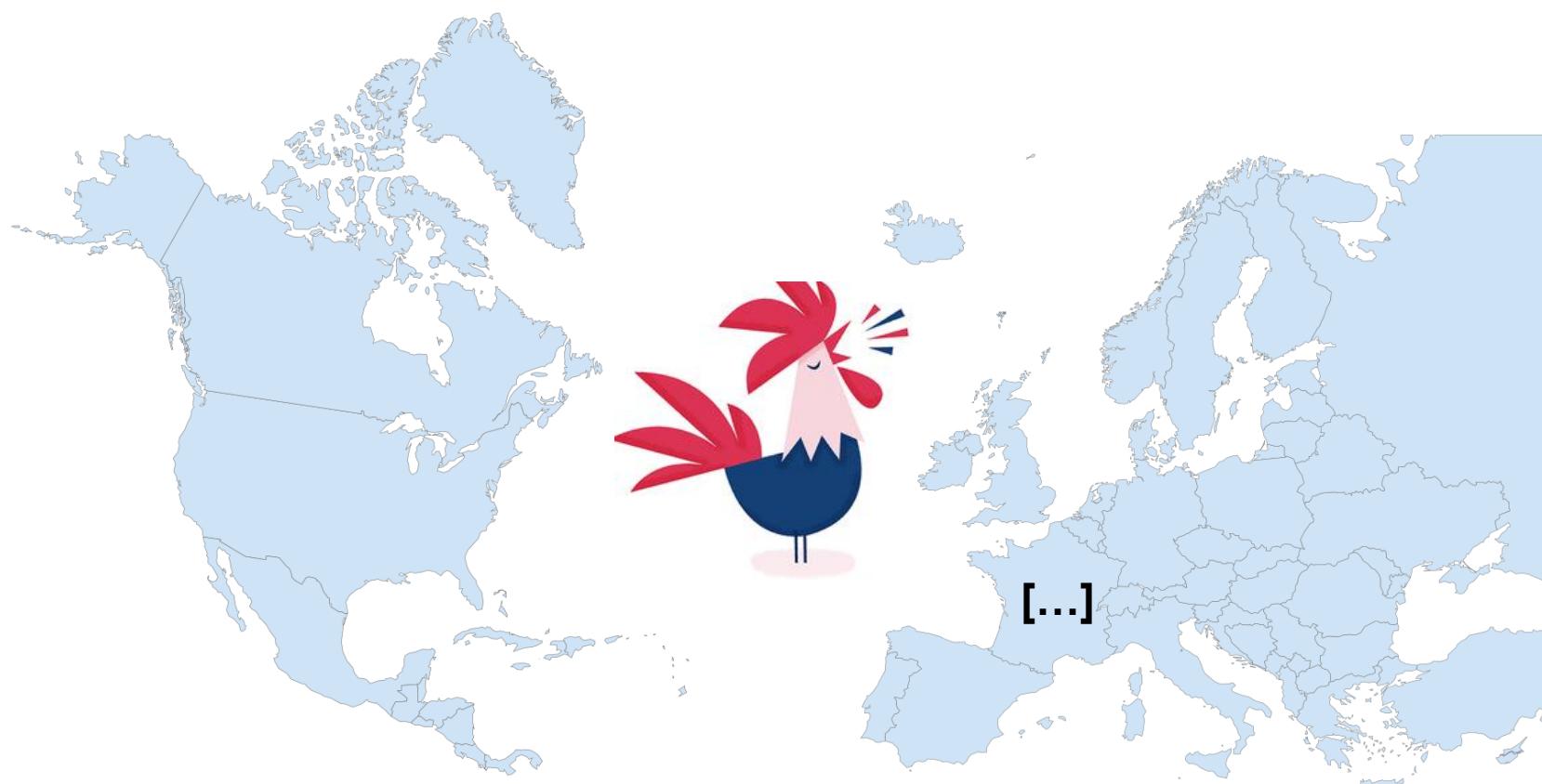


## Modeling in structural bioinformatics: structure – dynamics – function

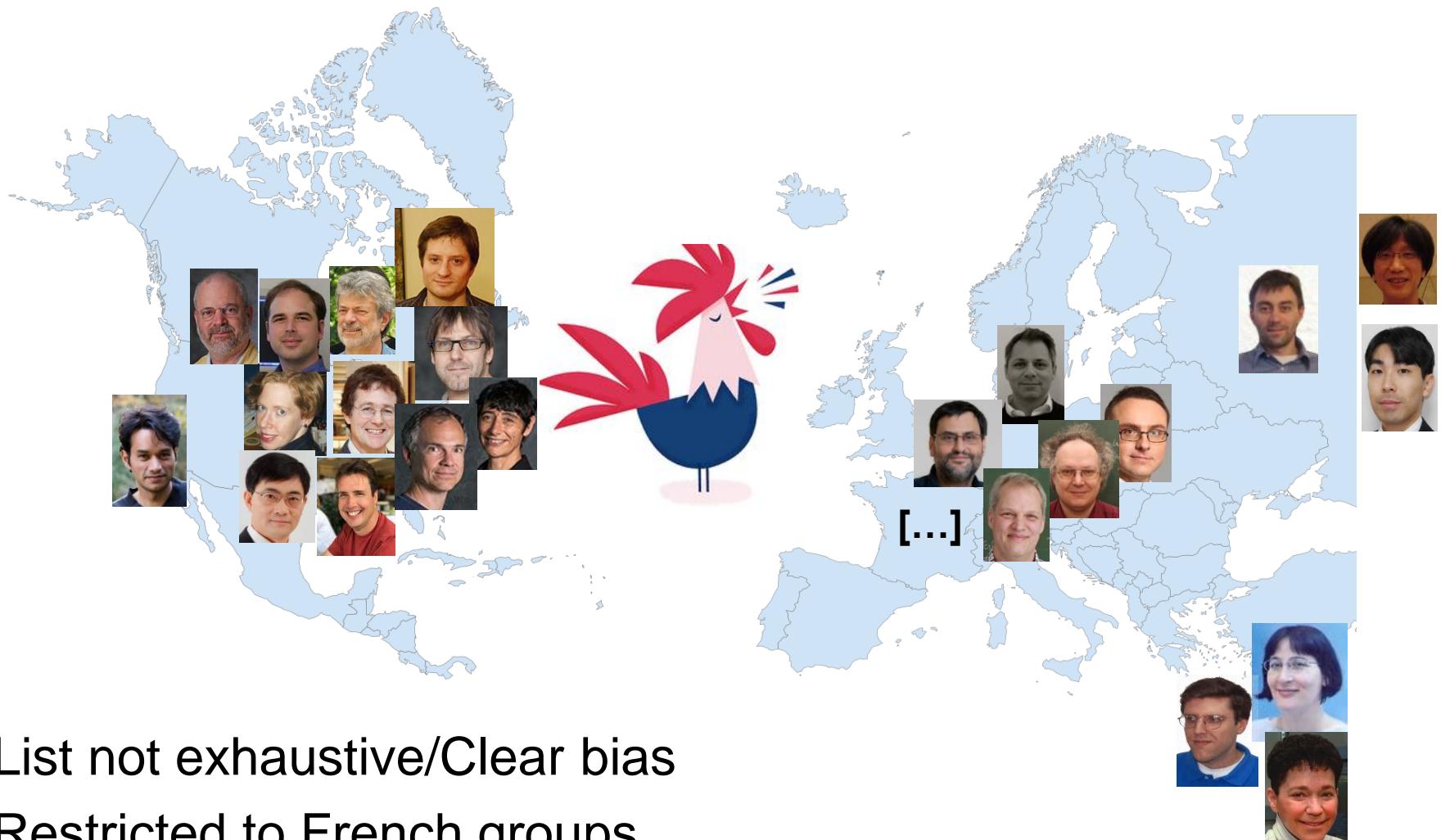
Part II: RNA bioinformatics, ze French way...

# The RNA (Bioinfo) world



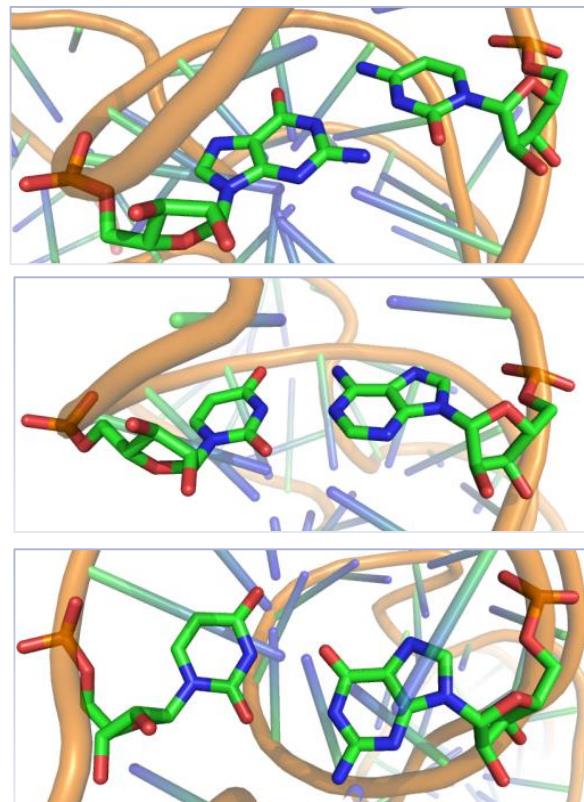
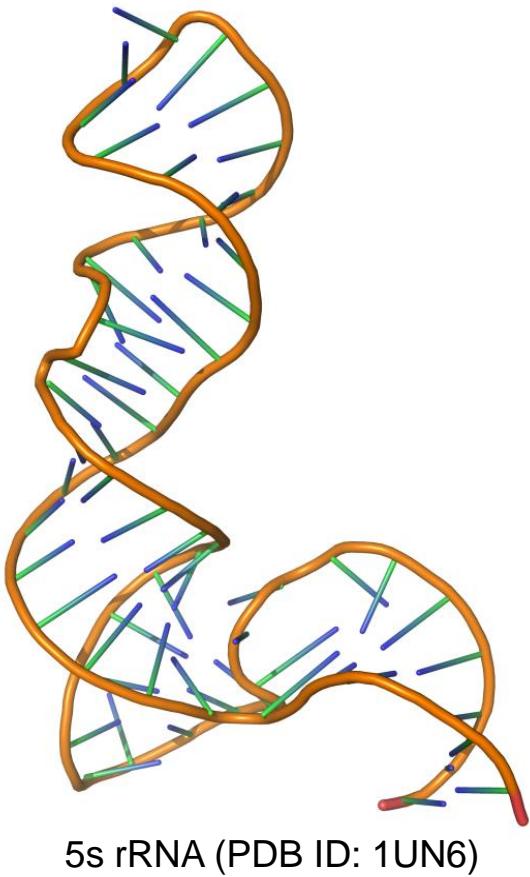
- ▶ List not exhaustive/Clear bias
- ▶ Restricted to French groups...
- ▶ ... focusing on speakers at the MASIM kickoff meeting

# The RNA (Bioinfo) world



- ▶ List not exhaustive/Clear bias
- ▶ Restricted to French groups...
- ▶ ... focusing on speakers at the MASIM kickoff meeting

# Why RNA folds



Canonical base-pairs

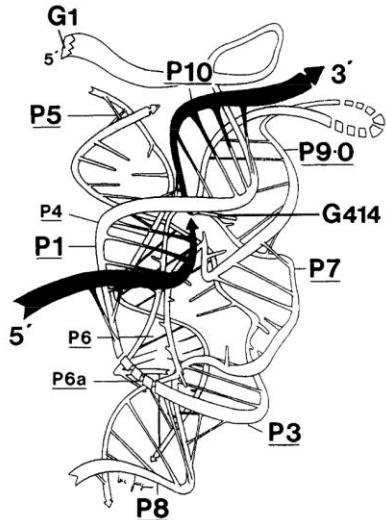
RNA folding = Hierarchical stochastic process driven by/resulting in the pairing (hydrogen bonds) of a subset of its bases.

# A French pioneer in RNA Structural Modeling

Tetrahymena group I intron



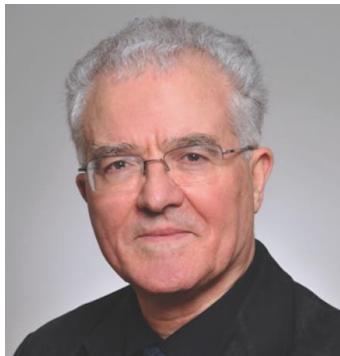
E Westhof  
IBMC, Strasbourg



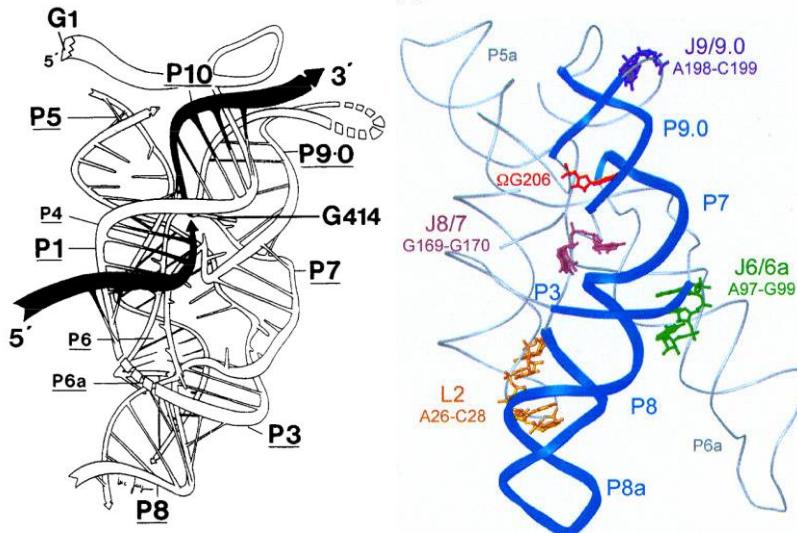
Predicted (manually)  
[Westhof Michel, 1990]

# A French pioneer in RNA Structural Modeling

## Tetrahymena group I intron



E Westhof  
IBMC, Strasbourg

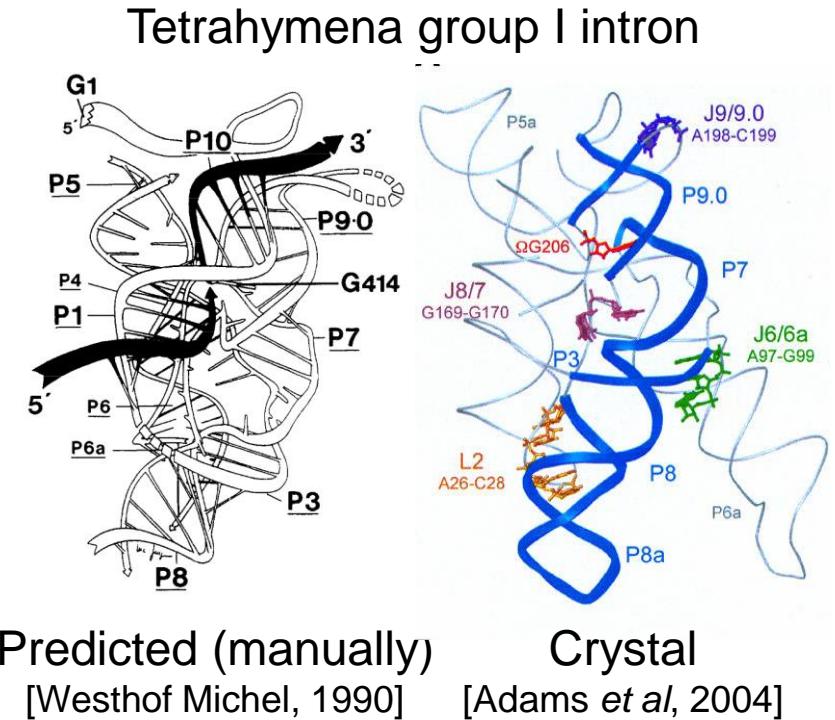


Predicted (manually)      Crystal  
[Westhof Michel, 1990]      [Adams et al, 2004]

# A French pioneer in RNA Structural Modeling



**E Westhof**  
IBMC, Strasbourg



P5a P5b P5c P5d P5e A-rich P5r  
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 P15 P16 P17 P18 P19 P20 P21 P22 P23 P24 P25 P26 P27 P28 P29 P30 P31 P32 P33 P34 P35 P36 P37 P38 P39 P40 P41 P42 P43 P44 P45 P46 P47 P48 P49 P50 P51 P52 P53 P54 P55 P56 P57 P58 P59 P60 P61 P62 P63 P64 P65 P66 P67 P68 P69 P70 P71 P72 P73 P74 P75 P76 P77 P78 P79 P80 P81 P82 P83 P84 P85 P86 P87 P88 P89 P90 P91 P92 P93 P94 P95 P96 P97 P98 P99 P100

P7.1 P7.1a P7.1 P7.1b P7.1'

P7.2 P7.2 P7.2

P7.3 P7.3 P7.3

P7.4 P7.4 P7.4

P7.5 P7.5 P7.5

P7.6 P7.6 P7.6

P7.7 P7.7 P7.7

P7.8 P7.8 P7.8

P7.9 P7.9 P7.9

P7.10 P7.10 P7.10

P7.11 P7.11 P7.11

P7.12 P7.12 P7.12

P7.13 P7.13 P7.13

P7.14 P7.14 P7.14

P7.15 P7.15 P7.15

P7.16 P7.16 P7.16

P7.17 P7.17 P7.17

P7.18 P7.18 P7.18

P7.19 P7.19 P7.19

P7.20 P7.20 P7.20

P7.21 P7.21 P7.21

P7.22 P7.22 P7.22

P7.23 P7.23 P7.23

P7.24 P7.24 P7.24

P7.25 P7.25 P7.25

P7.26 P7.26 P7.26

P7.27 P7.27 P7.27

P7.28 P7.28 P7.28

P7.29 P7.29 P7.29

P7.30 P7.30 P7.30

P7.31 P7.31 P7.31

P7.32 P7.32 P7.32

P7.33 P7.33 P7.33

P7.34 P7.34 P7.34

P7.35 P7.35 P7.35

P7.36 P7.36 P7.36

P7.37 P7.37 P7.37

P7.38 P7.38 P7.38

P7.39 P7.39 P7.39

P7.40 P7.40 P7.40

P7.41 P7.41 P7.41

P7.42 P7.42 P7.42

P7.43 P7.43 P7.43

P7.44 P7.44 P7.44

P7.45 P7.45 P7.45

P7.46 P7.46 P7.46

P7.47 P7.47 P7.47

P7.48 P7.48 P7.48

P7.49 P7.49 P7.49

P7.50 P7.50 P7.50

P7.51 P7.51 P7.51

P7.52 P7.52 P7.52

P7.53 P7.53 P7.53

P7.54 P7.54 P7.54

P7.55 P7.55 P7.55

P7.56 P7.56 P7.56

P7.57 P7.57 P7.57

P7.58 P7.58 P7.58

P7.59 P7.59 P7.59

P7.60 P7.60 P7.60

P7.61 P7.61 P7.61

P7.62 P7.62 P7.62

P7.63 P7.63 P7.63

P7.64 P7.64 P7.64

P7.65 P7.65 P7.65

P7.66 P7.66 P7.66

P7.67 P7.67 P7.67

P7.68 P7.68 P7.68

P7.69 P7.69 P7.69

P7.70 P7.70 P7.70

P7.71 P7.71 P7.71

P7.72 P7.72 P7.72

P7.73 P7.73 P7.73

P7.74 P7.74 P7.74

P7.75 P7.75 P7.75

P7.76 P7.76 P7.76

P7.77 P7.77 P7.77

P7.78 P7.78 P7.78

P7.79 P7.79 P7.79

P7.80 P7.80 P7.80

P7.81 P7.81 P7.81

P7.82 P7.82 P7.82

P7.83 P7.83 P7.83

P7.84 P7.84 P7.84

P7.85 P7.85 P7.85

P7.86 P7.86 P7.86

P7.87 P7.87 P7.87

P7.88 P7.88 P7.88

P7.89 P7.89 P7.89

P7.90 P7.90 P7.90

P7.91 P7.91 P7.91

P7.92 P7.92 P7.92

P7.93 P7.93 P7.93

P7.94 P7.94 P7.94

P7.95 P7.95 P7.95

P7.96 P7.96 P7.96

P7.97 P7.97 P7.97

P7.98 P7.98 P7.98

P7.99 P7.99 P7.99

P7.100 P7.100 P7.100

P7.101 P7.101 P7.101

P7.102 P7.102 P7.102

P7.103 P7.103 P7.103

P7.104 P7.104 P7.104

P7.105 P7.105 P7.105

P7.106 P7.106 P7.106

P7.107 P7.107 P7.107

P7.108 P7.108 P7.108

P7.109 P7.109 P7.109

P7.110 P7.110 P7.110

P7.111 P7.111 P7.111

P7.112 P7.112 P7.112

P7.113 P7.113 P7.113

P7.114 P7.114 P7.114

P7.115 P7.115 P7.115

P7.116 P7.116 P7.116

P7.117 P7.117 P7.117

P7.118 P7.118 P7.118

P7.119 P7.119 P7.119

P7.120 P7.120 P7.120

P7.121 P7.121 P7.121

P7.122 P7.122 P7.122

P7.123 P7.123 P7.123

P7.124 P7.124 P7.124

P7.125 P7.125 P7.125

P7.126 P7.126 P7.126

P7.127 P7.127 P7.127

P7.128 P7.128 P7.128

P7.129 P7.129 P7.129

P7.130 P7.130 P7.130

P7.131 P7.131 P7.131

P7.132 P7.132 P7.132

P7.133 P7.133 P7.133

P7.134 P7.134 P7.134

P7.135 P7.135 P7.135

P7.136 P7.136 P7.136

P7.137 P7.137 P7.137

P7.138 P7.138 P7.138

P7.139 P7.139 P7.139

P7.140 P7.140 P7.140

P7.141 P7.141 P7.141

P7.142 P7.142 P7.142

P7.143 P7.143 P7.143

P7.144 P7.144 P7.144

P7.145 P7.145 P7.145

P7.146 P7.146 P7.146

P7.147 P7.147 P7.147

P7.148 P7.148 P7.148

P7.149 P7.149 P7.149

P7.150 P7.150 P7.150

P7.151 P7.151 P7.151

P7.152 P7.152 P7.152

P7.153 P7.153 P7.153

P7.154 P7.154 P7.154

P7.155 P7.155 P7.155

P7.156 P7.156 P7.156

P7.157 P7.157 P7.157

P7.158 P7.158 P7.158

P7.159 P7.159 P7.159

P7.160 P7.160 P7.160

P7.161 P7.161 P7.161

P7.162 P7.162 P7.162

P7.163 P7.163 P7.163

P7.164 P7.164 P7.164

P7.165 P7.165 P7.165

P7.166 P7.166 P7.166

P7.167 P7.167 P7.167

P7.168 P7.168 P7.168

P7.169 P7.169 P7.169

P7.170 P7.170 P7.170

P7.171 P7.171 P7.171

P7.172 P7.172 P7.172

P7.173 P7.173 P7.173

P7.174 P7.174 P7.174

P7.175 P7.175 P7.175

P7.176 P7.176 P7.176

P7.177 P7.177 P7.177

P7.178 P7.178 P7.178

P7.179 P7.179 P7.179

P7.180 P7.180 P7.180

P7.181 P7.181 P7.181

P7.182 P7.182 P7.182

P7.183 P7.183 P7.183

P7.184 P7.184 P7.184

P7.185 P7.185 P7.185

P7.186 P7.186 P7.186

P7.187 P7.187 P7.187

P7.188 P7.188 P7.188

P7.189 P7.189 P7.189

P7.190 P7.190 P7.190

P7.191 P7.191 P7.191

P7.192 P7.192 P7.192

P7.193 P7.193 P7.193

P7.194 P7.194 P7.194

P7.195 P7.195 P7.195

P7.196 P7.196 P7.196

P7.197 P7.197 P7.197

P7.198 P7.198 P7.198

P7.199 P7.199 P7.199

P7.200 P7.200 P7.200

P7.201 P7.201 P7.201

P7.202 P7.202 P7.202

P7.203 P7.203 P7.203

P7.204 P7.204 P7.204

P7.205 P7.205 P7.205

P7.206 P7.206 P7.206

P7.207 P7.207 P7.207

P7.208 P7.208 P7.208

P7.209 P7.209 P7.209

P7.210 P7.210 P7.210

P7.211 P7.211 P7.211

P7.212 P7.212 P7.212

P7.213 P7.213 P7.213

P7.214 P7.214 P7.214

P7.215 P7.215 P7.215

P7.216 P7.216 P7.216

P7.217 P7.217 P7.217

P7.218 P7.218 P7.218

P7.219 P7.219 P7.219

P7.220 P7.220 P7.220

P7.221 P7.221 P7.221

P7.222 P7.222 P7.222

P7.223 P7.223 P7.223

P7.224 P7.224 P7.224

P7.225 P7.225 P7.225

P7.226 P7.226 P7.226

P7.227 P7.227 P

# A French pioneer in RNA Structural Modeling

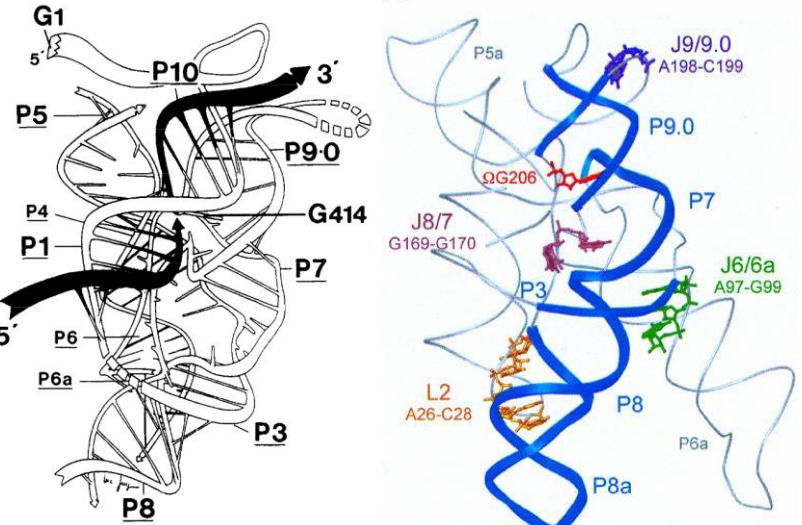
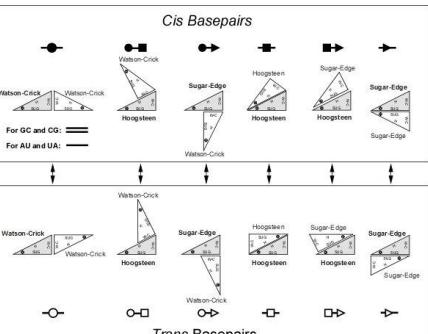
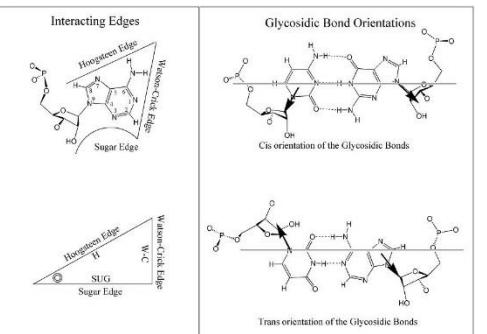


**E Westhof**  
**IBMC, Strasbourg**

# Non-canonical base pairs and modules

## Nomenclature

[Leontis Westhof 2001]



Predicted (manually)	Crystal
[Westhof Michel, 1990]	[Adams <i>et al</i> , 2004]

	P7.1	P7.1'	P7.2	P7.2'
52	-----	-----	-----	-----
53	-----	-----	-----	-----
54	-----	-----	-----	-----
55	-----	-----	-----	-----
56	-----	-----	-----	-----
57	-----	-----	-----	-----
58	-----	-----	-----	-----
59	-----	-----	-----	-----
60	-----	-----	-----	-----
61	-----	-----	-----	-----
62	-----	-----	-----	-----
63	-----	-----	-----	-----
64	-----	-----	-----	-----
65	-----	-----	-----	-----
66	-----	-----	-----	-----
67	-----	-----	-----	-----
68	-----	-----	-----	-----
69	-----	-----	-----	-----
70	-----	-----	-----	-----
71	-----	-----	-----	-----
72	-----	-----	-----	-----

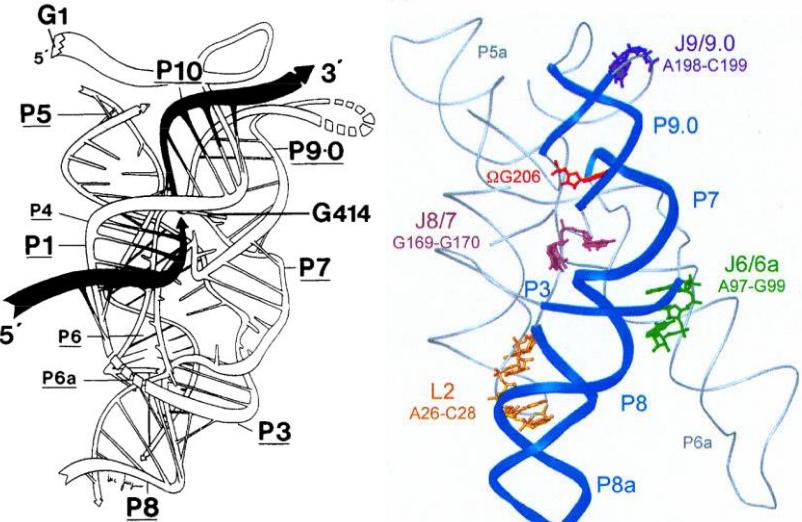
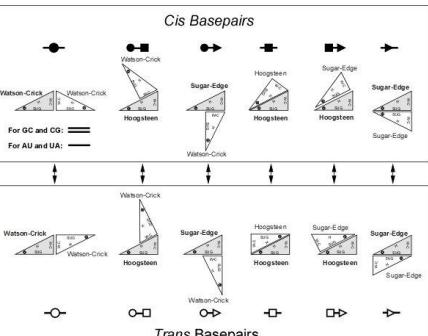
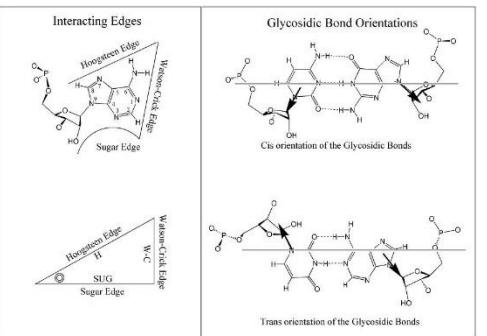
# A French pioneer in RNA Structural Modeling



**E Westhof**  
**IBMC, Strasbourg**

# Non-canonical base pairs and modules

## Nomenclature



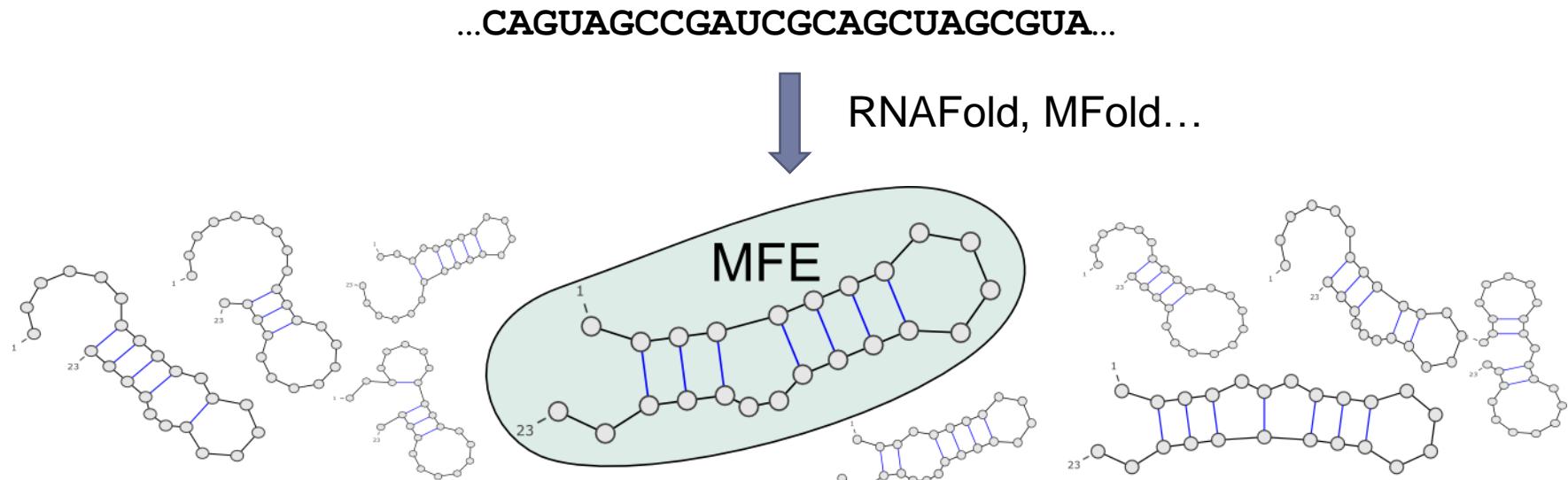
Predicted (manually)      Crystal  
[Westhof Michel, 1990]      [Adams *et al*, 2004]



# Computational paradigms and methods

# Minimal Free-Energy (MFE) Folding

Despite an exponential #conformations, the Minimal Free Energy secondary structure can be predicted exactly in polynomial time in the nearest neighbor model



[Nussinov & Jacobson, PNAS 1980; Zuker & Stiegler, NAR 1981]

*Poster-child* of dynamic programming in bioinformatics

# Energy-based *Ab initio* folding methods: Do they *really* work?

- ▶ Generally **yes**, but some variations **across studies**

Benchmark: 1919 non-multimer/non-pseudoknotted structures from RNAstrand  
[Andronescu *et al*, BMC Bioinformatics 2008]

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Program	Sensitivity	PPV	MCC	F-measure
<b>RNAfold 2.1.9</b>	<b>0.742</b>	<b>0.795</b>	<b>0.767</b>	0.765
RNAfold 2.1.8	0.740	0.792	0.764	0.762
RNAfold 1.8.5	0.711	0.773	0.740	0.737
UNAfold 3.8	0.693	0.767	0.727	0.725
RNAstructure 5.7	0.716	0.781	0.746	0.744

[Lorenz *et al*, Algo Mol Biol 2011]

# Optimization methods can be sensitive to fluctuations of the energy model

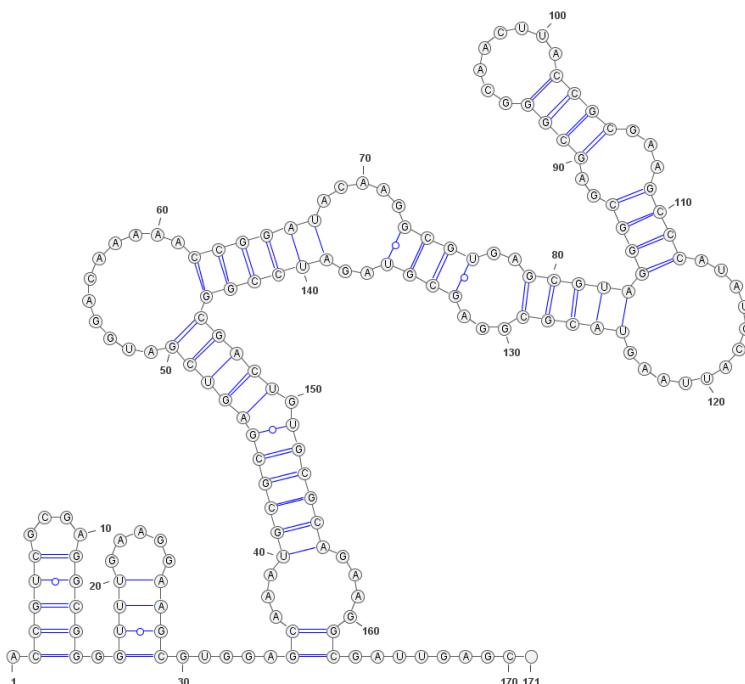
Example:

- ▶ RFAM *A. capsulatum* D1-D4 domain of the Group II intron
- ▶ Run RNAFold using:
  - ▶ Default energy model (Turner 2004)

# Optimization methods can be sensitive to fluctuations of the energy model

Example:

- ▶ RFAM *A. capsulatum* D1-D4 domain of the Group II intron
- ▶ Run RNAFold using:
  - ▶ Default energy model (Turner 2004)

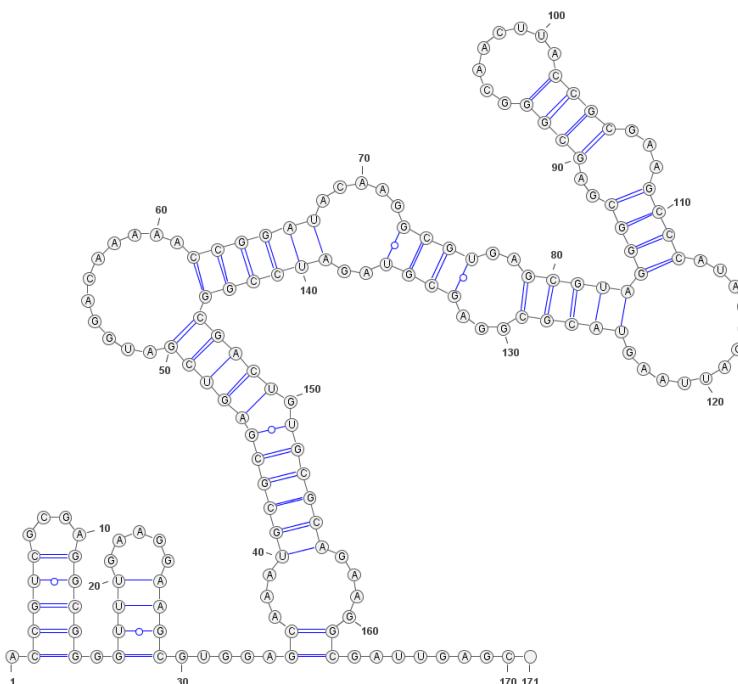


Turner 2004

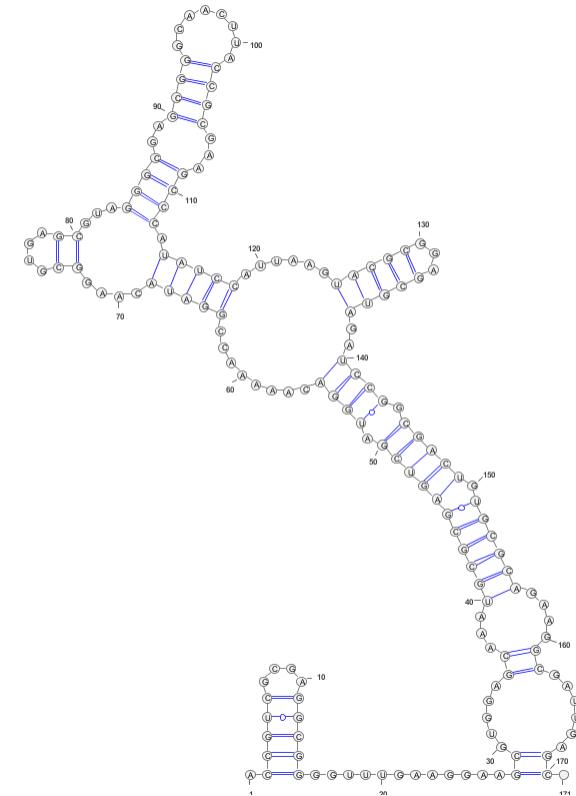
# Optimization methods can be sensitive to fluctuations of the energy model

Example:

- ▶ RFAM *A. capsulatum* D1-D4 domain of the Group II intron
- ▶ Run RNAFold using:
  - ▶ Default energy model (Turner 2004)
  - ▶ Data-driven model (Andronescu 2007)



Turner 2004

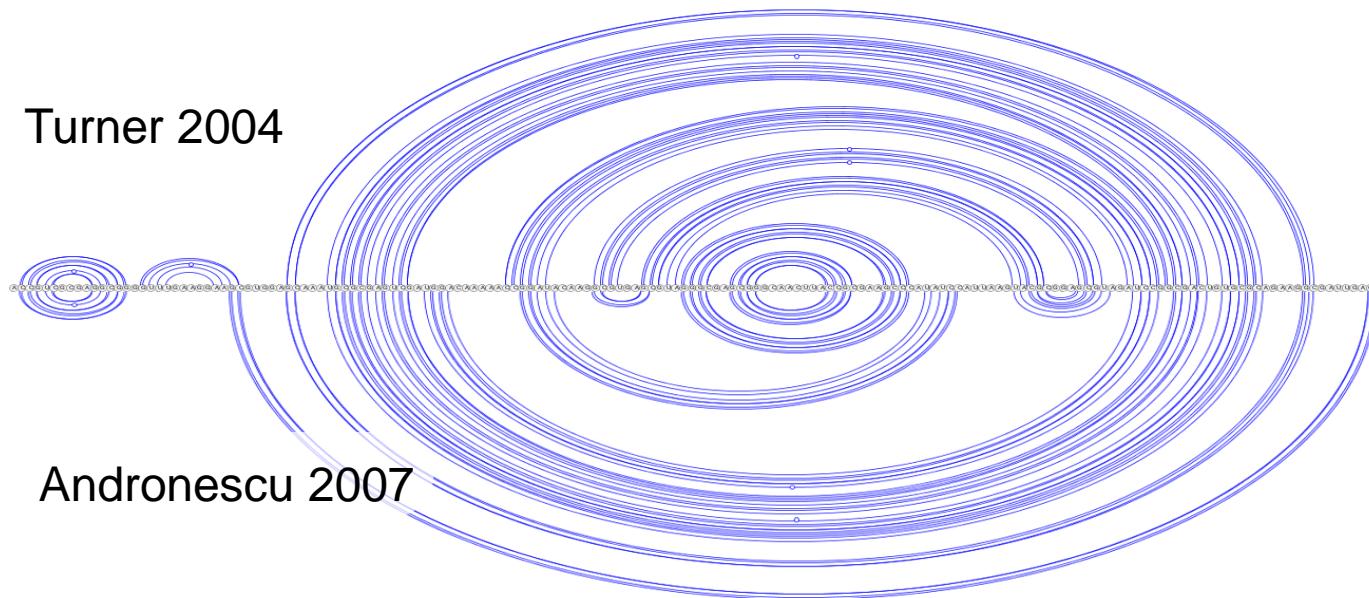


Andronescu 2007

# Optimization methods can be overly sensitive to fluctuations of the energy model

Example:

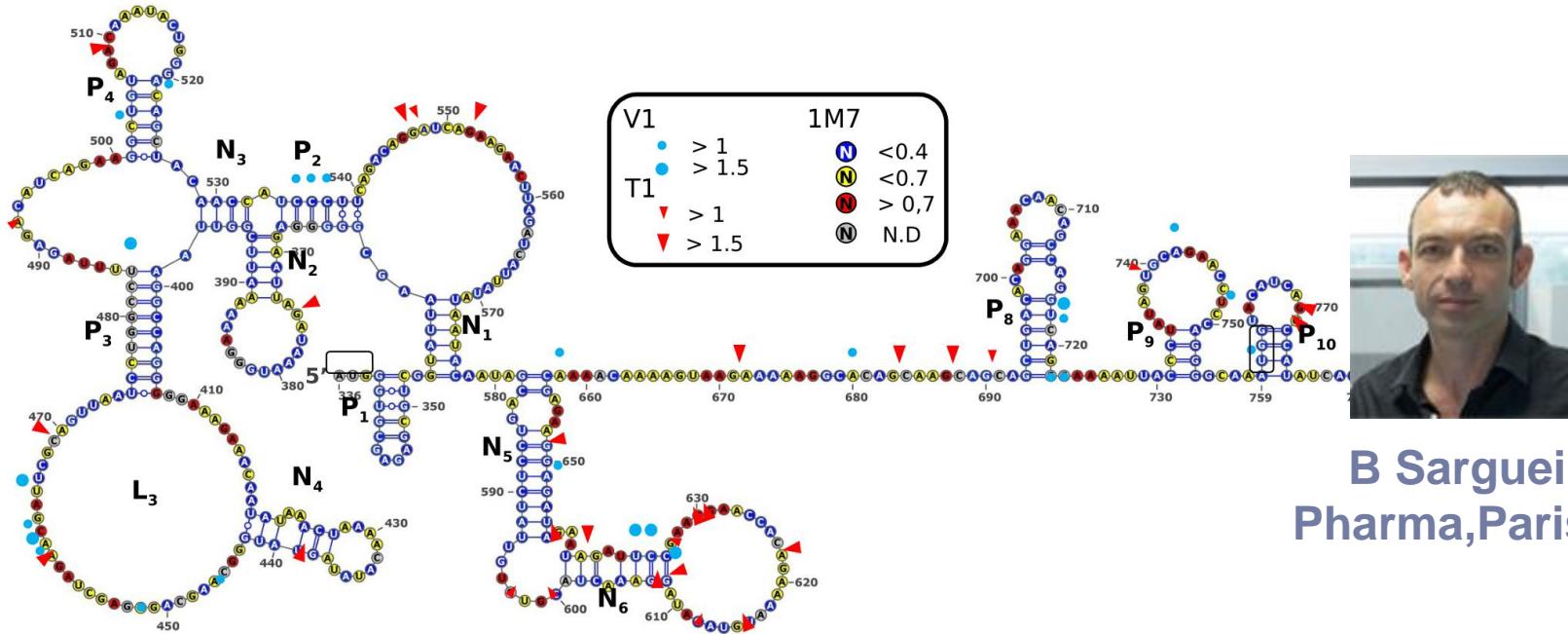
- ▶ RFAM A. *capsulatum* D1-D4 domain of the Group II intron
- ▶ Run RNAFold using:
  - ▶ Default energy model (Turner 2004)
  - ▶ Data-driven model (Andronescu 2007)



Discrepancy not as embarrassing as it first looks...  
... still substantial!

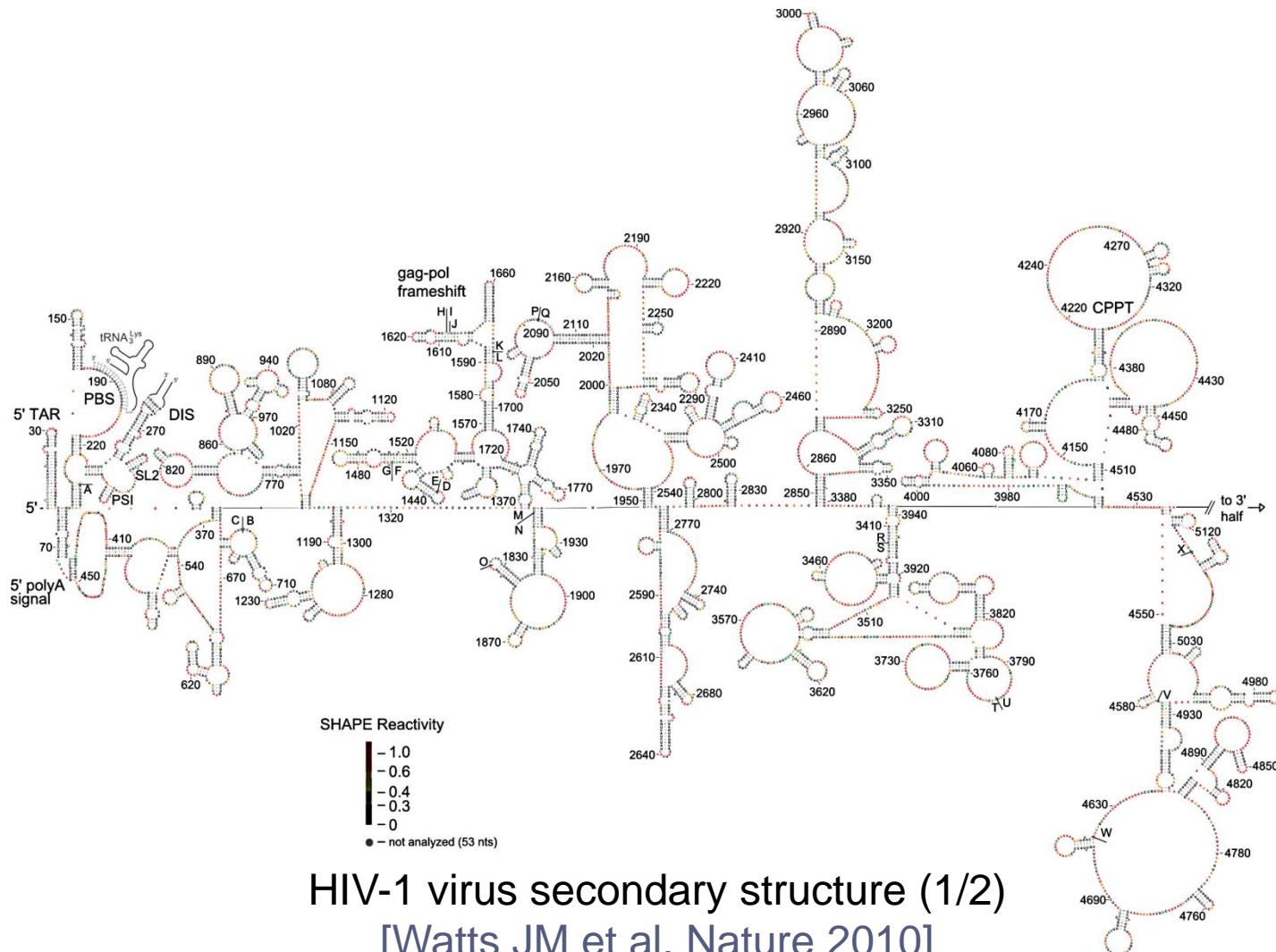
# Chemical/enzymatic probing to model 2D

- ▶ High-throughput secondary structure determination
- ▶ Reactivity/accessibility guide manual modeling choices

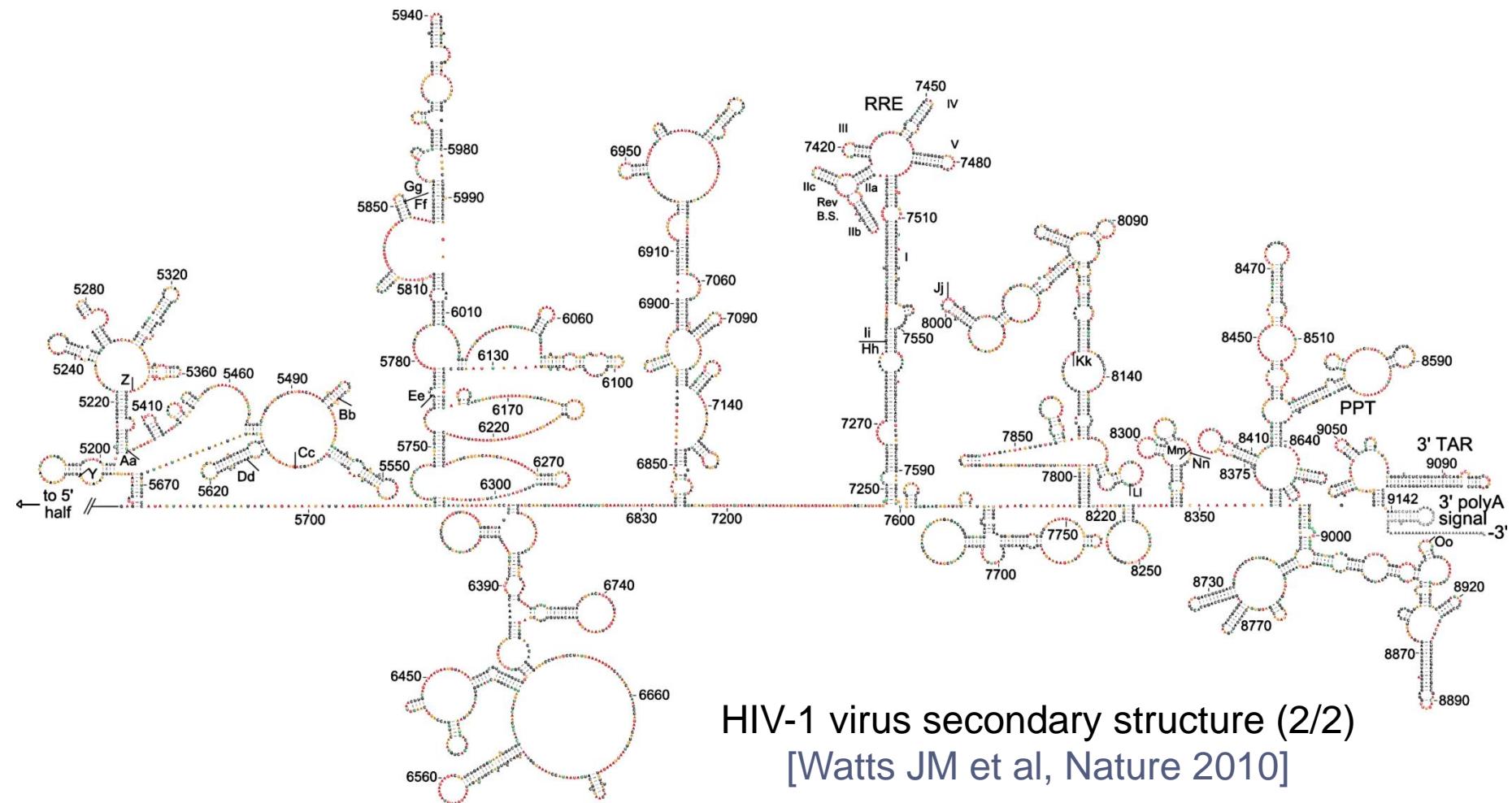


- ▶ Inclusion as pseudo potentials within energy-models  
[Lorenz *et al*, Bioinformatics 2015]

# SHAPE probing to model 2D



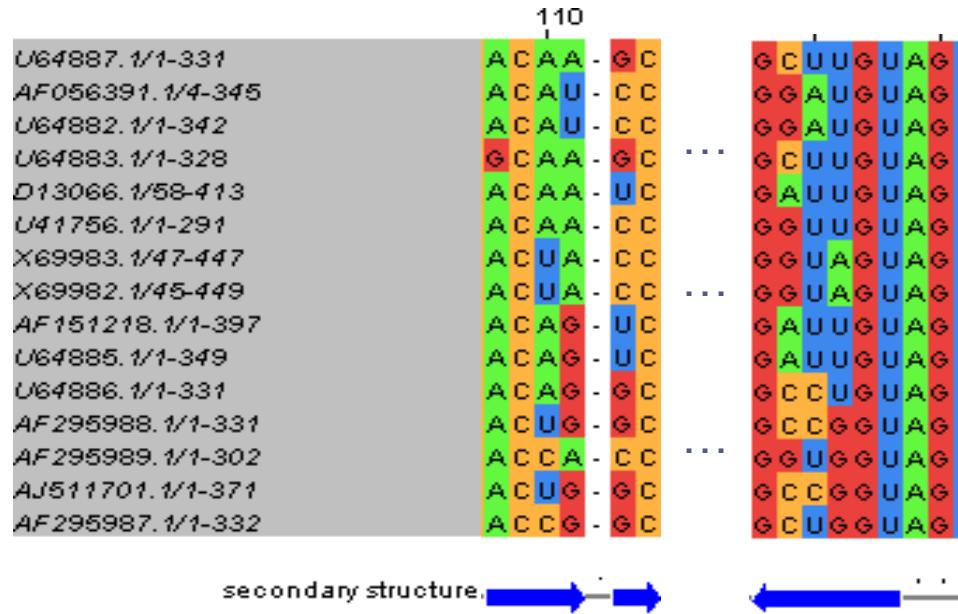
# SHAPE probing to model 2D



# Evolution to the rescue: Comparative approaches for structured RNAs



F Tahí  
IBISC Evry



RFAM Bacterial RNase P class B Alignment  
RF00011, rendered using JalView



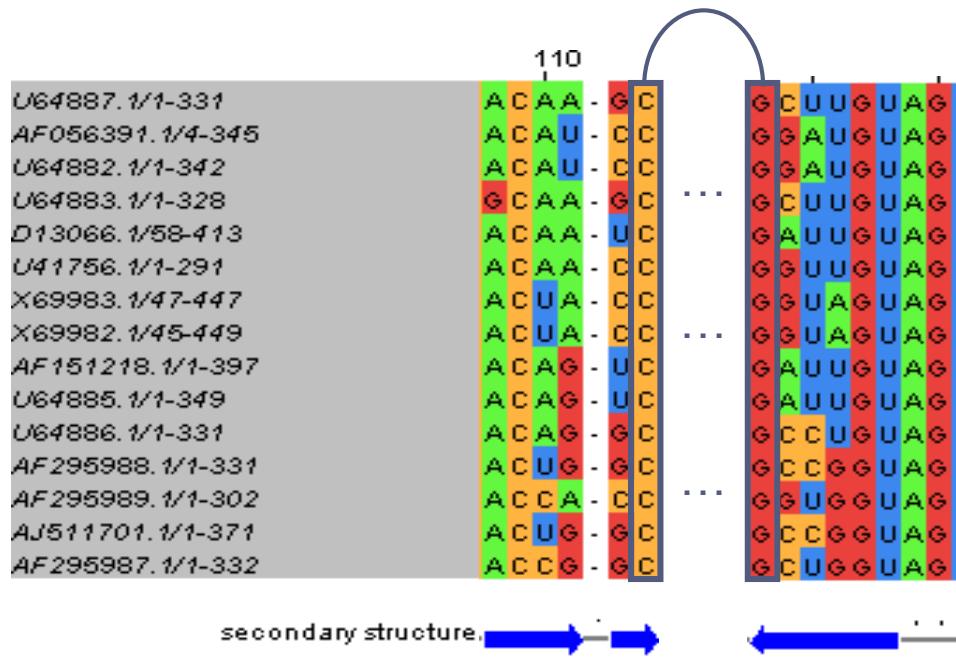
H Touzet  
Cristal, Lille I

- ▶ **Structure** ( $\approx$  phenotype) typically more conserved than **sequence**
- ▶ Covariations/compensatory mutations hint towards **shared structure**

# Evolution to the rescue: Comparative approaches for structured RNAs



F Tahí  
IBISC Evry



H Touzet  
Cristal, Lille I

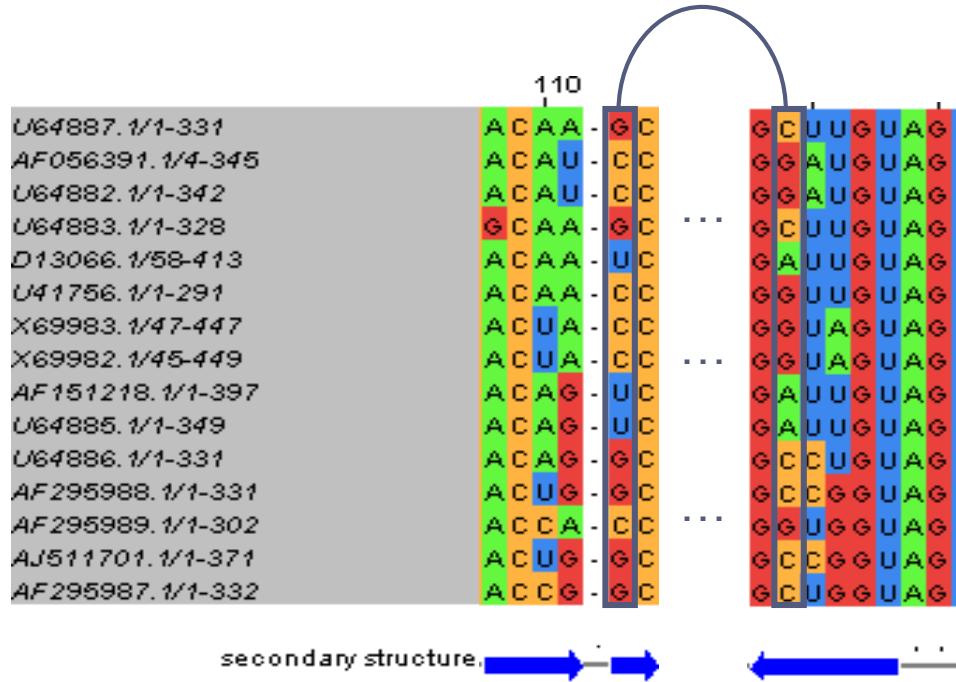
RFAM Bacterial RNase P class B Alignment  
RF00011, rendered using JalView

- ▶ **Structure** ( $\approx$  phenotype) typically more conserved than **sequence**
- ▶ Covariations/compensatory mutations hint towards **shared structure**

# Evolution to the rescue: Comparative approaches for structured RNAs



F Tahí  
IBISC Evry



RFAM Bacterial RNase P class B Alignment  
RF00011, rendered using JalView



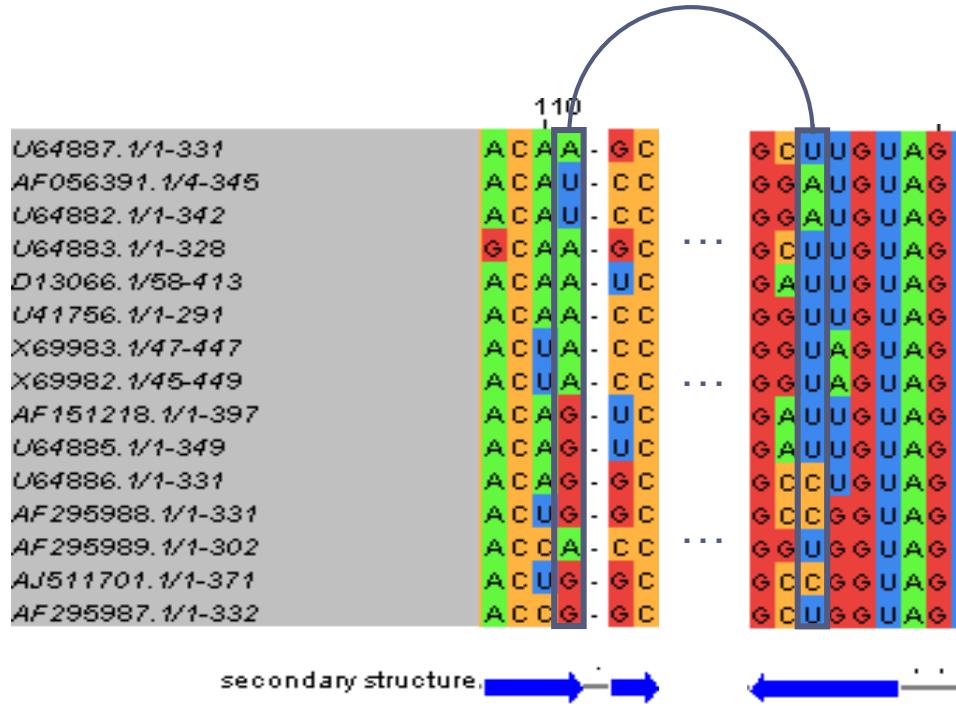
H Touzet  
Cristal, Lille I

- ▶ **Structure** ( $\approx$  phenotype) typically more conserved than **sequence**
- ▶ Covariations/compensatory mutations hint towards **shared structure**

# Evolution to the rescue: Comparative approaches for structured RNAs



F Tahí  
IBISC Evry



RFAM Bacterial RNase P class B Alignment  
RF00011, rendered using JalView



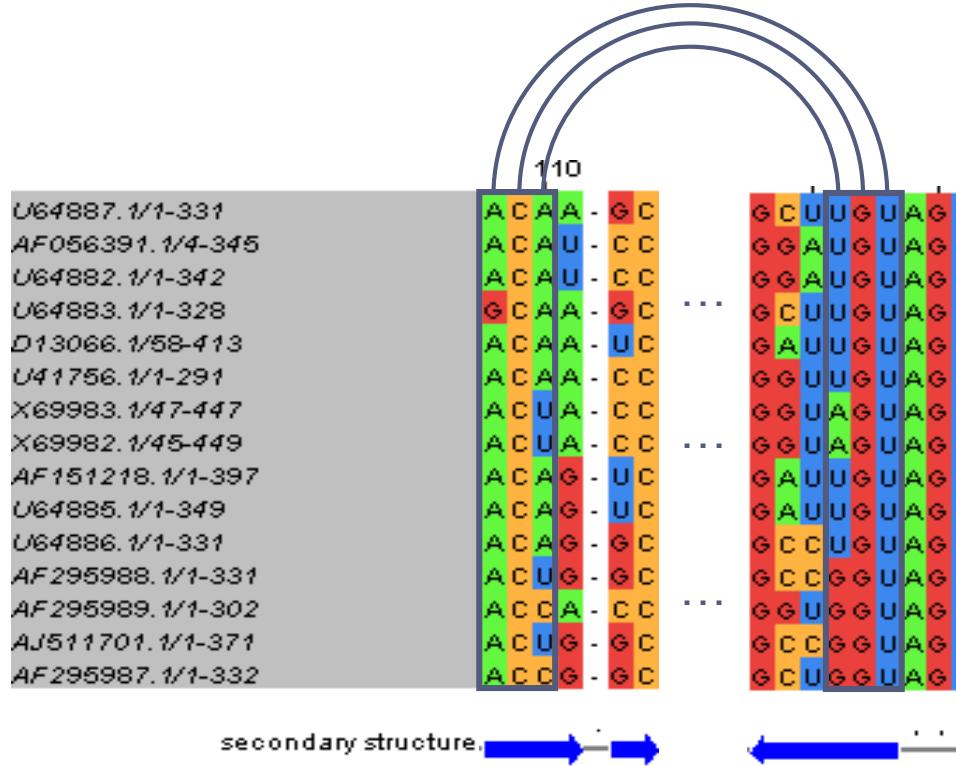
H Touzet  
Cristal, Lille I

- ▶ **Structure** ( $\approx$  phenotype) typically more conserved than **sequence**
- ▶ Covariations/compensatory mutations hint towards **shared structure**

# Evolution to the rescue: Comparative approaches for structured RNAs



F Tahí  
IBISC Evry



RFAM Bacterial RNase P class B Alignment  
RF00011, rendered using JalView

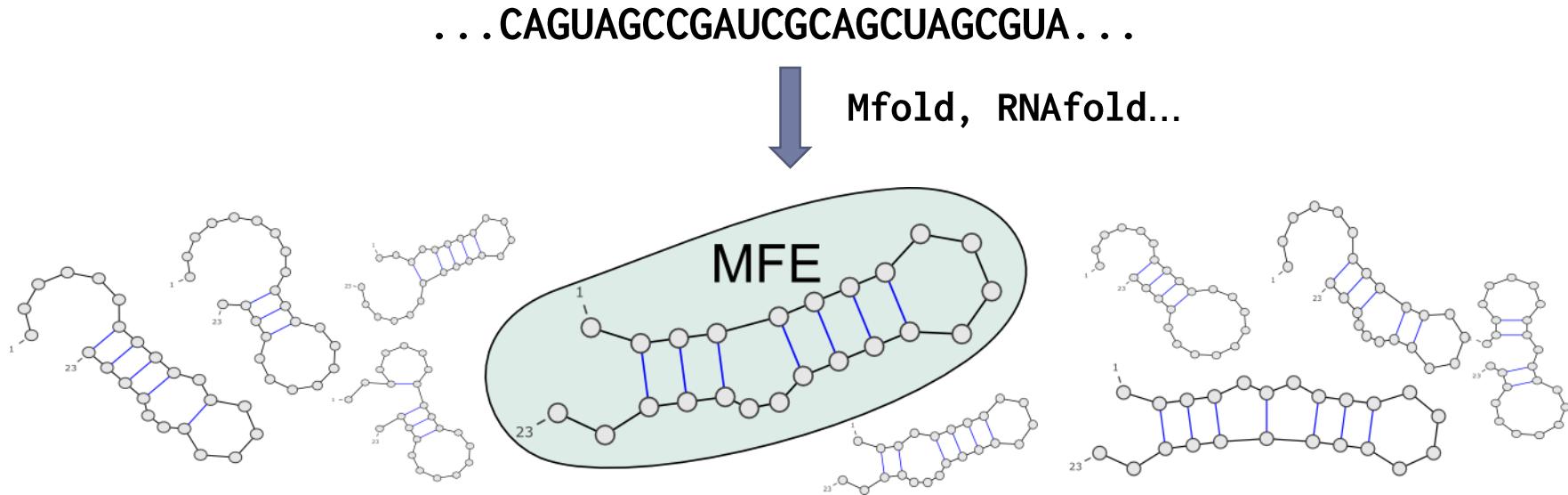


H Touzet  
Cristal, Lille I

- ▶ **Structure** ( $\approx$  phenotype) typically more conserved than **sequence**
- ▶ Covariations/compensatory mutations hint towards **shared structure**

# Ensemble approaches in RNA folding

- ▶ RNA *in silico* paradigm shift:
  - ▶ From single structure, minimal free-energy folding...

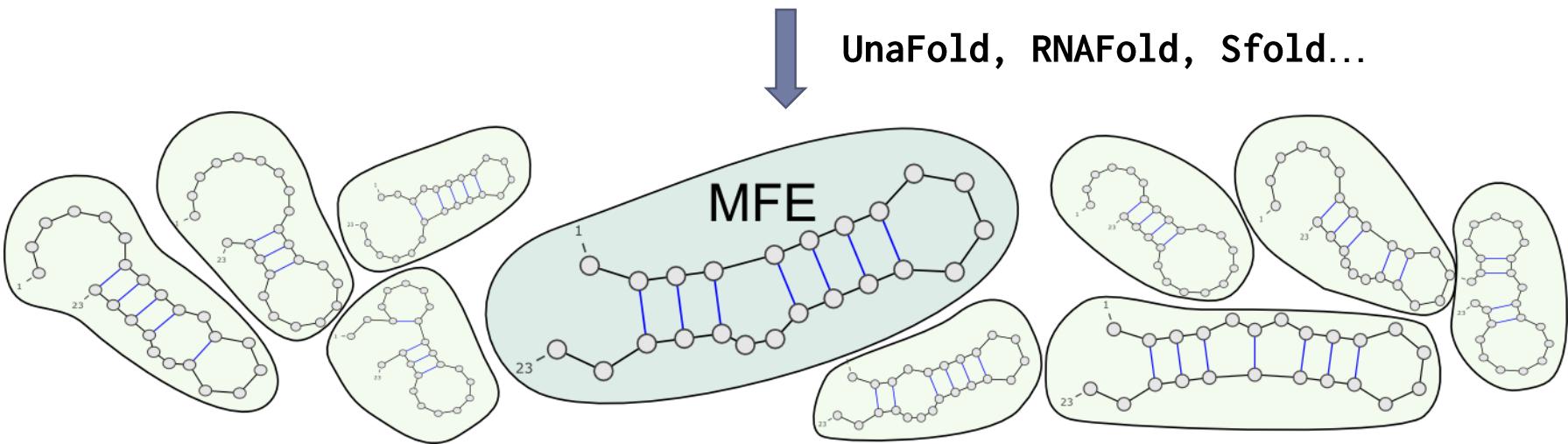


# Ensemble approaches in RNA folding

## ► RNA *in silico* paradigm shift:

- From single structure, minimal free-energy folding...
- ... to ensemble approaches.

... CAGUAGCCGAUCGCAGCUAGCGUA ...



Thermodynamic equilibrium: Each secondary structure has probability

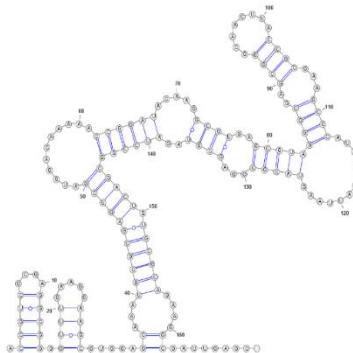
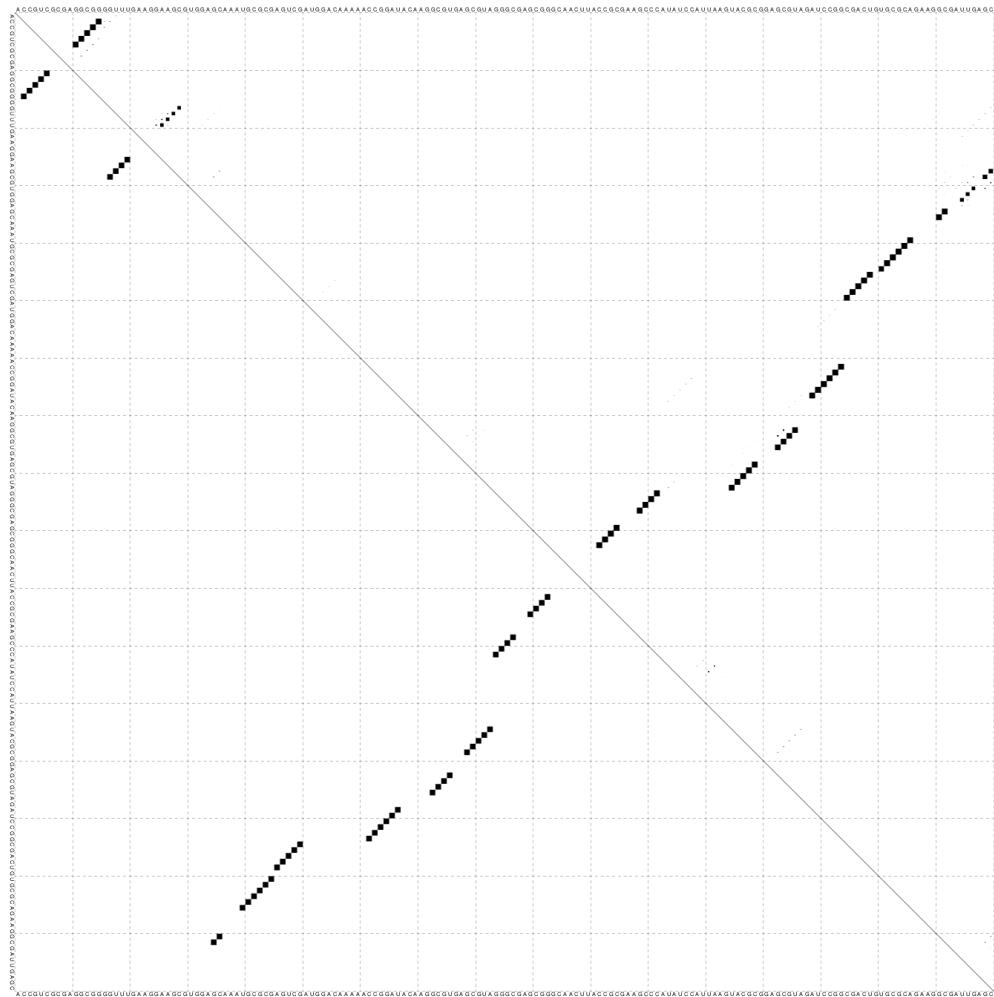
$$\text{Boltzmann Probability } Prob(S) = \frac{e^{-\Delta G(S)/kT}}{Z}$$

$$\text{Partition Function } Z = \sum_{Struc S} e^{-\Delta G(S)/kT}$$

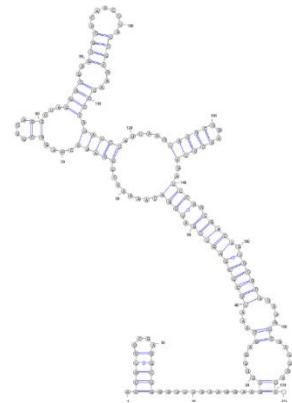
[McCaskill, Biopolymers 1990 ]

# Ensemble-based methods are more robust than energy-minimizing approaches

Back to D1D4 example...



Turner 2004

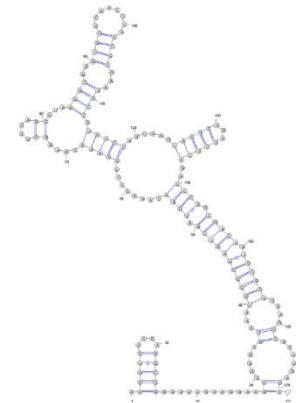
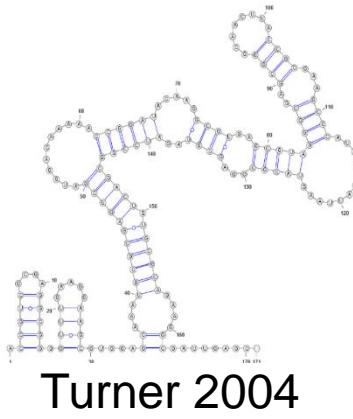
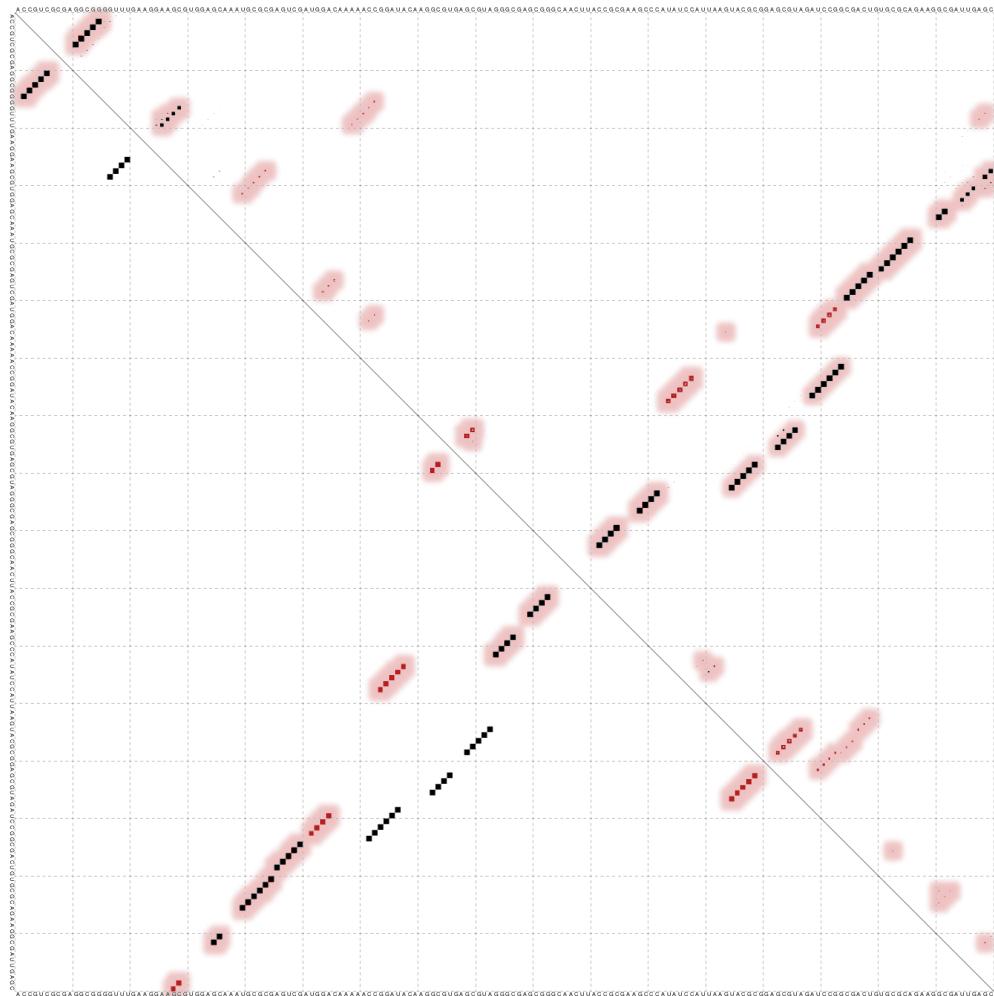


Andronescu 2007

BP>99% → PPV>90%  
BP>90% → PPV>83%...  
[Mathews, RNA 2004]

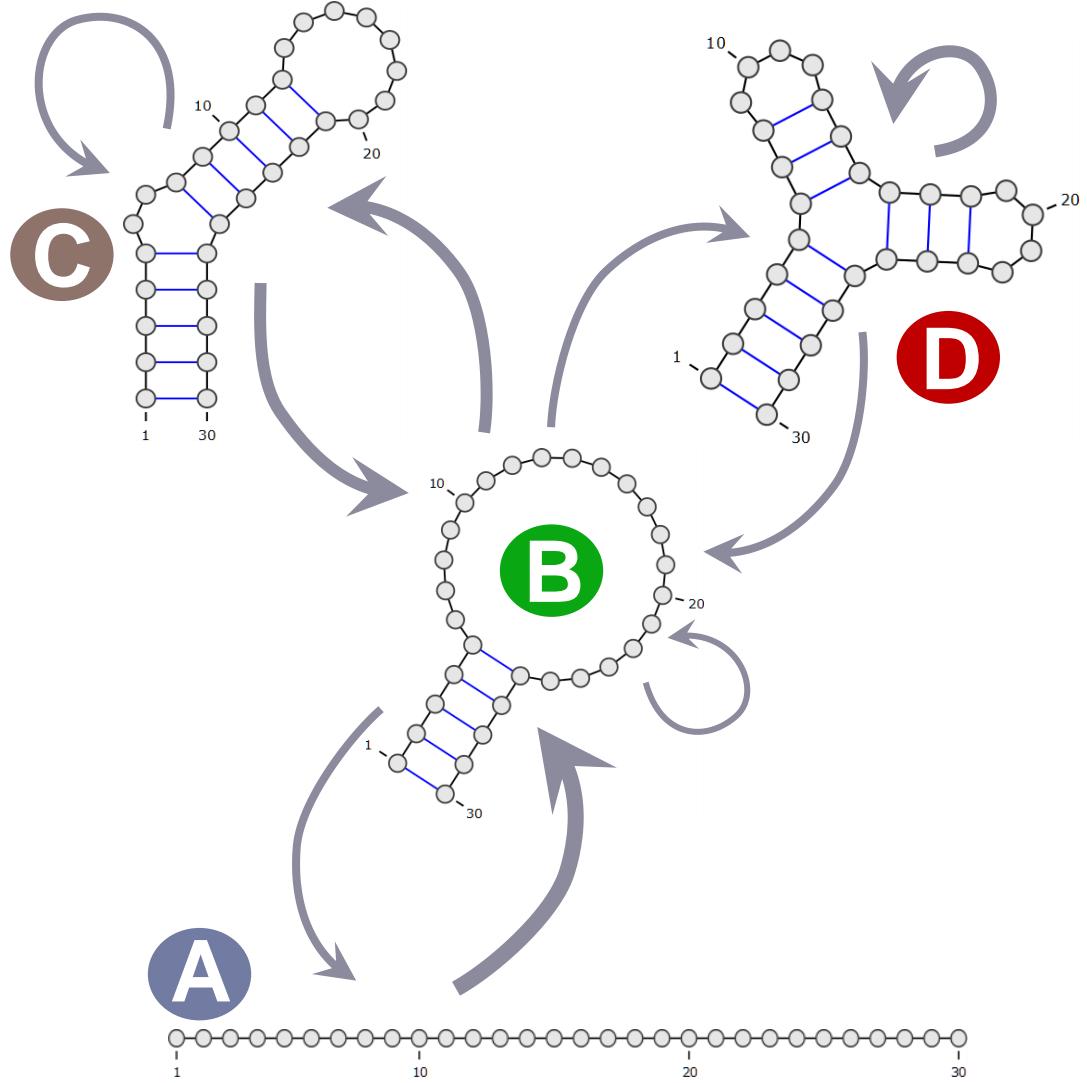
# Ensemble-based methods are more robust than energy-minimizing approaches

Back to D1D4 example...

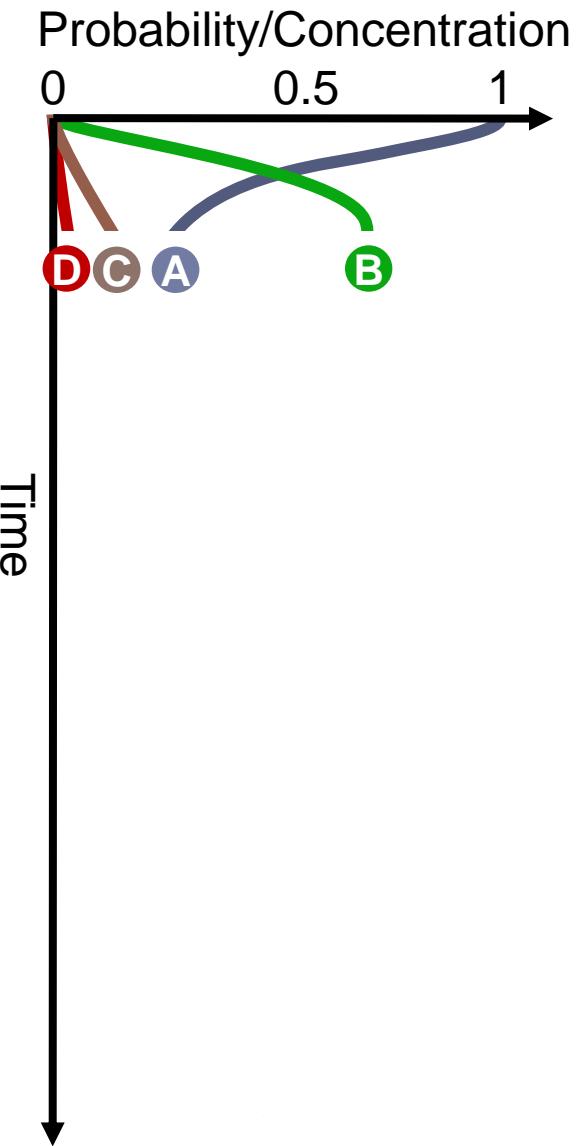
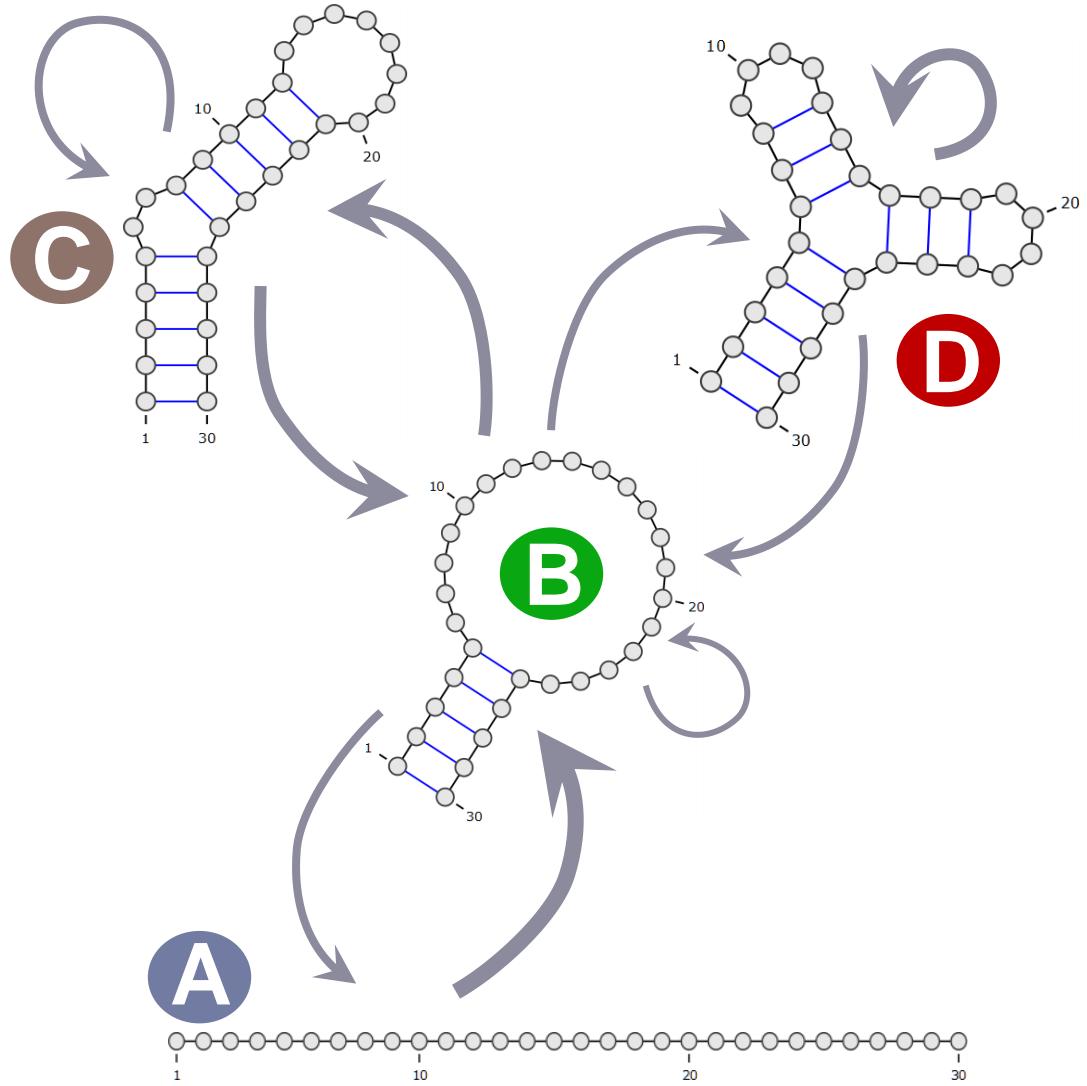


BP>99% → PPV>90%  
BP>90% → PPV>83%...  
[Mathews, RNA 2004]

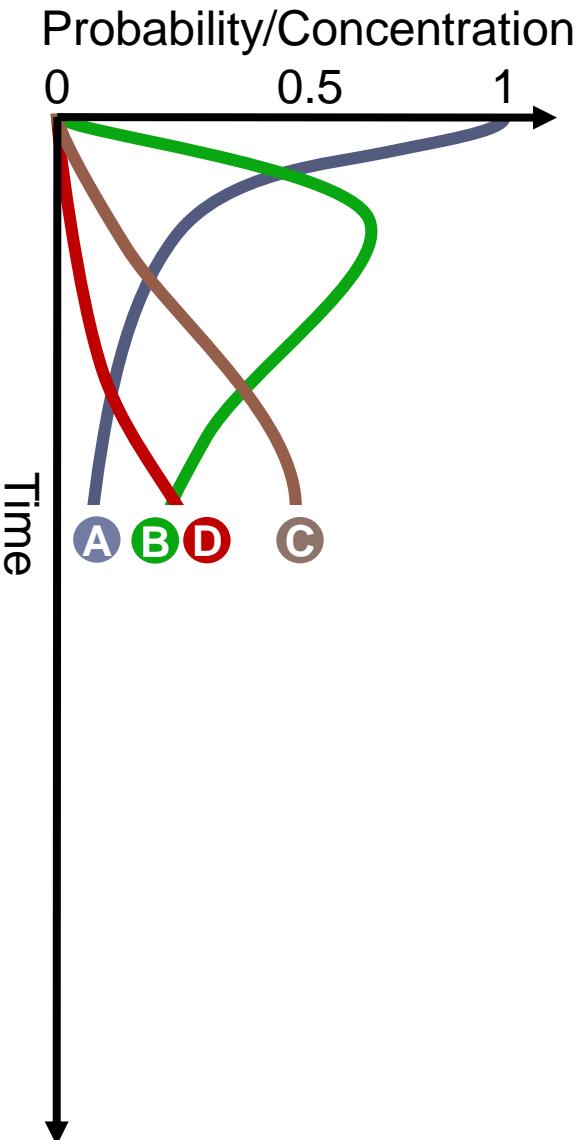
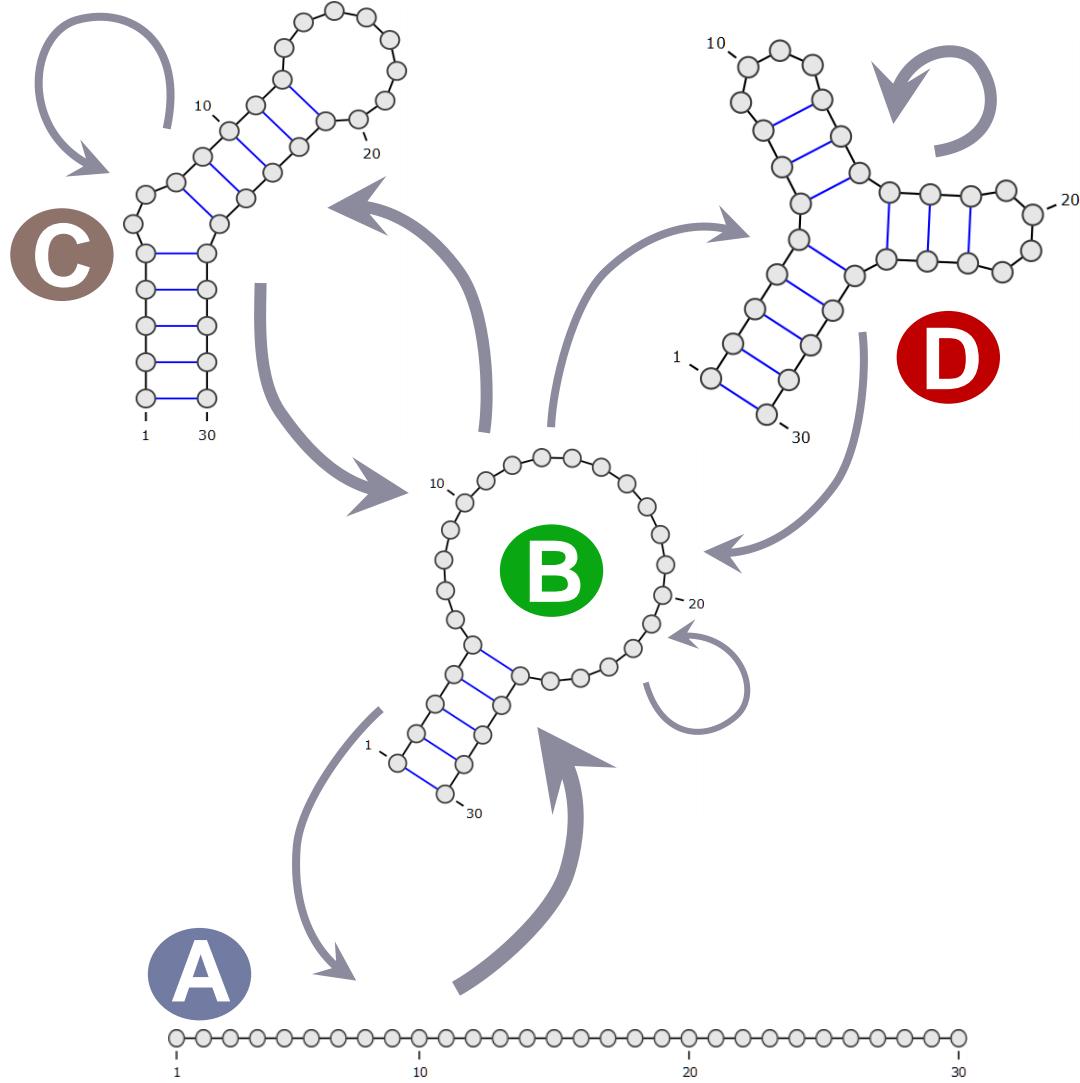
# RNA kinetics



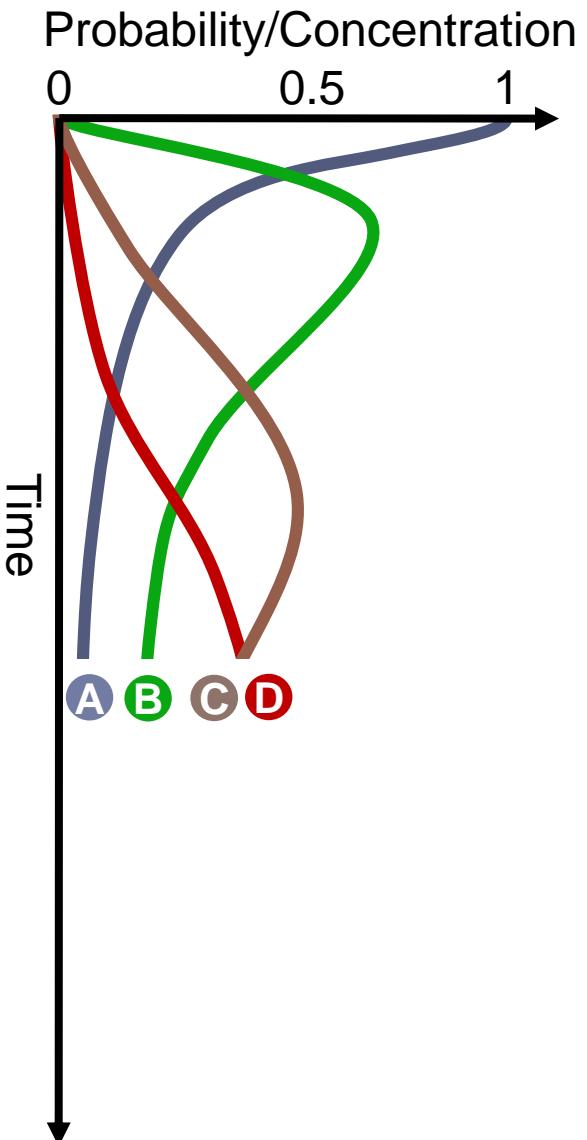
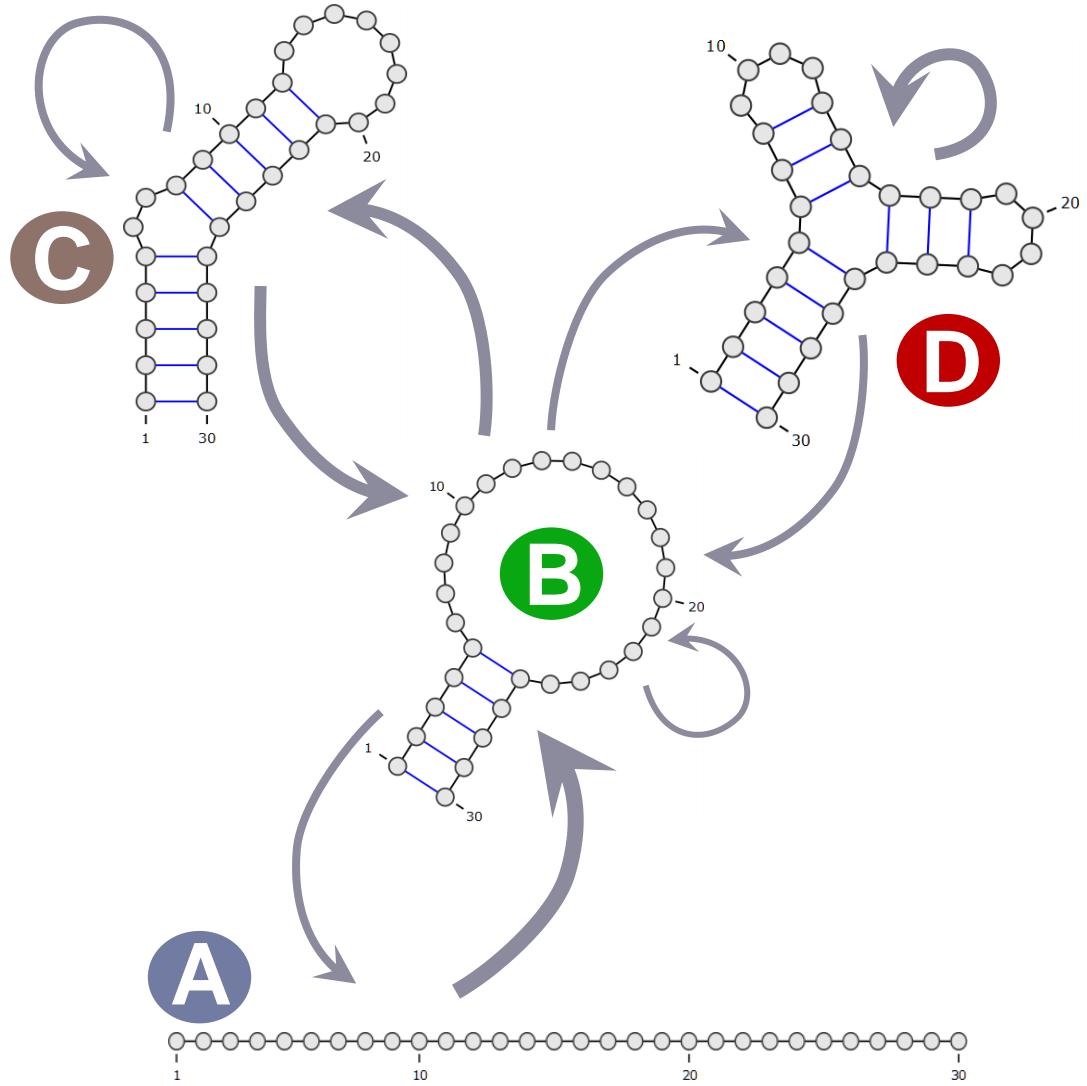
# RNA kinetics



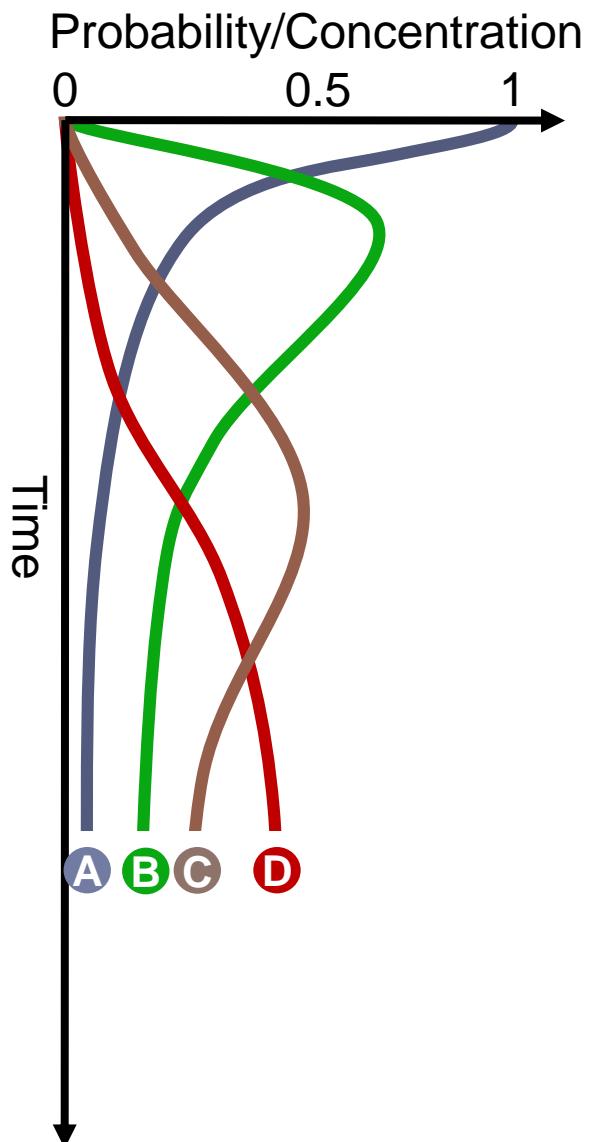
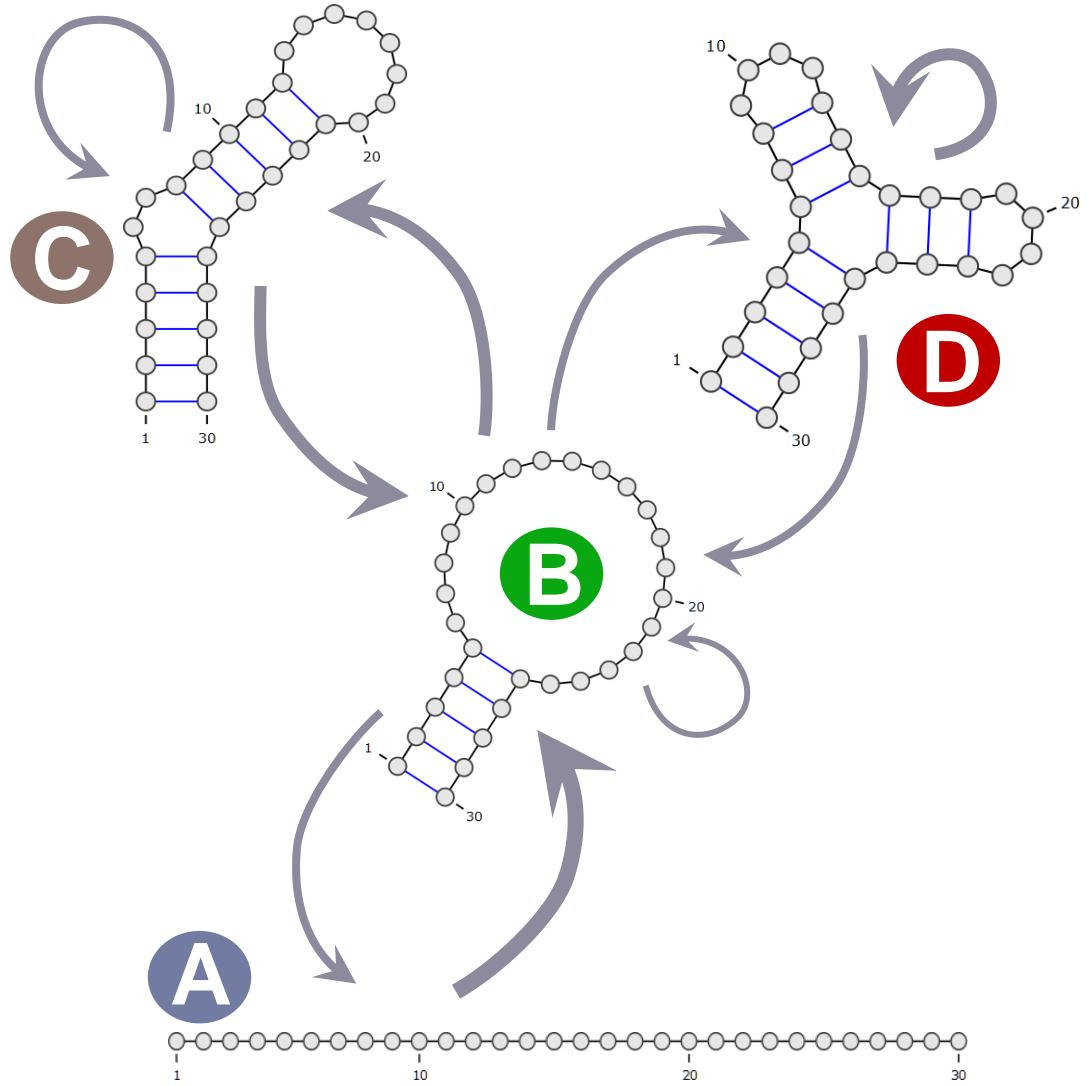
# RNA kinetics



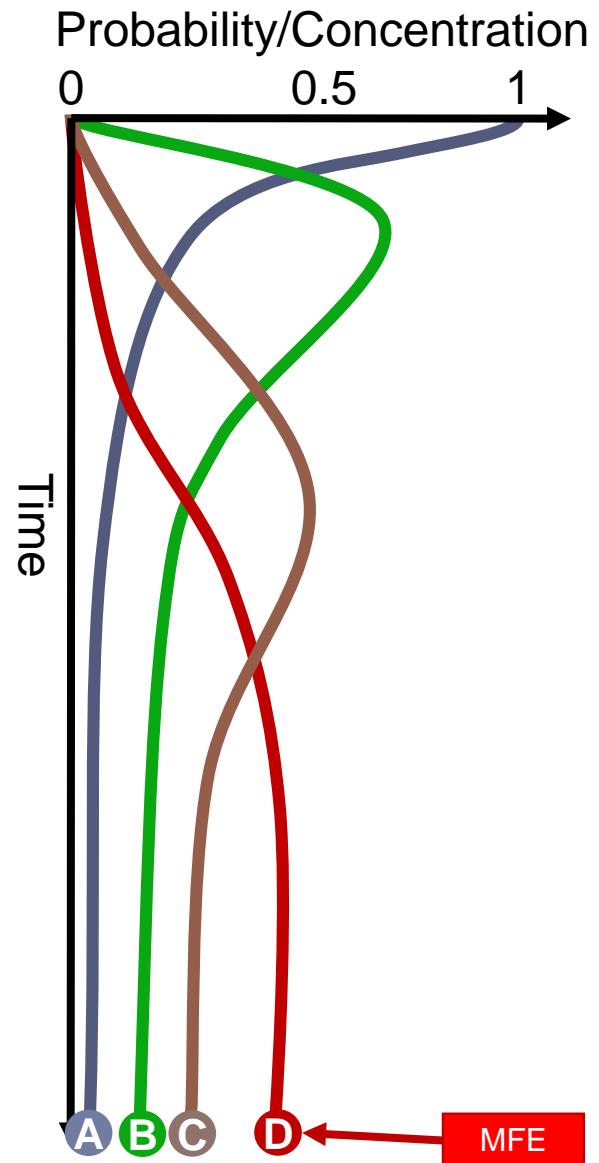
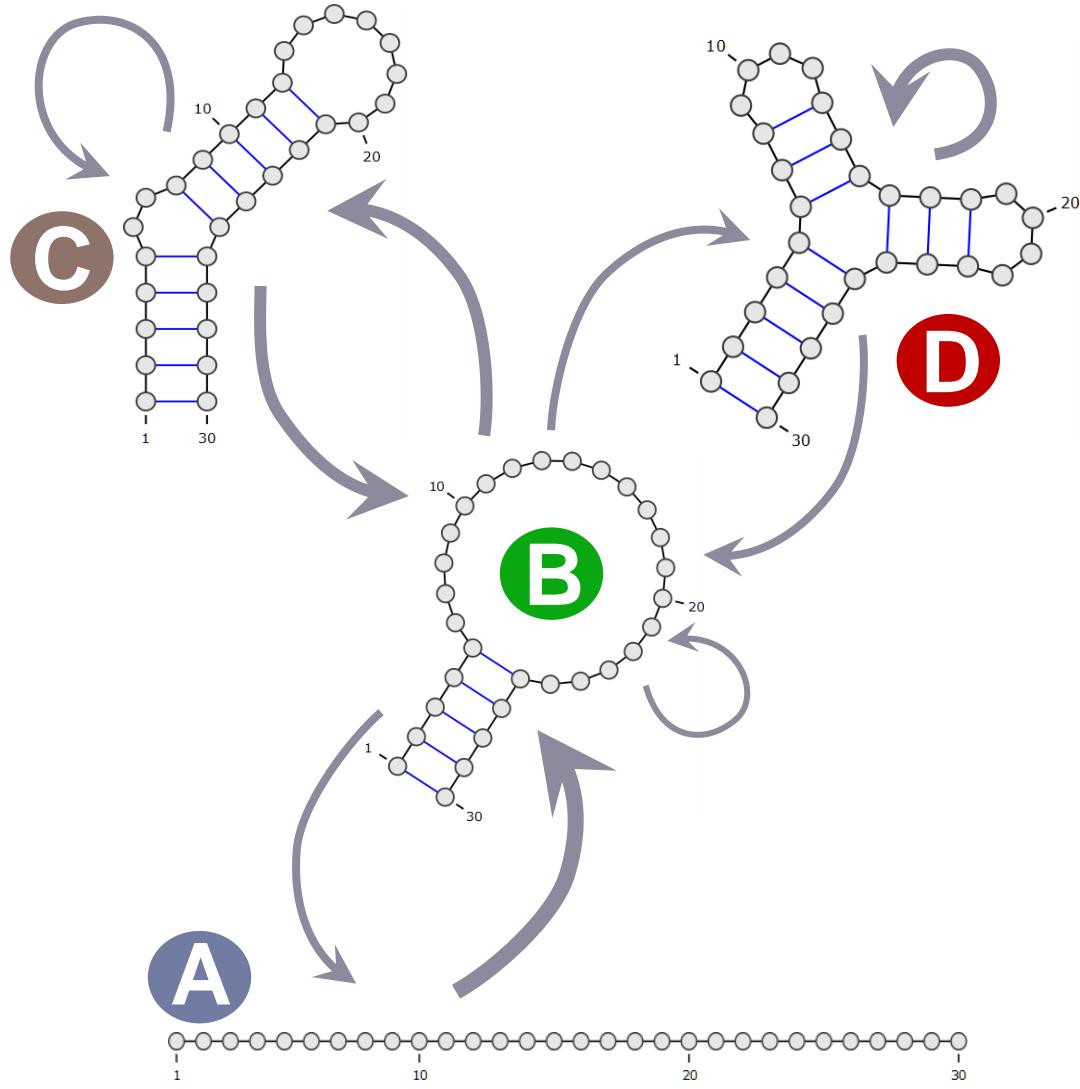
# RNA kinetics



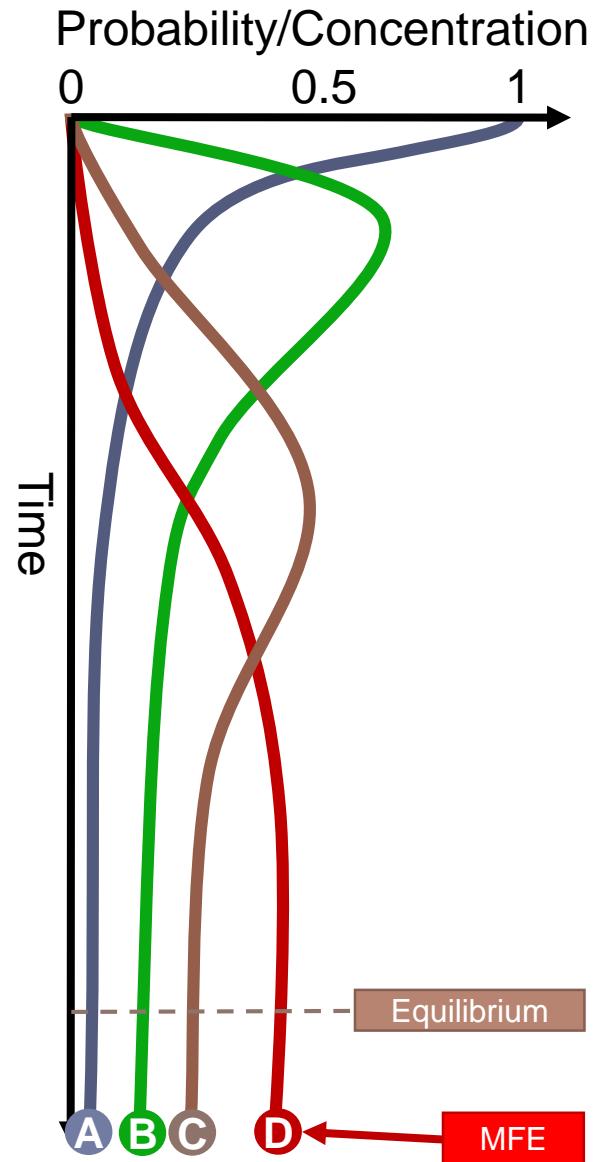
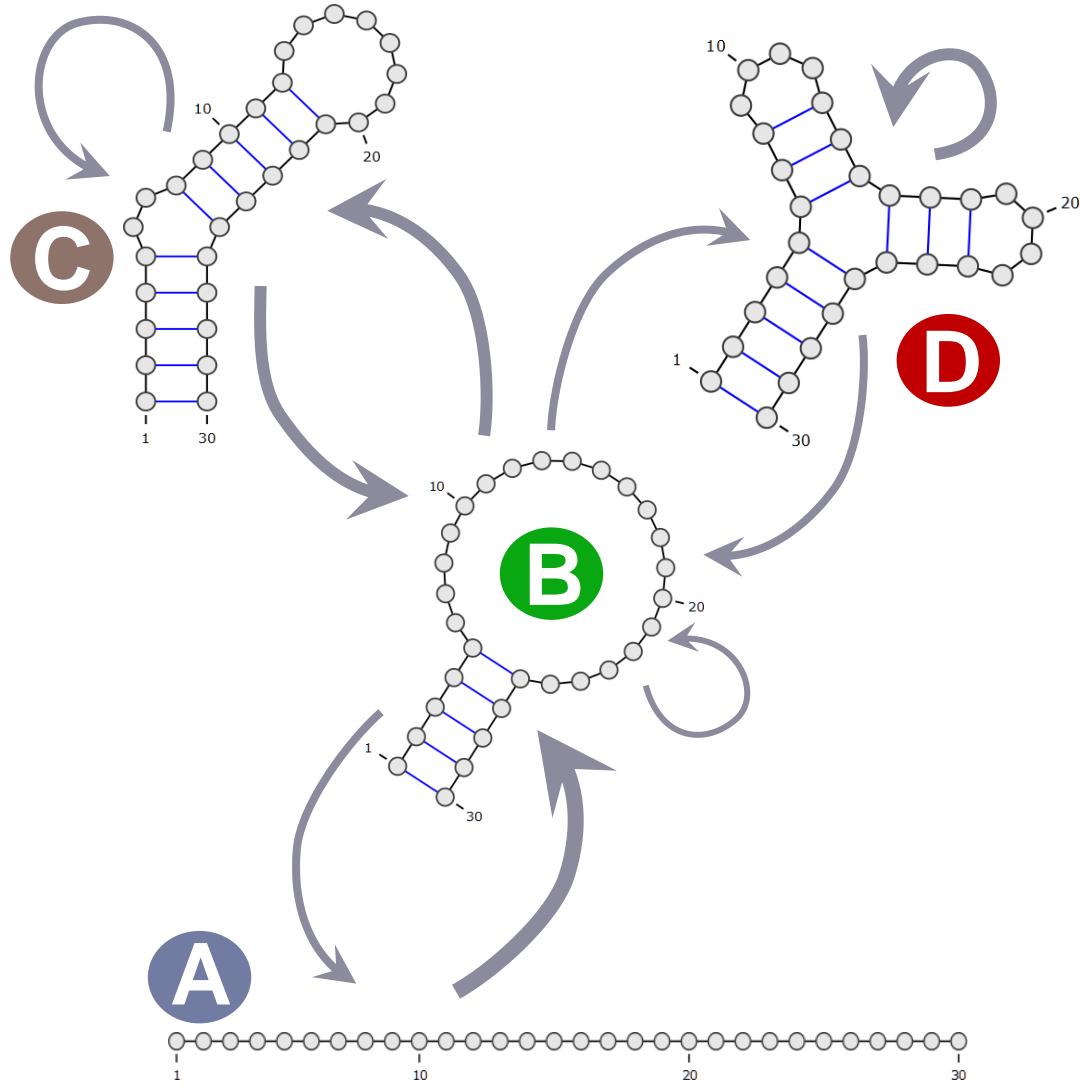
# RNA kinetics



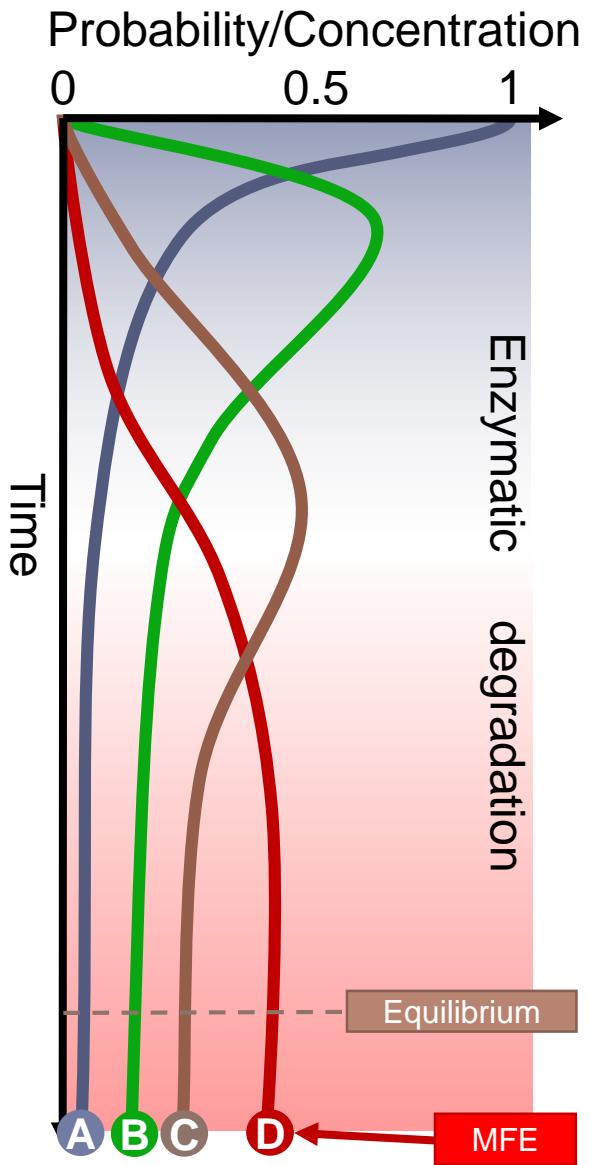
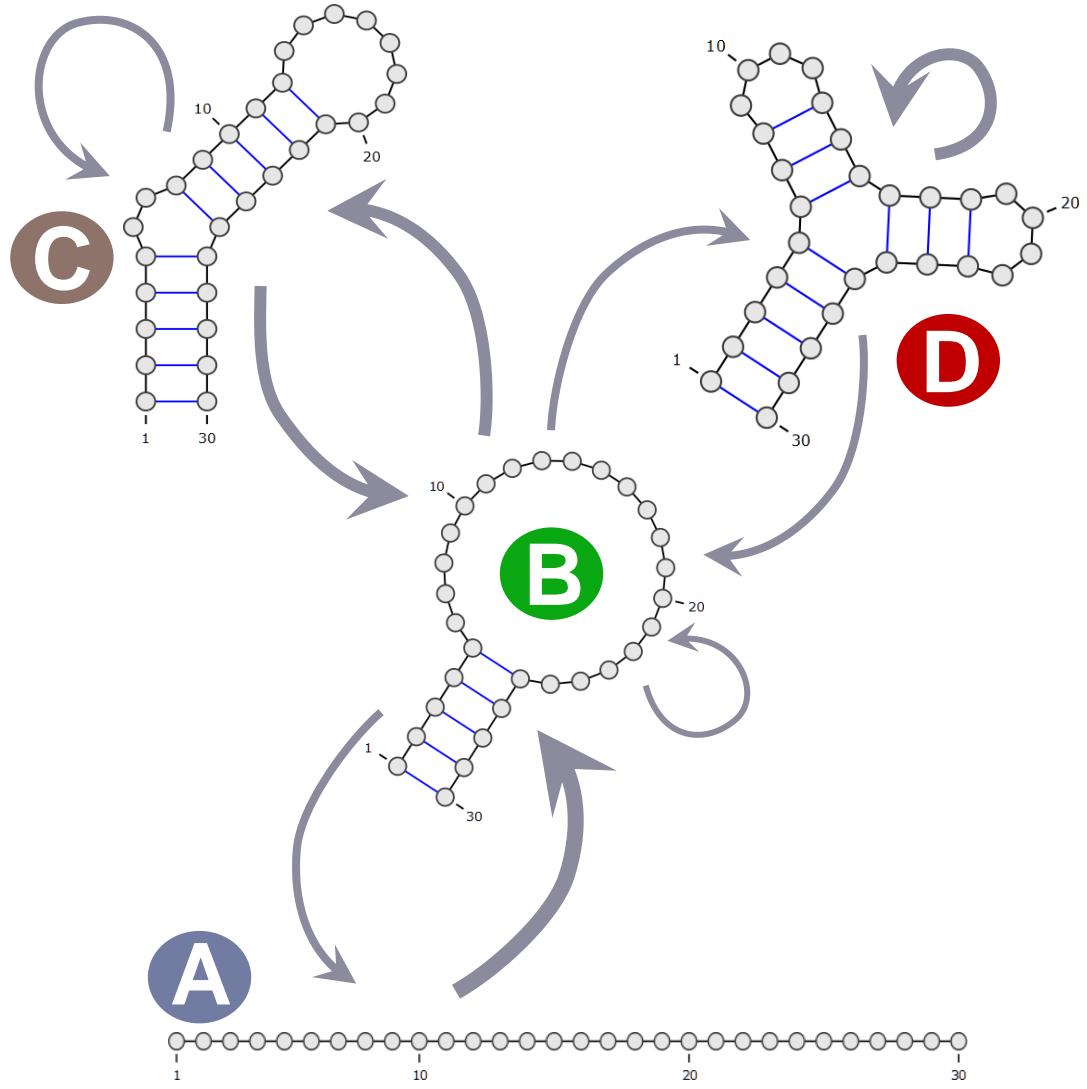
# RNA kinetics



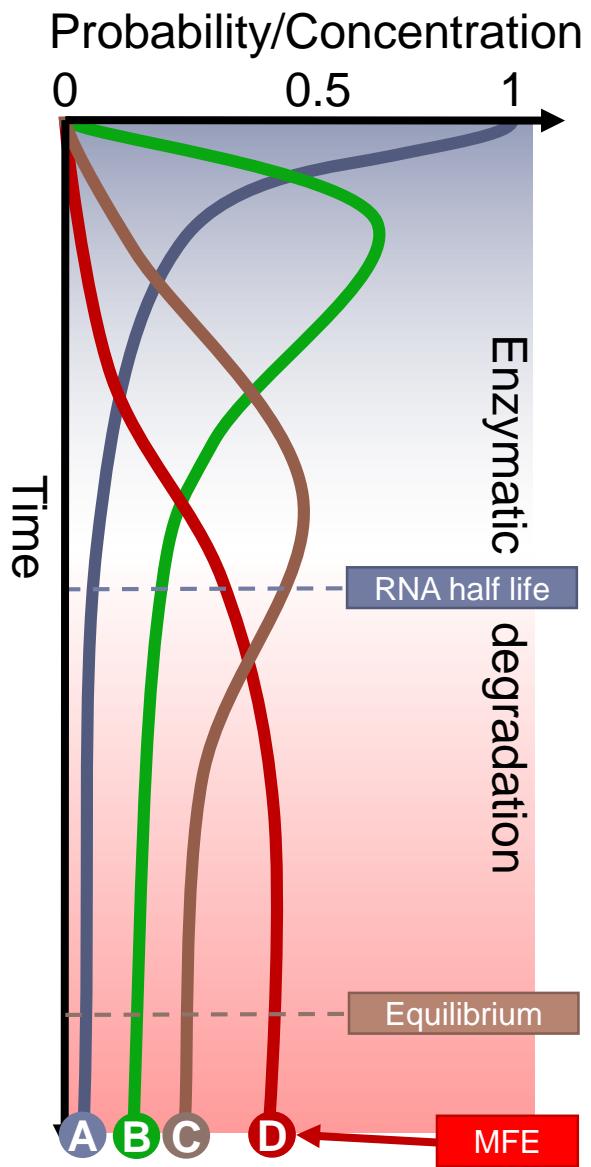
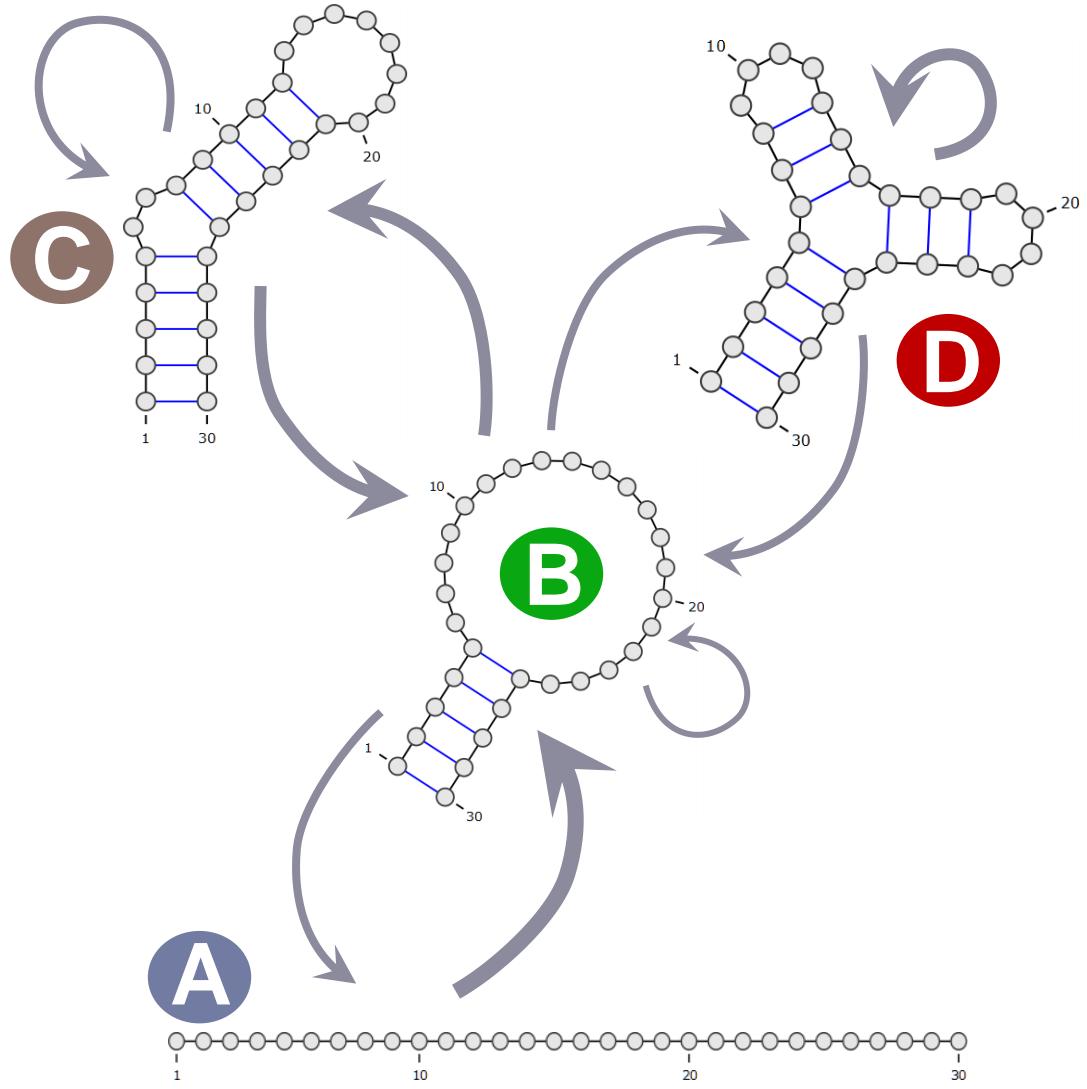
# RNA kinetics



# RNA kinetics



# RNA kinetics



# Beyond 2D

- ▶ Molecular dynamics (coarse-grained)



**S Pasquali**

Pharma, Paris Descartes

# Beyond 2D

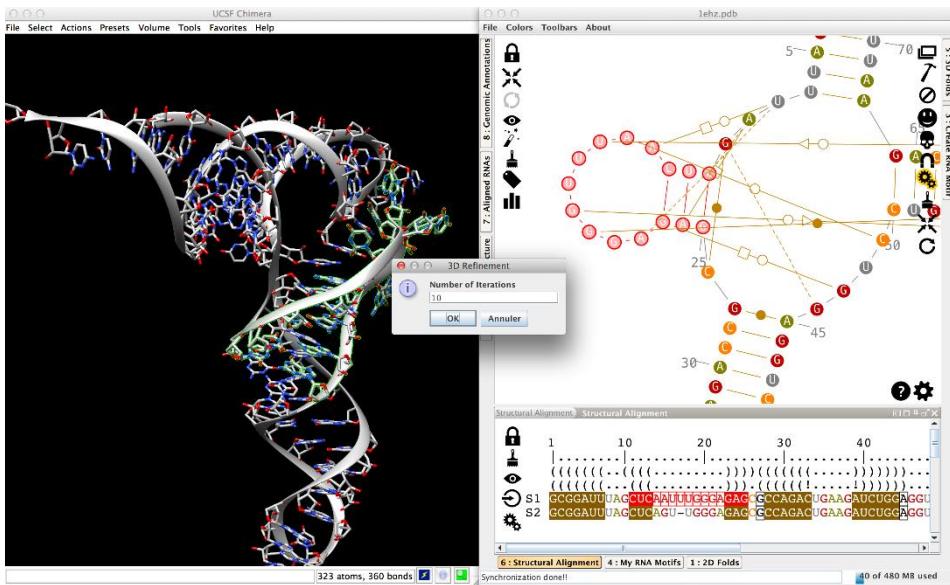
- ▶ Molecular dynamics (coarse-grained)



S Pasquali

Pharma, Paris Descartes

- ▶ Computer-assisted modeling (Assemble)



F Jossinet

Univ. Strasbourg

# Beyond 2D

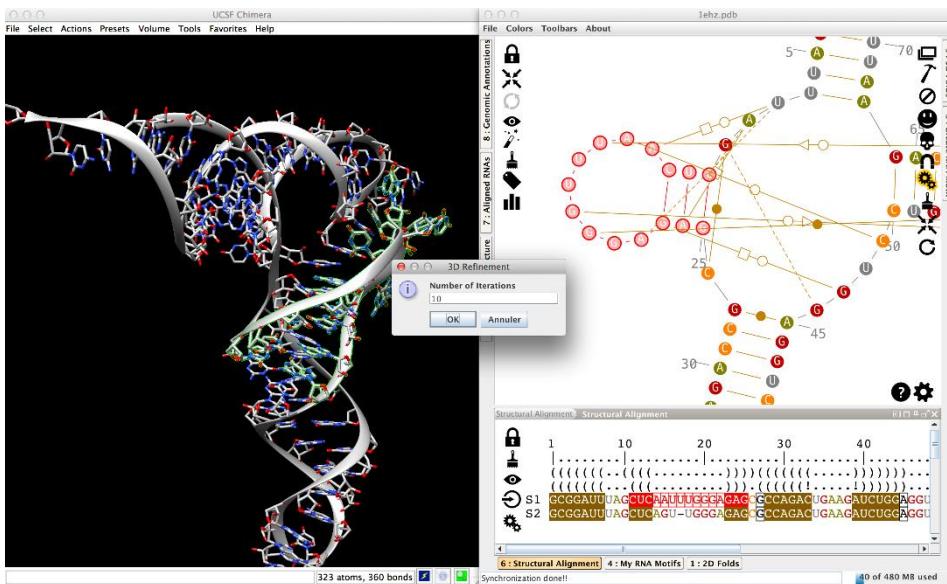
- ▶ Molecular dynamics (coarse-grained)



S Pasquali

Pharma, Paris Descartes

- ▶ Computer-assisted modeling (Assemble)



F Jossinet

Univ. Strasbourg



F Leclerc

I2BC, Paris-Sud

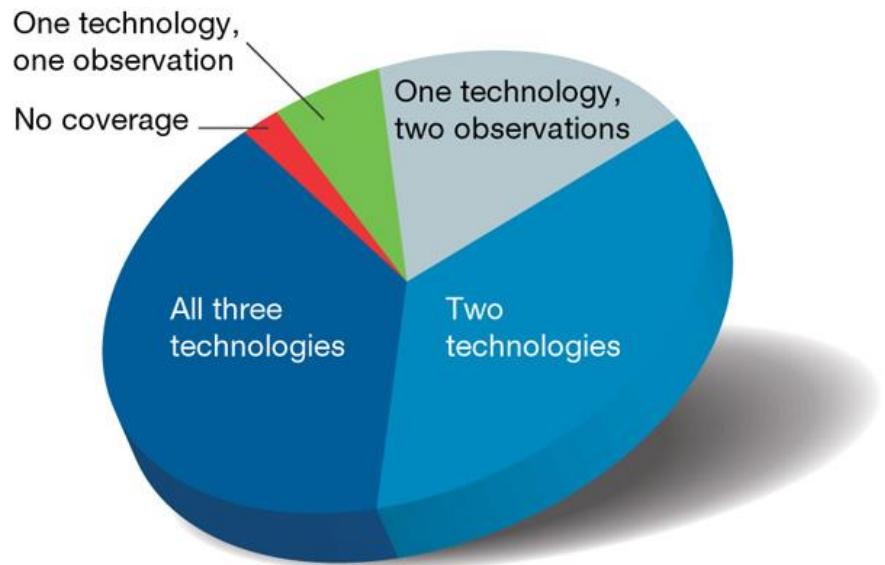
- ▶ RNA-Protein Interactions



Sounds good, but ...  
why should we care?

# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed

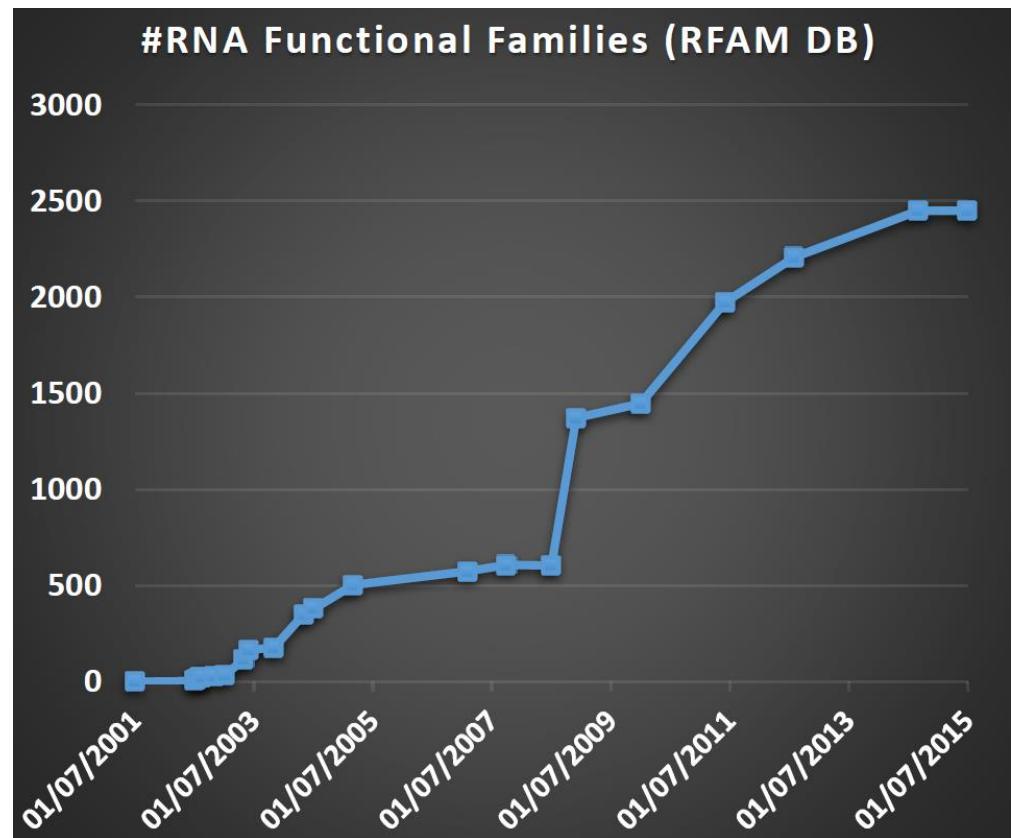


The human genome is **pervasively transcribed**, such that the majority of its bases are associated with at least one primary transcript and many transcripts link distal regions to established protein-coding loci.

*ENCODE Analysis of 1% of the human genome Nature 2007*

# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
  - Carriers
  - Transporter
  - Enzymatic
  - Processing
  - Regulatory
  - ssRNA genomes (HIV)
  - Immune system (CRISPR)
  - And more

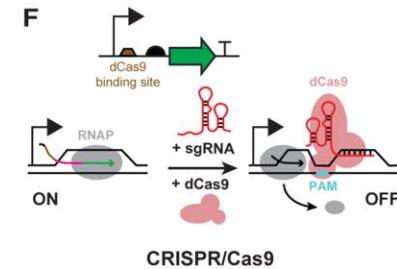
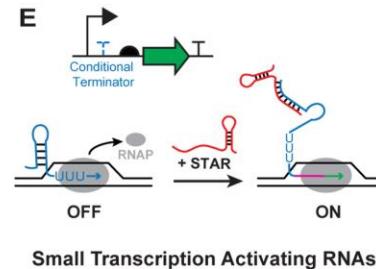
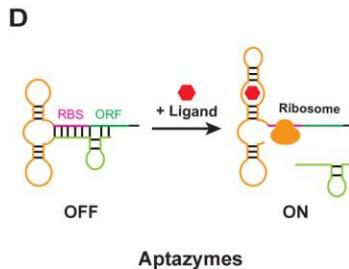
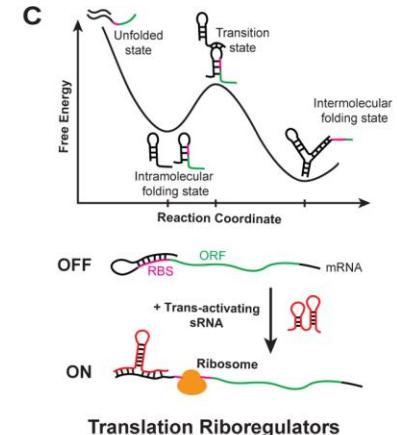
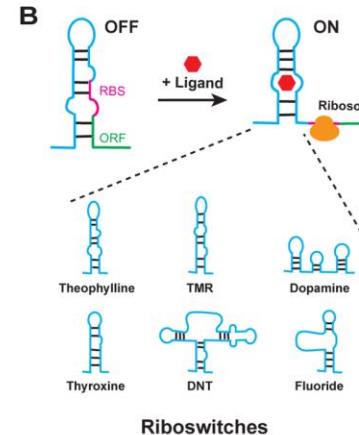
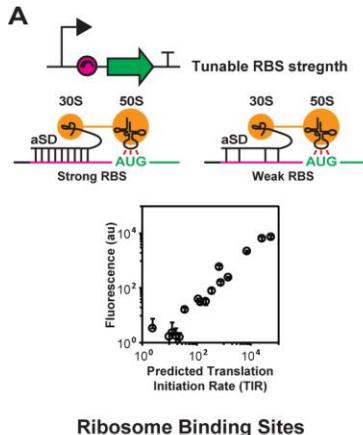


# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology



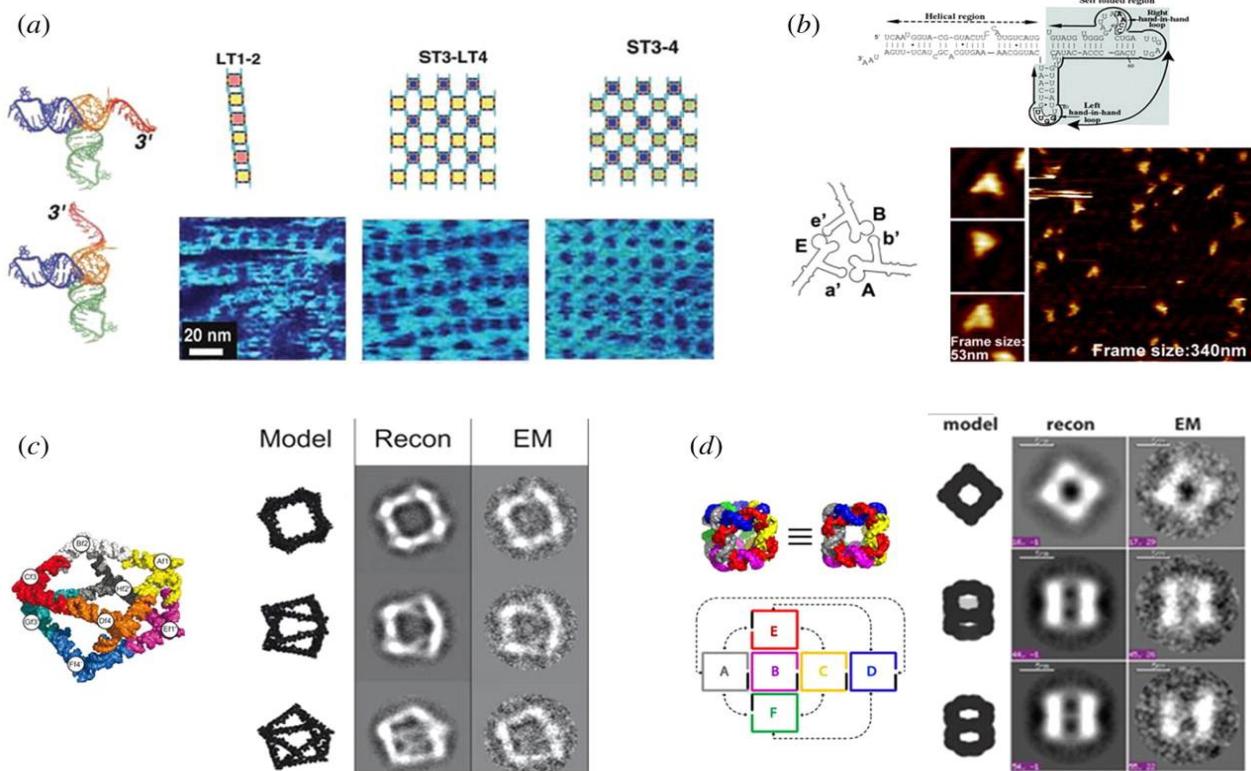
A Jaramillo  
Evry/Warwick



[Kushwaha *et al*, ACS Synthetic Bio. 2016]

# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology
  - ▶ Nanotech

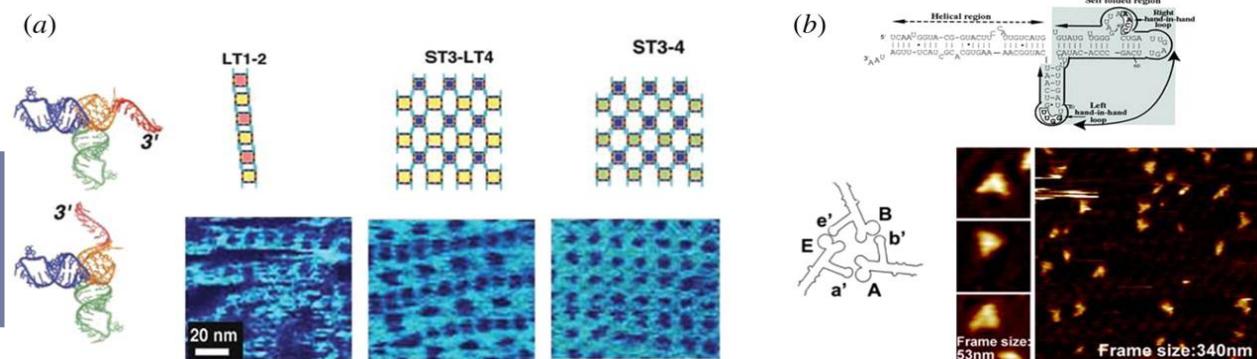


RNA-based Nanoarchitectures  
[Li H *et al*, Interface Focus 2011]

# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle

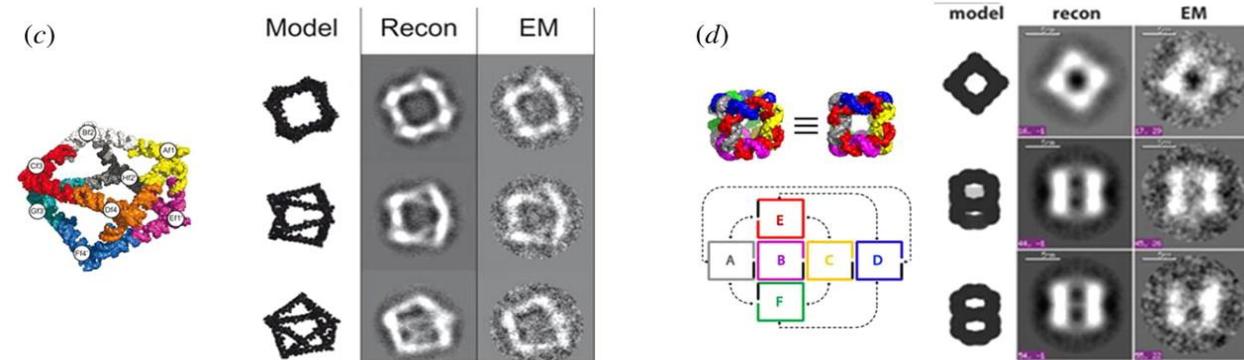
- ▶ Synthetic biology
- ▶ Nanotech



## Rational Design



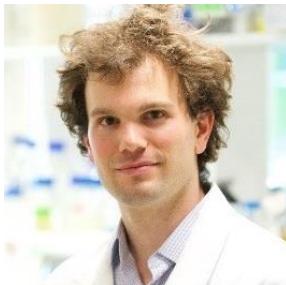
AMIBio  
LIX Polytechnique  
/Inria Saclay



RNA-based Nanoarchitectures  
[Li H *et al*, Interface Focus 2011]

# Why RNA is **totally awesome!**

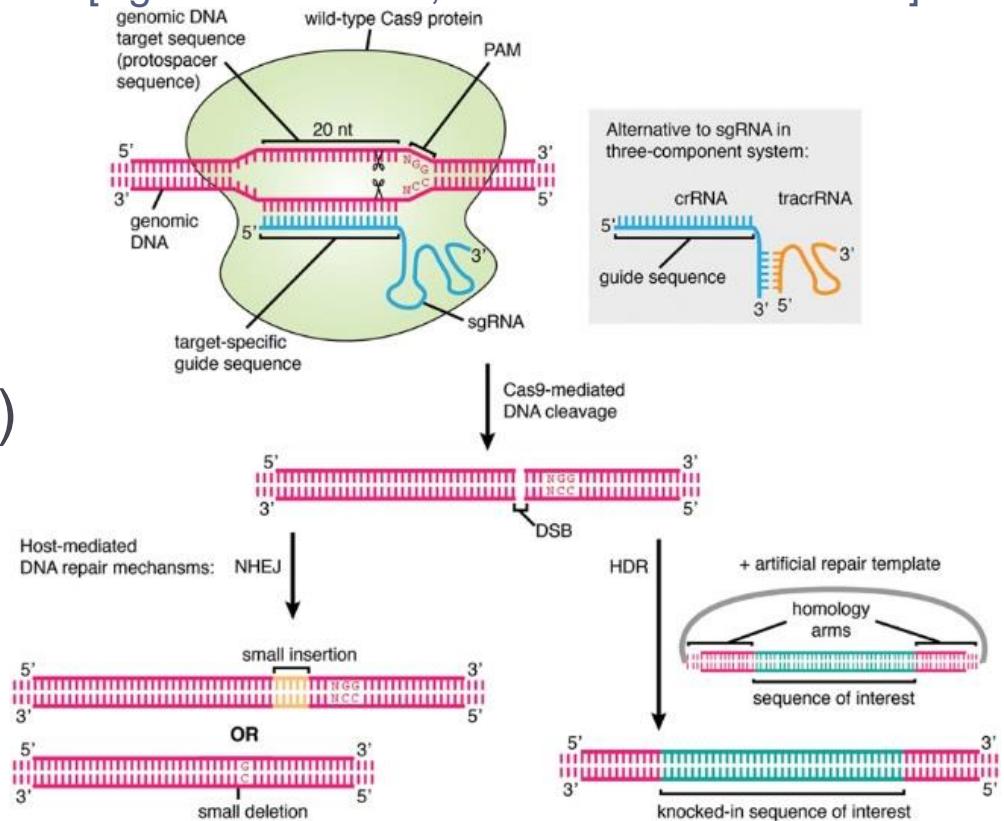
- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology
  - ▶ Nanotech
  - ▶ Genome editing (CRISPR)



D Bikard  
Pasteur

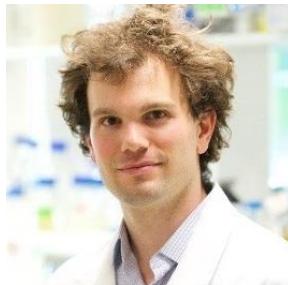
Blooming therapeutic RNAI...  
... making way for CRISPR!

[Agrotis & Ketteler, Frontiers Genetics 2015]



# Why RNA is **totally awesome!**

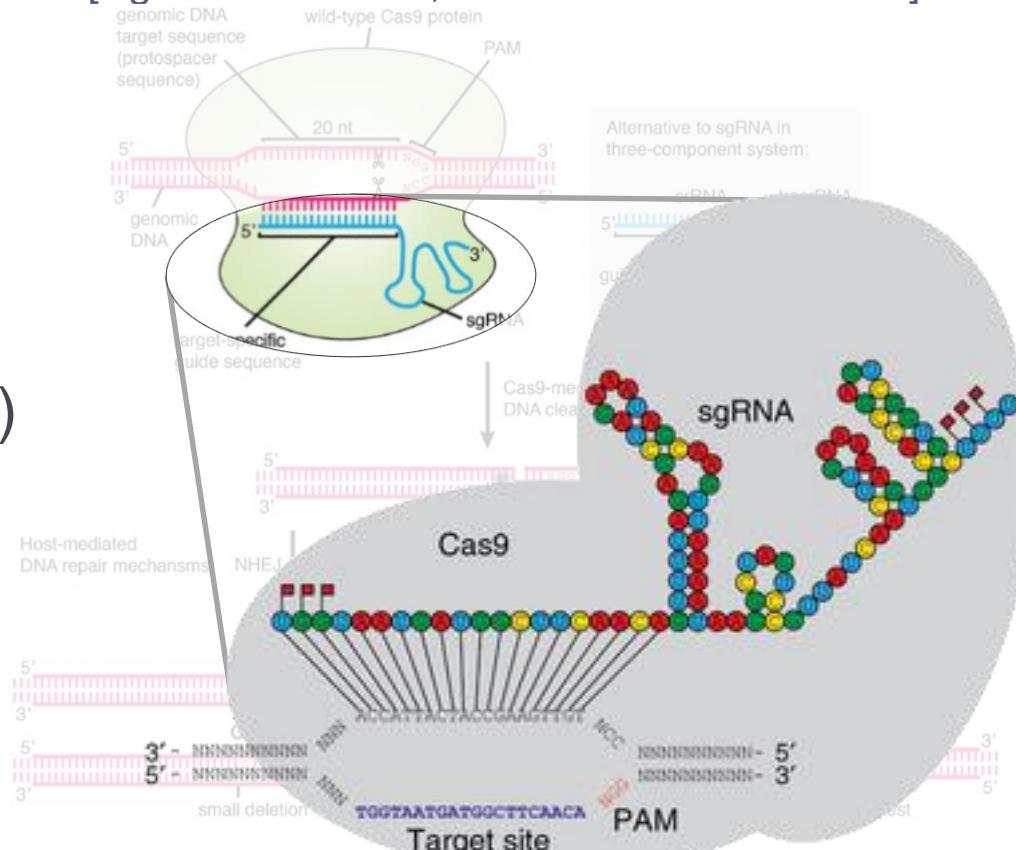
- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology
  - ▶ Nanotech
  - ▶ Genome editing (CRISPR)



D Bikard  
Pasteur

Blooming therapeutic RN...  
... making way for CRISPR!

[Agrotis & Ketteler, Frontiers Genetics 2015]



[Hendel et al, Nature Biotech. 2015]

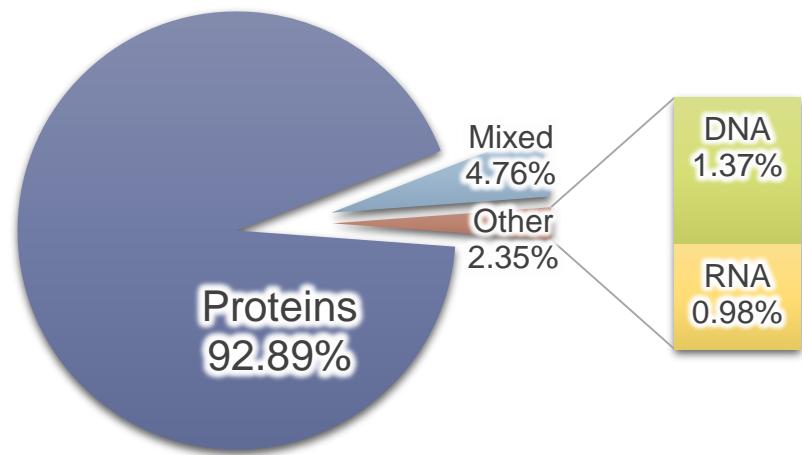
# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology
  - ▶ Nanotech
  - ▶ Genome editing (CRISPR)
  - ▶ Computationally fun  
(but still challenging...)



D Barth  
DAVID, UVSQ

PDB: 117,022 entries (March 2016)



**(Initial) lack of structural data**

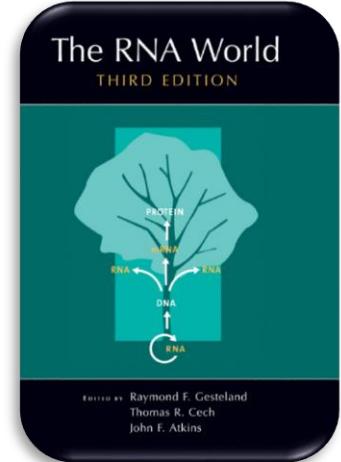
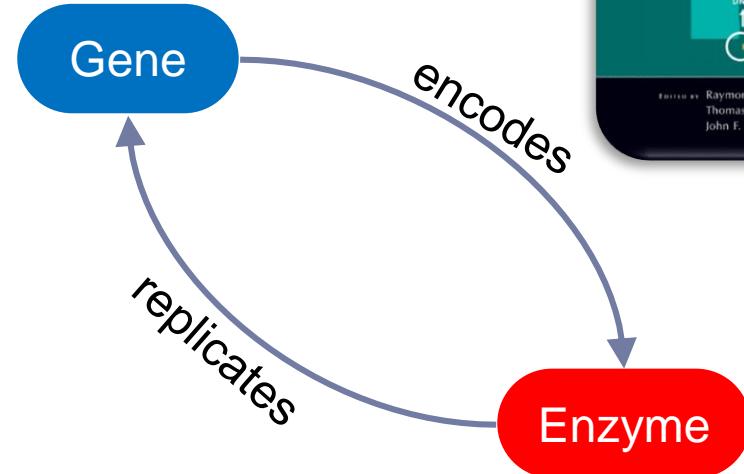
Experiment-based energy models  
+ Secondary structure  
+ Efficient combinatorial algorithms

⇒ Mature *ab initio* prediction tools  
(Mfold, RNAfold...)

# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology
  - ▶ Nanotech
  - ▶ Therapeutics and genetic engineering (CRISPR)
  - ▶ Computationally fun (but still challenging)
- ▶ RNA at the origin of life!?

*Chicken vs egg paradox  
at the origin of life*



This is the RNA World:

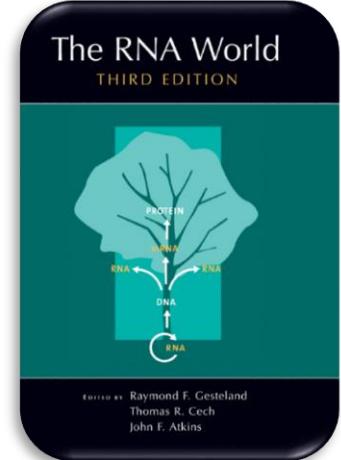
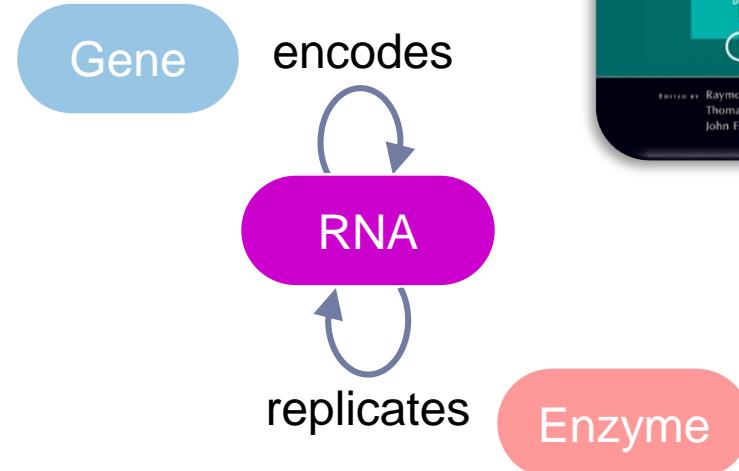
- **Proteins** are good at being enzymes but bad at being replicators;
- **DNA** is good at replicating but bad at being an enzyme;
- **RNA** might just be good enough at both roles to break out of the Catch-22.

**R. Dawkins.** The Ancestor's tale

# Why RNA is **totally awesome!**

- ▶ Ubiquitous
- ▶ Pervasively expressed
- ▶ Versatile
- ▶ Easy to handle
  - ▶ Synthetic biology
  - ▶ Nanotech
  - ▶ Therapeutics and genetic engineering (CRISPR)
  - ▶ Computationally fun (but still challenging)
- ▶ RNA at the origin of life!?

*Chicken vs egg paradox  
at the origin of life*



This is the RNA World:

- **Proteins** are good at being enzymes but bad at being replicators;
- **DNA** is good at replicating but bad at being an enzyme;
- **RNA** might just be good enough at both roles to break out of the Catch-22.

**R. Dawkins.** The Ancestor's tale

# GT MASIM

Méthodes et Algorithmes pour les  
Structure(s) et Interactions des Macromolécules

Depuis Sept. 2017

## Objectifs :

Accompagner un dialogue **interdisciplinaire**  
entre **méthodologues et praticiens**  
**en bioinformatique structurale**

- ▶ Participer à l'animation d'une communauté nationale de (jeunes) chercheurs en bioinformatique structurale;
- ▶ Accompagner le développement et la diffusion de méthodologies pérennes, basées sur des fondements algorithmiques bien maîtrisés;
- ▶ Recenser l'offre logicielle au niveau national ainsi que les bonnes pratiques tant au niveau développement qu'utilisation;
- ▶ Catalyser les échanges avec l'industrie
- ▶ ...

Pas besoin d'être affilié au GDR BiM pour participer !

## Quelques problèmes et leurs méthodes:

- ▶ Analyse cristallographiques : géométrie, apprentissage (régression, classification)
- ▶ Analyse RMN: géométrie des distances, modèles Bayésiens
- ▶ Analyse comparative : information mutuelle, corrélations structure - séquence
- ▶ Analyse de structures cryo-EM : analyse d'images, clustering
- ▶ Comparaison de structures : matching et graphes produits, recherche de structure
- ▶ Classification de structures tertiaires : classification, clustering
- ▶ Modélisation d'assemblages (cryo-EM, spectrométrie de masse) : optimisation multi-critères, inférence de contacts, et algo/théorie des graphes
- ▶ Recherche de structures stables - paysages énergétiques: algorithmes d'échantillonnage, optimisation, analyse de processus Markoviens
- ▶ Prédiction de structures secondaires (ARN, protéines, chromatine?) : modèles discrets (graphes, arbres), énumération, programmation dynamique
- ▶ Modèles pour la thermodynamique : algorithmes d'échantillonnage (dynamique moléculaire et variantes, méthodes de Monte Carlo et variantes), réduction de dimension, modèles harmoniques, analyse spectrale; régression pour la prédiction d'affinité
- ▶ Modèles pour la cinétique : modèles de Markov multi-échelle, path sampling
- ▶ Parallélisation des codes de calcul
- ▶ ...

Contacts : Frédéric Cazals (Inria Sophia) & Yann Ponty (LIX)

## Actions :

- ✓ Site Web dédié
- ✓ Journées annuelles du GT
- ✓ Mailing list communautaire
- ⌚ Accompagnement organisation école thématiques
- ⌚ Répertoire de l'offre logicielle nationale
- ✗ Newsletter (*highlights* scientifiques)
  - ▶ Et ??? ...

En coordination avec le GGMM !

Contacts : Frédéric Cazals (Inria Sophia) & Yann Ponty (LIX)

# Journées de lancement 16-17 Nov 2017



- Inria Paris
- ~50 participants
- 27 exposés
- 7 sessions
- 18+ litres de café
- 2 organisateurs épuisés
- ...



# Journées de lancement 16-17 Nov 2017



- Inria Paris
- ~50 participants
- 27 exposés
- 7 sessions
- 18+ litres de café
- 2 organisateurs épuisés
- ...



**But : Couverture nationale de la communauté**

# Journées de lancement 16-17 Nov 2017



- Inria Paris
- ~50 participants
- 27 exposés
- 7 sessions
- 18+ litres de café
- 2 organisateurs épuisés
- ...



**But : Couverture nationale de la communauté**