What do we mean by “contingency”?

Occurrence of a “robust” fact as a result of fundamental physical processes.

“Robust” refers to something which depends only on few key ingredients ➔ largely independent of details.

Physics of Proteins: Origin of Protein Structures and Molecular Evolution

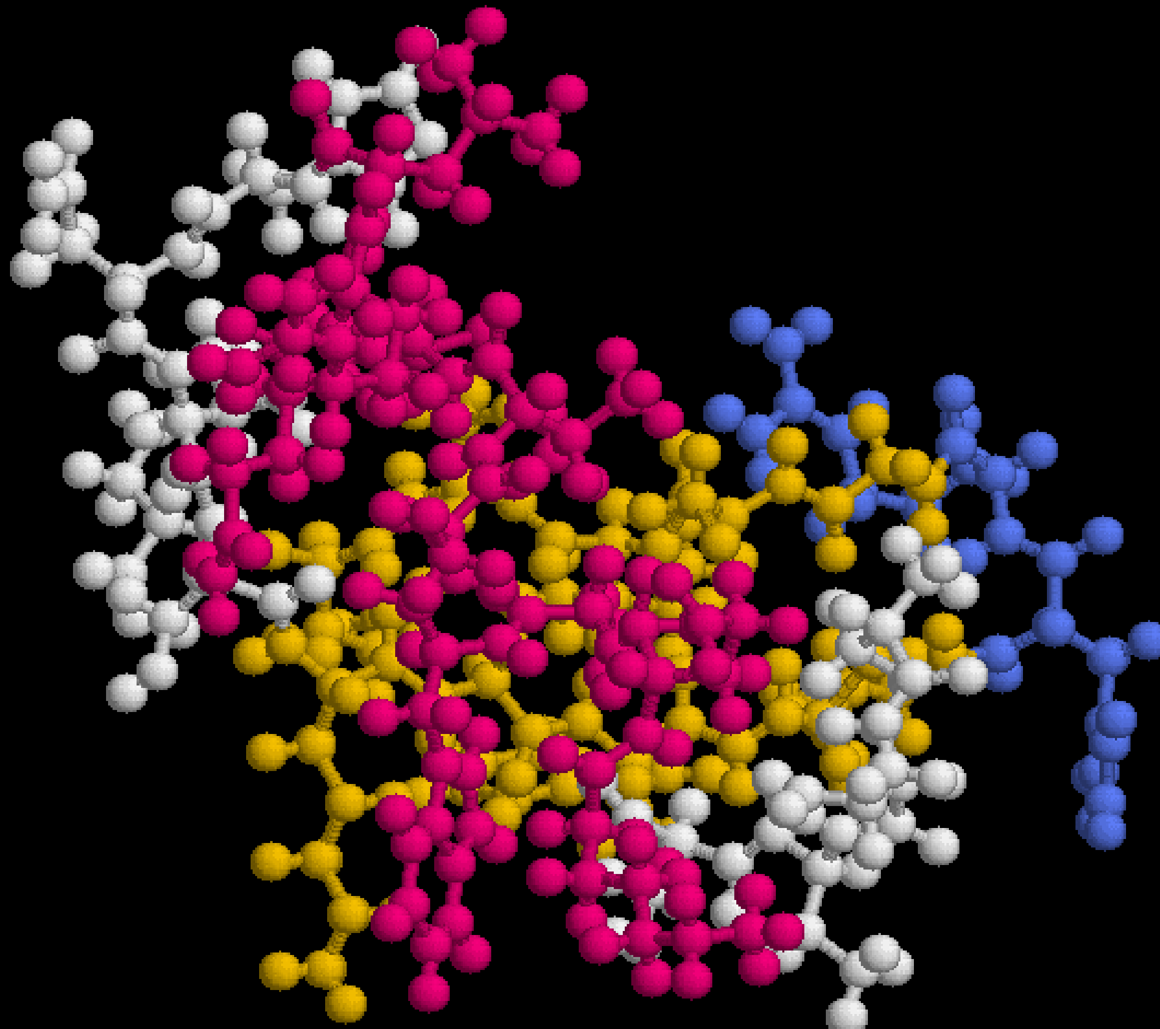
UNIVERSITA' DEGLI STUDI DI PADOVA
DIPARTIMENTO DI FISICA "GALILEO GALILEI"

Jayanth R. Banavar PennState - USA

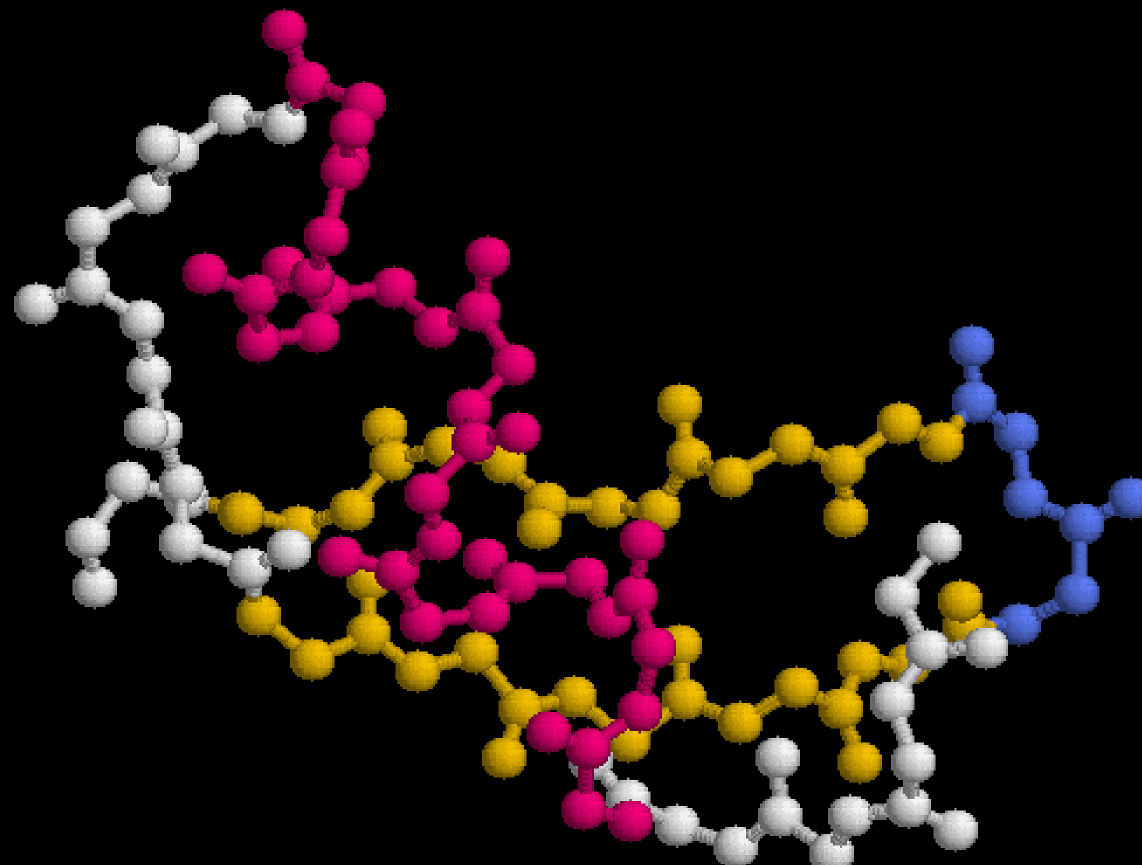
Sandro Azaele Padova - I
Davide Marenduzzo Edinburgh - UK
Flavio Seno Padova - I
Antonio Trovato Padova - I
Igor Volkov PennState - USA

Outline of the I part

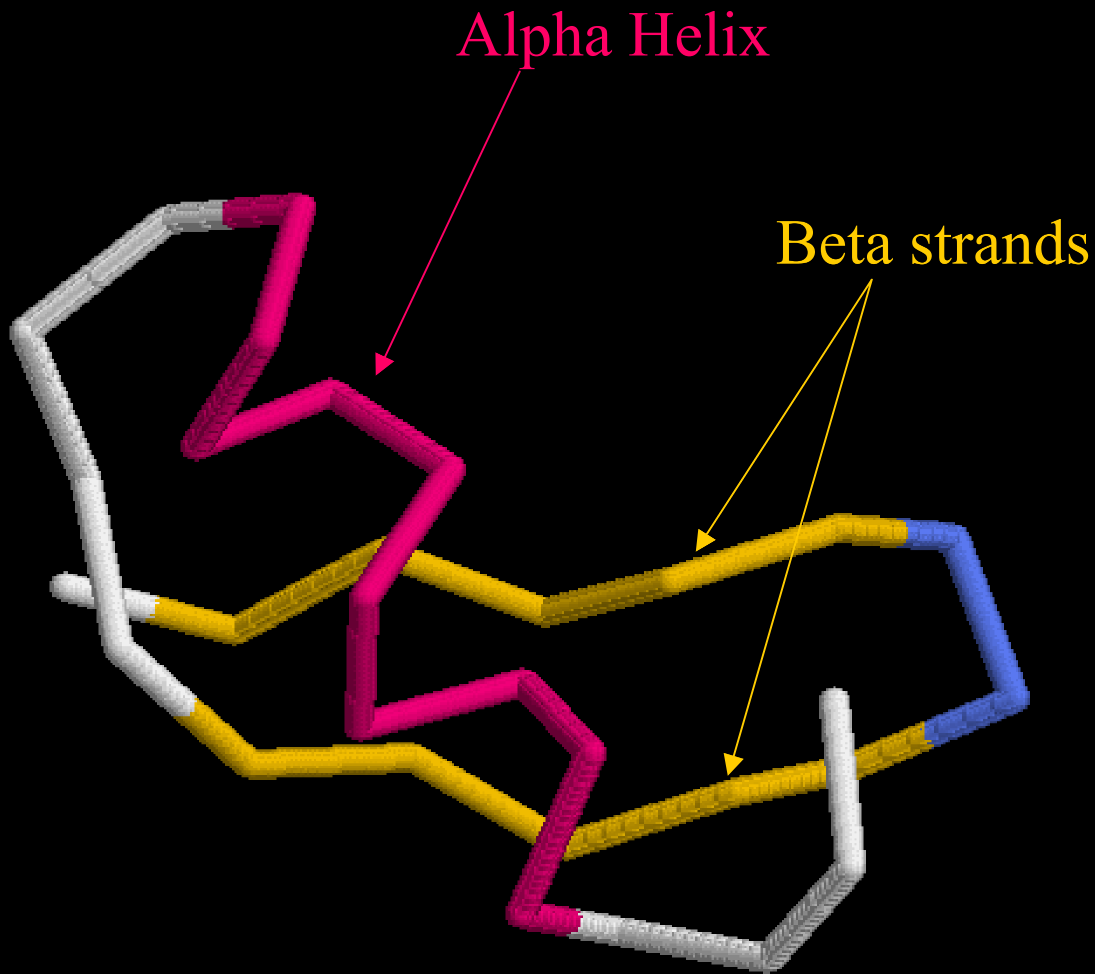
- Known facts about homo-polymers
- Known facts about proteins
- Where do secondary motifs come from?
- Compact phase of spheres and polymers
- The missing ingredient – new phase of matter (contingency) used by bio-polymers !



All Atom but H



All Backbone Atoms but H

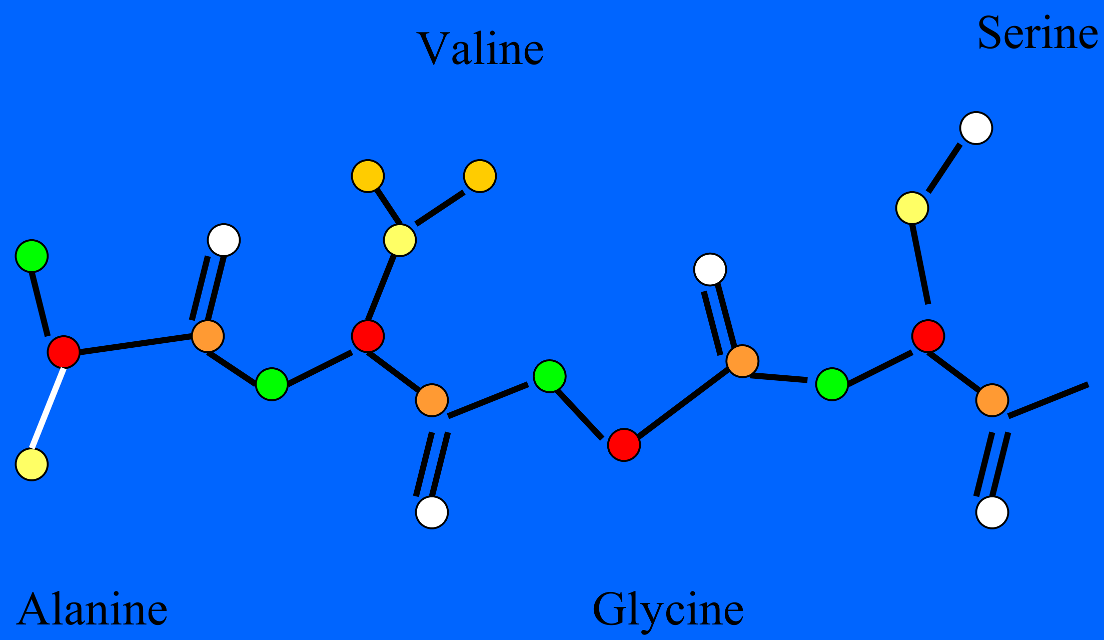


Joining C_alpha Atoms

Proteins are heteropolymers and are compact due to their hydrophobicity

Carbon alpha, Carbon’, Nitrogen, Oxigen …. Hydrogen omitted

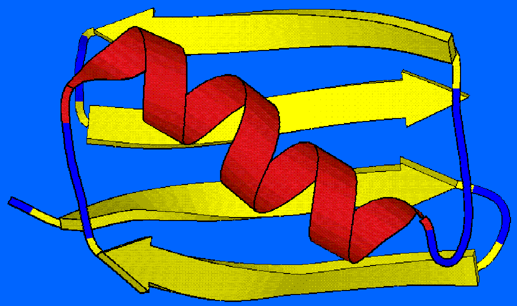
20 kinds of amino-acids



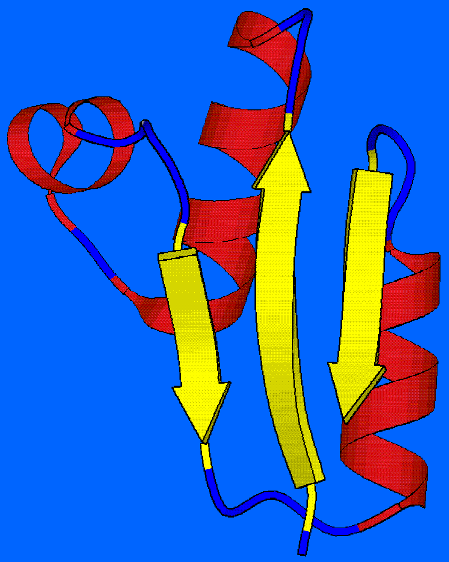
Side-Chains

Compactness-Hydrophobicity

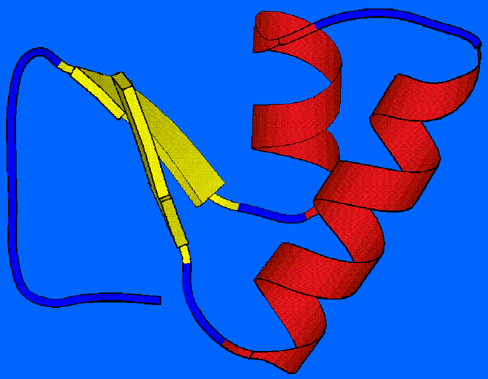
Some Example of Protein Structures - Folds



1GB1

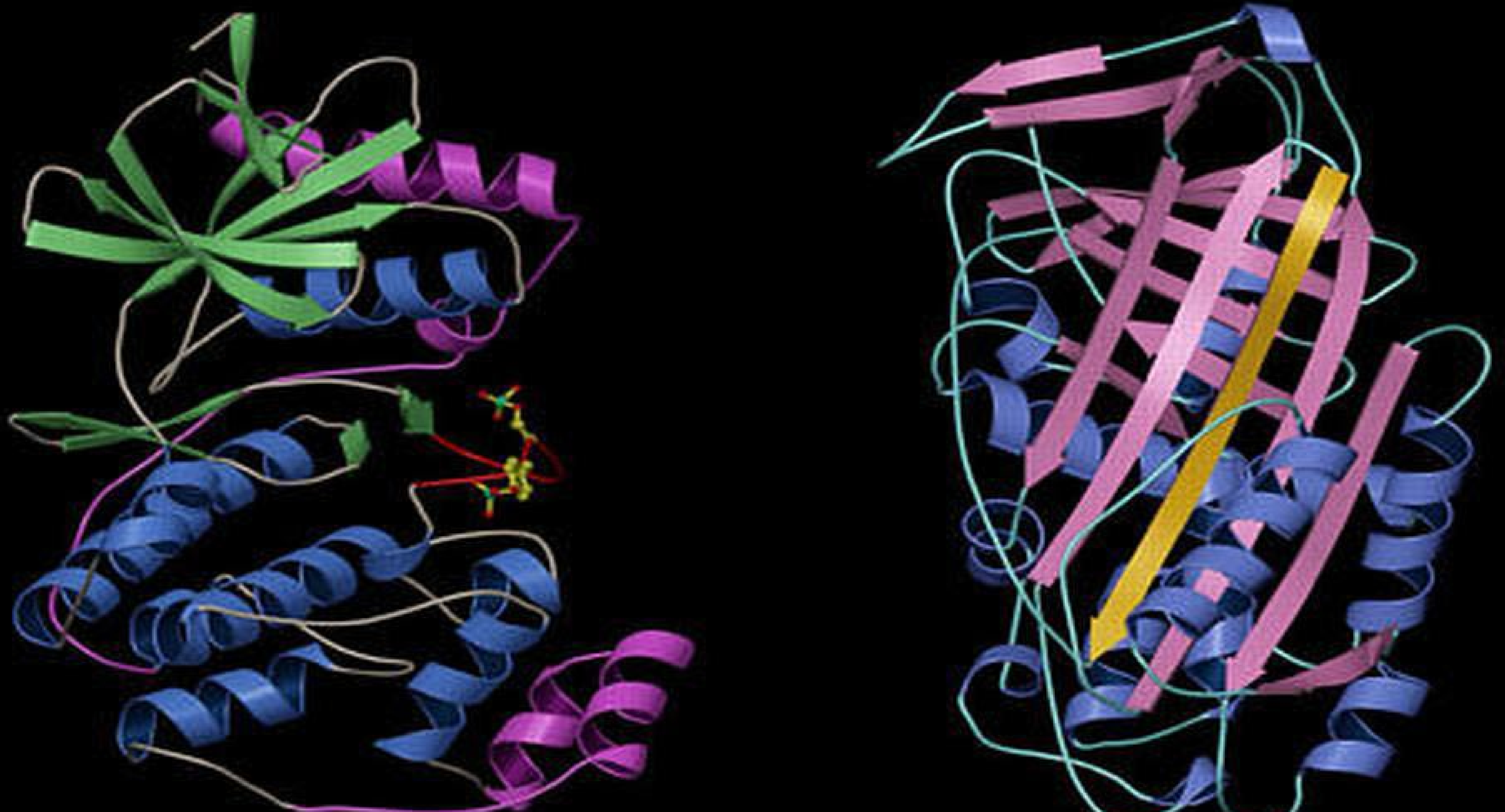


1CTF

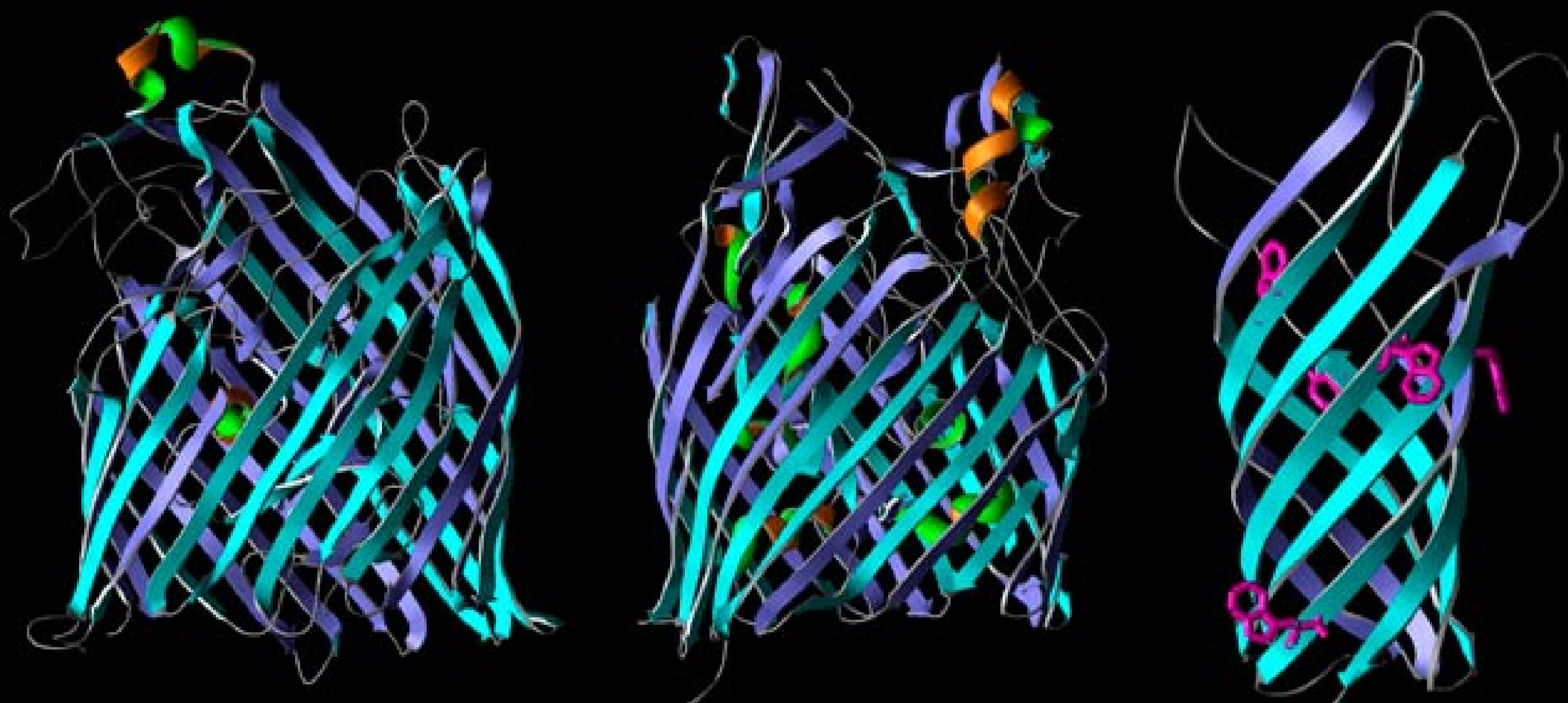


1CRN

High content of secondary motifs: helices and beta sheets



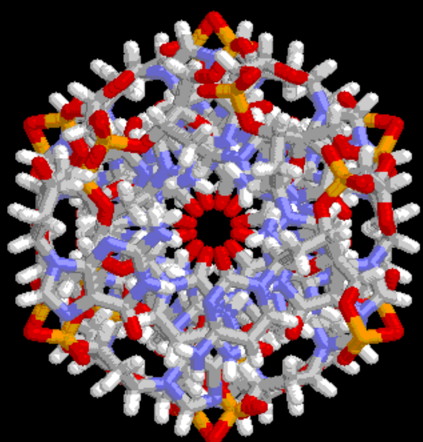
Crystal Structures of Some Outer Membrane Proteins



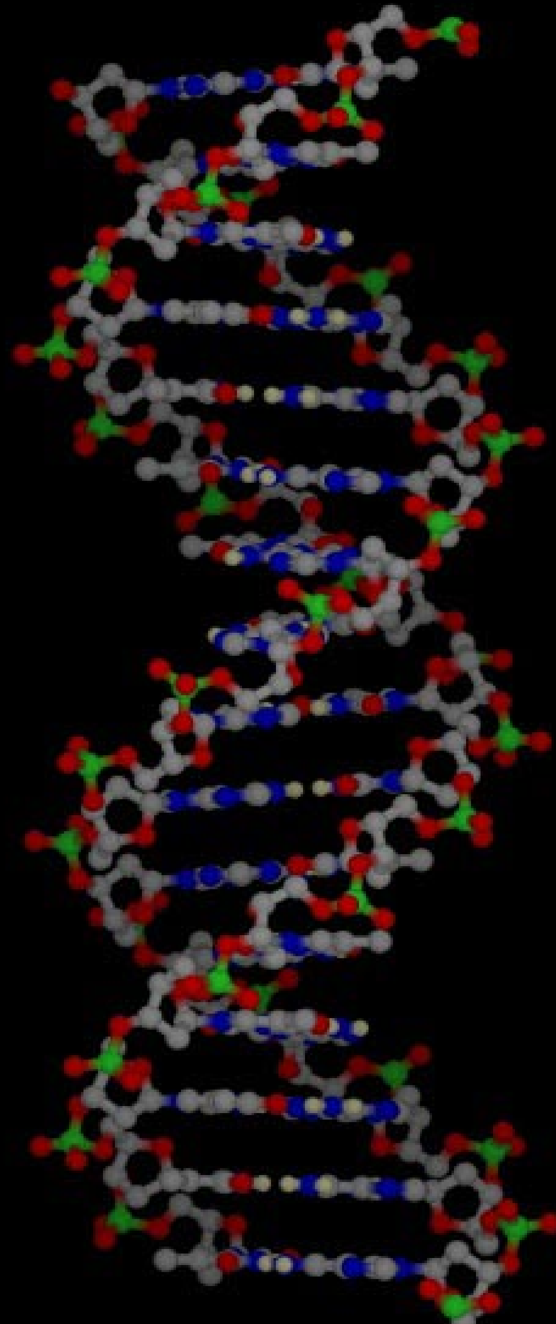
FhuA
(Ferguson et al., 1998)

FepA
(Buchanan et al., 1999)

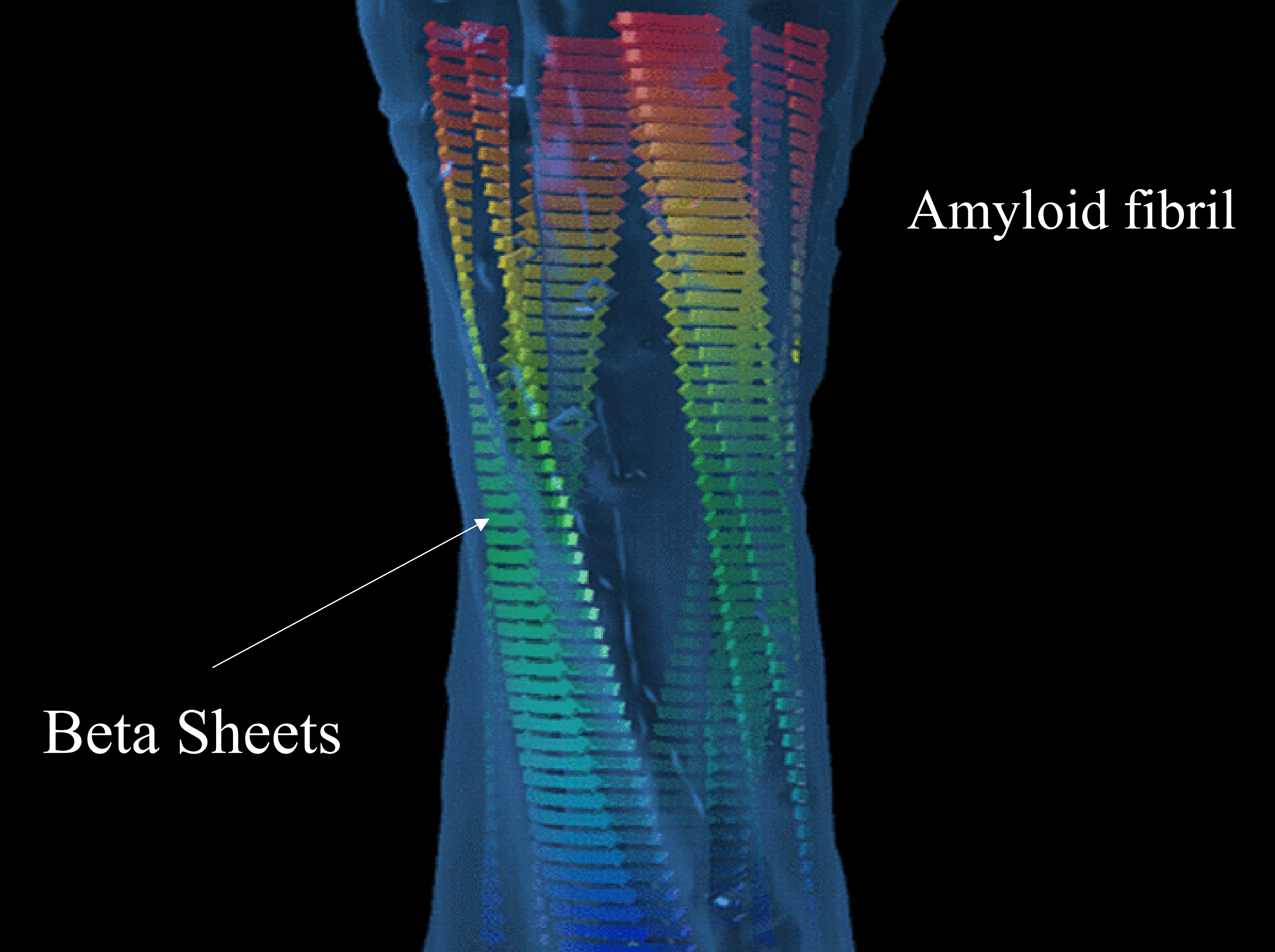
OmpA
(Pautsch & Schulz, 1998)



Top view



DNA


Amyloid fibril
Beta Sheets

- **Why have sequences evolved but not protein folds? → Neutral evolution;**
- **Origin of a finite (discrete) menu of protein native states and amyloid aggregation ← Stability, Sensitivity and Diversity.**

Another example where Nature has used discreteness

Origin of "discrete" species?
Stability, Diversity and Interactions in ecosystem:

EXPERIMENTAL RESULTS

Common Characters of Proteins

1. Proteins fold rapidly
2. The geometry of native states affects functionality
3. Distinct folds are only few thousand
4. Many sequences → same native-state fold → not sensitive to mutations
5. Multiple protein functionalities within the same fold
6. The folding rate is not vastly different and transition states are also similar in proteins sharing the same native-state topology
7. Protein folds are modular forms made up of simple building blocks: helices and almost-planar sheets … domains.
8. Protein structures are flexible allowing proteins to carry out a wide variety of tasks.
9. Proteins interact with each other and with ligands in a versatile yet robust manner, and they act as molecular targets of evolution.
10. Sequences have evolved but not protein folds (→Neutral evolution).

➔ Sequence alone does not shape the structure

➔ menu of protein folds

“Any effective picture of protein structure must provide at the same time for the common character of all proteins as exemplified by their many chemical and physical similarities, and for the highly specific nature of each protein type.”

J. D. Bernal (1939)

“Synthetic analogs of globular proteins are unknown. The capability of adopting a dense globular configuration stabilized by self interactions and of transforming reversibly to the random coil are peculiar to the chain molecules of globular proteins alone.”

Paul Flory

Challenge

Identification of the key common attributes that determine this menu.

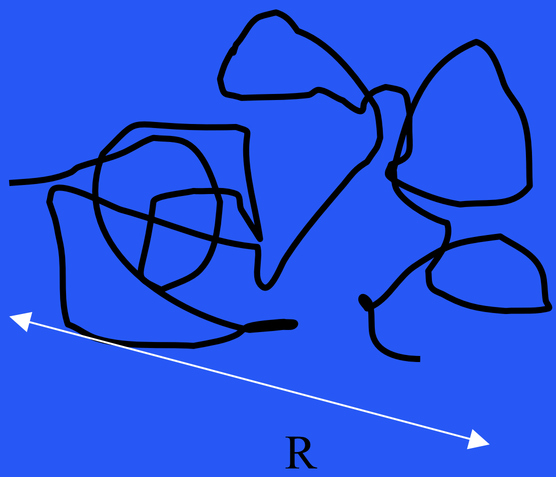
Question 1

Is the ability to perform so many functions in a so synergistic manner exclusive to proteins or can also be realized in artificial devices, too?

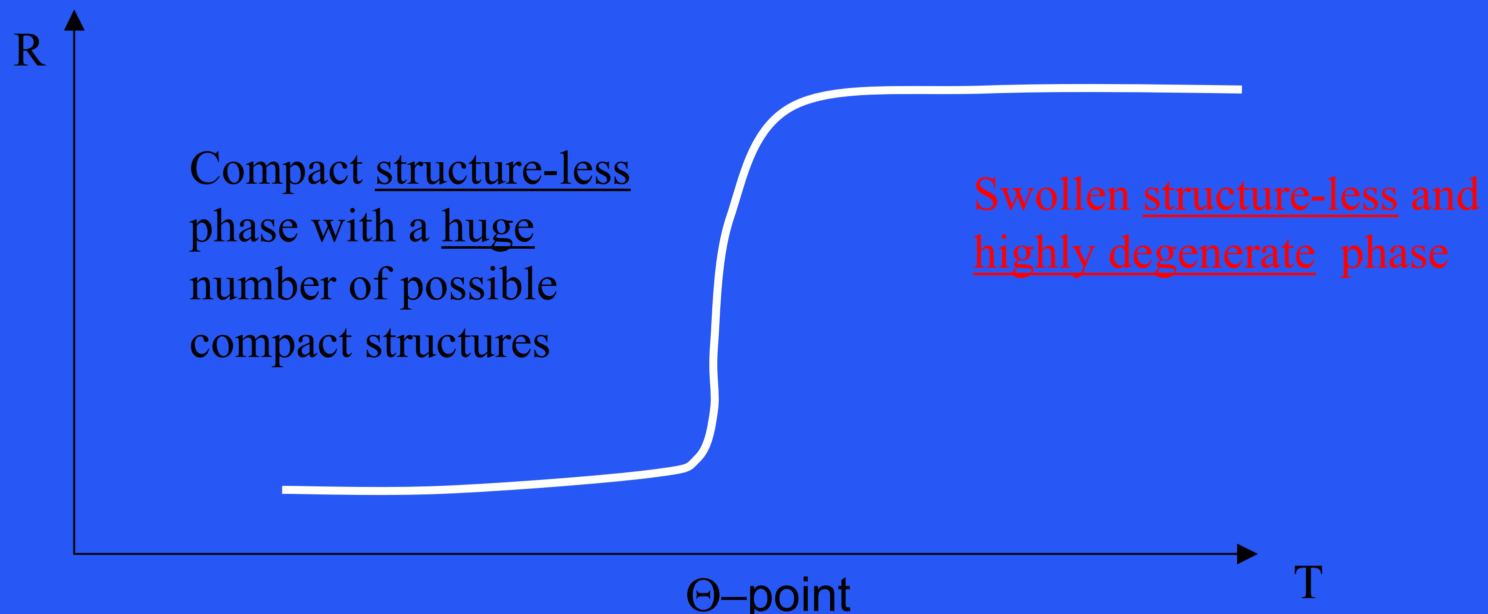
Question 2

Are standard models of polymer physics able to explain the origin of the menu of native states?

Homo-polymers: Known Facts



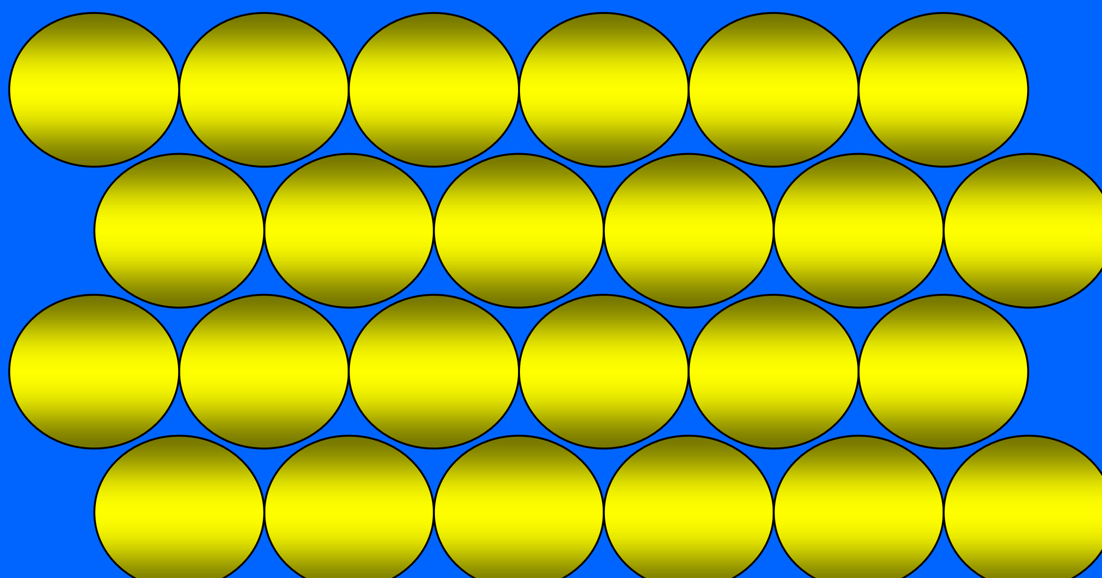
Self-interacting polymer in a good/bad solvent



Preamble: Origin of Crystals

Kepler's conjecture: optimal packing of cannon balls!

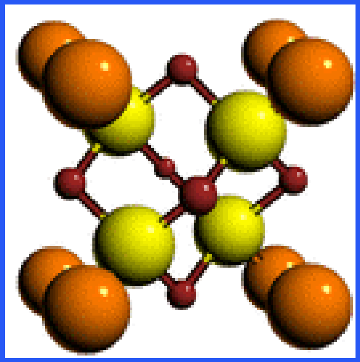
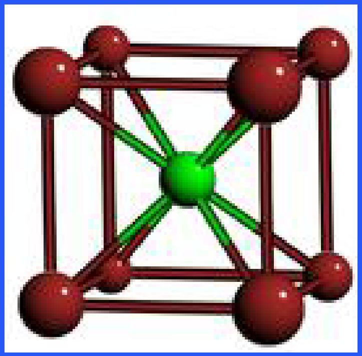
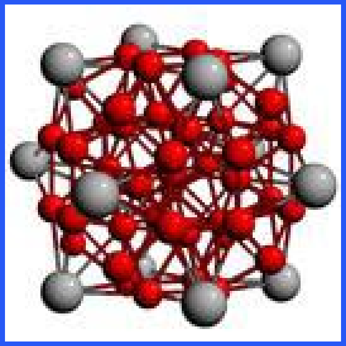
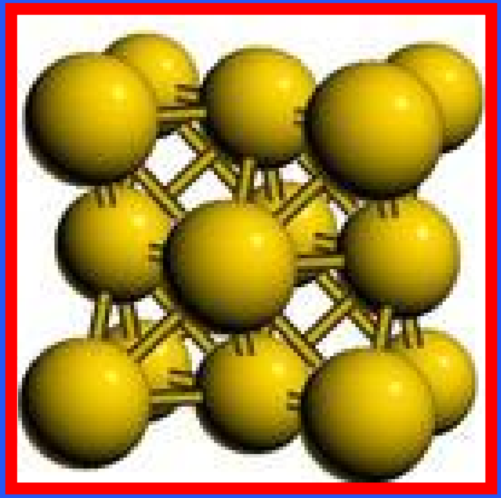
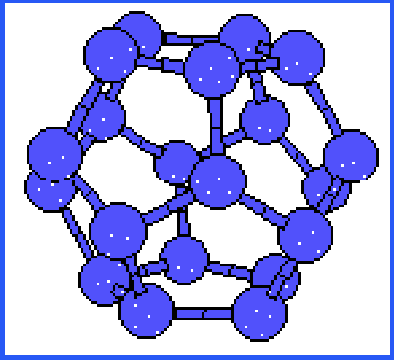
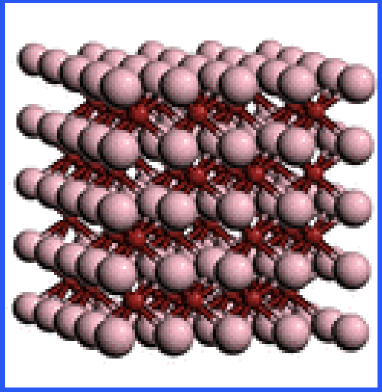
Solids of noble elements but helium



"Only" 28% on empty space in 3-D

Fiodorov (1895):

Periodicity & Symmetry => *Menu* of only **230** Groups !

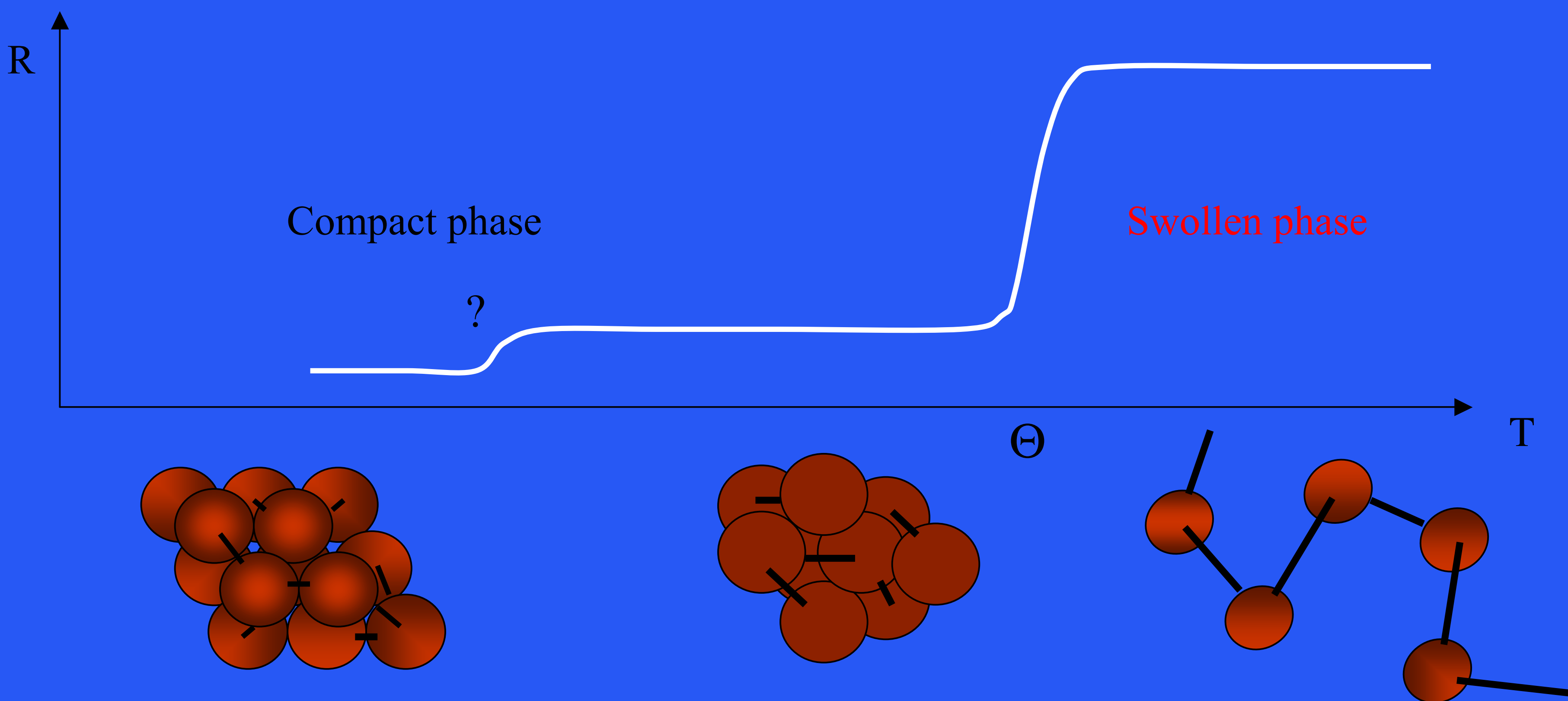







Packing of hard spheres

Could Compactness Alone Induce Secondary Structure?

Compact Phases of Standard Polymers

String and beads model



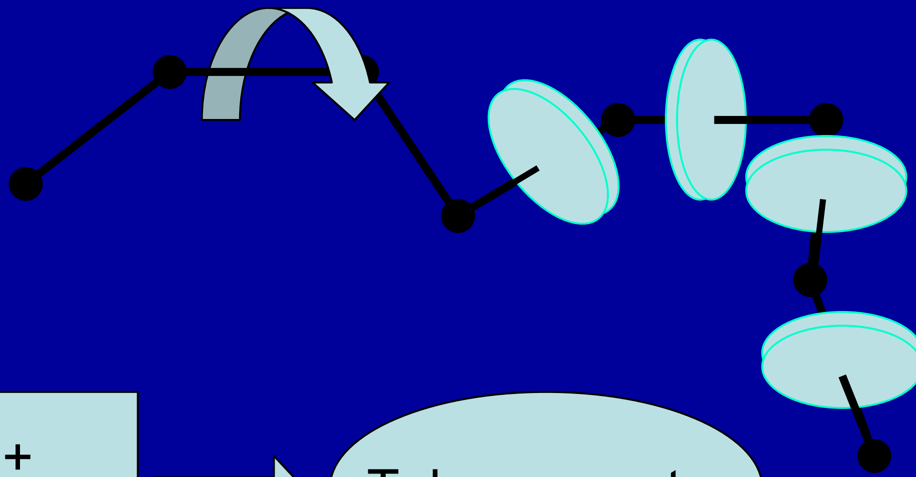
Crystalline phase: Hamiltonian walks

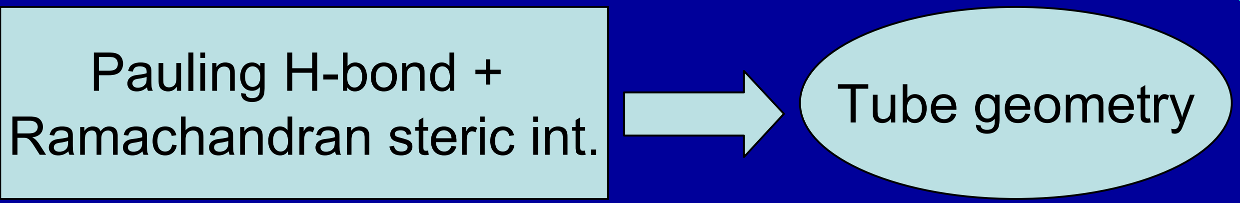
Compact disordered phase

Swollen phase

The missing key ingredient

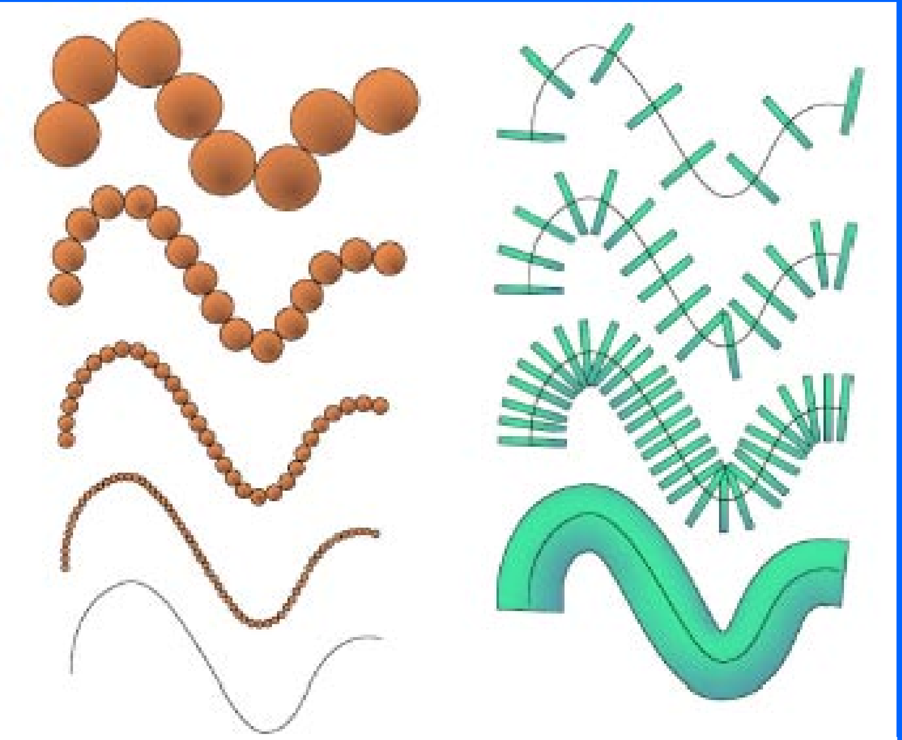
Local residual rotational invariance → no spheres but <u>coins</u>





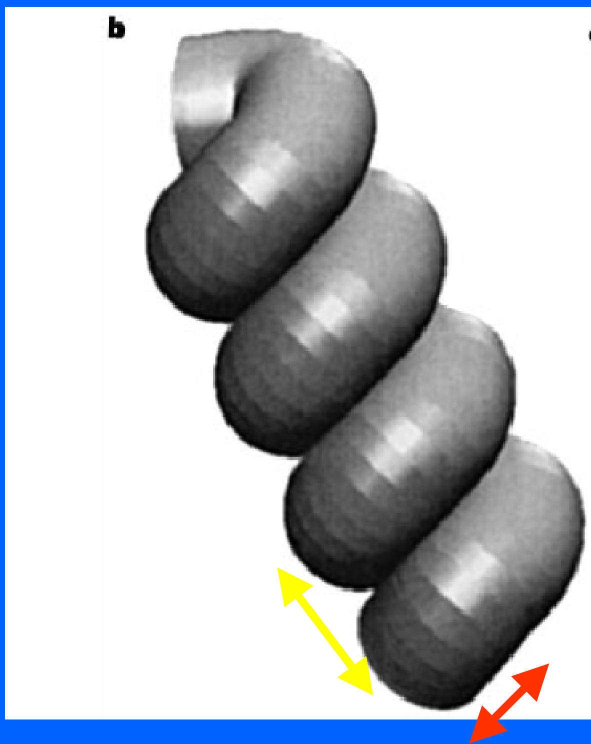
In the continuum limit …

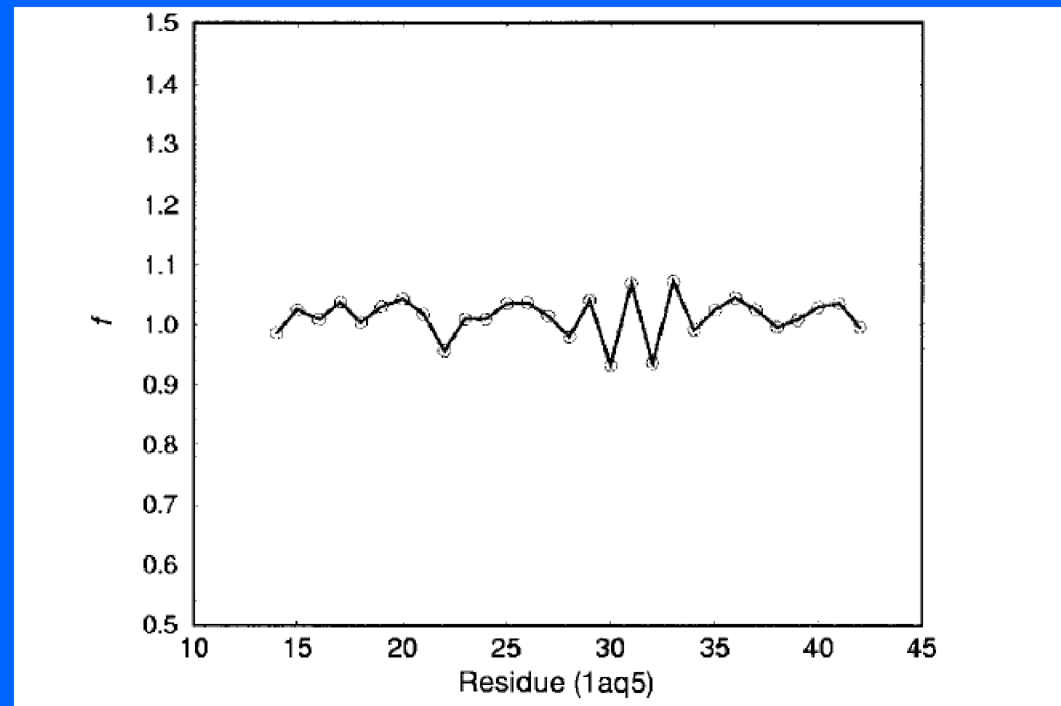
String of beads and string of coins in the continuum limit



Zero thickness string   Finite thickness tube

Squeezing a tube in the marginally compact phase
The Magic Helix

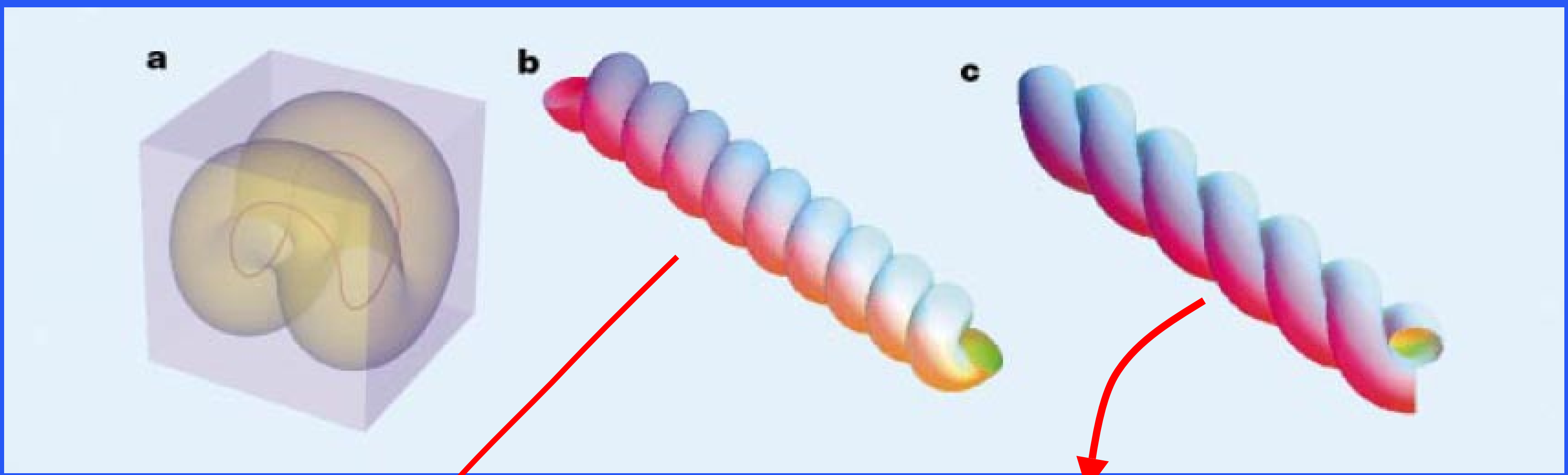




Radius to **pitch** ratio within 5% of the value in alpha helices

Maritan, Trovato, Micheletti & Banavar **Nature** 2000;
Banavar & Maritan **Rev. Mod. Phys**. 2003; Marenduzzo et al. **CoPlexUs** 2003 and **Polymers** 2004; Hoang et al. **PNAS** 2004: Marsella et al **PNAS** 2006; …..

Squeezing a tube in the marginally compact phase; The Magic Helix: from alpha to double helix

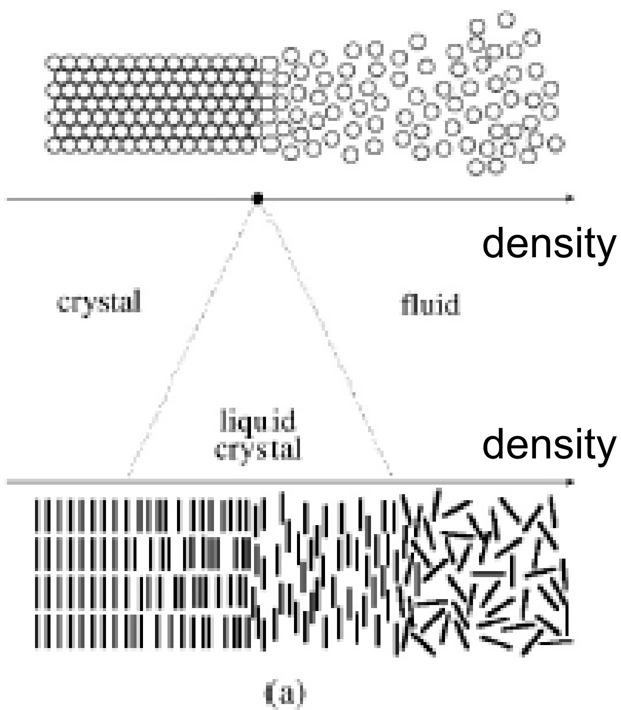


Densest double helix → pitch/radius = 2π
For DNA pitch/radius = 6.03   4% difference!!
News and View by Stasiak and Maddocks

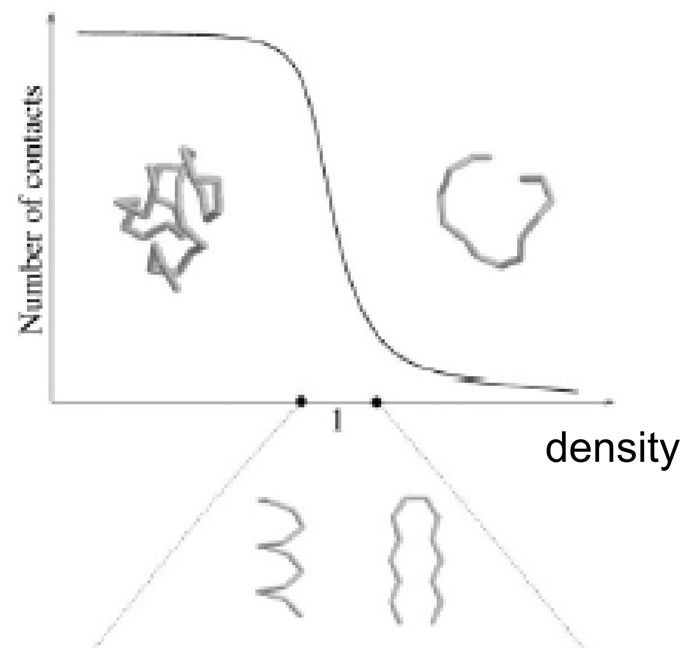
Densest single helix → pitch/radius = 2.512…
within 5% from alpha helices (with Trovato, Micheletti & Banavar **Nature** 2000)

Similar to what happens for liquid crystals

SPHERES



RODS



Marginal compact phase

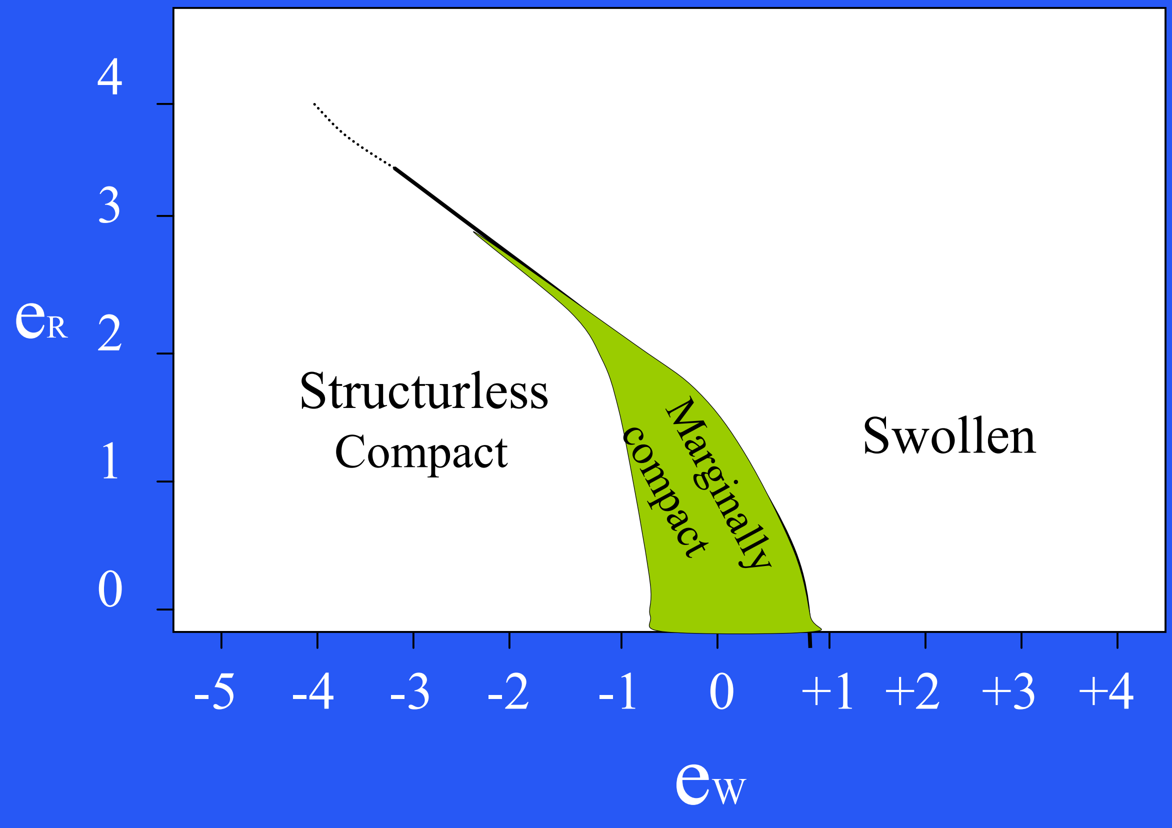
THICK POLYMER

Ground State Phase Diagram

e_R = curvature - Ramachandran

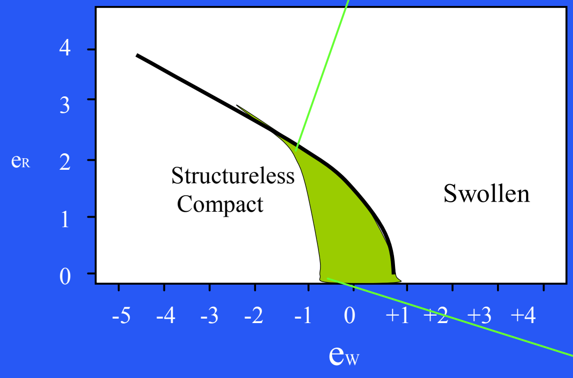
e_W = water mediated hydrophobic interaction

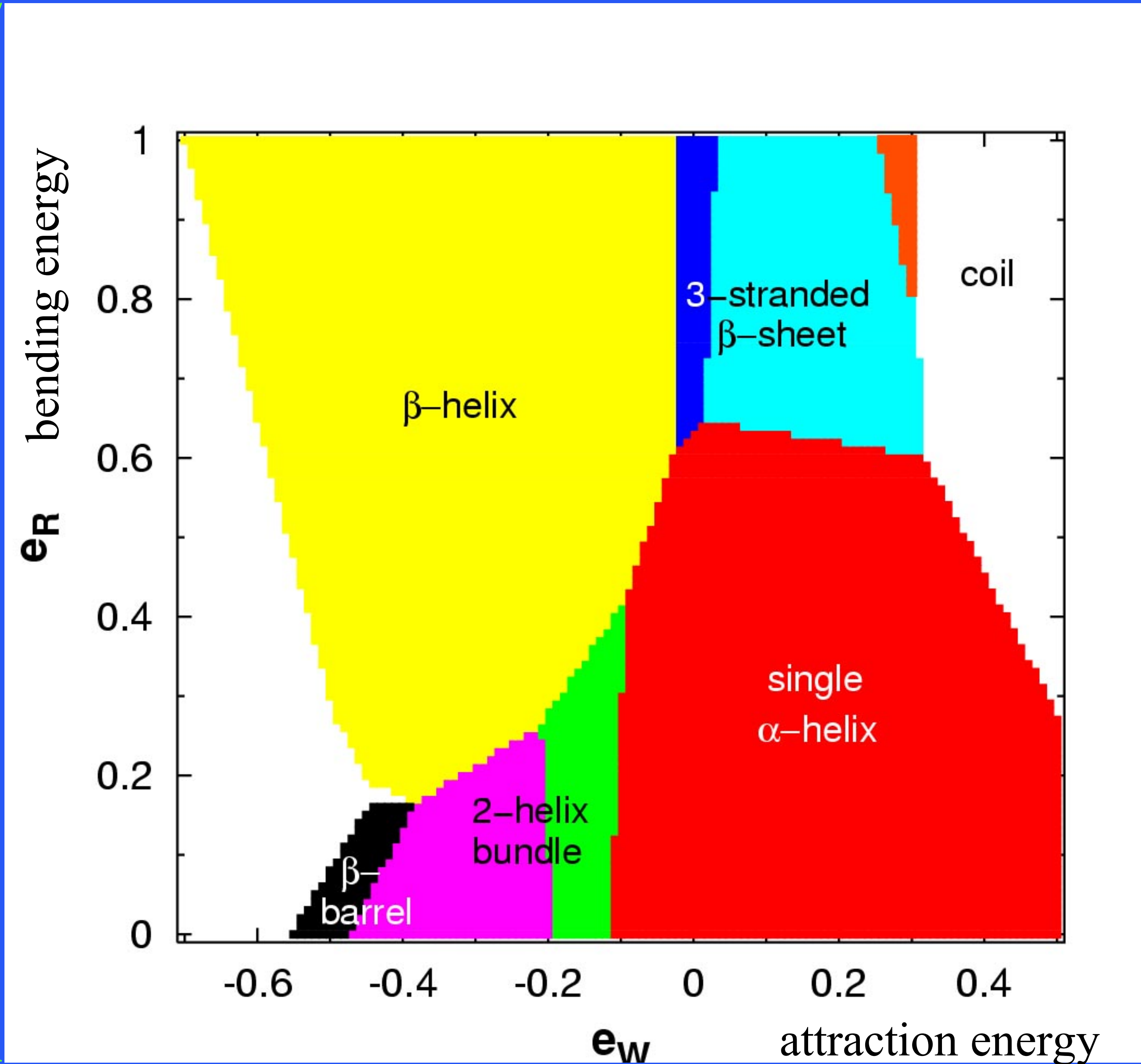
No sequence specificity: HOMOPOLYMER

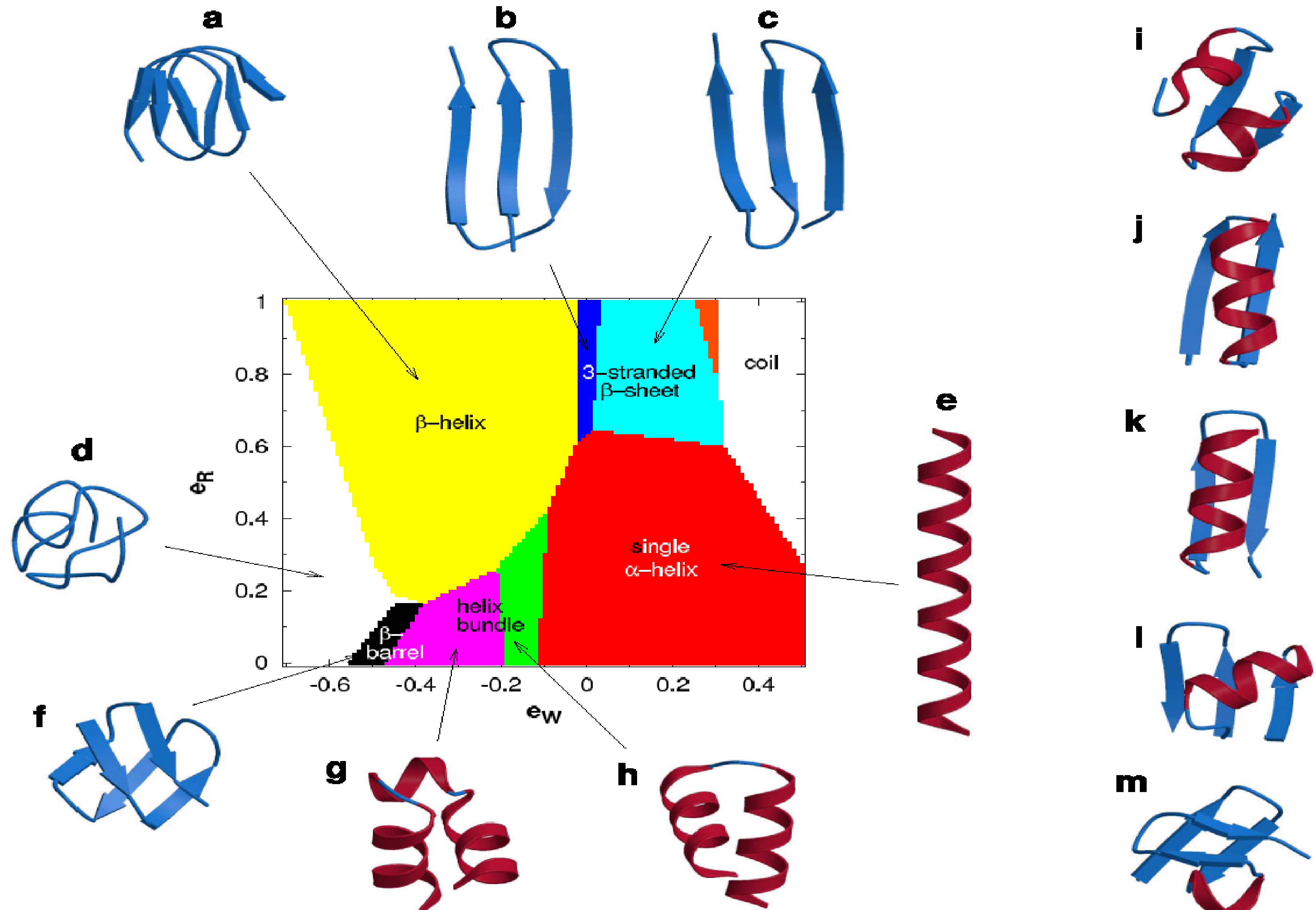


Ground State Phase Diagram

HOMOPOLYMER

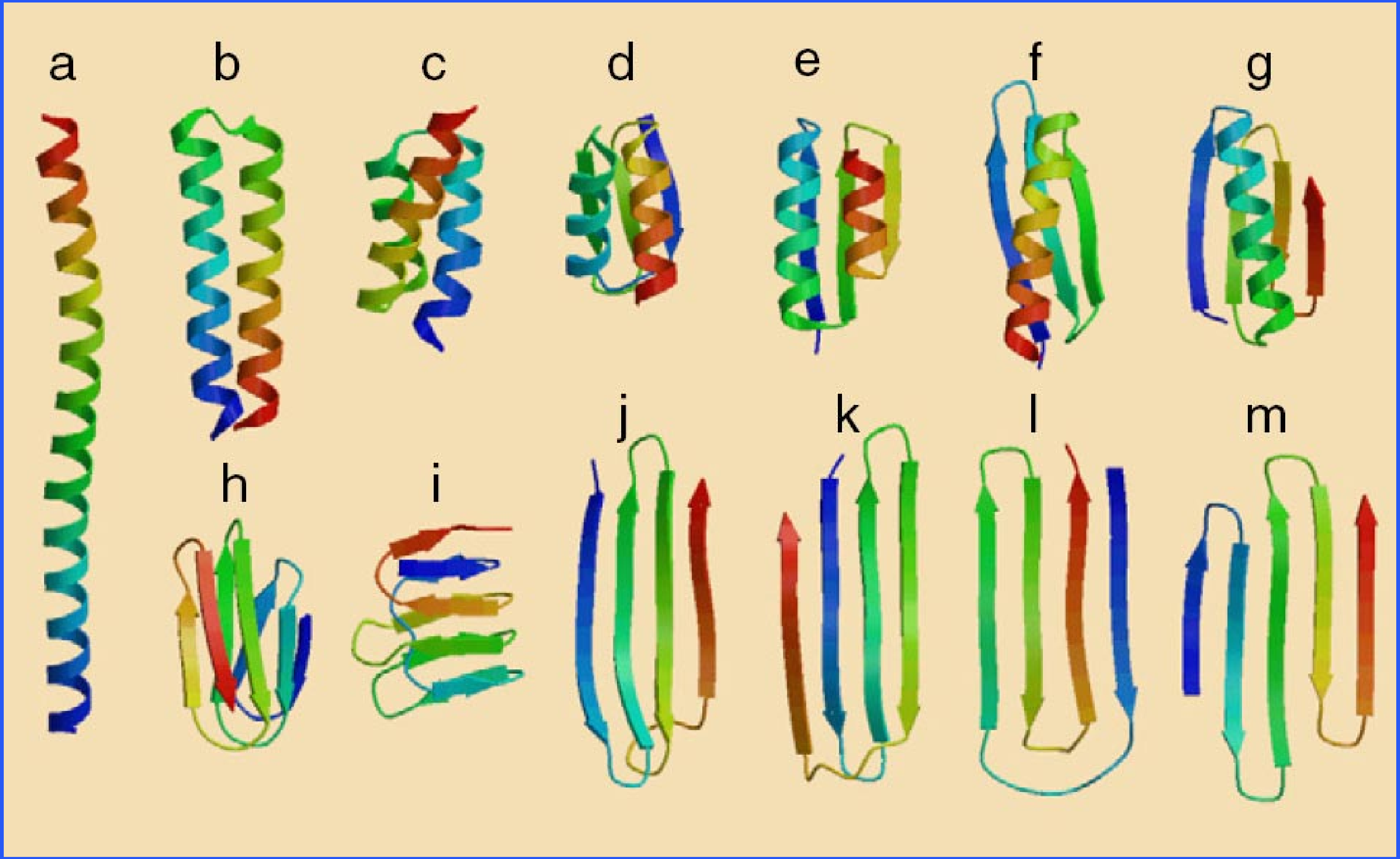




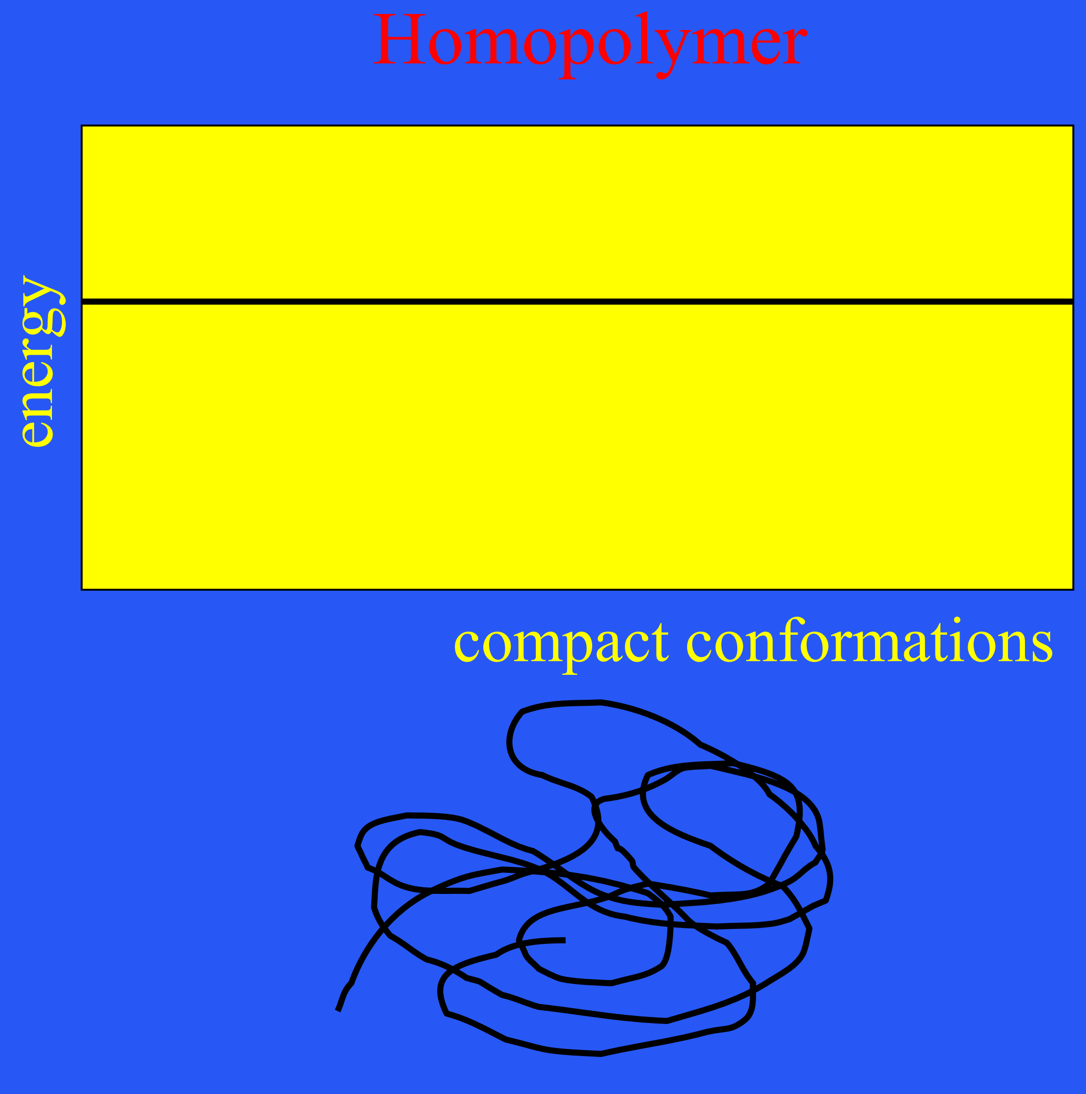


a
b
c
d
e
f
g
h
i
j
k
l
m
β–helix
3–stranded
β–sheet
coil
single
α–helix
helix
bundle
β–
barrel
e_R
e_W
1
0.8
0.6
0.4
0.2
0
-0.6
-0.4
-0.2
0
0.2
0.4

Menu of folds for sequences of 48 amino acids

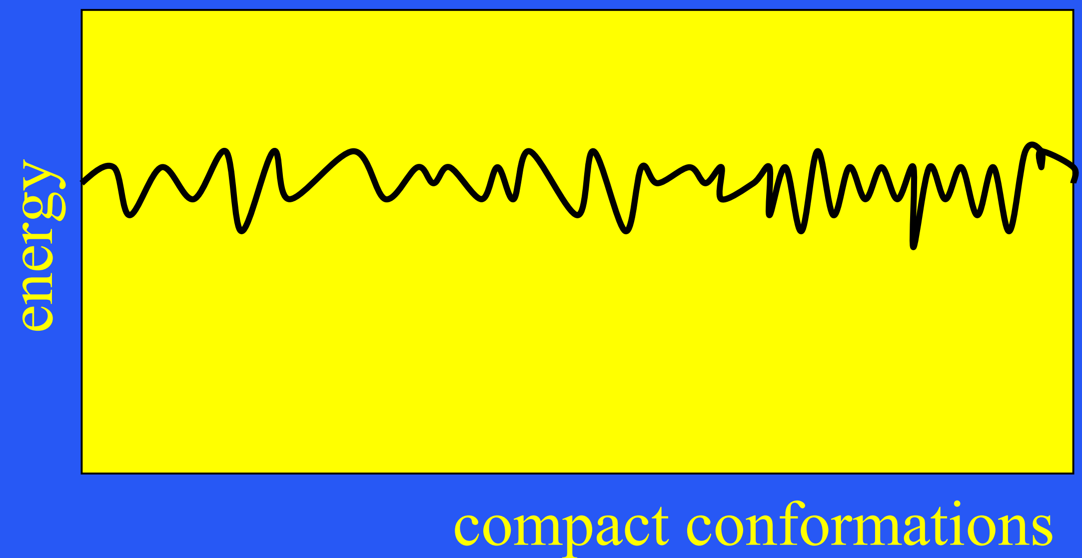


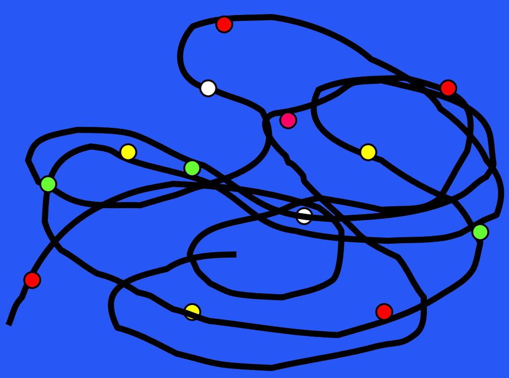
Compact *versus* marginally compact phase



Compact *versus* marginally compact phase

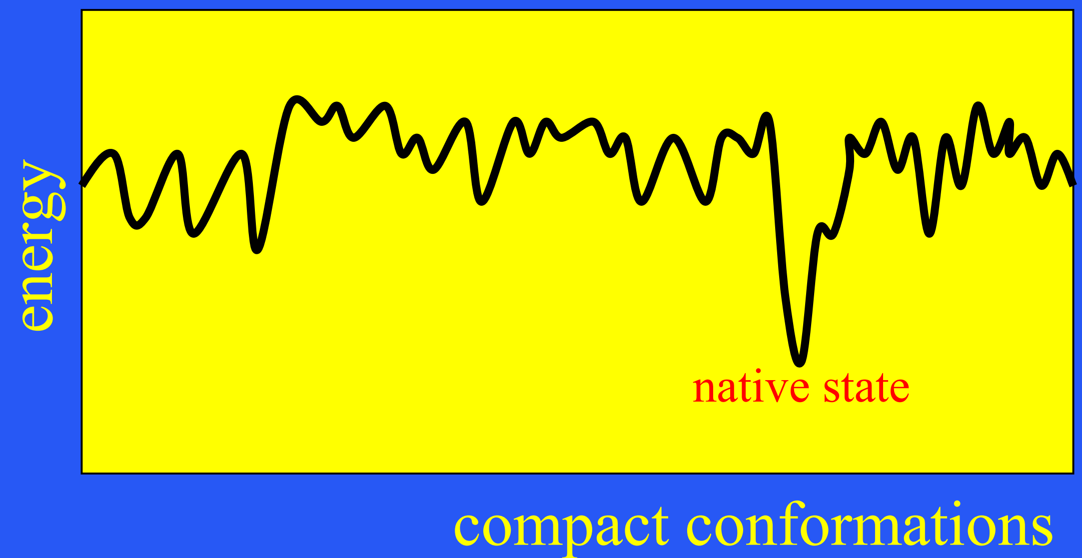
Hetero-polymer

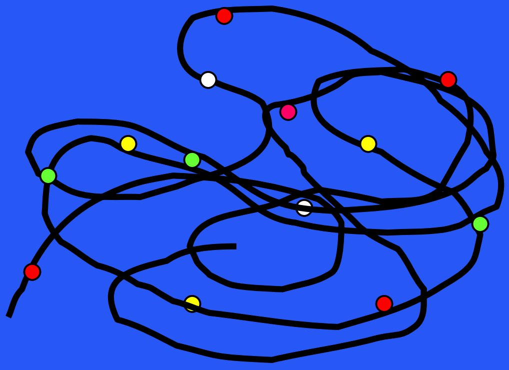




Compact *versus* marginally compact phase

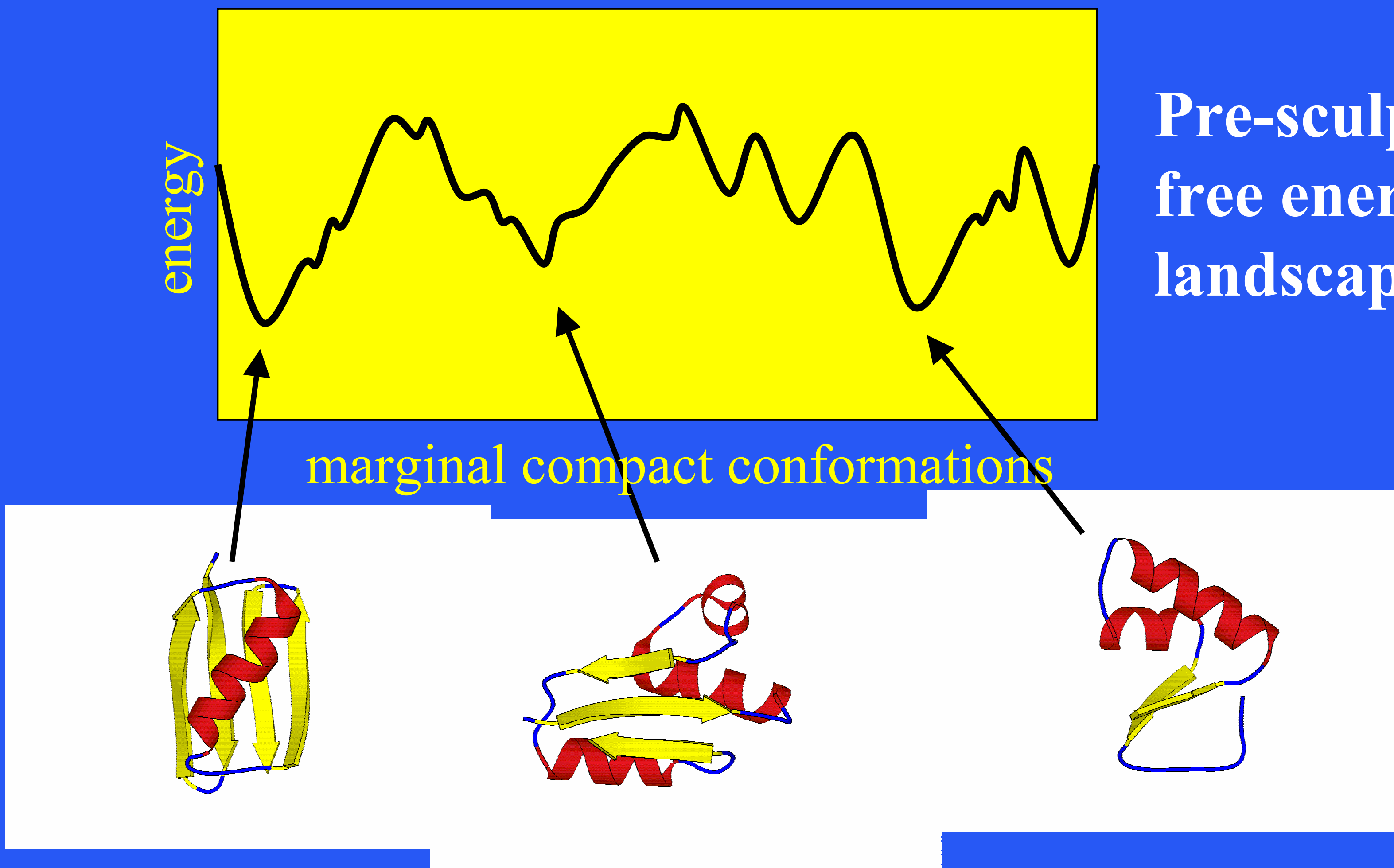
Protein-like sequence





Compact *versus* marginally compact phase

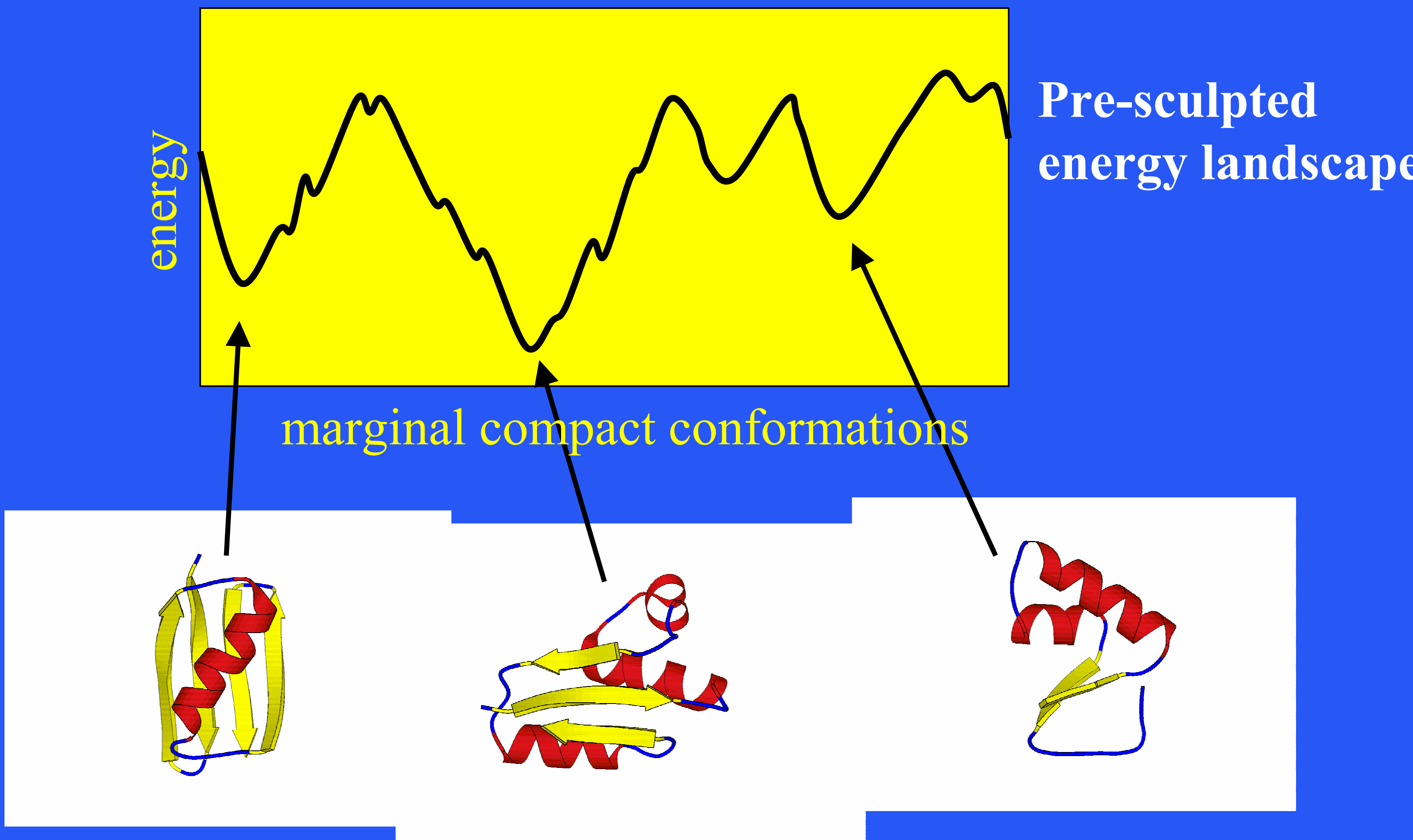
Homopolymer



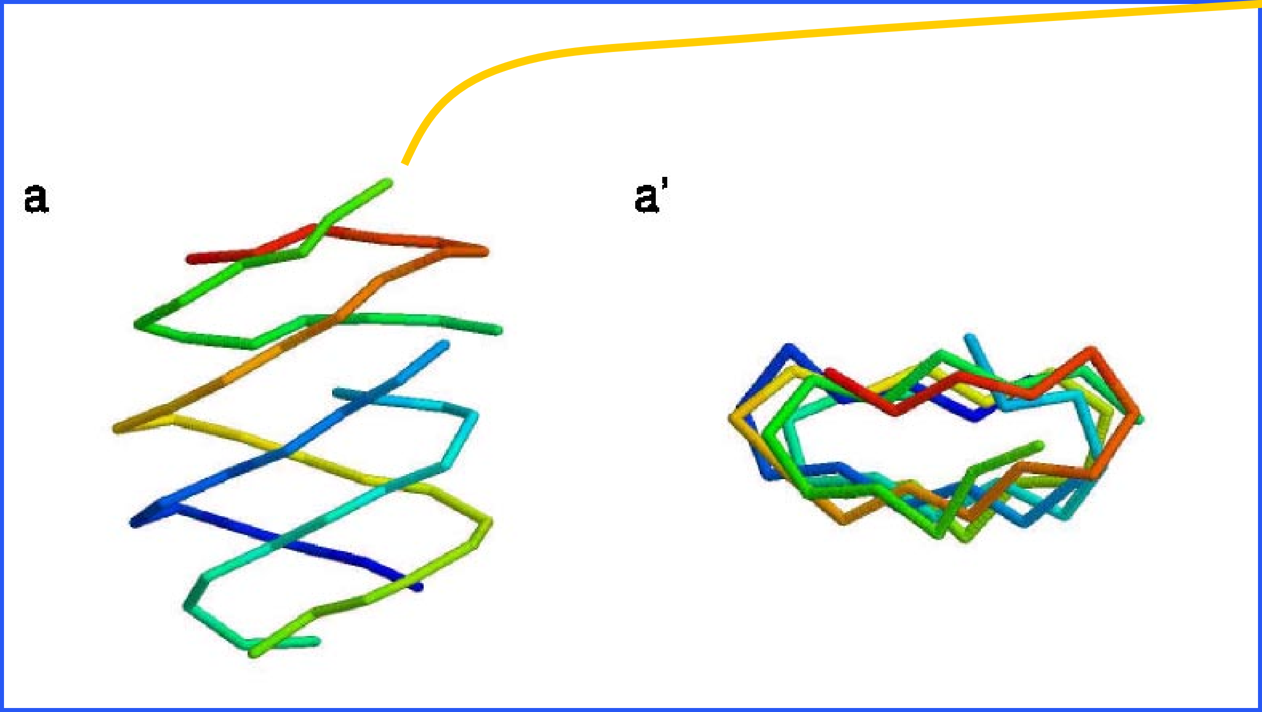
Pre-sculpted free energy landscape

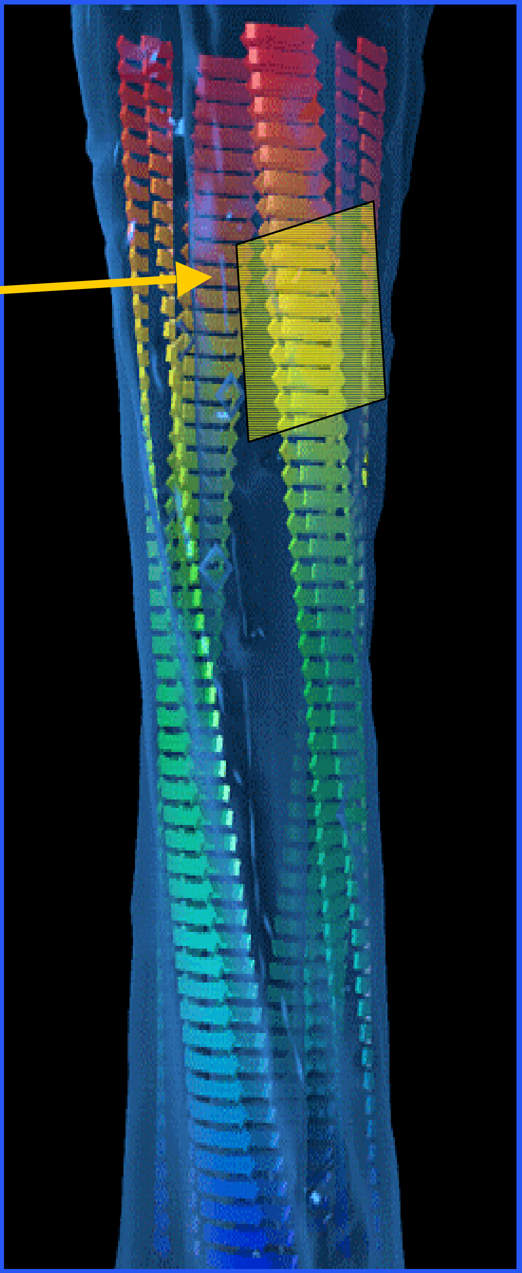
Compact *versus* marginally compact phase

Protein-like sequence



Amyloid aggregation →
The thermodynamic (stable) phase!
6 homopolimeric 12 a.a. chain.

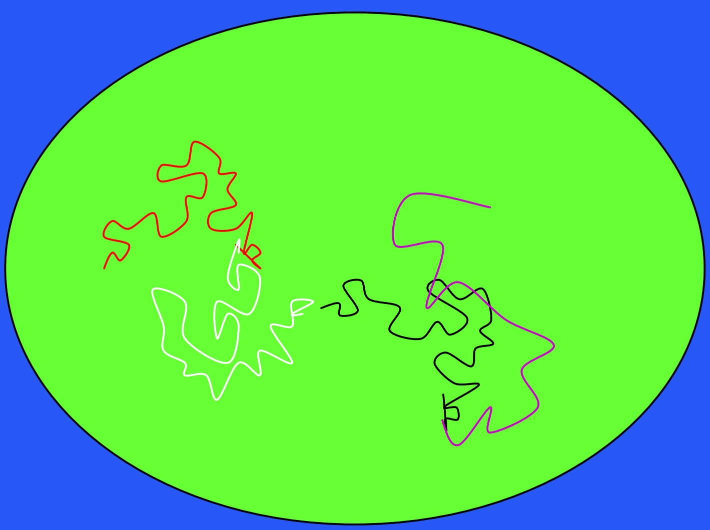


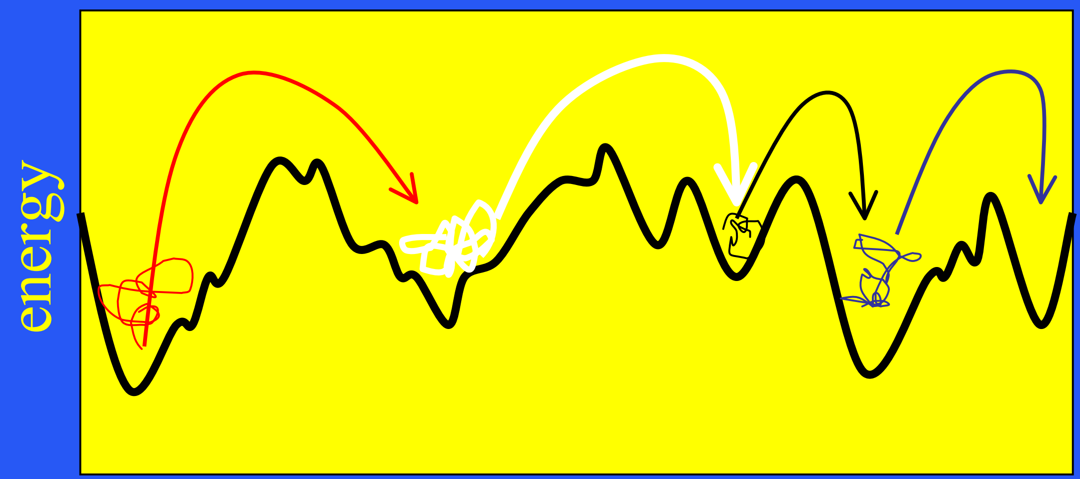


New structure for a many chain systems belonging to the menu

Molecular Evolution:

Random walk in sequence and conformation space





Sequence space

Random walk of useful sequences
→ sequences evolve

Corresponding random walk in the pre-sculpted landscape – minima are <u>immutable</u>

Implications

The contingency- the fixed backdrop-
for evolution of sequences and functionalities

A marginal compact phase for thick homo-polymers (tubes) exists with a pre-sculpted free energy landscape with not too many and not too few fundamental “folds” with secondary motifs → diversity.

The pre-sculpted landscape provides the finite (discrete) menu of folds which does not evolve.

Being a phase it provides stability → Neutral evolution.

The marginality provides the sensitivity (flexibility) to interactions and external perturbations.

In this scenario a sequence chooses its native state from the menu rather than sculpting its own funnel ! They adapt to the pre-defined folds and “evolution” refines protein interaction. Folding is also easier on such a landscape.

From the point of view of the protein-protein interaction co-evolution of folds would be very inefficient.

Another instance of contingency

Reasons for

- Ecosystem complexity-diversity;
- Ecosystem stability;
- Species interactions.

Role of

- Stochasticity;
- Space;
- Speciation.

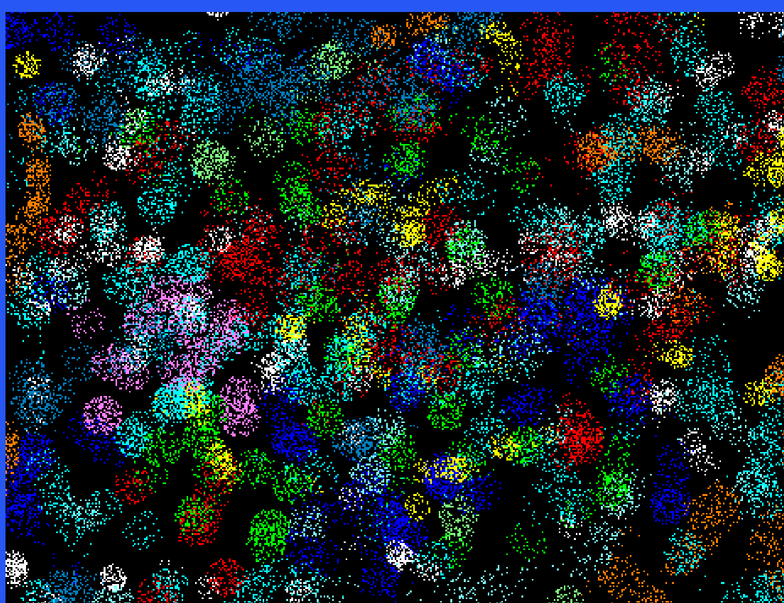
Role of stochastic noise

Deterministic equations (e.g Lotka-Volterra like eqs.) for time evolution of species populations are unable to explain “diverse” ecosystem (May, Nature, 1972).

The introduction of any amount of noise leads to complete extinction on time scale inversely proportional to the noise strength

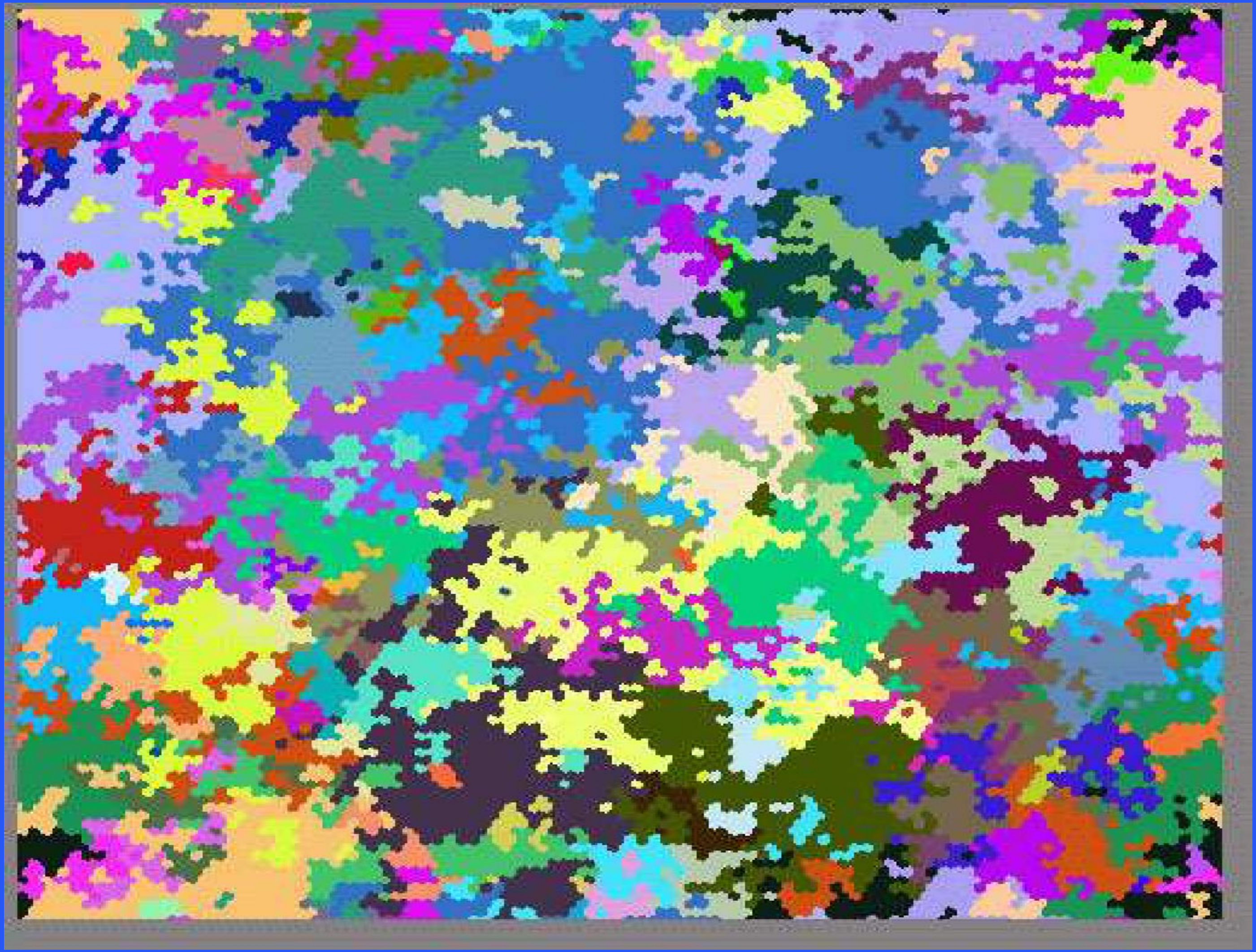
Role of spatial dimensionality – species competing for space

1 dimensional case: species segregate with almost sharp boundary→ low probability of interaction but high efficiency in spatial exploration.



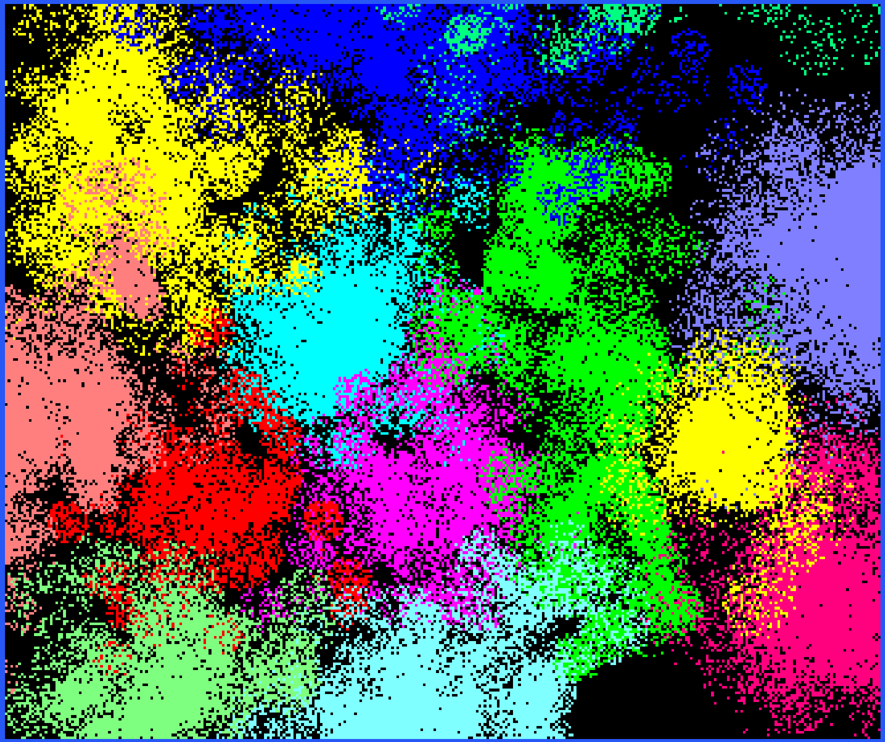
3 dimensional case: Species mix very well → high probability to interact, low efficiency to explore all resources

2 dimensional case



Intermediate between 1 and 3 D case → Species mix enough for stabilizing interactions (on time scales smaller than speciation/extinction occurs) and resources are explored efficently.

2 dimensional case



Intermediate between 1 and 3 D case → Species mix enough for stabilizing interactions (on time scales smaller than speciation/extinction occurs) and resources are explored efficently.

Implications

- Infinite (spatial) size system → no extinction below a noise threshold! ➔ Finite system has a non-trivial transient with diversity, similar to the infinite size system, before extinction occurs (time to extinction grows with system size).
- 2 D finite system progeny has the possibility to explore resources and develop stabilizing interactions which delays further extinction

- BUT…. we have assumed species are "discrete"!

General model with few key ingredients