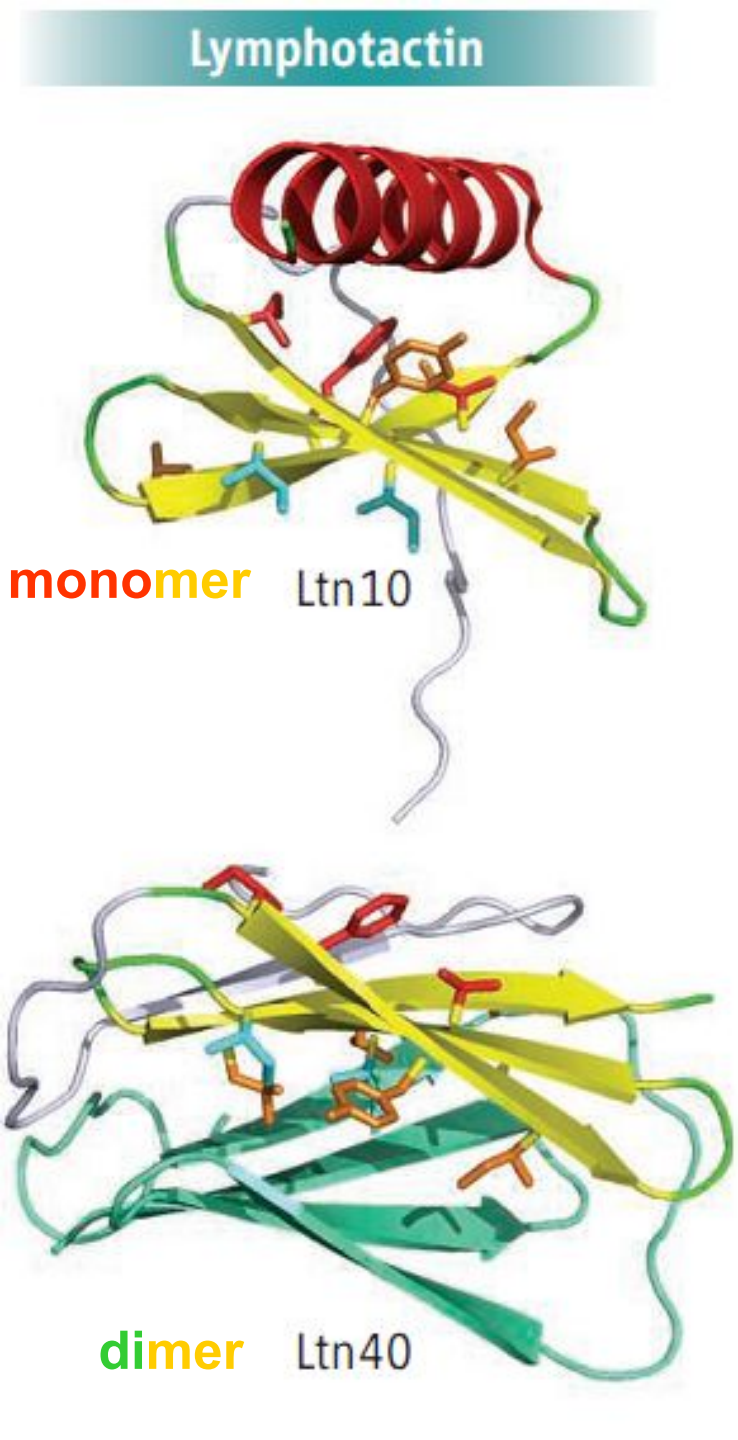
2^O STRUCTURE,
3D STRUCTURE,
SIDE CHAIN ROTAMERS,
S-S BONDS, etc.

“Unique” fold?

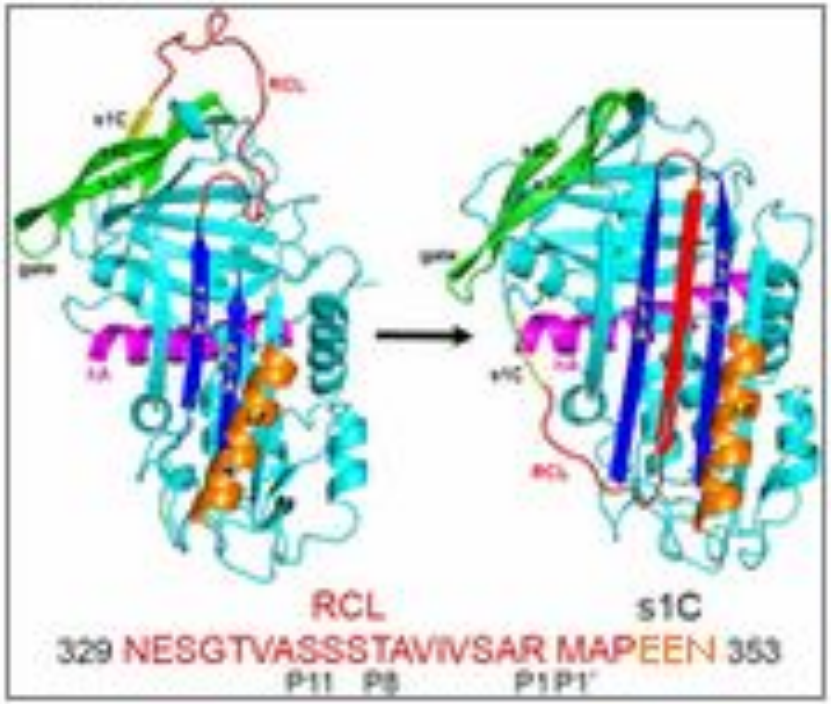
Dimerization
involves an isomerization of the β-sheet.

Structurally equivalent residues are few and contribute either to the Ltn10 core (red) or to the dimeric interface of Ltn40 (cyan).

Other nonpolar residues (orange) change sides, such that the formation of the dimeric interface on one side of the β-sheet destroys the hydrophobic core on the other side and vice versa.



“Unique” fold?



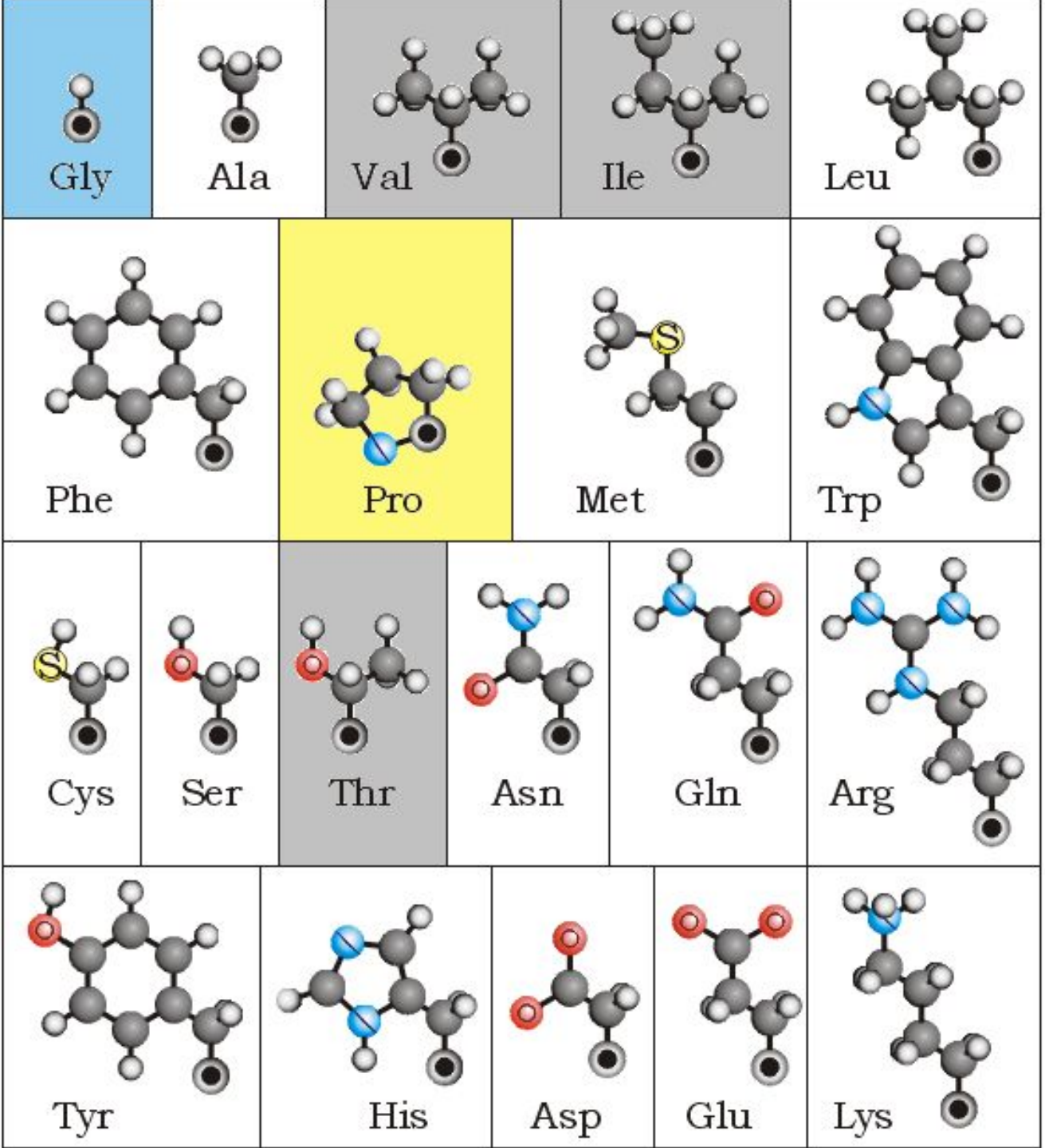
Serpin latency transition at atomic resolution

G. Cazzolli, F. Wang, S. a Beccara, A. Gershenson, P. Faccioli, P. L. Wintrode

PNAS October 28, 2014 vol. 111 no. 43 15414-15419

active METASTABLE form (~ 30 min.)

INactive STABLE form



Side chains

no C^{β}
C^{β}, ≤one γ
C^{β}, two γ
imino-acid

no C^{β}: coil

C^{β}, ≤1 γ: α, β, coil

C^{β}, 2 γ: β

imino: coil, turn

C^{α}
C
H
N
O
S

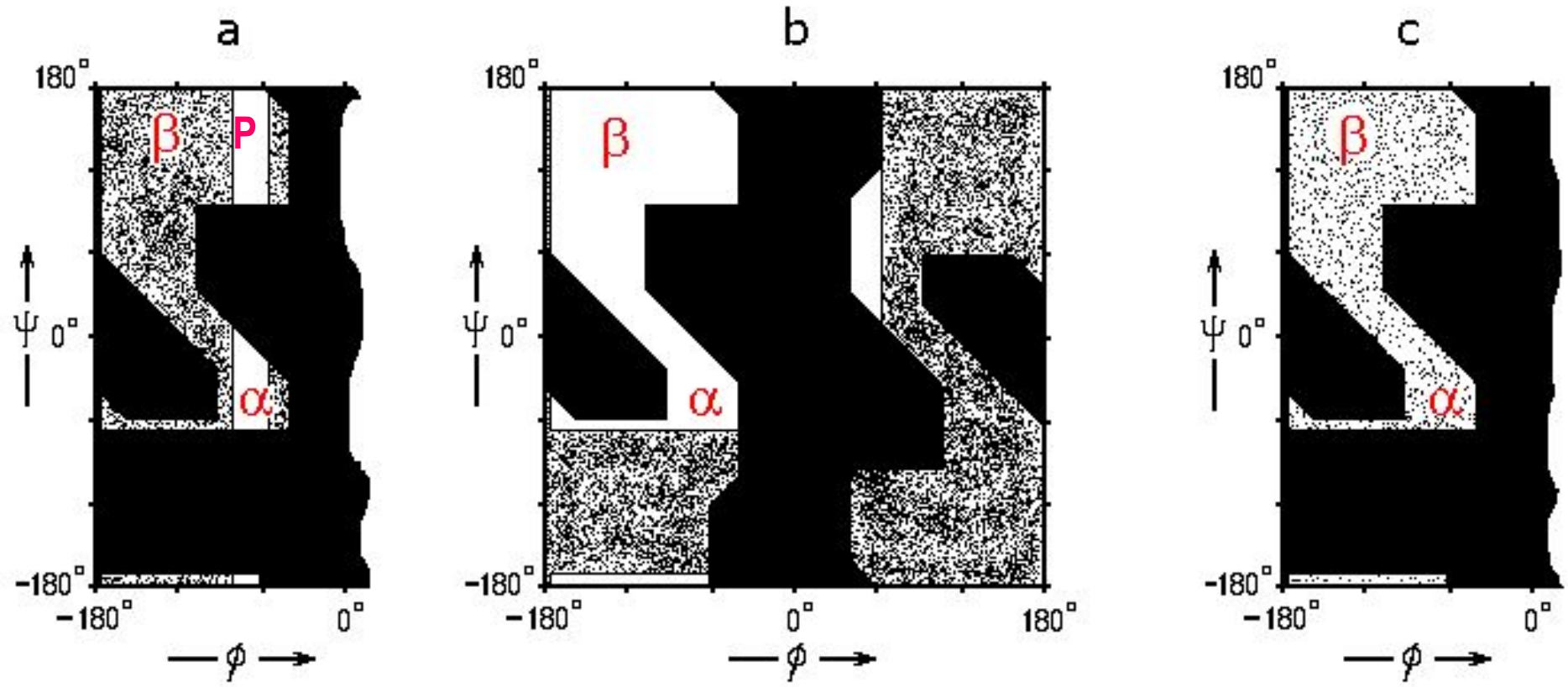
Pro

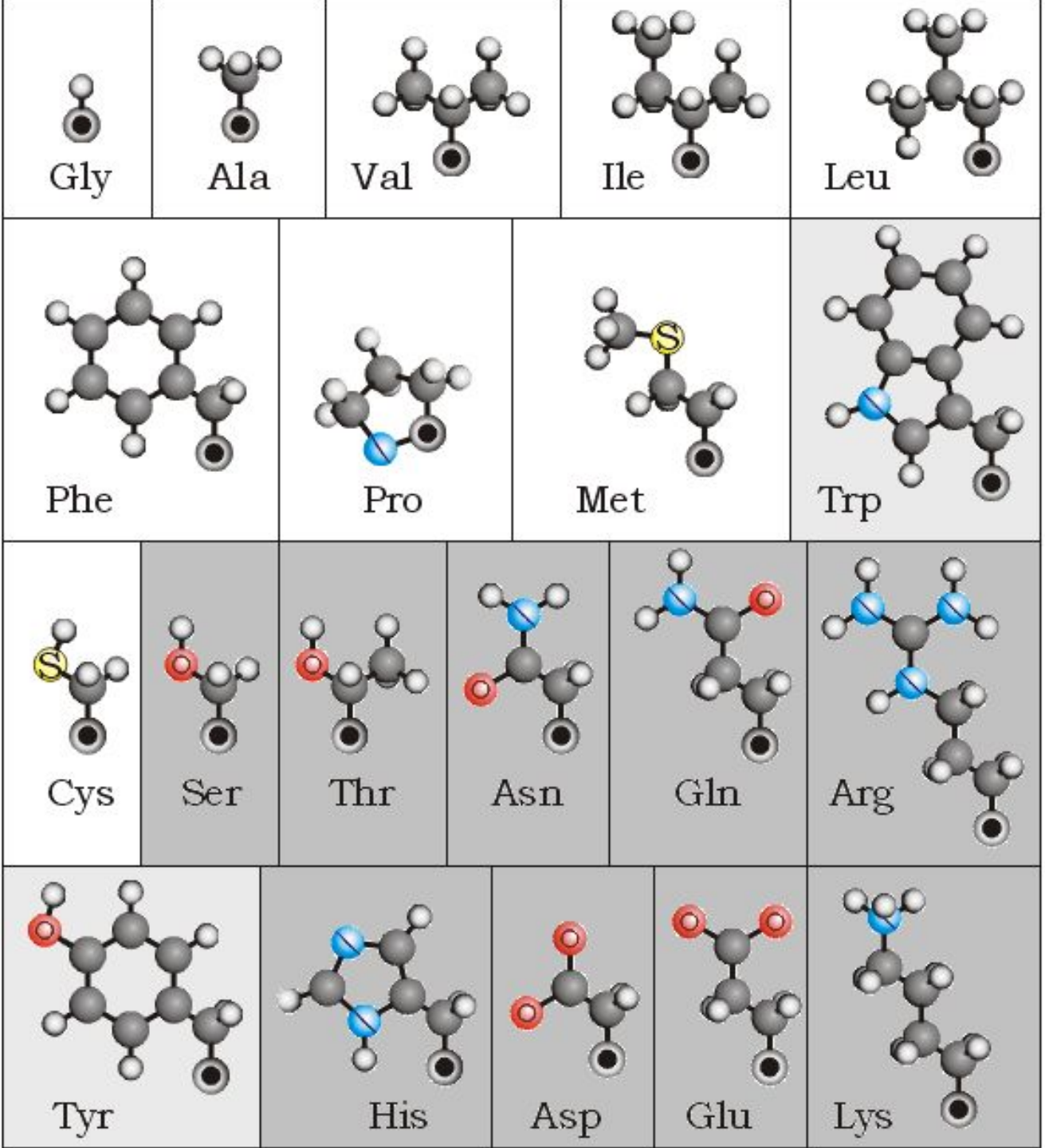
1,2,3 rot.

imino:
coil, turn, α_N

no C^β: coil
C^β, ≤1 γ: α, β, coil

C^β, 2 γ: β





Side chains

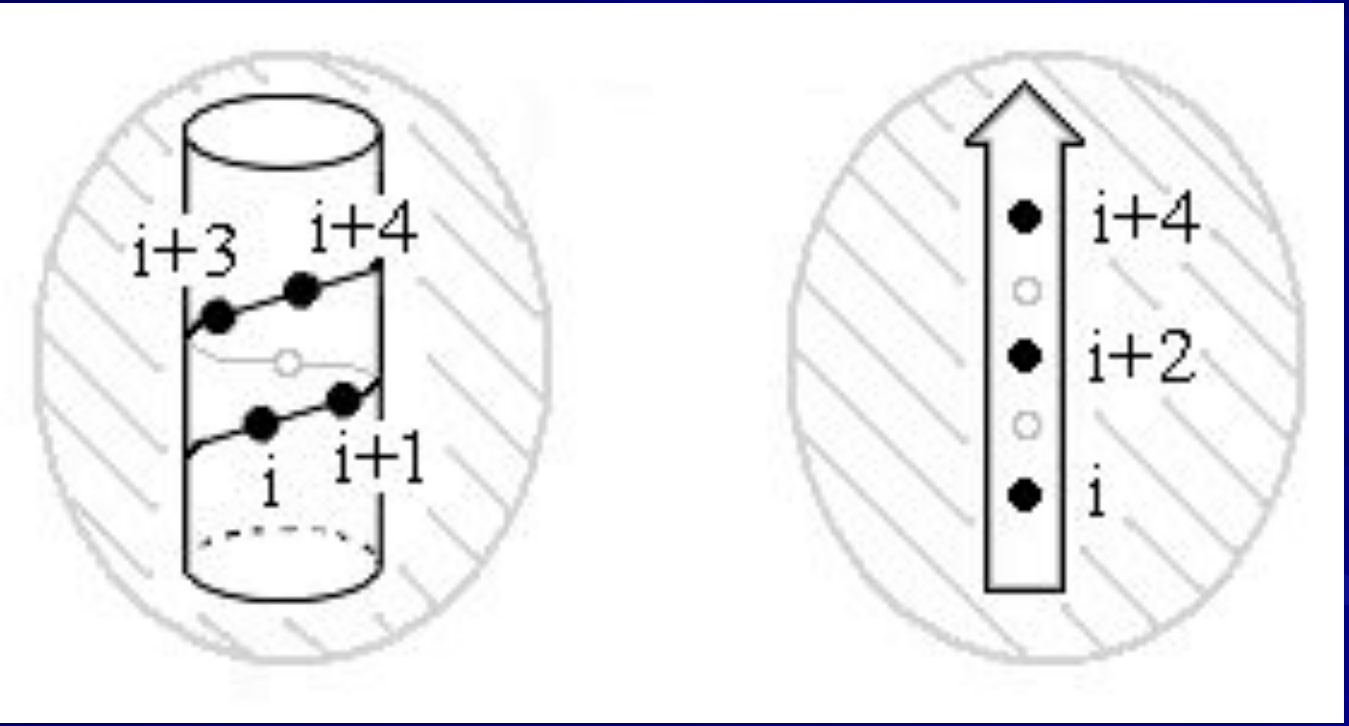
non-polar
polar

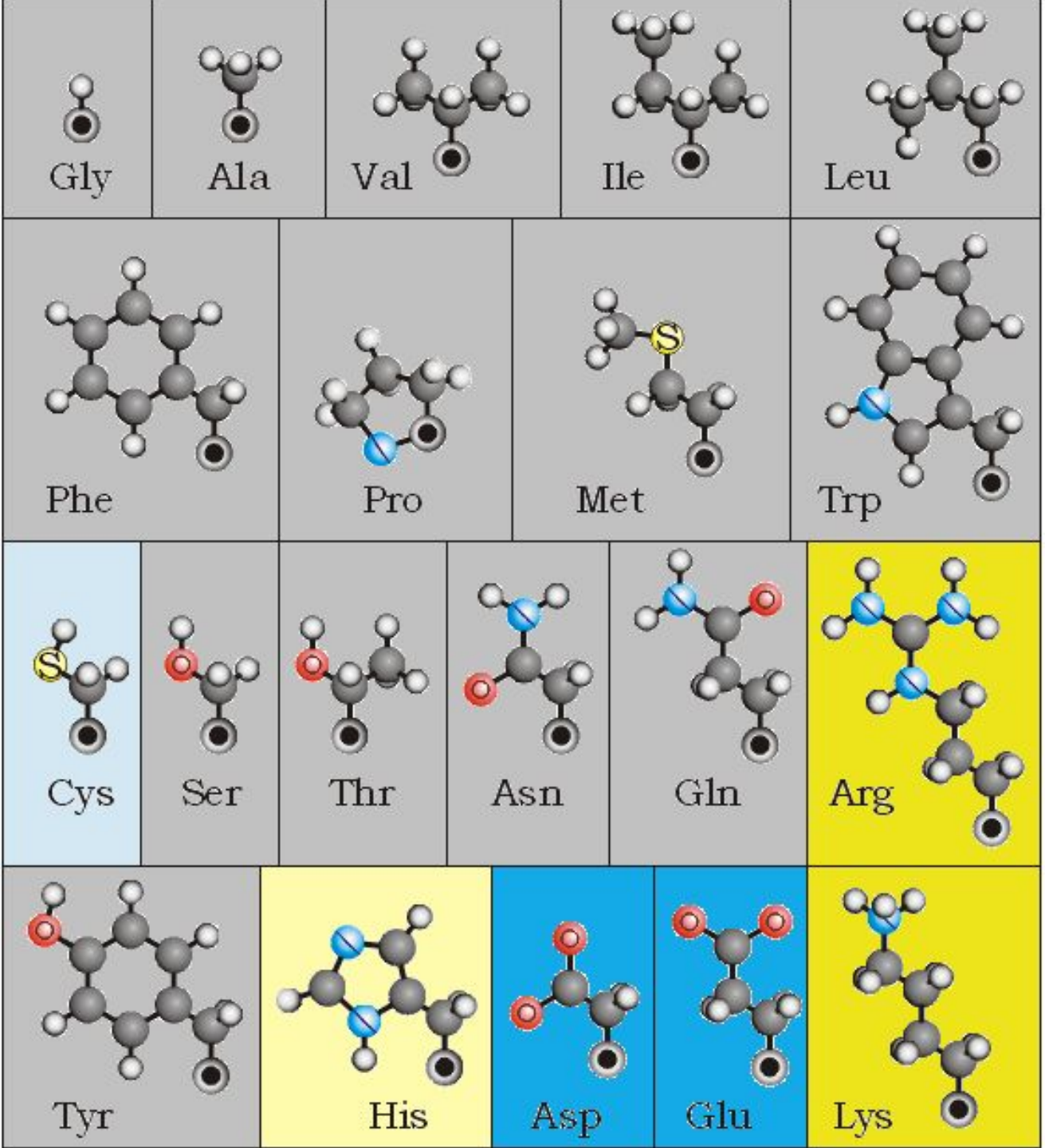
non-polar: core

polar: surface

non-polar: in the core

polar: at the surface





Side chains

charged −

charged +

charged −: coil, α_N

charged +: coil, α_C

Half-charged: active sites

C^{α}

C

H

N

O

S

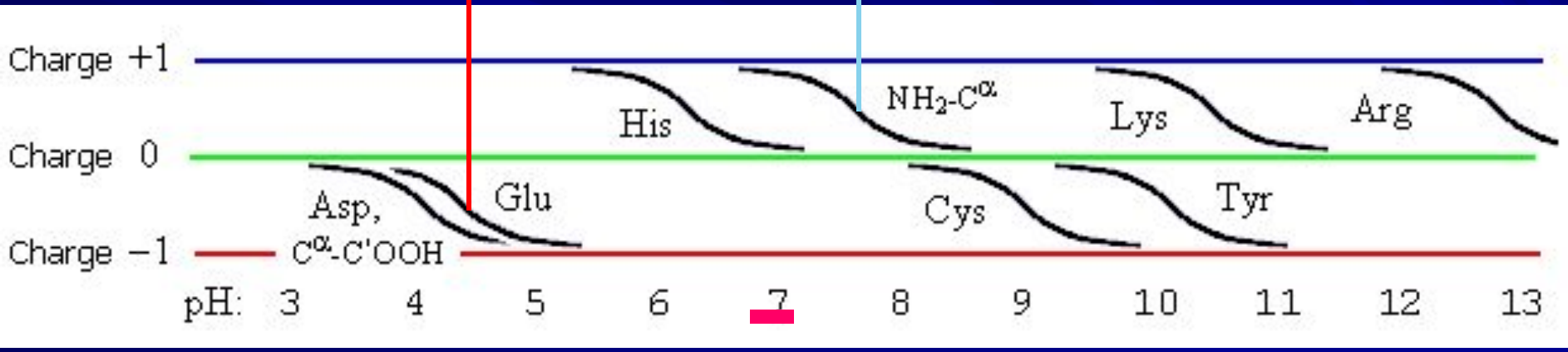
$$P_{charged} + P_{uncharged} = 1$$

$$P_{charged} / P_{uncharged} = 10^{-(pKa - pH)}$$

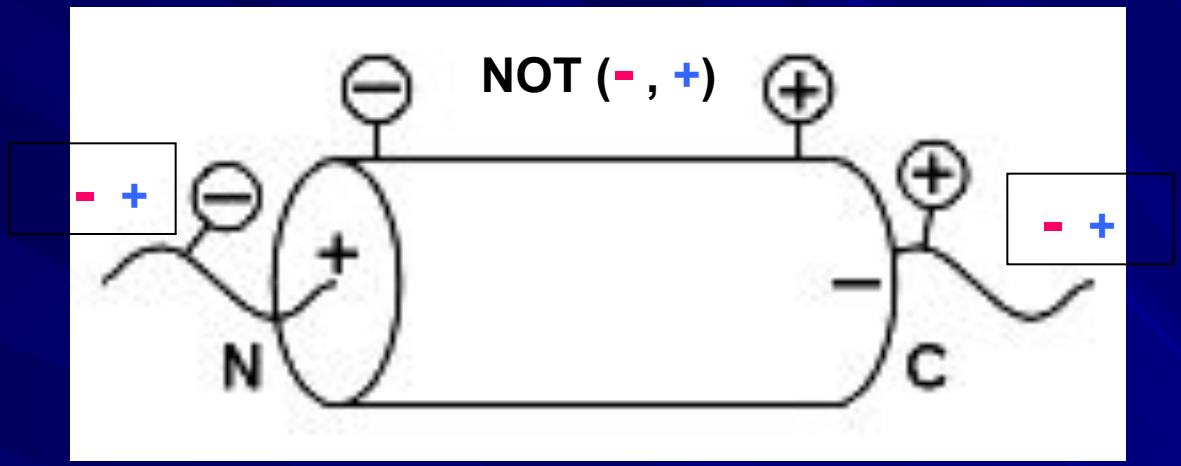
$$P_{charged} / P_{uncharged} = 10^{+(pKa - pH)}$$

Acids (charge –)

Bases (charge +)



Half-charged: active sites



charged –: coil, α_N

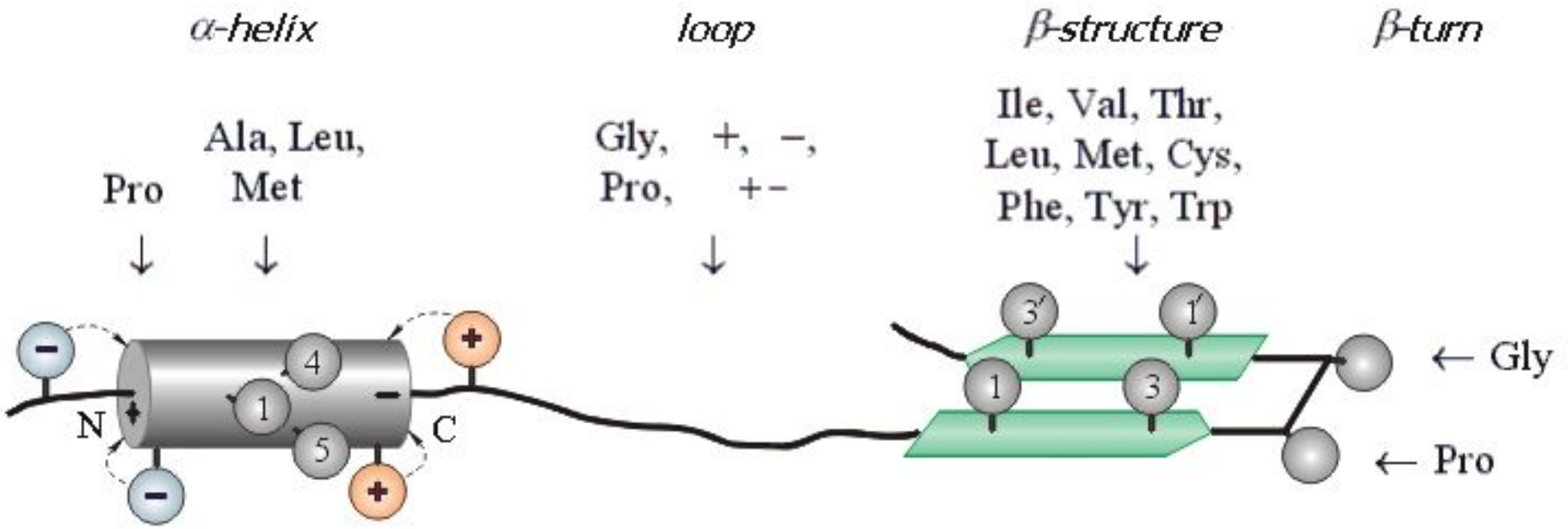
====

charged +: coil, α_C

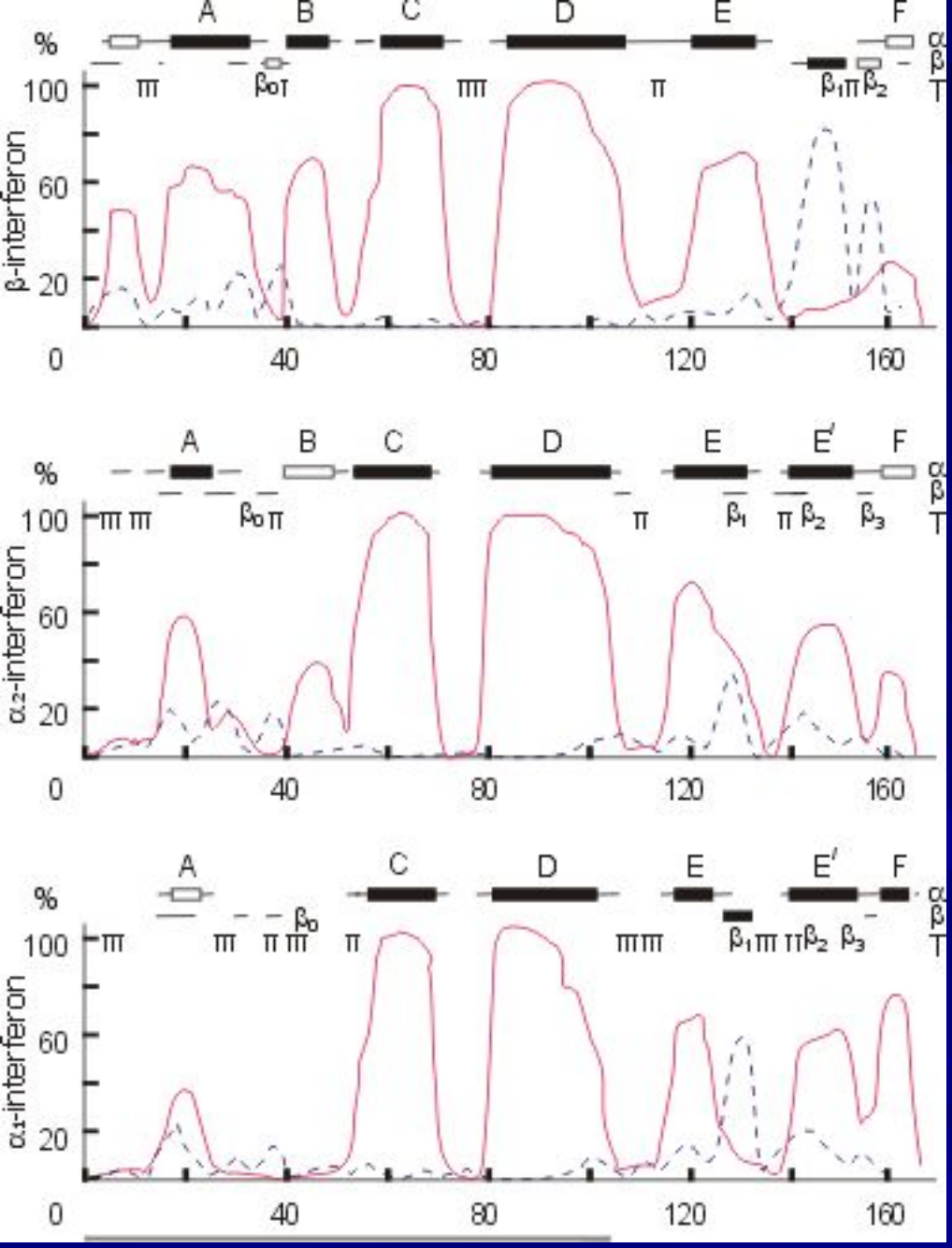
====

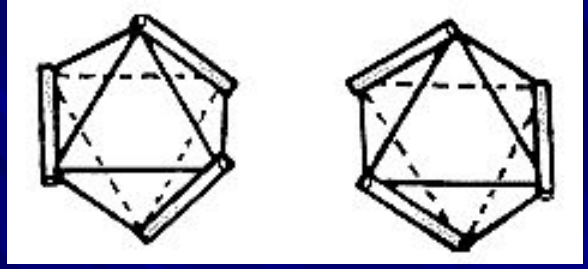
PREDICTION FROM PHYSICS (OR PROTEIN STATISTICS)

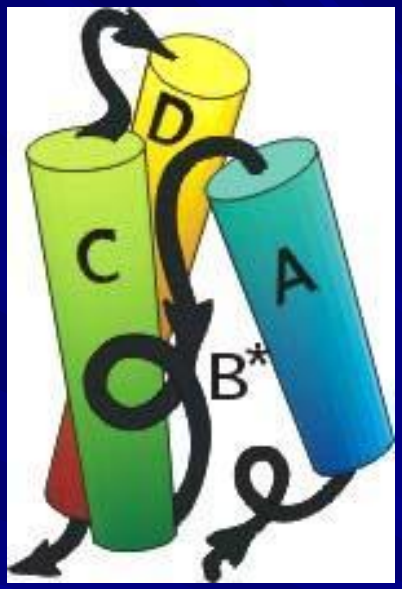
2^{O} STRUCTURES

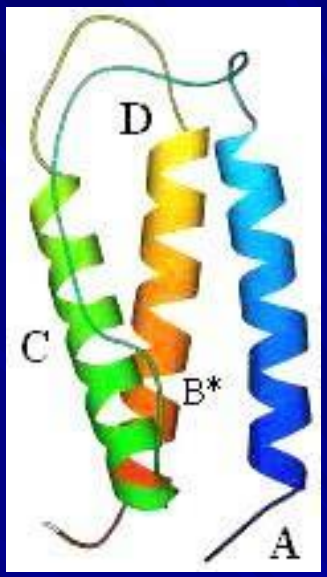


USUALLY, THIS WORKS WELL, BUT…



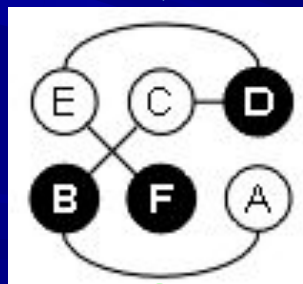




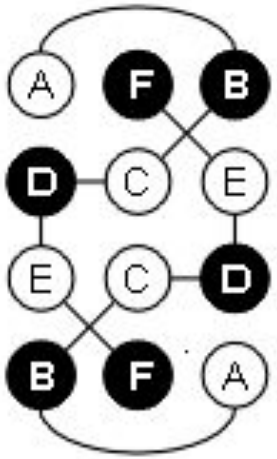


Prediction, 1985

X-ray str.,1990



β



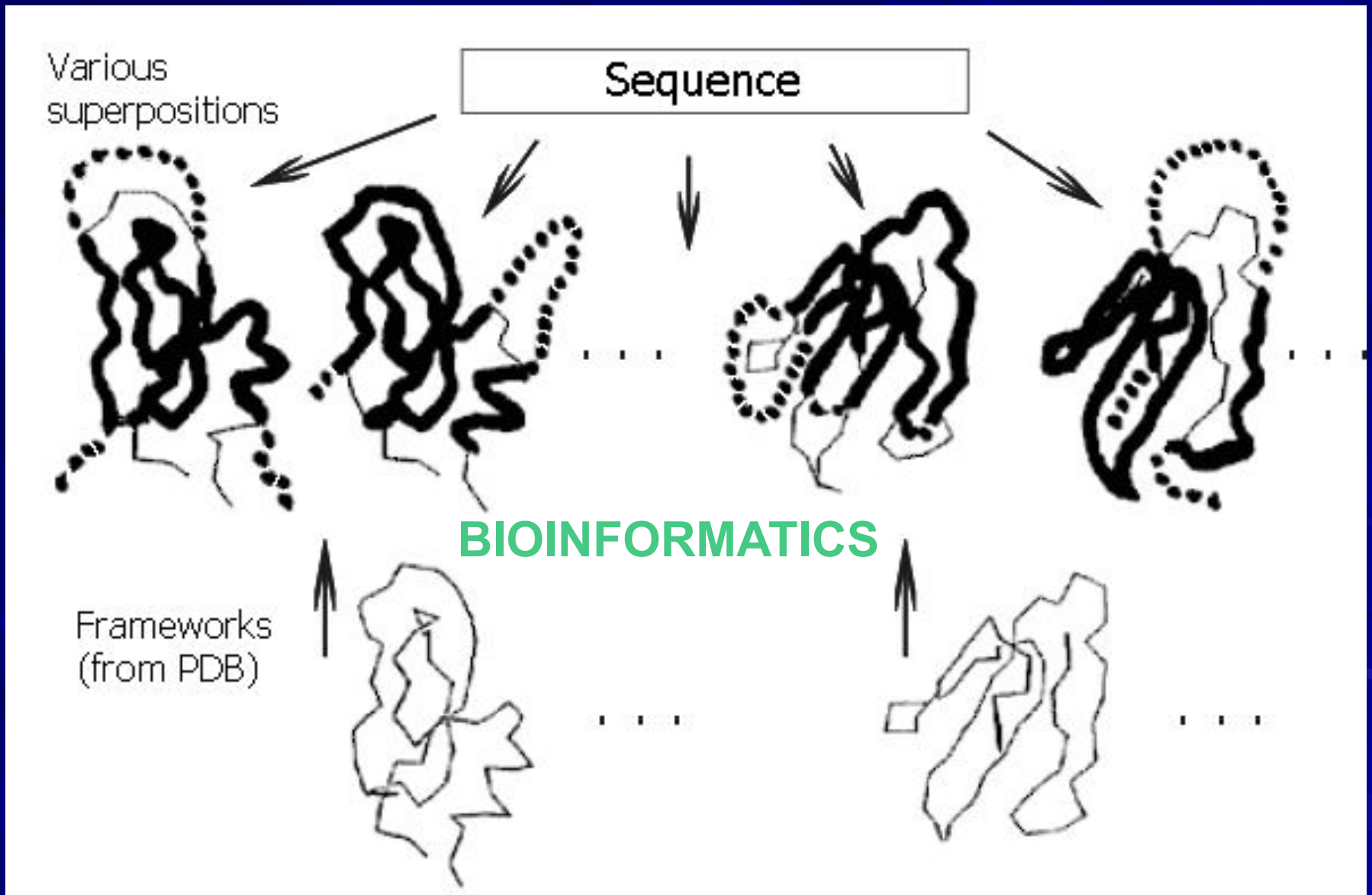
γ

A B C D .---different---

THREADING

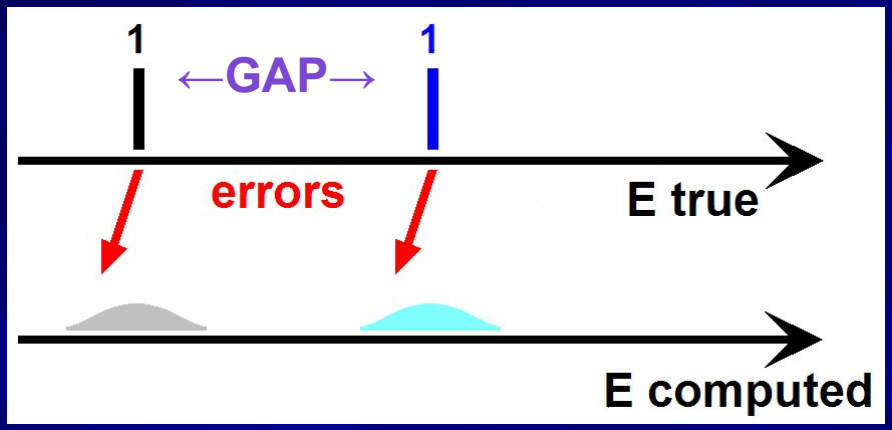
helps, when sequence identity is low (<10-20%)

Finkelstein, Reva, 1990-91 (*Nature*); Bowie, Lüthy, Eisenberg, 1991 (*Science*)

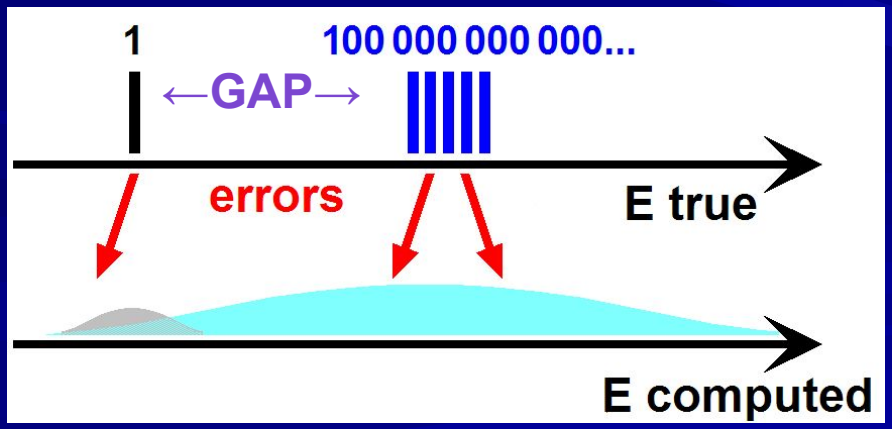


… but one still cannot reliably predict 3D protein structure from the a. a. sequence without homologues… **WHY??**

choice of one structure out of two:
DOES NOT require too precise estimate of interactions

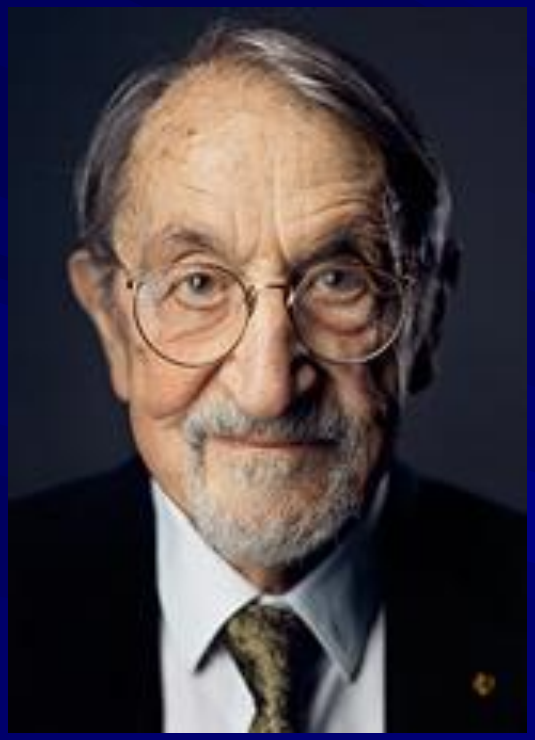


choice of one structure out of zillions:
REQUIRES very precise estimate of interactions



HOT POINTS IN PROTEIN PHYSICS

The Nobel Prize in Chemistry 2013



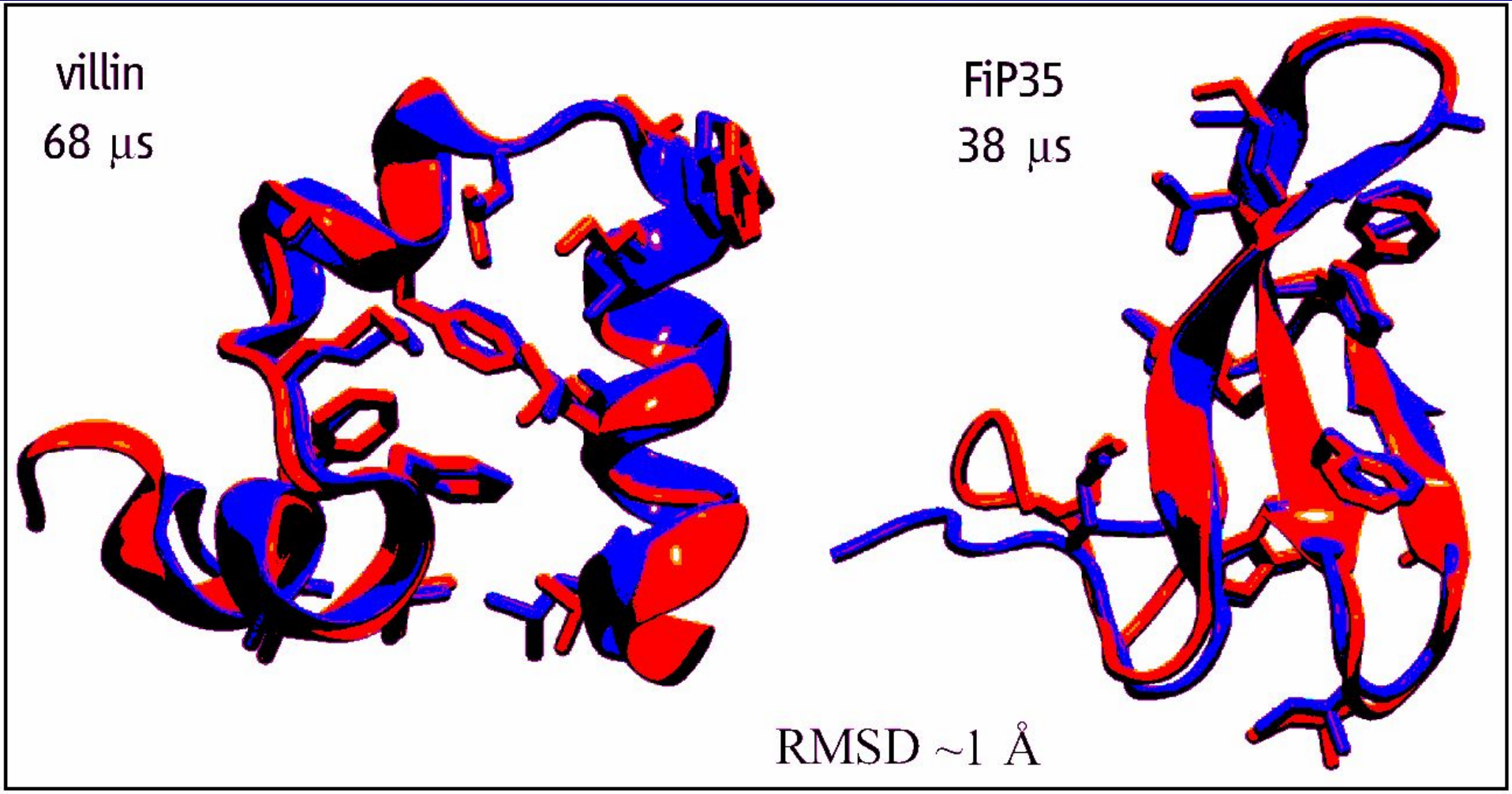
Martin Karplus,
1930

Michael Levitt,
1947

Arieh Warshel,
1940

"for the development of multiscale models for complex chemical systems"

Predicting 3D structures of small proteins



modified version of the Amber ff99SB force field:

K. Lindorff-Larsen *et al.*, *Proteins* **78**, 1950 (2010)

D.E. Shaw et al., Oct. 2010, *Science* **330**, 341

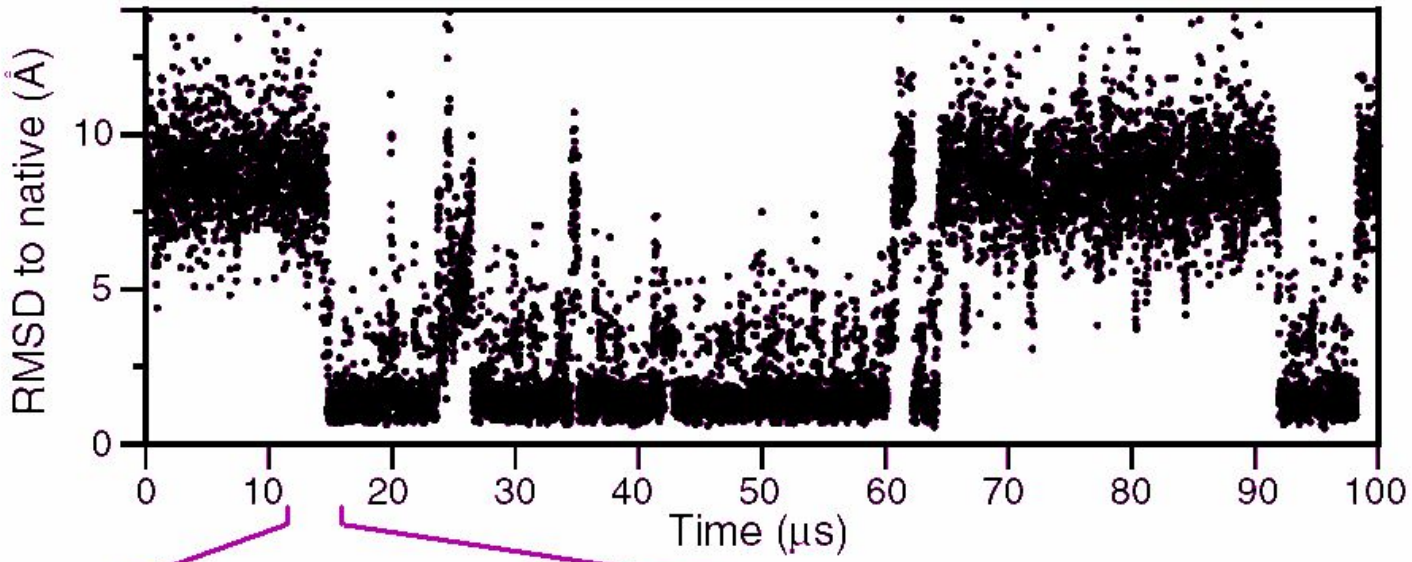
HOT POINTS IN PROTEIN PHYSICS

David E. Shaw, 1951

“D. E. Shaw Research”
US$ 3.5 billion
Supercomputer “Anton”

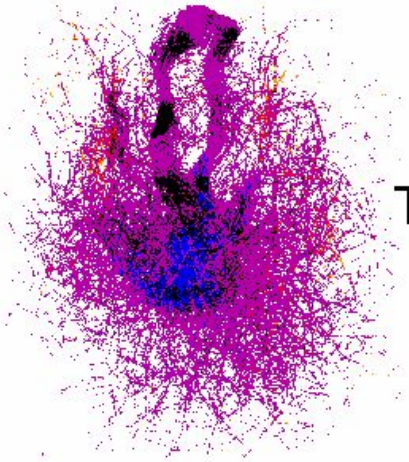
FiP35 protein: simulation of folding

D.E. Shaw et al., Oct. 2010, *Science* **330**, 341



folding pathway

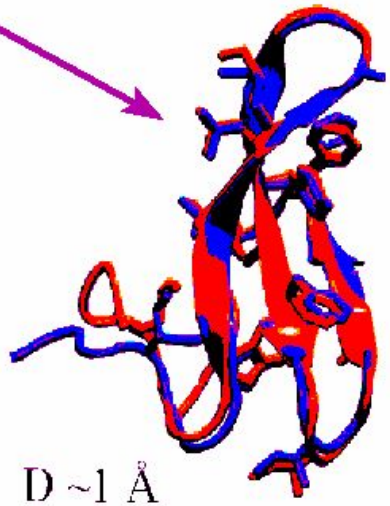
Coil



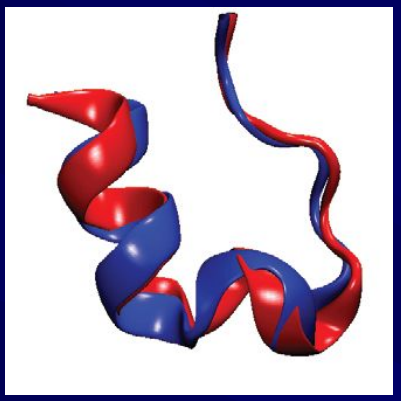
TS ensamble

phase separation

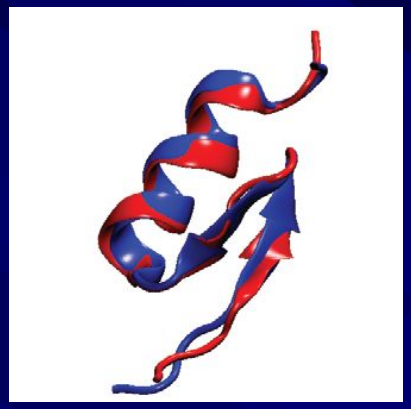
Native



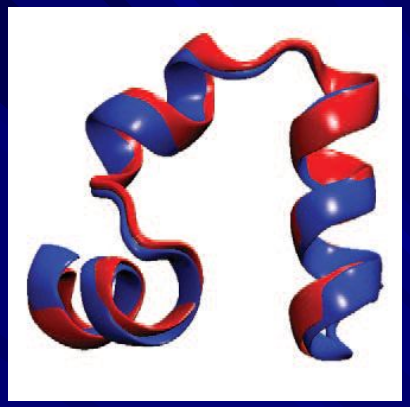
D ~1 Å



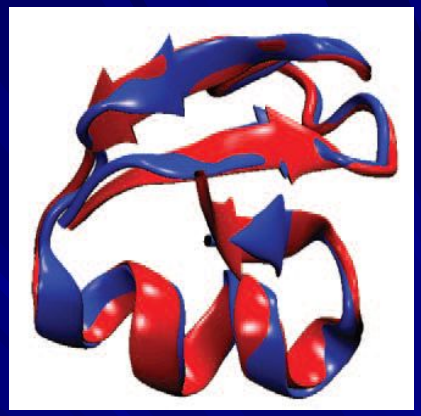
Trp-cage 208μs
1.4Å 14μs



BBA 325μs
1.6Å 18μs

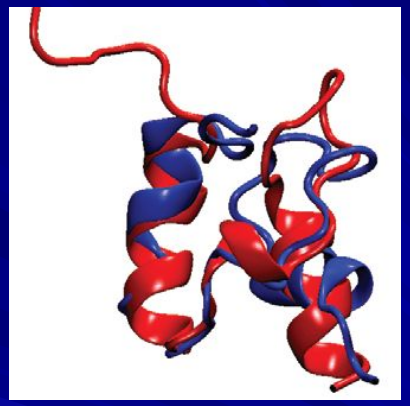


Villin 125μs
1.3Å 2.8μs



NTL9 3936μs
0.5Å 29μs

In total - 12 proteins

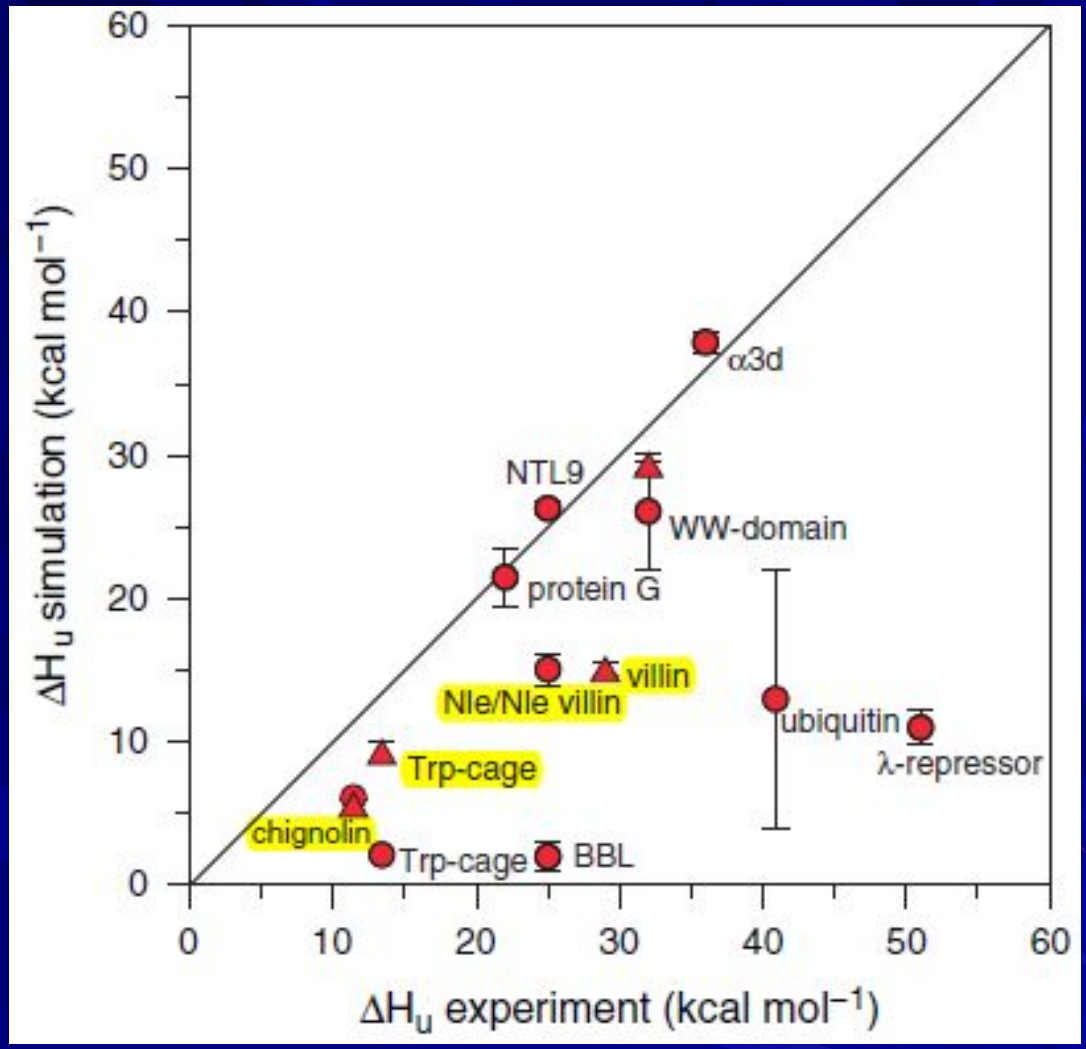


BBL 429μs
4.8Å 29μs

K. Lindorff-Larsen, S. Piana, R.O. Dror, D. E. Shaw (2011)
How Fast-Folding Proteins Fold. *Science* 334, 517

BUT:
comparison of experimental and simulation-derived unfolding enthalpies
shows very large differences…

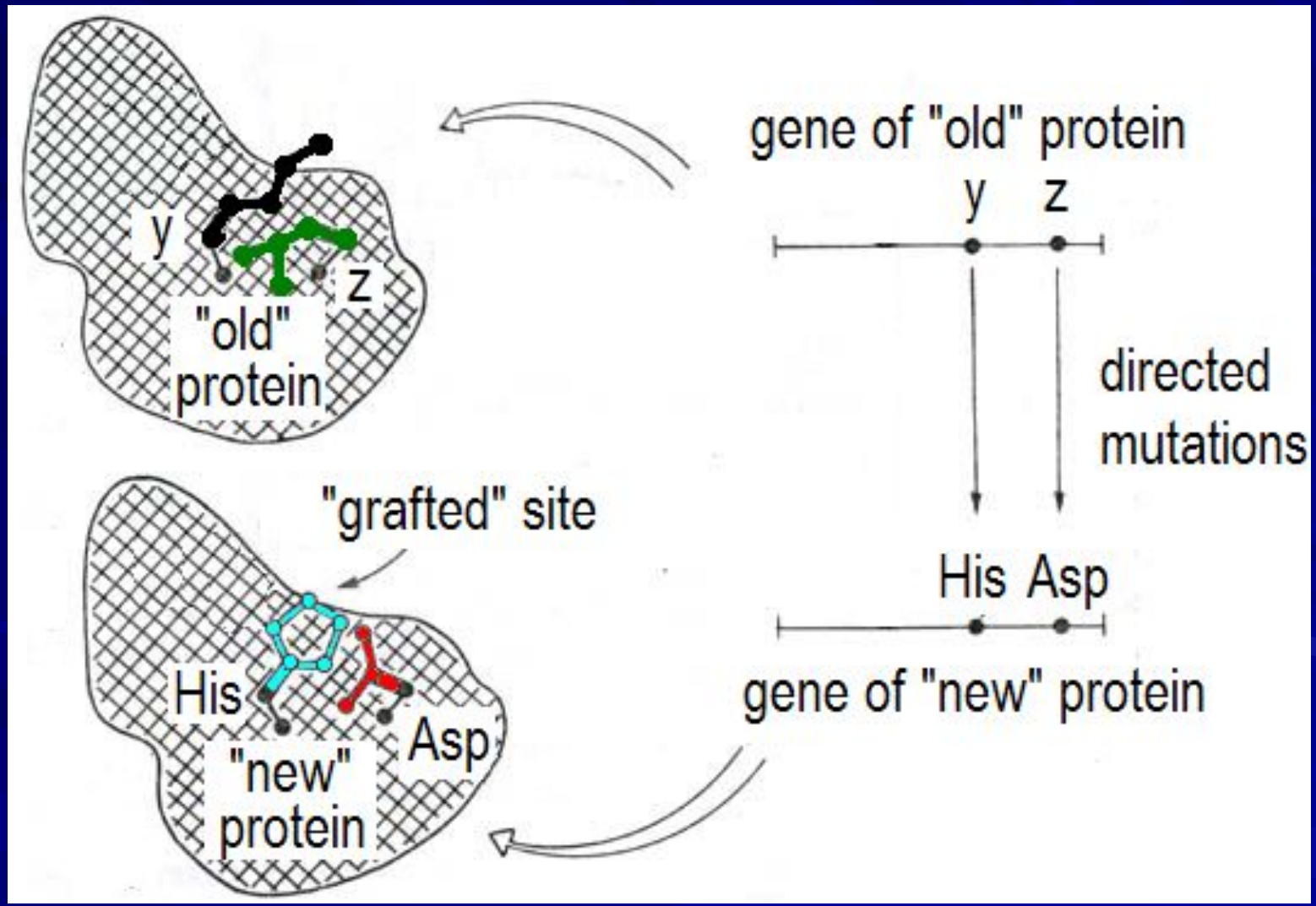
Improvement in the potential-energy function is needed!



S. Piana, J.L. Klepeis, D.E Shaw
Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations
Current Opinion in Structural Biology 2014, 24:98–105

Protein engineering

Wanted: new protein with additional salt bridge (e.g., His^+:::Asp^-)



Daniela Röthlisberger[1]*, Olga Khersonsky[4]*, Andrew M. Wollacott[1]*, Lin Jiang[1,2], Jason DeChancie[6], Jamie Betker[3], Jasmine L. Gallaher[3], Eric A. Althoff[1], Alexandre Zanghellini[1,2], Orly Dym[5], Shira Albeck[5], Kendall N. Houk[6], Dan S. Tawfik[4] & David Baker[1,2,3]

Kemp elimination catalysts by computational enzyme design

David Baker

doi:10.1038/nature06879

2008

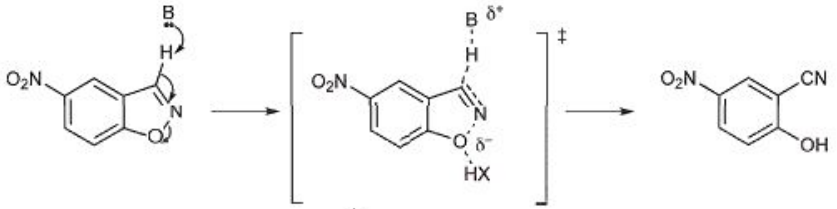
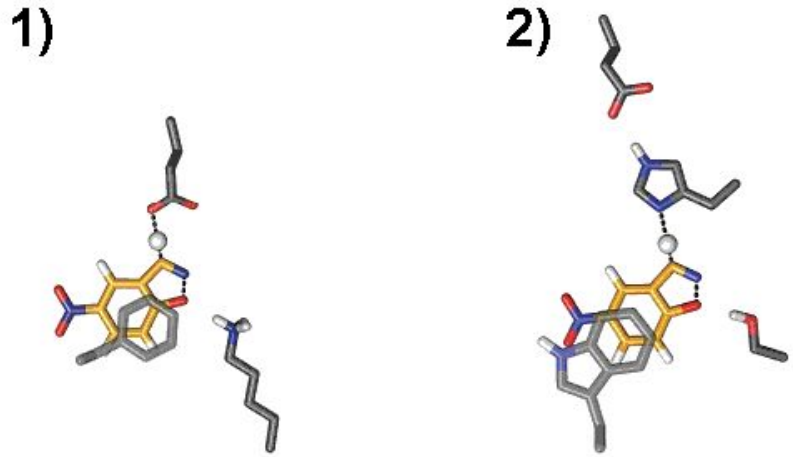


Figure 1 | Reaction scheme and catalytic motifs used in design.



$k_{cat}/k_{uncat} \sim 10^6$

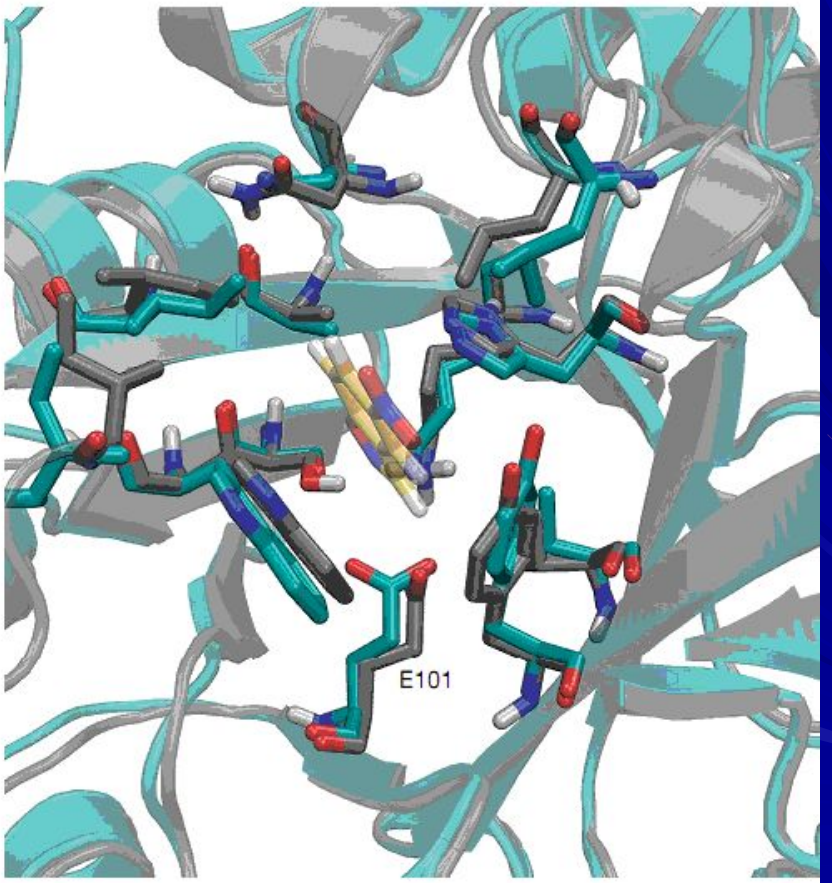
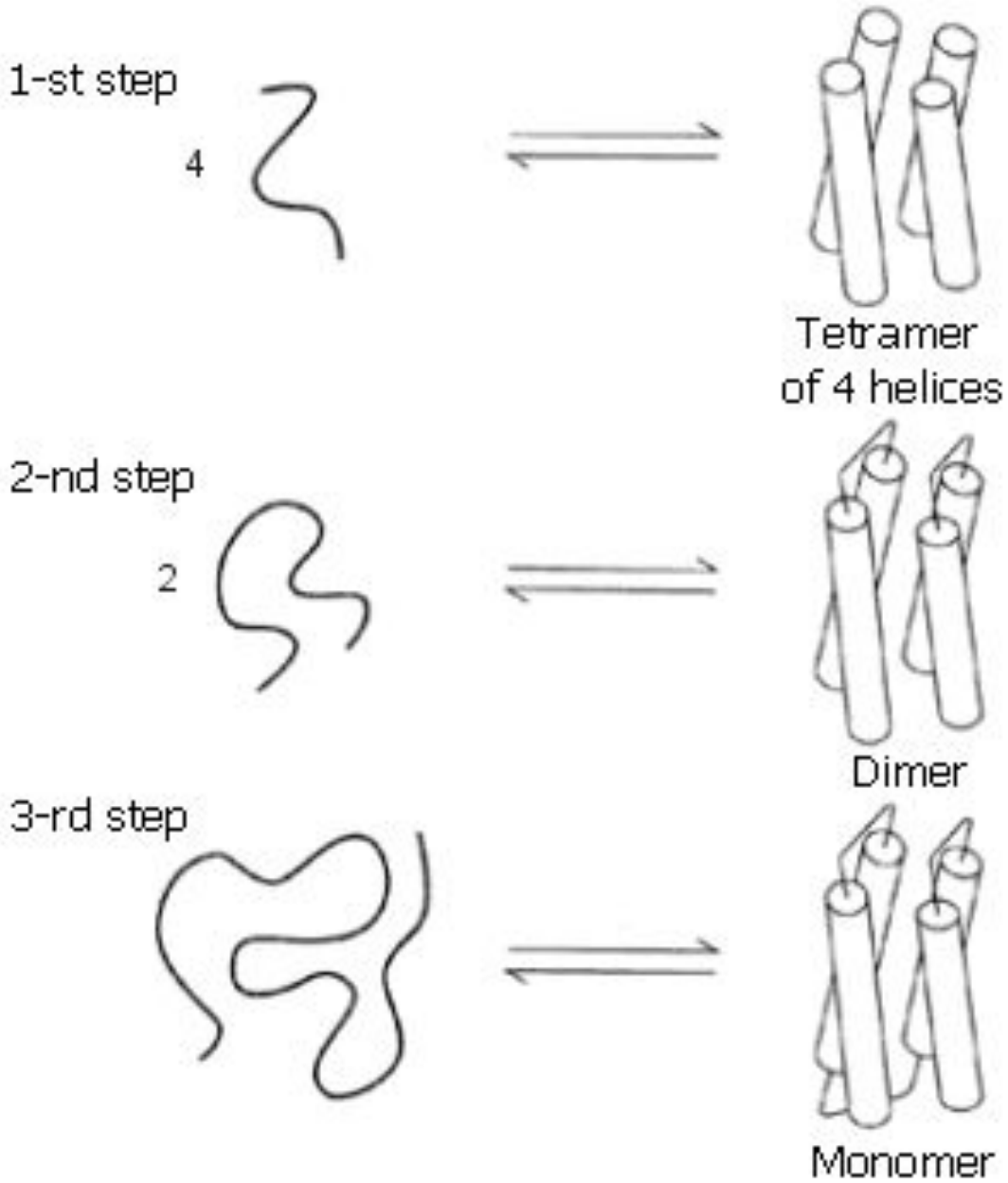


Figure 4 | Comparison of the designed model of KE07 and the crystal structure.

DESIGN



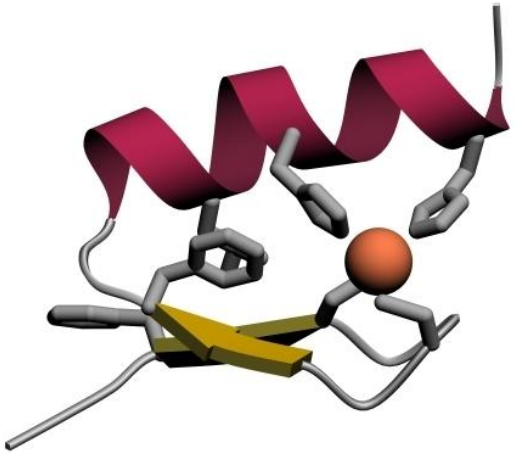
DeGrado, 1989

DOES NOT MELT !
MOLTEN GLOBULE…

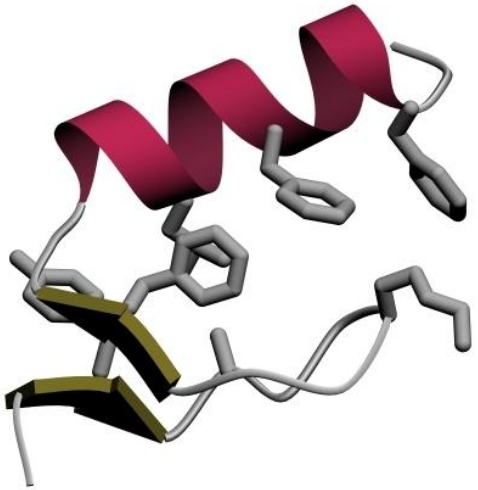
+ ION BINDING SOLID

DESIGN

a Natural protein (with Zn ion)



b Designed without ion: Mayo, **1997**



Stephen L. **Mayo**

Later, in 2003,
David **Baker** (1962) *et al.*
designed and made a new,
„unnatural“ fold

DESIGN

Ptitsyn
Dolgikh
Finkelstein
Fedorov
Kirpichnikov
1987-97

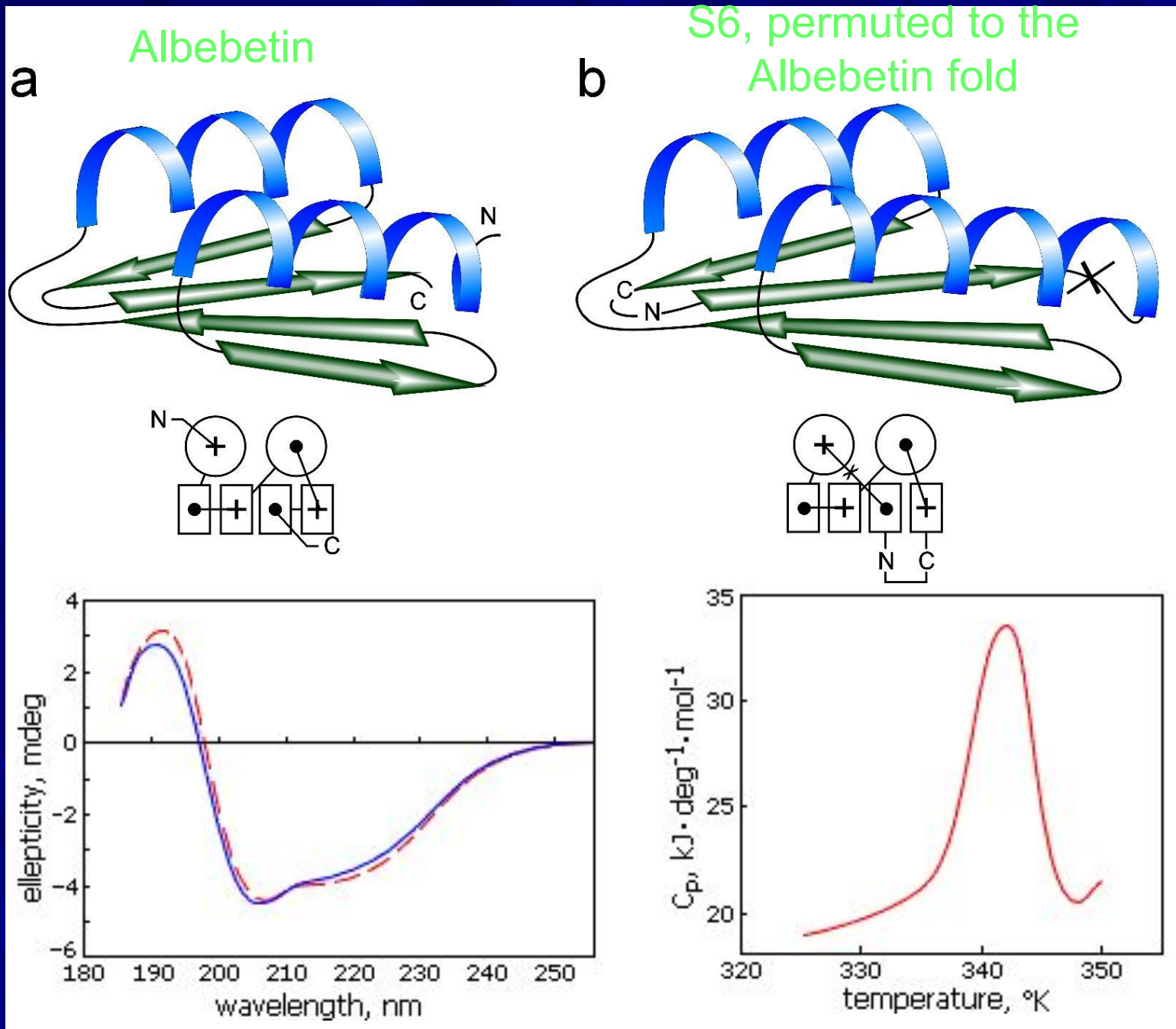
Albebetin;

⇓

Albeferon,

…

(grafting functional groups)

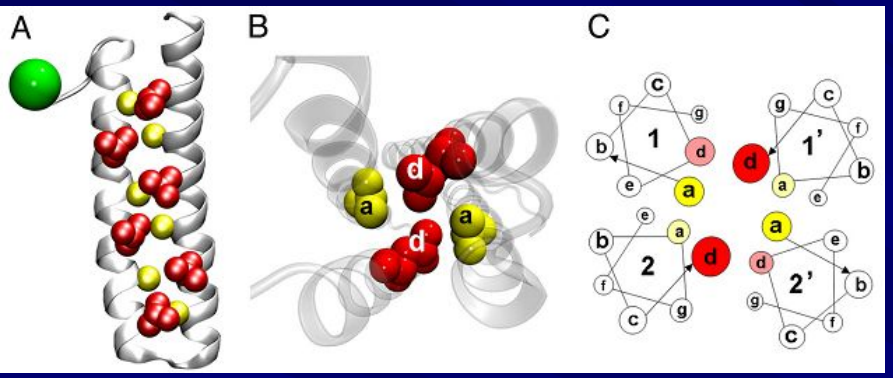


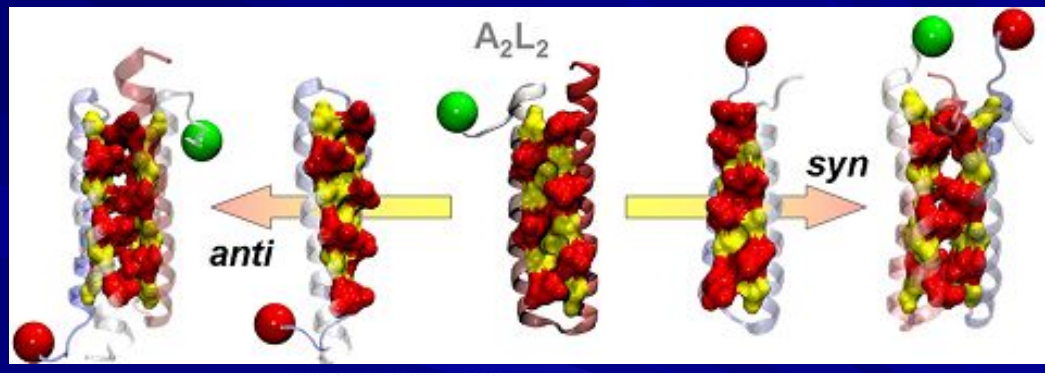
DESIGN OF A “HAMELION” PROTEIN:

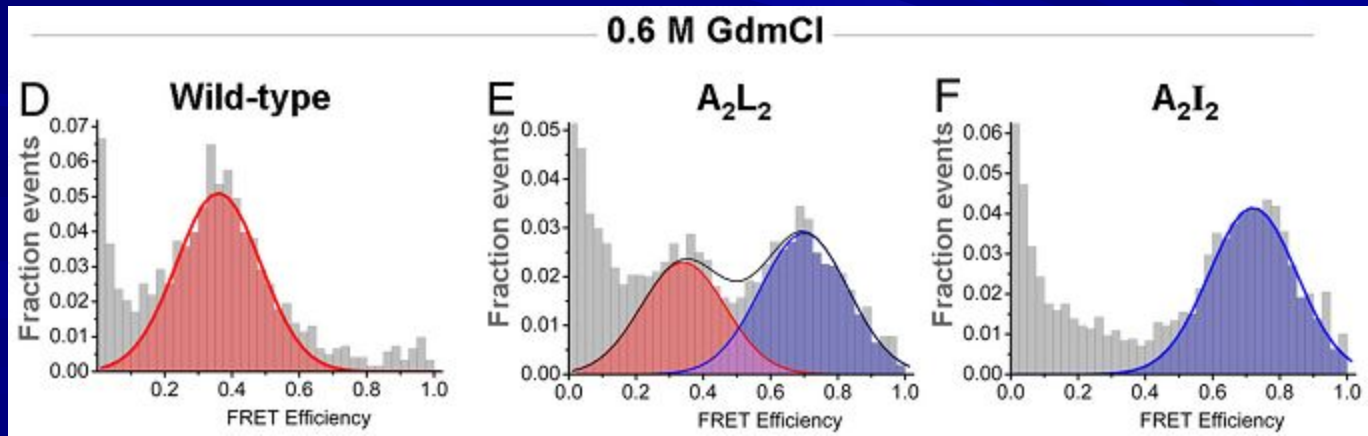
Direct single-molecule observation of a protein living in two opposed native structures

Y.Gambin, A.Schug, E.A.Lemke, J.J.Lavinder, A.C.M.Ferreon, T.J.Magliery, J.N.Onuchic, A.A.Deniz

PNAS, 2009 v.106, 10153–8

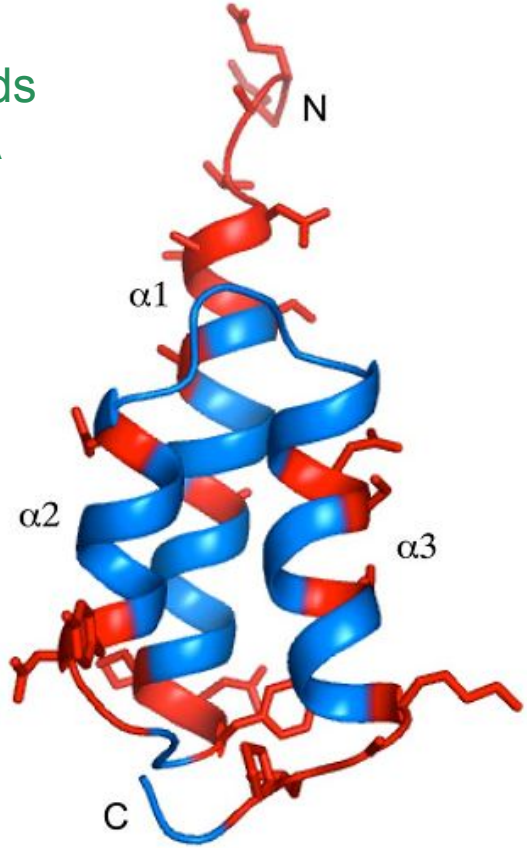






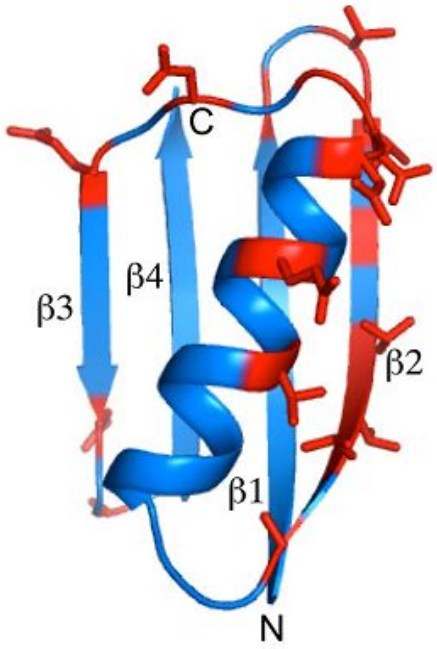
A

G_A binds to HSA



B

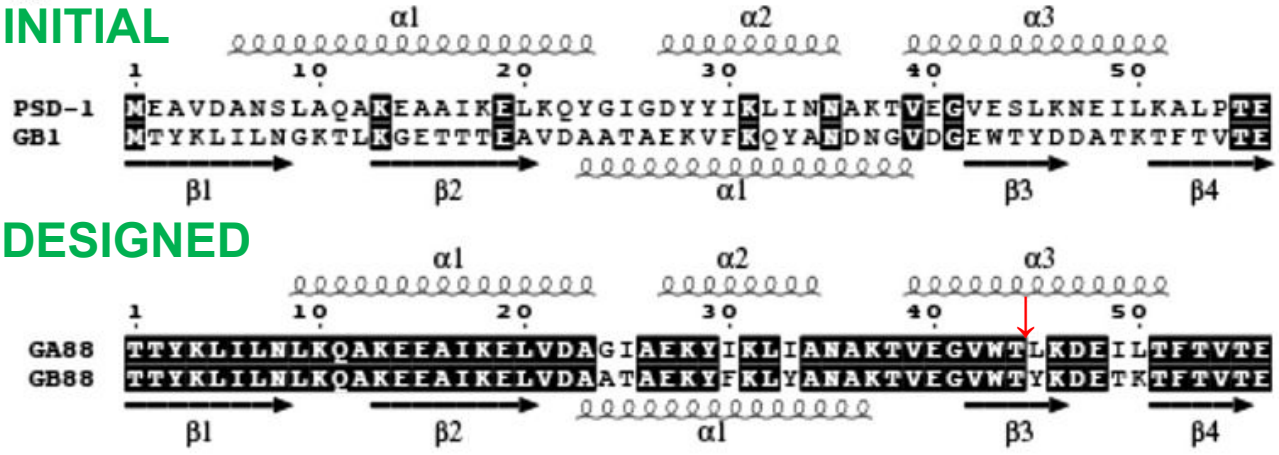
G_B binds to IgG Fc region



C

INITIAL

DESIGNED



Protein design

Wanted: new protein fold

P.A.Alexander, Y.He, Y.Chen, J.Orban, P.N.Bryan

PNAS, 2007, **104**, 11963-8

The design and characterization of two proteins with 88% sequence identity but different structure and function

Y.He, Y.Chen, P.Alexander, P.N.Bryan, J.Orban

PNAS, 2008, **105**, 14412-7

NMR structures of two designed proteins with high sequence identity but **different fold** and **function**

2012 (*Structure*, **20**, 283-91): one-residue difference

PROTEIN STRUCTURE:

PREDICTION
ENGINEERING
DESIGN