Progress in RNA 3D Modelling

E. Westhof

Institut de biologie moléculaire et cellulaire du CNRS

Université de Strasbourg

2012

Automatic Modelling requires

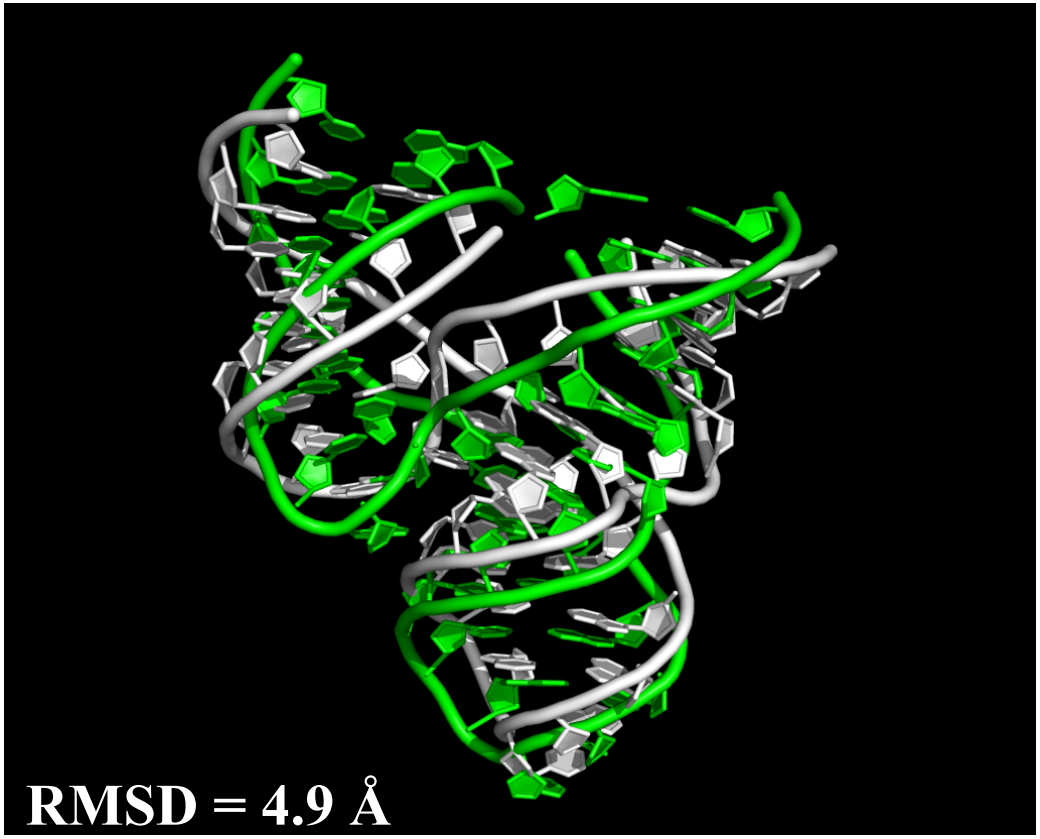
1/ metrics for comparisons

2/ blind testing

Beyond "Single Number" metrics

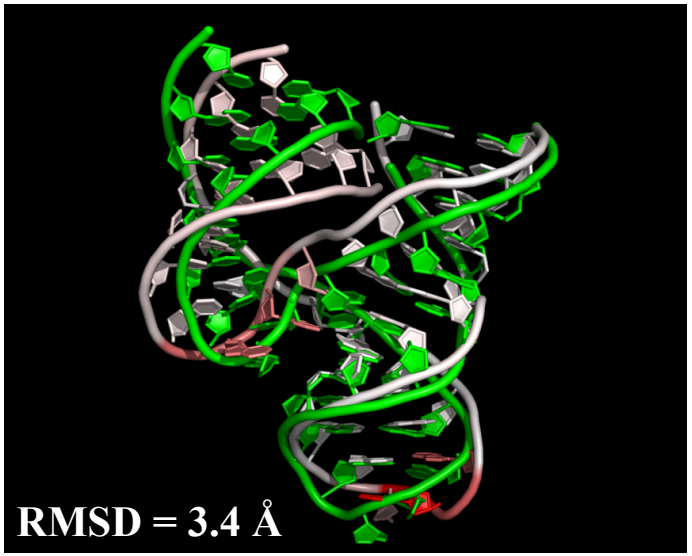
No "single number" metric is able to describe the differences between models.

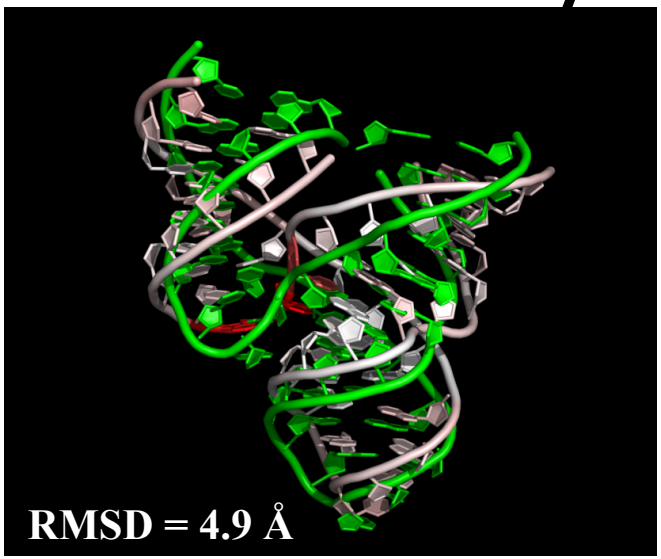
Where are the differences between the models below?



Metrics for RNA Structure Comparisons

Which is the best model and why?





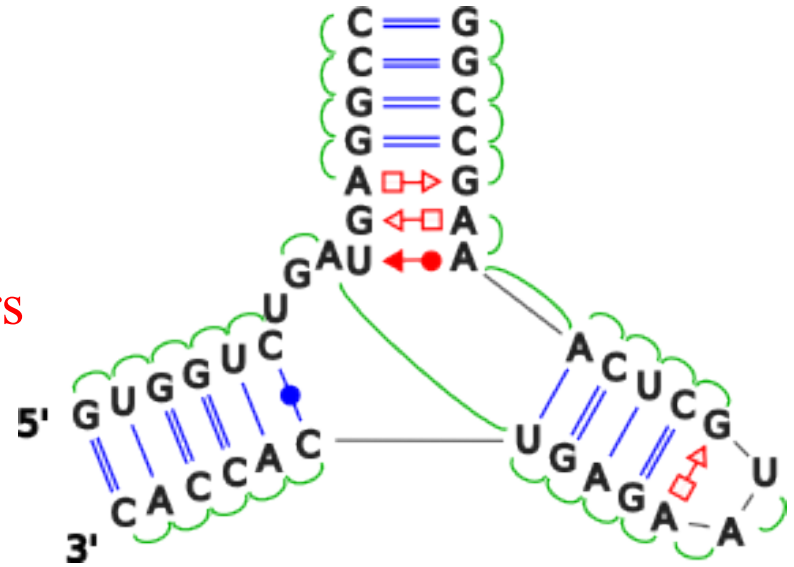
RMSD <u>is good!</u>
- Easy to compute.
- Easy to compare.

<u>...But</u>
- Spreads errors over whole structure.
- Provides little information on:
  - Local deviations;
  - Base-pairing and base-stacking patterns;
  - Intra and inter domain deviations.
- Difficulties for the localization of modeling defects.

Interaction Networks

Set of base-base interactions constitute the elementary Structural building units of RNA molecules

- Watson-Crick base pairs
- Non Watson-Crick base pairs
- Base stacking



Good Structure Prediction

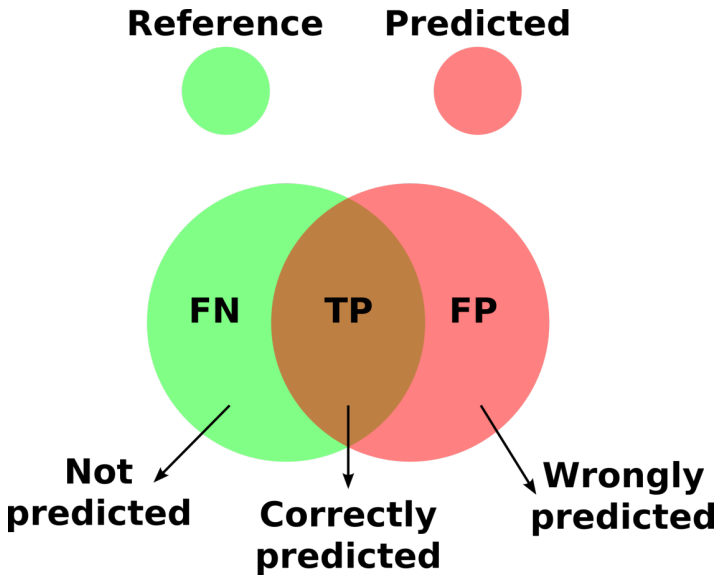
|| ^

v ||

Correct Interaction Network

INF – Interaction Network Fidelity

INF: How well the predicted model (**P**) fits the interaction network of the reference model (**R**)?



$$INF = MCC(P,R) = \sqrt{\frac{|TP|}{|TP|+|FP|} \times \frac{|TP|}{|TP|+|FN|}} \in [0,1]$$

Parisien M, Cruz JA, Westhof E and Major F, RNA 15, 2009

DI – Deformation Index

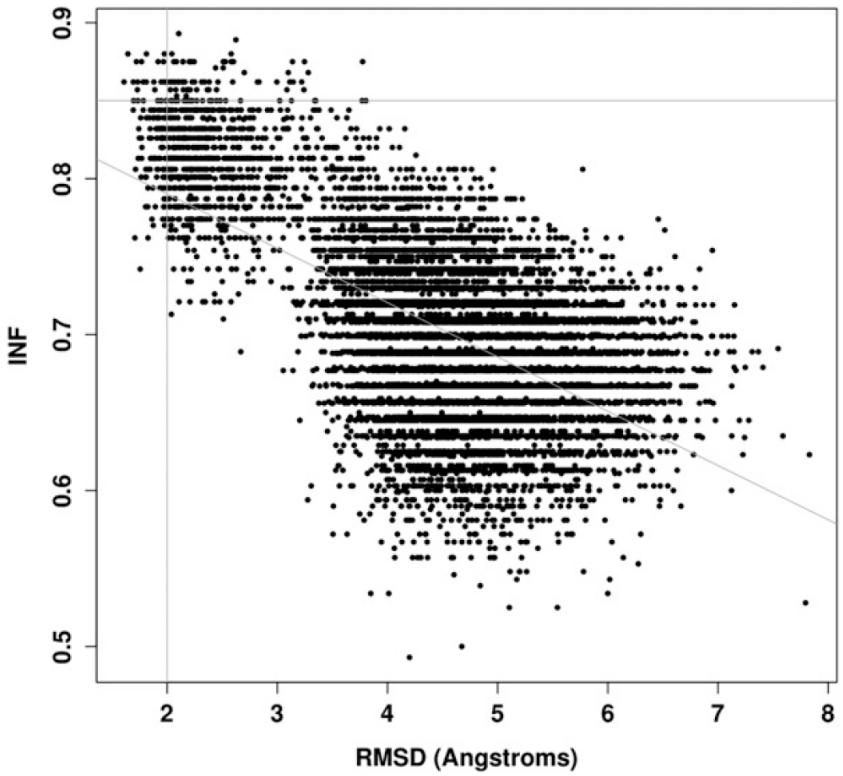
A good RNA prediction model must:

- Geometrically resemble the reference model: **low RMSD**
- Predict the correct base interactions: **high INF**

$$DI(P,R) = \frac{RMSD(P,R)}{INF(P,R)}$$

Parisien M, Cruz JA, Westhof E and Major F, RNA 15, 2009

DI – Deformation Index

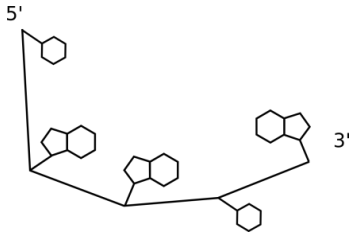


The distribution of (RMSD, INF) values shows some correlation (Pearson corr. coef.=0.6), although, for a given RMSD threshold we have a wide range of INF values (values for 9847 structures).

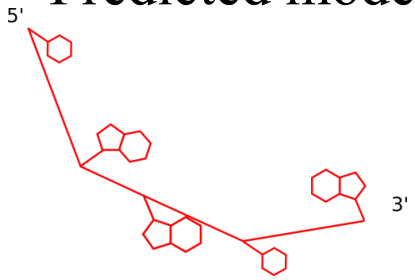
Parisien M, Cruz JA, Westhof E and Major F, RNA 15, 2009

DP – Deformation Profile

Reference model



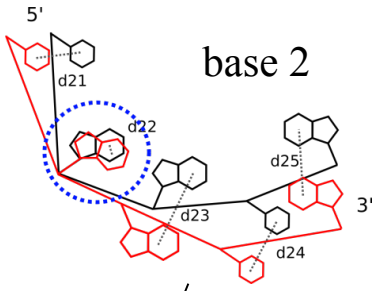
Predicted model



Both models have the same number of bases

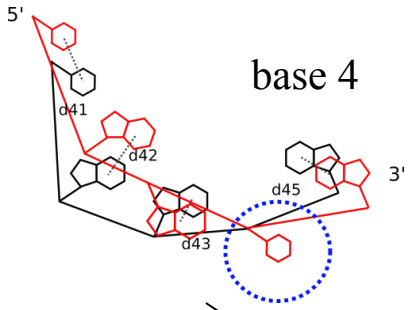
Superimpose both models base by base:

...



base 2

...

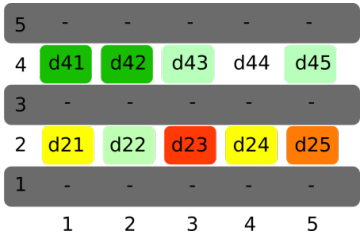


base 4

...

Update row 2 with the individual distances of each base

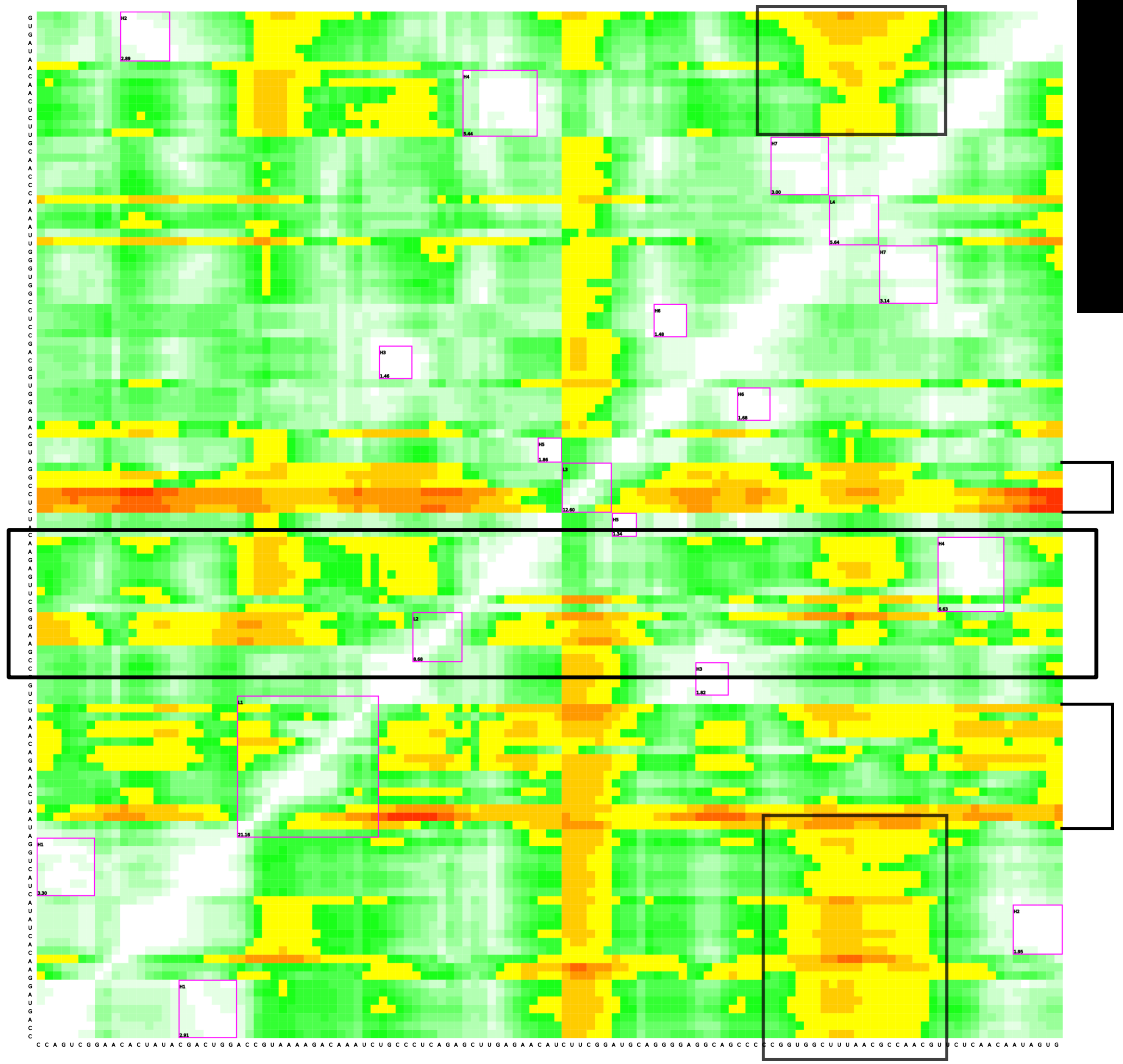
DP =



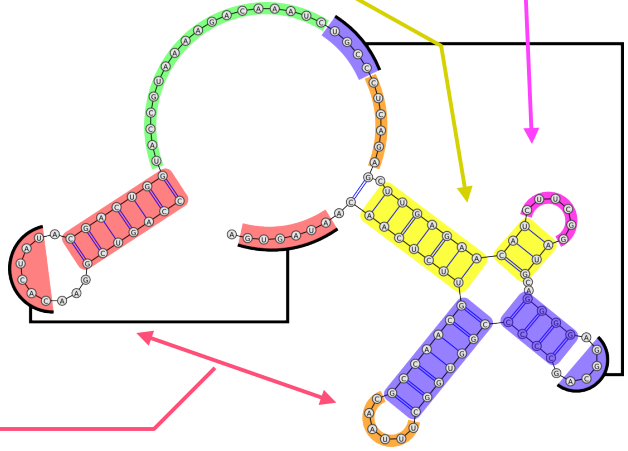
Update row 4 with the individual distances of each base

Parisien M, Cruz JA, Westhof E and Major F, RNA 15, 2009

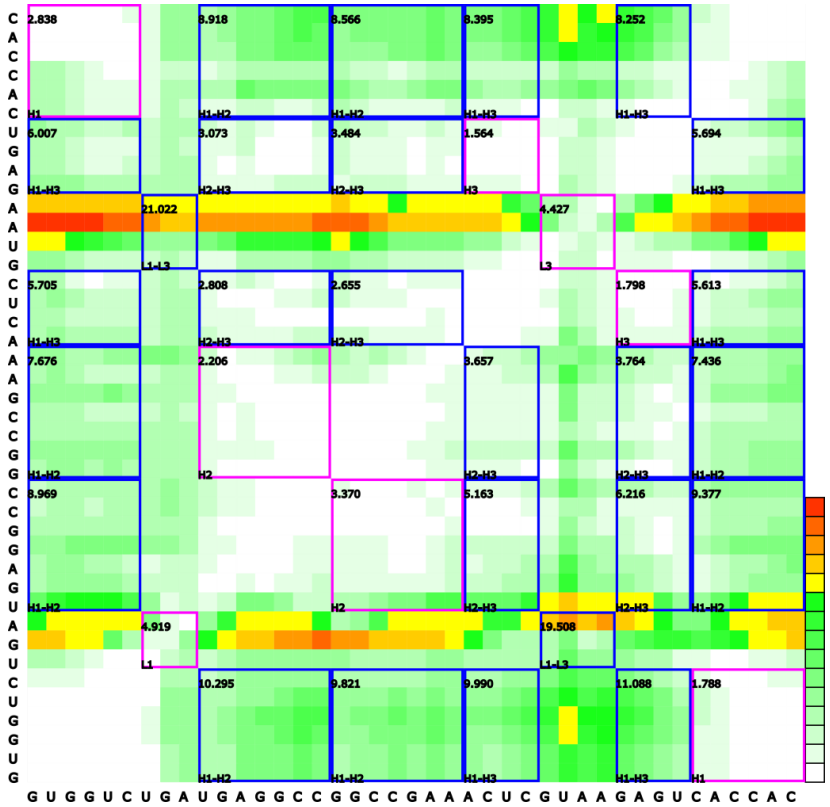
Relative domain position

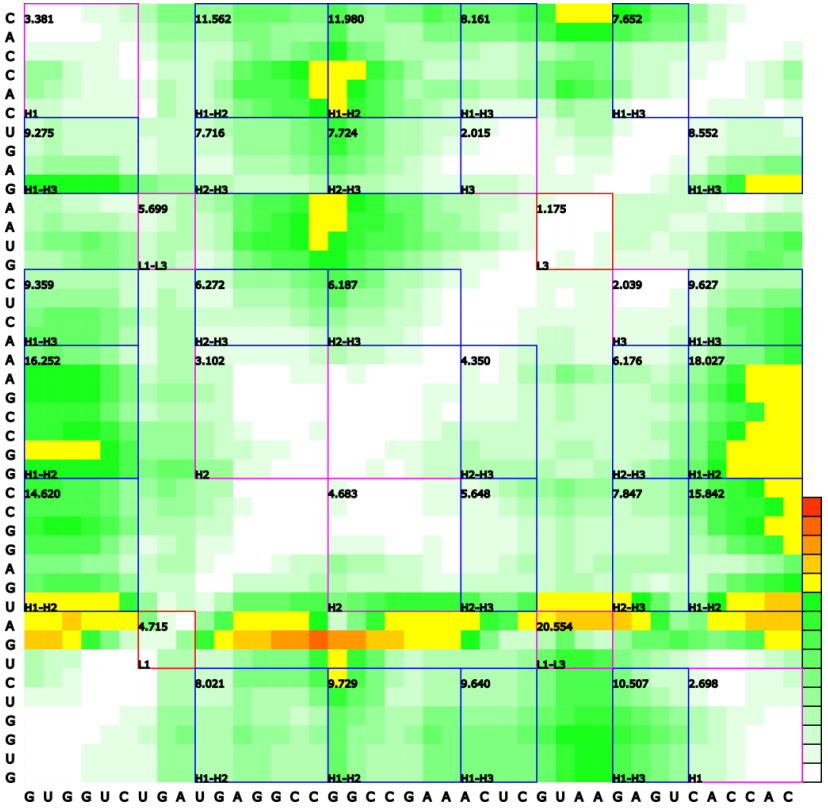


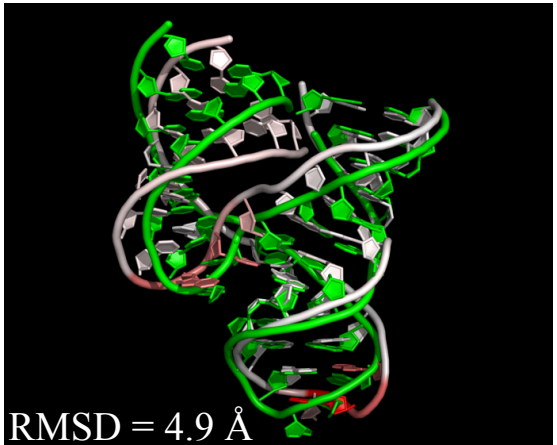
Class I Ligase
Ribozyme

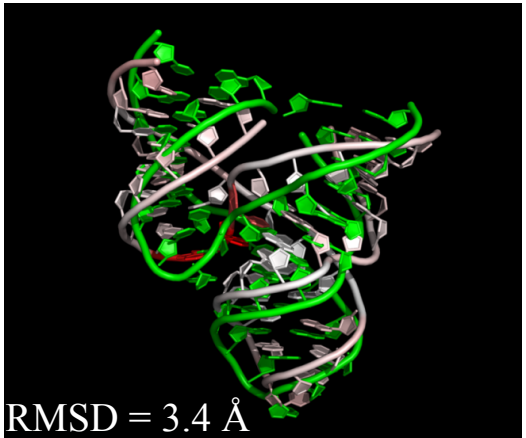


DP – Deformation Profile





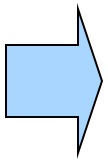




RNA-Puzzles

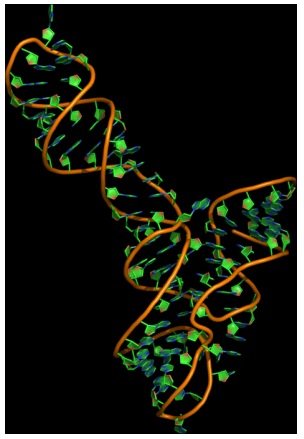
A CASP-Like collective experiment for RNA structure Prediction
(**C**ritical **A**ssessment of protein **S**tructure **P**rediction)

Sequence information is provided to interested predition groups with no other (or very limited) information.



CUCUGGAGAGAA
CCGUUUAAUCGG
UCGCCGAAGGAG
CAAGCUCUGCGC
AUAUGCAGAGUG
AAACUCUCAGGC
AAAAGGACAGAG

X-ray structure information is kept unpublished until the end of the prediction submission period.

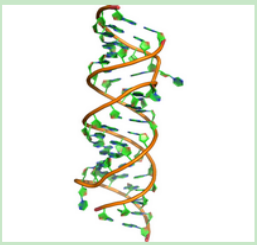


Sequence information and crystal structure kindly provided by: Thomas Hermann, Dinshaw Patel and Alexander Serganov

RNA-Puzzles

RNA Puzzles

Home | Open Challenges | Past Challenges | The Participants



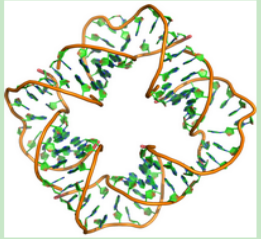
Challenge 1 (November 2011)

What is the structure of the following sequence: 5'-CCGCCGCGCCAUGCCUGUGGCGG-3' knowing that the crystal structure shows a homodimer that contains two strands of the sequence. The strands hybridize with blunt ends (C-G closing base pairs).

Crystal structure kindly provided by Thomas Hermann:

Dibrov SM, McLean J, Hermann T. 2011. Structure of an RNA dimer of a regulatory element from human thymidylate synthase mRNA., Acta Cryst. D, Biological Crystallography 67, 97-104.

results >



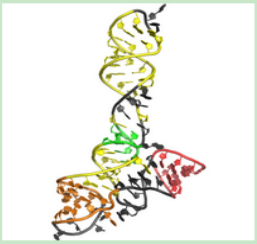
Challenge 2 (November 2011)

The crystal structure shows a 100 nt square that assembles from four inner and four outer strands. The secondary structure shown was used for the design of the square. Actual base pairing in the crystal may deviate. 3D coordinates of the nucleotides in the inner strands (B,D,F,H) were provided. What are the structures of the outer strands (A,C,E,G)?

Crystal structure kindly provided by Thomas Hermann:

Dibrov SM, McLean J, Parsons J, Hermann T. 2011. Self-assembling RNA square. PNAS 108, 6405-6408.

results >



Challenge 3 (November 2011)

A domain of a riboswitch was crystallized. The sequence is the following:
5'-CUCUGGAGAGAACCGUUUAAUCGGUCGCCGAAGGAGCAAGCU
CUGCGCAUAUGCAGAGUGAAACUCUCAGGCAAAAGGACAGAG-3'
The crystallized sequence was slightly different (an apical loop was replaced by a GAAA loop) but it was not mentioned to protect the crystallographers.

Crystal structure kindly provided by Dinshaw Patel:

Huang L, Serganov A, Patel DJ. 2010. Structural Insights into Ligand Recognition by a Sensing Domain of the Cooperative Glycine Riboswitch. Mol. Cell. 40, 774-786.

results >

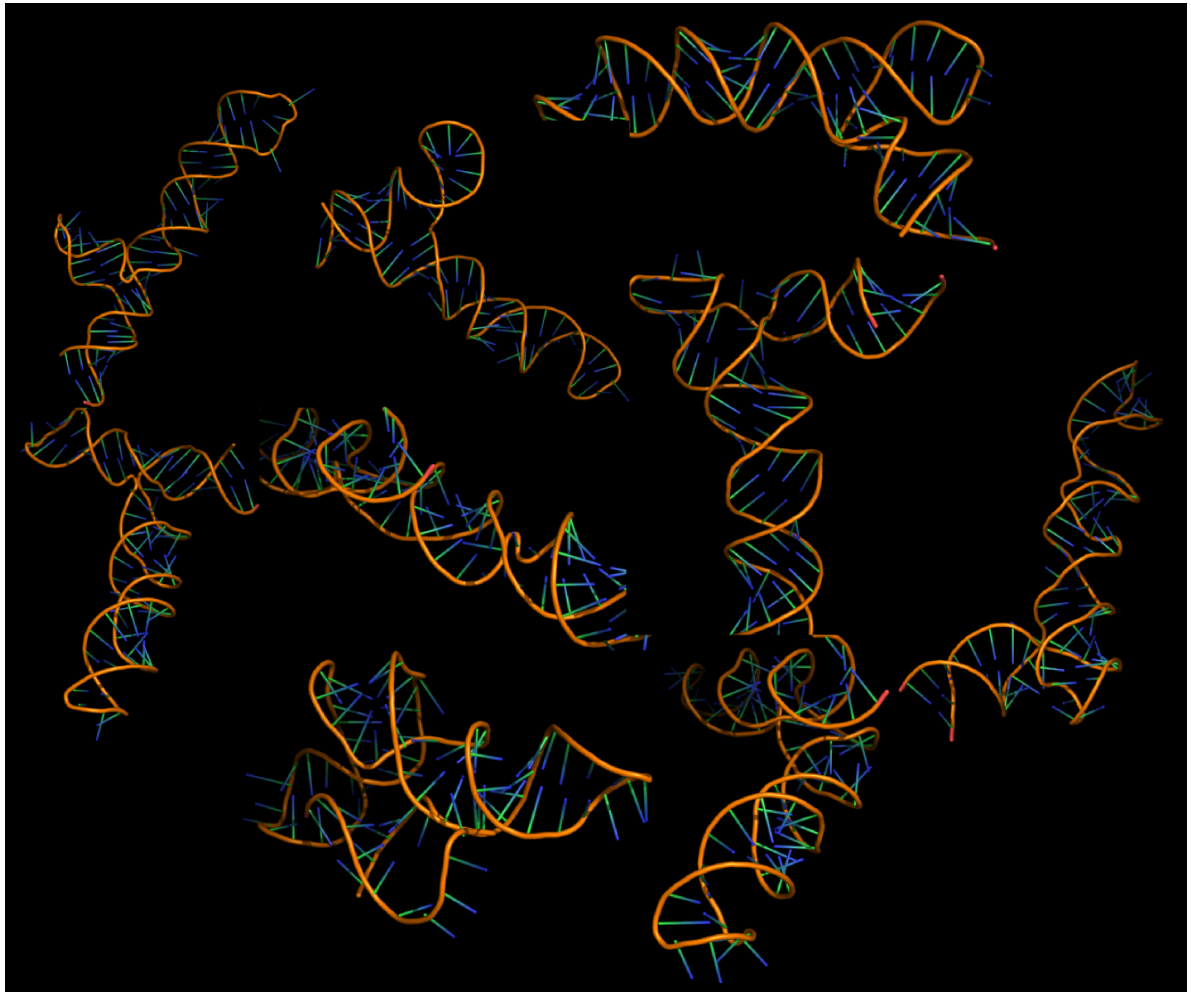
More informations: e.westhof [AT] ibmc-cnrs.unistra.fr
Last update May 20th, 2011

http://http://paradise-ibmc.u-strasbg.fr/rnapuzzles/

Cruz JA et al. (in prep.)

RNA-Puzzles

Seven groups sent their predicted models:



Models from Bujnicki, Chen, Das, Dokholyan, Major, Santalucia and Flores labs

Most methods based on template-based and fragment assembly with 2D constraints and energy refinement (often AMBER)

Bujnicki : ModeRNA

Chen : use of 3D coarse-grained scaffold on 2D

Das : stepwise assembly (Rosetta all-atom energy function)

Dokholyan : coarse-grained discrete MD (energies and filters)

Flores : RNABuilder (internal-coordiante mechanics)

Major : MC-Fold and MC-Sym pipeline.

Santalucia : RNA123 (motif library and score function)

Problem 1 : What is the structure of the following sequence: 5'-CCGCCGCGCCAUGCCUGUGGCGG-3' knowing that the crystal structure shows a homodimer that contains two strands of the sequence. The strands hybridize with blunt ends (C-G closing base pairs).

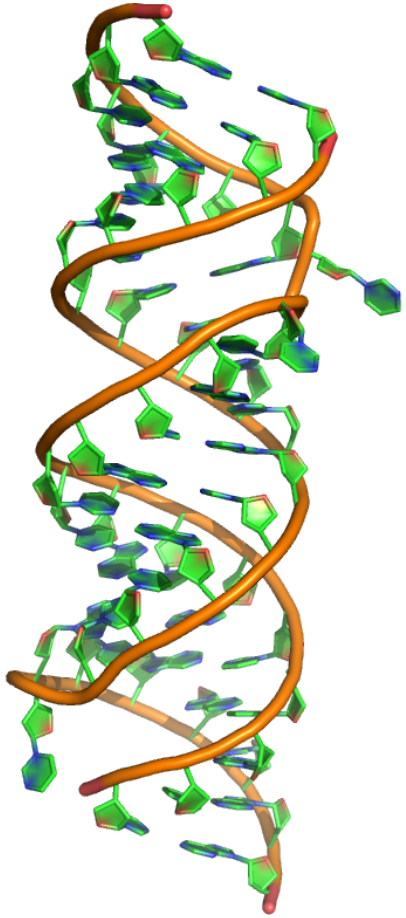
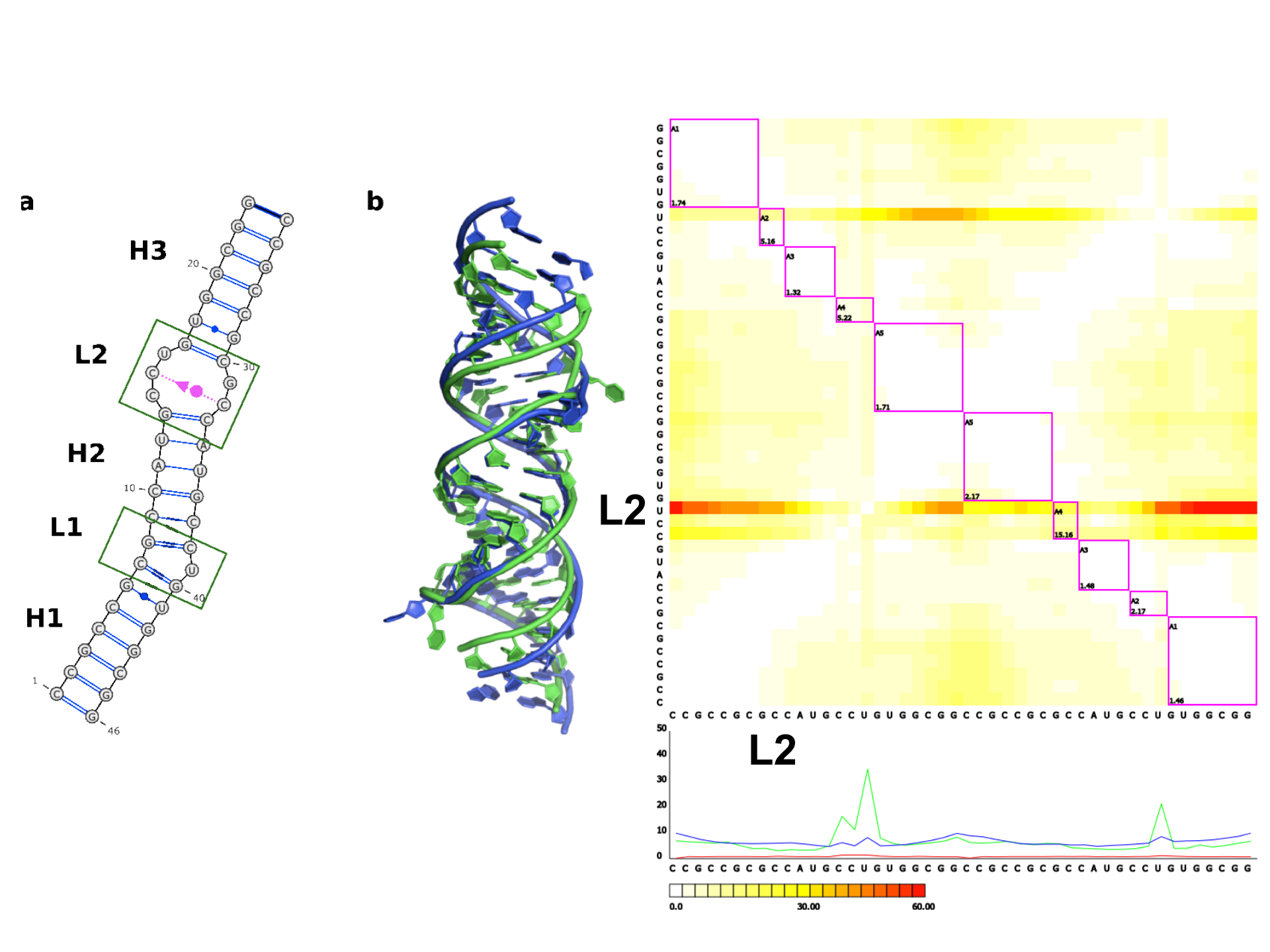


TABLE 1. Summary of the results for Puzzle 1

Problem 1 Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF stack[h]	Rank[d]	Clash Score[i]	Rank[d]	L1[j]
Das	3	3.41	1	3.66	1	0.93	1	0.95	2	0.92	1	0.00	5	x
Das	1	3.58	2	3.89	2	0.92	3	0.95	1	0.91	2	0.00	3	x
Das	4	3.91	3	4.31	3	0.91	4	0.91	8	0.91	4	0.00	4	
Major	1	4.06	4	4.57	4	0.89	5	0.95	6	0.87	5	66.40	11	
Chen	1	4.11	5	5.01	6	0.82	9	0.87	11	0.80	8	0.68	6	
Das	2	4.34	6	4.70	5	0.92	2	0.95	4	0.91	3	1.36	7	x
Das	5	4.56	7	5.36	7	0.85	7	0.88	10	0.84	7	0.00	2	
Bujnicki	3	4.66	8	5.75	9	0.81	11	0.95	3	0.74	14	54.73	10	x
Bujnicki	4	4.74	9	6.59	11	0.72	14	0.65	14	0.75	13	83.33	14	
Bujnicki	5	4.89	10	6.26	10	0.78	13	0.78	13	0.80	9	81.98	13	
Bujnicki	1	5.07	11	5.75	8	0.88	6	0.93	7	0.86	6	0.00	1	x
Bujnicki	2	5.43	12	6.75	12	0.80	12	0.90	9	0.77	12	71.57	12	x
Santalucia	1	5.69	13	6.75	13	0.84	8	0.95	5	0.79	11	39.86	9	
Dokholyan	1	6.94	14	8.55	14	0.81	10	0.86	12	0.79	10	31.74	8	
Mean		4.67		5.56		0.85		0.89		0.83				
Standard deviation		0.93		1.34		0.06	N	0.09		0.07				
										X-Ray Model		1.35		

Values in each row correspond to a predicted model.

Cruz JA et al. RNA 2012


a
H3
L2
H2
L1
H1
1
10
20
30
40
46
b
L2
A1
A2
A3
A4
A5
1.74
5.16
1.32
5.22
1.71
2.17
15.16
1.48
2.17
1.46
C C G C C G C G C C A U G C C U G U G G C G G C C G C C G C G C C A U G C C U G U G G C G G
L2
50
40
30
20
10
0
0.0
30.00
60.00

Problem 2 : The crystal structure shows a 100 nt square that assembles from four inner and four outer strands. The secondary structure shown was used for the design of the square. Actual base pairing in the crystal may deviate. 3D coordinates of the nucleotides in the inner strands (B,D,F,H) were provided. What are the structures of the outer strands (A,C,E,G)?

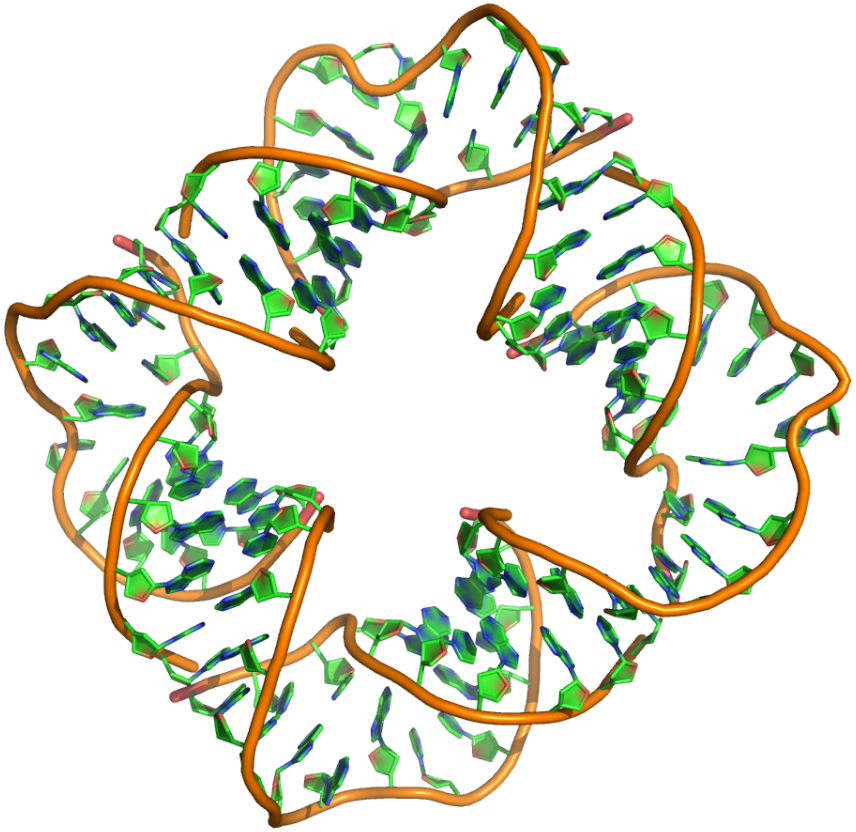


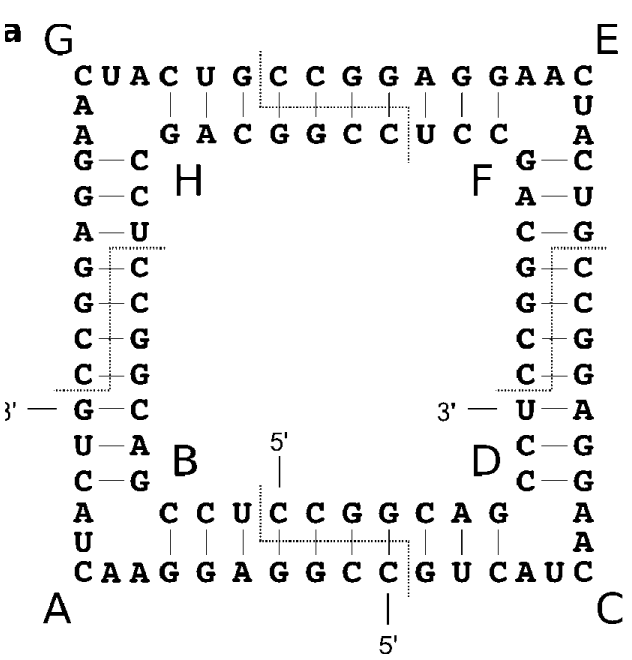
TABLE 2. Summary of the results for Puzzle 2

Problem 2

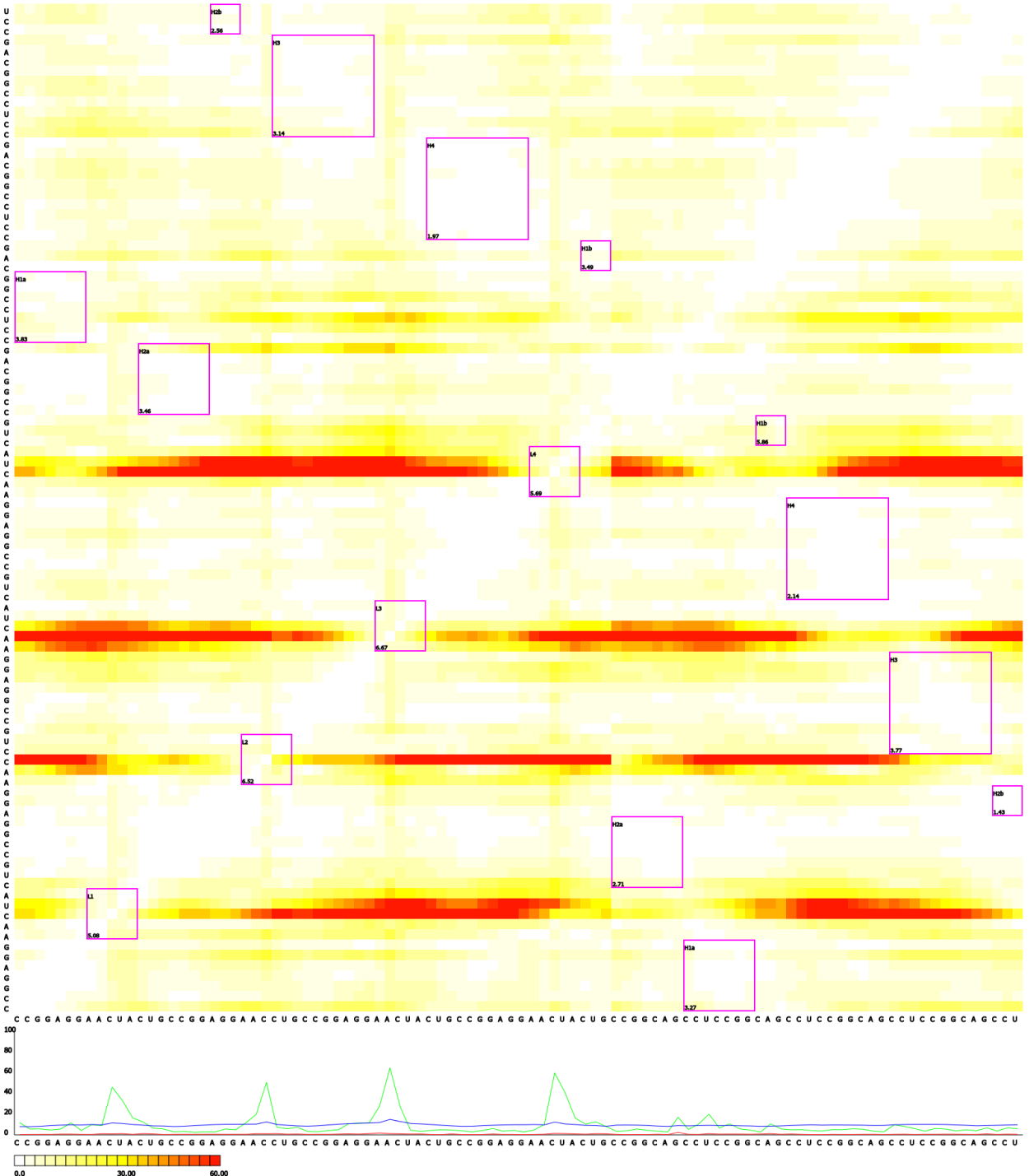
Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]
Bujnicki	2	2.3	1	2.83	1	0.81	8	0.92	9	0	13	0.79	7	14.54	2
Bujnicki	3	2.33	2	2.9	3	0.8	10	0.91	10	0	2	0.77	9	0.62	1
Das	1	2.5	3	2.9	2	0.86	2	0.96	5	0	8	0.85	2	19.8	5
Dokholyan	1	2.54	4	3.09	5	0.82	6	0.9	11	0	1	0.8	5	19.85	6
Bujnicki	1	2.65	5	2.99	4	0.89	1	0.96	4	0	3	0.86	1	15.47	3
Chen	1	2.83	6	3.74	9	0.76	13	0.9	12	0	9	0.69	13	18.66	4
Das	4	2.83	7	3.46	6	0.82	7	0.97	3	0	12	0.78	8	23.82	8
Major	1	2.98	8	3.82	10	0.78	12	0.95	7	0	10	0.71	12	134.26	12
Das	3	3.03	9	3.67	7	0.83	5	0.97	1	0	6	0.8	6	25.37	10
Das	2	3.05	10	3.69	8	0.83	4	0.97	2	0	7	0.81	3	23.51	7
Das	5	3.46	11	4.18	11	0.83	3	0.96	6	0	11	0.81	4	24.75	9
Flores	1	3.48	12	4.4	12	0.79	11	0.89	13	0	5	0.77	10	165.57	13
Santalucia	1	3.65	13	4.54	13	0.81	9	0.92	8	0	4	0.75	11	25.73	11
Mean		2.90		3.55		0.82		0.94		0.00		0.78			
Standard deviation		0.44		0.59		0.03		0.03		0.00		0.05			
												X-Ray Model		36.10	

Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model.

Cruz JA et al. RNA 2012



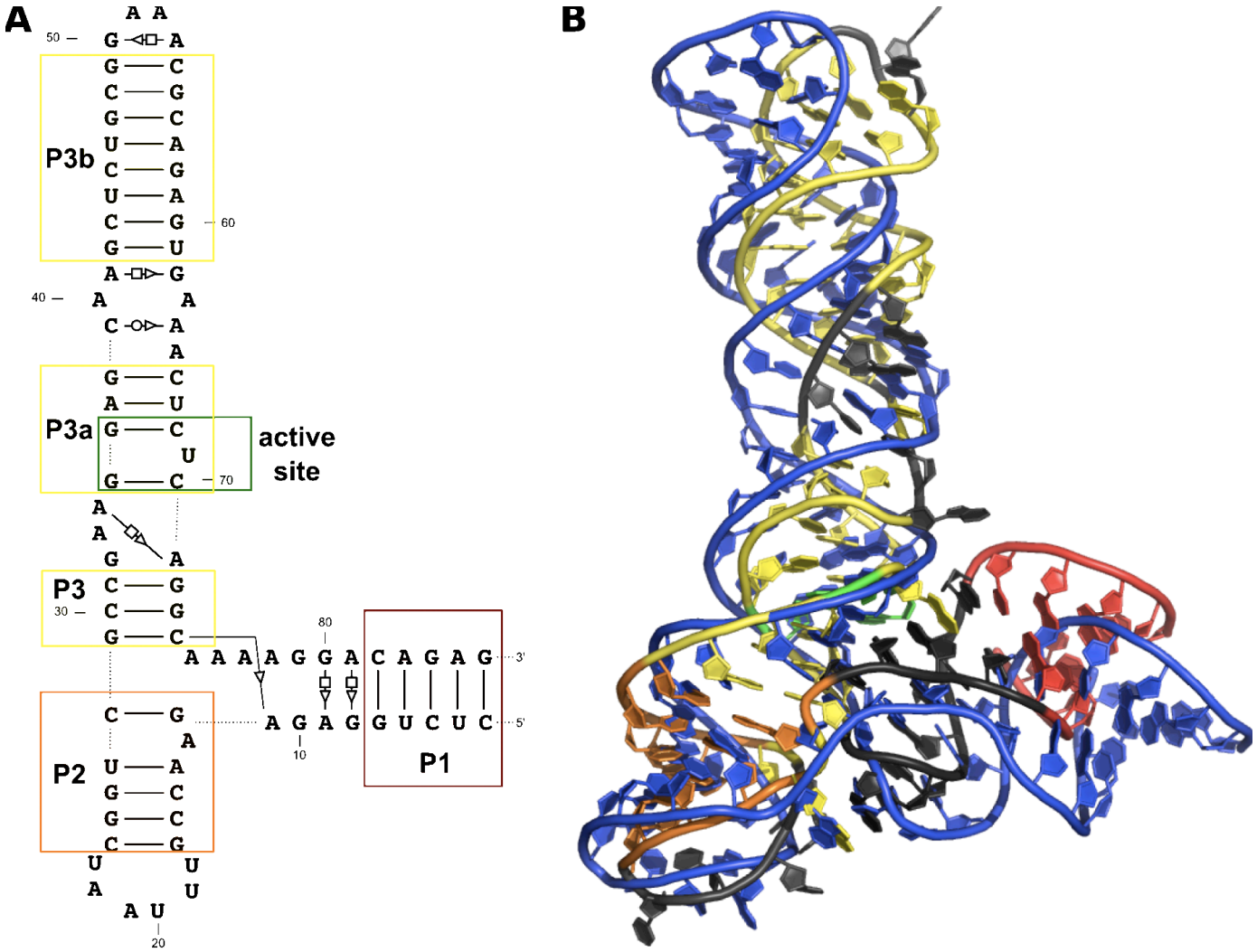
AACUA at the corners



Problem 3 : A domain of a riboswitch was crystallized. The sequence is the following:

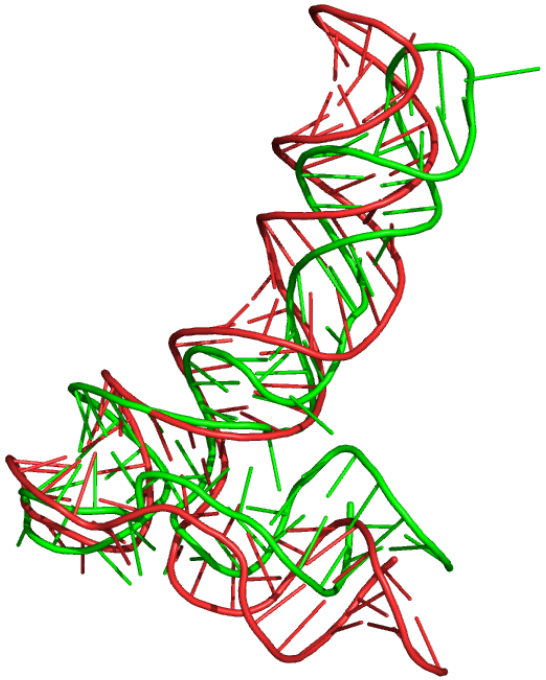
5'CUCUGGAGAGAACCGUUUAAUCGGUCGCCGAAGGA
GCAAGCUCUGCGCAUAUGCAGAGUGAAACUCUCAGG
CAAAGGACAGAG-3'

The crystallized sequence was slightly different (an apical loop was replaced by a GAAA loop) but it was not mentioned to protect the crystallographers.


A
B
P3b
P3a
active
site
P3
P2
P1
50
60
40
70
30
80
10
20
3'
5'

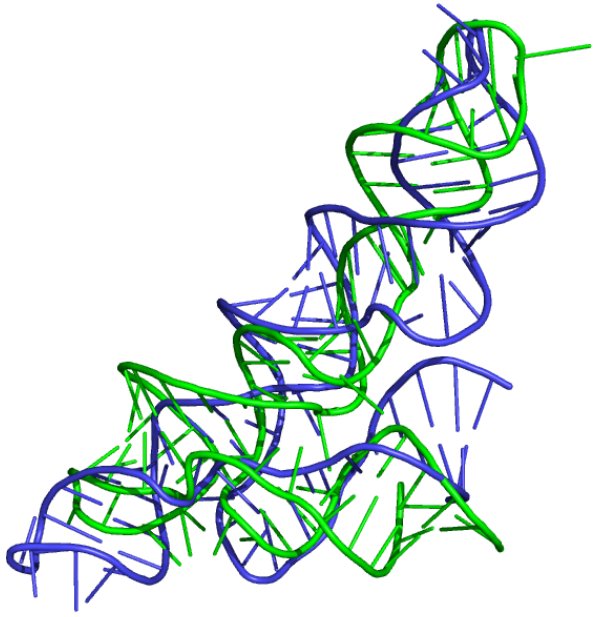
RNA Puzzles

Chen's Lab



RMSD: 7.241 Å
INF: 0.736

Dokholyan's Lab



RMSD: 11.46 Å
INF: 0.712

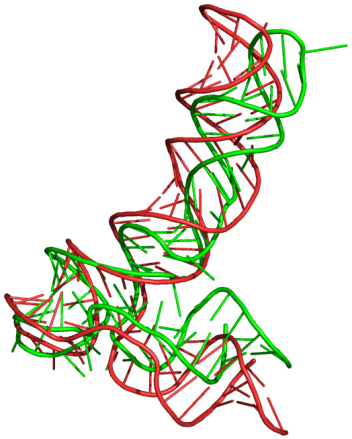
TABLE 3. Summary of the results for Puzzle 3

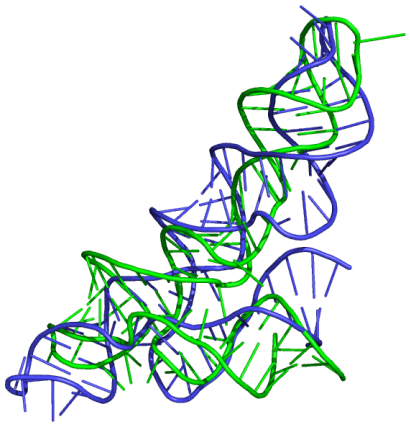
Problem 3 Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]	*P*-value[k]	Rank[d]
Chen	1	7.24	1	9.84	1	0.74	2	0.86	5	0	6	0.73	1	1.1	3	2.01E-05	1
Dokholyan	2	11.46	2	16.1	2	0.71	6	0.82	9	0	9	0.71	6	41.21	10	3.90E-02	2
Das	5	11.97	3	16.42	3	0.73	5	0.9	1	0.36	5	0.71	3	1.1	4	6.92E-02	3
Bujnicki	1	12.19	4	17.49	5	0.7	7	0.82	10	0	10	0.7	7	14.72	8	8.71E-02	4
Das	2	12.2	5	16.6	4	0.74	3	0.86	6	0.4	2	0.73	2	0.74	2	8.83E-02	5
Major	2	13.7	6	23.33	10	0.59	11	0.67	11	0	8	0.61	10	93.52	12	3.03E-01	6
Bujnicki	2	14.06	7	22.51	7	0.62	10	0.83	8	0	7	0.59	11	5.15	7	3.75E-01	7
Das	1	15.48	8	20.9	6	0.74	1	0.87	4	0.57	1	0.71	5	0	1	6.81E-01	8
Dokholyan	1	15.92	9	23.28	9	0.68	9	0.9	2	0	12	0.66	9	39.37	9	7.629E-01	9
Das	3	16.95	10	23.17	8	0.73	4	0.89	3	0.4	3	0.71	4	1.47	5	9.02E-01	10
Das	4	18.3	11	26.55	11	0.69	8	0.85	7	0.38	4	0.67	8	2.21	6	9.79E-01	11
Major	1	22.99	12	45.27	12	0.51	12	0.39	12	0	11	0.59	12	75.11	11	1.00E+00	12
Mean		14.37		21.79		0.68		0.80		0.18		0.68					
Standard deviation		3.99		8.69		0.07		0.14		0.22		0.05					
												X-Ray Model		1.83			

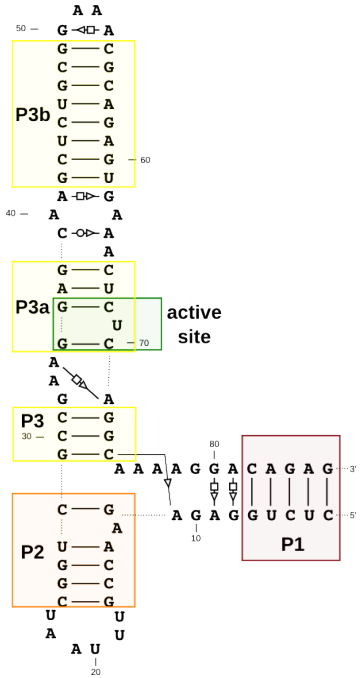
Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model

Cruz JA et al. RNA 2012

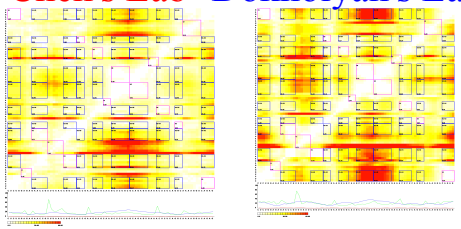
RNA Puzzles

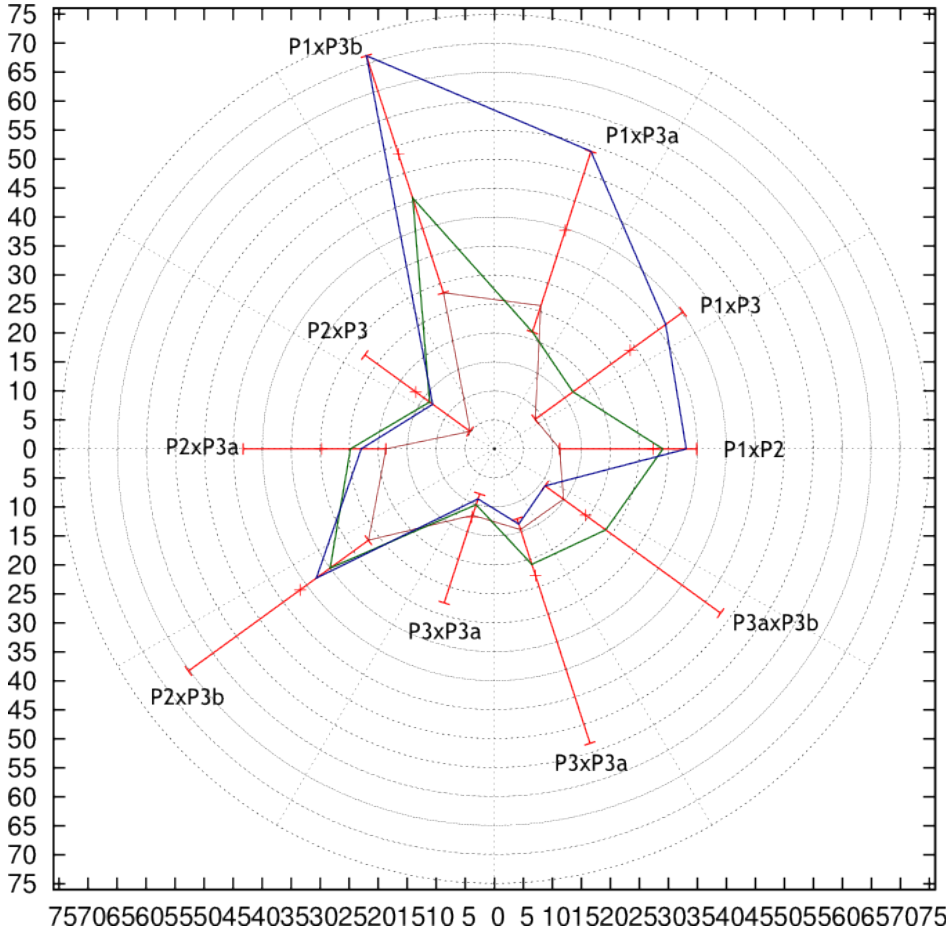






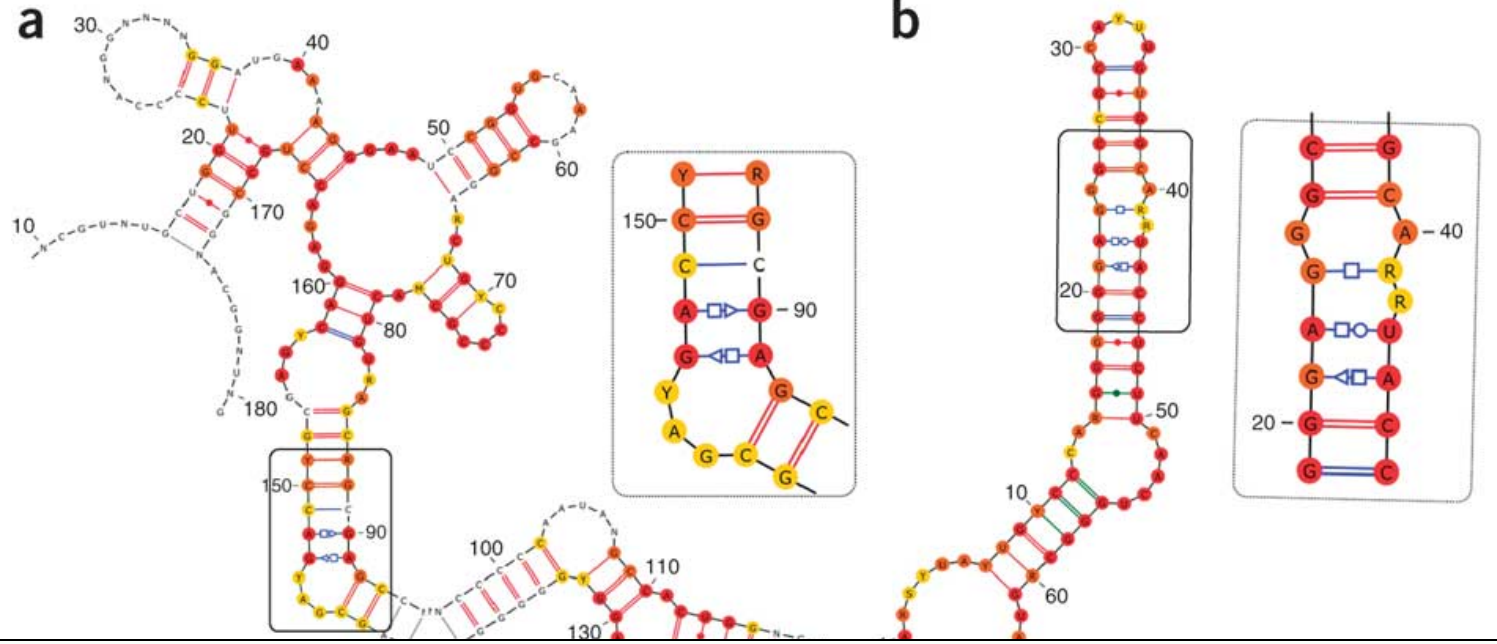
Chen's Lab Dokholyan's Lab



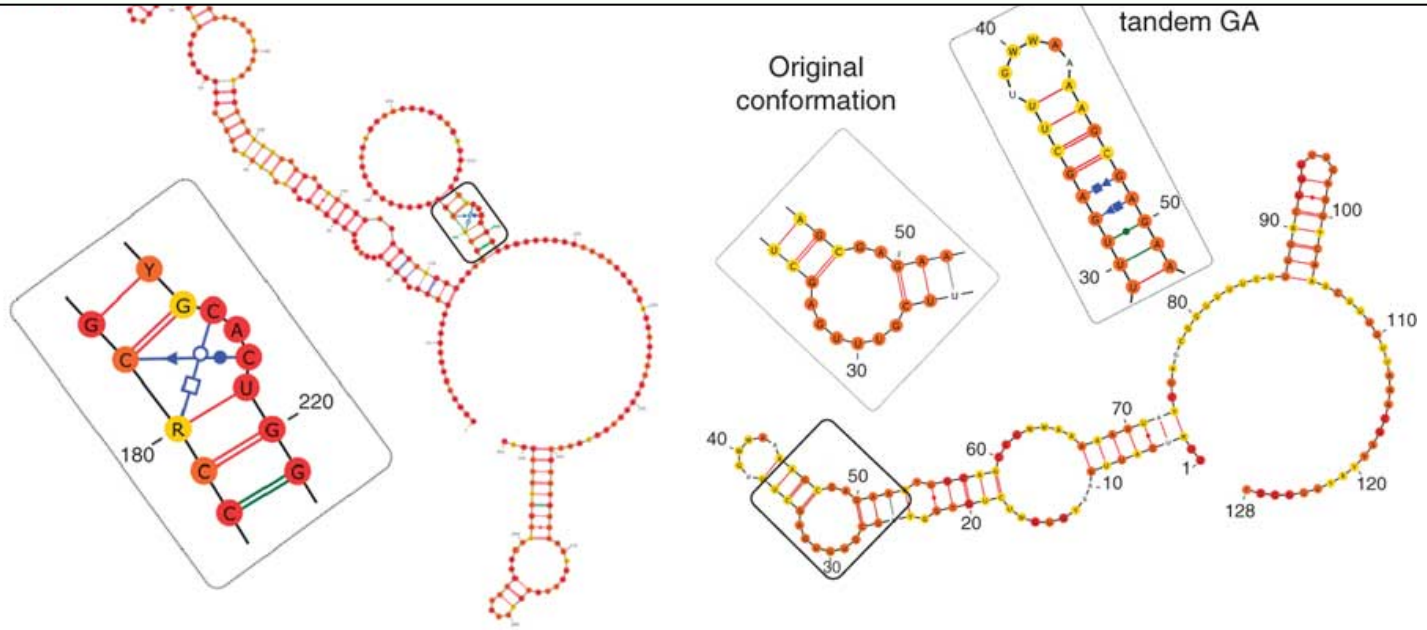


Cruz JA et al. RNA 2012

Puzzle	RMSD (Å)	INF
1: Das	3.41	0.93
2: Bujnicki	2.3	0.81
	2.65	0.89
3: Chen	7.24	0.74



The detection of RNA 3D Modules would improve considerably the modelling accuracy



UPR 9002 du CNRS,
Architecture et Réactivité de l'ARN,
Institut de Biologie Moléculaire et Cellulaire,
Université de Strasbourg

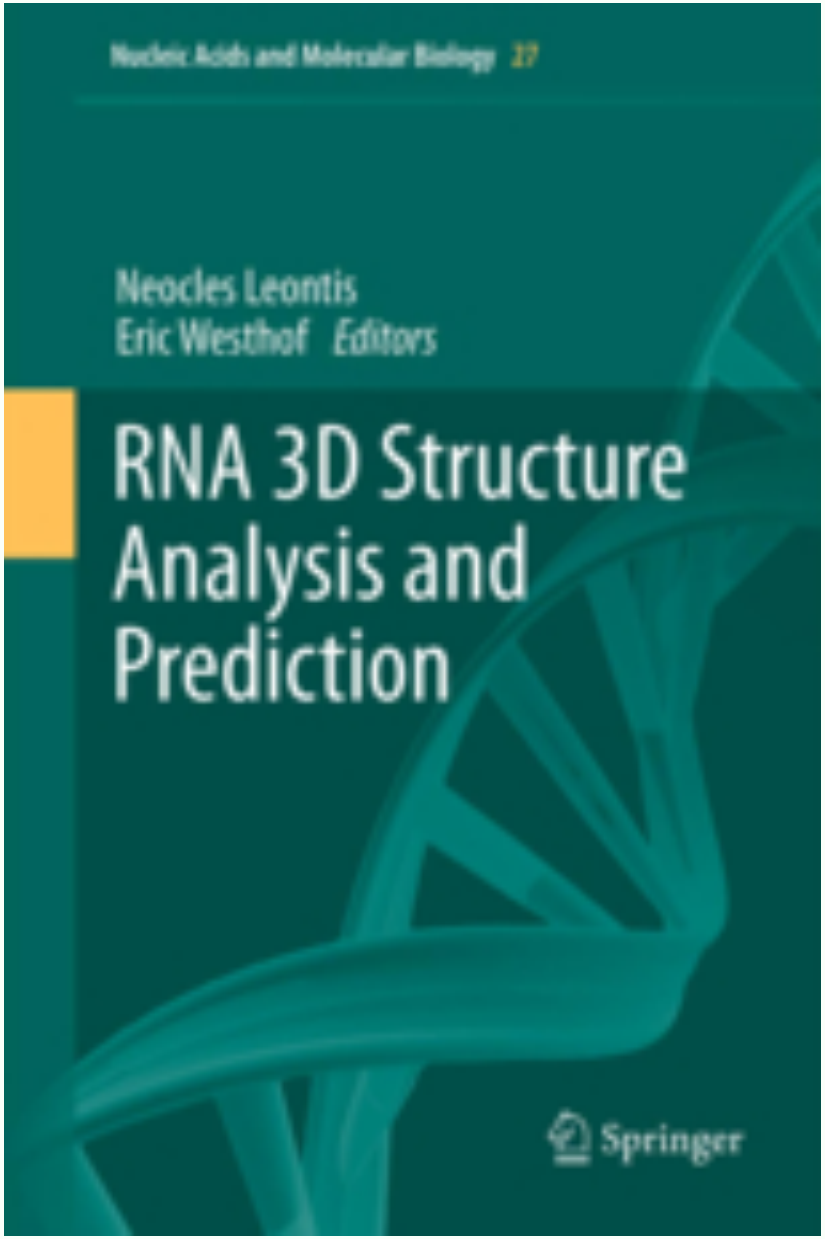
José Cruz
Fabrice Jossinet
Neocles Leontis (OBGU)
Jesse Stombauch (OBGU)
ALL the RNA-Puzzles team
and esp the structuralists

Assemble

http://bioinformatics.org/assemble

S2S

http://bioinformatics.org/s2s


Nucleic Acids and Molecular Biology 27
Neocles Leontis
Eric Westhof Editors
RNA 3D Structure Analysis and Prediction
Springer