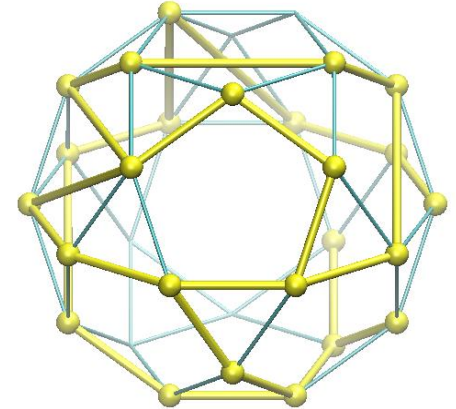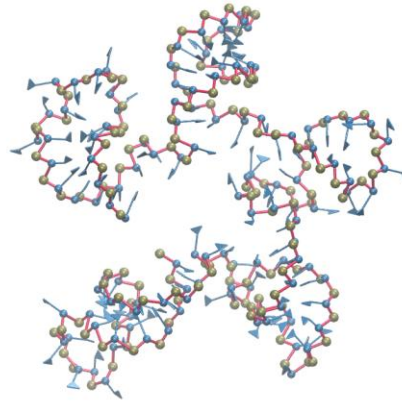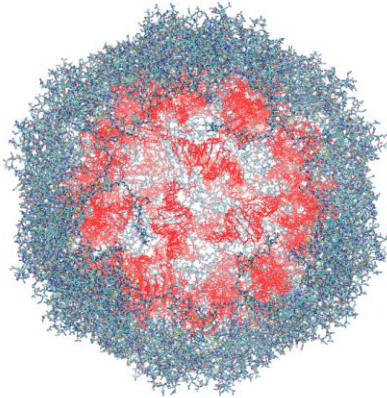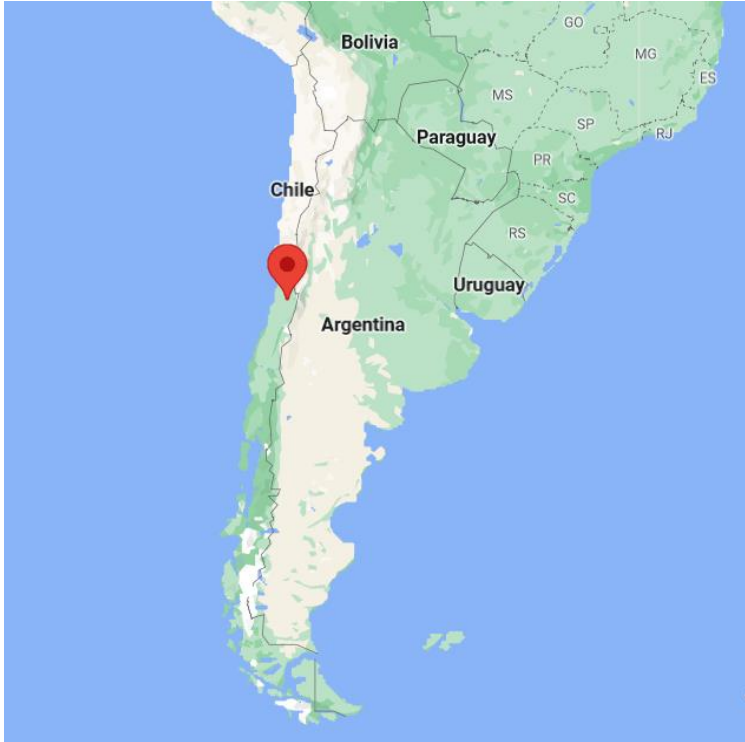# Modeling RNA structure using multiscale simulations
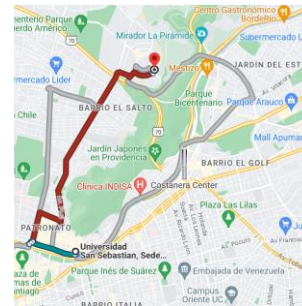
Simón Poblete
Fundación Ciencia y Vida – Universidad San Sebastián - Chile
Seminar on Mechanics, Institute of Fundamental Technological Research
Polish Academy of Sciences. Warsaw, Poland, 2023.

A bit of my working place

Fundación Ciencia y Vida / Universidad San Sebastián

Outline

- Introduction : computational modelling

- RNA and its complex structure

- The SPQR model

- Fixing structures

- Viral modeling
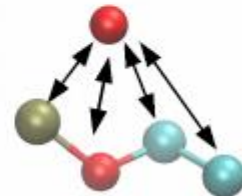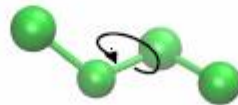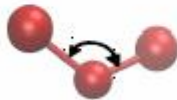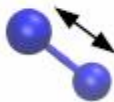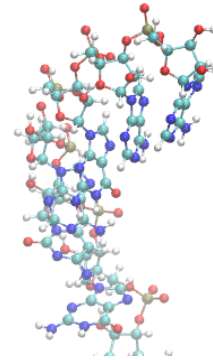
Computational modeling

Molecular Dynamics simulations : accurate, detailed

$$m_i \ddot{\vec{r}}_i = \sum_j F_i$$

$$F_i = -\nabla_i U(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$$
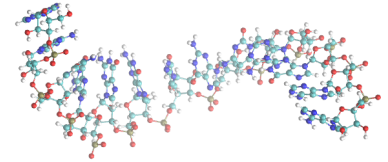
$$U(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) =$$

$$\sum_{bonds} k_b (l - l_0)^2 \;\; + \sum_{angles} k_a (\theta - \theta_0)^2 \;\; + \sum_{torsions} \sum_n \frac{1}{2} U_n [1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \{ \epsilon_{ij} [(\frac{r_{0ij}}{r_{ij}})^{12} - 2(\frac{r_{0ij}}{r_{ij}})^6]\} + \frac{q_i q_j}{4 \pi \epsilon_0 r_{ij}}$$
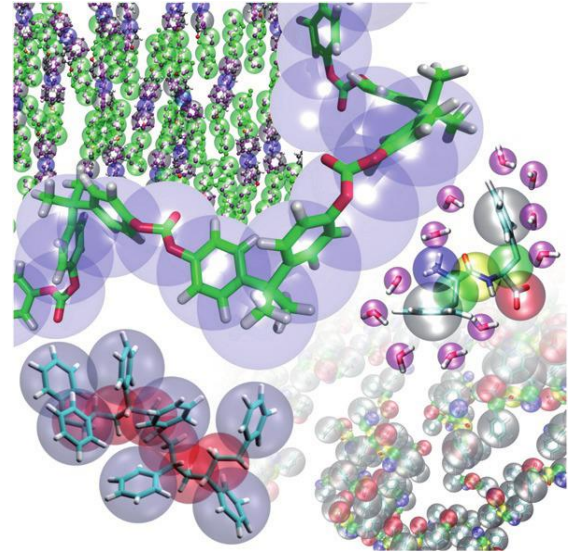
Computational modeling

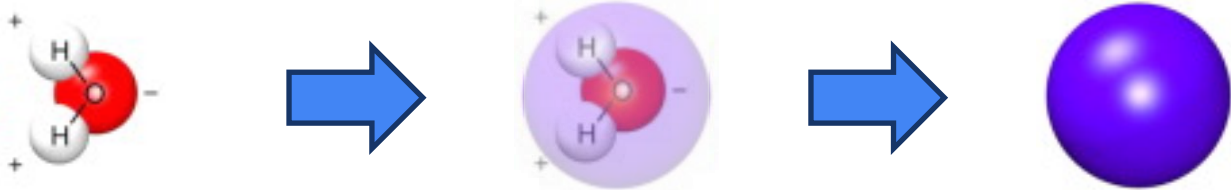Molecular Dynamics simulations : accurate, detailed

COARSE-GRAINED MODELS:
Faster, more generic, taylored for a specific problem.

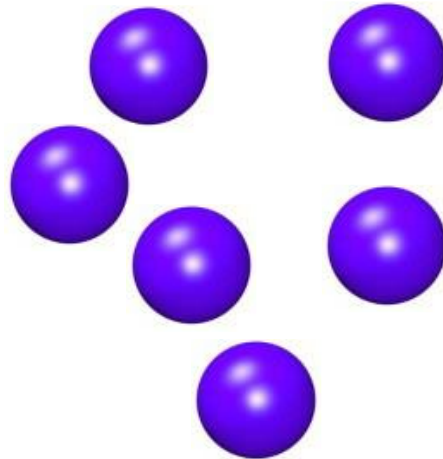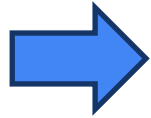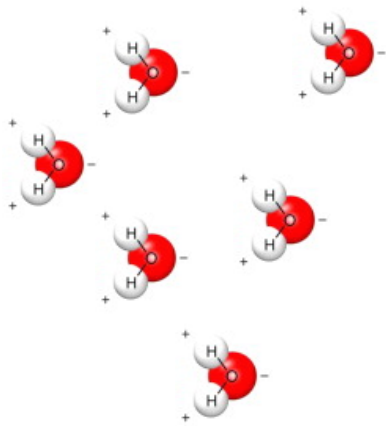# Computational modeling

COARSE-GRAINED MODELS:
Faster, more generic, taylored for a specific problem.
e.g. a water molecule

Computational modeling

COARSE-GRAINED MODELS:
Faster, more generic, taylored for a specific problem.
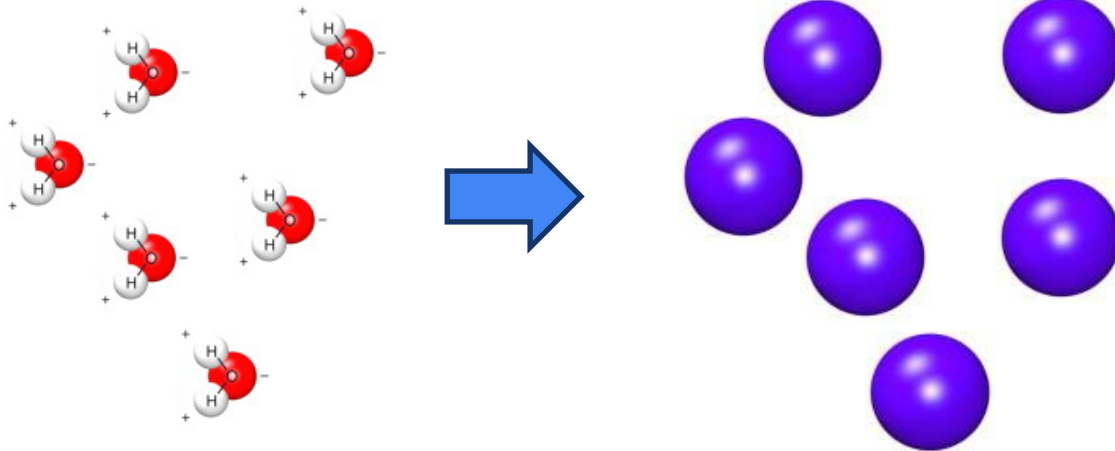e.g. a water molecule

Is this water??

Computational modeling

COARSE-GRAINED MODELS:
Faster, more generic, taylored for a specific problem.
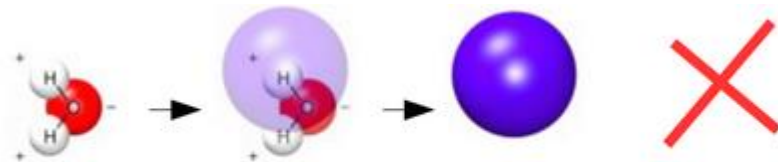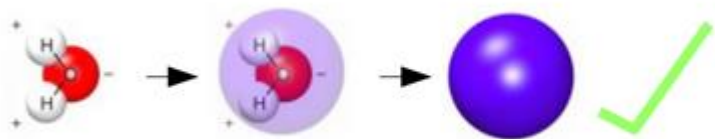e.g. a water molecule

Is this water??

We must develop interactions acordingly : diffusion, structure, thermodynamics, etc.

Coarse-grained simulations

Reduction of number of degrees of freedom
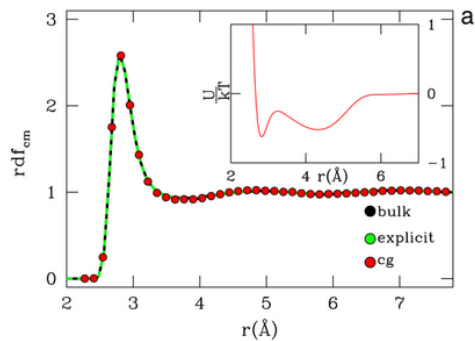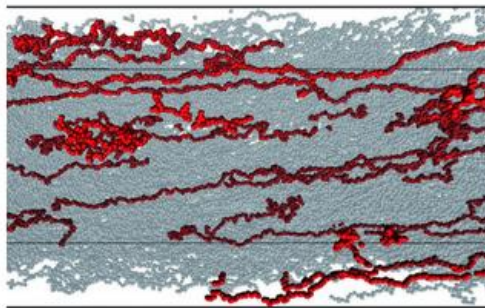
How to map the system?

Coarse-grained simulations
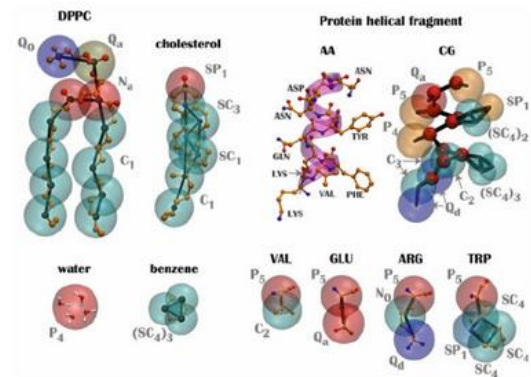
Reduction of number of degrees of freedom

How to define the interactions?

Structure (constant T)

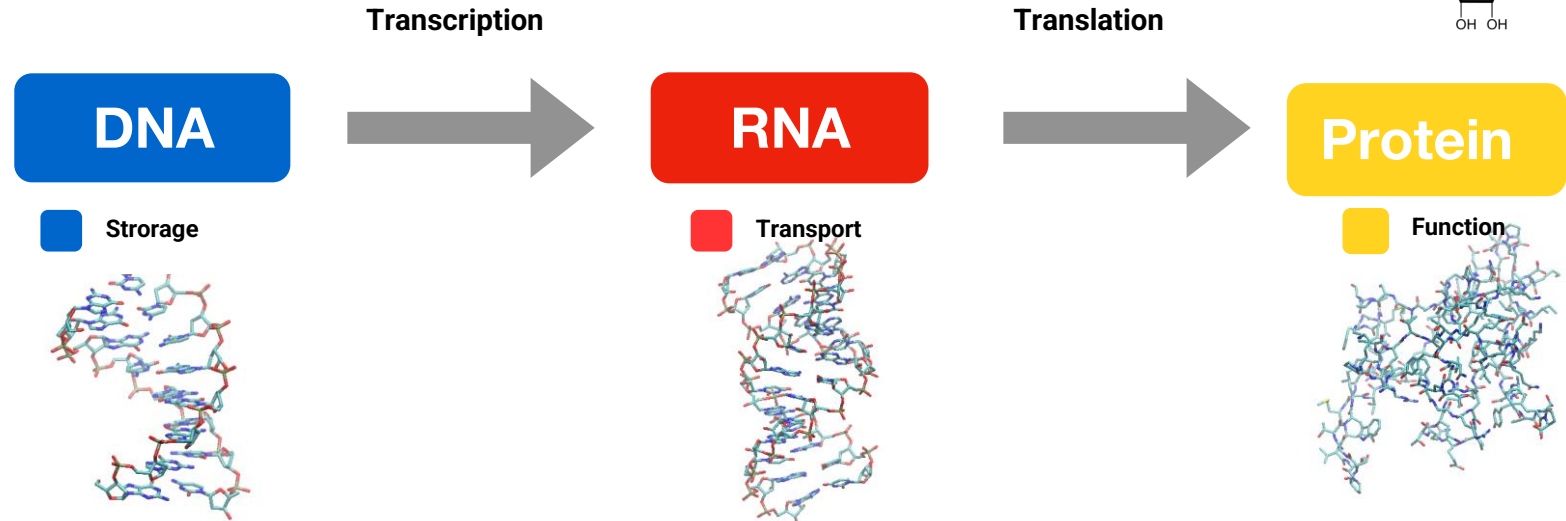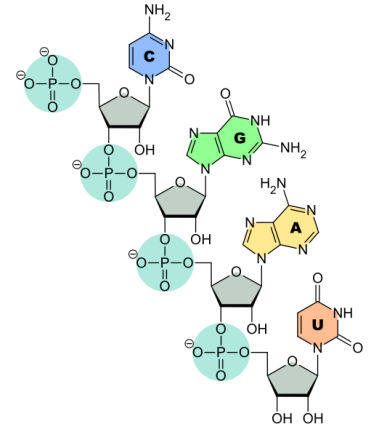**Transport**

Thermodynamics

Etc…PROBLEM DEPENDENT

What is RNA?

- Biopolymer(ribose,phosphate,nitrogenous base)

- 4 letter alphabet: A, C, G, U

Transcription
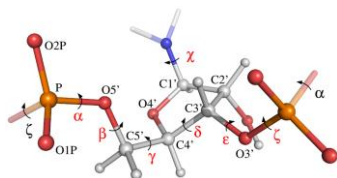
Translation

DNA

Strorage

RNA

Transport

Protein

Function

What is RNA?

- Biopolymer(ribose,phosphate,nitrogenous base)

- 4 letter alphabet: A, C, G, U

Transcription (80%)

DNA → RNA

Retrotranscription

Translation (1%)

RNA → Proteins
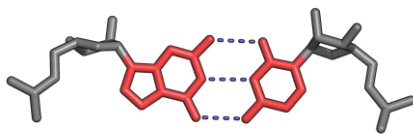
Storage

Transport

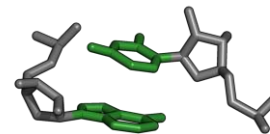Function

Function

Function
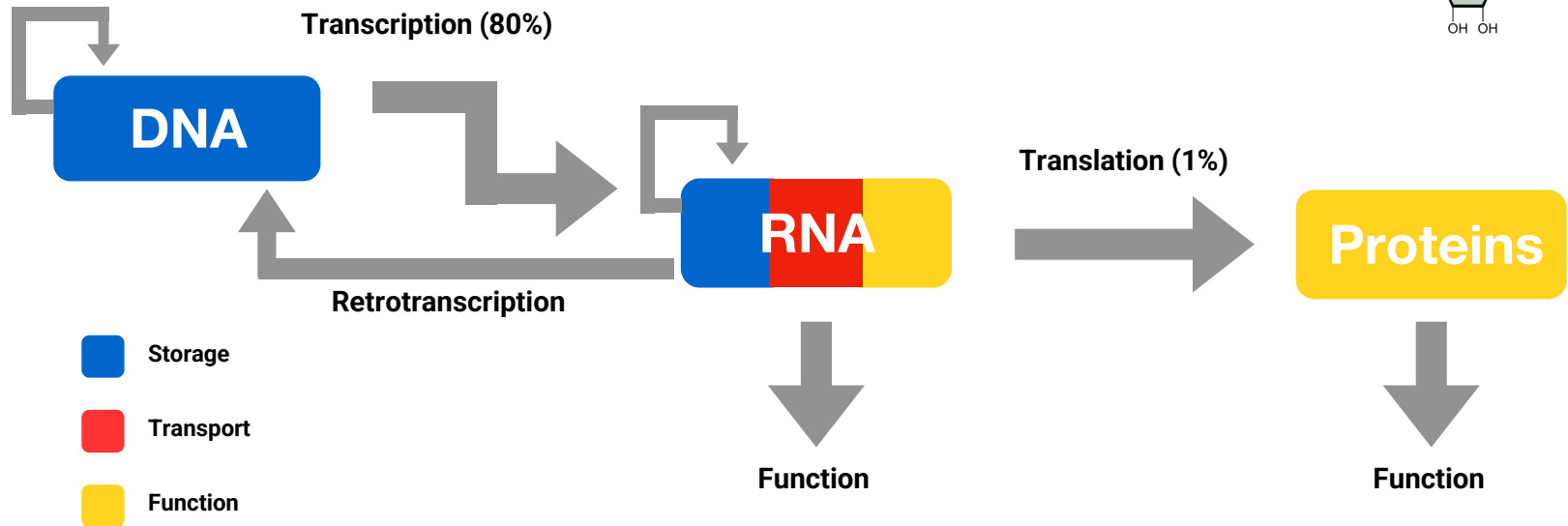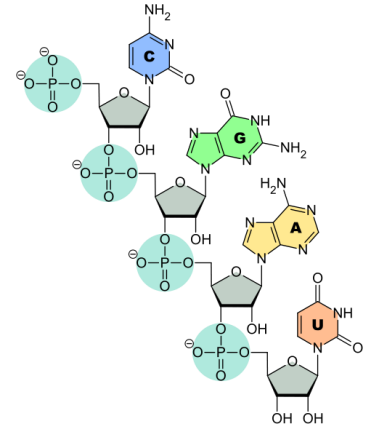
What is RNA?

Structural Complexity

7 dihedrals
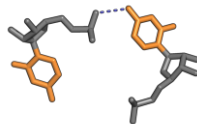
Canonical pair A-U, C-G

Stacking
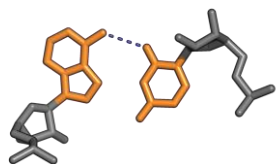
Non-canonical
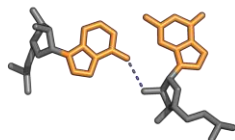
Base-ribose

Base-phosphate

Triplex

Quadruplex

What is RNA?

Structural Complexity

Structure and function depend on the sequence

Primary structure   Secondary structure   Tertiary structure (3D)

GGGCGCAAGCCU

GGCGCGGCACCGUCCGCGGAACAAACGG

Persistence length : 1nm (ss)-62nm(ds)
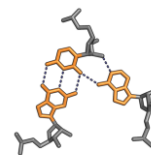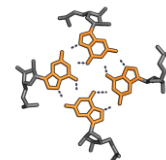
Coarse-grained model

SPlit and conQueR (SPQR)

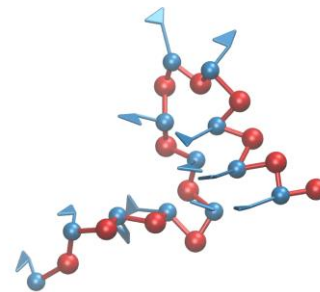Nucleotide-level representation

Nucleoside
Phosphate
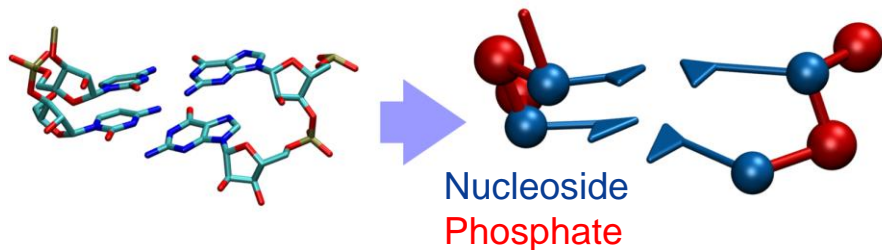
Essential conformations and interactions:
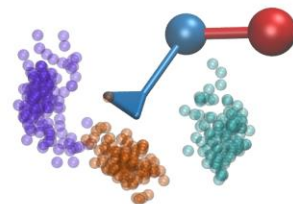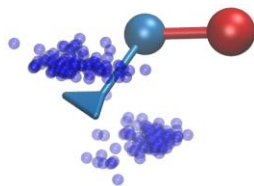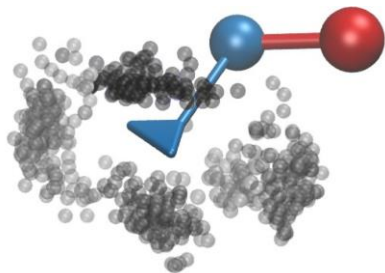stacking, EV, base pairing, base-phosphate

GGGCGCAAGCCU

Monte Carlo simulation.
Developed for structure
prediction

Poblete, Bottaro and Bussi
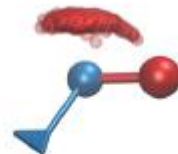Nucleic Acids Res. 2018

Coarse-grained model

SPlit and conQueR (SPQR)

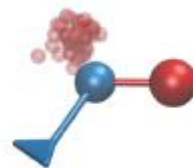Sampling from a large structure database

Identify stacking and base pairs…
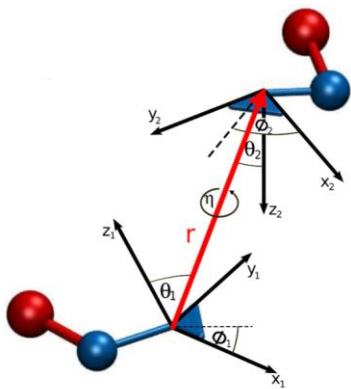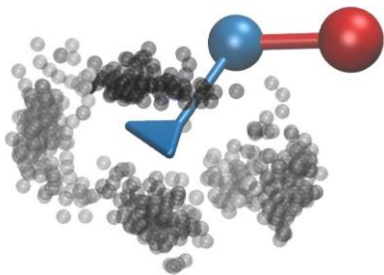
…and backbone conformations.

Pucker C3', endo   Pucker C2', endo   Pucker C3', syn

Coarse-grained model

SPlit and conQueR (SPQR)

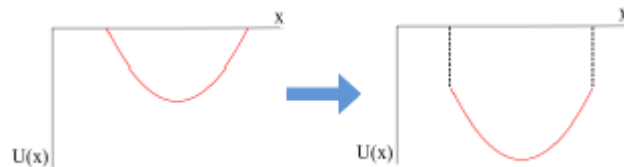$$U(\vec{r}) = -k_B \log P(\vec{r})$$

P(r) sampled from Structural Database *

$$P(r, \theta_1, \phi_1, \theta_2, \phi_2, \eta) \approx r^2 P_\eta(\eta) \frac{P_1^{\mathcal{E}}(r, \theta_1, \phi_1) P_2^{\mathcal{E}}(r, \theta_2, \phi_2)}{P_r(r)}$$
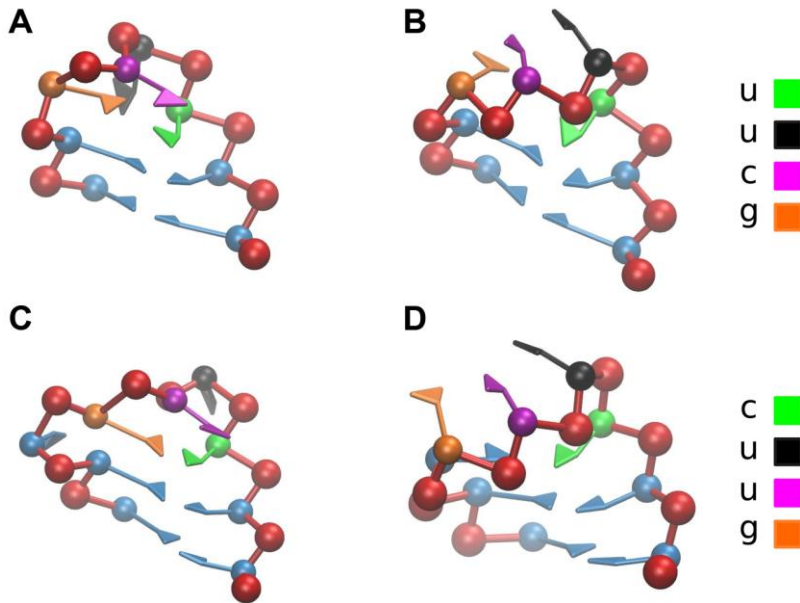
$$U_a(\vec{r}) \rightarrow U_a(\vec{r}) - \epsilon_a$$
$$U_{pb}(\vec{r}) \rightarrow U_{pb}(\vec{r}) - \epsilon_{pb}$$

Coarse-grained model

SPlit and conQueR (SPQR)

Loops require freedom on pucker and glycosidic bond angle to adopt correct structure.
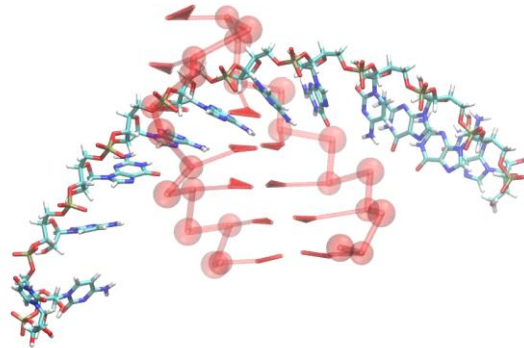
Coarse-grained model

SPlit and conQueR (SPQR)

ERMSD steering : minimize the ERMSD distance between predicted and all-atom structure.
ERMSD depends both on relative positions and orientations of nucleobases.
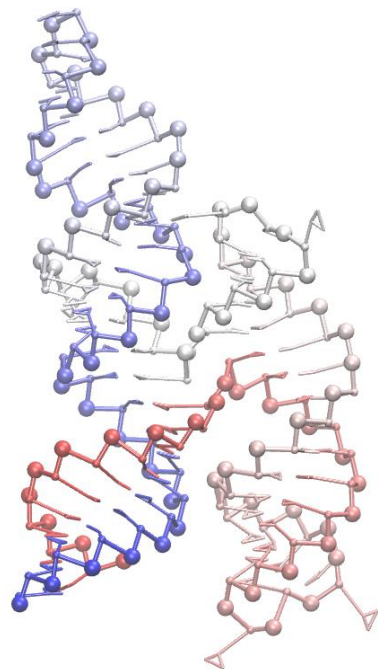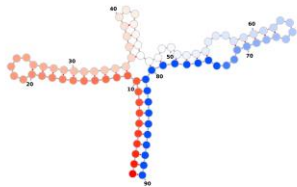
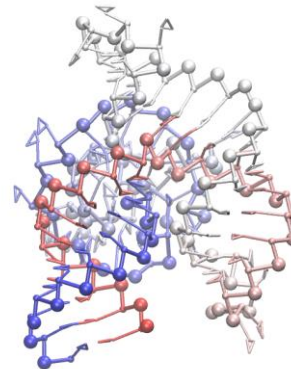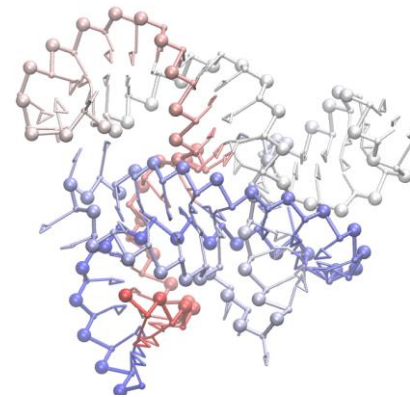$$U = \frac{1}{2} K \, (ERMSD)^2$$

Structure prediction

Sampling orientations of stems
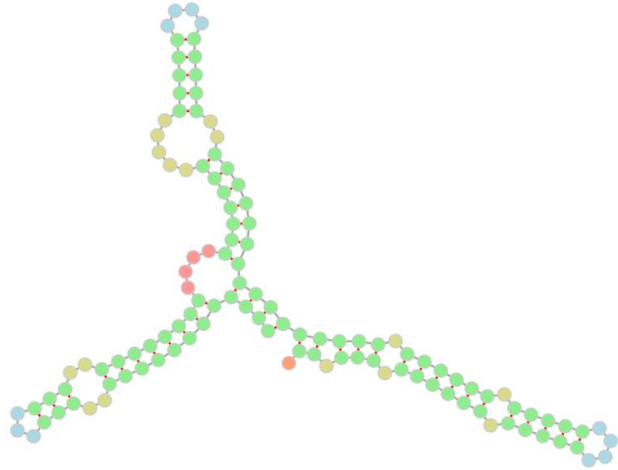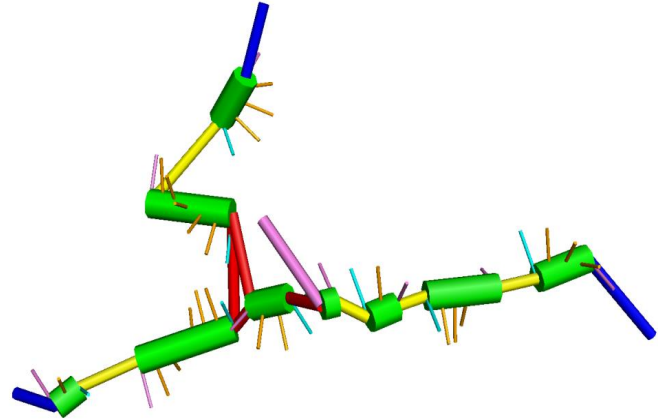
Example : PZ39

TARGET

SAMPLES

Structure prediction

Ernwin
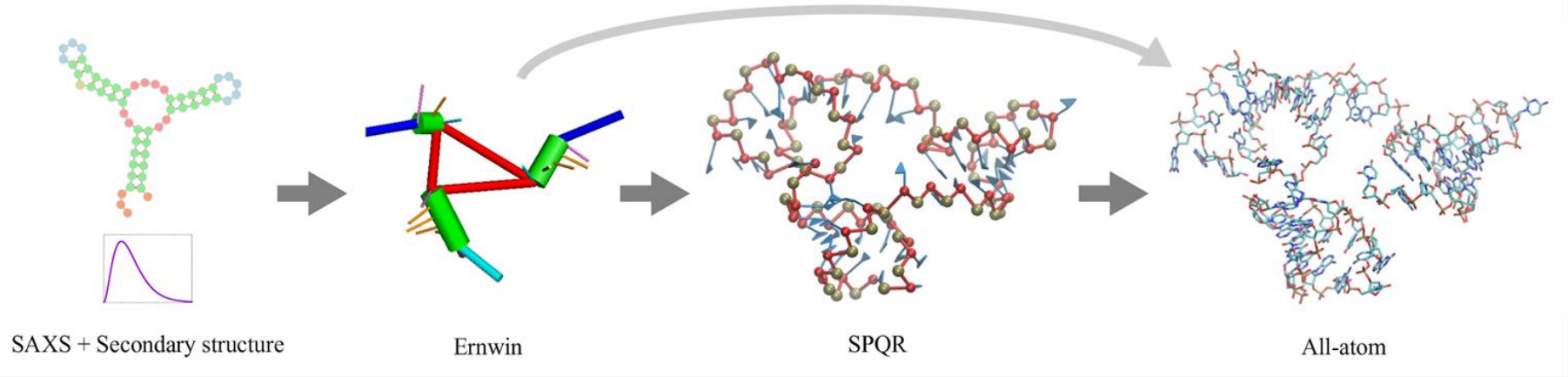

Secondary structure elements.
Knowledge-based interactions.
A-minors and minimize radius of gyration.

Kerpedjiev et al., RNA 2015
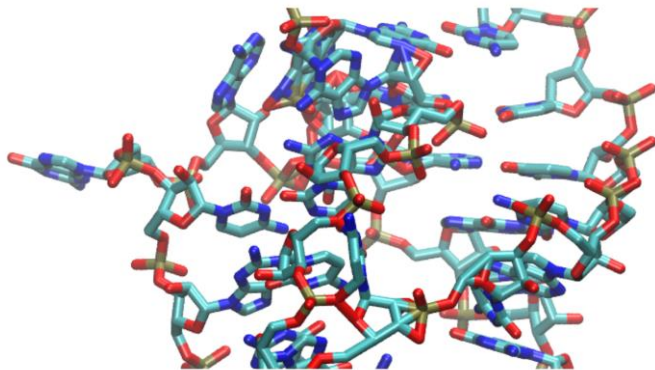
Structure prediction
Multiscale modeling

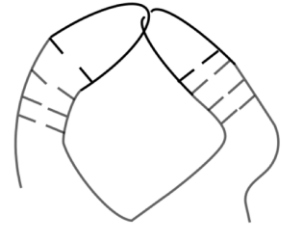SAXS + Secondary structure   Ernwin   SPQR   All-atom
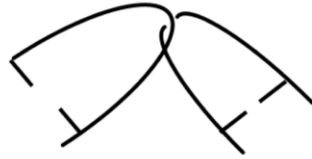
Structure prediction

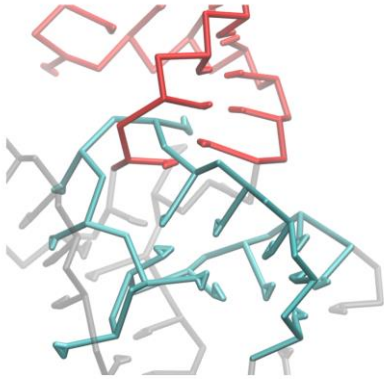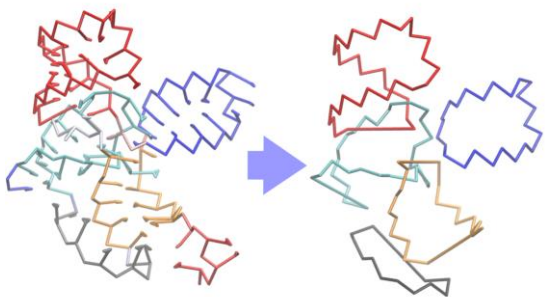Fragment assembly produces clashes!

Structure prediction

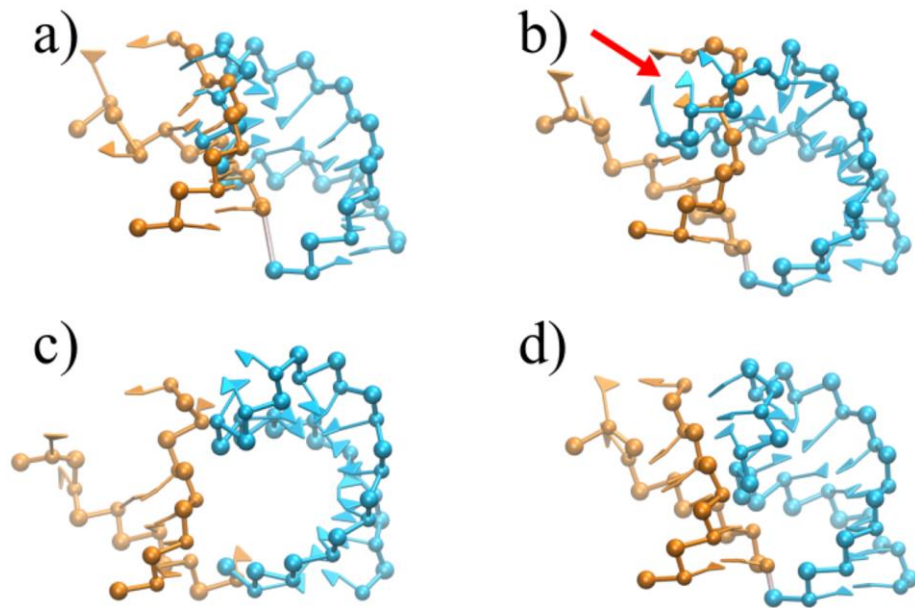Links between secondary structure elements!

Structure prediction

Energy and topology refinement

$$L_N(c_i, c_j) = \frac{1}{4\pi} \oint_{c_i} \oint_{c_j} \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \cdot (d\mathbf{r}_i \times d\mathbf{r}_j)$$

Thiel et al., submitted
Popenda et al. 2021
Nyemiska et al. 2020

Structure prediction

Test on 4 deposited structures and one SAXS profile. 30 decoys each.

Structure	Length	Linked	RMSD (A)*	Contacts	Clash Score
1L9A	126	12	3.2 ± 0.9	30/95	128/5
2R8S	120	23	4.3 ± 1.1	69/128	128/7
3R4F	66	7	2.0 ± 0.2	0/0	90/4
4PQV	68	16	2.5 ± 0.8	5/13	118/6
SASDK34*	118	10	2.0 ± 0.8	-	115/6

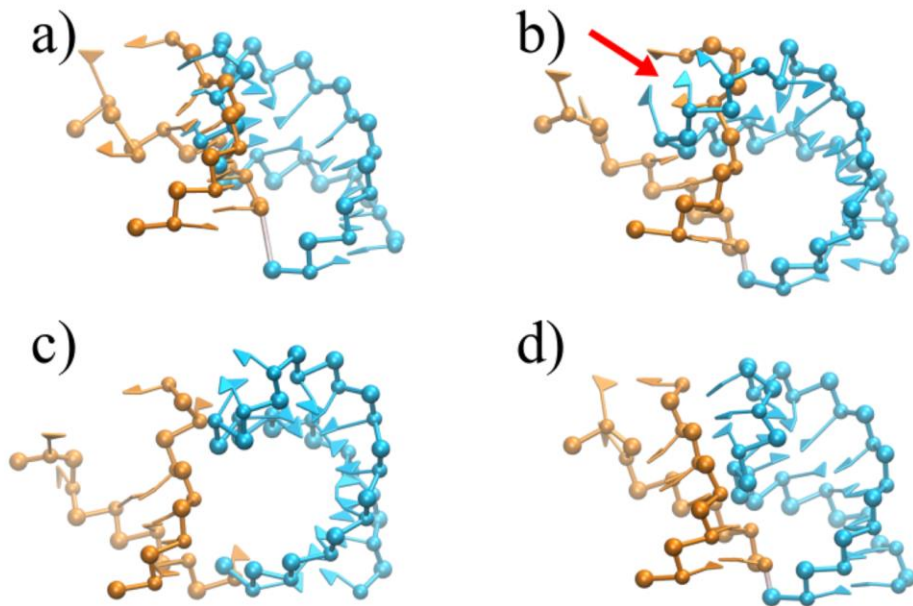Refinement does not affect significantly the SAXS profile agreement.

Refinement procedure

- Energy minimization.
- Link removal
- ERMSD steering towards initial structure
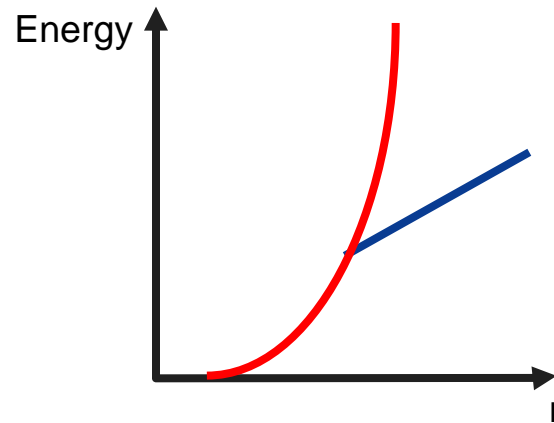- Energy minimization
- MD energy minimization
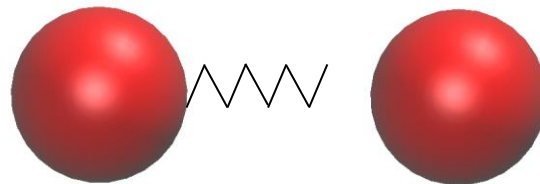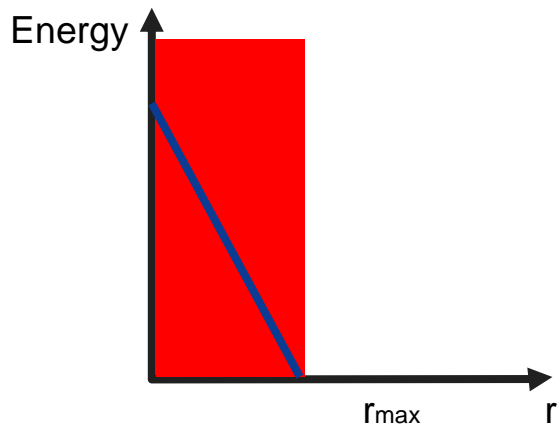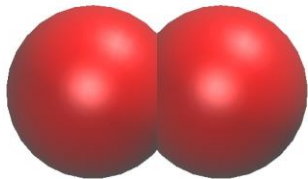
Install SPQR (Mac or Linux).

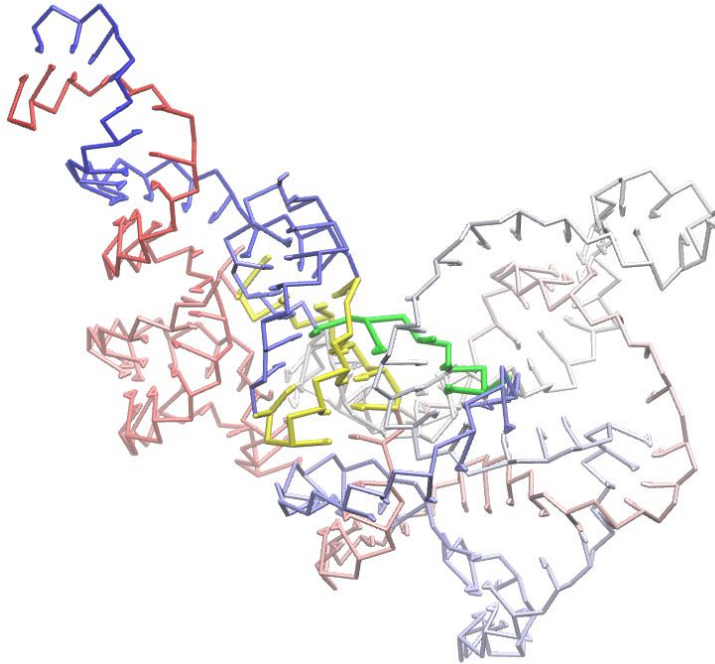Run

```
./SPQR_REFINE -t ss.fa -i struct.pdb
```

Energy minimization

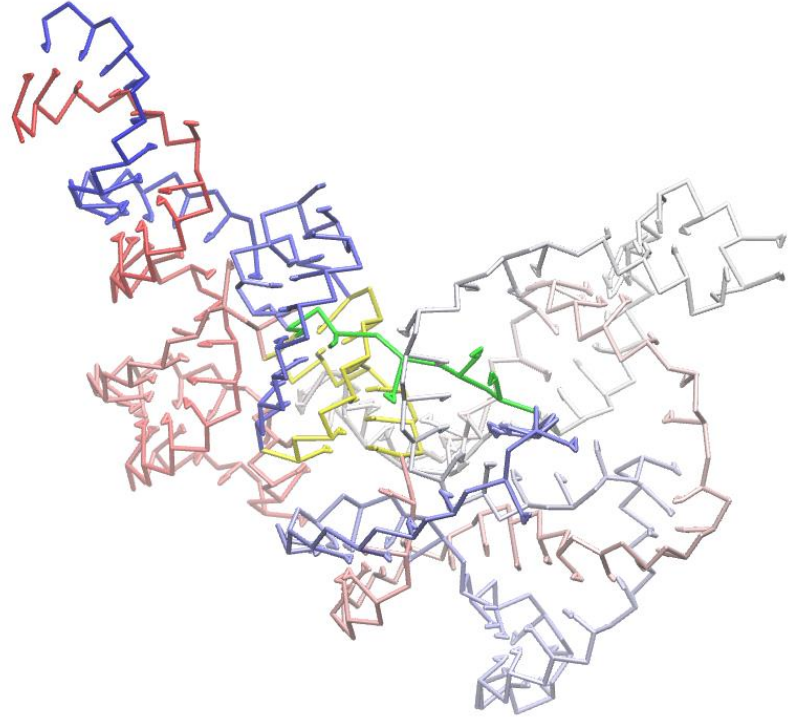Clashes and broken bonds

Link removal

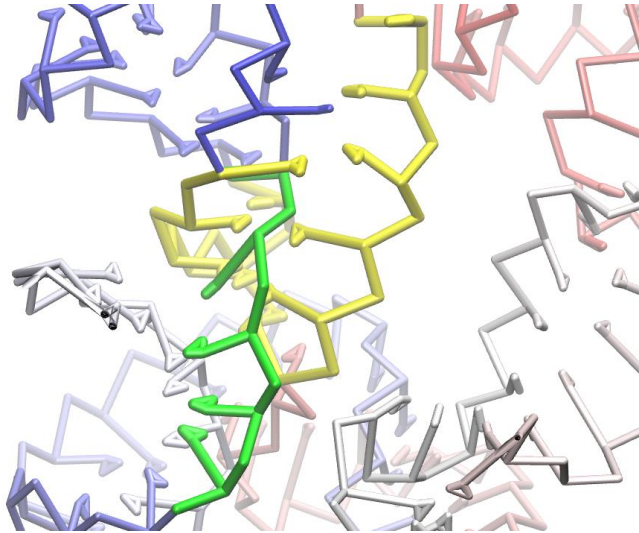In RNA puzzles : PZ05A1
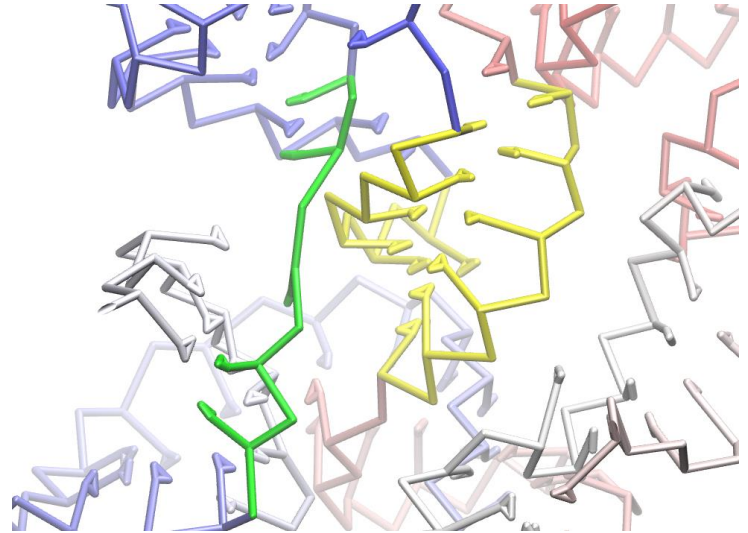

Linked   Unlinked

Link removal

In RNA puzzles : PZ05A1
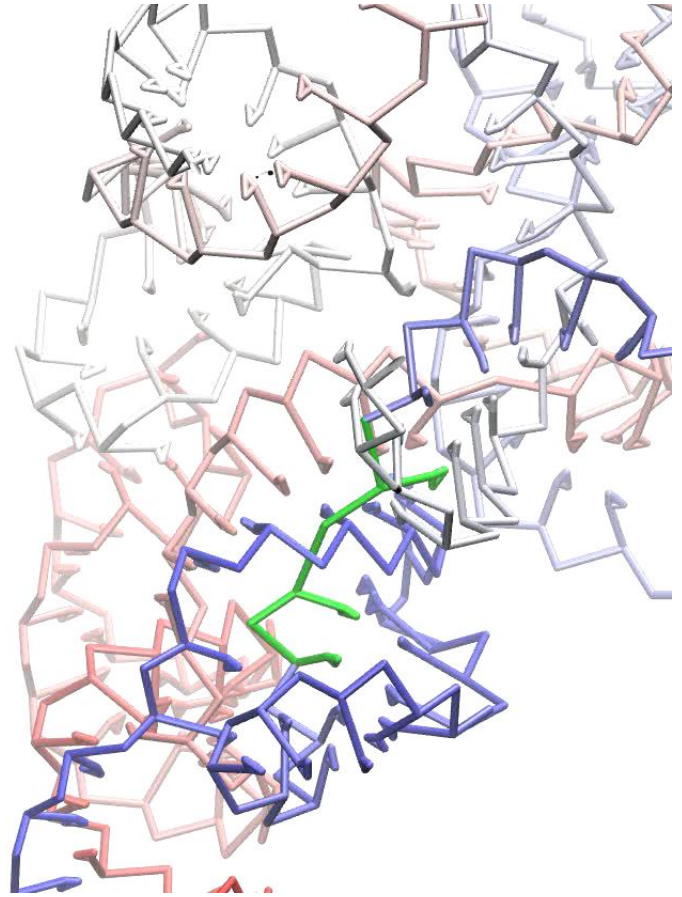

Linked   Unlinked

Link removal
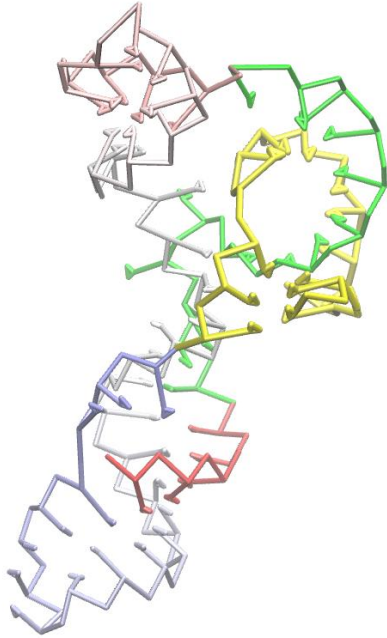
In RNA puzzles : PZ05A1

(most relevant part highlighted)
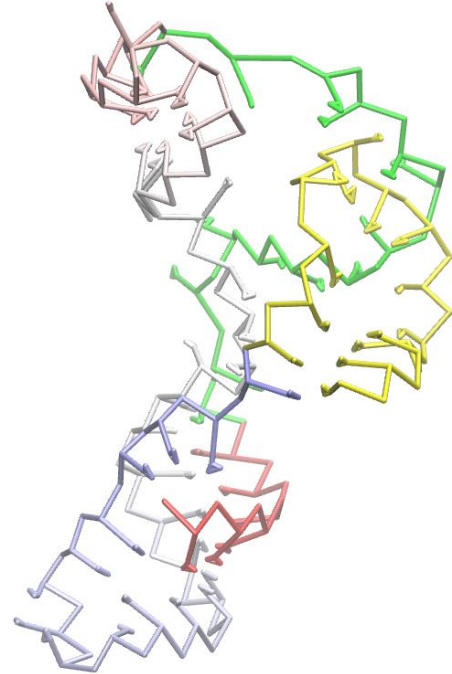
Link removal

In RNA puzzles : PZ13D8

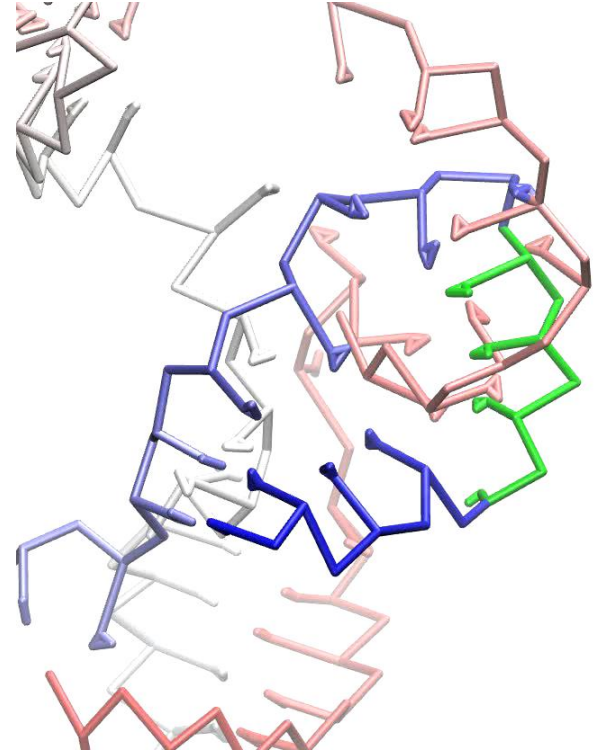Linked   Unlinked
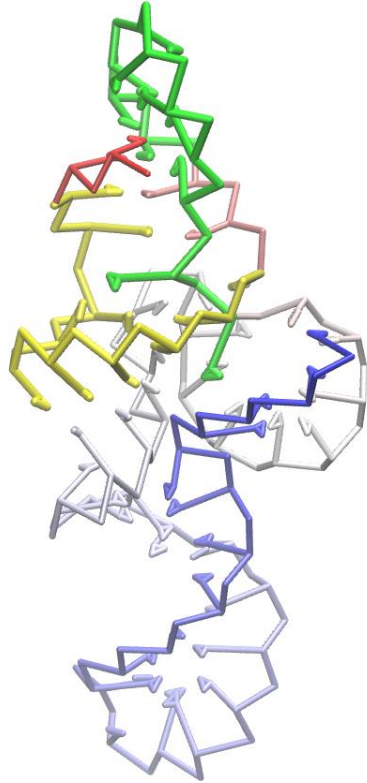
Link removal
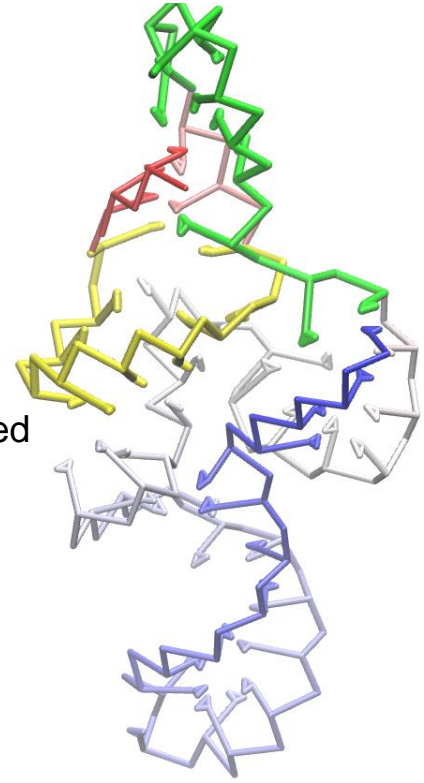
In RNA puzzles : PZ13D8

(most relevant part highlighted)

Link removal
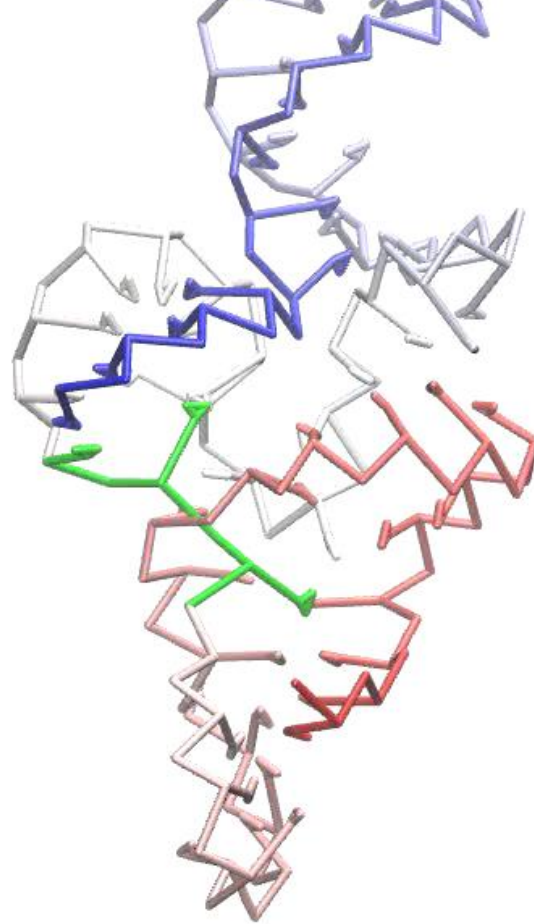
In RNA puzzles : PZ17C5

Linked

Unlinked

Link removal

In RNA puzzles : PZ17C5

(most relevant part highlighted)

Viral genome modelling

- dsDNA : double strand pumped inside

- ssRNA: Collaborative process.

Perlmutter et al. 2014
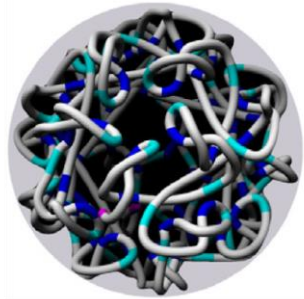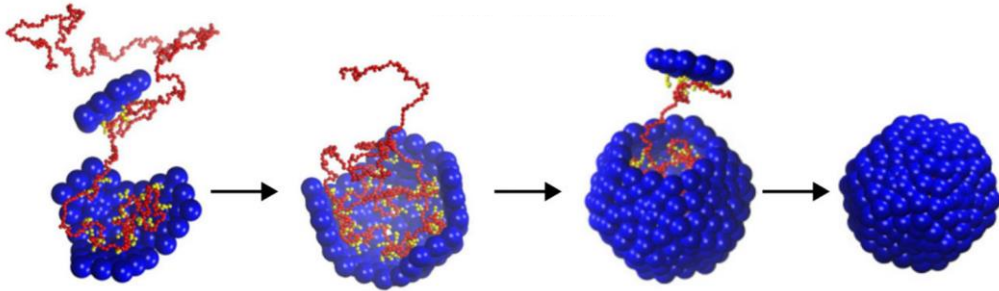
Viral genome modelling

SATELLITE TOBACCO MOSAIC VIRUS (STMV)
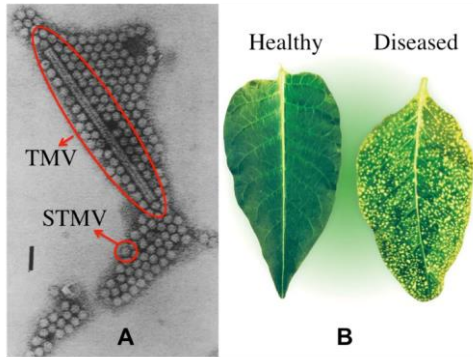
Small virus (17 nm, 1058 nt)

Known genome

Secondary structure determined both in virio and in vitro.

60% of its structure known by X-ray spectroscopy - 30 double helices of 9 base pairs

Viral genome modelling
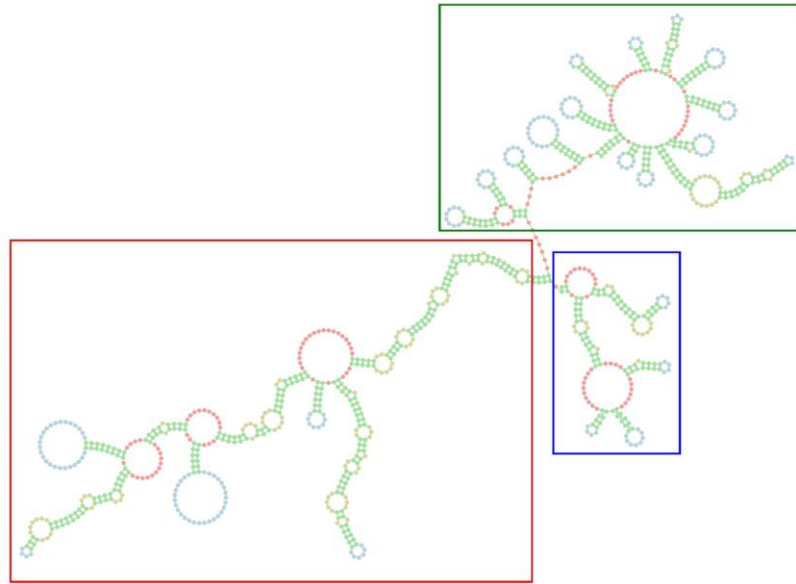
Modelling a structured system

Brute-force compression

Secondary structure
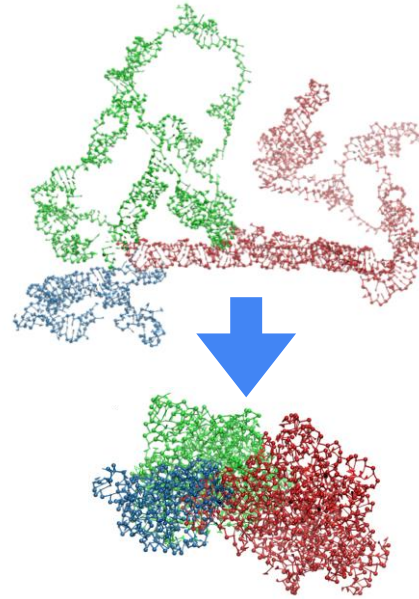
Nucleotide representation

Viral genome modelling

Modelling a structured system

Secondary structure
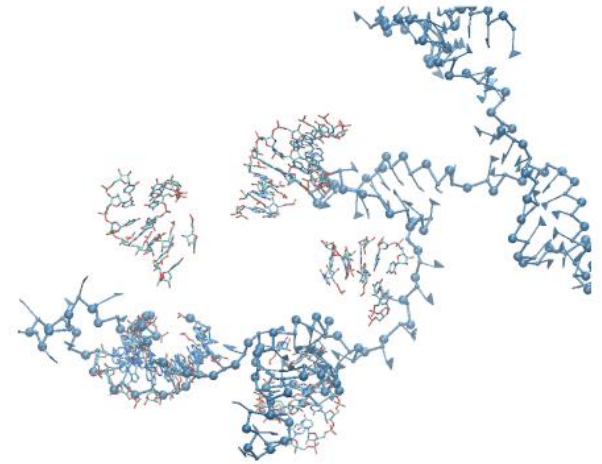
ERMSD pulling

Viral genome modelling

Use X-ray fragments as scaffold
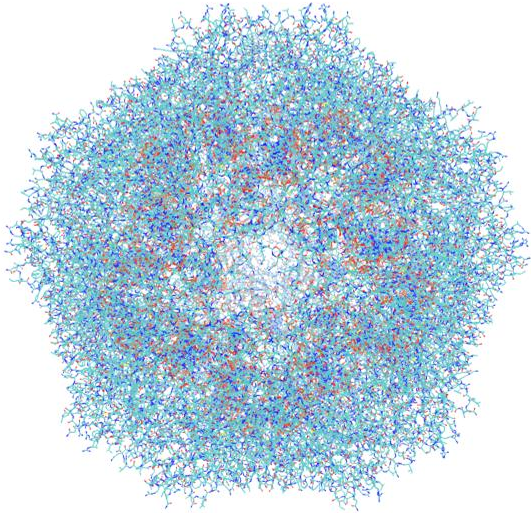
Secondary structure
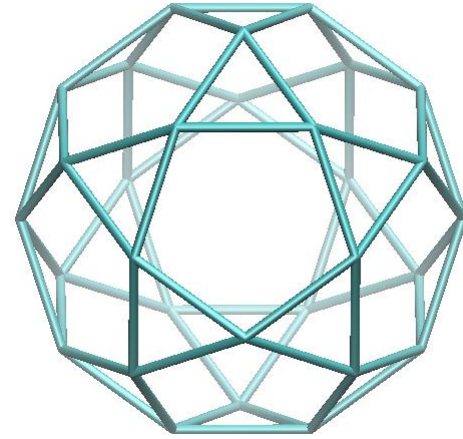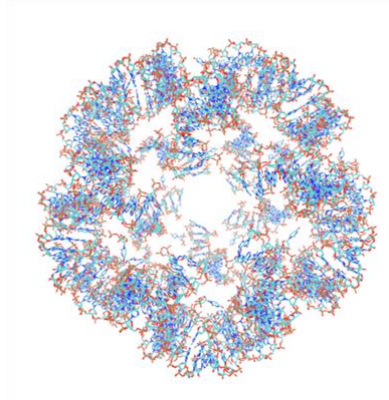
Lattice polymer

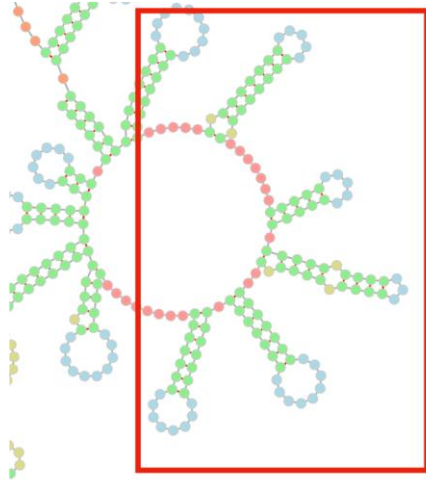# Viral genome modelling

An icosahedral lattice polymer

Real secondary structure

Larman et a., 2017
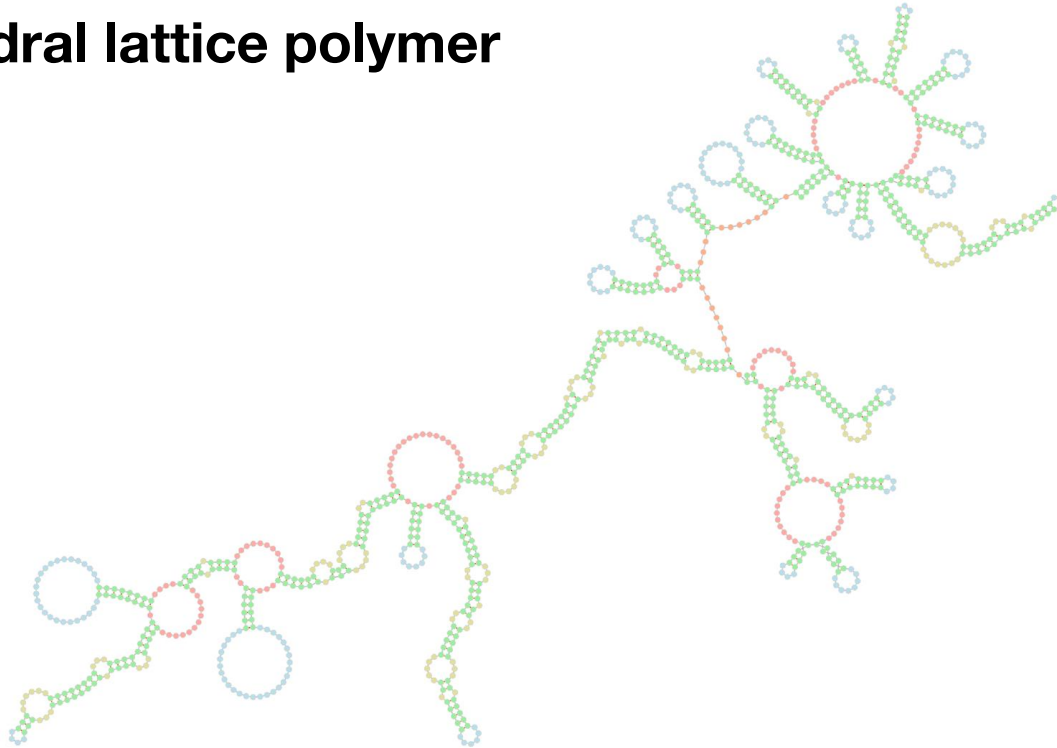
Viral genome modelling

An icosahedral lattice polymer
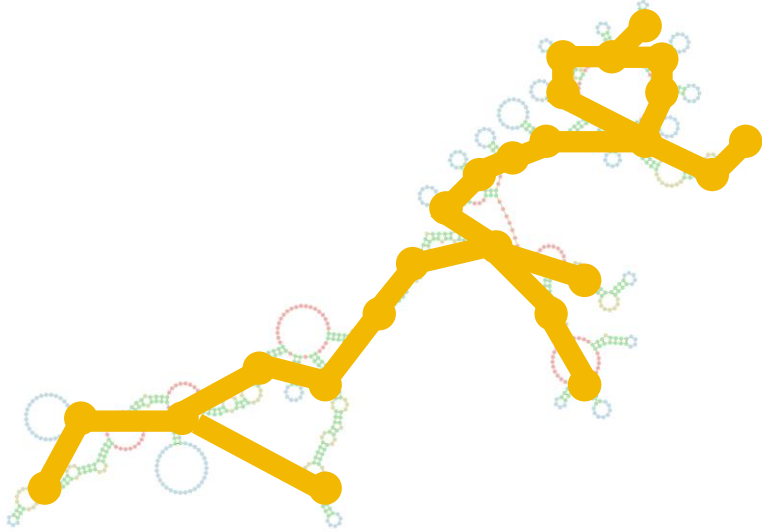
Real secondary structure
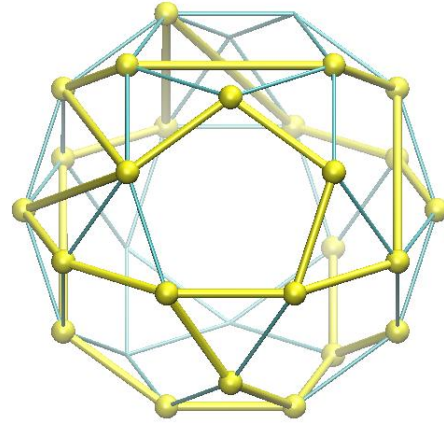
Lattice polymer

Viral genome modelling

Exploring the conformations

Monte Carlo trial moves

Wedge
(3-body)
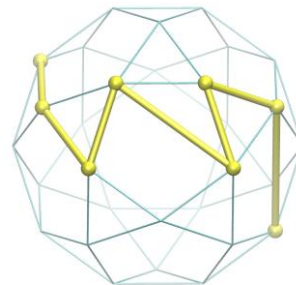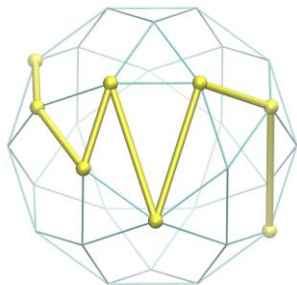
Torsion
(4-body)
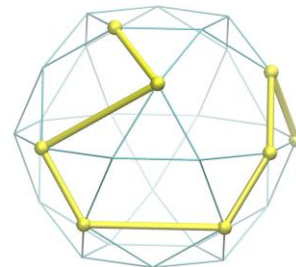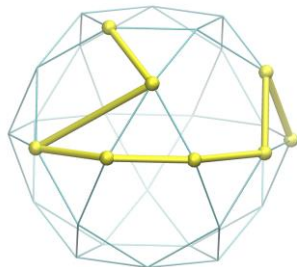
Viral genome modelling
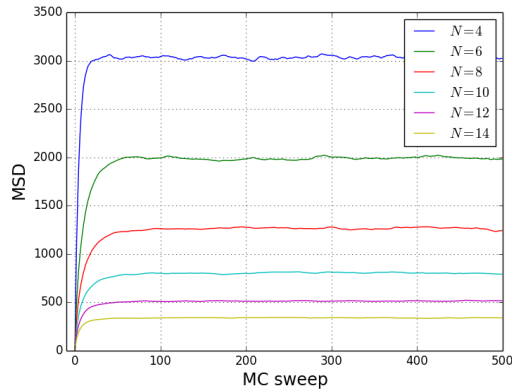
Lattice polymer results, 10000 Monte Carlo sweeps
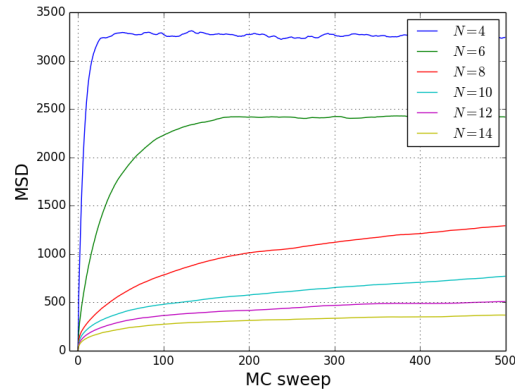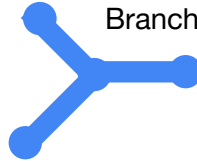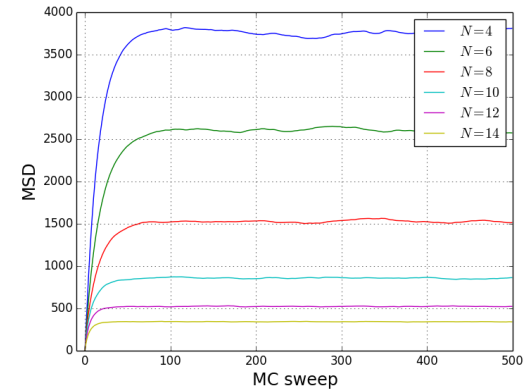
Viral genome modelling

Lattice polymer results, 10000 Monte Carlo sweeps, 6 initial conditions

In the lattice, a 60° angle involves 3 capsid proteins - stable, suitable for guiding capsid assembly

N_T : number of 60° angles
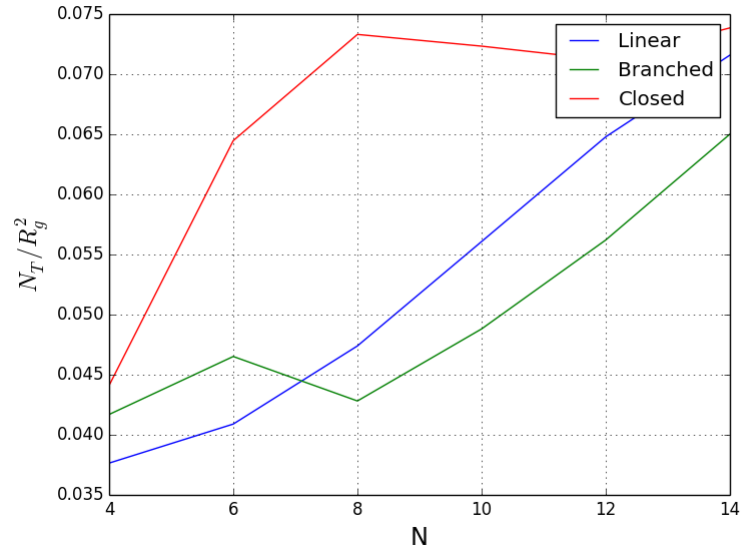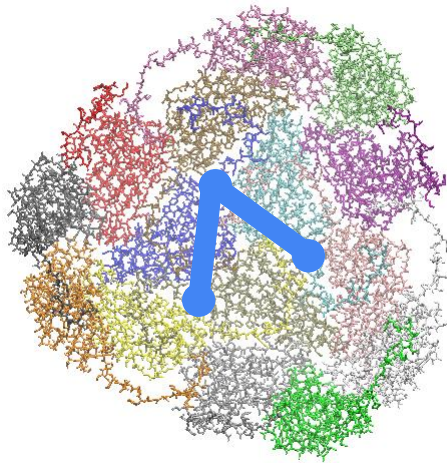
Closed rings stabilize more efficiently

Viral genome modelling

Massive sampling of conformations of each pair of consecutive stems in the lattice

Evaluate energy

Assemble final model
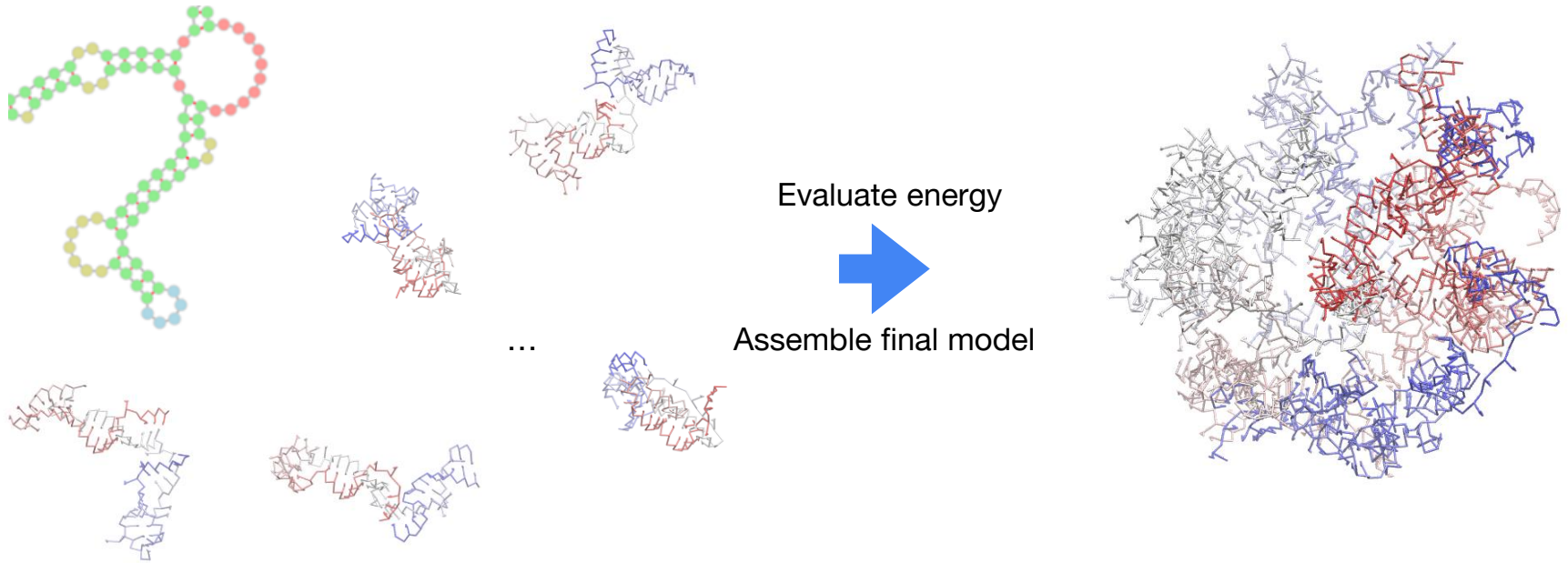
Porcine Circovirus 2 (PCV2)

Collaboration with Prof. Sergio Pantano, Pasteur Institute, Montevideo, Uruguay.

Variant PCV2d of clinical importance. PCV associated disease is responsible for diarrhea, respiratory diseases and death.

Icosahedral DNA virus. **Unstructured genome.**
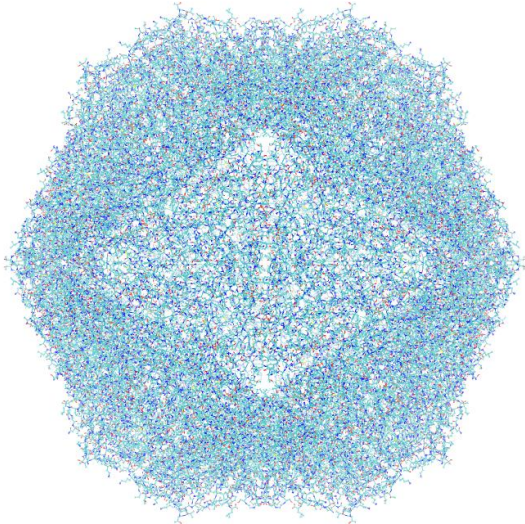
Viral genome modelling

X-ray structure contains fragments, but which ones? Sequence RRYY
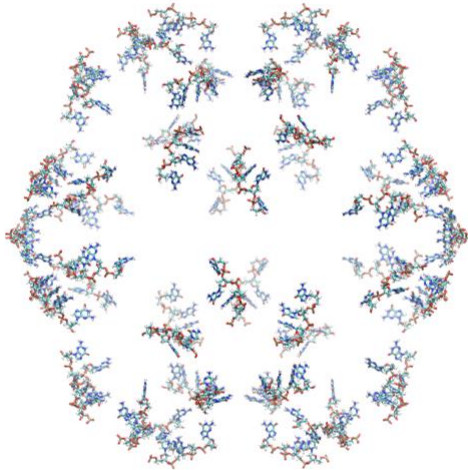

```
ACCAGCGCACTTCGGCAGCGGCAGCACCTCGGCAGCACTTCAGCAGCAACATGCCCAGCAAAAAGAATGG
AAGAAGCGGACCCCAACCACACAAAAGGTGGGTGTTCACGCTGAATAATCCTTCCGAGGACGAGCGCAAG
AAAATACGGGAGCTTCCAATCTCCCTTTTTGATTATTTTATTGTTGGCGAGGAGGGTAATGAGGAAGGAC
GAACACCCCACCTCCAGGGGTTCGCTAATTTTGTGAAGAAGCAAACATTTAATAAAGTGAAATGGTATTT
CGGTGCCCGCTGCCACATCGAGAAAGCGAAAGGAACTGATCAGCAGAATAAAGAATATTGCAGTAAAGAA
GGCAACTTACTGATTGAATGTGGAGCTCCTAGATCTCAAGGACAACGGAGTGACCTGTCTACTGCTGTGA
GTACCTTGCTGGCGAGCGGGAGTCTGGTGACCGTTGCAGAGCAGCACCCTGTAACGTTTGTCAGAAATTT
CCGCGGGCTGGCTGAACTTTTGAAAGTGAGCGGGAAAATGCAGAACGGTGATTGGAAGACTAATGTACAC
GTCATTGTGGGGCCACCTGGGTGTGGTAAAAGCAAATGGGCTGCTAATTTTGCAGACCCGGAAACCACAT
ACTGGAAACCACCTAGAAACAAGTGGTGGGATGGTTACCATGGTCAAGAATGGTTGTTATTGATGACTT
TTATGGCTGGCTGCCCTGGGATGATCTACTGAGACTGTGTGATCGATATCCATTGACTGTAGAGACTAAA
GGTGGAACTGTACCTTTTTTGGCCCGCAGTATTCTGATTACCAGCAATCAGACCCCGTTGGAATGGTACT
CCTCAACTGCTGTCCCAGCTGTAGAAGCTCTTTATCGGAGGATTACTTCCTTGGTATTTTGGAAGAATGC
TACAGAACAATCCACGGAGGAAGGGGGCCAGTTCGTCACCCTTTCCCCCCATGCCCTGAATTTCCATAT
GAAATAAATTACTGAGTCTTTTTTATCACTTCGTAATGGTTTTTATTATTCACTTAGGGTTAAGTGGGGG
GTCTTTAAGAATAAATTCTCTGAATTGTACATACATGGTTATACGGATATTGTAGTCCTGGTCGTATATA
CTGTTTTCGAACGCAGTGCCGAGGCCTACATGGTCTACATTTCCAGTAGTTTGTAGTCTCAGCCAGAGTT
GATTTCTTTTGTTATTGGGTTGGAAGTAATCGATTGTCCTATCAAGGACAGGTTCGGGGTAAAGTATCG
GGAGTGGTAGGAGAAGGGCTGGGTTATGGTATGGCGGGAGGAGTAGTTTACATAGGGGTCATAGGTTAGG
GCATTGGCCTTTGTTACAAAGTTATCATCTAGAATAACAGCAGTGGAGCCCACTCCCCTGTCACCCTGGG
TGATTGGGGAGCAGGGCCAGAATTCAACCTTAACCTTCCTTATTCTGTAGTATTCAAAGGGCACAGTGAG
GGGGTTTGAGCCCCCTCCTGGGGGAAGAAAATCATTAATATTAAATCTCATCATGTCCACATTCCAGGAG
GGCGTTCTGACTGTGGTTTTCTTGACAGTATAACCGATGGTGCGGGAGAGGCGGGTGTTGAAGATGCCAT
TTTTCCTTCTCCAGCGGTAACGGTGGCGGGGGTGGACGAGCCAGGGGCGGCGGCGAGGATTTGGCCAAG
ATGGCTGCGGGGGCGGTGTCTTCGTCTGCGGAAACGCCTCCTTGGATACGTCATCGCTGAAAACGAAAGA
AGTGCGCTGTAAGTATT
```
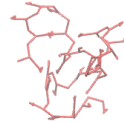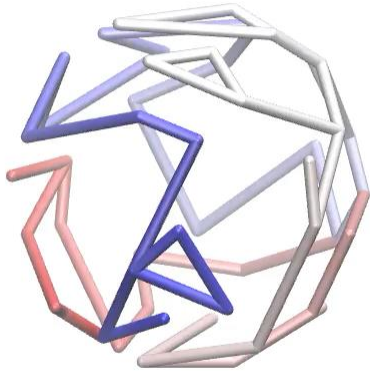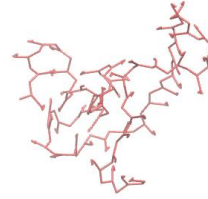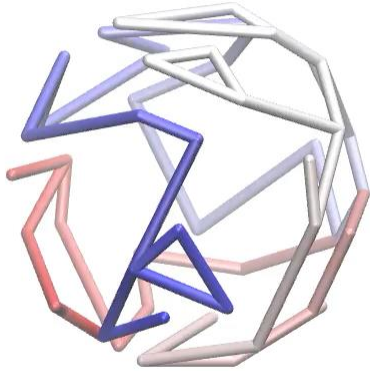
Select 30 occurrences
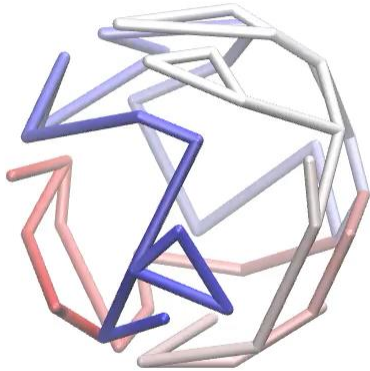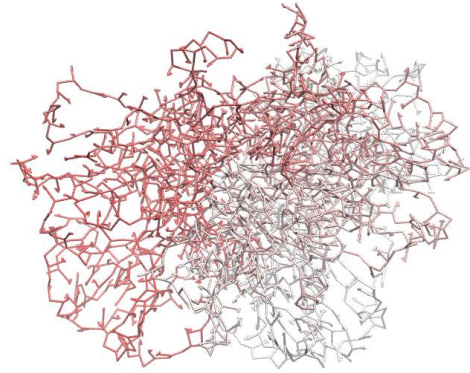
Viral genome modelling

A possible candidate…

Viral genome modelling

A possible candidate…

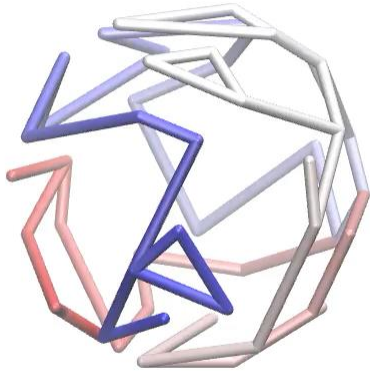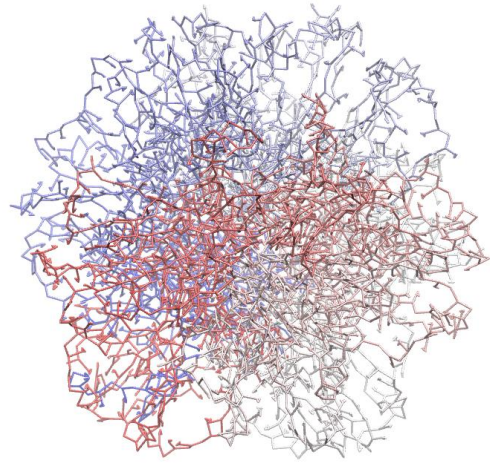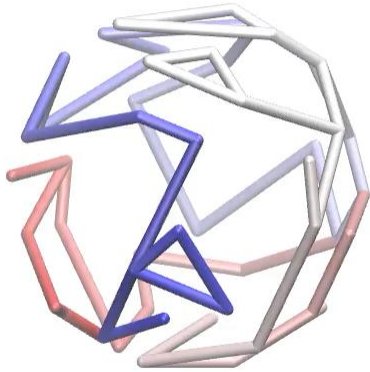# Viral genome modelling

A possible candidate…
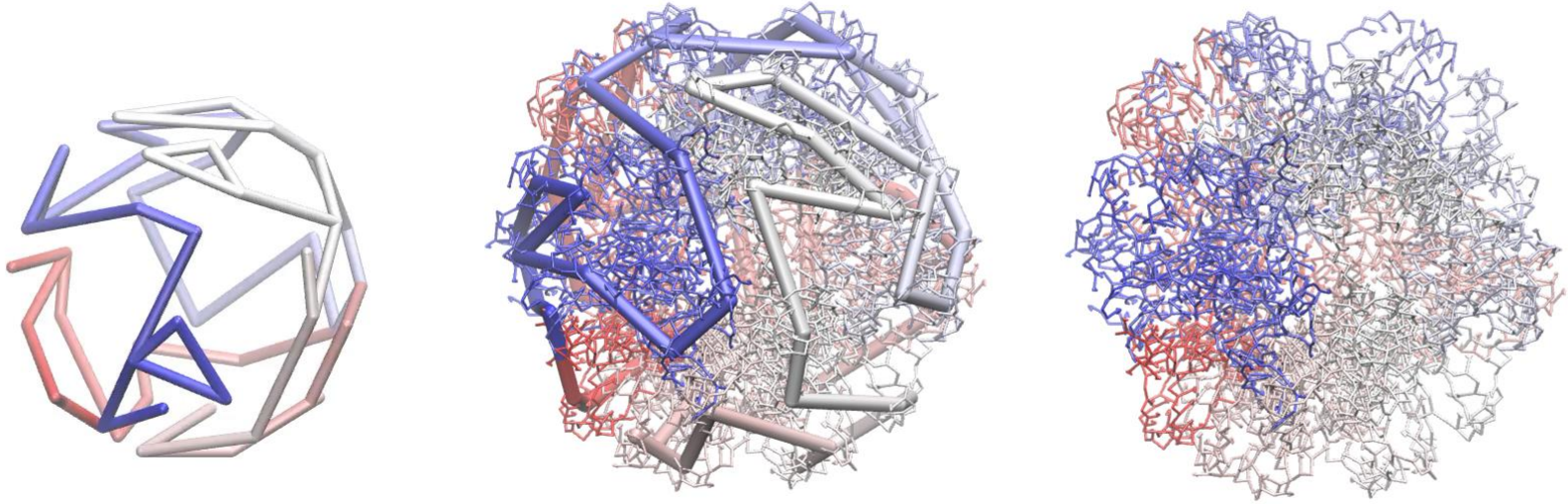
Viral genome modelling

A possible candidate…

Viral genome modelling

A possible candidate…

Viral genome modelling

A possible candidate…

…which of course, requires an automatic disentanglement.

Conclusions

- Coarse Graining of RNA molecules can greatly help in structure prediction.
- Topological artifacts are relevant for structure design.
- Large structures as viral genomes are prone to be modelled by these means.

Acknowledgements