Computer-Aided Conformation-Dependent Design of Copolymer Sequences

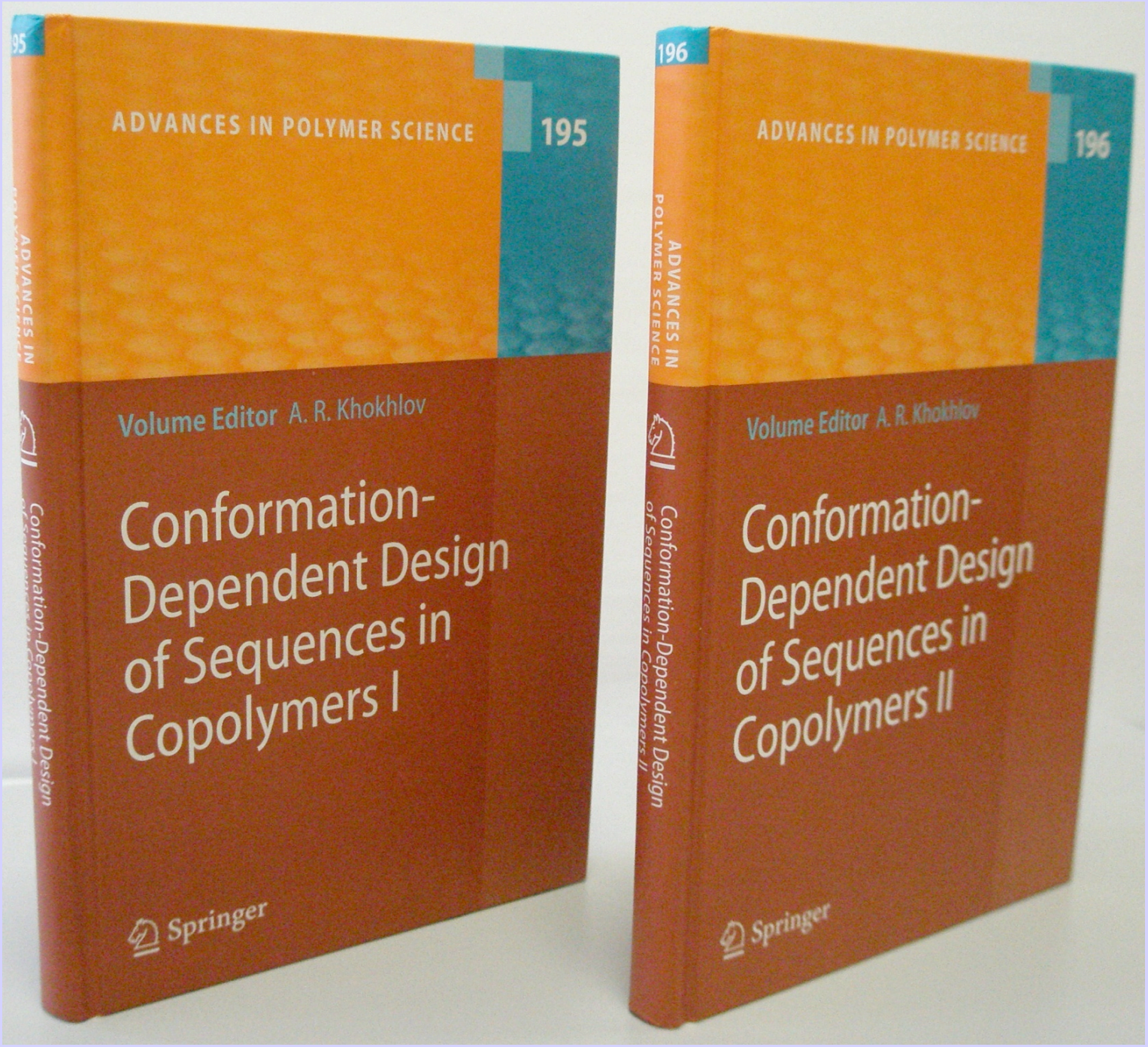
Alexei R.Khokhlov
Moscow State University

Aim: polymers with sophisticated functions

Two approaches:

1) Synthesis of novel monomer units where the required function is linked to the chemical structure of these units.

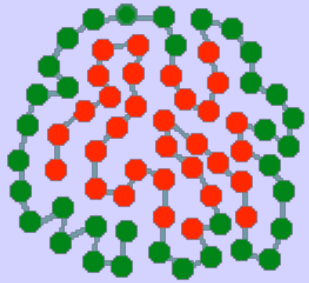
2) Design of sequences of known monomer units in copolymer chains (biopolymers in living systems followed this way in the course of molecular evolution => biomimetic approach).


ADVANCES IN POLYMER SCIENCE
195
Volume Editor A. R. Khokhlov
Conformation-Dependent Design of Sequences in Copolymers I
Springer
ADVANCES IN POLYMER SCIENCE
196
Volume Editor A. R. Khokhlov
Conformation-Dependent Design of Sequences in Copolymers II
Springer

Globular proteins - enzymes

1. Are soluble in aqueous media

2. Adopt the globular state in aqueous media



For homopolymers and random copolymers these requirements contradict to each other

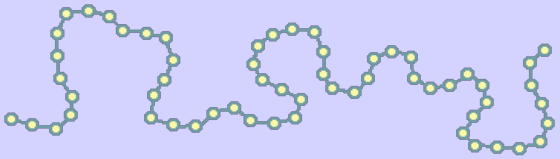
Hydrophobic A-units (red) form the dense core of the globule, while hydrophilic B-units (green) - constitute the stabilizing envelope for the core.

Question: is it possible to design such an AB-sequence of a synthetic copolymer, so that in its most dense globular conformation all A-units would be in the core of the globule, and all B-units would form the envelope of this core ?

→ protein-like AB-copolymers

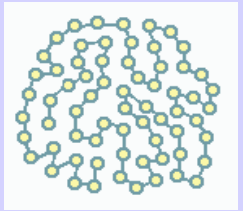
Computer realization of protein-like AB-copolymers

Stage 1



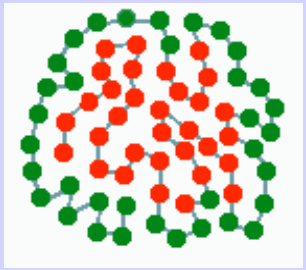
Homopolymer coil with excluded volume.

Stage 2



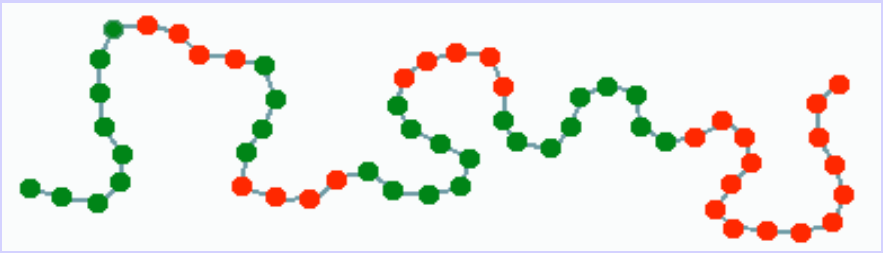
The strong attraction of links is switched on. The homopolymer globule is formed.

Stage 3



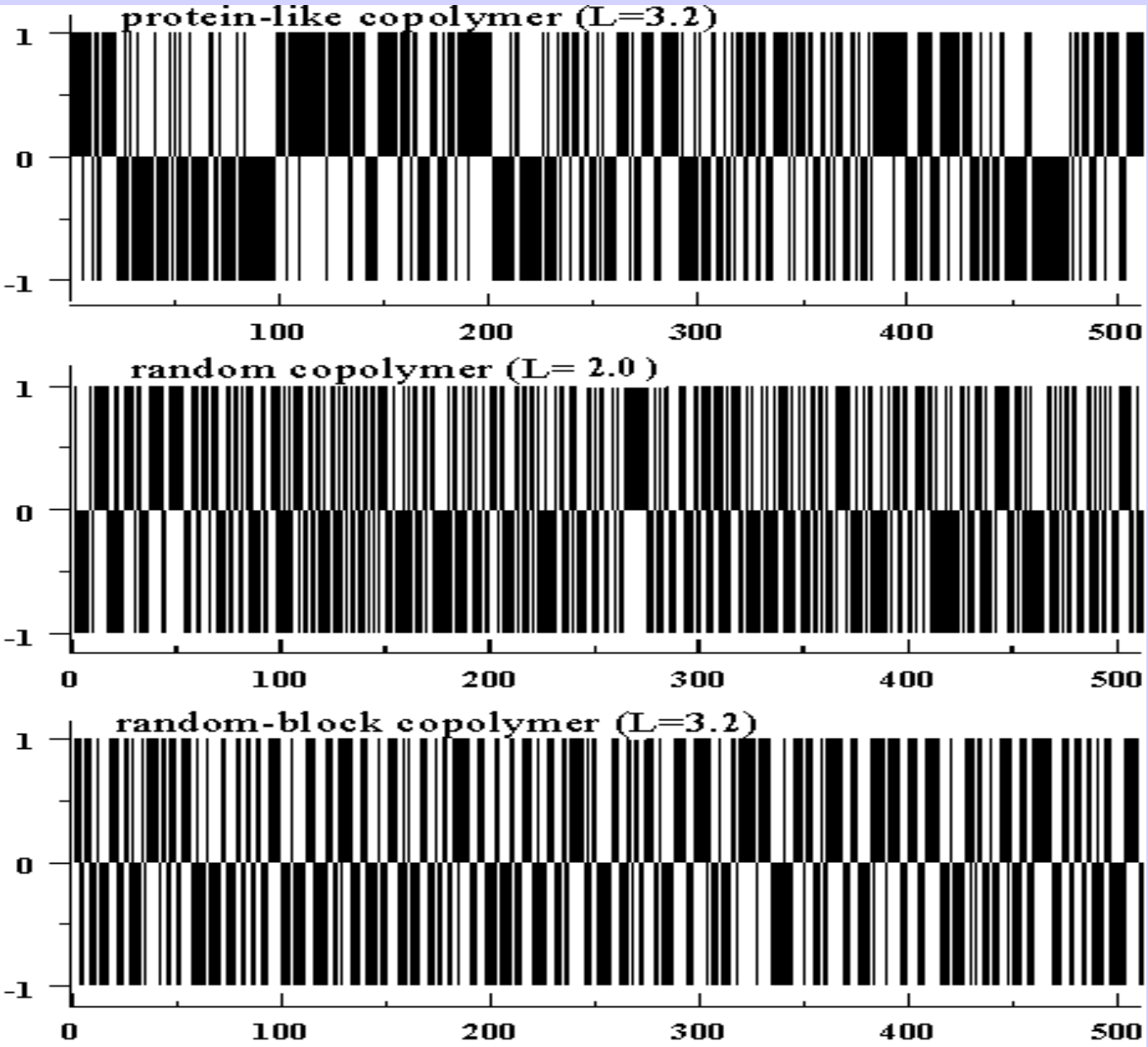
"Instant photo" of the globule is considered. Links on the surface are colored in green and called hydrophilic. Links in the core are colored in red and called hydrophobic. Then the primary structure is fixed.

Stage 4



The uniform strong attraction is removed and different interaction potentials are introduced for green and red links. The protein-like copolymer is ready.

Analysis of the primary structure



Poisson distribution for the length x of A and B sequences: $W(x) = \exp(-L)L^x / x!$

Statistical correlations in the protein-like sequence

...ABBAABAAAABAABBBBABAABB...

L

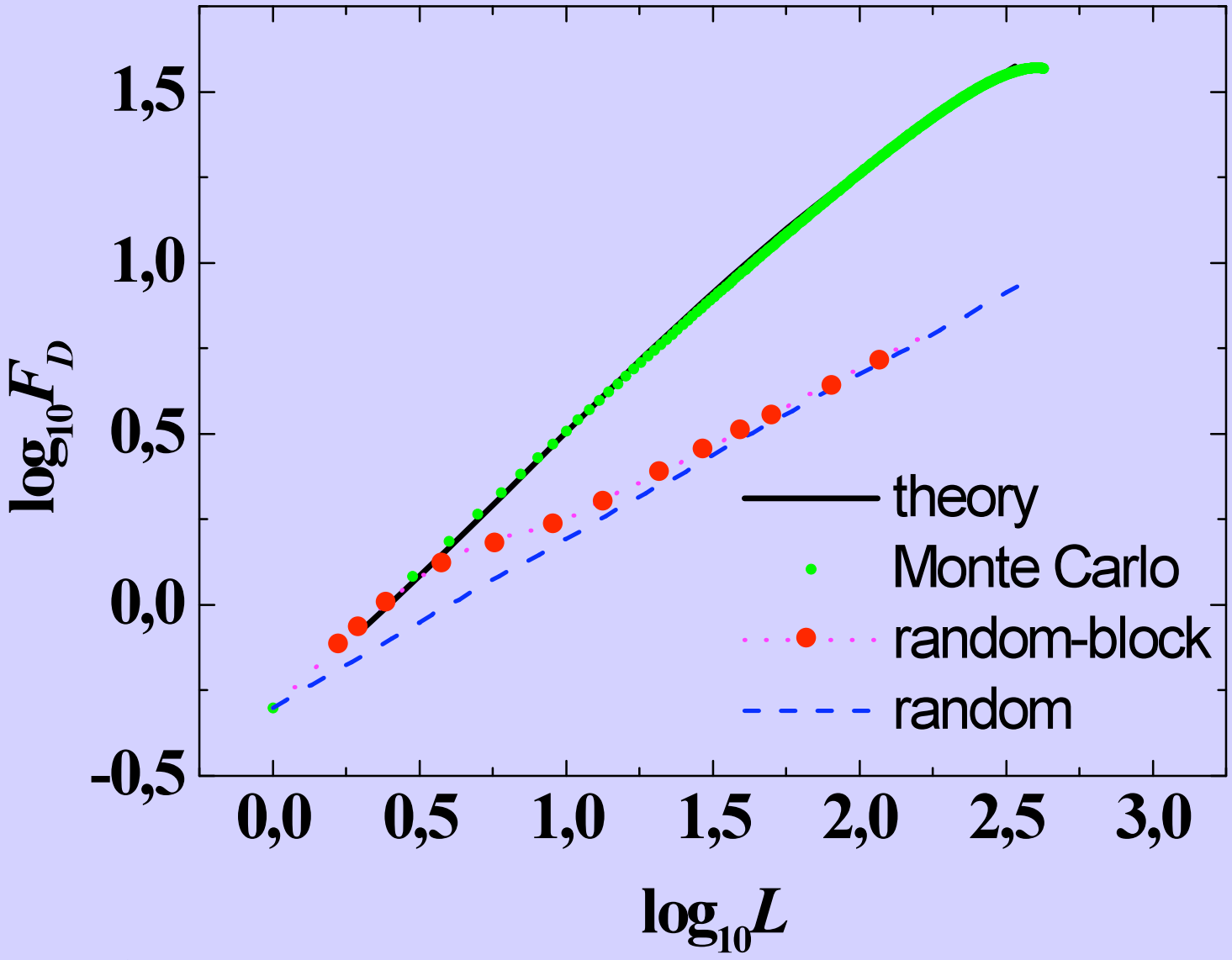
The number of A-units in the "window" of length L is a random variable, depending on the "window" position.

Dispersion of the distribution of this random variable, F_D :

for a random copolymer $F_D \sim L^{1/2}$

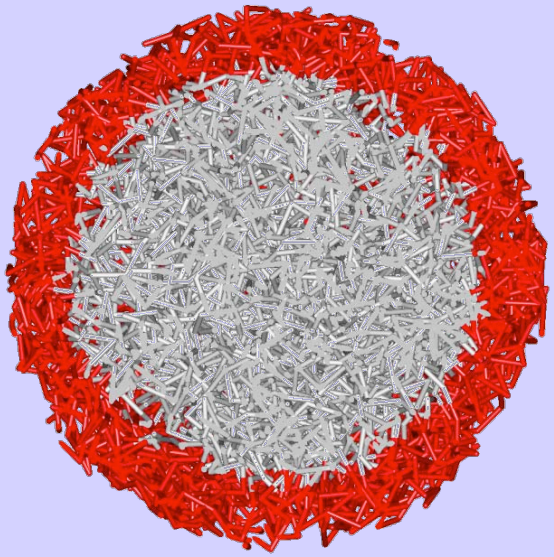
for a copolymer with long-range power-law correlations

$F_D \sim L^{\alpha}$, $\alpha \neq 1/2$

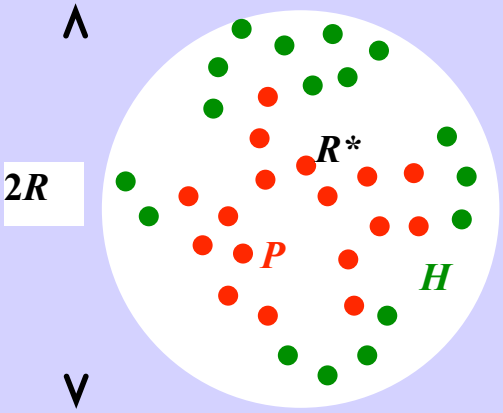


for random copolymers $\alpha = 0.5$

for protein-like copolymers $0.75 < \alpha < 1$



$G(\vec{r},k\,|\,\vec{r}_0)$ is the statistical weight of random walk trajectories starting at r_o and arriving at r after k steps



Function $G(\vec{r},k\,|\,\vec{r}_0)$ satisfies the diffusion equation:

$$\frac{\partial G(\vec{r},k\,|\,\vec{r}_0)}{\partial k}=\frac{a^2}{6}\Delta G(\vec{r},k\,|\,\vec{r}_0)+\delta(k)\delta(\vec{r}-\vec{r}_0)$$

Boundary conditions:

$$\vec{\nabla}_{\vec{r}}G(\vec{r},k\,|\,\vec{r}_0)\Big|_{r=R}=0$$

$$G_{H(P)}(\vec{r},k\,|\,\vec{r}_0)\Big|_{|\vec{r}|=R^*}=0$$

Solution:

1. The probability distributions of lengths k of hydrophobic $P_H(k)$ and polar $P_P(k)$ sections

$$P_H(k) = \frac{\pi a^2}{3R^* r_0} \sum_{n=1}^{\infty} n(-1)^{n+1} \exp\left[-\frac{a^2}{6}\left(\frac{n\pi}{R^*}\right)^2 k\right] \sin\left(n\pi \frac{r_0}{R^*}\right)$$

$$P_P(k) = \frac{a^2 R^*}{3\left(R - R^*\right) r_0} \sum_{n=1} \exp\left[-\frac{a^2}{6}\left(\frac{\zeta_n}{R - R^*}\right)^2 k\right] \frac{\zeta_n^2 \sin\left(\zeta_n \dfrac{r_0 - R^*}{R - R^*}\right)}{\zeta_n - \sin\zeta_n \cos\zeta_n}$$

ζ_n satisfies the equation $\zeta_n = (1-R^*/R)\ \mathrm{tg}\ \zeta_n$.

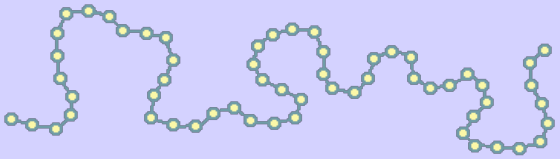
Asymptotic forms:

$$P_{H(P)} \cong \begin{cases} k^{-3/2}, & \text{for } 1 < k < (d_{H(P)}/a)^2; \\ \left(\dfrac{d_{H(P)}}{a}\right)^{-3} \exp(-\lambda_{H(P)} k), & \text{for } k > (d_{H(P)}/a)^2 \end{cases}$$

These probability distributions correspond to so-called **Levy-flight statistics**.

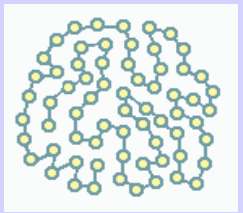
Computer realization of protein-like AB-copolymers

Stage 1



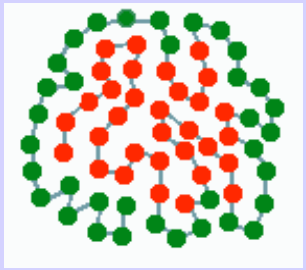
Homopolymer coil with excluded volume.

Stage 2



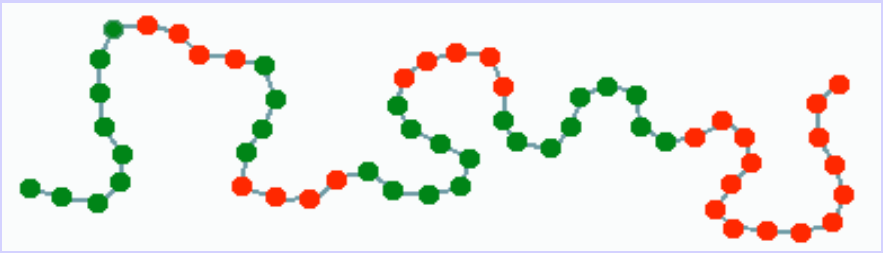
The strong attraction of links is switched on. The homopolymer globule is formed.

Stage 3

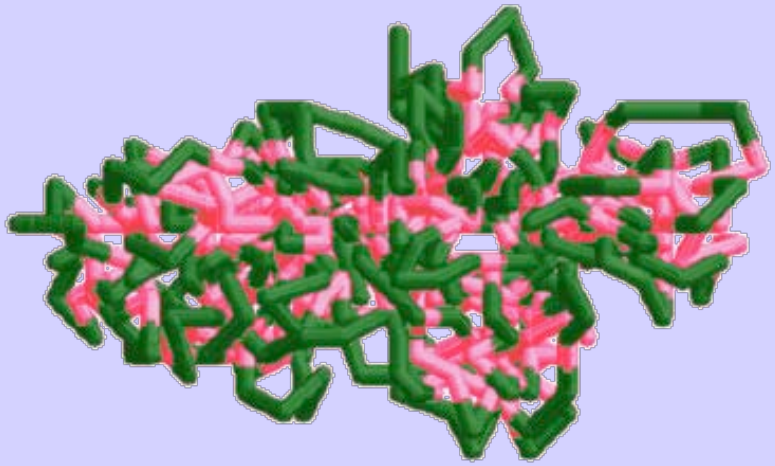


“Instant photo” of the globule is considered. Links on the surface are colored in green and called hydrophilic. Links in the core are colored in red and called hydrophobic. Then the primary structure is fixed.

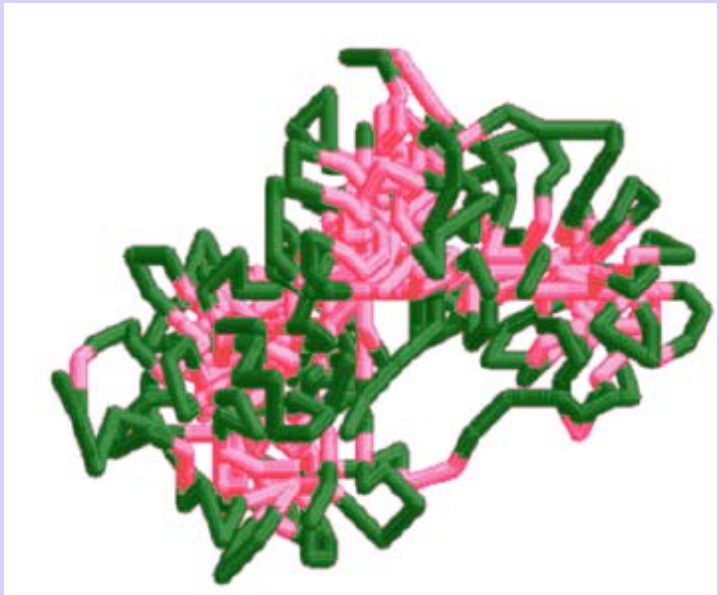
Stage 4



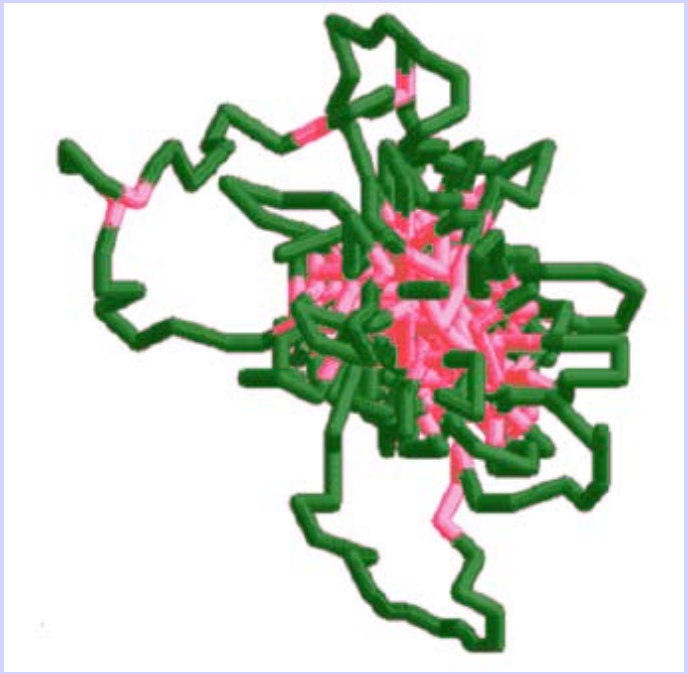
The uniform strong attraction is removed and different interaction potentials are introduced for green and red links. The protein-like copolymer is ready.



Random copolymer $\langle R_g^2 \rangle_{core}$= 106.6



Random-block copolymer $\langle R_g^2 \rangle_{core}$=99.4

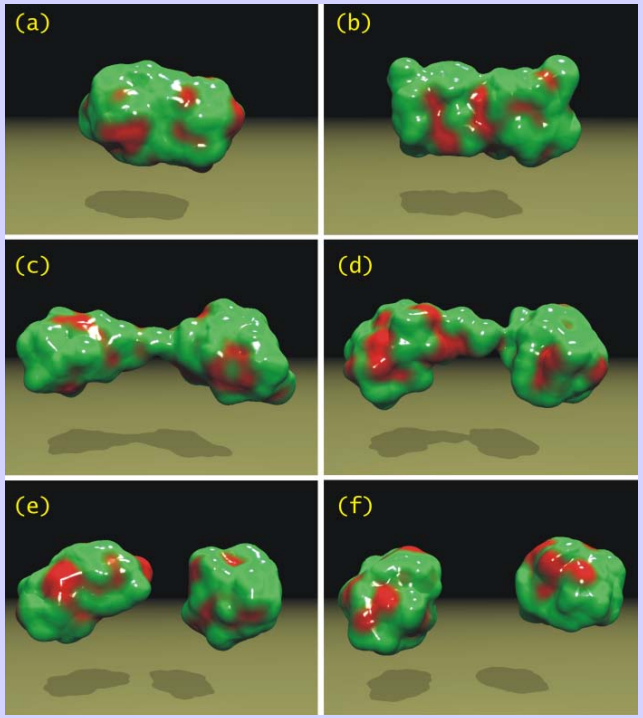


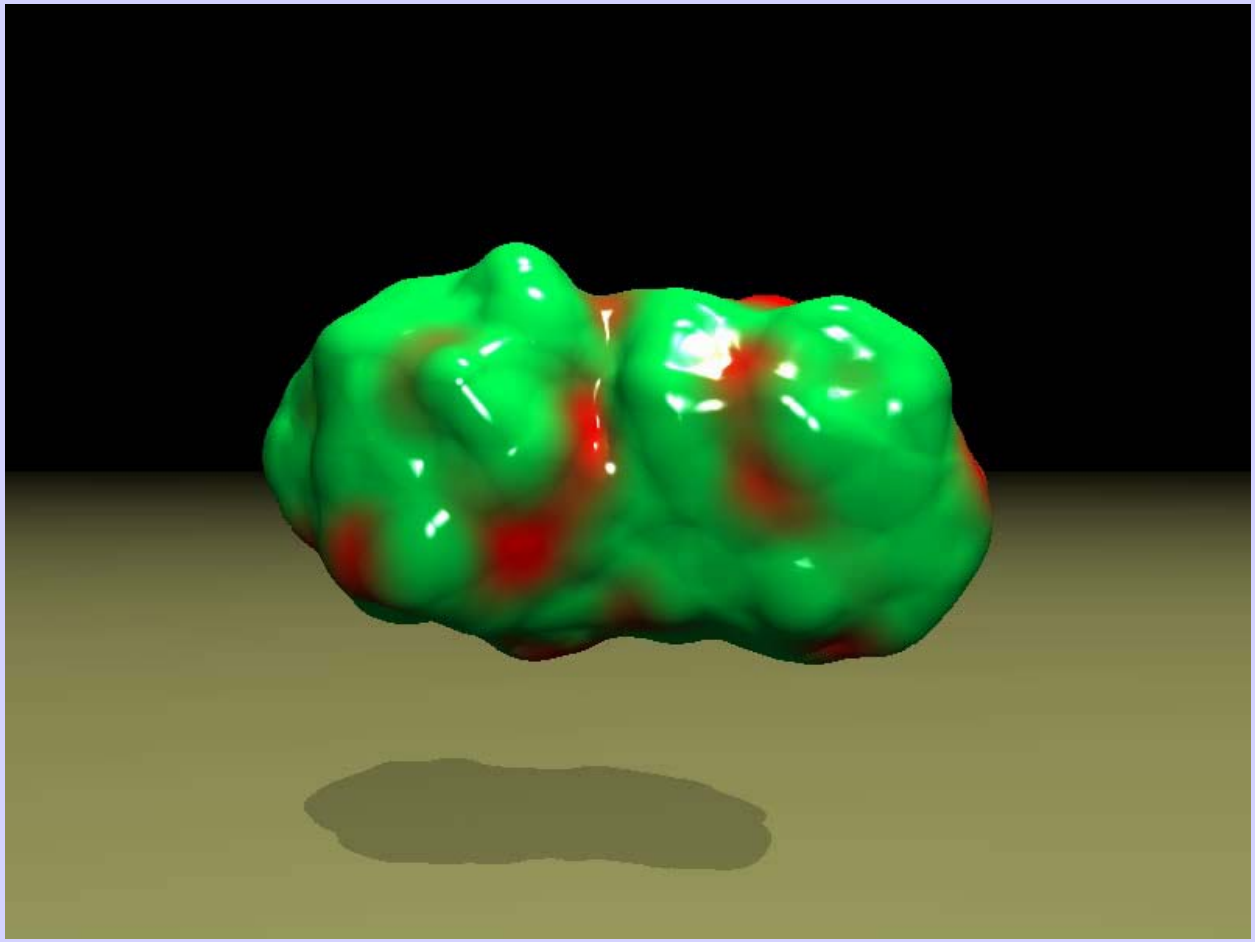
Protein-like copolymer

$\langle R_g^2 \rangle_{core}$=74.1

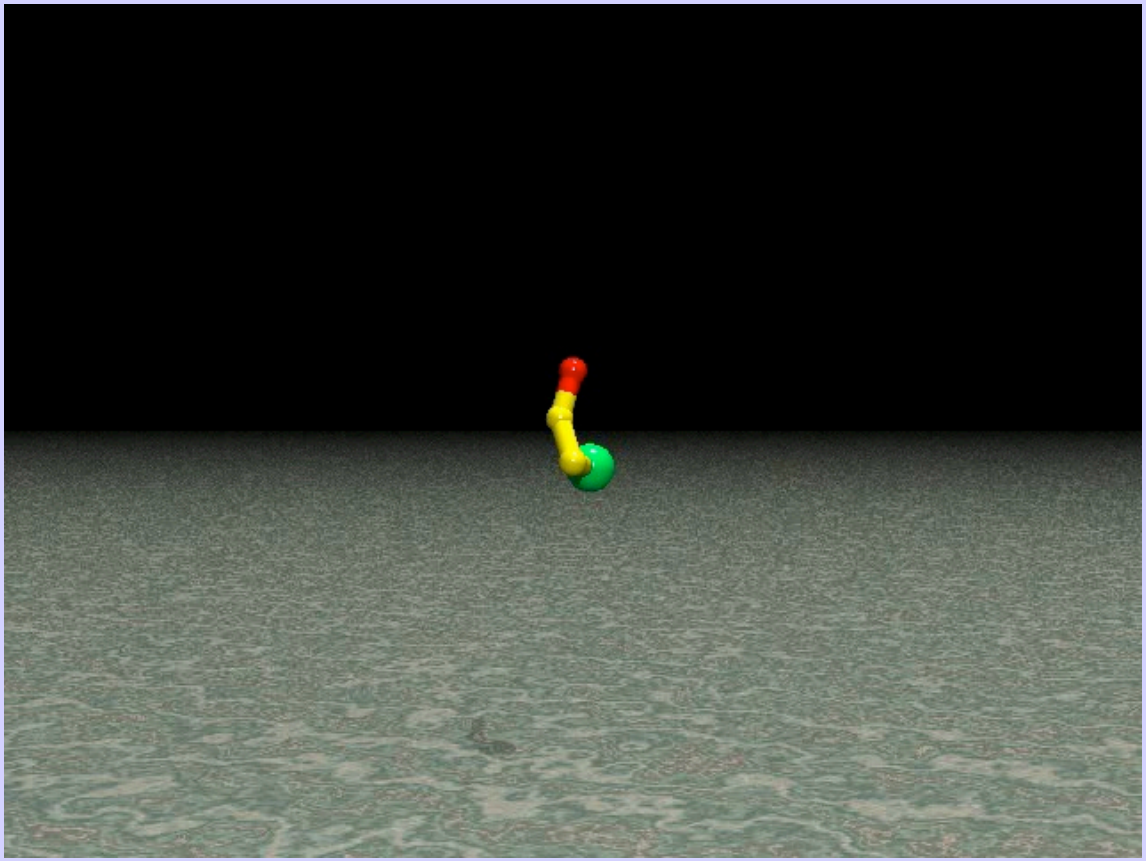
Stabilization vs. Aggregation

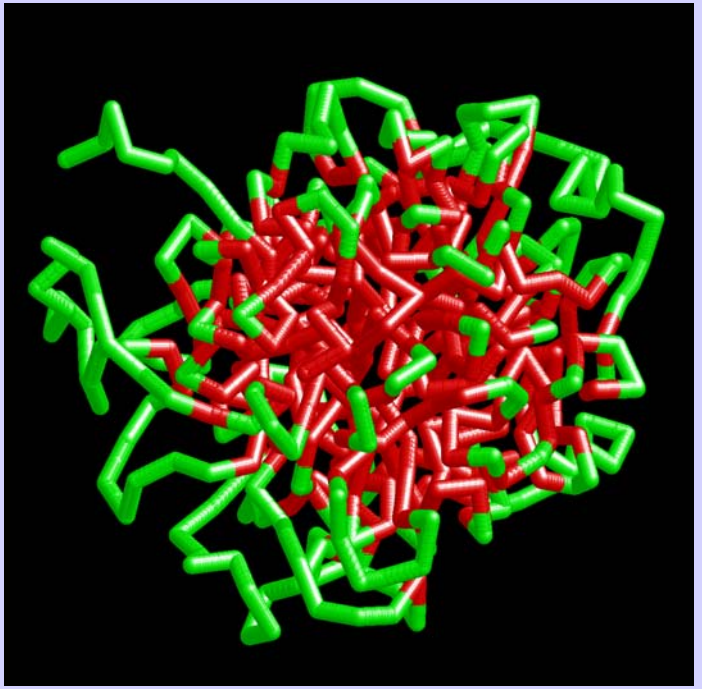
Two proteinlike globules at poor solvent conditions





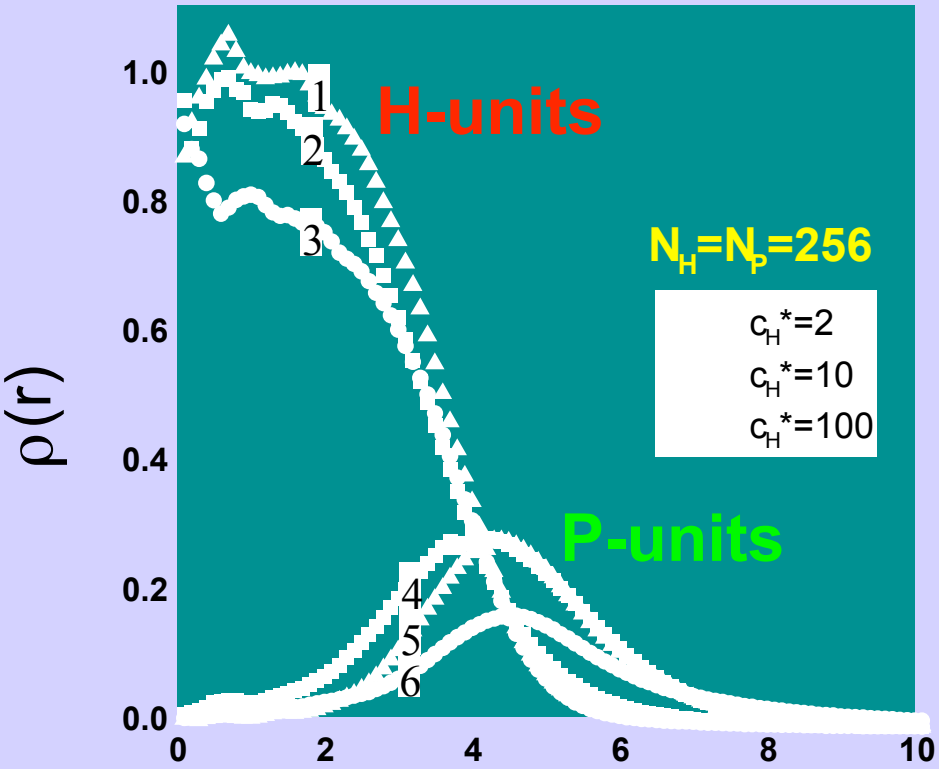
Copolymerization with simultaneous globule formation





Copolymerization with simultaneous globule formation

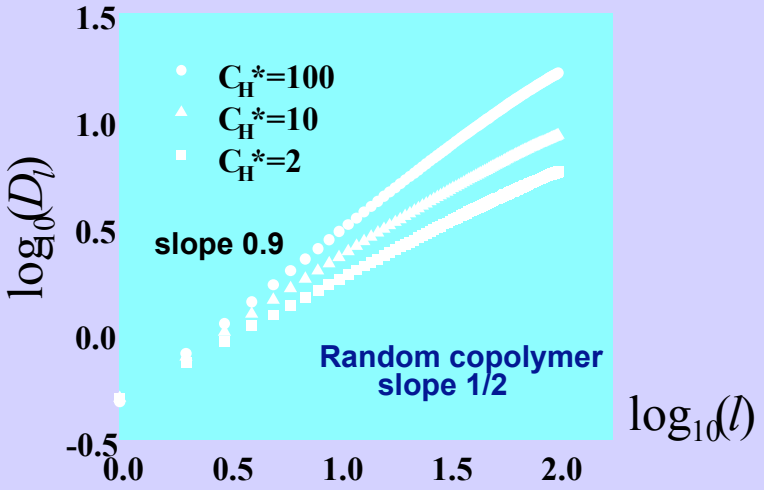
RADIAL DISTRIBUTION of H and P UNITS



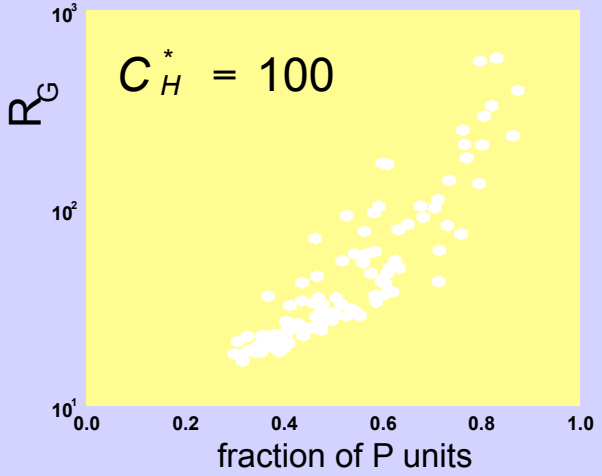
P-units are outside and
H-units are inside the globule

LONG-RANGE CORRELATIONS

Dispersion of H-units as a fulnction of lag length



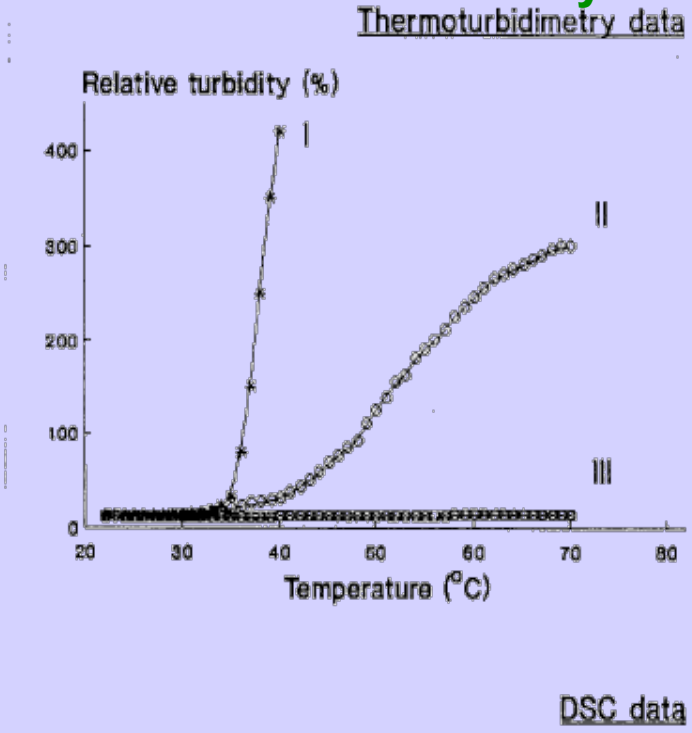
MEAN-SQUARE GYRATION RADIUS



Copolymerization with Simultaneous Globule Formation

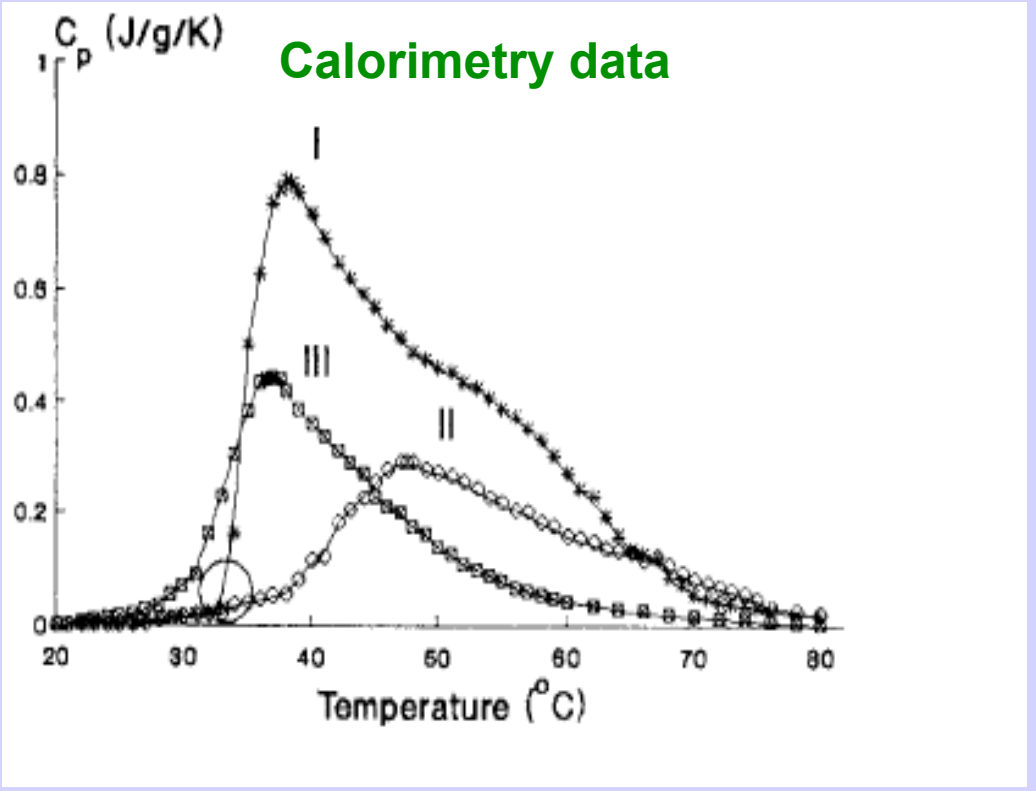
Copolymerization of N-Vinylcaprolactam and N-Vinylimidazole was performed for different temperatures: I – 20 °C; II – 45 °C; III – 65 °C.

Thermoturbidimetry data



DSC data

Calorimetry data



Conclusion: for the case III coil-globule transition takes place without precipitation of globules

Concept of Evolution in Polymer Science

- Biopolymers (proteins, DNA, RNA) possess complicated sequences of monomer units which encode their functions and/or structure.

- These sequences should be statistically different from random ones, primarily from the viewpoint of information content.

- On the other hand, first copolymers at the very beginning of molecular evolution can be only random (zero information content).

Question: how to describe the increase of information complexity of copolymer sequences in the course of molecular evolution

Since information content is a mathematically defined quantity, this question is quantitative.

Because of the lack of information on the early prebiological evolution, this question is very difficult. Therefore, of particular interest are "toy models" of evolution of sequences which show different possibilities of appearance of statistical complexity in the sequence.

This can be achieved via coupling of polymer chain conformation and evolution of sequences.

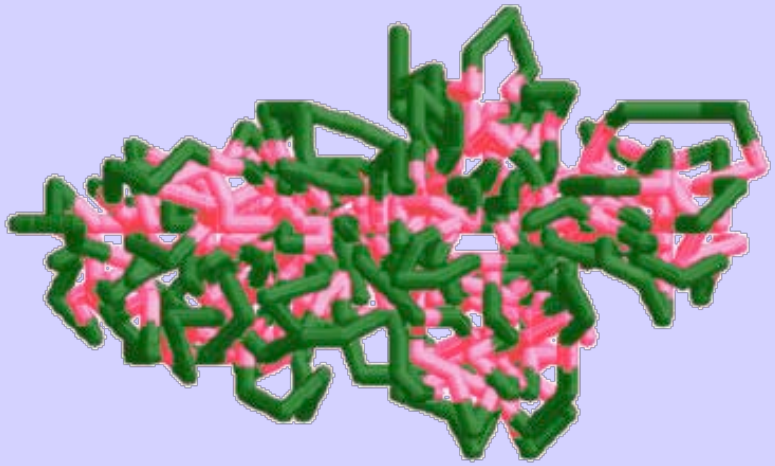
How to introduce explicitly the concept of evolution of sequences into the scheme of generation of protein-like copolymers ?

Formation of initial protein-like sequence ▸ **Refolding to a new globule due to the attraction of H-units** ▸ **"Recoloring" in the newly formed globule** ▸ **Transition to a coil** ▸

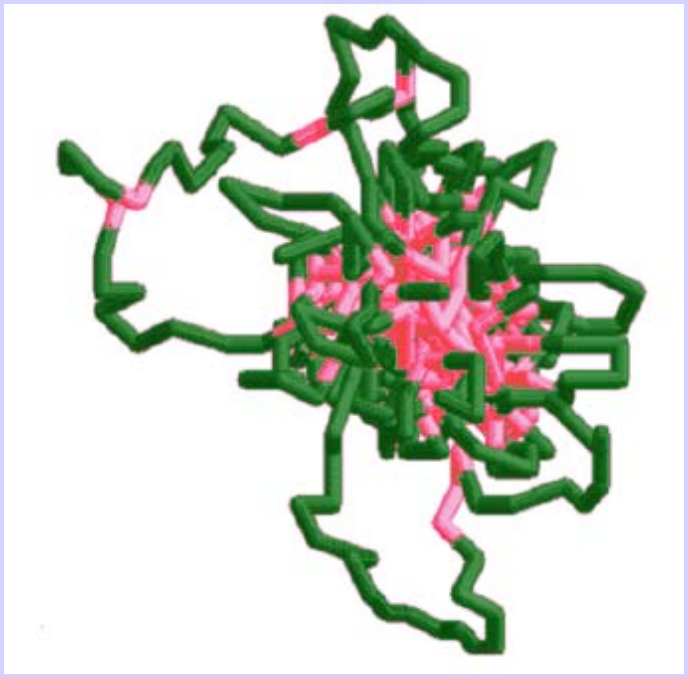
▸ **Refolding to a new globule** ▸ **etc**

As a result, we obtain some evolution of sequences which depend on the interaction parameters of the refolding process.

Question: whether this evolution leads to the increase of complexity (ascending branch of the evolution) or we will end up with some trivial sequence (descending branch of the evolution)?

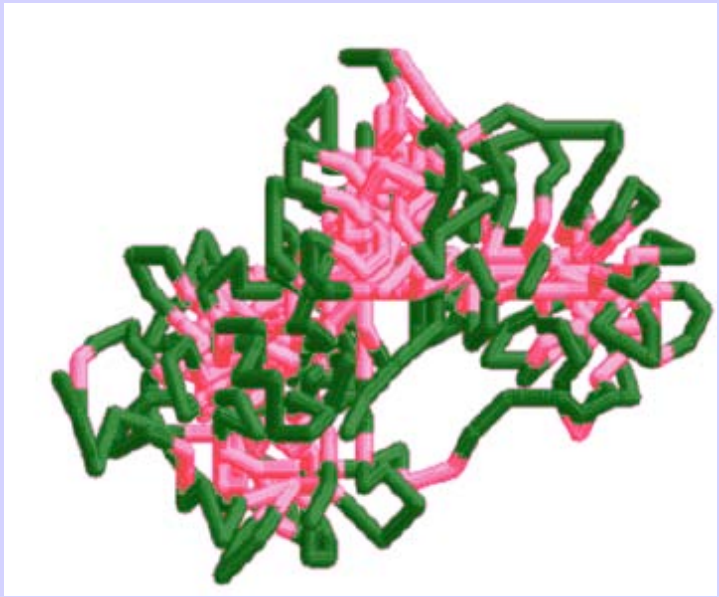


Random copolymer $\langle R_g^2 \rangle_{core} = 106.6$



Protein-like copolymer

$\langle R_g^2 \rangle_{core} = 74.1$



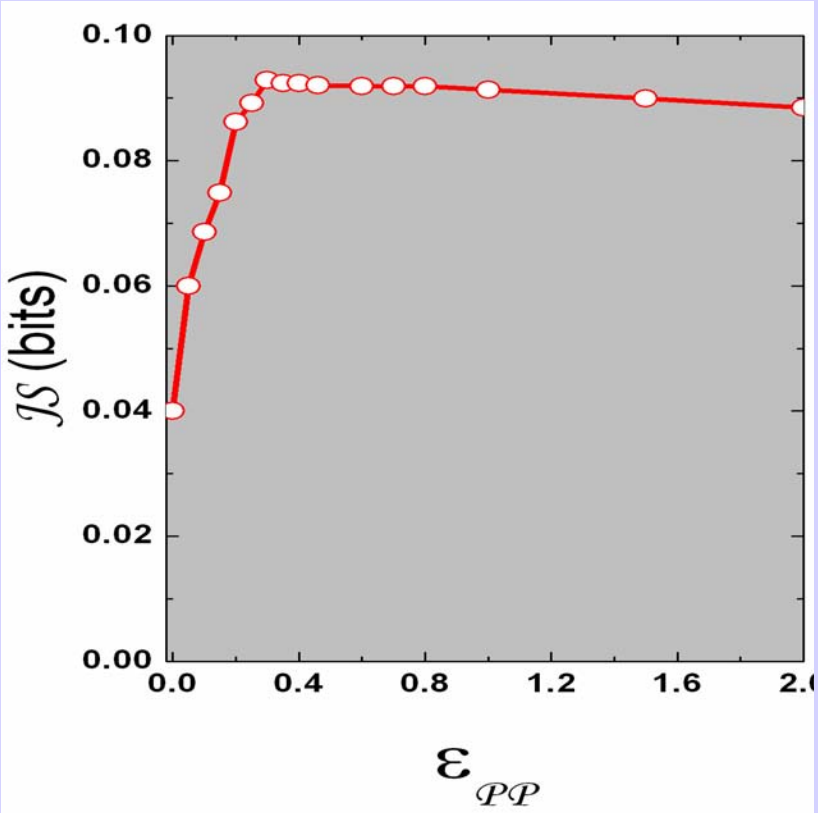
Random-block copolymer $\langle R_g^2 \rangle_{core} = 99.4$

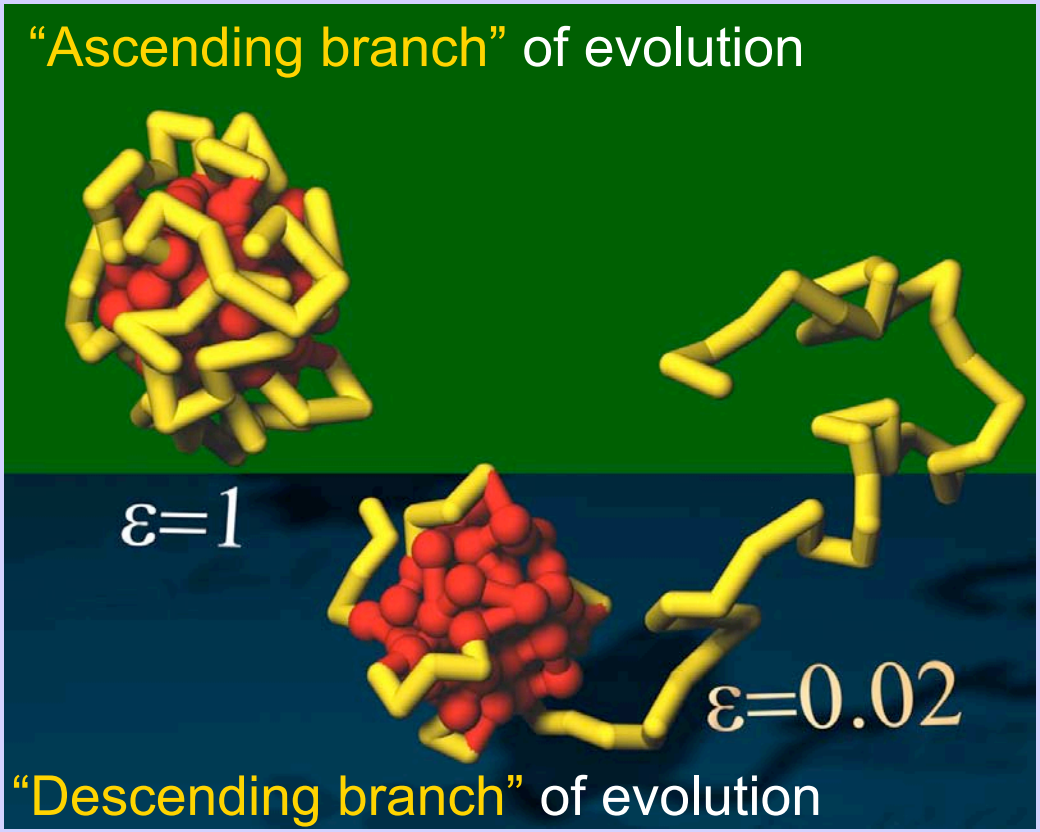
“REPEATED COLORING” =

coloring + equilibration (Molecular dynamics) +

new coloring + … etc.

$\varepsilon_{HH} = 2kT;$ $\varepsilon_{PP} = \varepsilon$ is variable (in kT units)





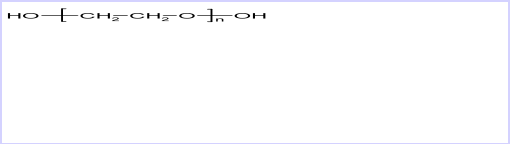
Aim: polymers with sophisticated functions

Two approaches:

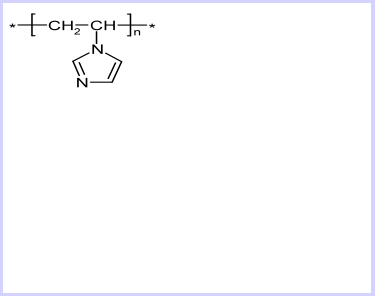
1) Synthesis of novel monomer units where the required function is linked to the chemical structure of these units.

2) Design of sequences of known monomer units in copolymer chains (biopolymers in living systems followed this way in the course of molecular evolution => biomimetic approach).

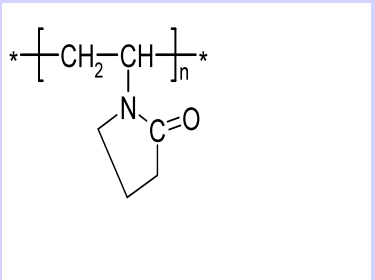
Most of "hydrophilic" monomer units are actually amphiphilic



- poly (ethylene oxide)

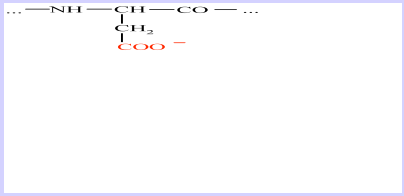


- poly(1-vinylimidazole)

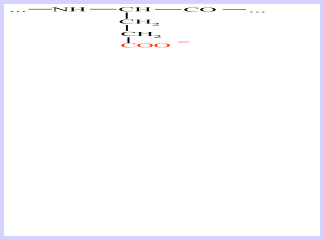


- poly (vinylpirrolidone)

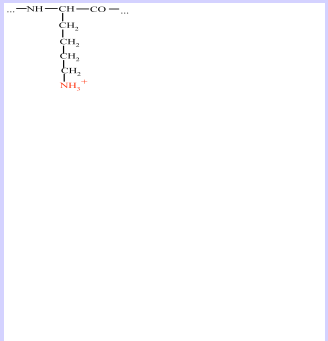
Most of "hydrophilic" monomer units in proteins are actually amphiphilic



Aspartic acid
(Asp)

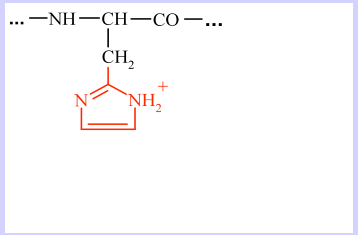


Glutamine
(Glu)

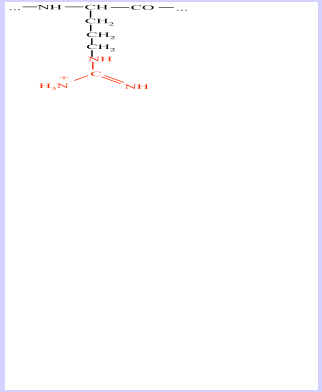


Lysine
(Lys)

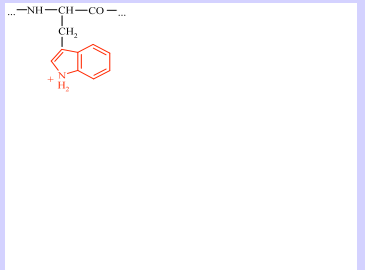
Hydrophilic charged



Histidine
(His)

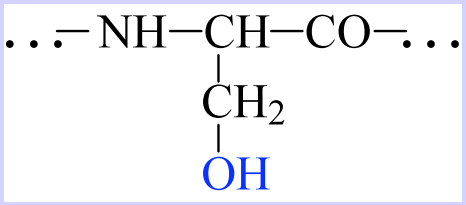


Arginine
(Arg)

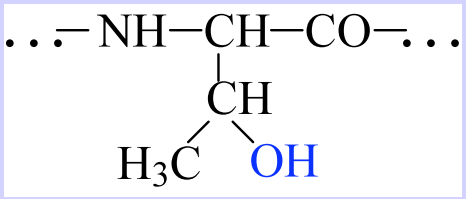


Tryptophane
(Trp)

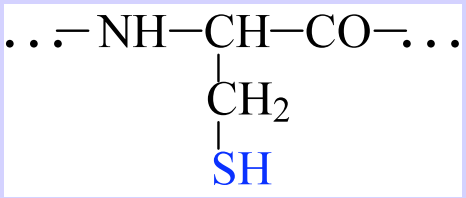
Most of "hydrophilic" monomer units in proteins are actually amphiphilic



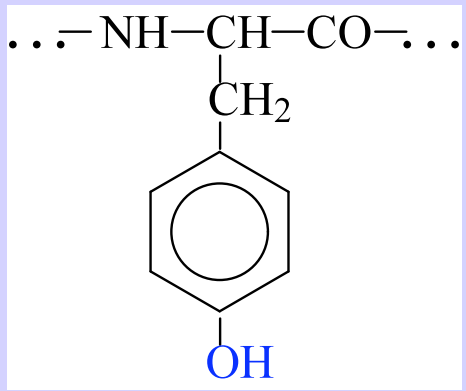
Serine (Ser)



Threonine (Thr)

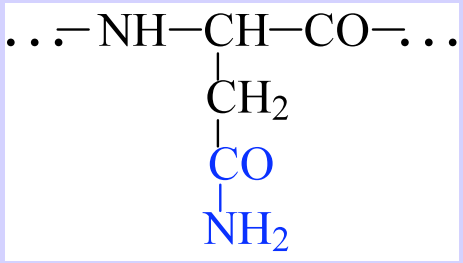


Cysteine (Cys)

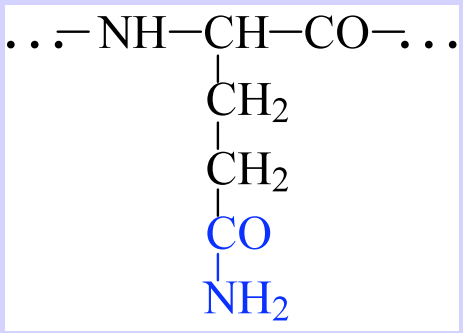


Tyrosine (Tyr)

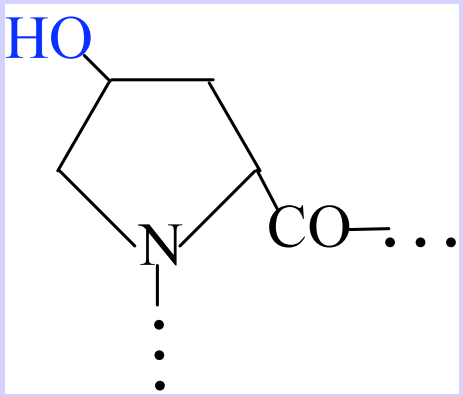
Hydrophilic non-charged



Asparagine (Asn)

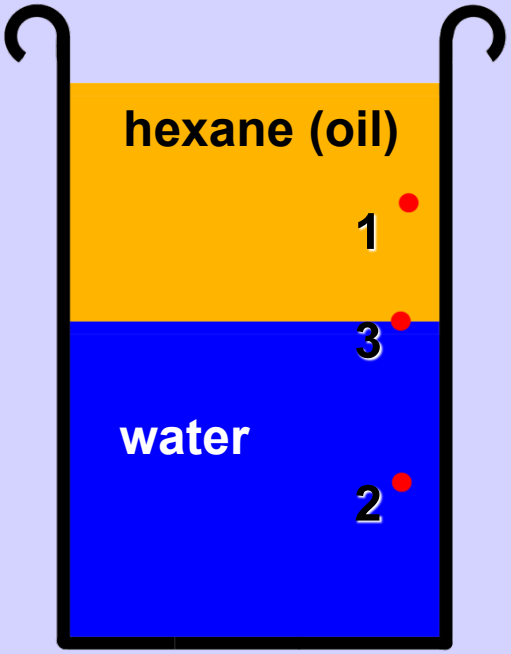


Glutamine (Gln)



Oxyproline (Hyp)

Amphiphilic monomers units should be surface active. It is necessary to take this into account, especially in biological and bio-inspired systems with abundance of water/organic media interfaces.

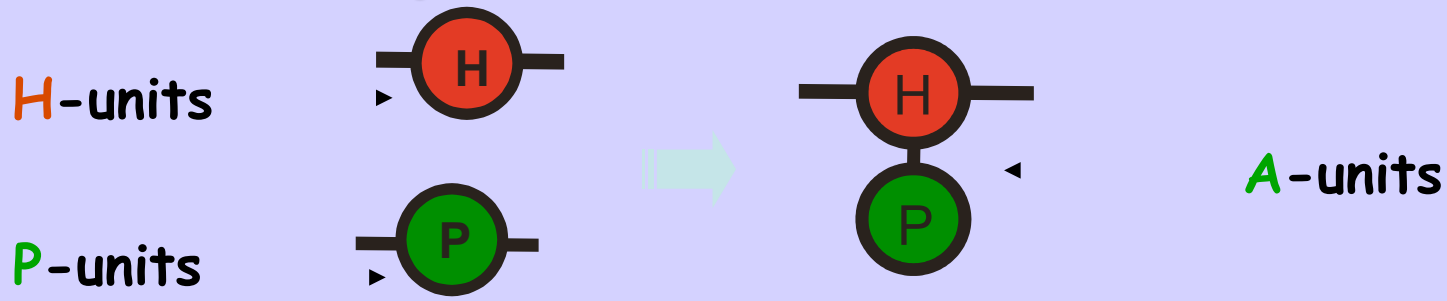


A given monomer unit in two-phase system has three (not two!) choices:

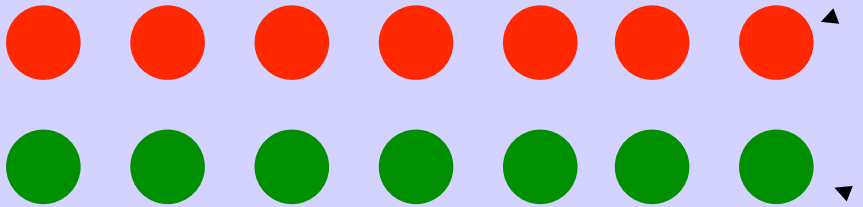
1. oil: free energy f_1
2. water: free energy f_2
3. Boundary oil/water: free energy f_3

If $f_1 < f_2, f_3$ – hydrophobic monomer unit
If $f_2 < f_1, f_3$ – hydrophilic monomer unit
If $f_3 < f_1, f_2$ – amphiphilic monomer unit

How to model amphiphilic monomer units? Simplest possibility is to represent it as dimer



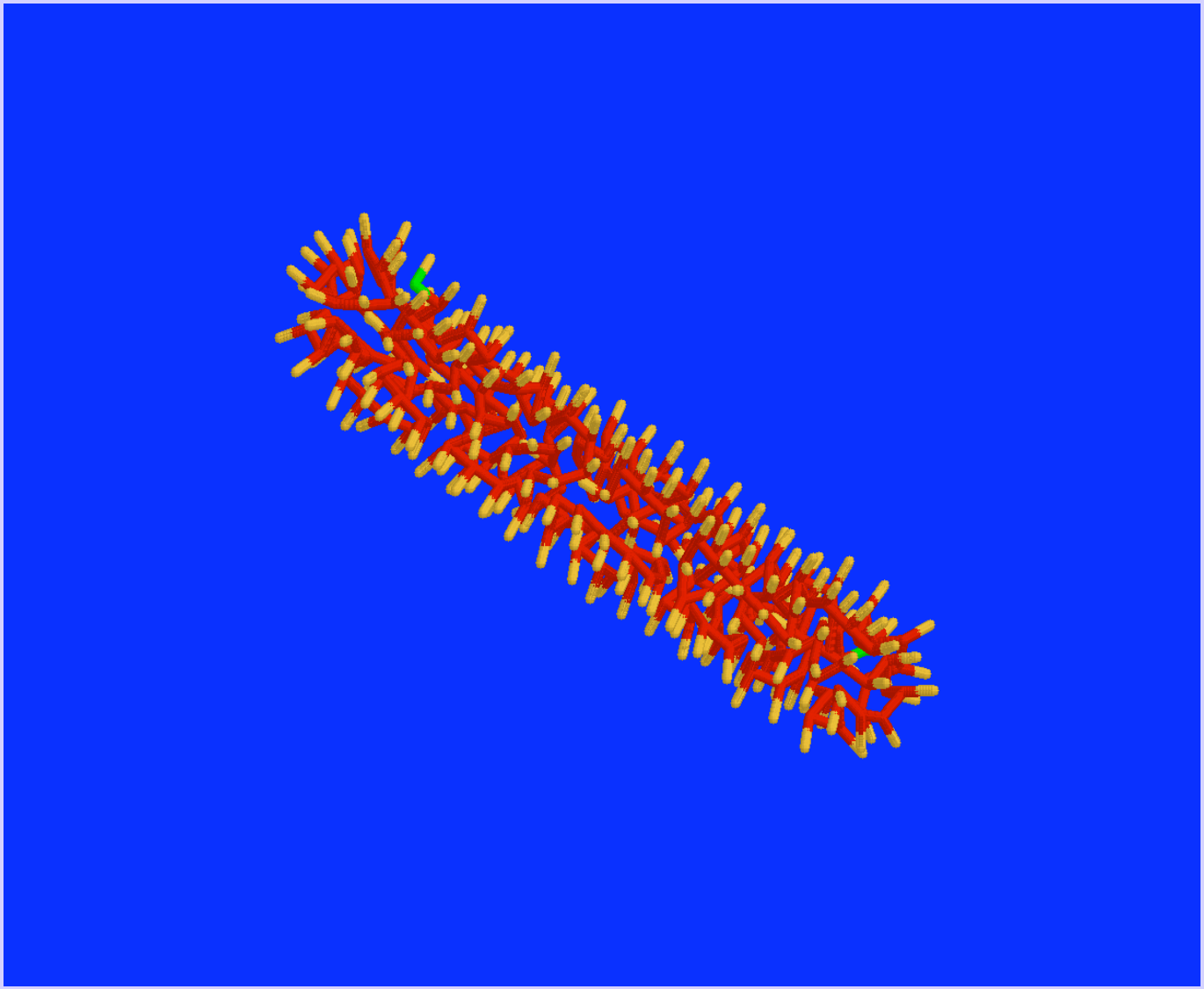
First, let us look at A-homopolymer



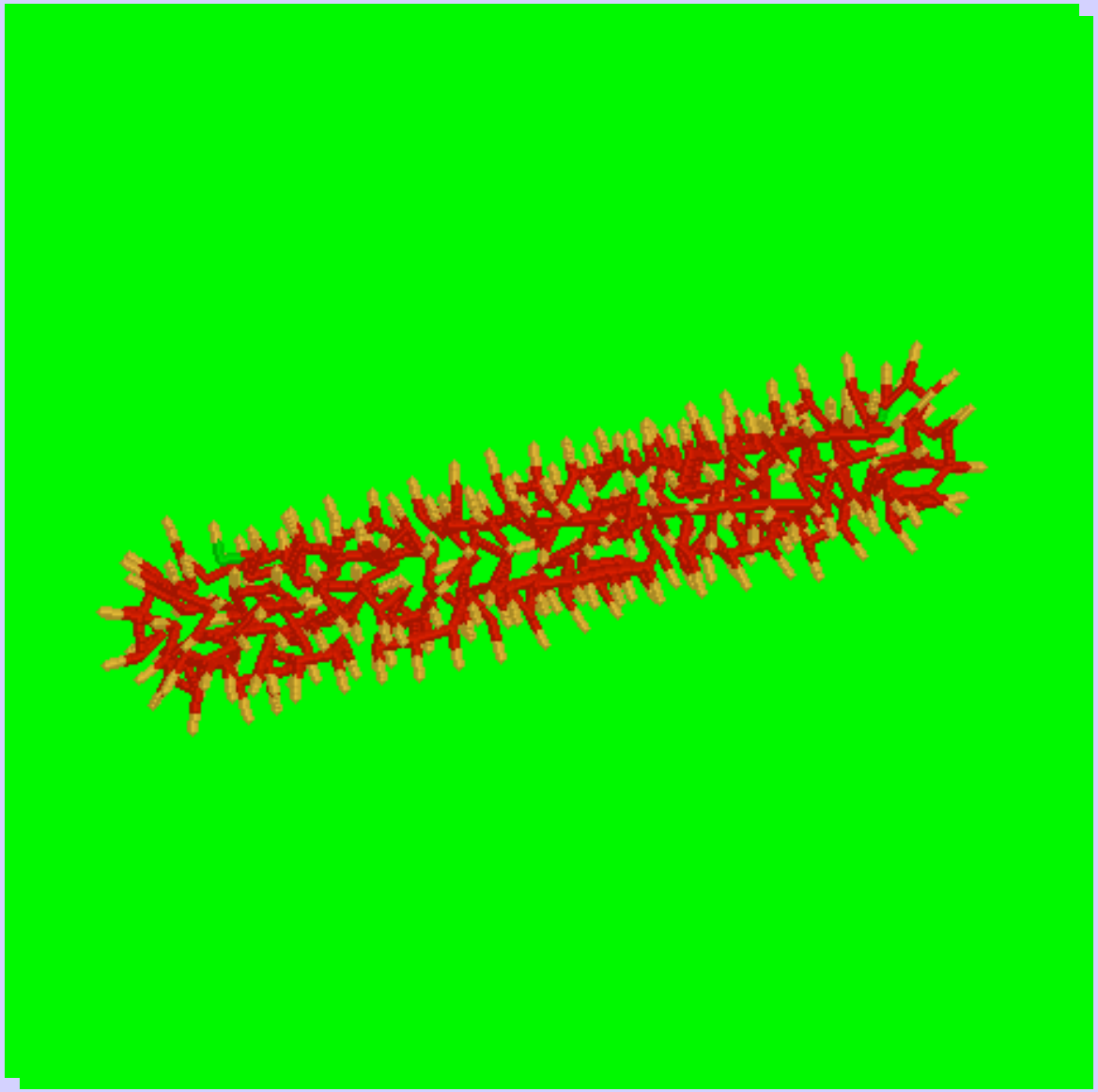
Attracting (hydrophobic) units

Repelling (hydrophilic) units

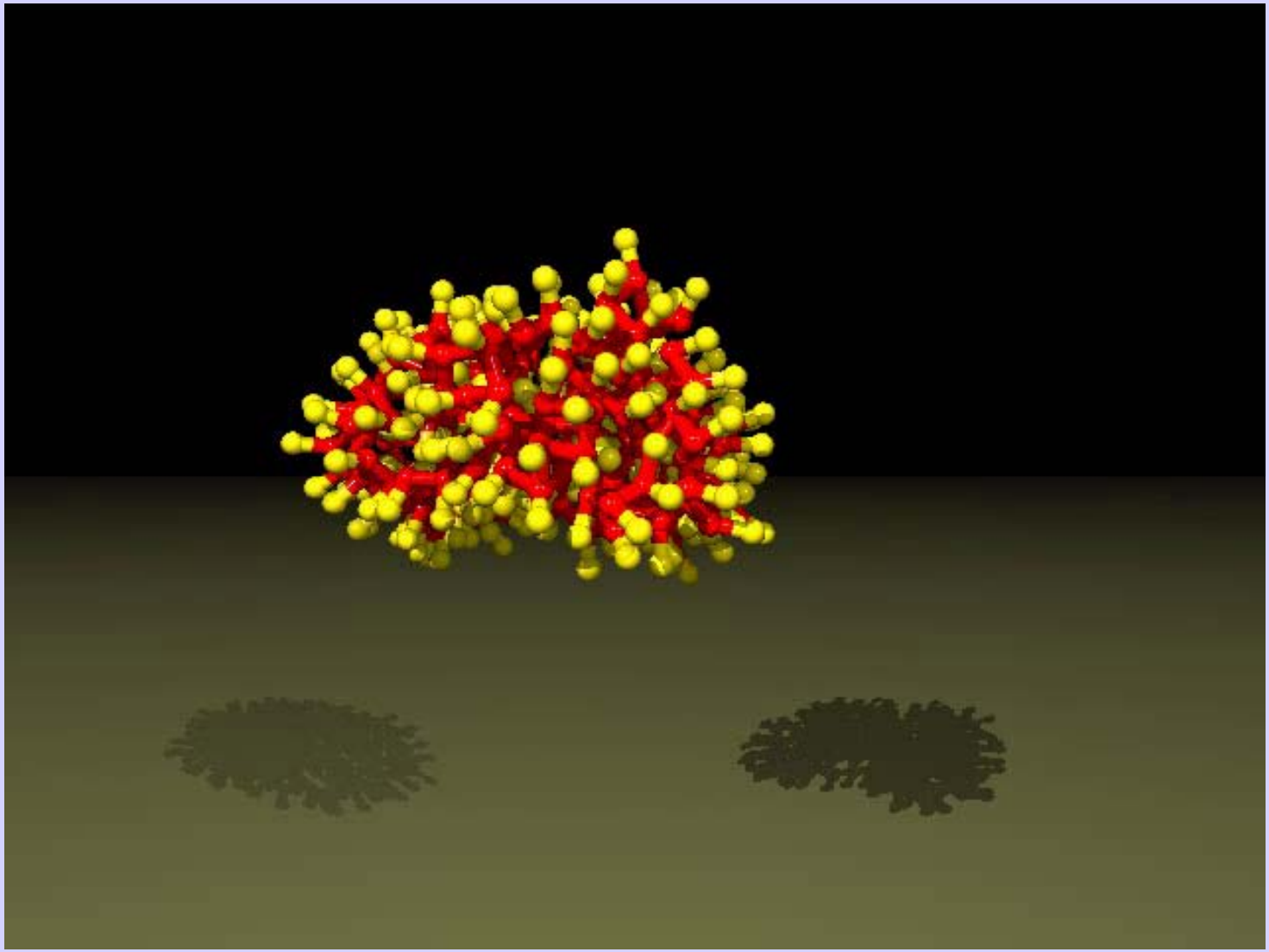
For macromolecules with amphiphilic monomer units a globule assumes cylindrical shape



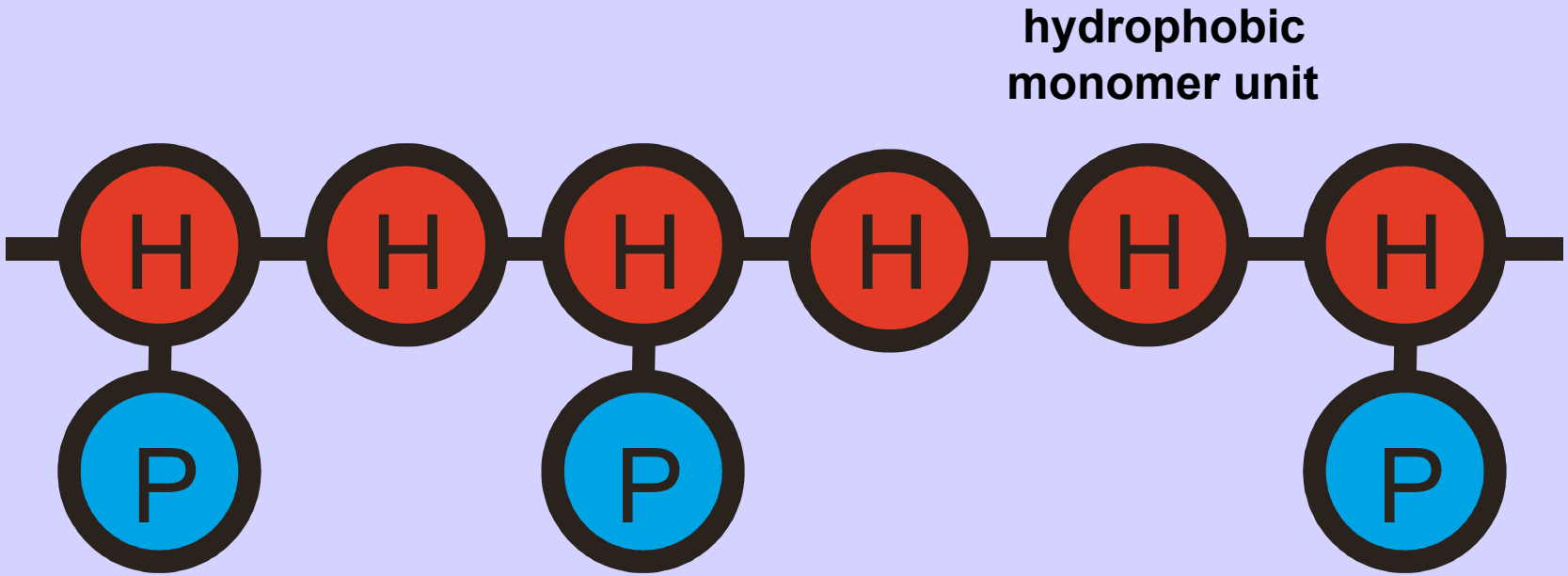
Typical conformations for macromolecules with amphiphilic monomer units at different energies of attractive interaction between hydrophobic groups



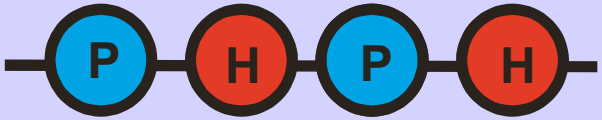
Typical conformations of the macromolecules with amphiphilic monomer units for different energies of attractive interaction between hydrophobic groups



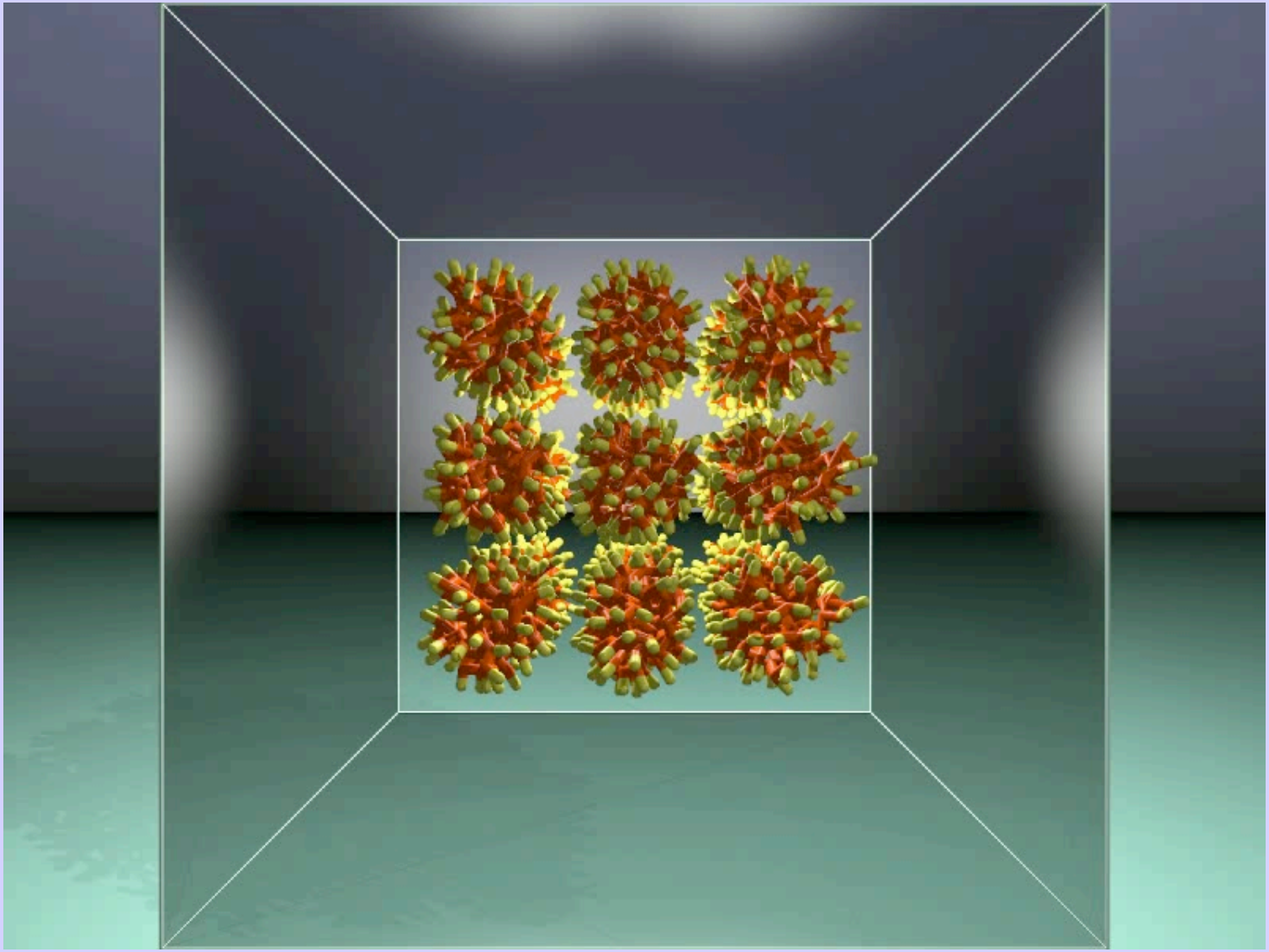
Since it is better to represent "hydrophilic" monomer units as amphiphilic, let us consider conformational properties of **HA** copolymers, instead of **HP** copolymers.

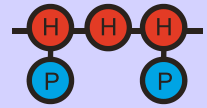


HP model



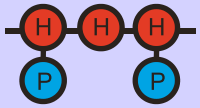
System of m=27 proteinlike HA copolymers does not show tendency to aggregation

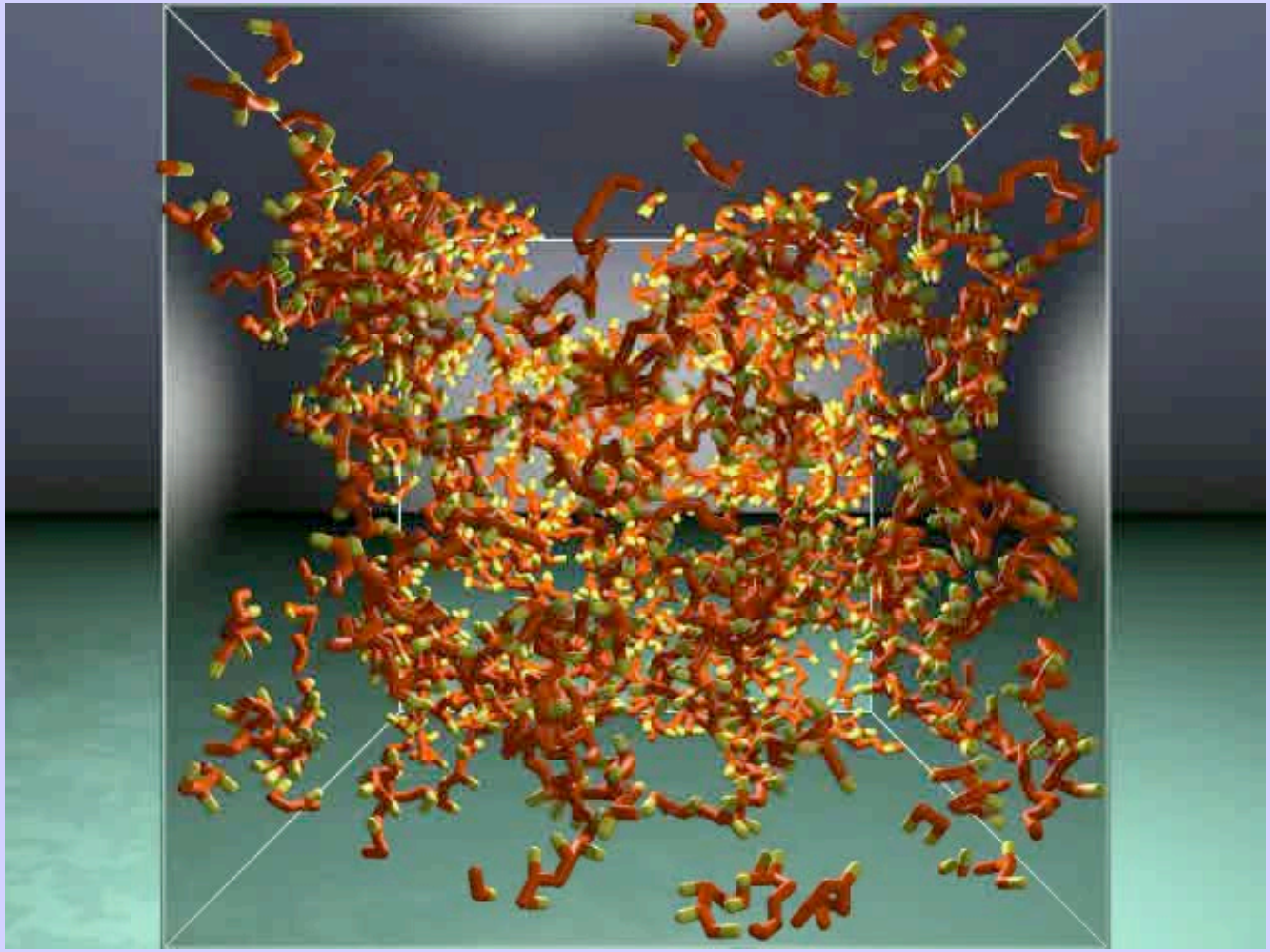




THERMOREVERSIBILITY for protein-like HA copolymer

Heating to T=4 and cooling back to T=1

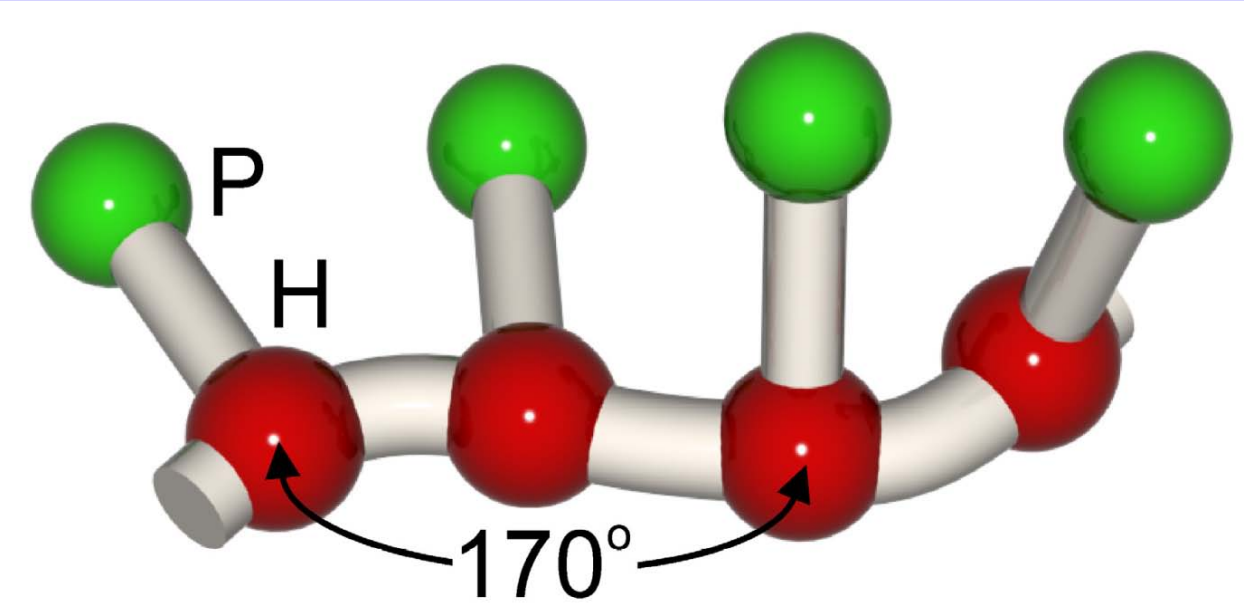




About 20% of globules form aggregates

Stiff amphiphilic macromolecules

$$U = \varepsilon_{st}(\theta - \theta_o)^2$$

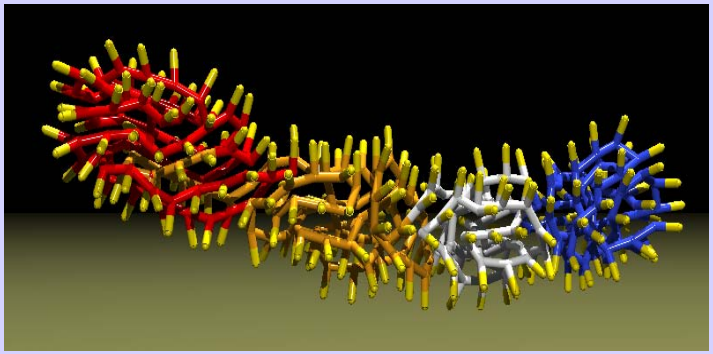


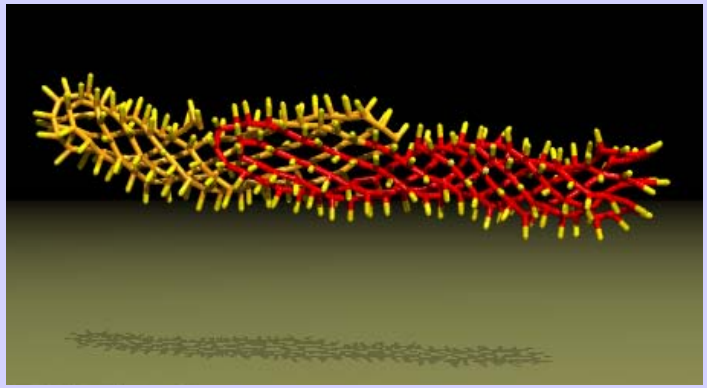
Cylindrical globule has blob-like structure

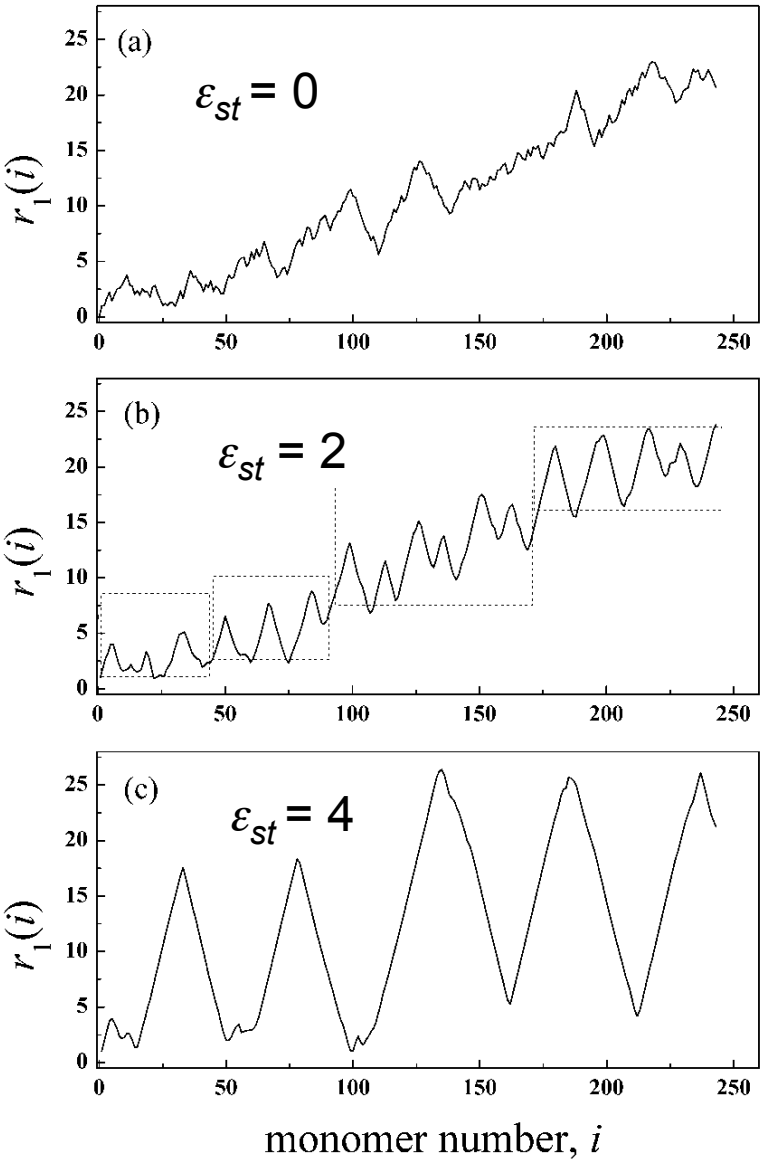
$r_1(i)$ - *distance between first and i-th monomer units*

For coil:

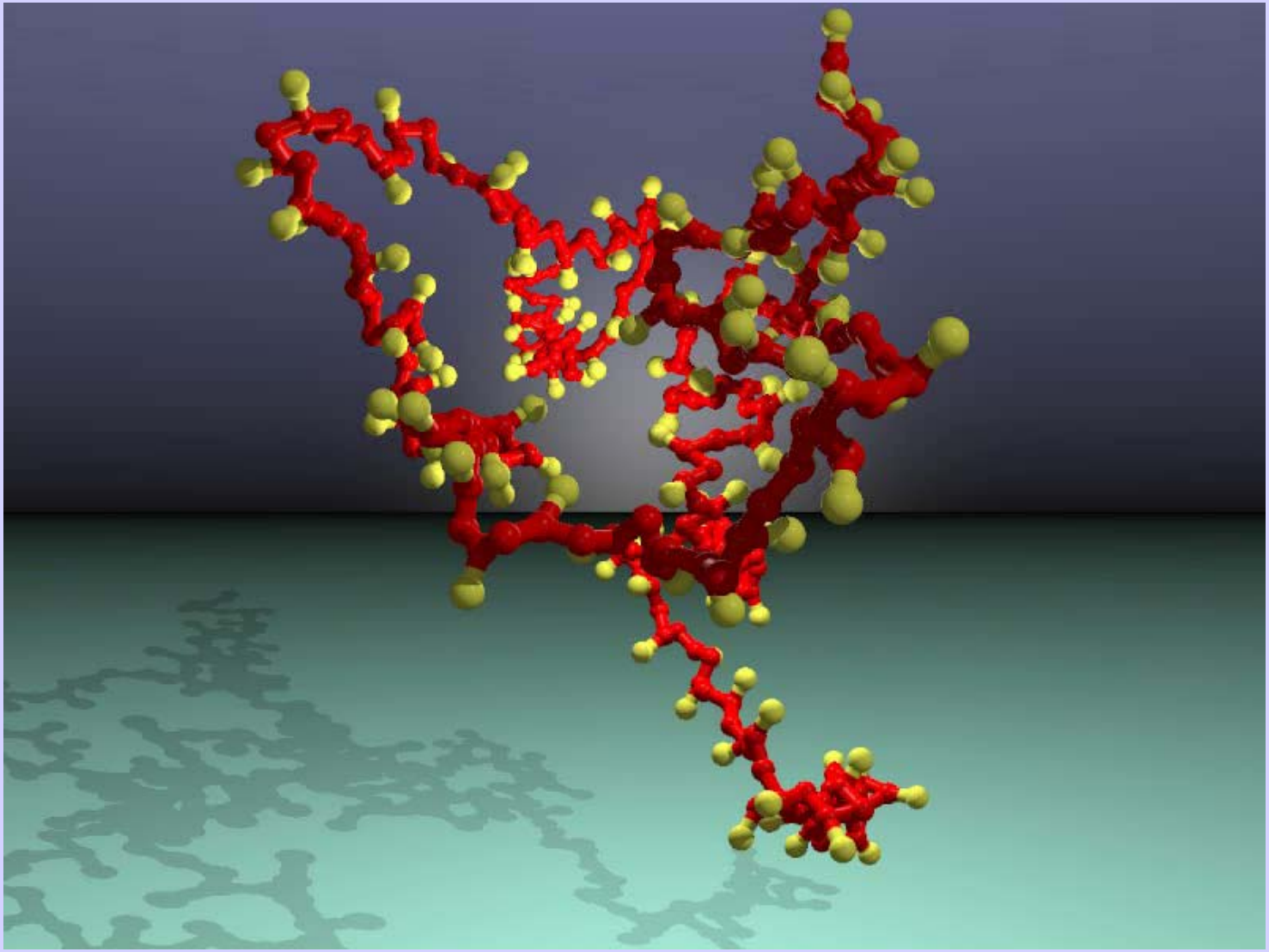
- $r_1(i)$ is strongly fluctuating function;
- $r_1(i) \sim i^{\nu}$; ν = 1/2 or 3/5

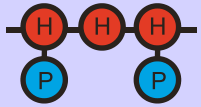




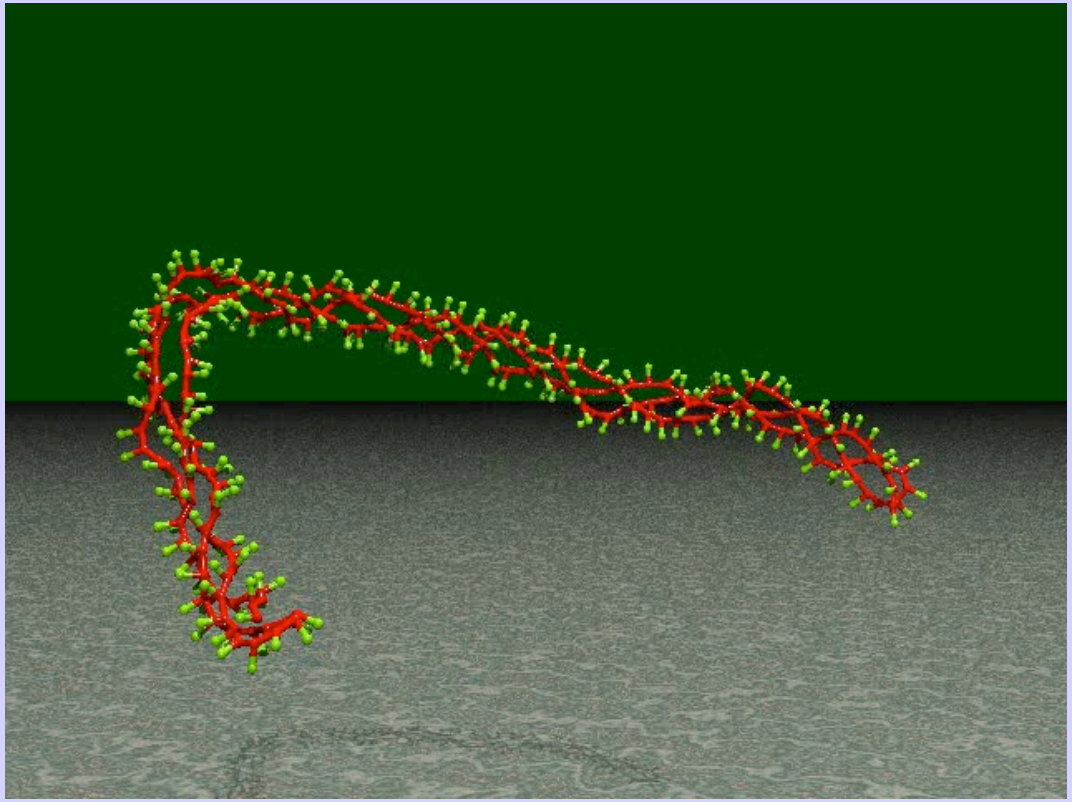


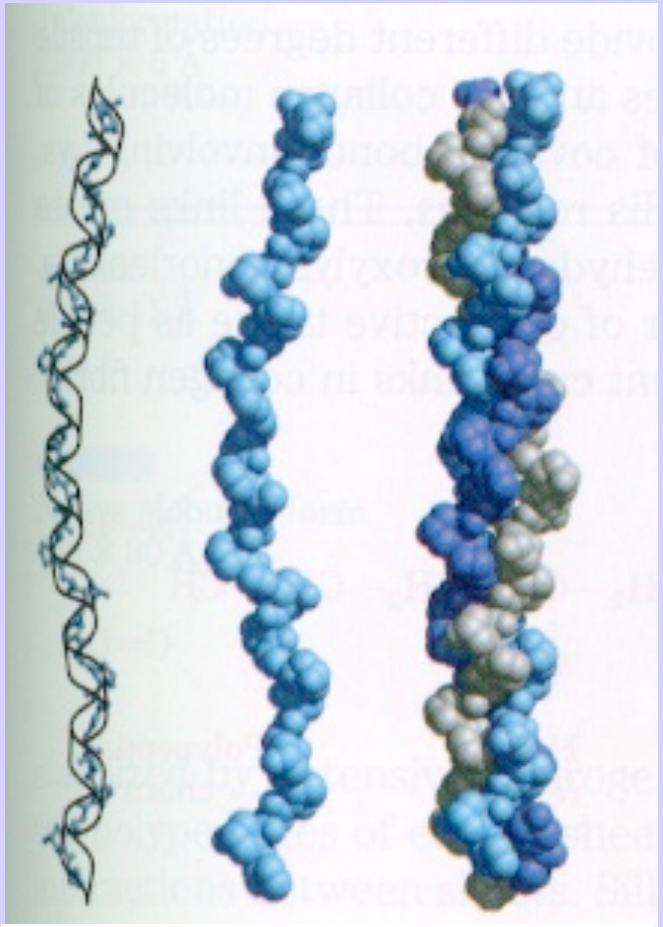
Formation of blob-like structure for regular HA copolymer





Collagen-like globule

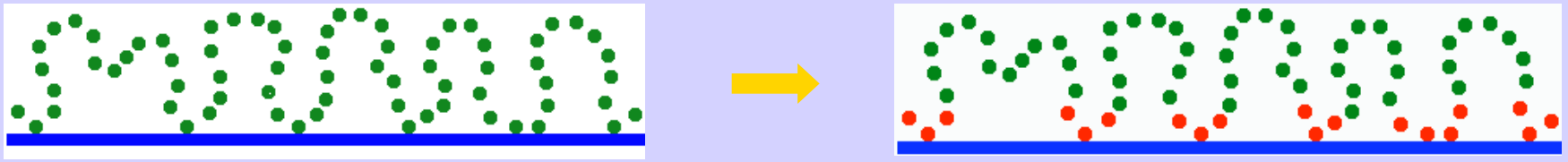


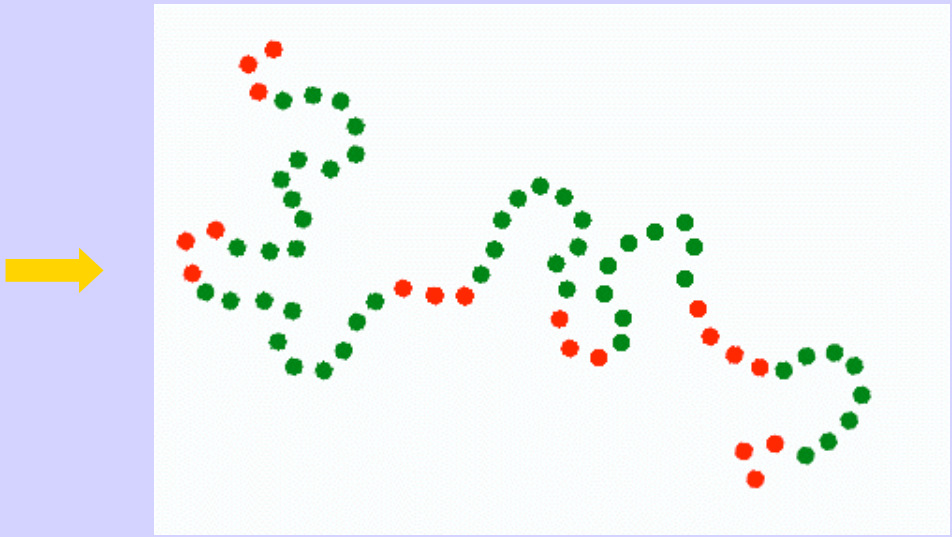


Some generalizations:

⇨**Primary AB sequence with long-range correlations can be obtained not only starting from globular conformations.**
Any specific conformation can play role of “parent conformation”.

❶ Chain adsorbed on the plane surface:

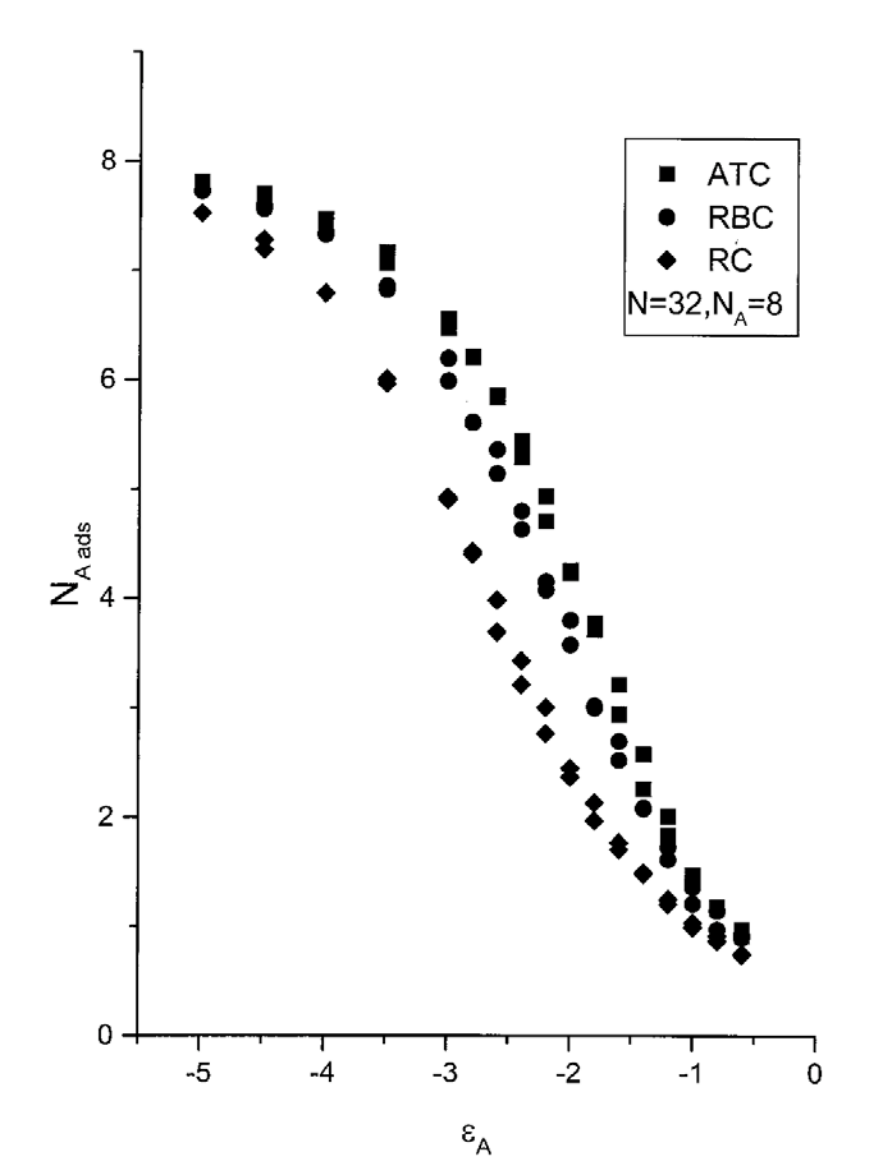




AB-copolymer
"tuned to adsorption"

For such AB-copolymer the critical energy of attraction of "red" units to the surface inducing the adsorption is smaller than for the random AB-copolymer with the same composition.

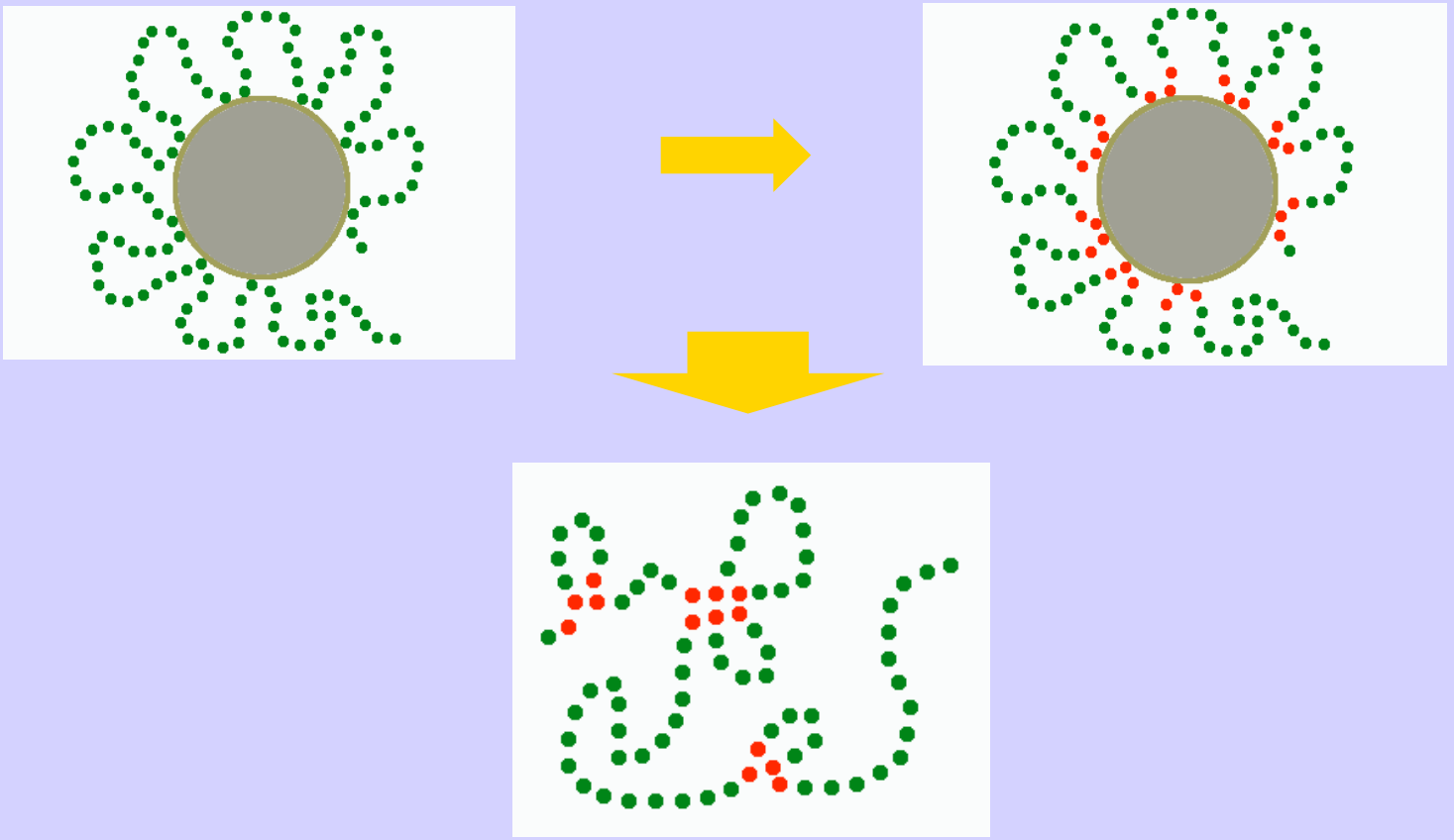
Adsorbtion-tuned copolymers



The number of adsorbed segments (at a given value of e_A) is always the highest for designed AB-copolymers

Molecular Dispensers

Chain adsorbed on the spherical colloidal particle.

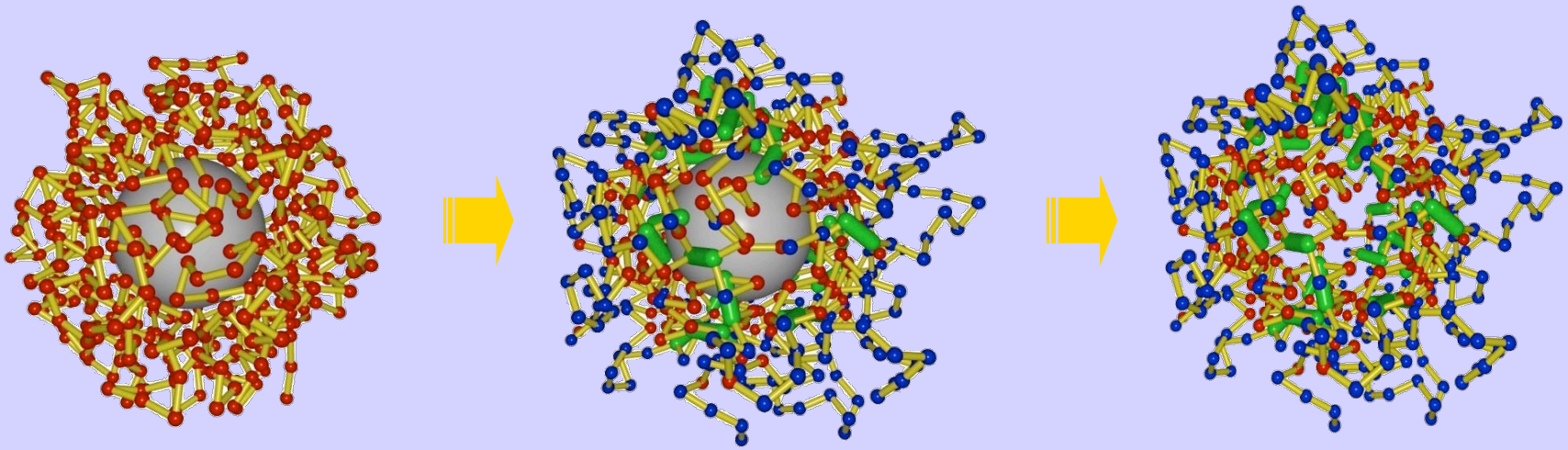


AB -copolymer “tuned to the adsorption of hydrophobic droplet of spherical size” - “molecular dispenser”

Molecular Dispensers

Envelope Preparation Scheme

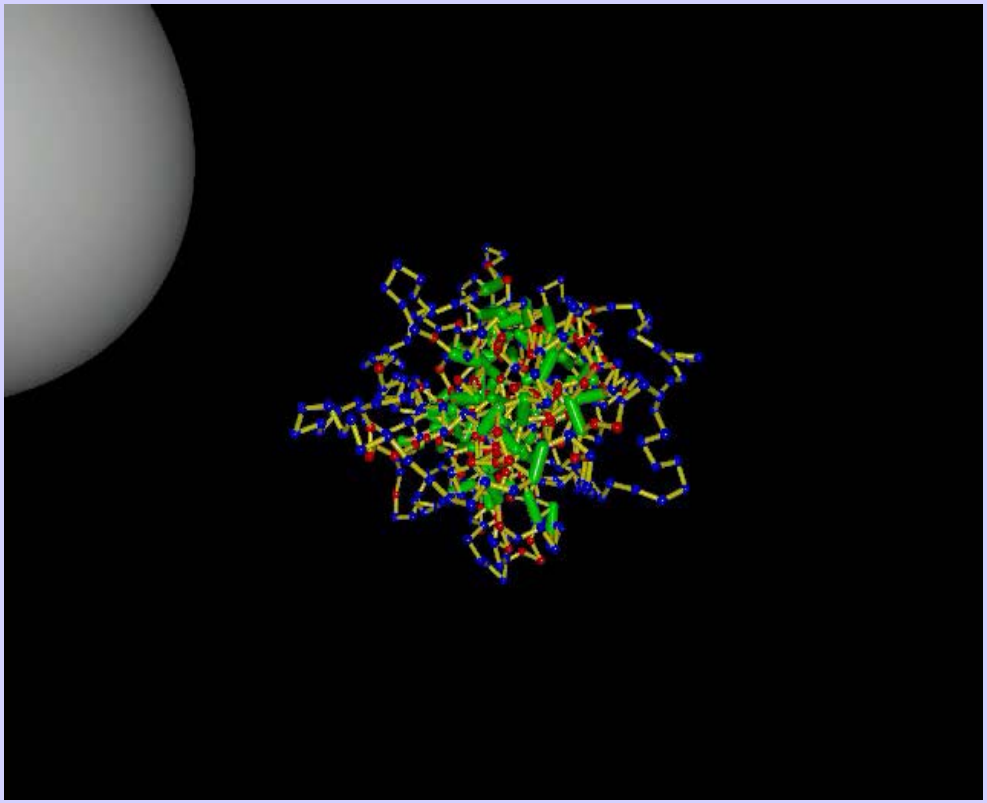
- Polymer chain adsorbs onto the particle
- Coloring of polymer chain
- Introduction of junctions between hydrophilic or hydrophobic monomer units.
- Particle removal



Molecular Dispensers

Copolymer-Particle Interaction

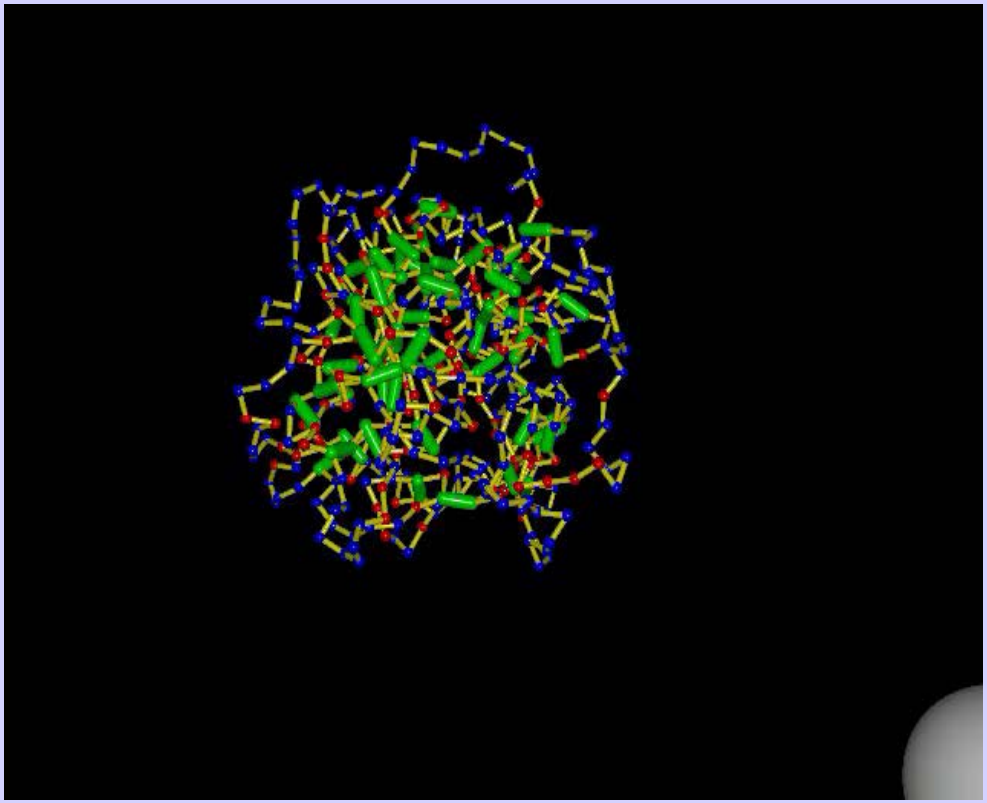
(particle's size = parental particle size)



Molecular Dispensers

Copolymer-Particle Interaction

(particle's size > parental particle size)



Selectivity factor

Probability to form complex with the particle of a given size