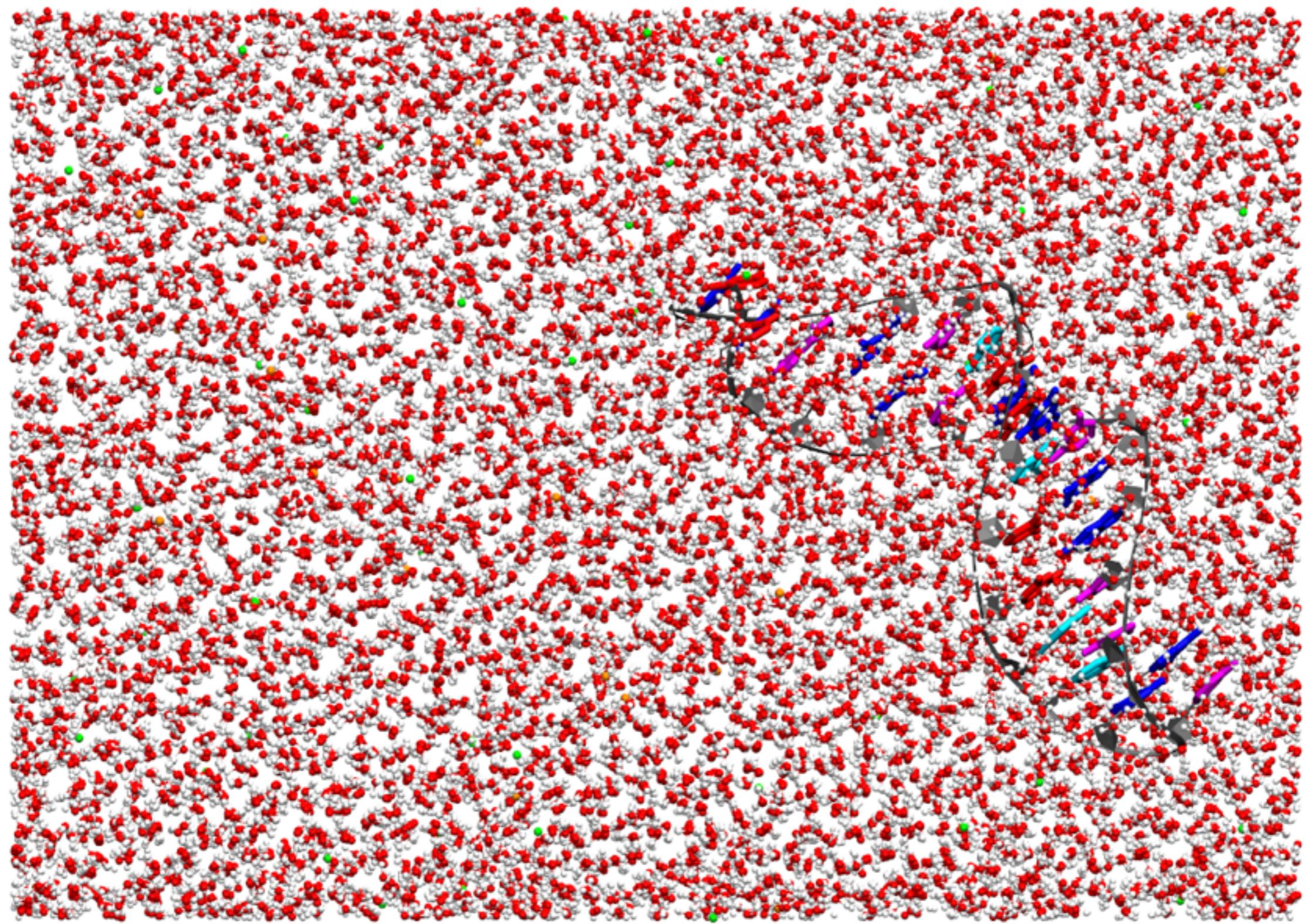
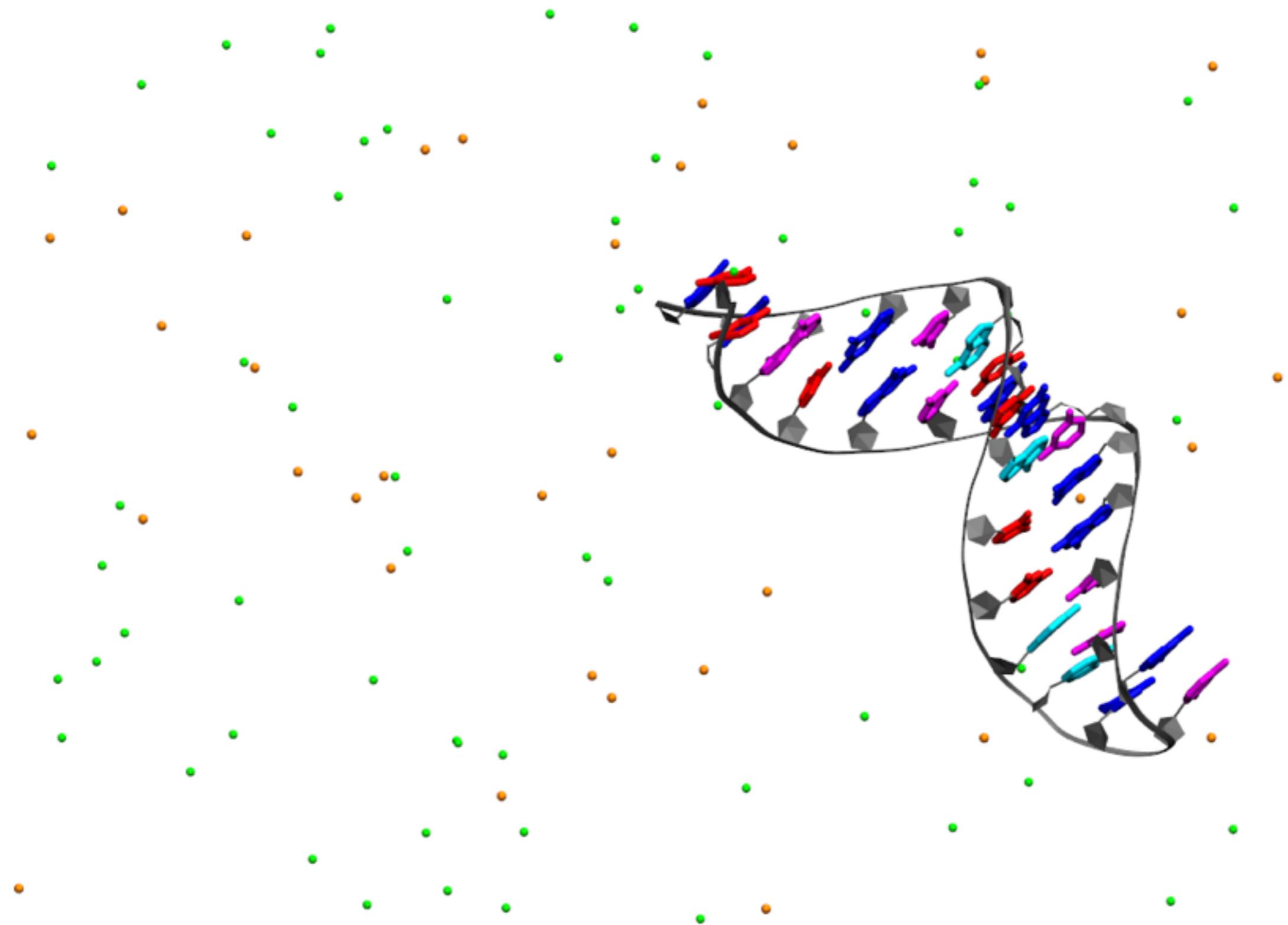


MHPC

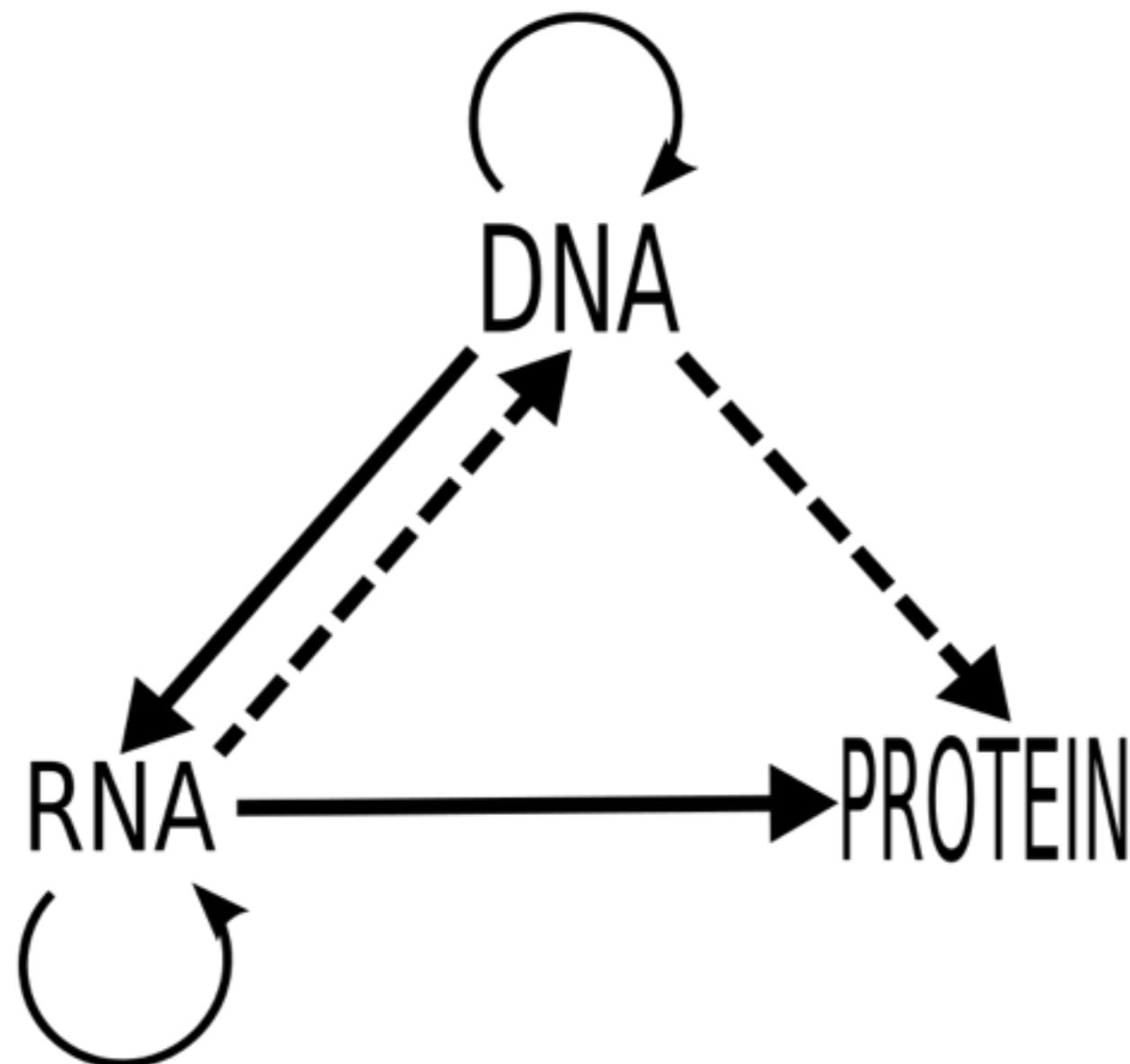
Molecular dynamics

Sabine Reißer



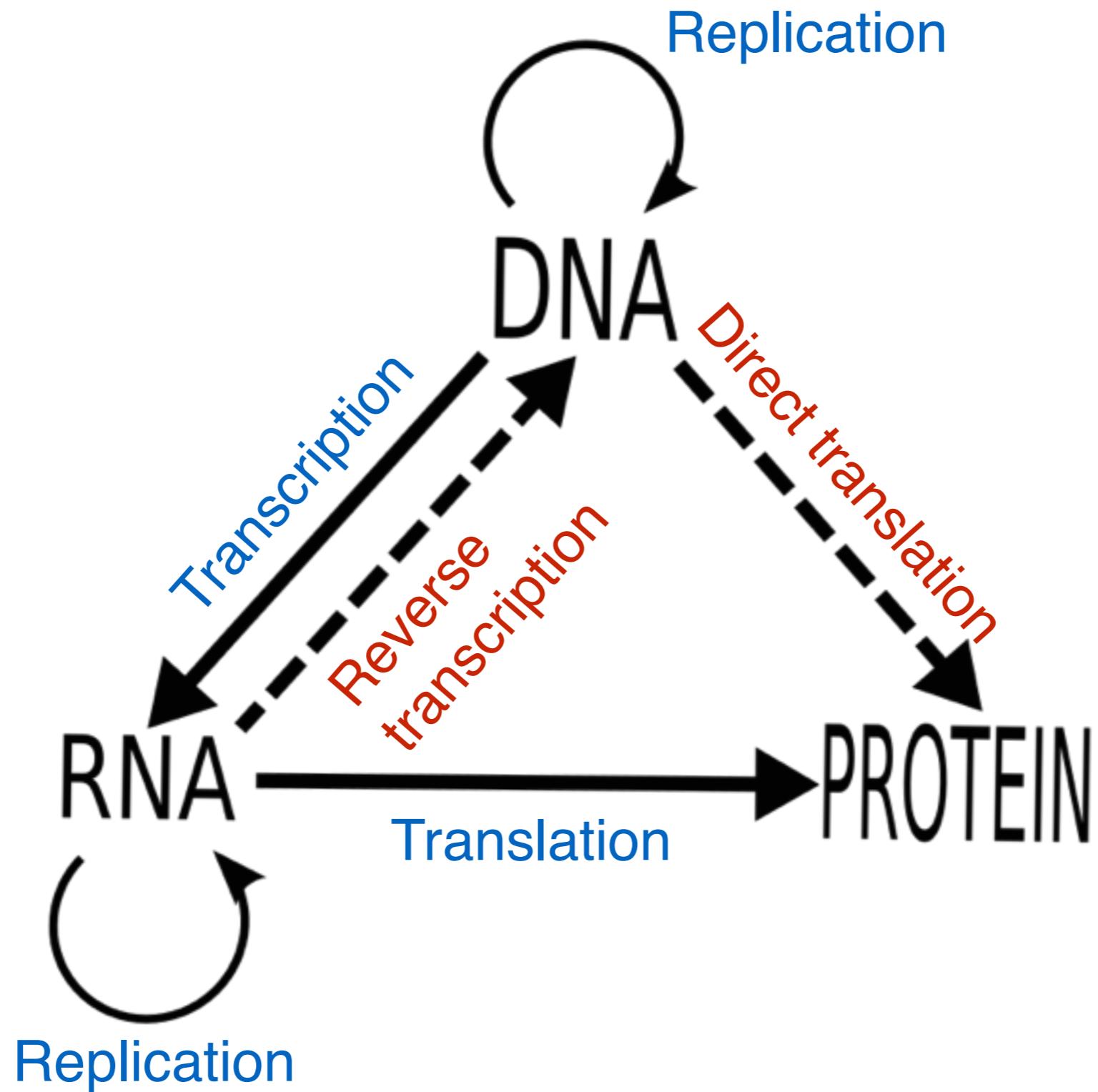


Central dogma of molecular biology



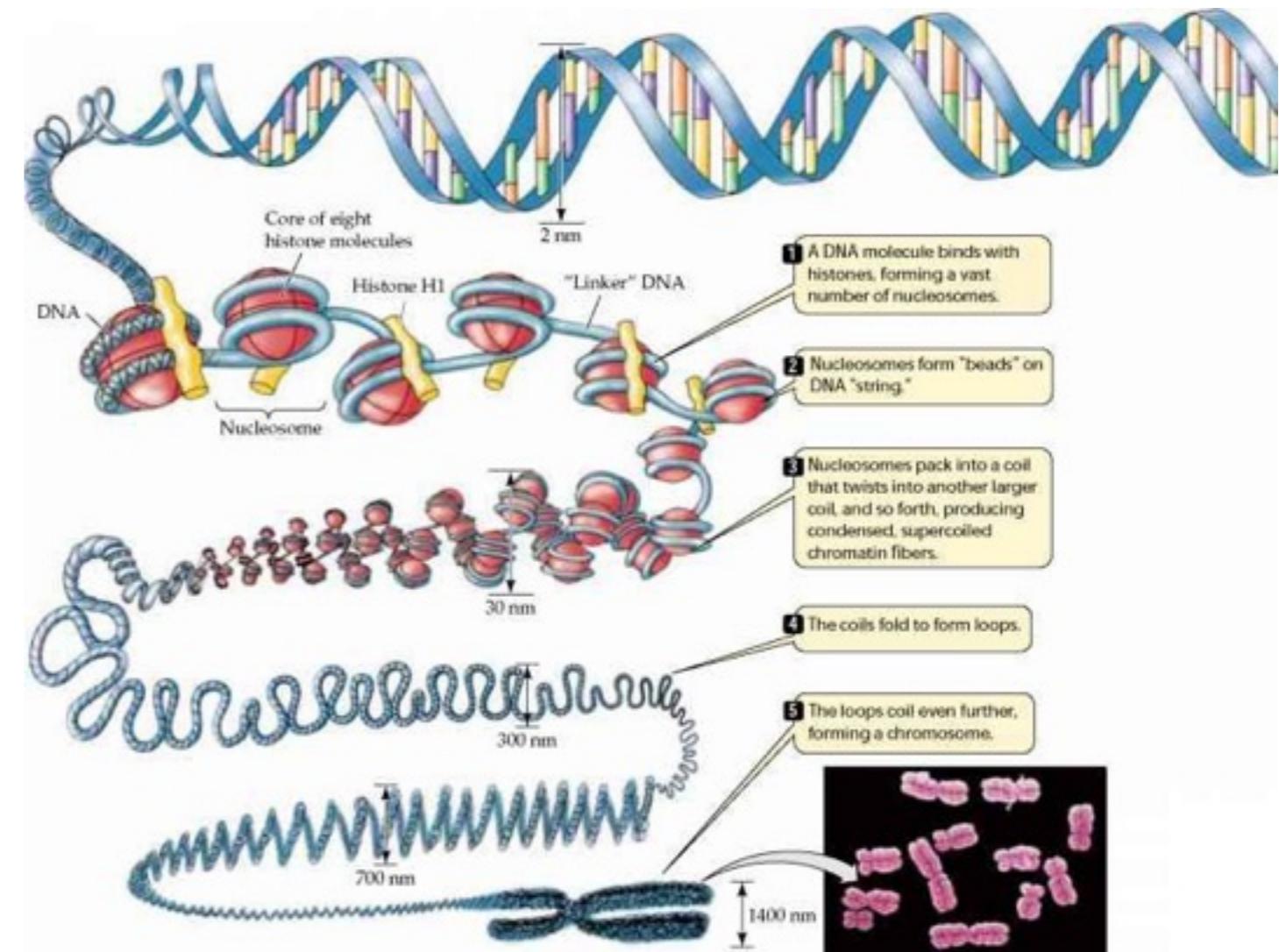
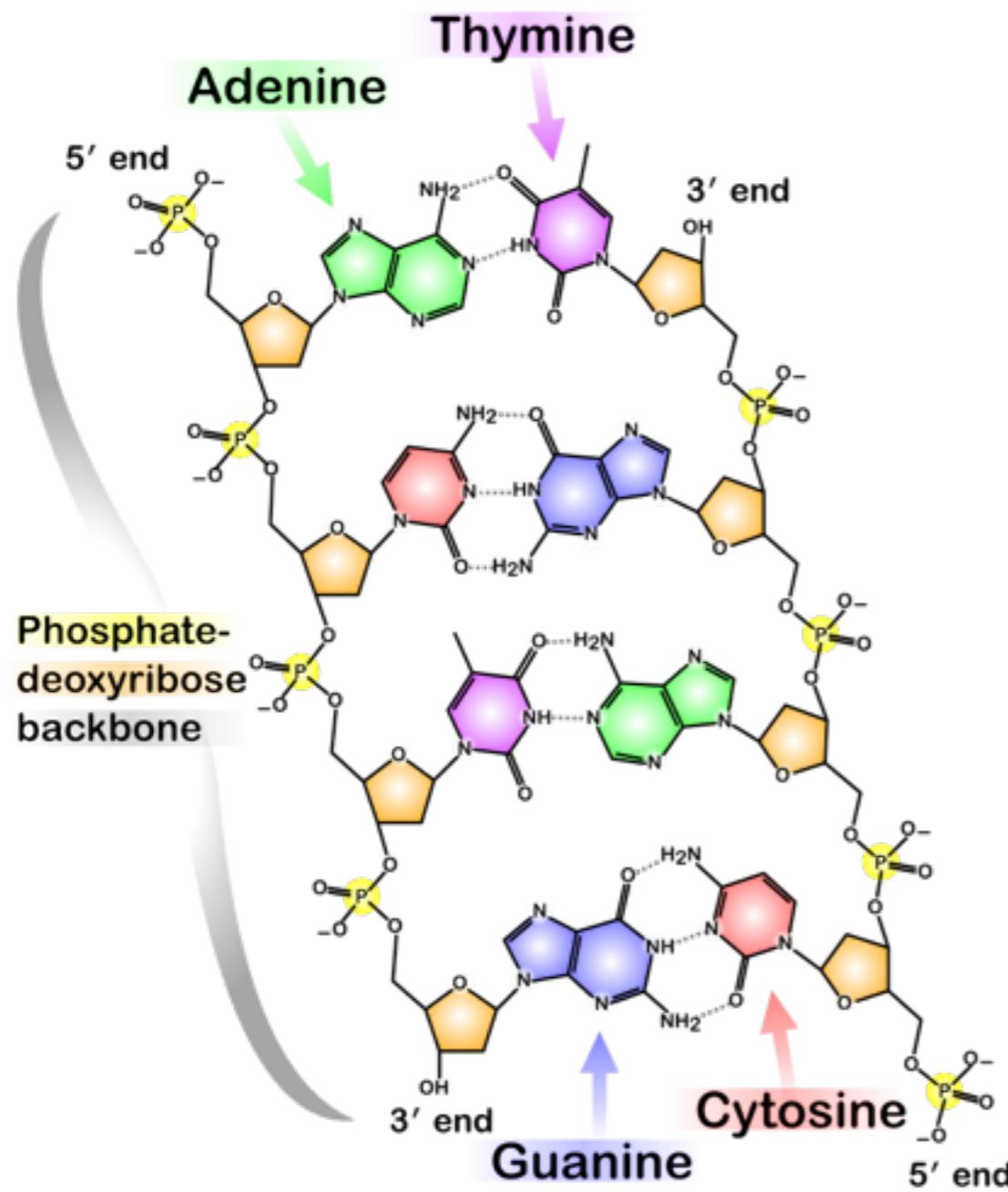
*Crick 1958

Central dogma of molecular biology



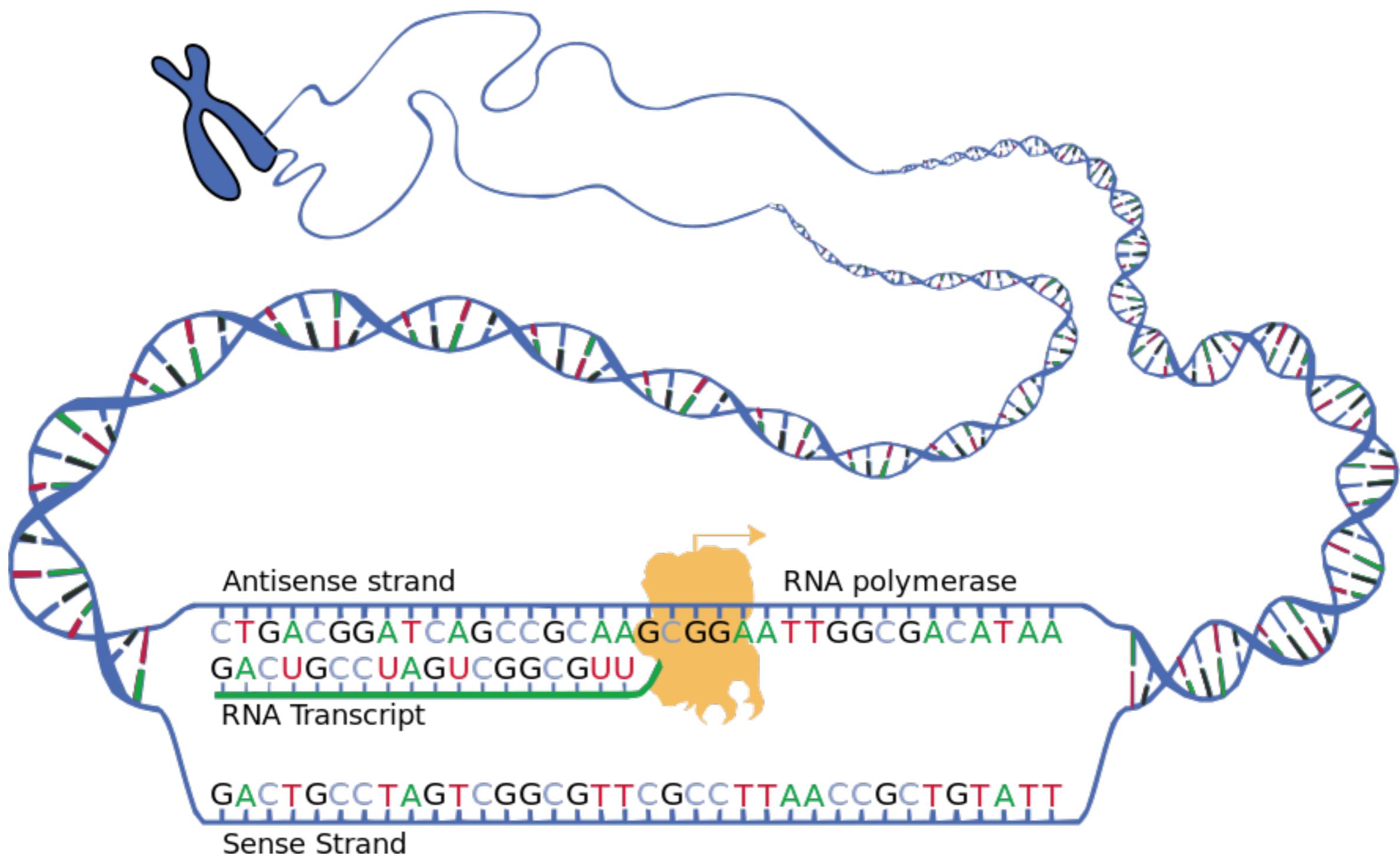
*Crick 1958

DNA

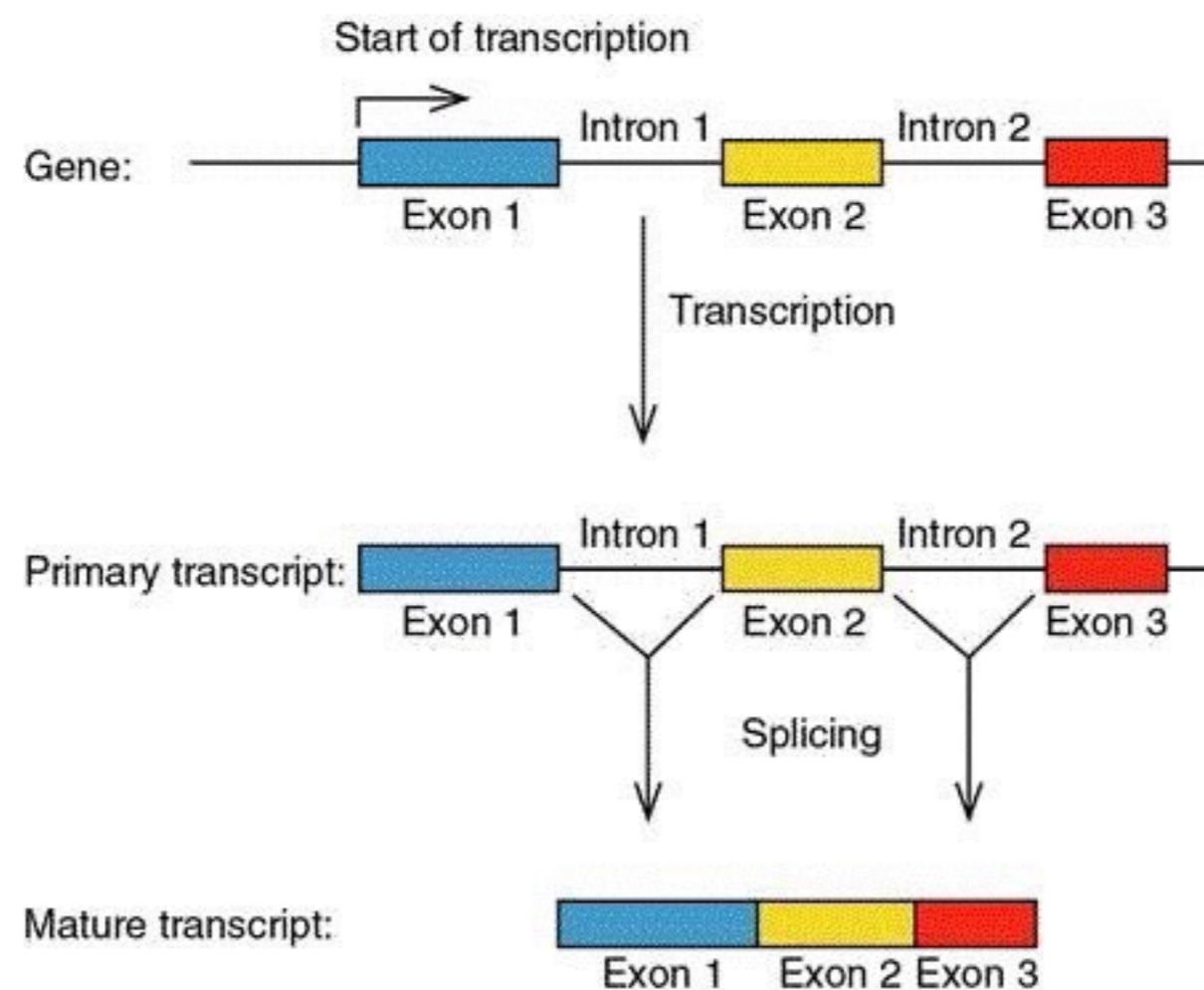


Amoeba: 670 000 Mbp
Homo sapiens: 3 080 Mbp
Bacteria: 0.13 - 14 Mbp

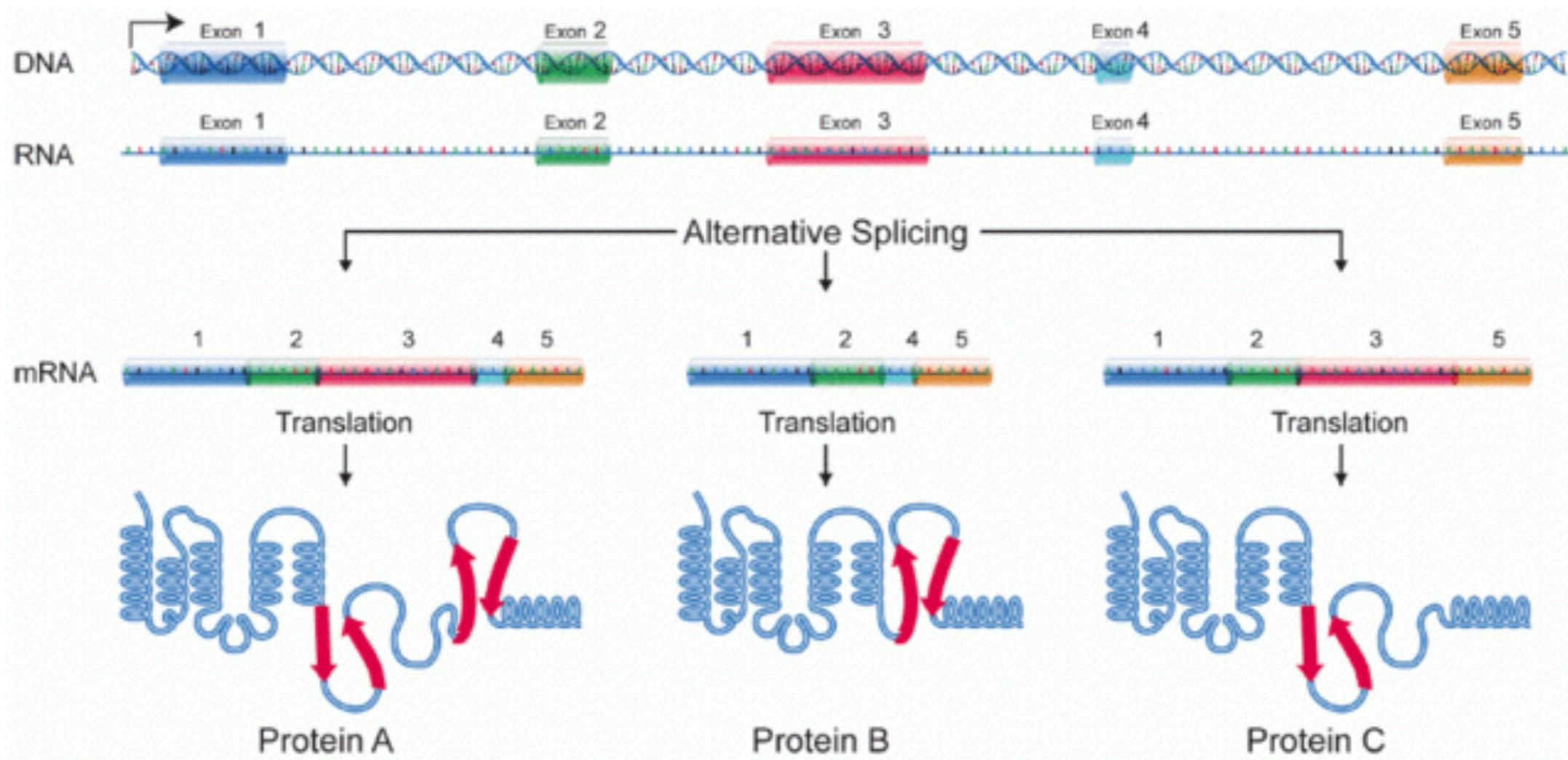
DNA transcription to RNA



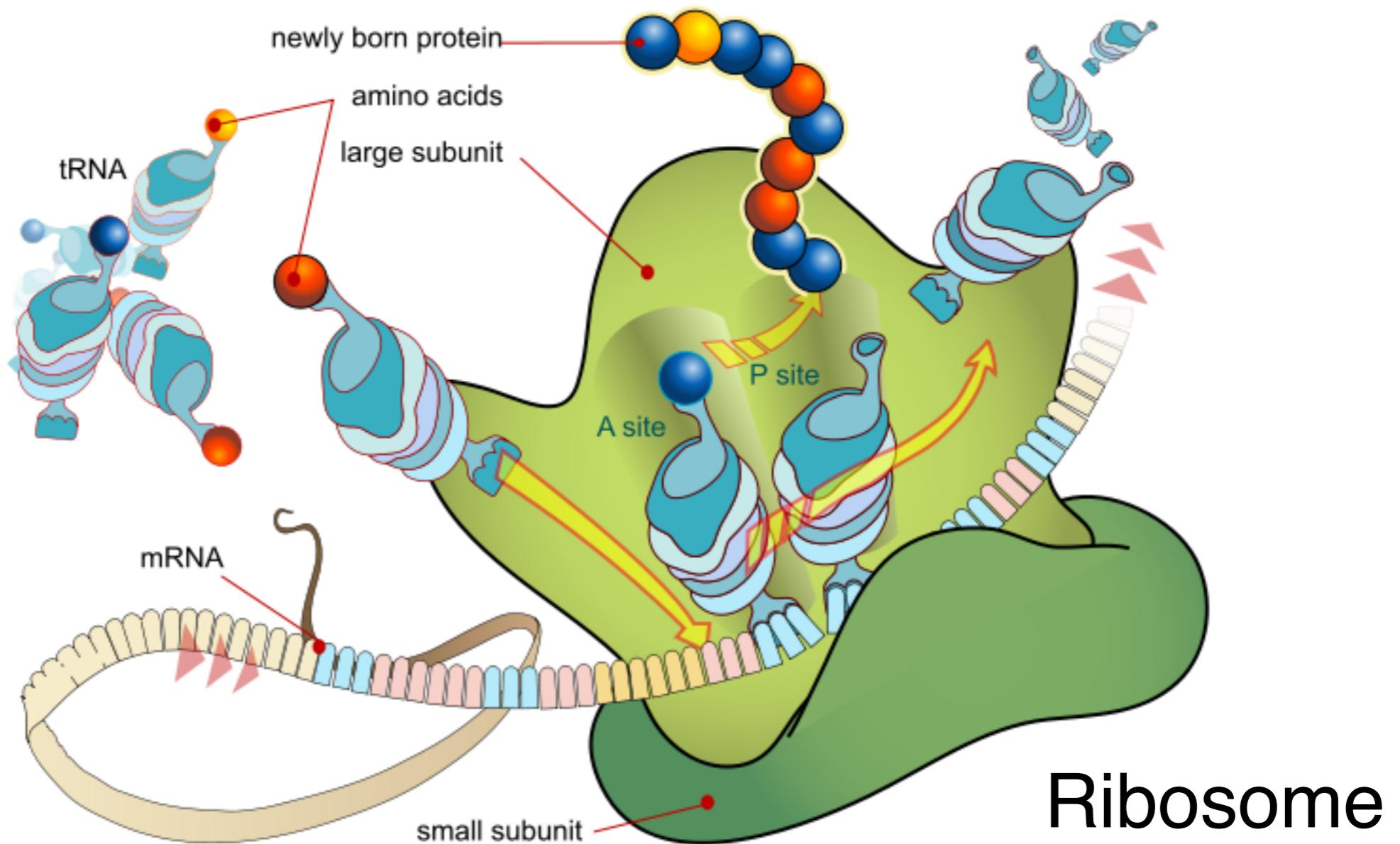
RNA splicing



Alternative Splicing



Translation



Translation code

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U	C
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA }	U	C
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U	C
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA }	U	C
Third letter							

Protein-coding genes

Organism	# of protein-coding genes	# of genes naïve estimate: (genome size /1000)	BNID
HIV 1	9	10	105769
<i>Influenza A virus</i>	10-11	14	105767
Bacteriophage λ	66	49	105770
Epstein Barr virus	80	170	103246
<i>Buchnera sp.</i>	610	640	105757
<i>T. maritima</i>	1,900	1,900	105766
<i>S. aureus</i>	2,700	2,900	105500
<i>V. cholerae</i>	3,900	4,000	105760
<i>B. subtilis</i>	4,400	4,200	111448
<i>E. coli</i>	4,300	4,600	105443
<i>S. cerevisiae</i>	6,600	12,000	105444
<i>C. elegans</i>	20,000	100,000	101364
<i>A. thaliana</i>	27,000	140,000	111380
<i>D. melanogaster</i>	14,000	140,000	111379
<i>F. rubripes</i>	19,000	400,000	111375
<i>Z. mays</i>	33,000	2,300,000	110565
<i>M. musculus</i>	20,000	2,800,000	100308
<i>H. sapiens</i>	21,000	3,200,000	100399, 111378
<i>T. aestivum</i> (hexaploid)	95,000	16,800,000	105448, 102713

1 mm



20%



<1%

Nucleic acids - RNA

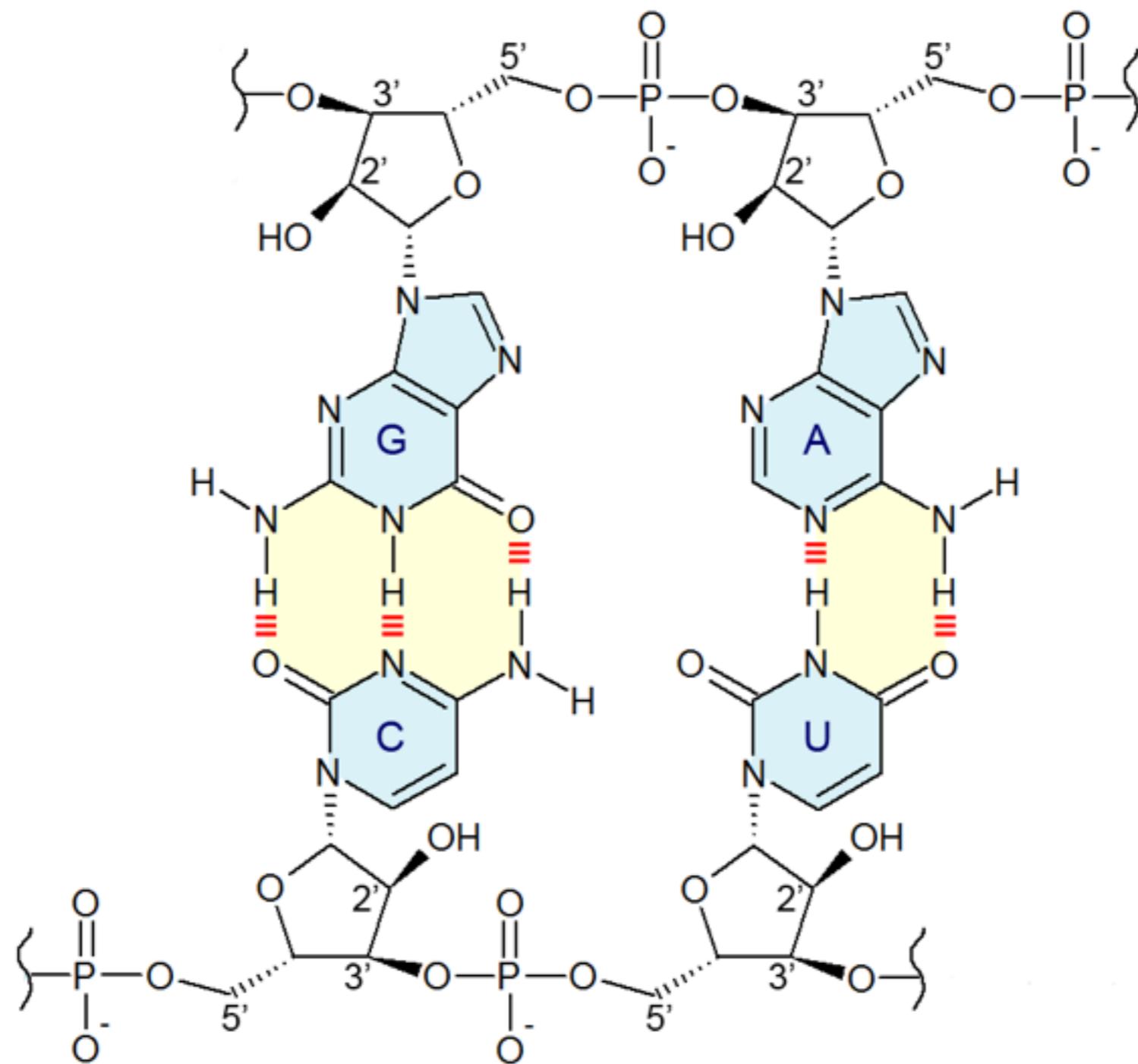
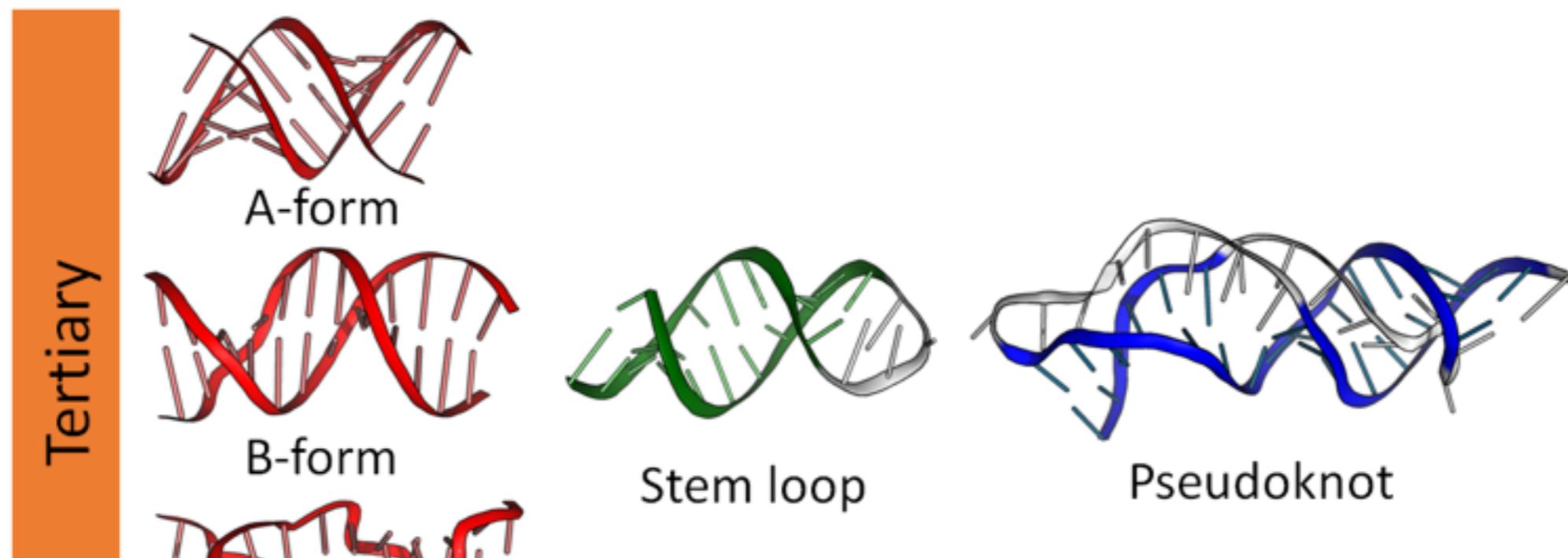
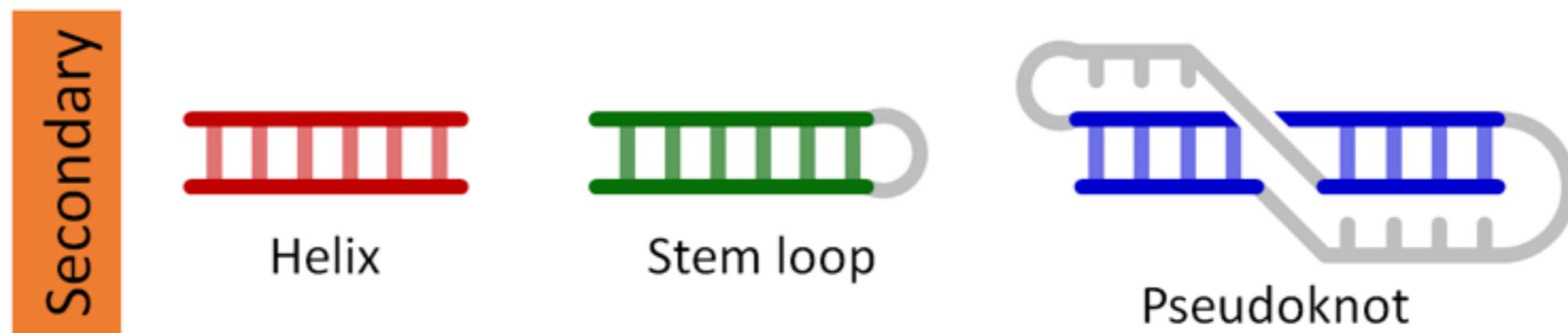
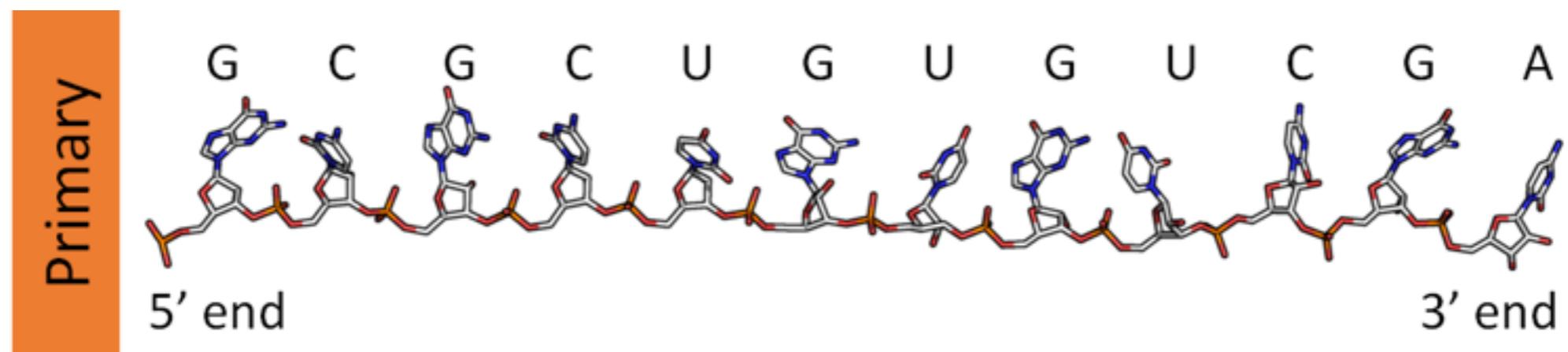


Figure: [wikipedia.org](https://en.wikipedia.org)

RNA structure



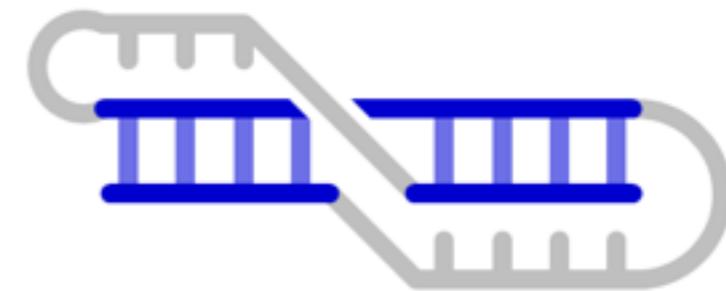
Secondary



Helix

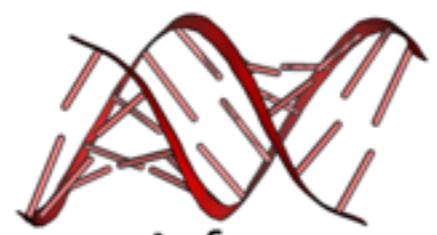


Stem loop

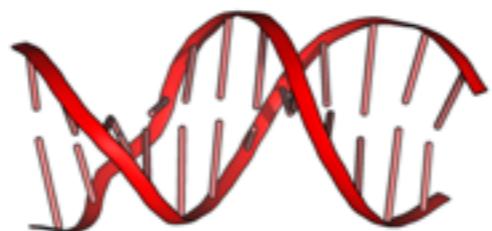


Pseudoknot

Tertiary



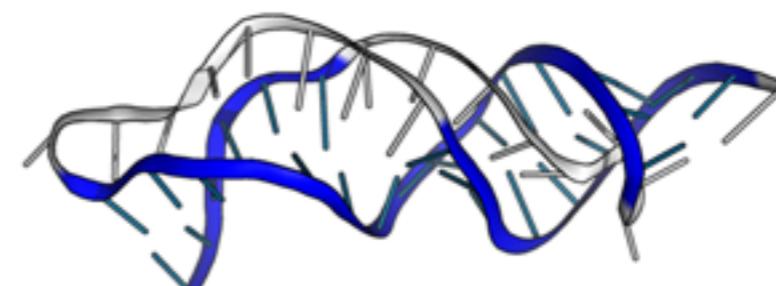
A-form



B-form



Stem loop

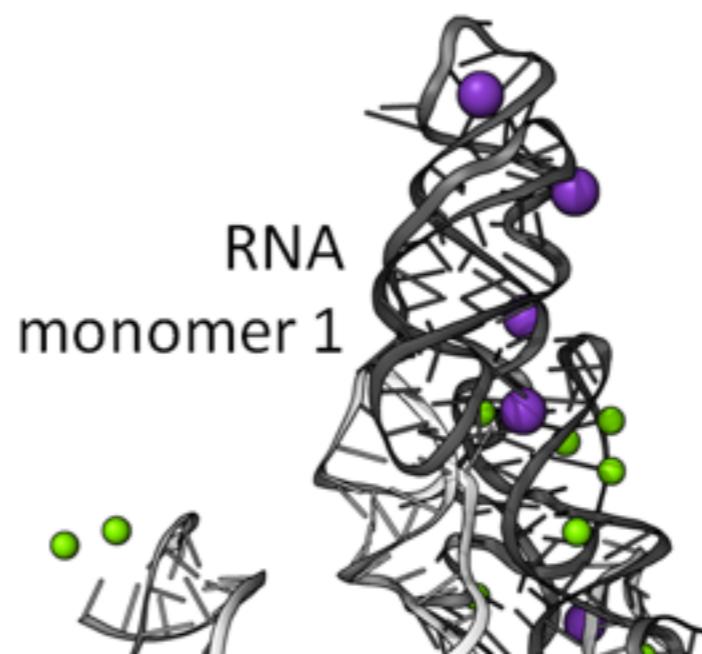
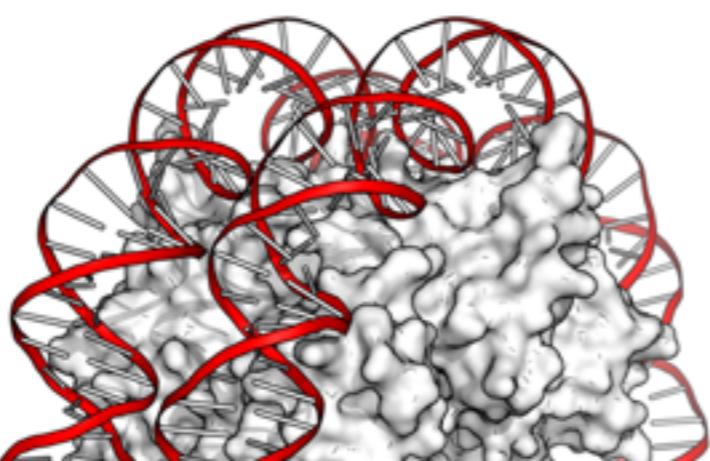


Pseudoknot

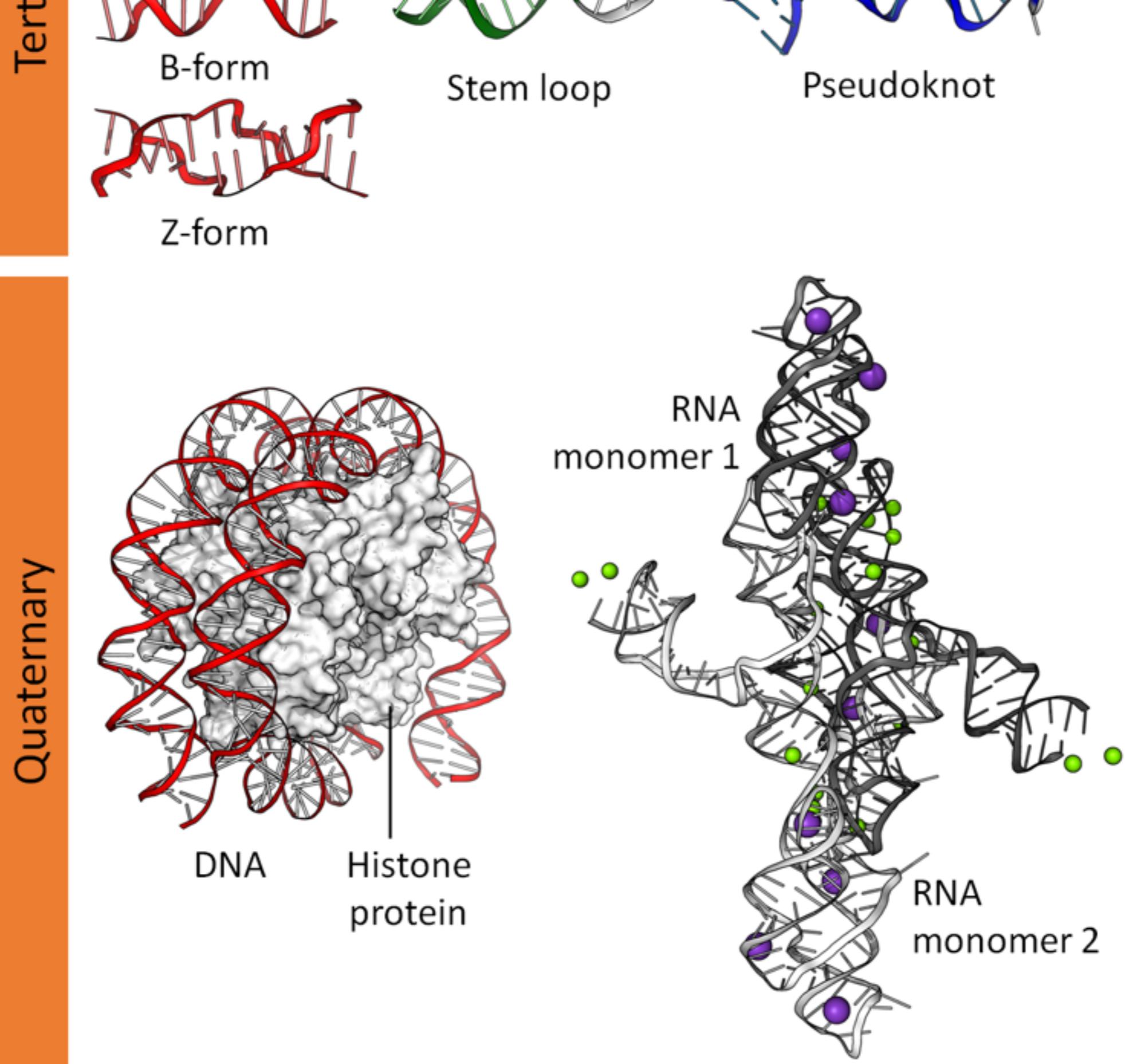


Z-form

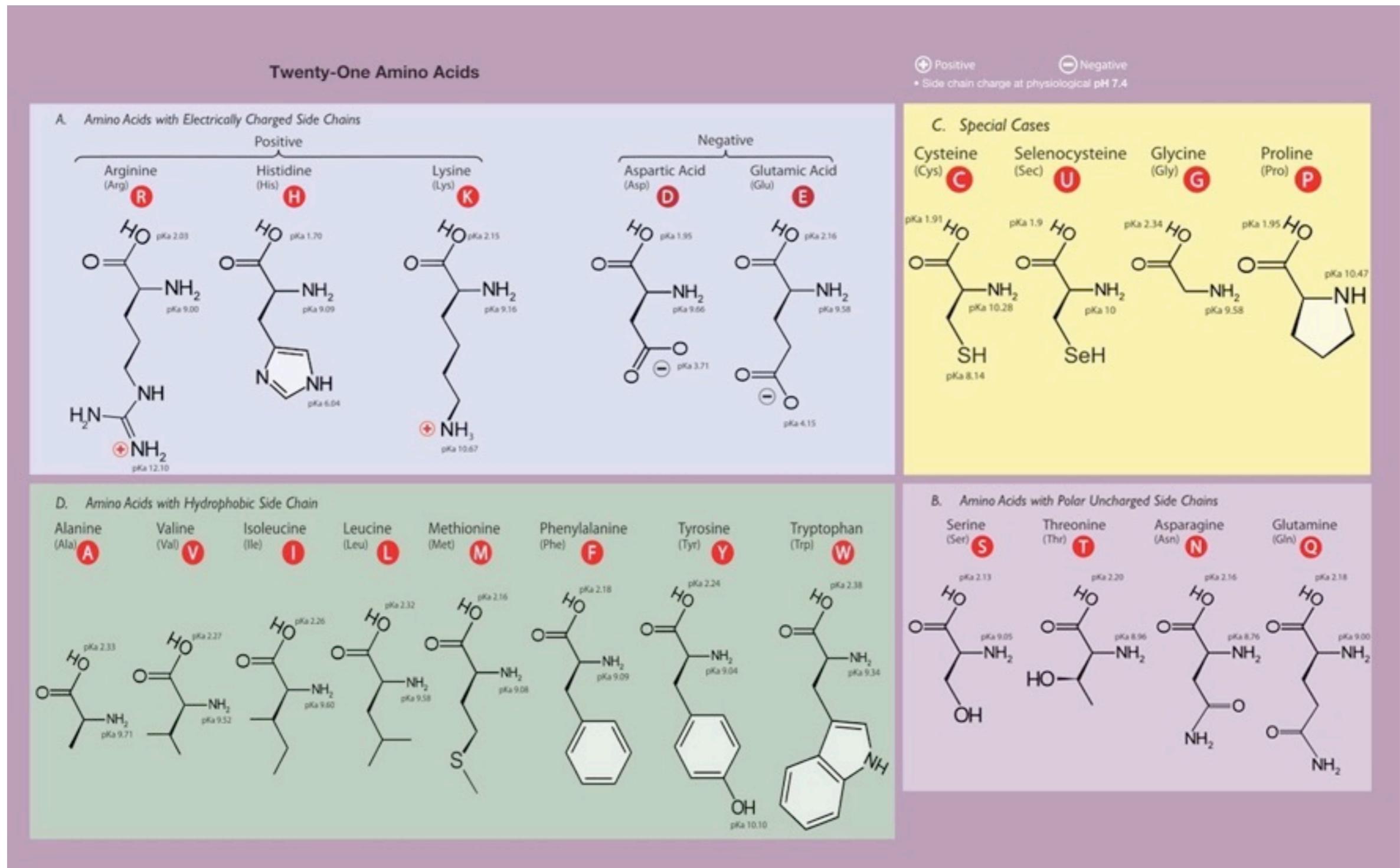
Tertiary



RNA
monomer 1

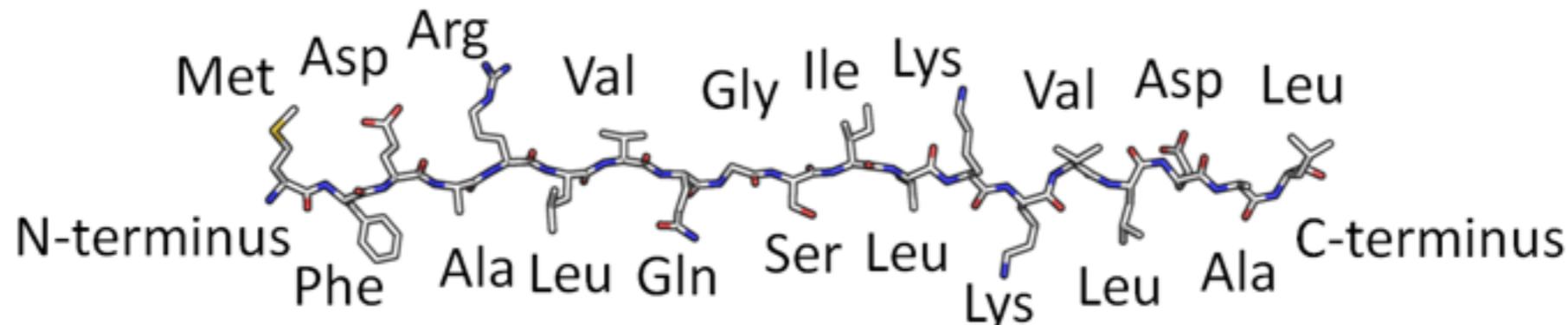


Amino acids

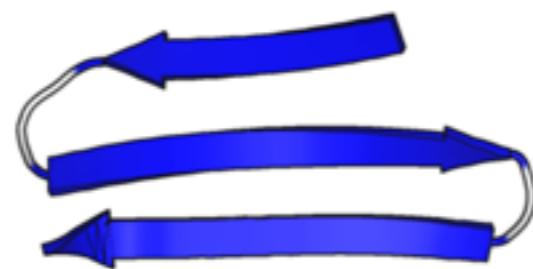


Protein structure

Primary



Secondary

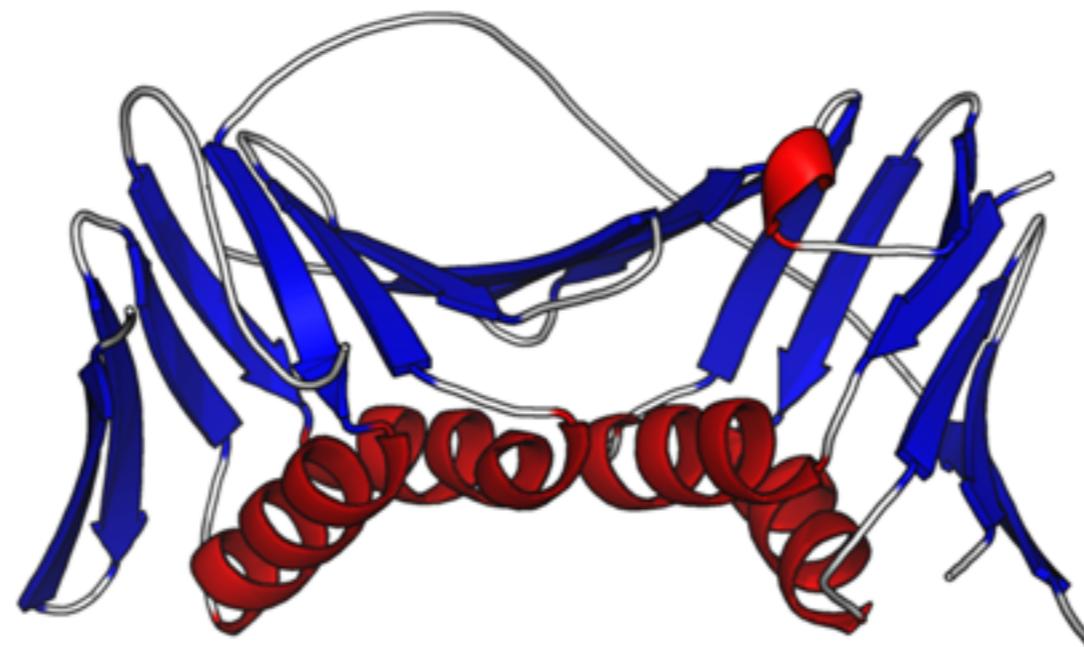


β -Sheet (3 strands)



α -helix

Tertiary



Quaternary

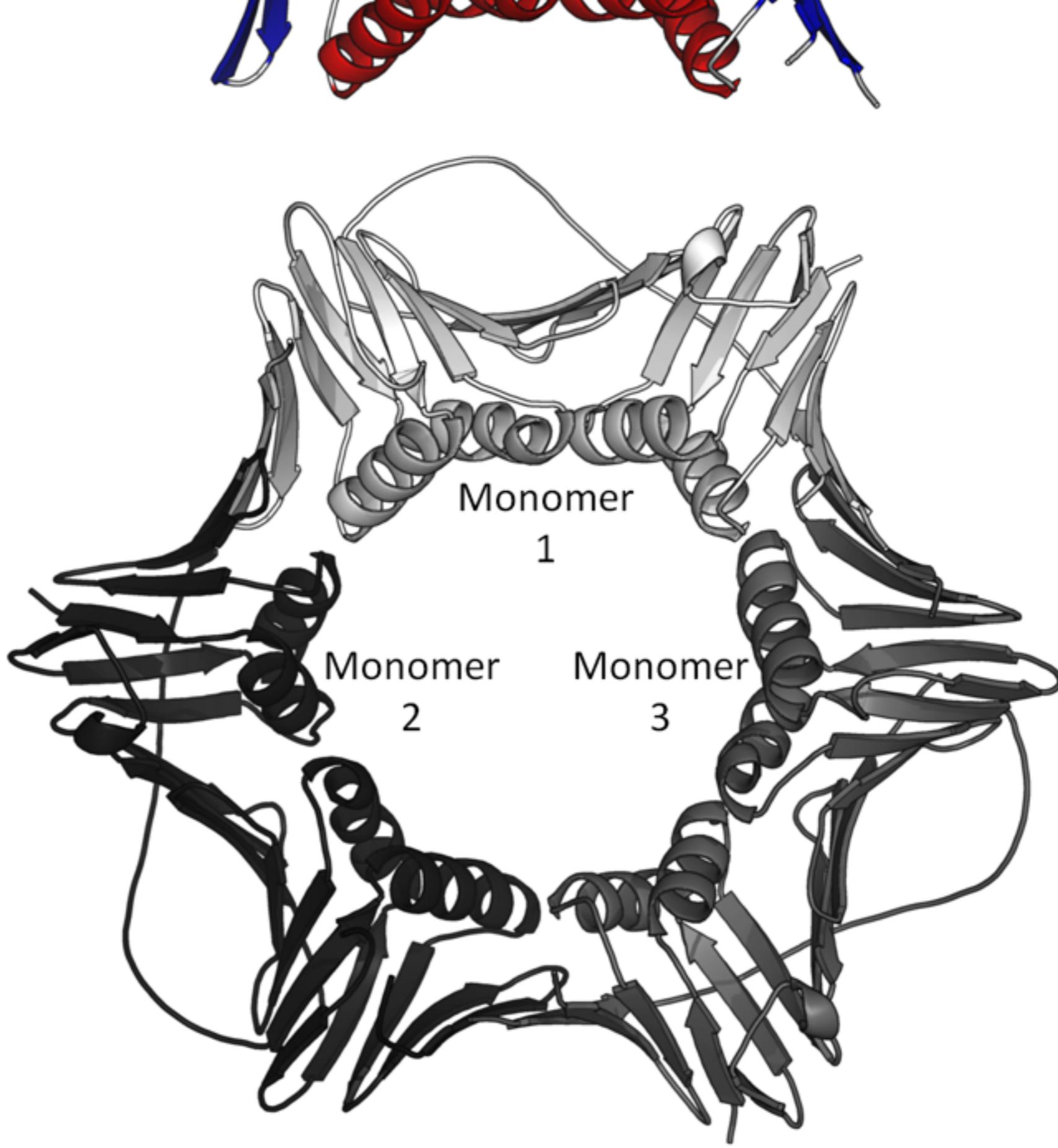


Figure: wikipedia.org

Experiments

1D

- Sequencing
DNA
RNA
Proteins

easy

3D

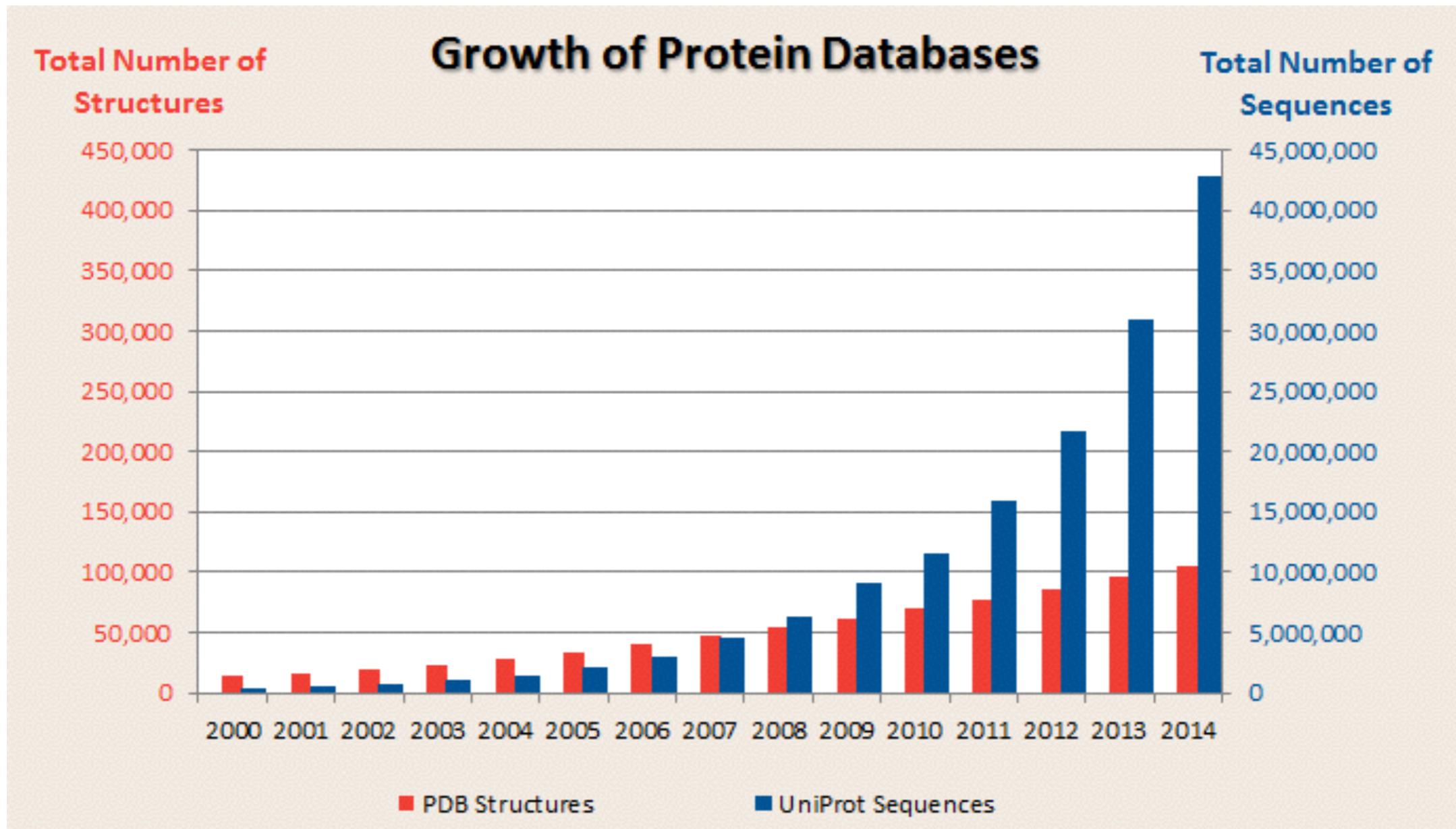
- X-ray crystallography
- NMR (ms)
- ESR (ms)
- circular dichroism (ms)
- x-ray laser (fs)

difficult and expensive

Experiments

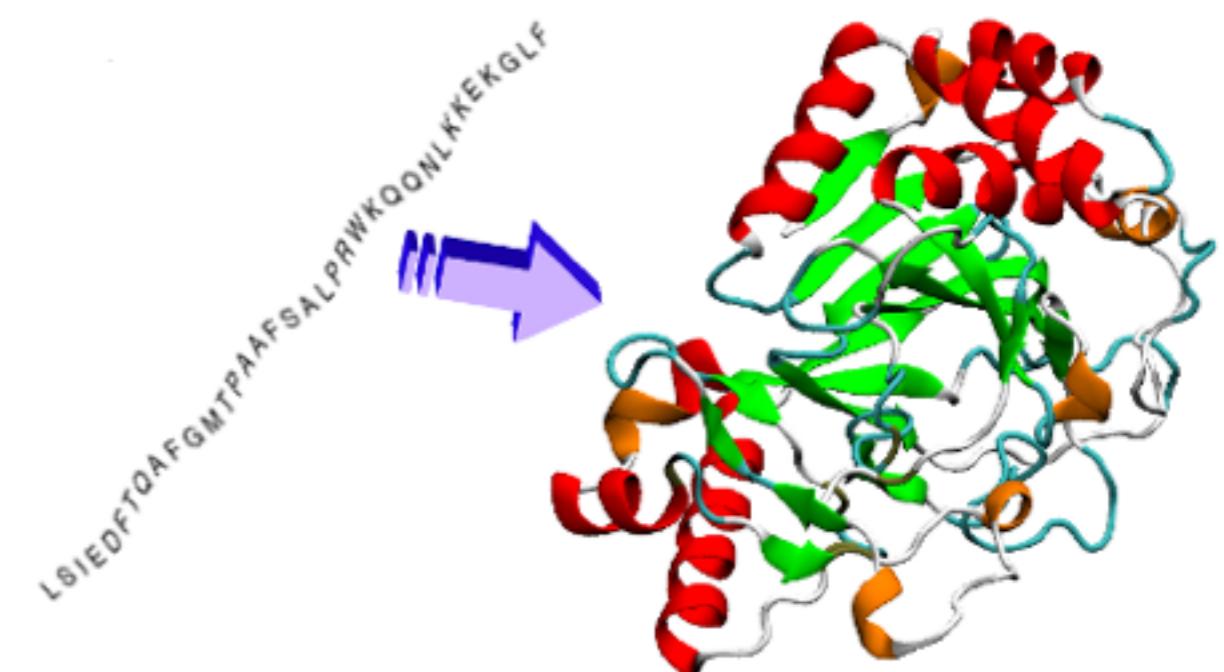
- **NMR**
 - distance between two nuclei
 - angles
 - configuration
 - in solid-state: orientation of bonds
 - BUT
 - averages over ms to s and over many molecules
 - Experiments expensive and time-consuming
 - signals can be mis-assigned
- **X-Ray crystallography**
 - 3D electron density
 - biomolecules are fit into electron density
 - BUT
 - expensive
 - protons are invisible
 - quality of structure depends on model → secondary data!

Protein structure vs. sequences



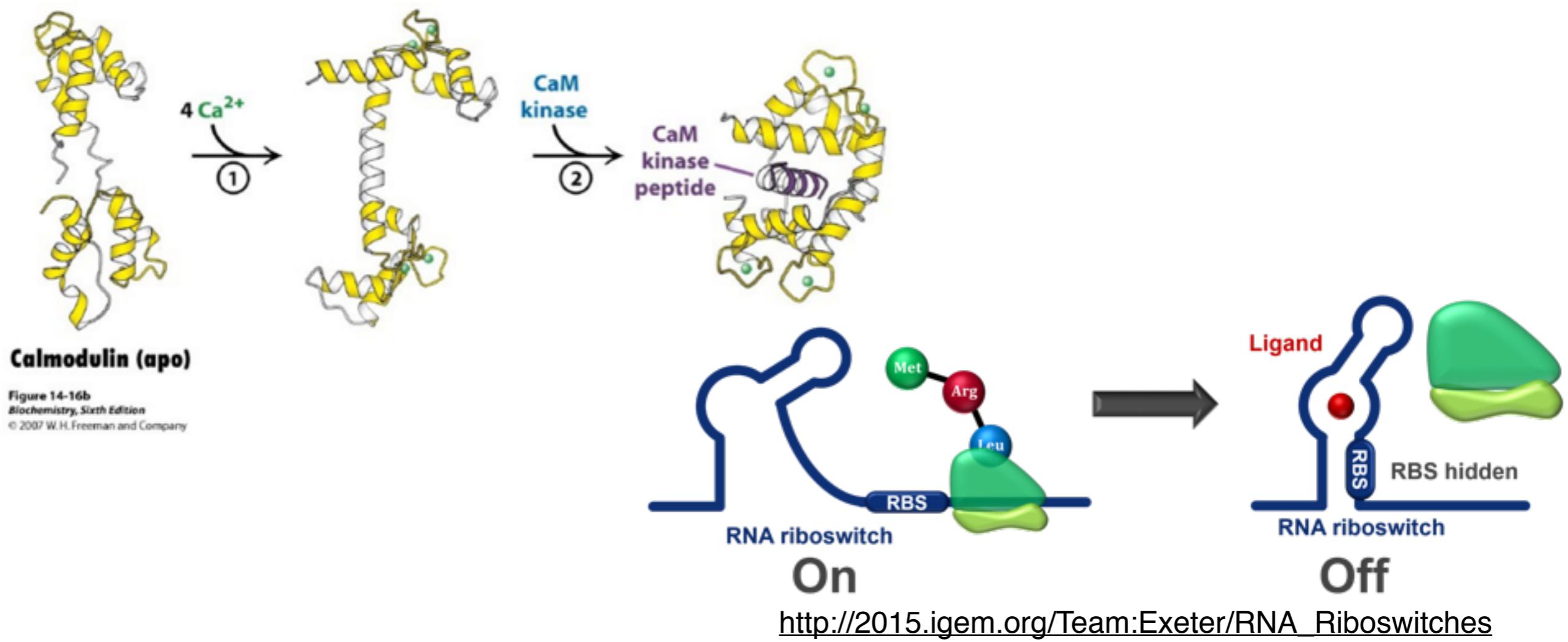
Computer “experiments”

- Use chemical knowledge to build a model
- Predict structure from sequence (proteins, RNA)
- Study internal (functional) dynamics



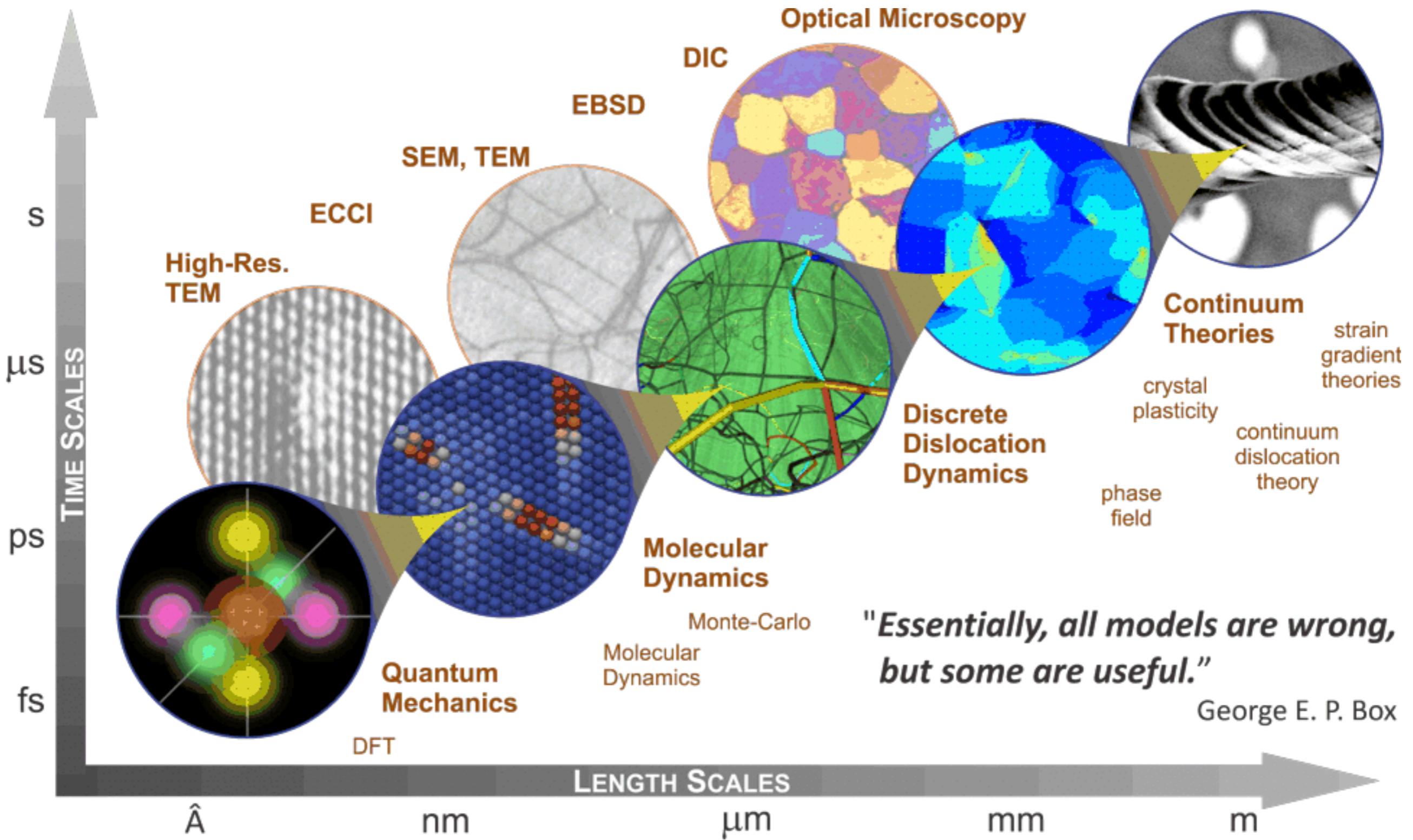
.

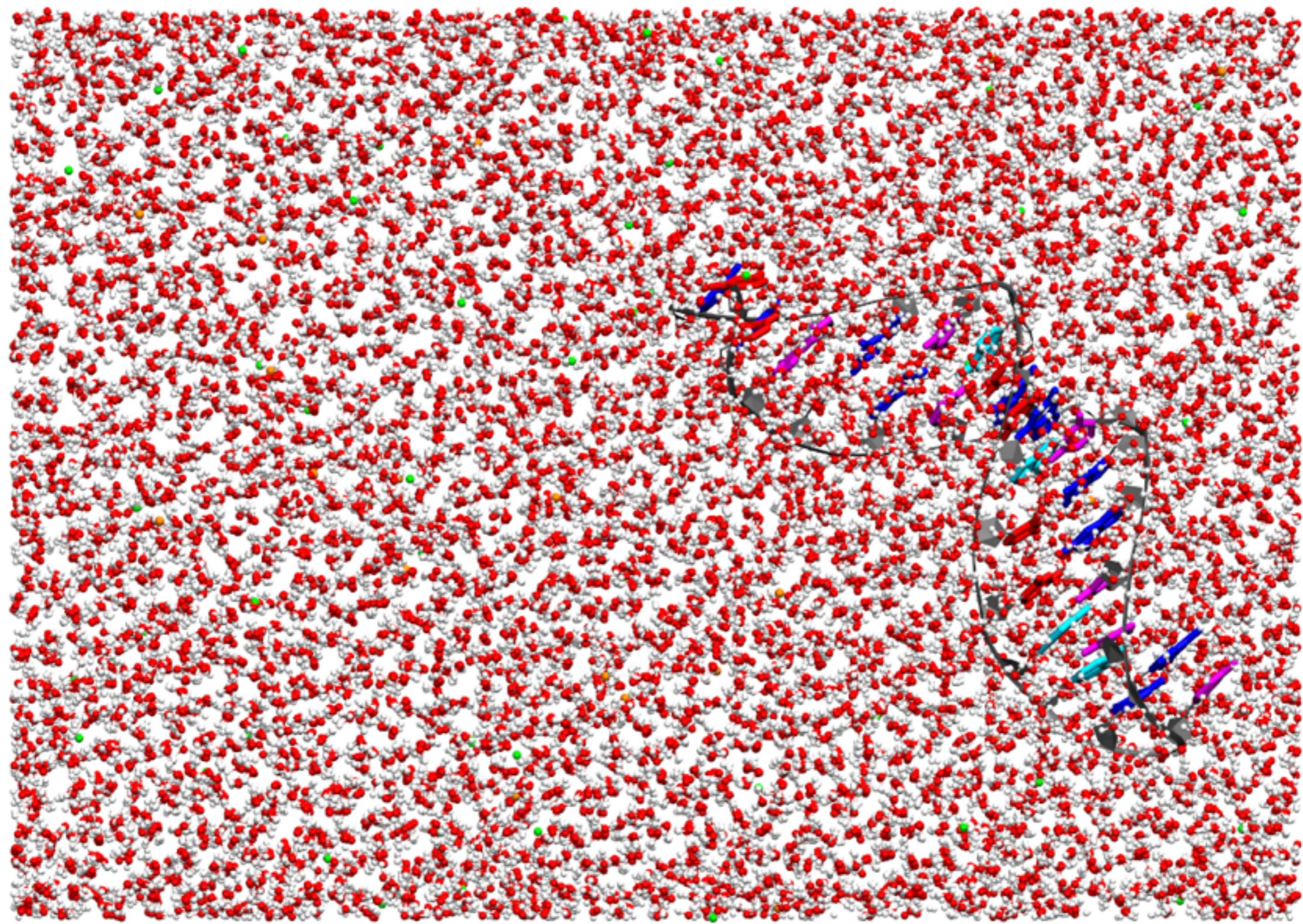
Dynamics

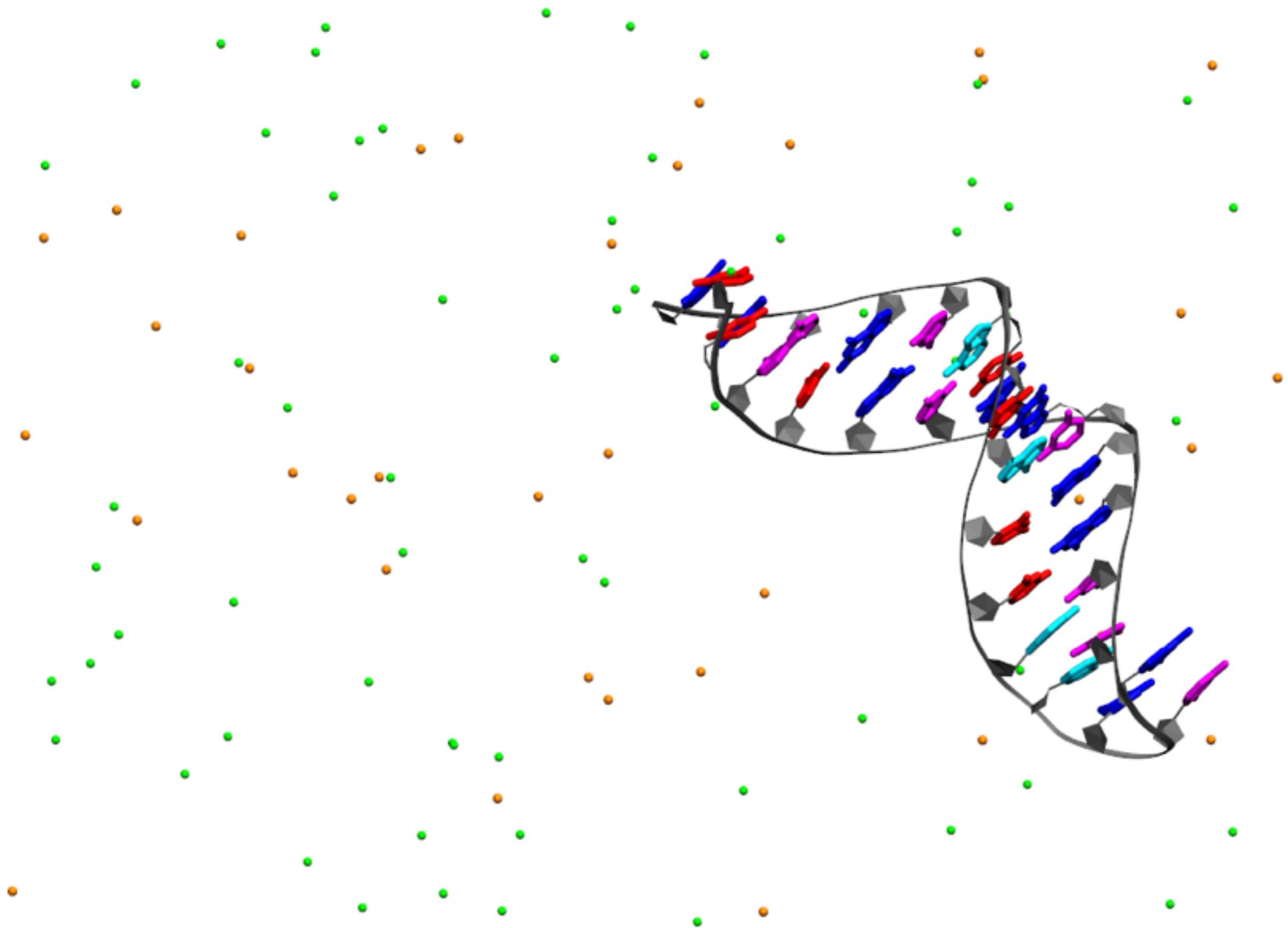


Dynamics difficult to study with experiments
→ Molecular dynamics simulations

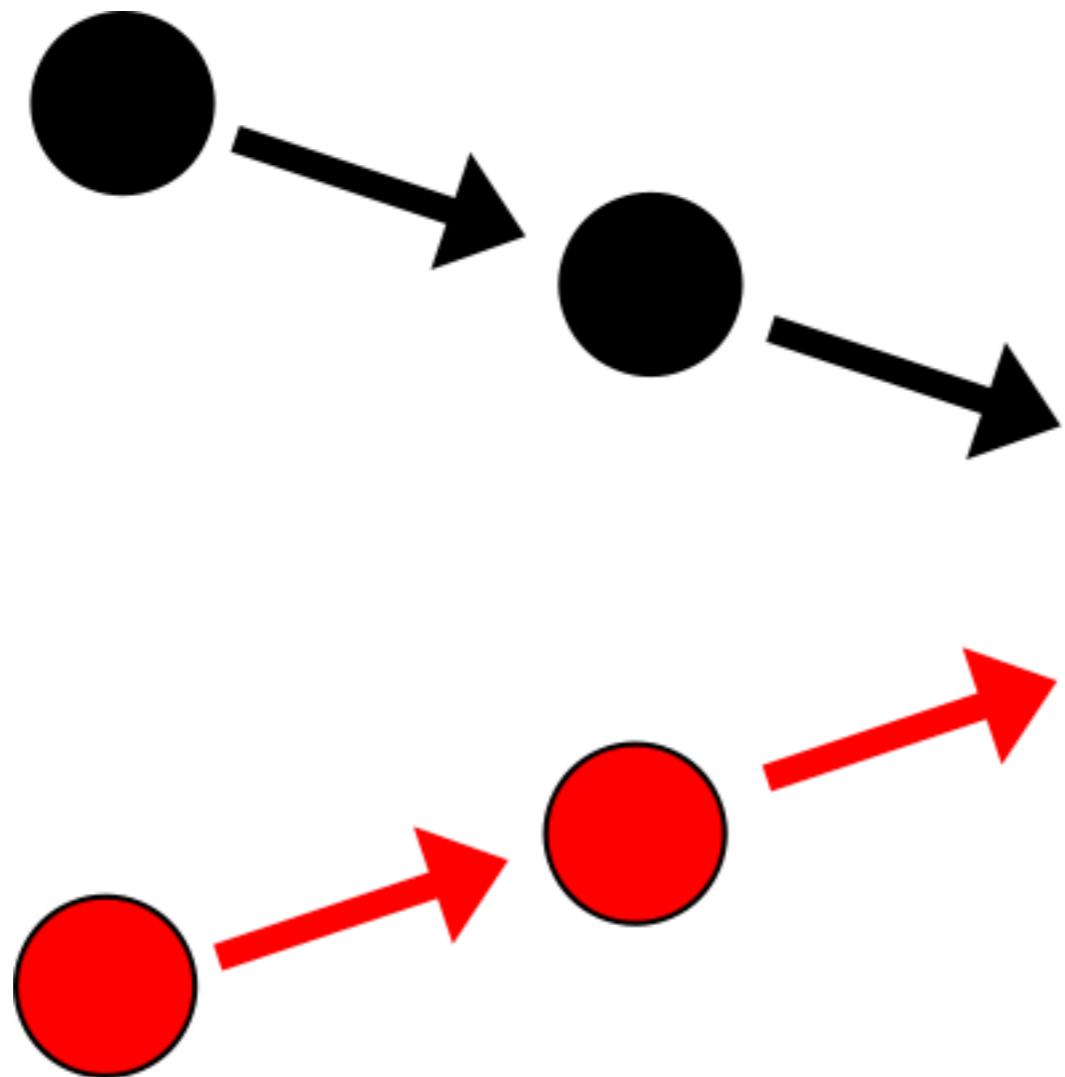
Time scales & length scales



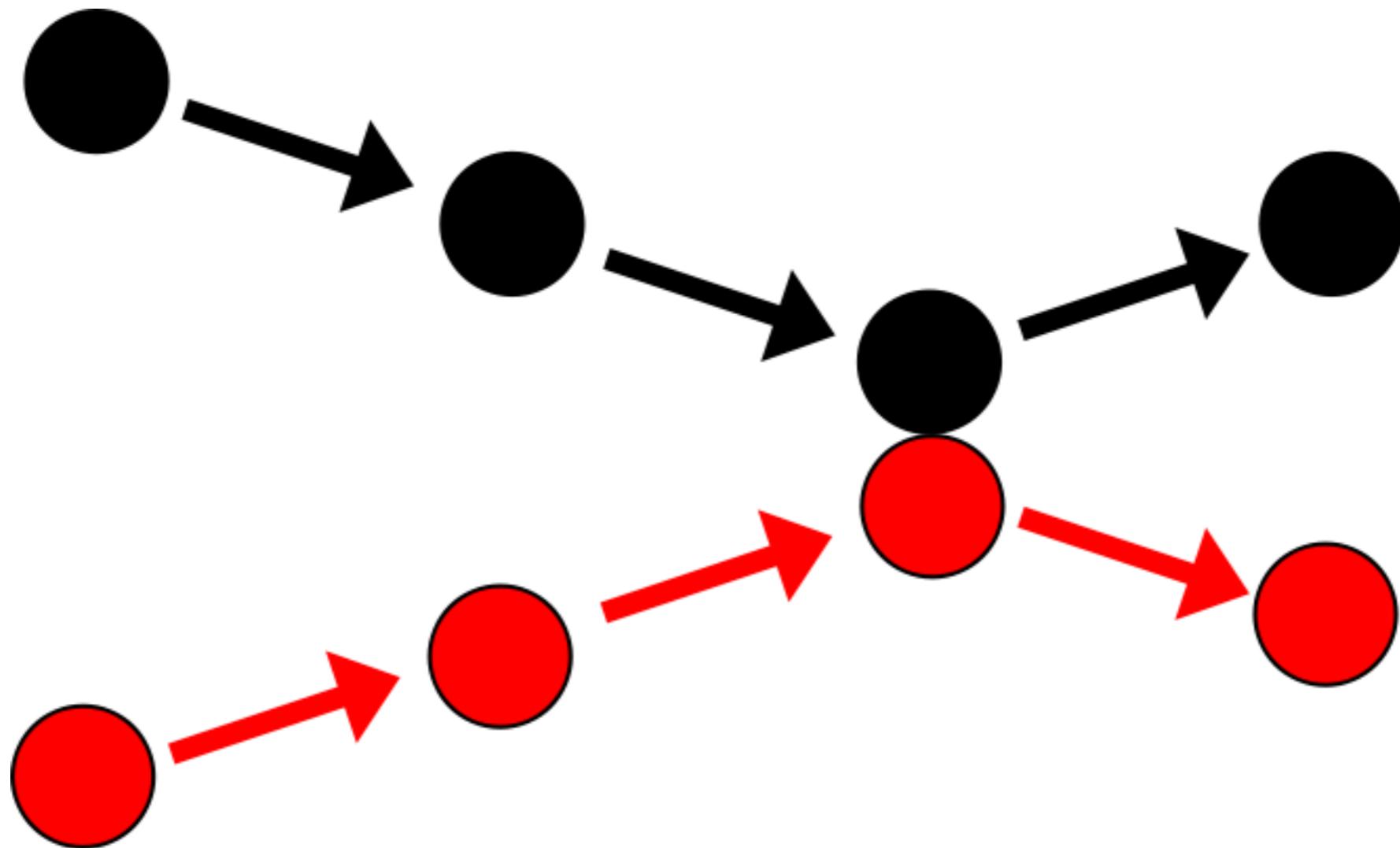




Trajectory



Trajectory



Calculate trajectory

$$r(t + \Delta t) = r(t) + v(t) \cdot \Delta t + \frac{1}{2} a(t) \cdot \Delta t^2$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}(a(t) + a(t + \Delta t)) \cdot \Delta t$$

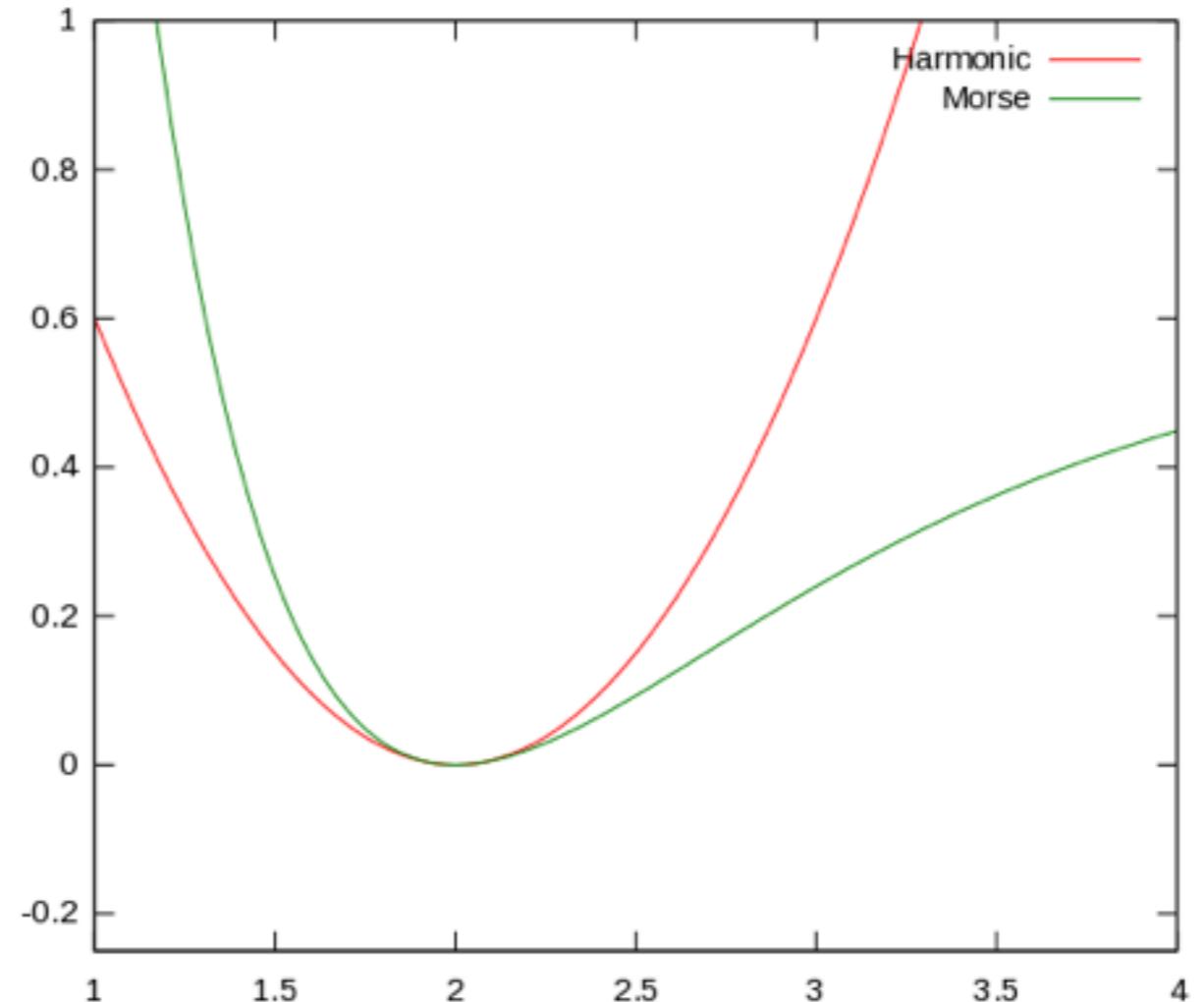
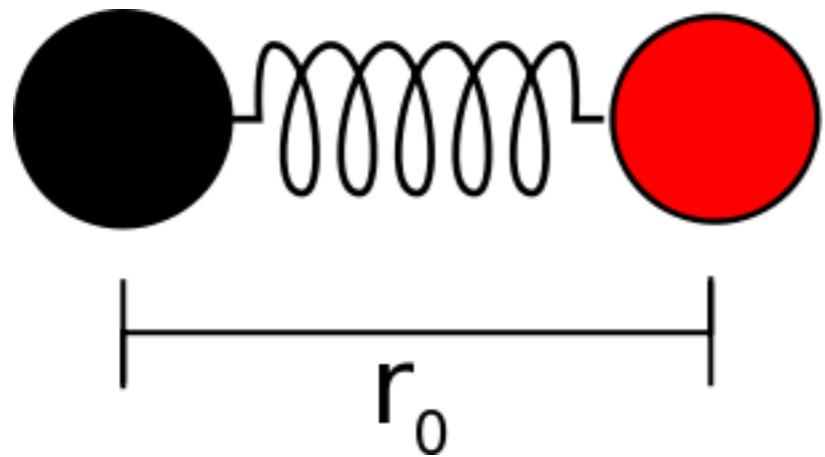
$$a = F/m$$

$$F(r) = -\frac{\partial V(R^N)}{\partial r}$$

Potential terms

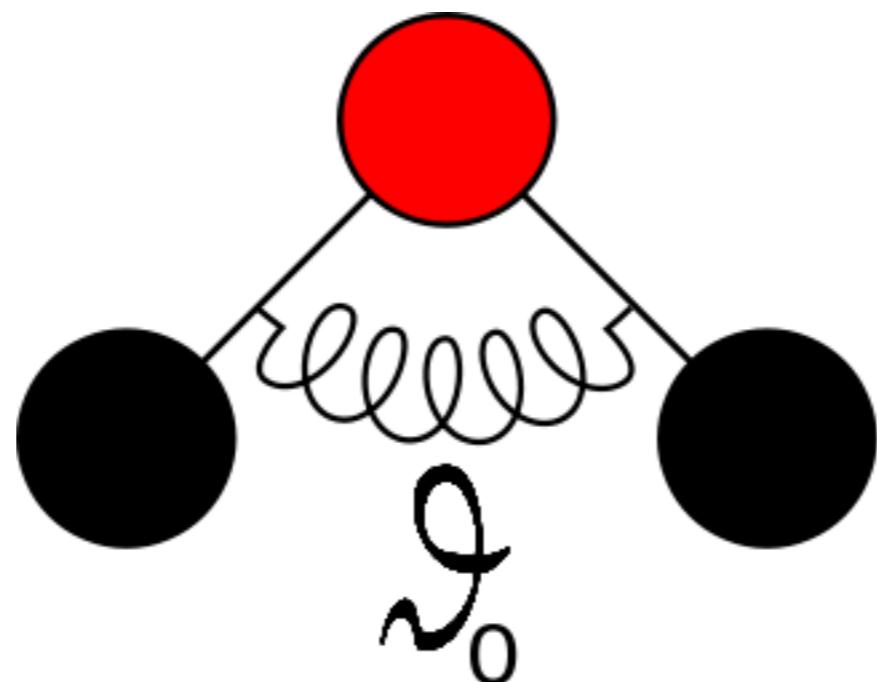
- Bonded interactions: bonds, angles, dihedrals
- Nonbonded interactions:
 - Coulomb (electrostatic)
 - Lennard-Jones (van der Waals)

Bonds



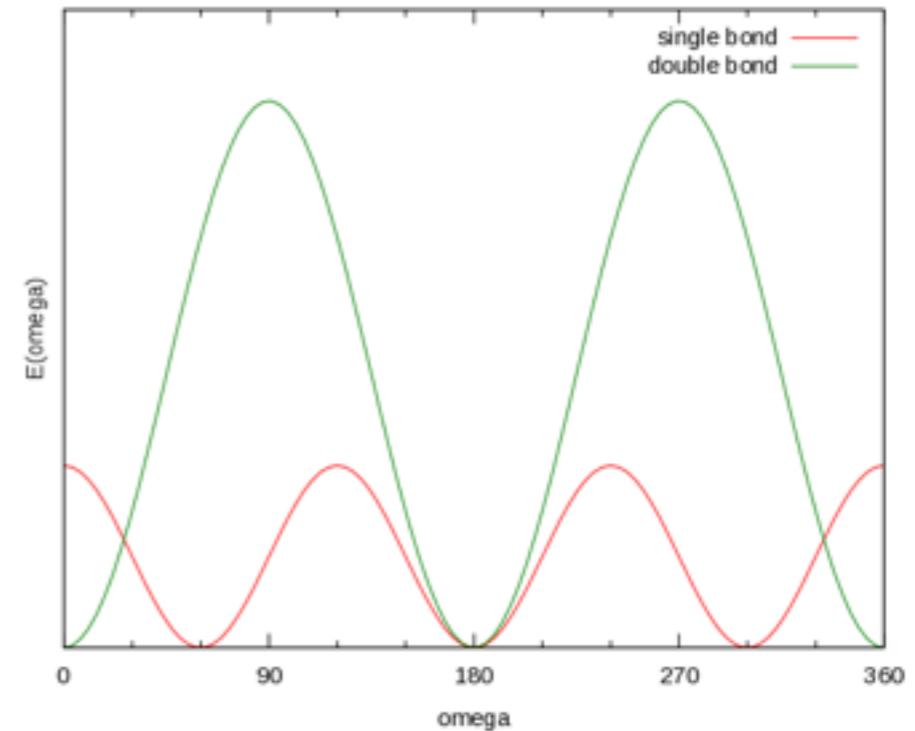
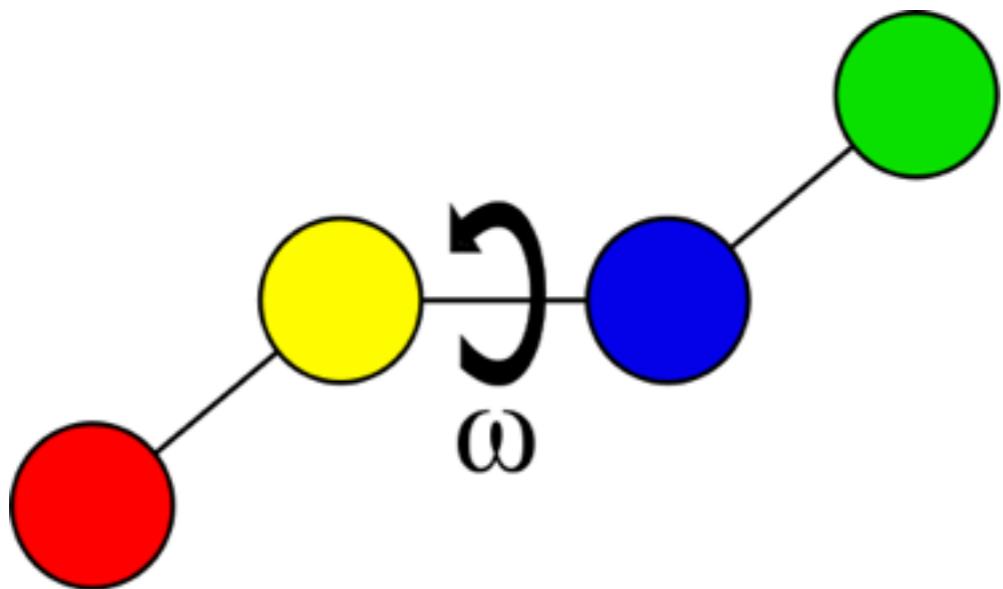
$$V_{bond}(r) = \frac{1}{2}k(r - r_0)^2$$

Angles



$$V_{angle}(\vartheta) = \frac{1}{2} k_\vartheta (\vartheta - \vartheta_0)^2$$

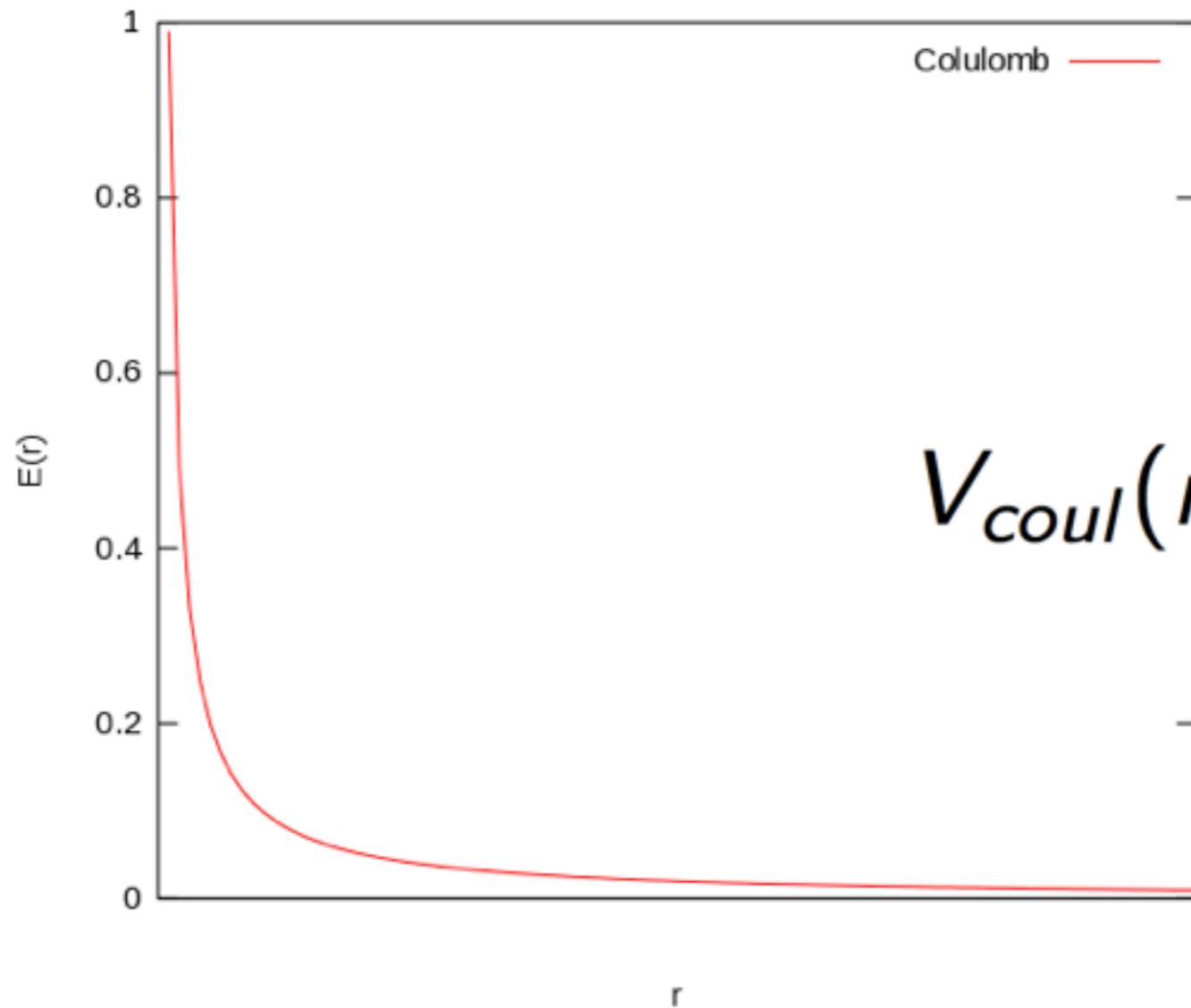
Dihedrals



$$V_{dih}(\omega) = \sum_{n=1,2,3,4,6} V_n \cos[n \cdot \omega - \gamma_n]$$

$$\begin{aligned} V_{C-C}(\omega) &= V \cdot \cos(3\omega) \\ V_{C=C}(\omega) &= V \cdot \cos(2\omega - 90^\circ) \end{aligned}$$

Electrostatics



$$V_{coul}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

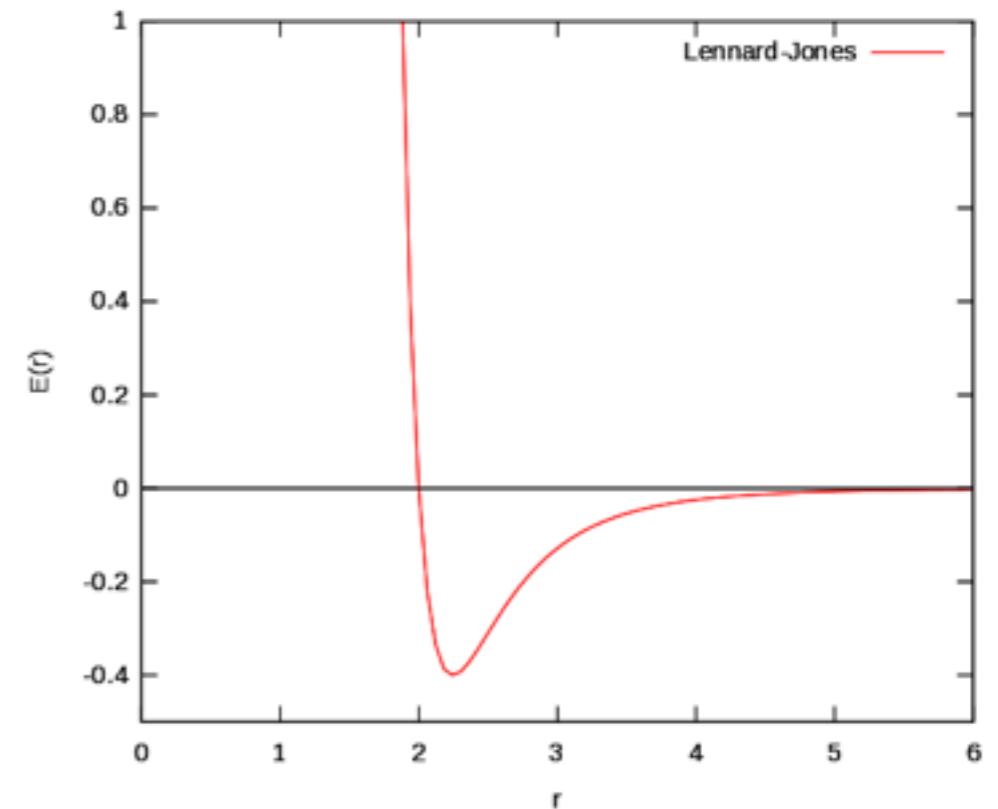
Coulomb potential

Van der Waals

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$\sigma_{ij} = (\sigma_i + \sigma_j)/2$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$



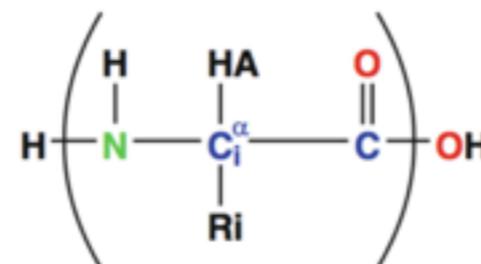
Lennard-Jones potential

Total energy

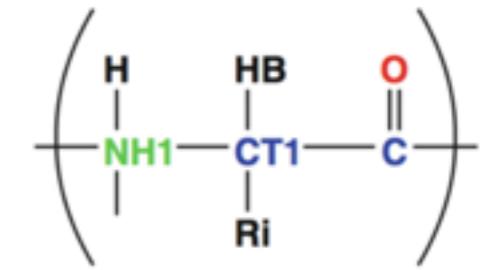
$$\begin{aligned} V(R^N) = & \frac{1}{2} \sum_{\text{bonds}} k_i (r_i - r_i^0)^2 \\ & + \frac{1}{2} \sum_{\text{angles}} k_{\vartheta i} (\vartheta_i - \vartheta_i^0)^2 \\ & + \frac{1}{2} \sum_{\text{dihedrals}} V_n \cdot \cos(n\omega - \gamma_n) \\ & + \sum_i^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\ & + \sum_i^N \sum_{j=i+1}^N \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \end{aligned}$$

Parameters

- Different atom types for different chemical environment

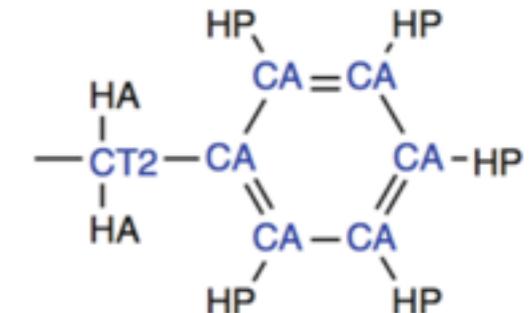
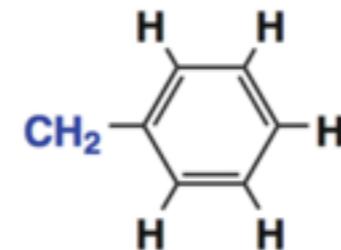


Polypeptide Building Block



Typical Nomenclature

Phenylalanine R:



- Parameters:
 $\sum_i^{bonds} (k_i, r_i^0)$, $\sum_i^{angles} (k_{\vartheta i}, \vartheta_i^0)$,
 $\sum_i^{dihedrals} (V_n, n, \gamma)$, $\sum_i^N (\epsilon_i, \sigma_i, q_i)$

Parameters bonds

```
[ bondtypes ]
```

;	i	j	func	b0	kb
OW	HW	1	0.09572	502080.0	
OW	LP	1	0.01750	753120.0	
C*	HC	1	0.10800	284512.0	
C	C3	1	0.15220	265265.6	
C_2	C3	1	0.15220	265265.6	
C_3	C3	1	0.15220	265265.6	

.

.

.

$$V_{bond}(r) = \frac{1}{2}k(r - r_0)^2$$

Parameters angles

[angletypes]						
;	i	j	k	func	th0	cth
HW		OW	HW	1	109.500	627.600
HW		OW	LP	1	54.750	418.400
OU		U	OU	1	180.000	1255.200
HC		C*	CW	1	126.800	292.880
HC		C*	CB	1	126.800	292.880
HC		CS	CW	1	126.800	292.880

.

.

.

$$V_{angle}(\vartheta) = \frac{1}{2} k_\vartheta (\vartheta - \vartheta_0)^2$$

Parameters dihedrals

```
[ dihedraltypes ]
; i j k l func coefficients
C C N H 3 20.50160 0.00000 -20.50160 0.00000 0.00000
0.00000
C C N CT 3 21.33840 -0.83680 -20.50160 0.00000 0.00000
0.00000
C C CT HC 3 0.17782 0.53346 0.00000 -0.71128 0.00000
0.00000
C C OH HO 3 29.28800 -6.27600 -23.01200 0.00000 0.00000
0.00000
C N C N 3 30.28798 -4.81160 -25.47638 0.00000 0.00000
0.00000
:
:
```

$$V_{dih}(\omega) = \sum_{n=1,2,3,4,6} V_n \cos[n \cdot \omega - \gamma_n]$$

Parameters Lennard-Jones

```
[ atomtypes ]
; full atom descriptions are available in ffoplsaa.atp
; name bond_type mass charge ptype sigma epsilon
opls_001 C 6 12.01100 0.500 A 3.75000e-01 4.39320e-01
opls_002 O 8 15.99940 -0.500 A 2.96000e-01 8.78640e-01
opls_003 N 7 14.00670 -0.570 A 3.25000e-01 7.11280e-01
opls_004 H 1 1.00800 0.370 A 0.00000e+00 0.00000e+00
opls_005 C2 6 14.02700 0.200 A 3.80000e-01 4.93712e-01
opls_006 CH 6 13.01900 0.200 A 3.80000e-01 3.34720e-01
.
```

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$\sigma_{ij} = (\sigma_i + \sigma_j)/2$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$

Charges, masses

```
[ moleculetype ]
; Name          nrexcl
DMPC            3

[ atoms ]
;   nr      type    resnr   residue   atom    cgnr    charge    mass
  1  opls_4031    1       DMPC     C1        0        0.400    15.0350
  2  opls_4031    1       DMPC     C2        0        0.400    15.0350
  3  opls_4031    1       DMPC     C3        0        0.400    15.0350
  4  opls_4030    1       DMPC     N4        0       -0.500    14.0067
  5  opls_4032    1       DMPC     C5        0        0.300    14.0270
  6  opls_4004    1       DMPC     C6        1        0.250    14.0270
  7  opls_4002    1       DMPC     O7        1       -0.750    15.9994
.
.
.
```

$$V_{coul}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

Limits of MD

- bonds can not be broken → no reactions, no (de-)protonation
- atomic partial charges are fixed
 - no polarization
 - no electrons
 - no quantum mechanics
- simulation temperature around 300 K
- computational limit $\sim \mu\text{s}$ (10^9 steps)

MD and HPC

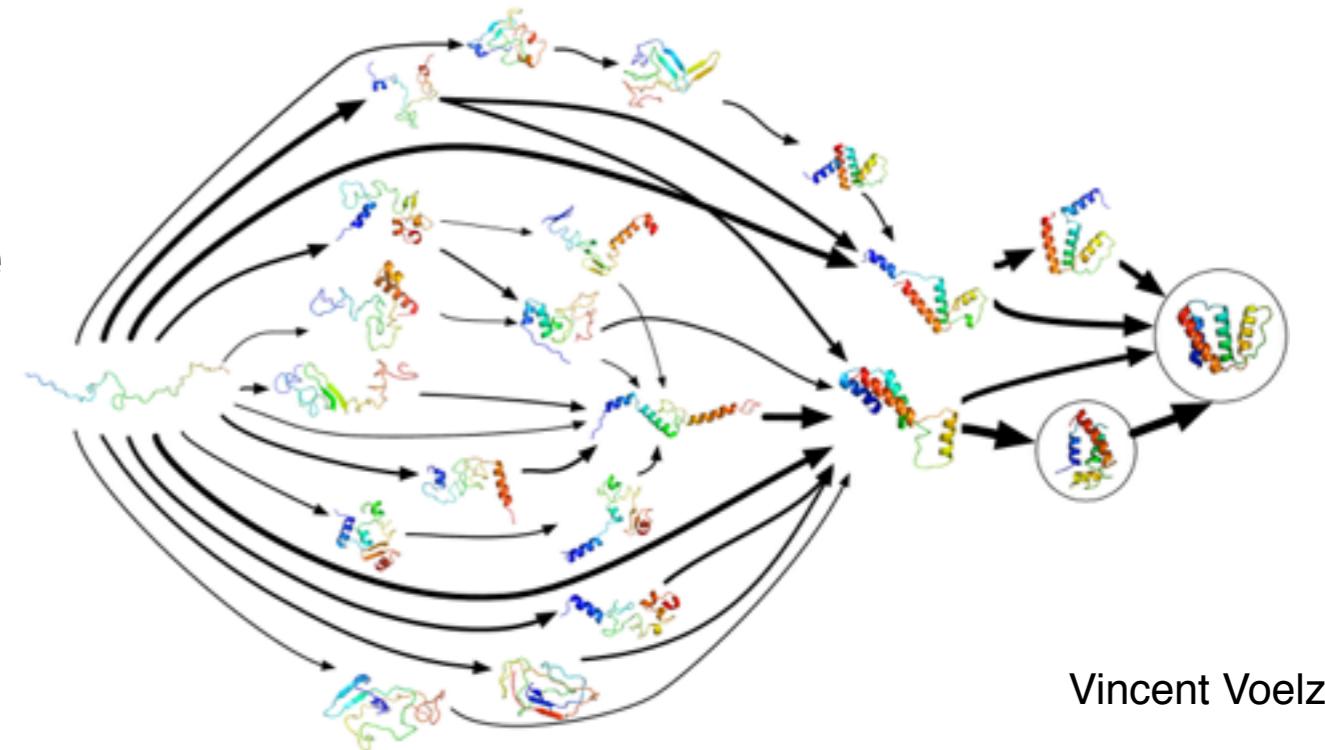
- calculate pairwise distances and forces between $\sim 10^6$ particles
- repeat for $10^8\text{-}10^9$ time steps

Efficiency is fundamental!

Two notable examples of parallel computing

folding@home

- distributed computing for disease research
- protein folding
- 200k PCs by volunteers



Anton

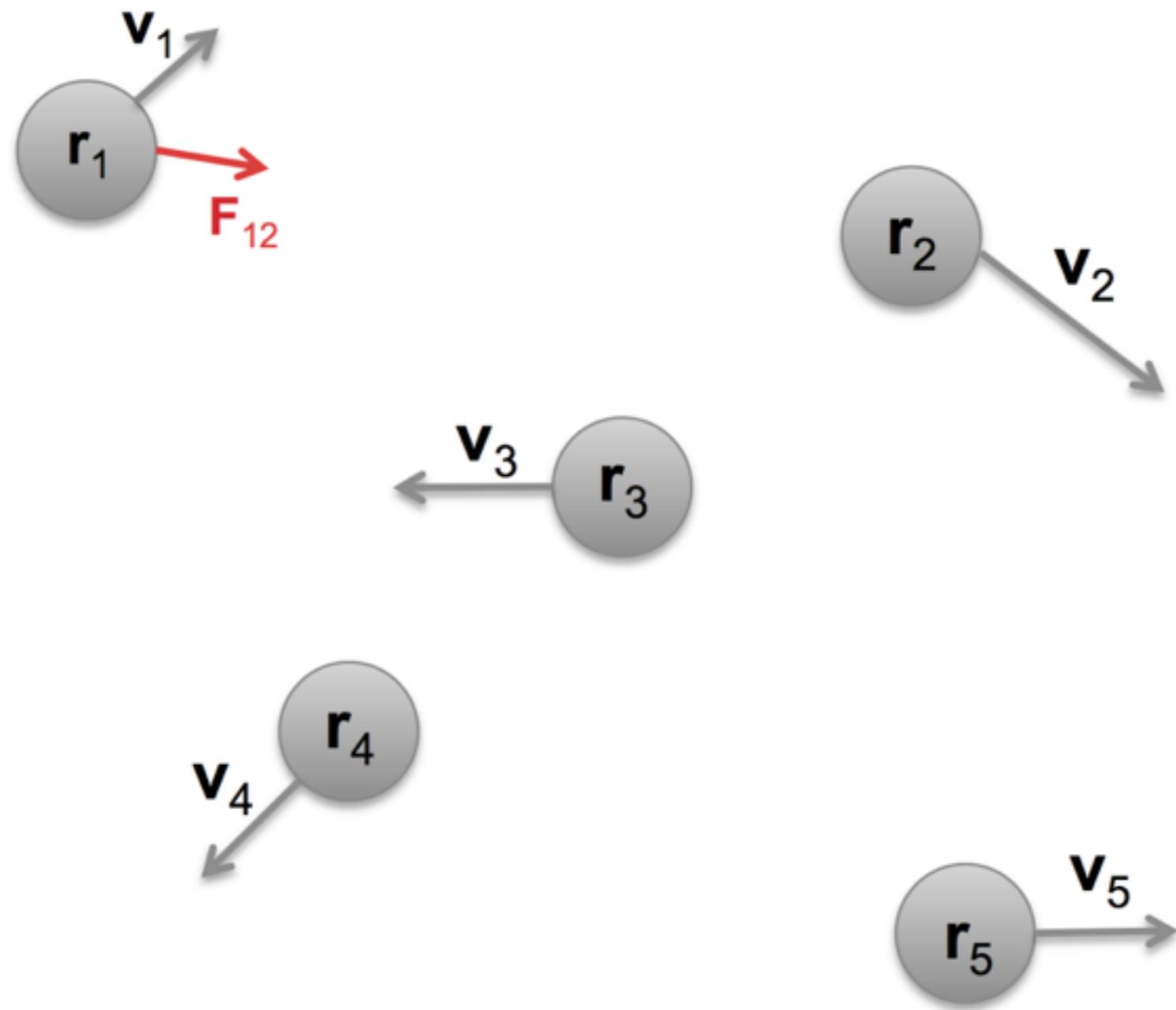
- massively parallel supercomputer by D. E. Shaw in New York
- specialized for MD on biomolecules
- Performance: 17 000 ns/d, ~25k atoms (standard HPC cluster ~100 ns/d)

MD and HPC

- calculate pairwise distances and forces between $\sim 10^6$ particles
- repeat for $10^8\text{-}10^9$ time steps

Efficiency is fundamental!

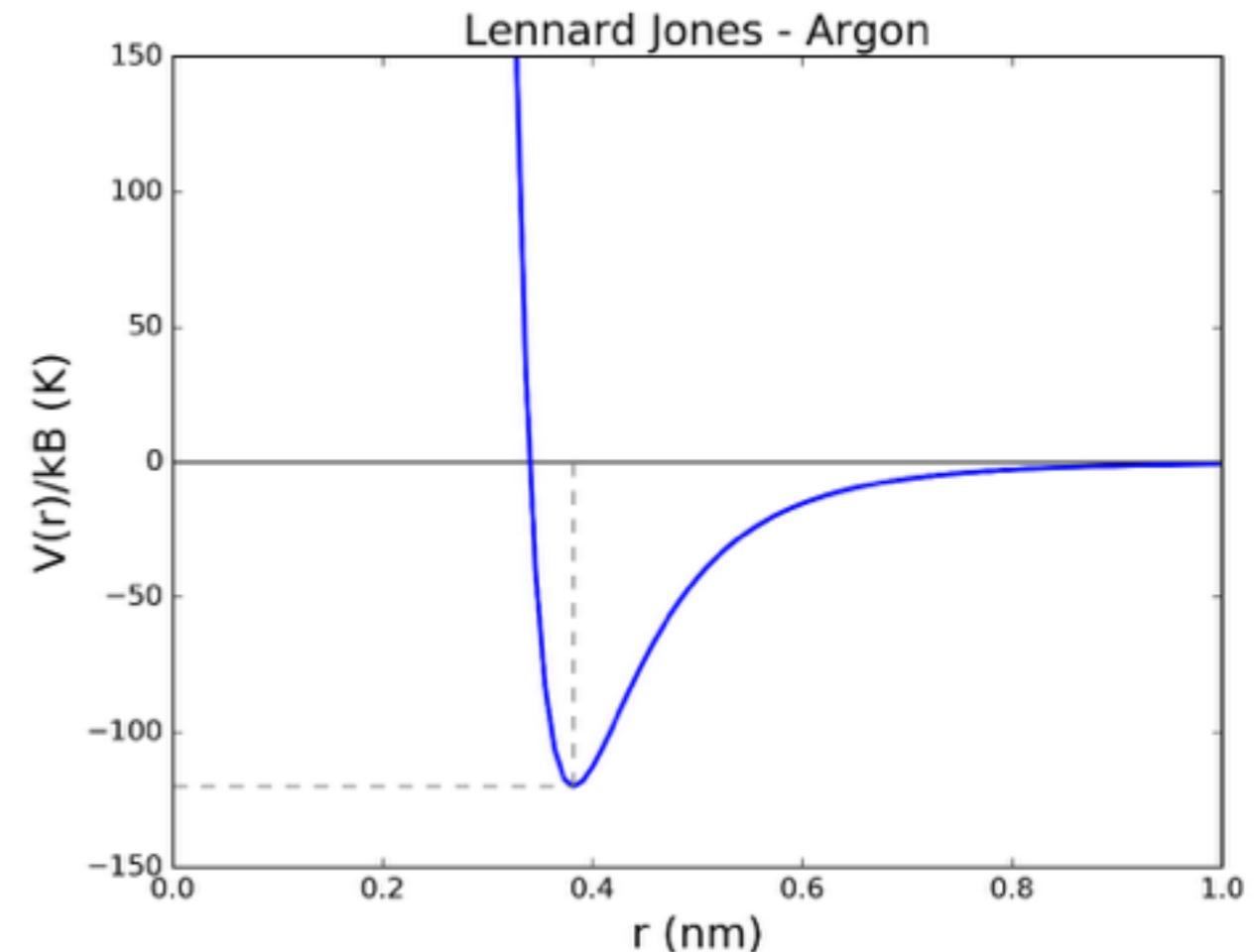
Force calculation



Lennard-Jones* potential

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

- Approx. interaction between neutral atoms
- Attractive van der Waals ($-1/r^6$)
- Pauli repulsion ($1/r^{12}$)
- minimum $-\epsilon$ at $r=2^{1/6}\sigma$



*Sir Lennard-Jones, 1924

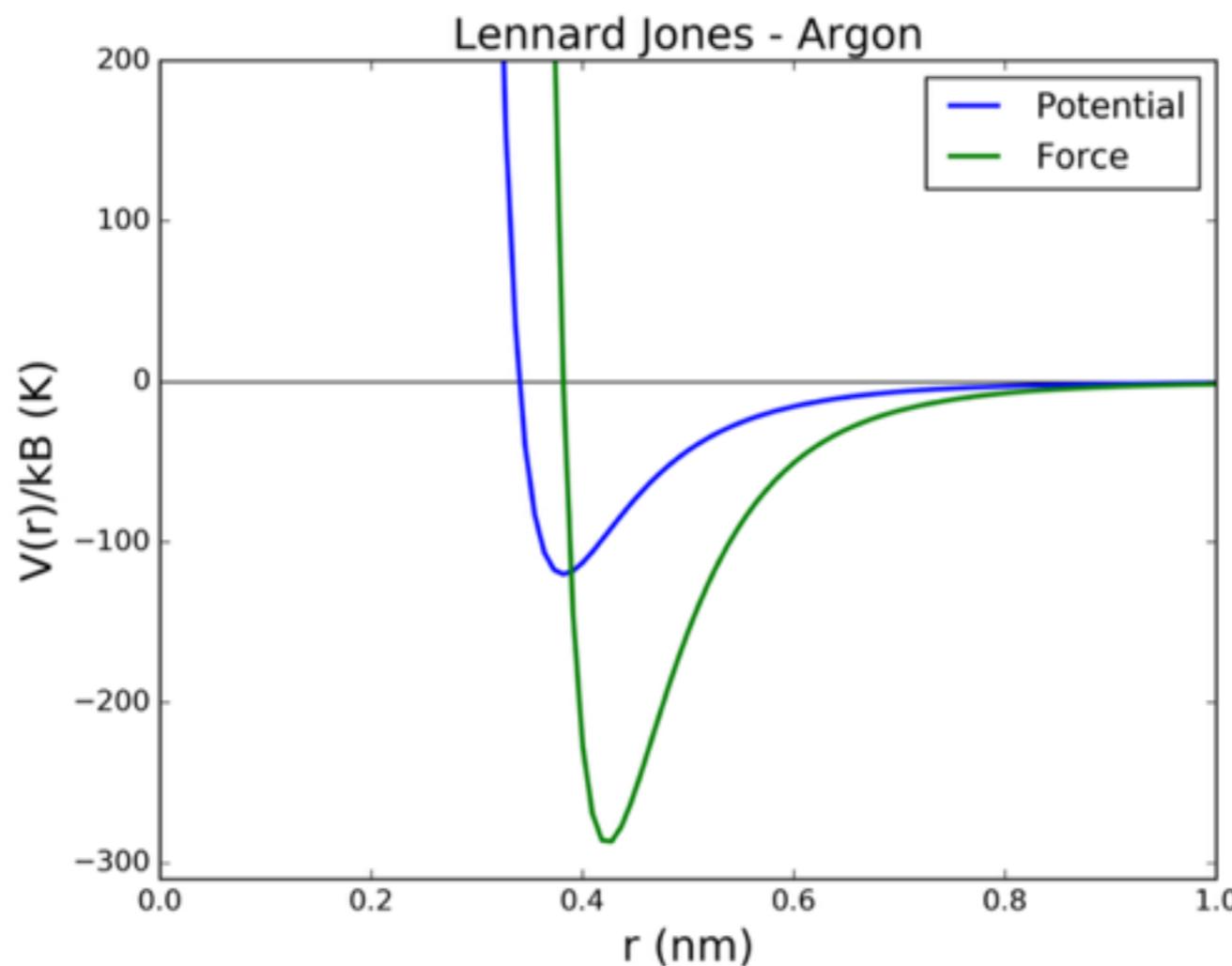
$\epsilon/k_B=120K$, $\sigma=0.34$ nm

Lennard-Jones potential - typical parameters

Lennard-Jones parameters, ε and σ , for various substances.

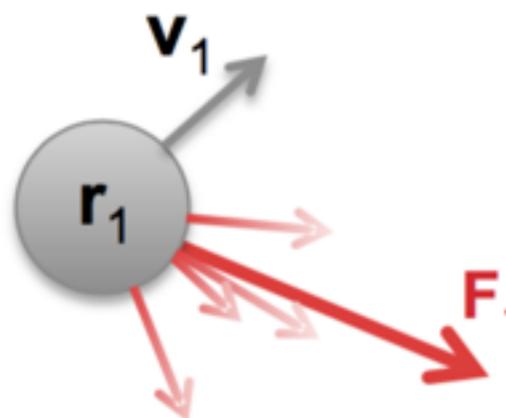
Species	$(\varepsilon/k_B)/K$	σ/pm	$(2\pi\sigma^3 N_A/3)/\text{cm}^3 \cdot \text{mol}^{-1}$
He	10.22	256	21.2
Ne	35.6	275	26.2
Ar	120	341	50.0
Kr	164	383	70.9
Xe	229	406	86.9
H ₂	37.0	293	31.7
N ₂	95.1	370	63.9
O ₂	118	358	57.9
CO	100	376	67.0
CO ₂	189	449	114.2
CF ₄	152	470	131.0
CH ₄	149	378	68.1
C ₂ H ₄	199	452	116.5
C ₂ H ₆	243	395	77.7
C ₃ H ₈	242	564	226.3
C(CH ₃) ₄	232	744	519.4

Calculate forces

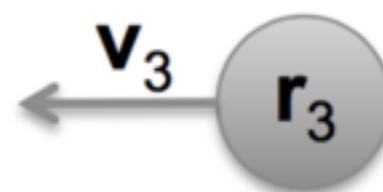


$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$
$$\mathbf{F}_{LJ}(r_{ij}) = -\nabla V = 4\epsilon_{ij} \frac{6}{r_{ij}} \left[2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \frac{\mathbf{r}_{ij}}{r_{ij}}$$

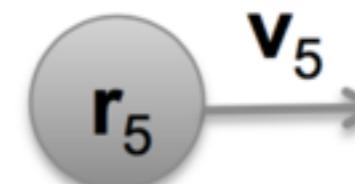
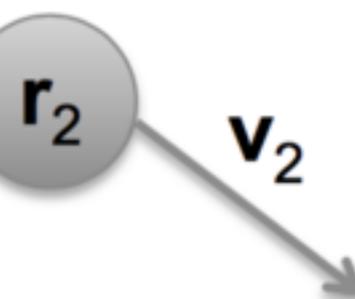
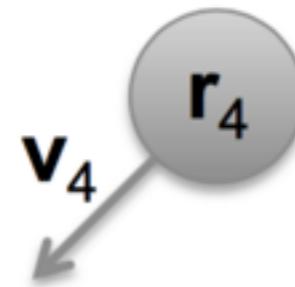
Calculate forces



$$\mathbf{F}_i = \sum_j \mathbf{F}(r_{ij})$$

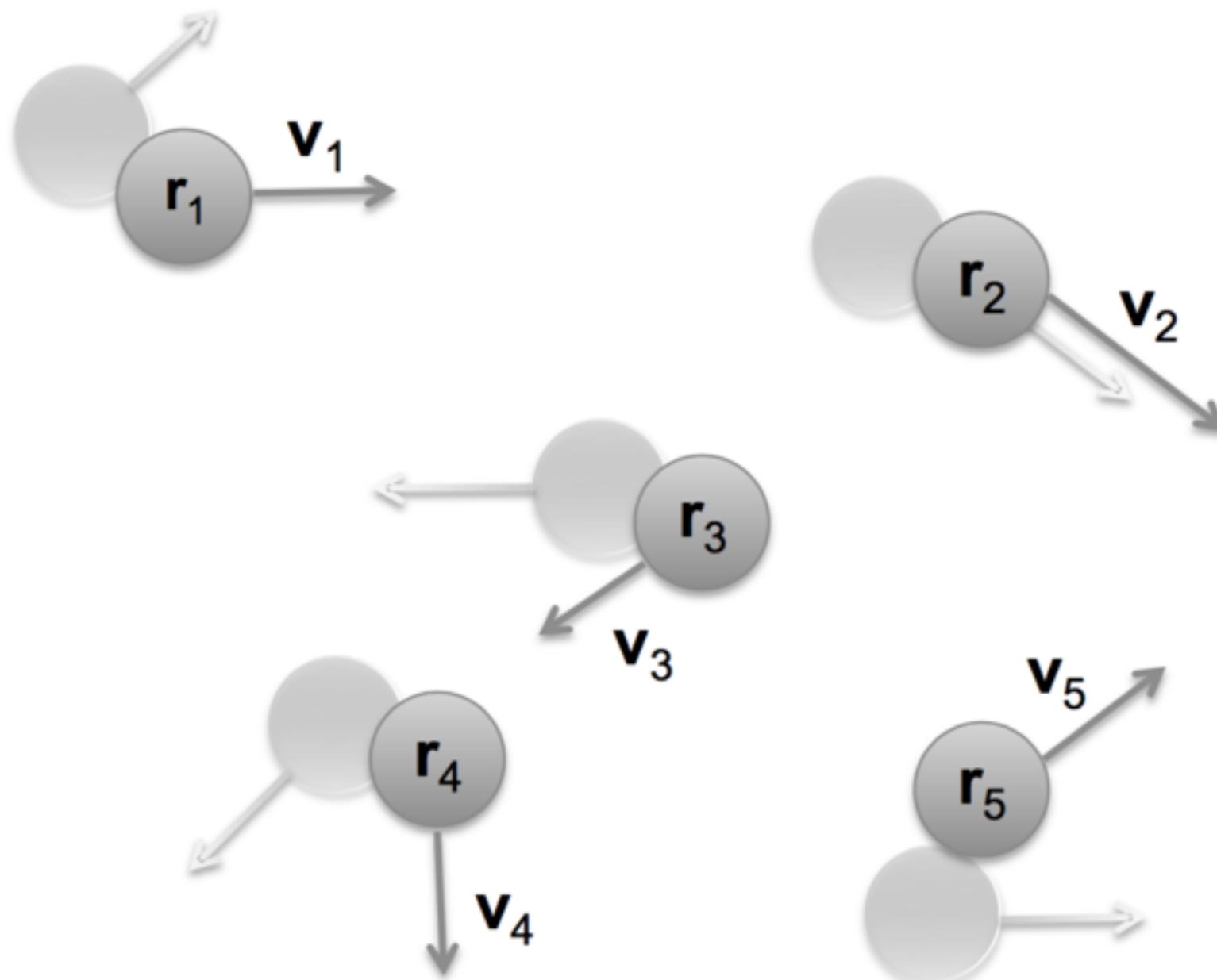


$$\mathbf{F}(r_{ij}) = -\mathbf{F}(r_{ji})$$



$n(n-1)/2$ terms

Update velocities and positions

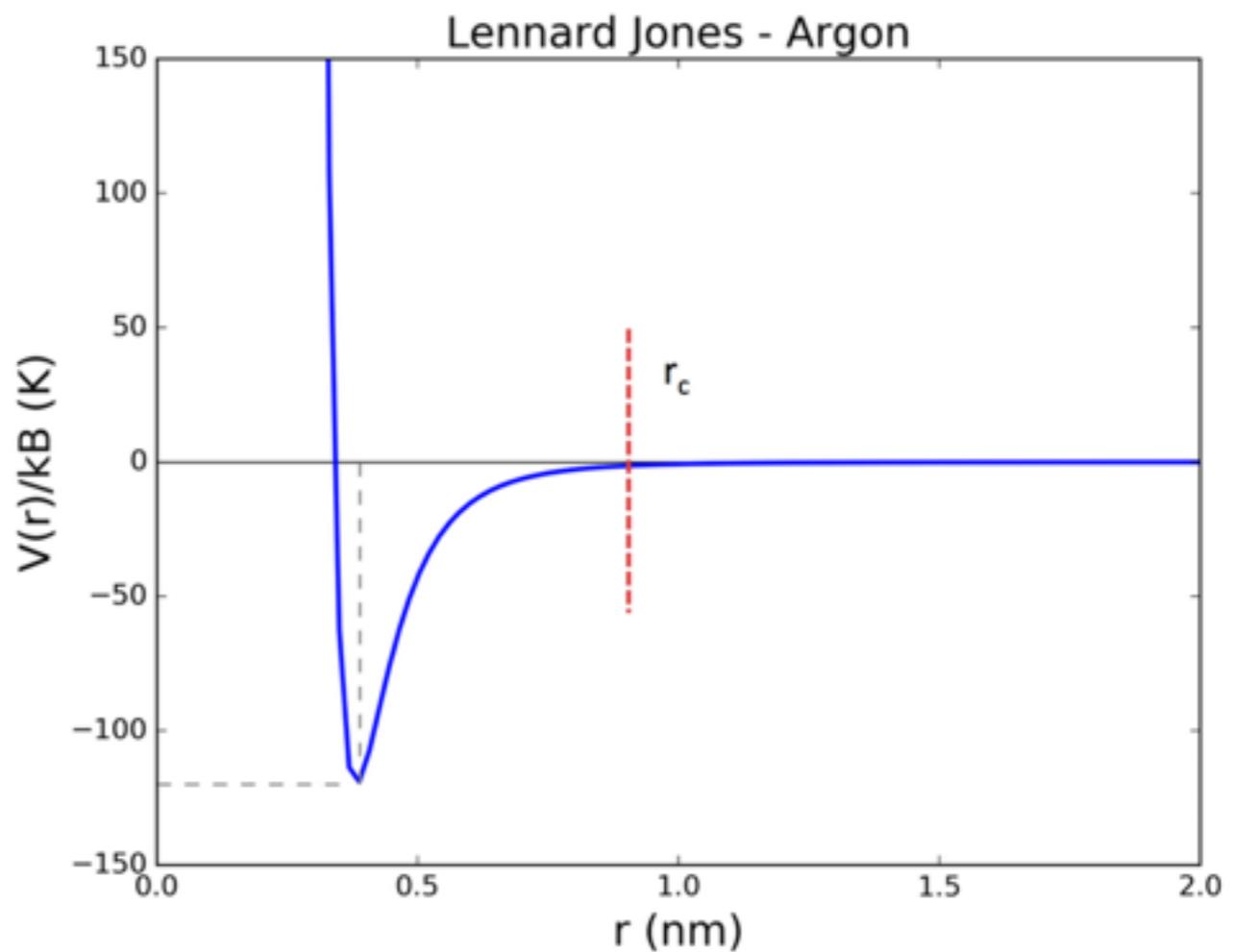


Long range cut-off

- calculation of $n(n-1)/2$ forces is most expensive part
- Lennard-Jones potential decays quickly
- Truncate potential
- Substantial computational speed-up

$$r_c \approx 2.5\sigma$$

$$r_c < L/2$$



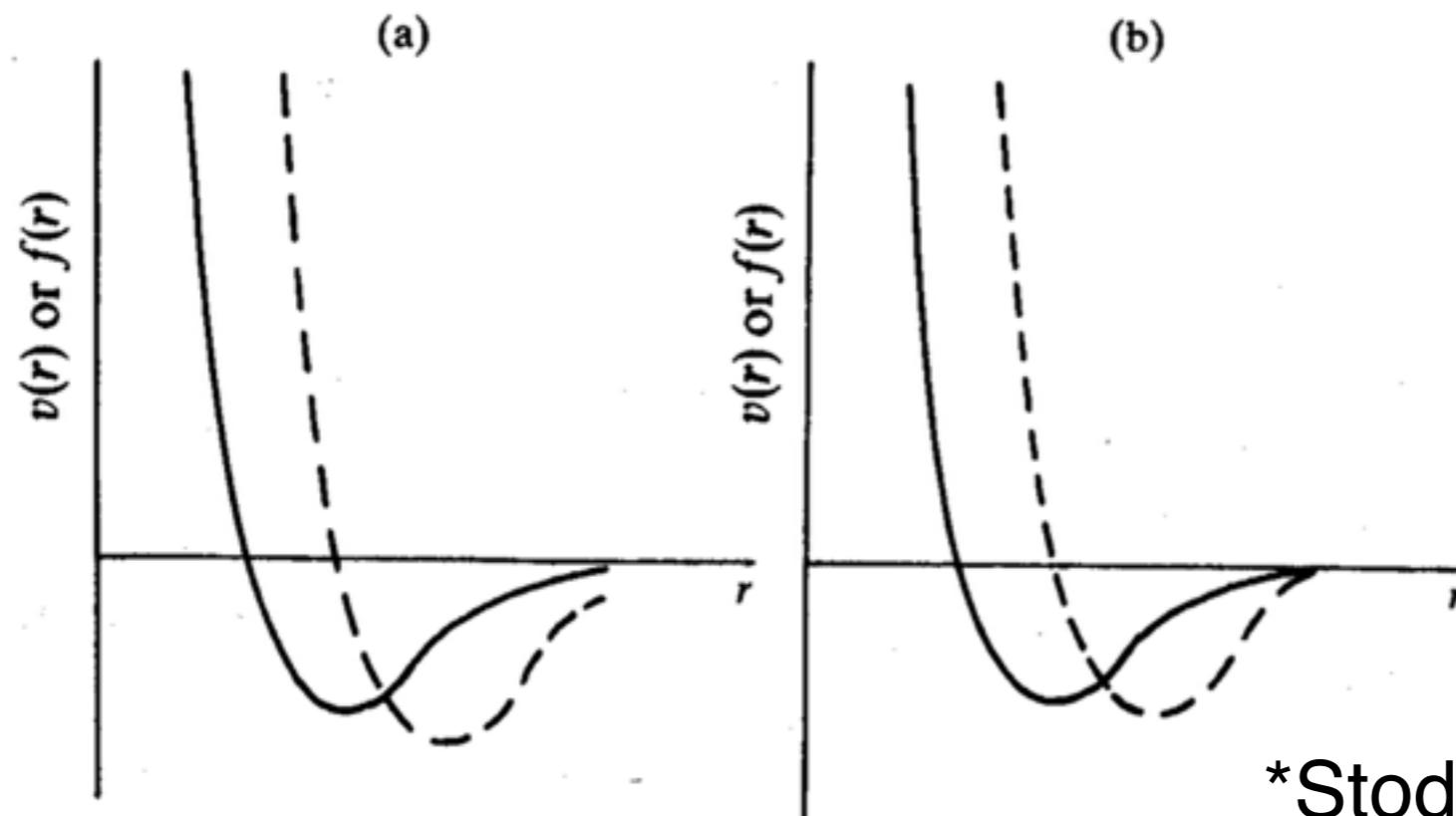
Cut-off correction

- Shifted Potential

$$V^S(r_{ij}) = \begin{cases} V_{LJ}(r_{ij}) - V_{LJ}(r_c) & \text{for } r_{ij} \leq r_c \\ 0 & \text{for } r_{ij} > r_c \end{cases}$$

- Shifted force* potential

$$V^{SF}(r_{ij}) = \begin{cases} V_{LJ}(r_{ij}) - V_{LJ}(r_c) - \left(\frac{dV}{dr}\right)_{r=r_c} (r - r_c) & \text{for } r_{ij} \leq r_c \\ 0 & \text{for } r_{ij} > r_c \end{cases}$$



*Stoddard and Ford, 1973

Shifted Potential

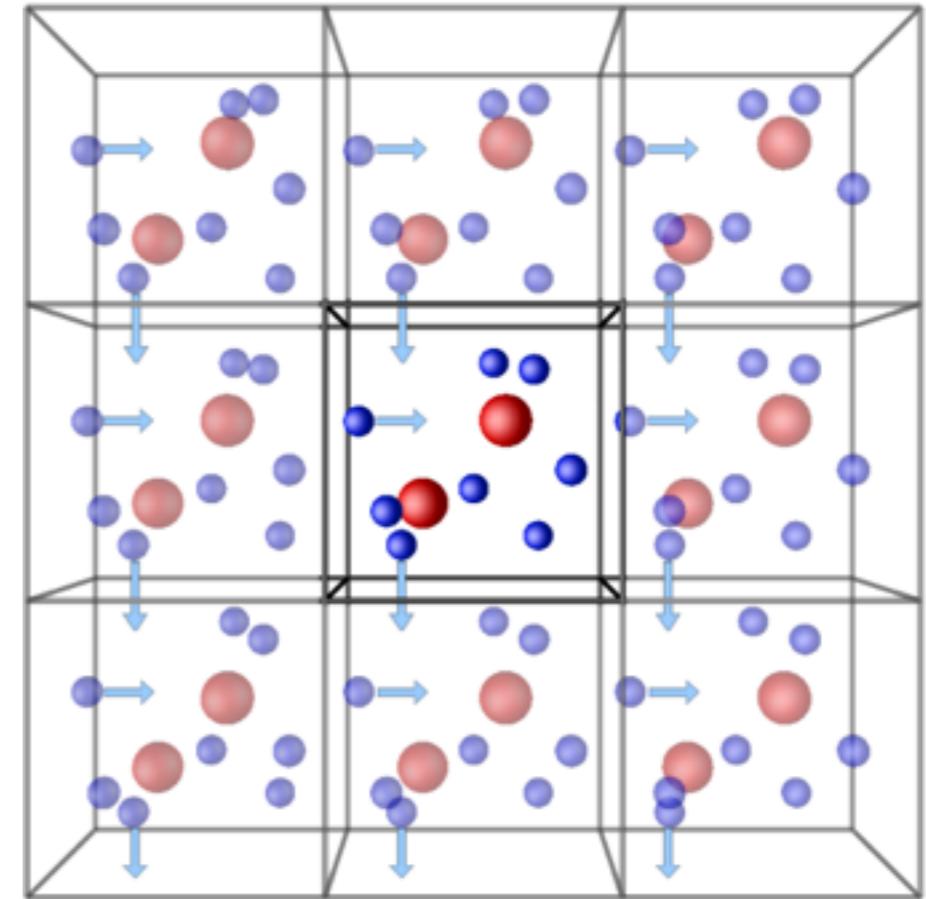
- simplemd:

```
engcorrection=4.0*(1.0/pow(forcecutoff2,6.0)-1.0/pow(forcecutoff2,3))
for iatom in atoms
    for jatom in neighbors_of_iatom
        // Total potential energy
        engconf += 4.0*(1./d12 - 1.0/d6) - engcorrection
```

Periodic boundary conditions

Problem

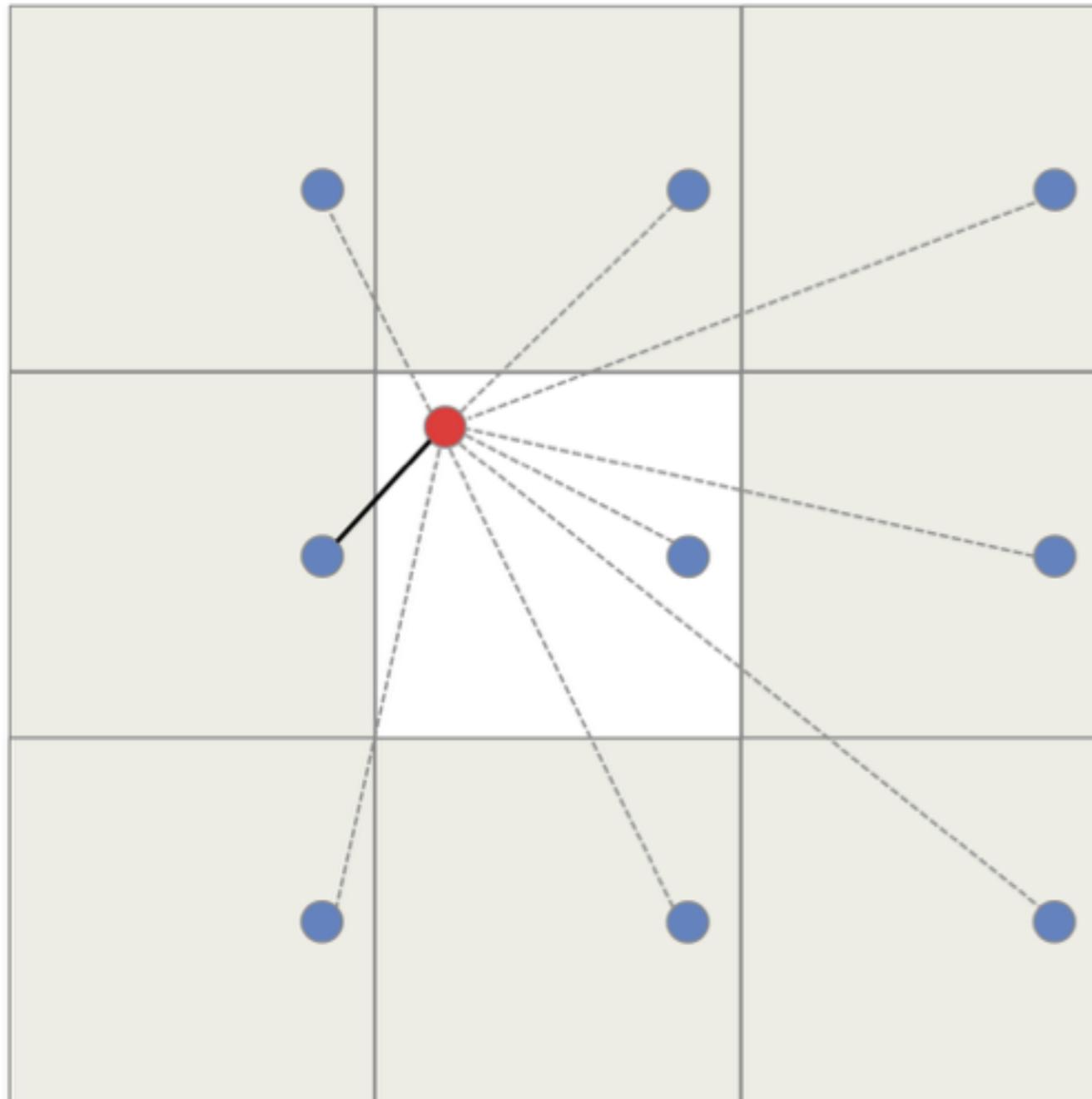
- in cubic box with 1000 particles, ~500 are close to wall
- not interested in surface effects



Solution: PBC

- unit cell is ‘replicated’ to form infinite lattice
- exiting particle enters on opposite face
- ok, if box length > interaction range

Minimum image convention

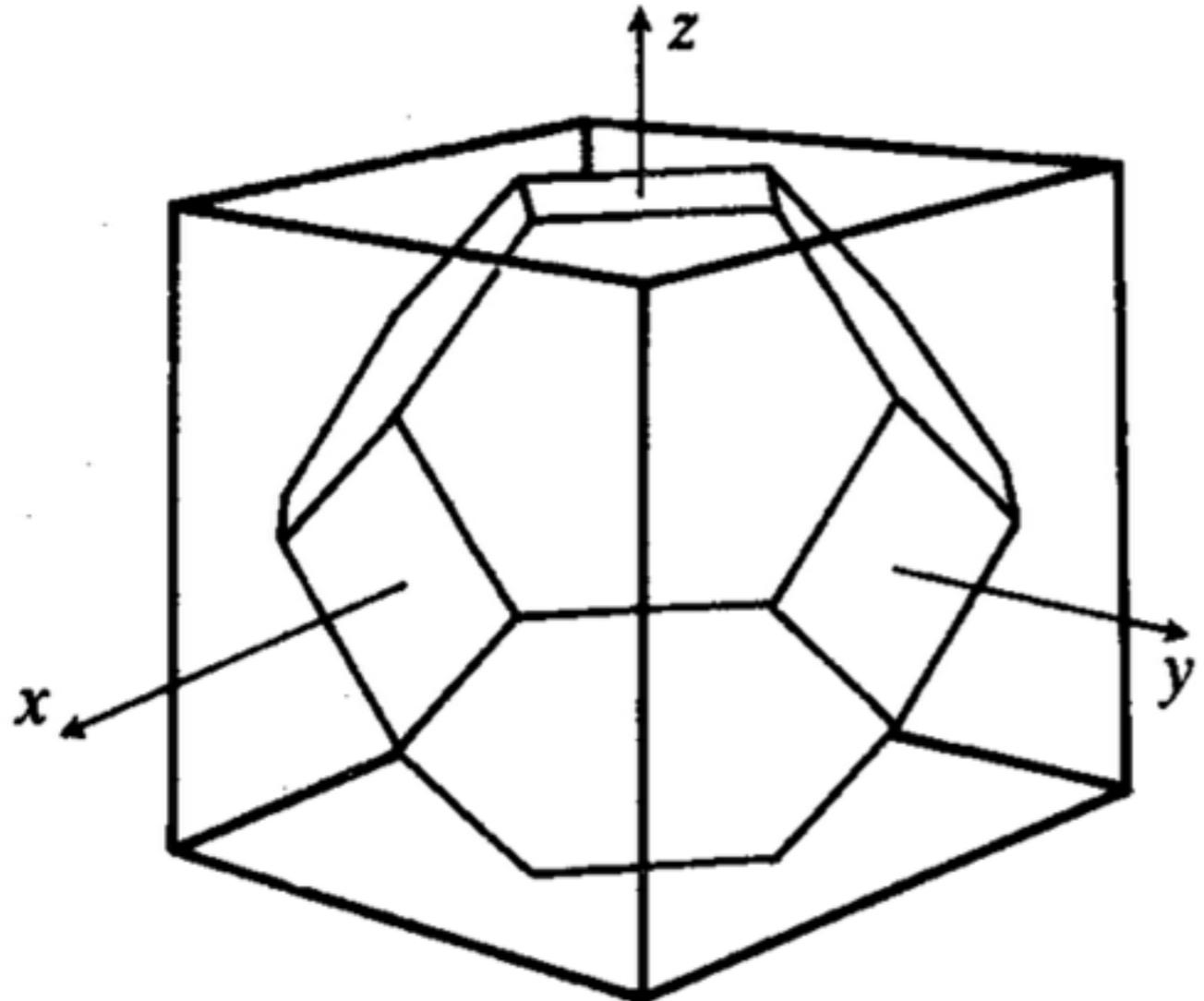


- Calculation of forces involves distance between all particle pairs
- Short-range potentials: consider only closest image

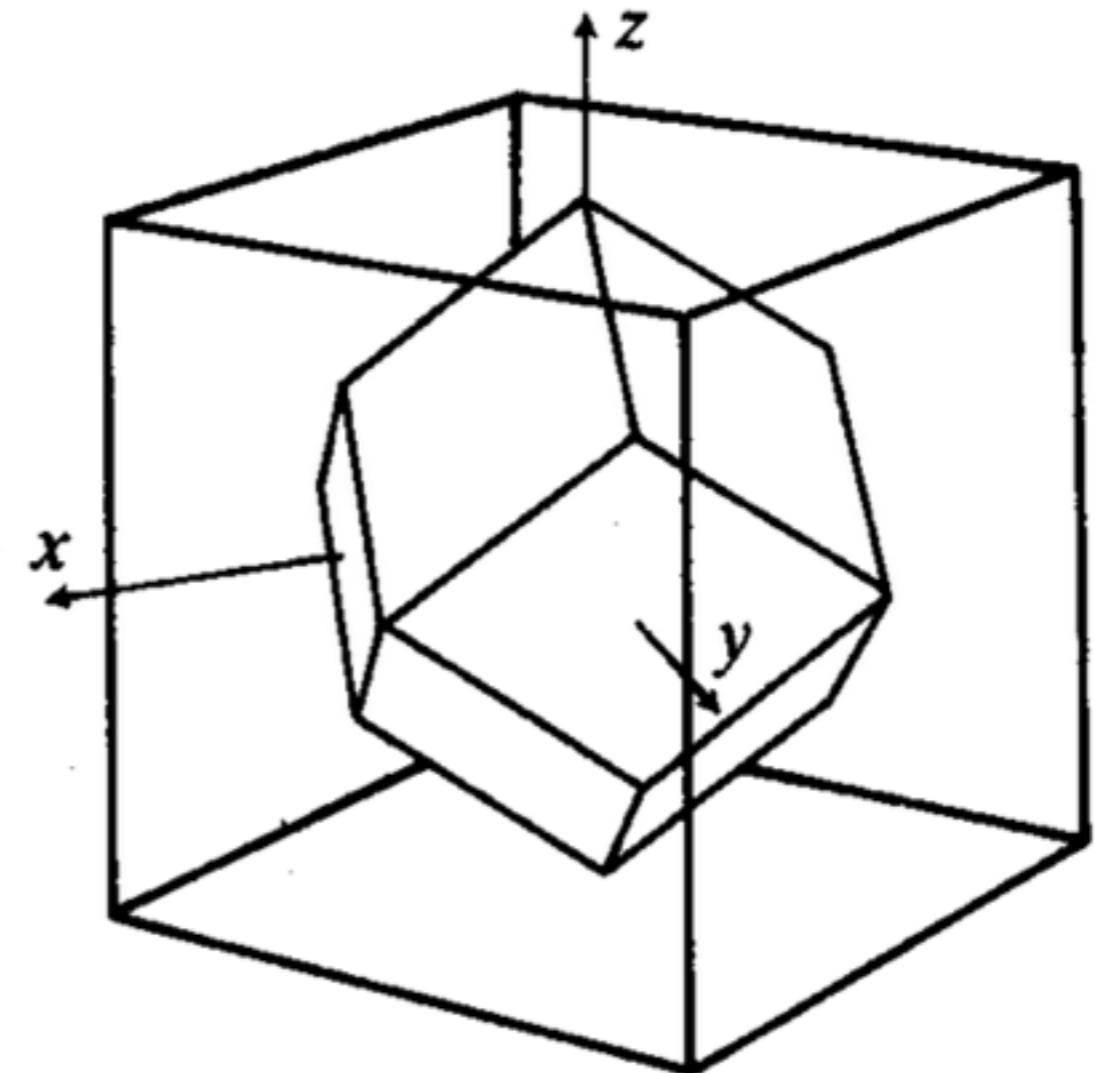
```
if( $r_x > L/2$   
     $r_x = r_x - L$   
if( $r_x < -L/2$   
     $r_x = r_x + L$ 
```

$$r_x = r_x - \text{floor}(r_x/L + 1/2)L$$

Simulation boxes



Truncated octahedron
(8 hexagons, 6 squares)



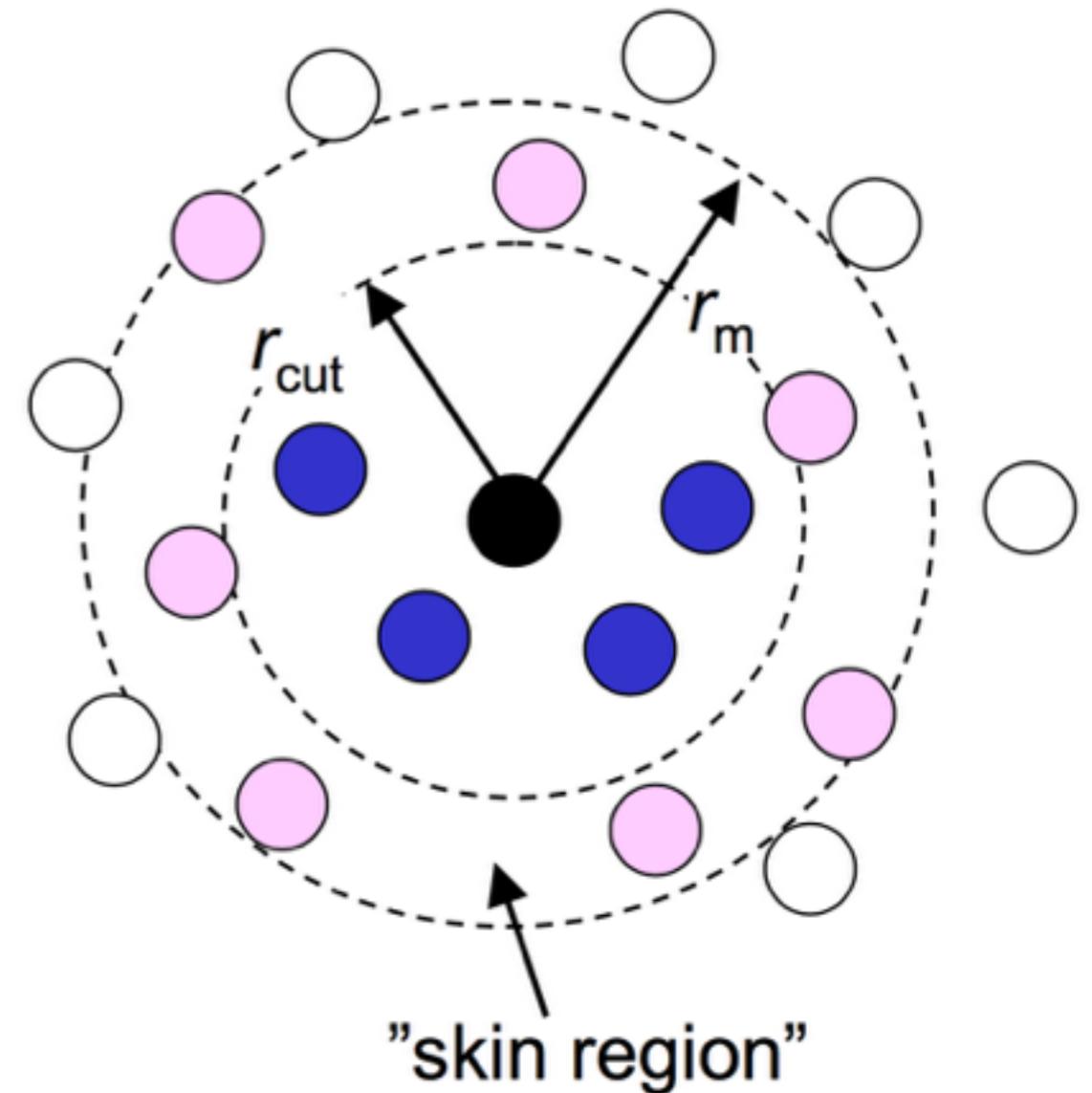
Rhombic dodecahedron
(12 congruent rhombic faces)

MD inner loop - force

```
for iatom=0... natoms-1
    for jatom=iatom+1...natoms
        # calculate distance vector
        d = position[jatom]- position[iatom]
        # apply PBC
        d_pbc = d - floor(d/L + 1/2)L
        # calculate distance (modulo squared)
        d_sq = d_pbc_x2+ d_pbc_y2 + d_pbc_z2
        # is distance larger than cutoff?
        if(d_sq < cutoff_sq)
            # calculate force between iatom and jatom
            f_ij = ...
            fi += fij
            fj -= fij
```

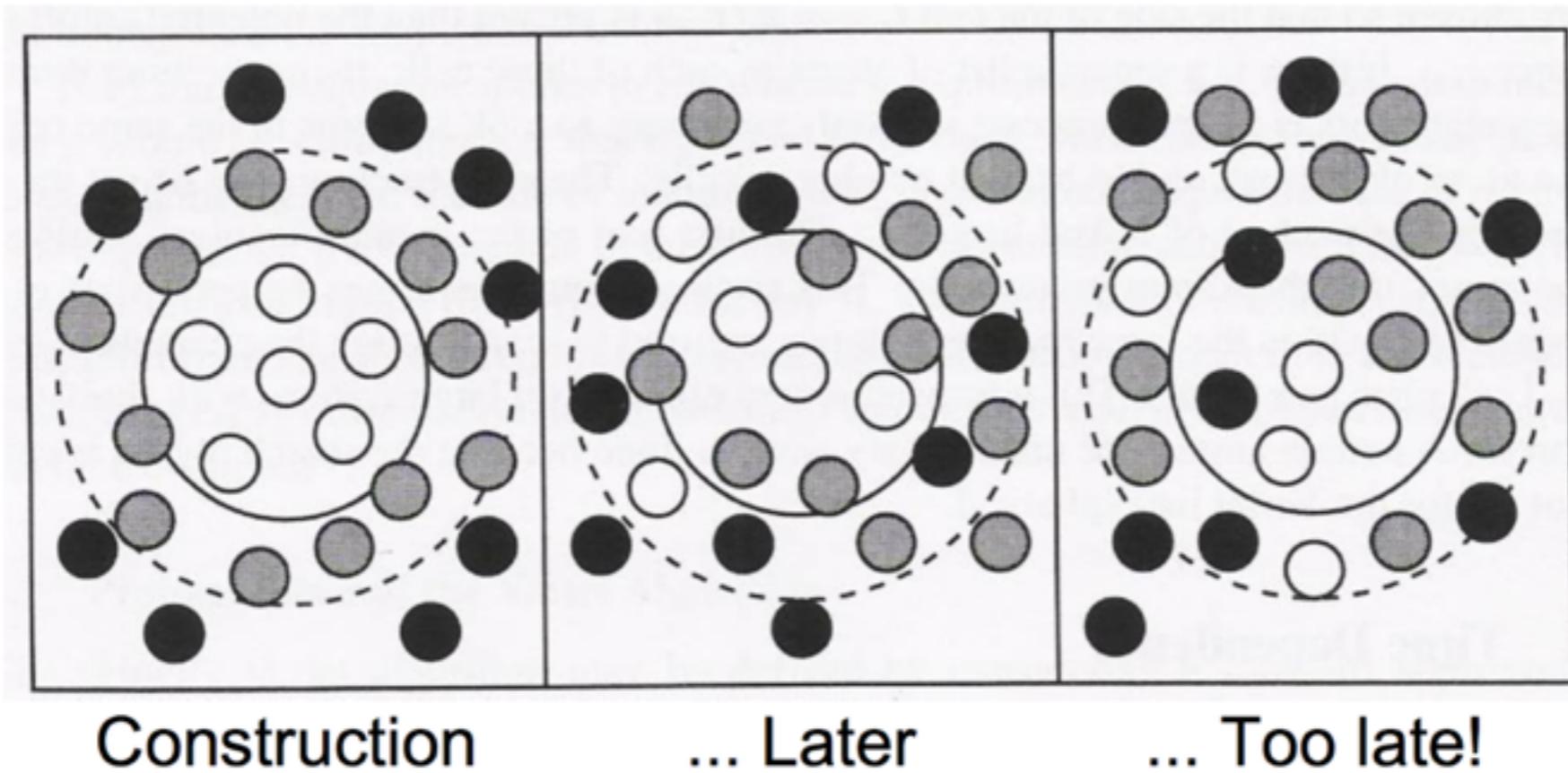
Verlet neighbor list

- Goal: list of neighbors within cutoff
- Create list of neighbors with $r < r_m$, with $r_m > r_c$
- Only atoms within list are used for force calculation



Verlet neighbor list

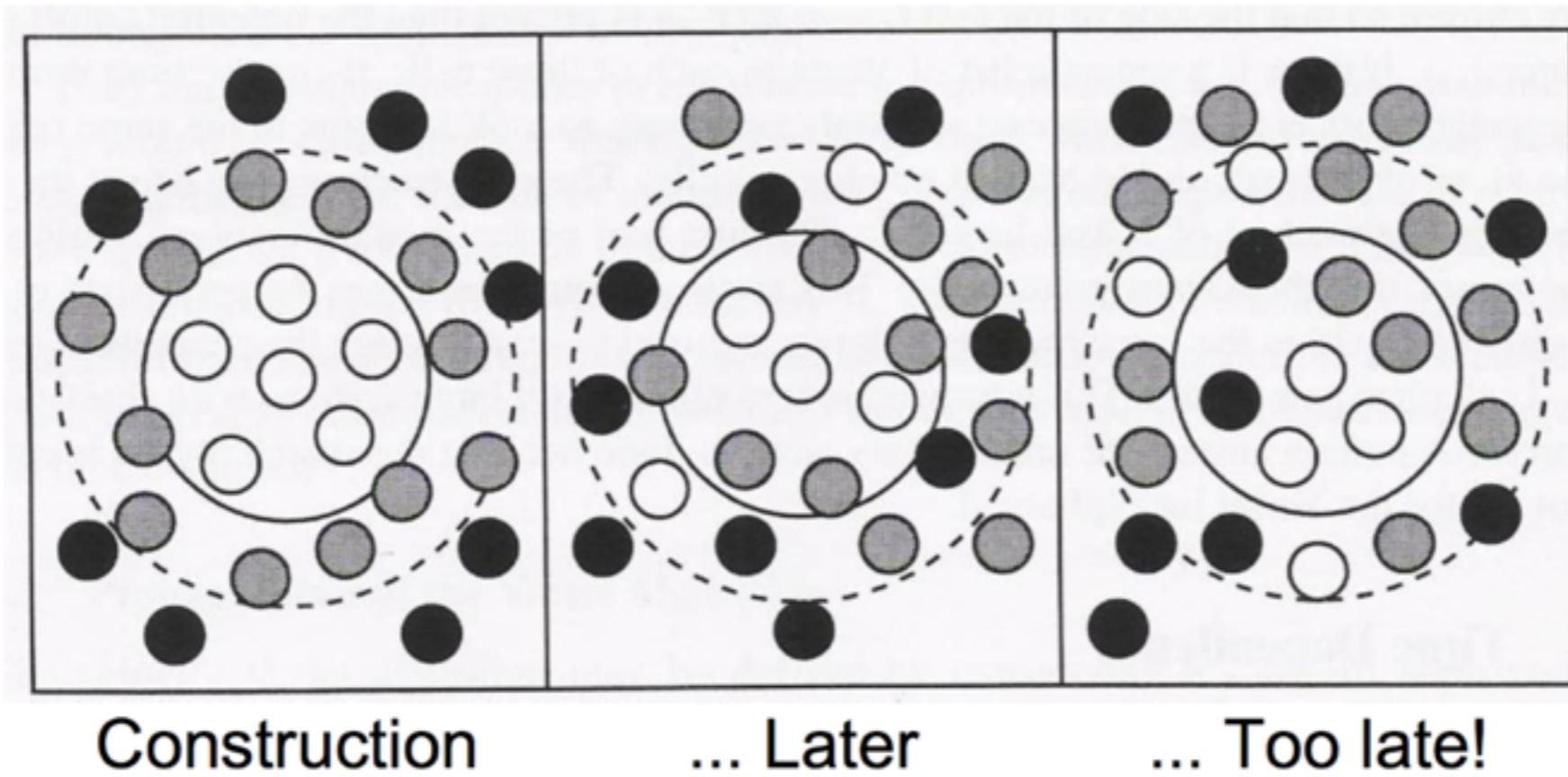
- Update list every n steps
- n and r_m are chosen such that $r_m - r_c > n\Delta t \cdot v$
- v typical speed of atom



Verlet neighbor list

- Automatic update
- Update when

$$|r(t_{update}) - r(t)| = \Delta r > (r_m - r_c)/2$$



MD outer loop

```
// Read initial positions
read_positions
// Set random velocities
randomize_velocities

// Compute initial neighbor list
compute_list
// compute initial forces
compute_forces

for(istep=0; istep<nstep; istep++) {

    thermostat
    velocity += (force*dt/2)/mass
    position += velocity*dt

    // Check whether the neighbour list has to be recomputed
    check_list

    if(recompute_list) {
        compute_list
    }

    compute_forces

    velocity += (force*dt/2)/mass
    thermostat
}
```