UNIVERSITY of VIRGINIA
SCHOOL OF MEDICINE

Midwest Center for Structural Genomics
New York Structural Genomics Research Consortium
Center for Structural Genomics of Infectious Diseases
CSGID

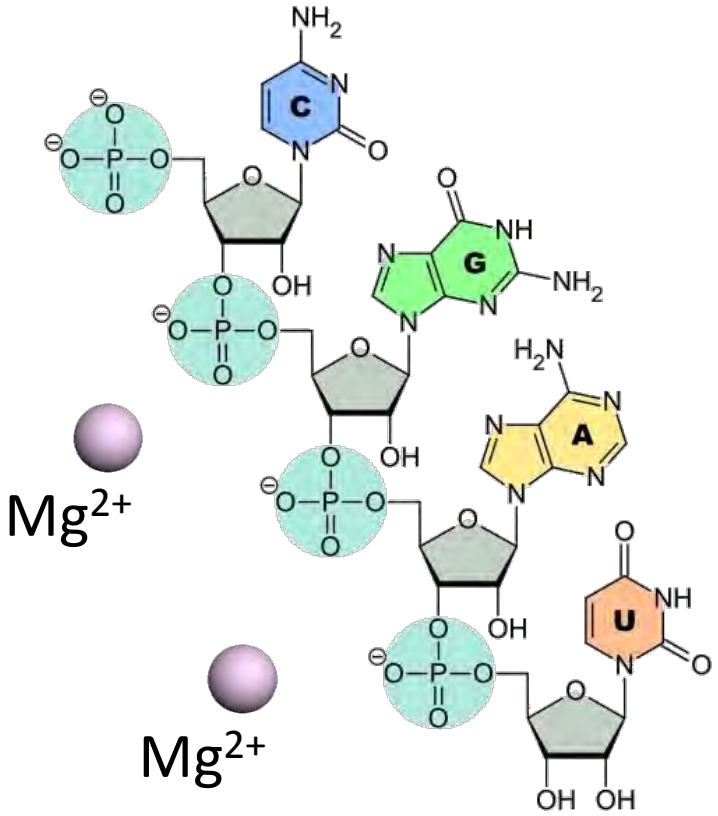
Department of Molecular Physiology and Biological Physics

Quality assessment and systematic classification of magnesium binding sites in RNA crystal structures

Ivan Shabalin[#], Heping Zheng[#], Katarzyna Mikolajczak and Wladek Minor
([#]equal contribution)

2013 ACA Meeting, Hawaii, July 20 -24, 2013

Role of magnesium ions in RNA

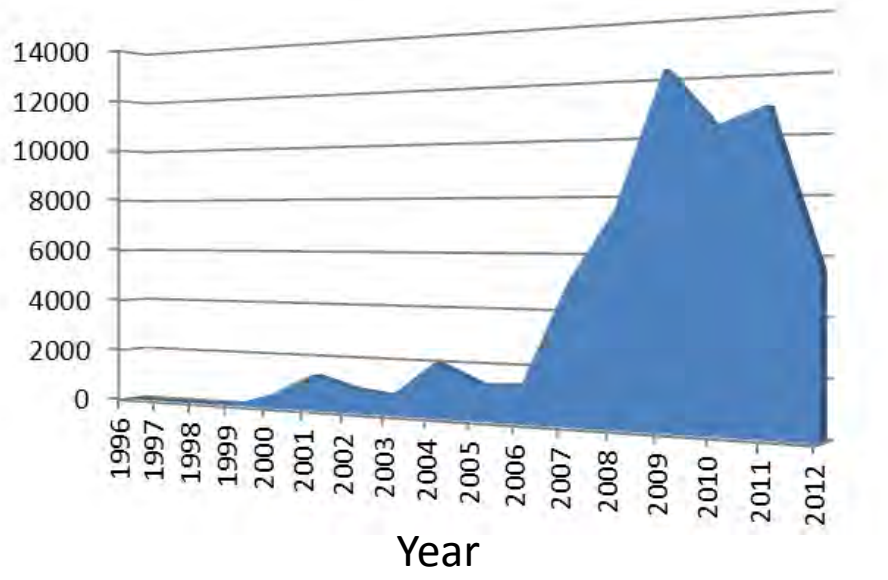


- Key factor governing RNA folding and maintaining 3D structure crucial for many diverse functions of RNA molecules
- Mediate recognition of binding partners in some RNA molecules
- Directly mediate catalytic processes in some ribozymes
- Most abundant metal in RNA crystal structures

- Charge compensation
- Magnesium clamps

Magnesium ions in RNA: goals

Number of magnesium ions in RNA structures in PDB



60145 magnesium ions from 778 RNA structures (December 2012)

- Create a comprehensive survey of magnesium binding sites in RNA
- Develop tools for annotation and comparison of magnesium-binding sites and motifs

Should facilitate understanding and prediction of RNA structure and function

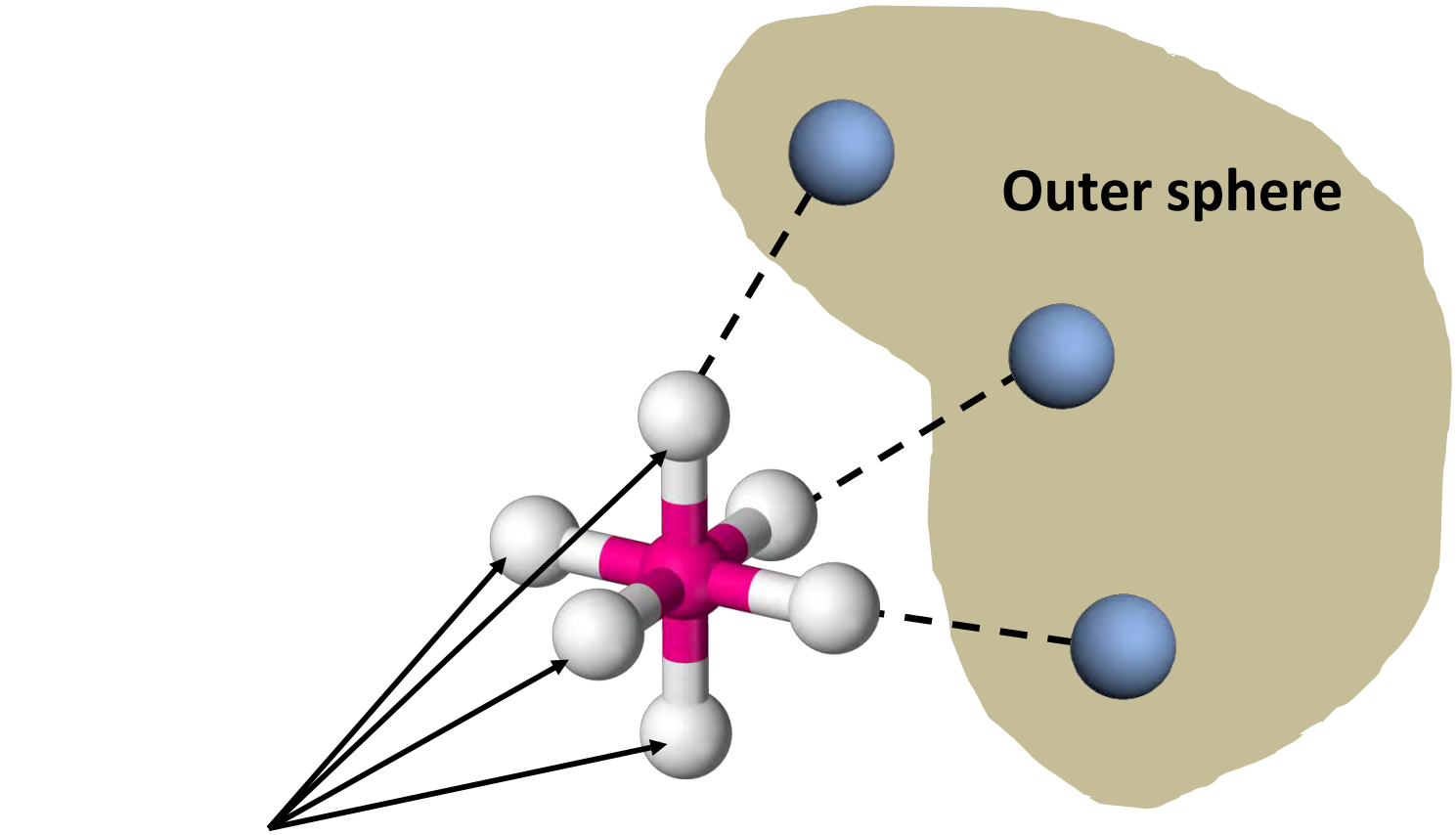
Talk outline

- Quality assessment of magnesium sites in RNA structures and benchmark dataset
- Systematic classification
- Detection of previously reported magnesium-binding motifs and the discovery of new motifs

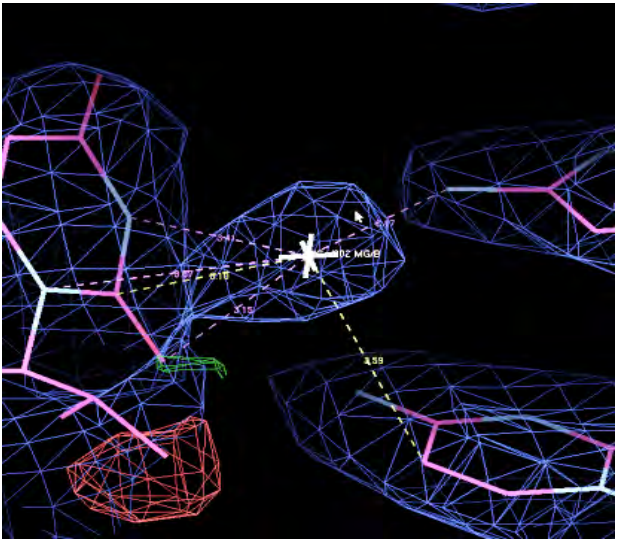
Part I

Quality assessment of magnesium sites in RNA structures

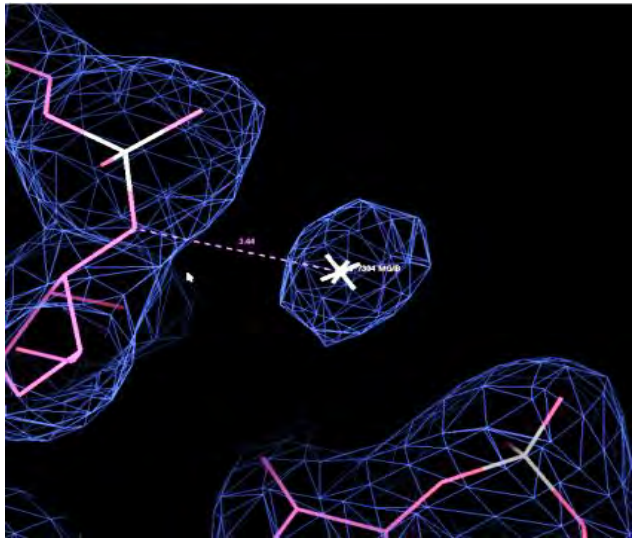
Definitions



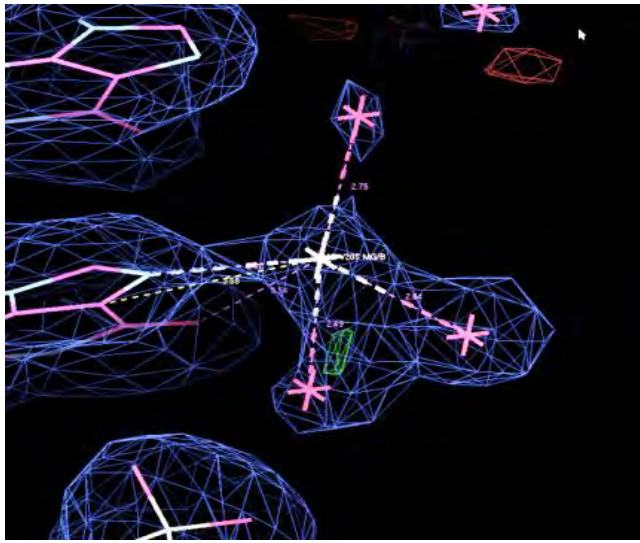
Many magnesium ions in RNA structures are poorly modeled or misidentified



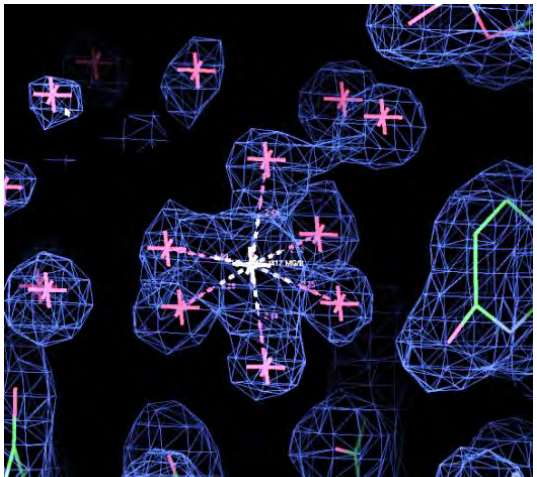
2h0s, 2.35 Å



2h0s, 2.35 Å



2h0s, 2.35 Å

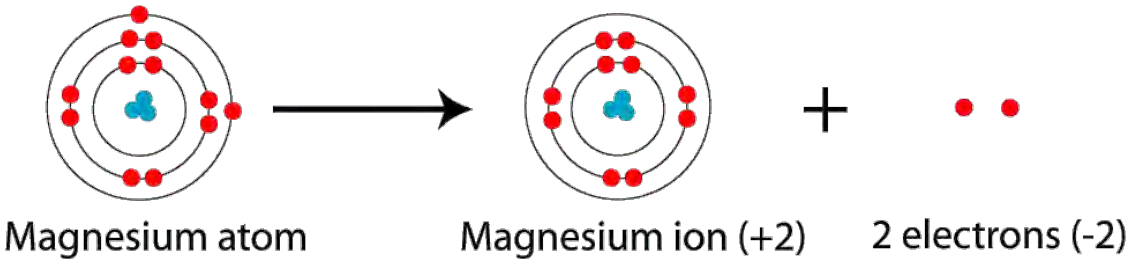


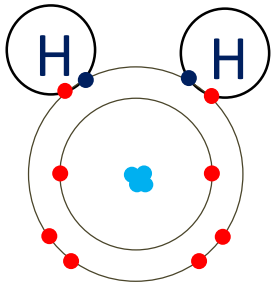
1nuv, 1.8 Å

Difficulties in Mg^{2+} identification

- Cannot be detected by anomalous diffraction

- Has the same number of electrons (10) as water molecules or sodium ions

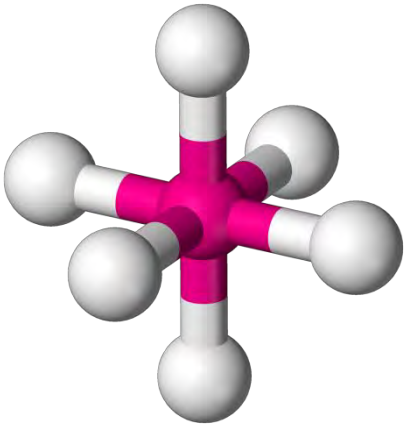


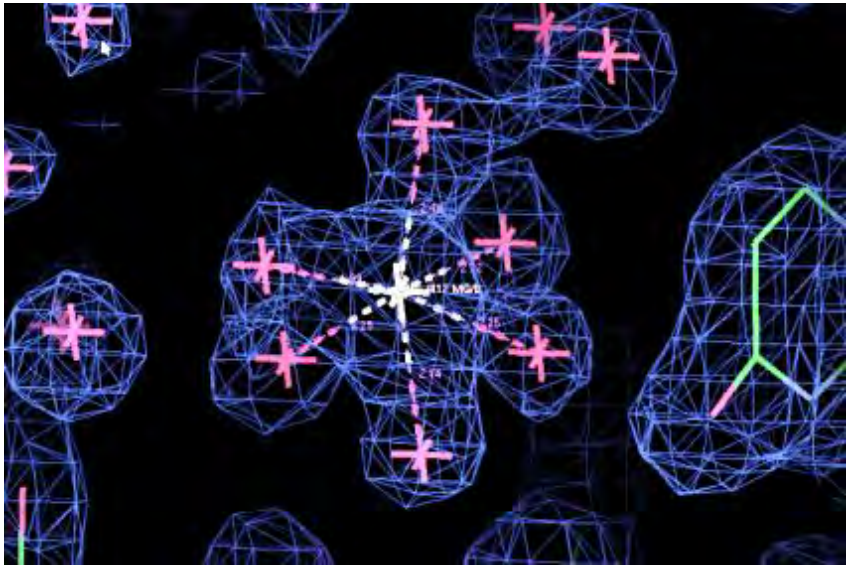


Water molecule

Possible way of Mg^{2+} identification

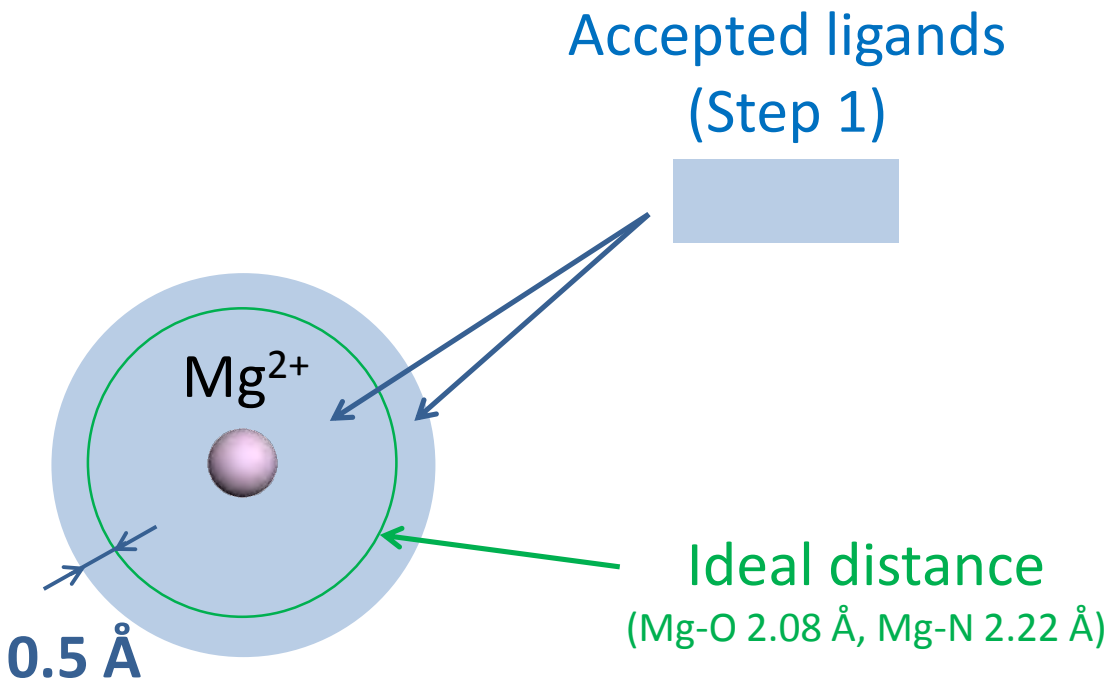
- Rigid octahedral arrangement of inner sphere ligands
- CN (coordination number) = 6
- Relatively short ideal Mg^{2+}-ligand distances (2.08 Å for Mg^{2+}-O interaction)





1nuv, 1.8 Å

Search for inner-sphere atoms
Step 1

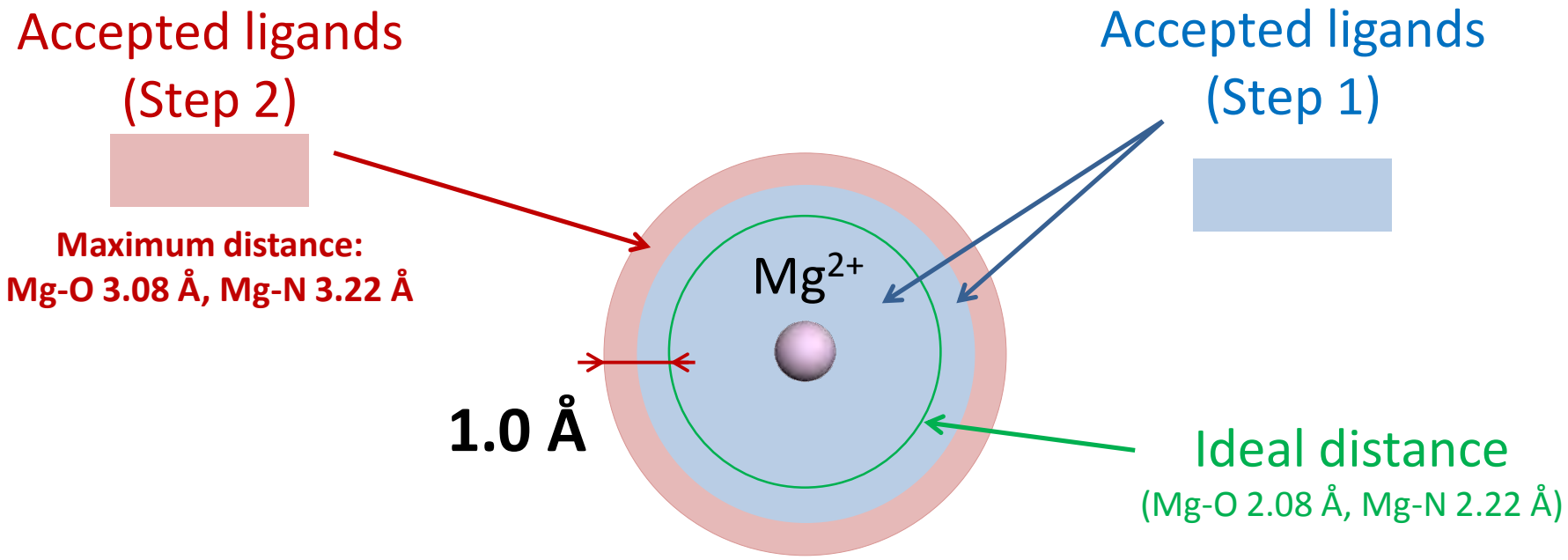


All oxygen and nitrogen atoms with a distance *d* to a Mg^{2+} ion

$$d \leq d_{ideal} + 0.5 \text{ Å}$$

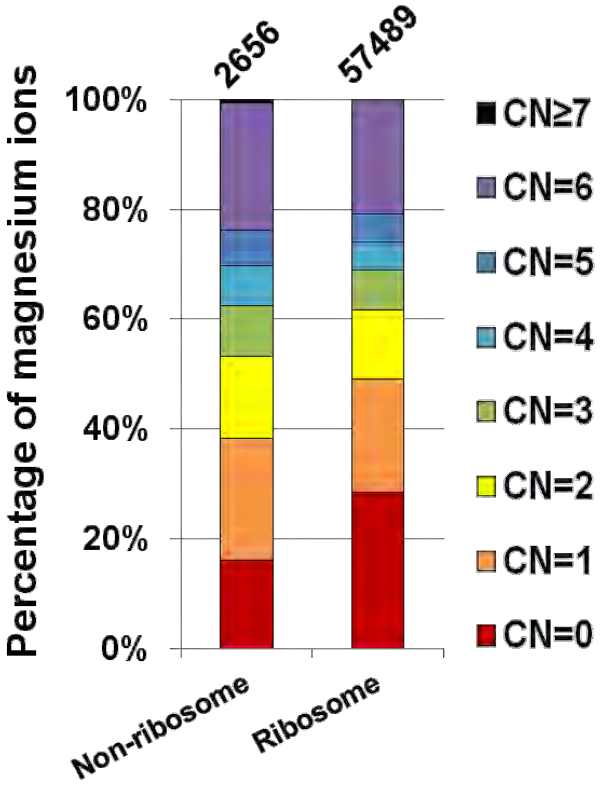
were identified as ligands

Search for inner-sphere atoms
Step 2

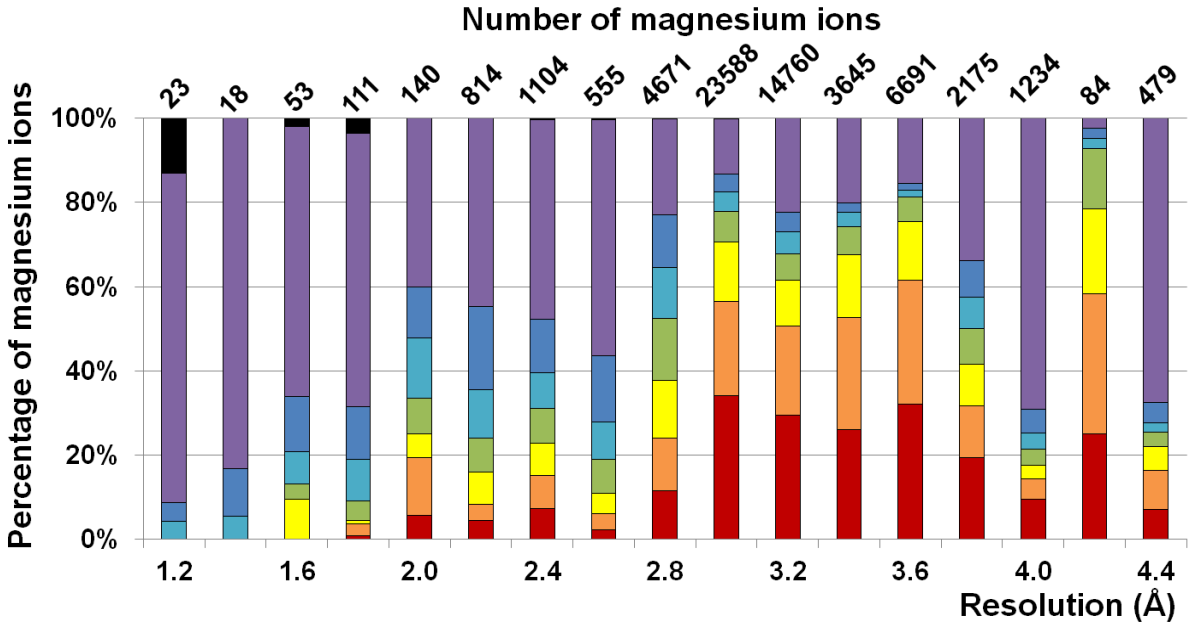


1) Search stops after 6 closest ligands are found
2) All new ligands should have a ligand-Mg-ligand angle of more than 50° with all previously found ligands.
3) Amongst nitrogen atoms only endocyclic nitrogen (–N=) without hydrogen bond(s) to a non-water molecule were accepted
4) No bidentate interaction with phosphate group

Prevalence of poorly coordinated magnesium ions in RNA structures



- More than half of the magnesium ions exhibit **a very incomplete inner-sphere** with CN < 3
- Magnesium ions with CN=4-6 comprised of **31%** of sites for ribosome, and **37%** of sites for non-ribosome
- Low CN even at **high resolution**

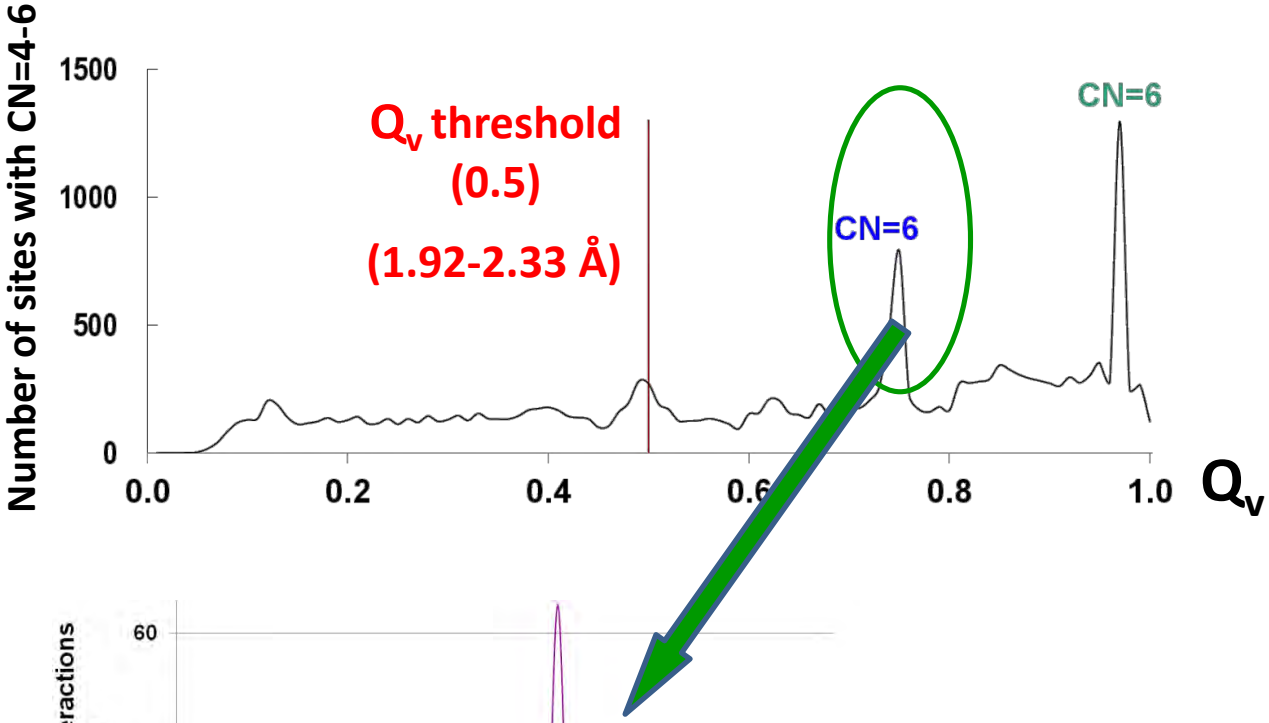


Quality assessment parameters

1. **4** ≤CN ≤ **6**
2. **Q**uality of **v**alence sum (Q_v)
3. **Q**uality of **s**ymmetry of ligands distribution (Q_s)
4. **Q**uality of B-factor correlation with B-factor of local **e**nvironment (Q_e)

2. Quality of valence sum (Q_v)

Agreement of the bond valence sum (ΣV_i) with the magnesium oxidation state (+2)

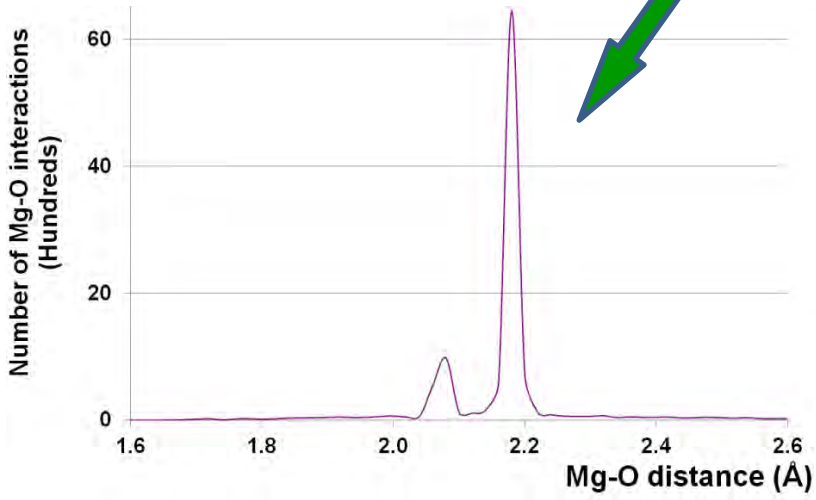


$$Q_v = 1 - \frac{|\sum V_i - 2|}{2}$$

Where,

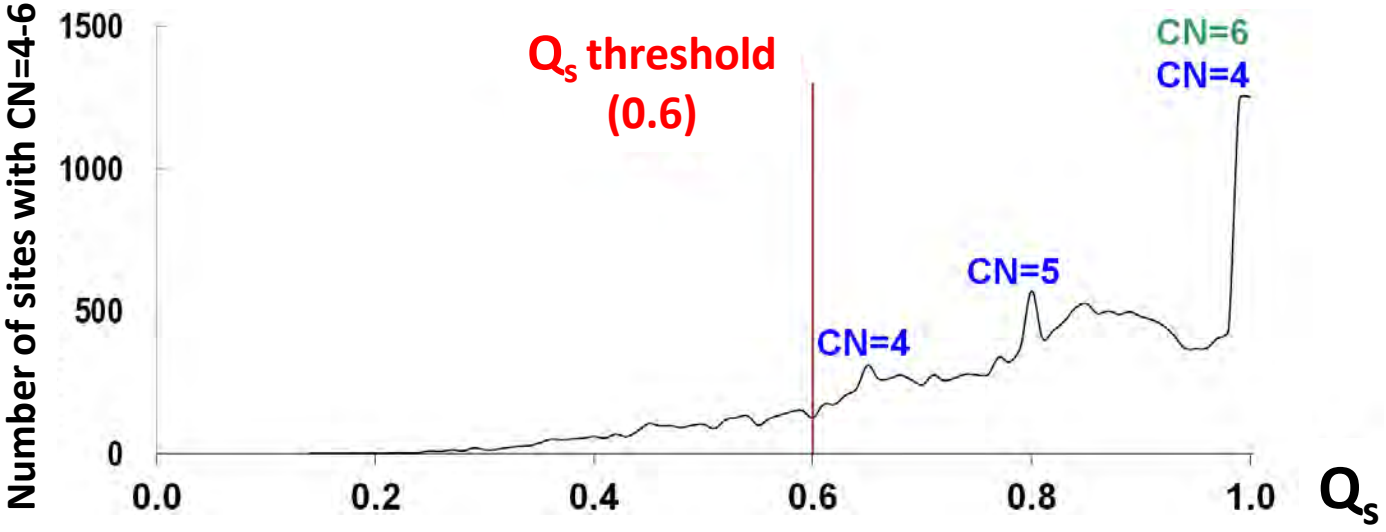
$$V_i = \exp\left(\frac{R_0 - R_i}{0.37\text{Å}}\right)^*$$

* Brown, I.D. (2009) Chem. Rev., 109, 6858-6919



Incorrect 2.18 Å Mg-water distance used in Phenix and REFMAC by default

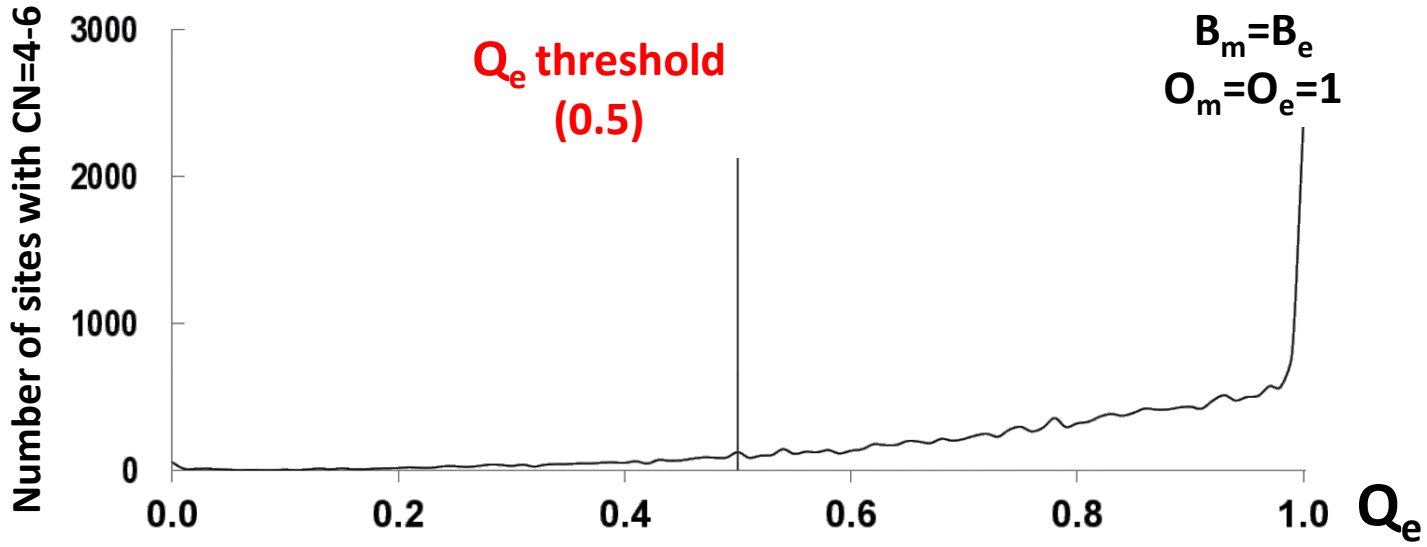
3. Quality of symmetry of ligands distribution around the magnesium (Q_s)



$$Q_s = 1 - \frac{|\sum \vec{V}_i|}{\sum V_i}$$

Q_s reflects the geometrical symmetry of the coordination sphere required for the octahedral geometry. It represents the summation of bond valence vectors ($\vec{V}_i$)

4. Quality of B-factor correlation with B-factor of local environment (Q_e)



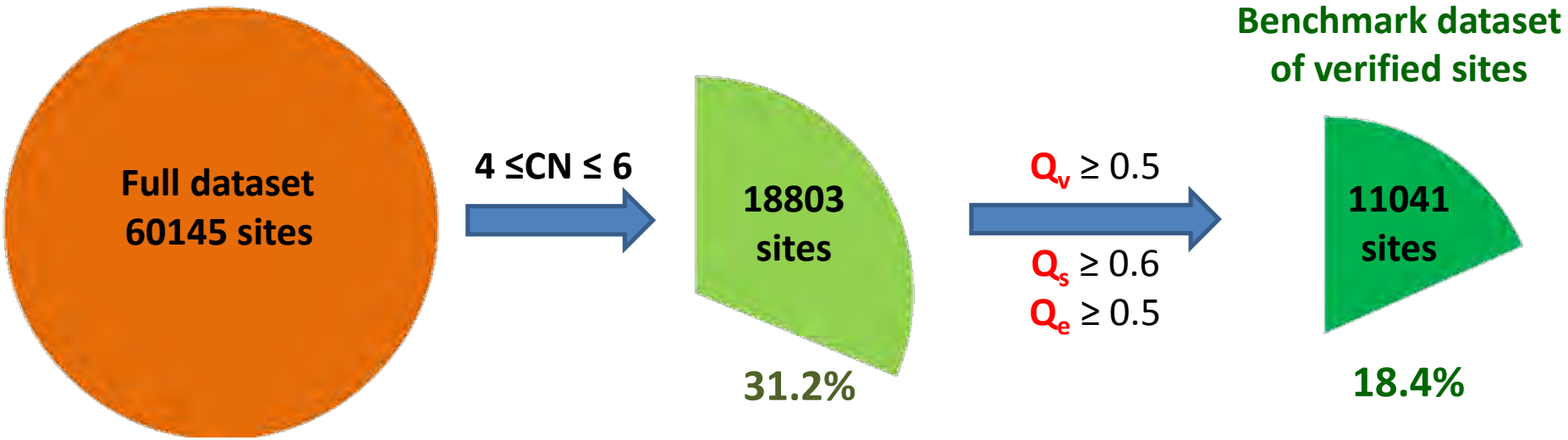
$$Q_e = \min(O_m, O_e) * \min\left(\frac{B_m/O_m}{B_e/O_e}, \frac{B_e/O_e}{B_m/O_m}\right)$$

$O_m=O_e=1$ →

$$Q_e = 1 * \min\left(\frac{B_m}{B_e}, \frac{B_e}{B_m}\right)$$

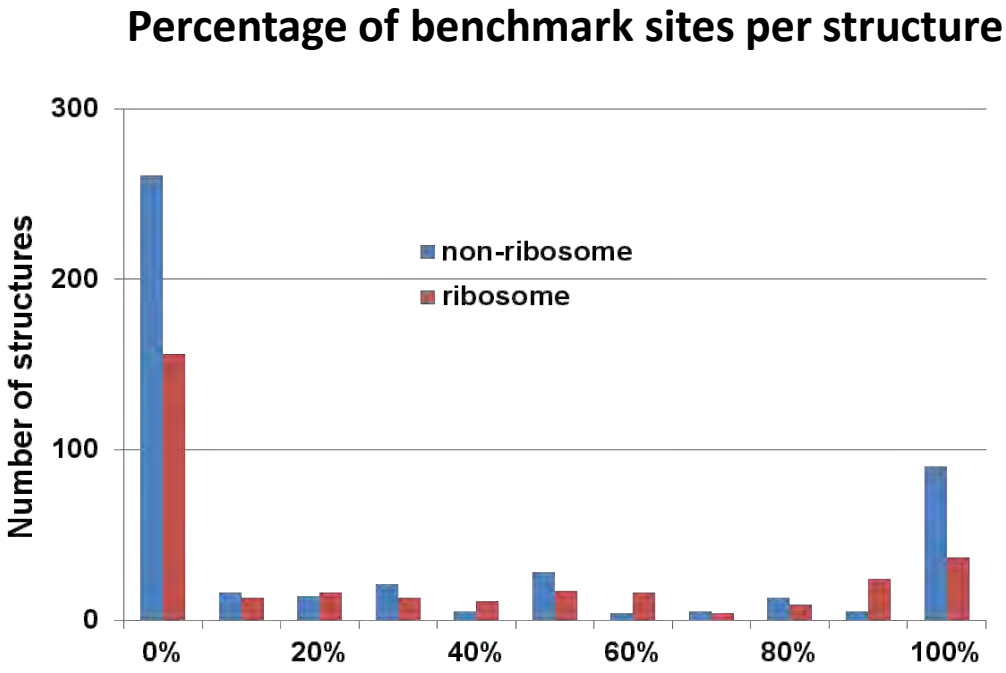
Q_e reflects the agreement of the magnesium B-factor (B_m) and occupancy (O_m) with its environment (B_e,O_e) within 4 Å of the magnesium

The benchmark dataset

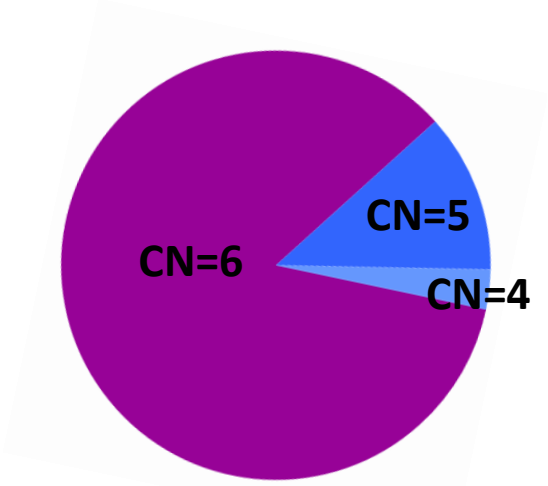


- 10499 sites from 179 ribosomal structures
- 542 sites from 164 non-ribosomal structures

The benchmark dataset

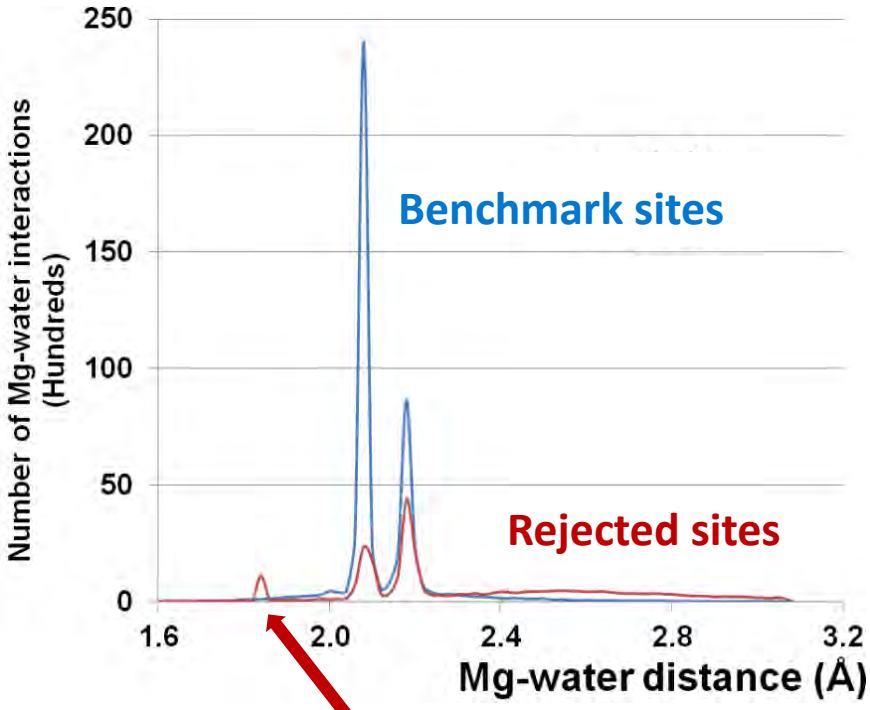


Tendency to have all good sites or all bad sites in a RNA structure

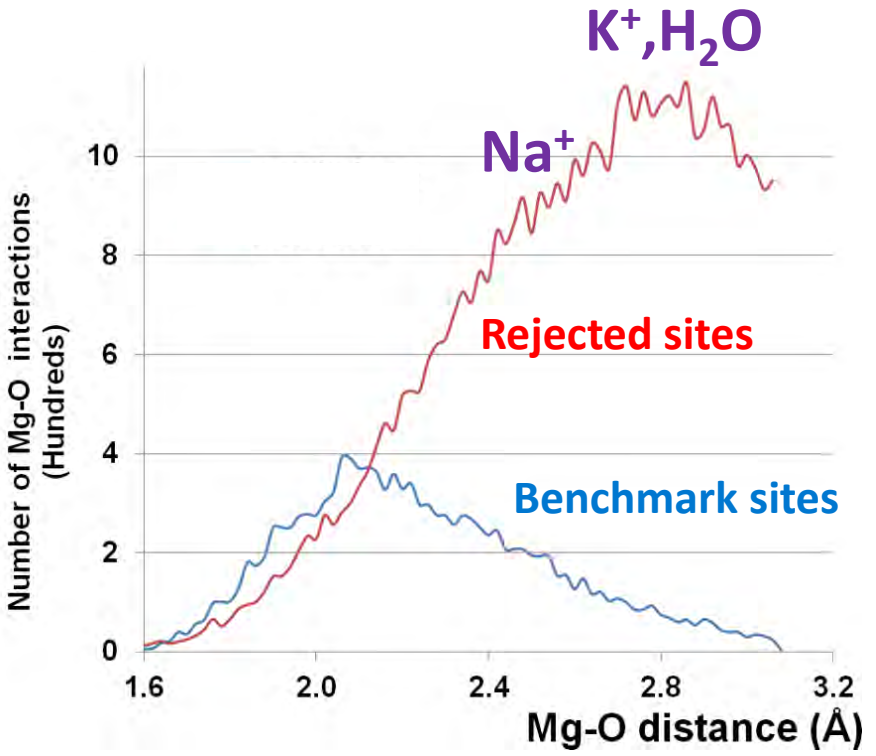


Coordination number of sites in the benchmark dataset

Distances in benchmark sites and in rejected sites

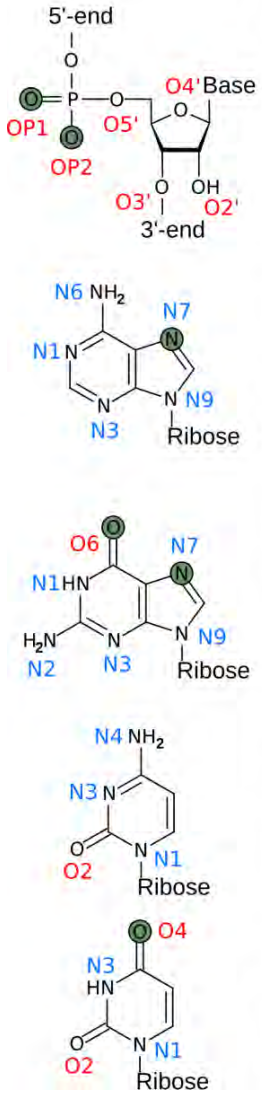


200 sites with 1.83 Å Mg^{2+}-water distance



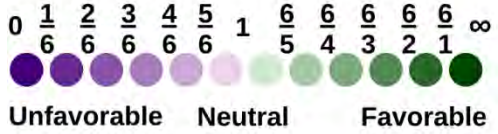
Easy-to-use restraints in refinement software are necessary !

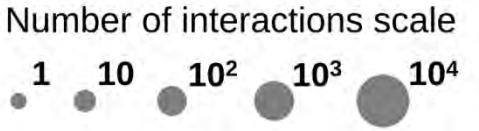
Normalized frequencies of nucleotide atoms for magnesium coordination



Atom name			Inner-sphere		Outer-sphere	
			F_{atom}	N	F_{atom}	N
O_P		OP1	4.30	5008	1.97	10326
		OP2	4.94	5759	2.96	15496
O_R		O2'	0.08	95	0.53	2775
		O3'	0.03	36	0.54	2802
		O4'	0.005	6	0.06	294
		O5'	0.04	48	0.54	2823
A	N_B	N1 (-N=)	0.06	16	0.60	750
		N3 (-N=)	0.004	1	0.18	228
		N6 (NH_2)	0	0	0.84	1048
		N7 (-N=)	0.75	208	2.41	3002
		N9 (R-N<)	0	0	0.0008	1
G	O_B	O6	1.47	564	3.77	6475
	N_B	N1 (-NH-)	0	0	0.19	333
		N2 (NH_2)	0.005	2	0.09	147
		N3 (-N=)	0	0	0.11	182
		N7 (-N=)	1.21	462	3.59	6175
		N9 (R-N<)	0	0	0.005	9
C	O_B	O2	0.13	39	0.37	490
	N_B	N1 (R-N<)	0	0	0.005	6
		N3 (-N=)	0	0	0.36	477
		N4 (NH_2)	0.003	1	0.44	587
U	O_B	O2	0.01	2	0.25	231
		O4	2.12	442	2.44	2291
	N_B	N1 (R-N<)	0	0	0.002	2
		N3 (-NH-)	0	0	0.07	68

$$F_{atom}(X) = \frac{N_{Mg\ interactions\ with\ atom\ X} / N_{\ all\ Mg\ interactions}}{N_{atoms\ X} / N_{all\ potential\ atoms}}$$

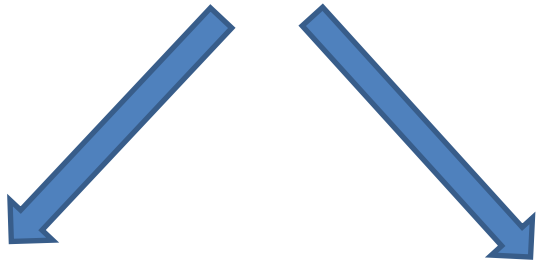




These preferences can be used in crystallographic model building and binding sites' prediction

Part II

Systematic classification of magnesium sites in RNA structures



RNA-inner
(at least one RNA atom in the inner sphere)

RNA-outer
(only water in inner sphere, but RNA present the outer sphere)

(sites with inner-sphere interactions with protein or small molecule were excluded)

RNA-inner class

Number of phosphate oxygens → **Number of ribose and base atoms** → **Combination of ribose and base atoms + Geometrical isomerism**

		Site type	Number of sites
$\#O_P=0$ (657)	$\#(O_R/O_B/N_B)=1$	O_R	17
		O_B	300
		N_B	187
	$\#(O_R/O_B/N_B)=2$	$2O_R$	5
		$O_R \bullet O_B$	1
		$2O_B$	10
		$O_B \bullet N_B$	13
		$2N_B$	122
.....			
$\#O_P=2$ (1843)	$\#(O_R/O_B/N_B)=0$	*cis*-$2O_P$	1403
		trans-$2O_P$	193
	$\#(O_R/O_B/N_B)=1$	*cis*-$2O_P \bullet O_R$	35
		cis-$2O_P \bullet O_B$	94
		cis-$2O_P \bullet N_B$	96
	$\#(O_R/O_B/N_B)=2$	*cis*-$2O_P \bullet 2O_R$	1
		cis-$2O_P \bullet O_R \bullet O_B$	2
		cis-$2O_P \bullet 2O_B$	18
		trans-$2O_P \bullet 2O_B$	1

Classification is based on combination of ligands in the inner sphere

Number of atoms		Site type	Number of sites
$\#O_P=0$ (657)	$\#(O_R/O_B/N_B)=1$	O_R	17
		O_B	300
		N_B	187
	$\#(O_R/O_B/N_B)=2$	$2O_R$	5
		$O_R \bullet O_B$	1
		$2O_B$	10
		$O_B \bullet N_B$	13
		$2N_B$	122
	$\#(O_R/O_B/N_B)=3$	$2O_B \bullet N_B$	1
	$\#(O_R/O_B/N_B)=4$	$2O_R \bullet 2O_B$	1
$\#O_P=1$ (3937)	$\#(O_R/O_B/N_B)=0$	O_P	410
		$O_P \bullet P_{out}$	913
		$O_P \bullet 2P_{out}$	1156
		$O_P \bullet 3P_{out}$	645
		$O_P \bullet 4P_{out}$	173
		$O_P \bullet 5P_{out}$	27
	$\#(O_R/O_B/N_B)=1$	$O_P \bullet O_R$	44
		$O_P \bullet O_B$	454
		$O_P \bullet N_B$	104
	$\#(O_R/O_B/N_B)=2$	$O_P \bullet 2O_R$	8
		$O_P \bullet 2O_B$	1
		$O_P \bullet O_B \bullet N_B$	1
	$\#(O_R/O_B/N_B)=3$	$O_P \bullet O_R \bullet 2O_B$	1
$\#O_P=2$ (1843)	$\#(O_R/O_B/N_B)=0$	*cis*-$2O_P$	1403
		trans-$2O_P$	193
	$\#(O_R/O_B/N_B)=1$	*cis*-$2O_P \bullet O_R$	35
		cis-$2O_P \bullet O_B$	94
		cis-$2O_P \bullet N_B$	96
	$\#(O_R/O_B/N_B)=2$	*cis*-$2O_P \bullet 2O_R$	1
		cis-$2O_P \bullet O_R \bullet O_B$	2
		cis-$2O_P \bullet 2O_B$	18
		trans-$2O_P \bullet 2O_B$	1
$\#O_P=3$ (796)	$\#(O_R/O_B/N_B)=0$	*fac*-$3O_P$	353
		mer-$3O_P$	365
	$\#(O_R/O_B/N_B)=1$	*fac*-$3O_P \bullet O_R$	3
		mer-$3O_P \bullet O_R$	19
		fac-$3O_P \bullet O_B$	55
		mer-$3O_P \bullet O_B$	1
$\#O_P=4$ (52)	$\#(O_R/O_B/N_B)=0$	$4O_P$	52

RNA-inner class

7285 sites

39 site types

Additional criteria for most populous site type (O_P) – number of phosphates in the outer sphere

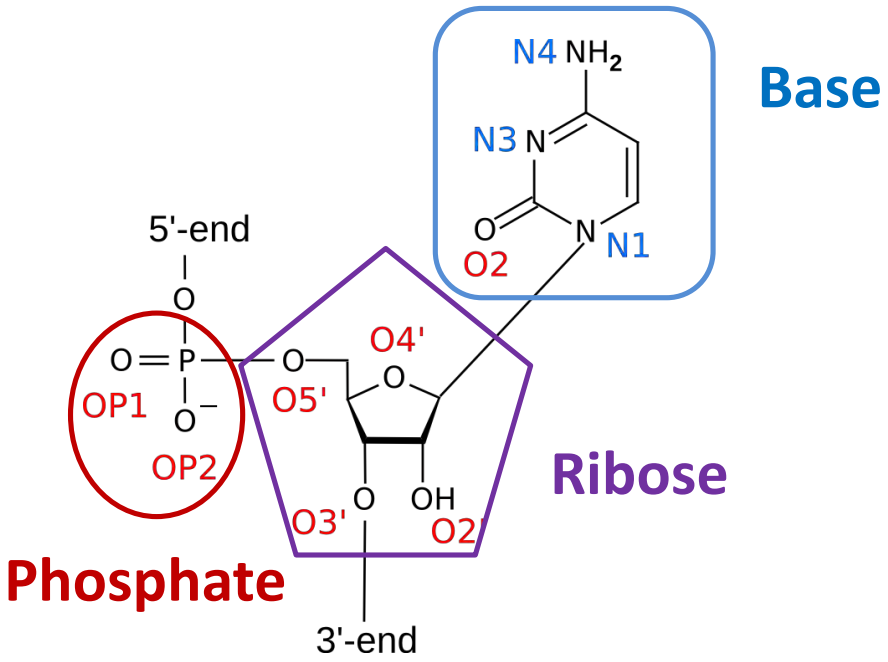
cis- conformation is 9 times more populous than *trans*-

The names of site types give extensive information about the ligands

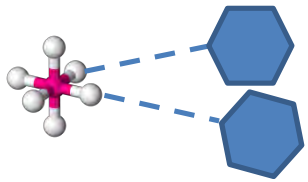
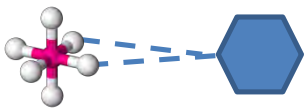
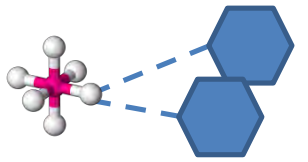
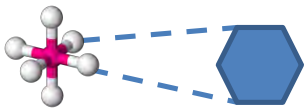
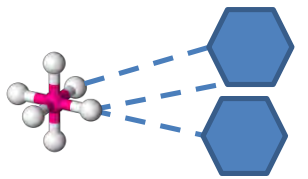
Maximal number of phosphates

RNA-outer class: Outer-sphere moieties

Definition



Outer-sphere moieties count

One moiety	Two moieties
	
	
	

RNA-outer class

3243 sites 80 site types

Number of moieties		Site type	Number of sites
Phosphate	Ribose/Base		
$\#P_{out}=0$ (927)	$\#(R_{out}/B_{out})=1$	R_{out}	19
		B_{out}	70
	$\#(R_{out}/B_{out})=2$	$2R_{out}$	1
		$R_{out}\bullet B_{out}$	7
		$2B_{out}$	404
	$\#(R_{out}/B_{out})=3$	$2R_{out}\bullet B_{out}$	2
		$R_{out}\bullet 2B_{out}$	9
		$3B_{out}$	230
	$\#(R_{out}/B_{out})=4$	$2R_{out}\bullet 2B_{out}$	1
		$R_{out}\bullet 3B_{out}$	12
		$4B_{out}$	62
	$\#(R_{out}/B_{out})=5$	$2R_{out}\bullet 3B_{out}$	3
		$R_{out}\bullet 4B_{out}$	5
		$5B_{out}$	27
	$\#(R_{out}/B_{out})=6$	$4R_{out}\bullet 2B_{out}$	1
		$R_{out}\bullet 5B_{out}$	1
		$6B_{out}$	68
	$\#(R_{out}/B_{out})=7$	$7B_{out}$	4
	$\#(R_{out}/B_{out})=8$	$4R_{out}\bullet 4B_{out}$	1
$\#P_{out}=1$ (843)	$\#(R_{out}/B_{out})=0$	P_{out}	57
	$\#(R_{out}/B_{out})=1$	$P_{out}\bullet R_{out}$	25
		$P_{out}\bullet B_{out}$	152
	$\#(R_{out}/B_{out})=2$	$P_{out}\bullet 2R_{out}$	1
		$P_{out}\bullet R_{out}\bullet B_{out}$	9
		$P_{out}\bullet 2B_{out}$	227
	$\#(R_{out}/B_{out})=3$	$P_{out}\bullet 2R_{out}\bullet B_{out}$	1
		$P_{out}\bullet R_{out}\bullet 2B_{out}$	28
		$P_{out}\bullet 3B_{out}$	222
	$\#(R_{out}/B_{out})=4$	$P_{out}\bullet 3R_{out}\bullet B_{out}$	1
		$P_{out}\bullet 2R_{out}\bullet 2B_{out}$	4
		$P_{out}\bullet R_{out}\bullet 3B_{out}$	17
		$P_{out}\bullet 4B_{out}$	81
	$\#(R_{out}/B_{out})=5$	$P_{out}\bullet 2R_{out}\bullet 3B_{out}$	3
		$P_{out}\bullet R_{out}\bullet 4B_{out}$	3
		$P_{out}\bullet 5B_{out}$	12
$\#P_{out}=2$ (728)	$\#(R_{out}/B_{out})=0$	$2P_{out}$	138
	$\#(R_{out}/B_{out})=1$	$2P_{out}\bullet R_{out}$	24
		$2P_{out}\bullet B_{out}$	117
	$\#(R_{out}/B_{out})=2$	$2P_{out}\bullet R_{out}\bullet B_{out}$	23
		$2P_{out}\bullet 2B_{out}$	174
	$\#(R_{out}/B_{out})=3$	$2P_{out}\bullet R_{out}\bullet 2B_{out}$	15
		$2P_{out}\bullet 3B_{out}$	145
	$\#(R_{out}/B_{out})=4$	$2P_{out}\bullet 2R_{out}\bullet 2B_{out}$	7
		$2P_{out}\bullet R_{out}\bullet 3B_{out}$	27
		$2P_{out}\bullet 4B_{out}$	44
	$\#(R_{out}/B_{out})=5$	$2P_{out}\bullet 2R_{out}\bullet 3B_{out}$	3
		$2P_{out}\bullet R_{out}\bullet 4B_{out}$	5
		$2P_{out}\bullet 5B_{out}$	3
	$\#(R_{out}/B_{out})=6$	$2P_{out}\bullet R_{out}\bullet 5B_{out}$	1
		$2P_{out}\bullet 6B_{out}$	2
$\#P_{out}=3$ (469)	$\#(R_{out}/B_{out})=0$	$3P_{out}$	107
	$\#(R_{out}/B_{out})=1$	$3P_{out}\bullet R_{out}$	39
		$3P_{out}\bullet B_{out}$	90
	$\#(R_{out}/B_{out})=2$	$3P_{out}\bullet 2R_{out}$	3
		$3P_{out}\bullet R_{out}\bullet B_{out}$	23
		$3P_{out}\bullet 2B_{out}$	84
	$\#(R_{out}/B_{out})=3$	$3P_{out}\bullet 2R_{out}\bullet B_{out}$	5
		$3P_{out}\bullet R_{out}\bullet 2B_{out}$	39
		$3P_{out}\bullet 3B_{out}$	25
	$\#(R_{out}/B_{out})=4$	$3P_{out}\bullet 2R_{out}\bullet 2B_{out}$	1
		$3P_{out}\bullet R_{out}\bullet 3B_{out}$	17
		$3P_{out}\bullet 4B_{out}$	18
	$\#(R_{out}/B_{out})=5$	$3P_{out}\bullet 2R_{out}\bullet 3B_{out}$	2
		$3P_{out}\bullet R_{out}\bullet 4B_{out}$	9
		$3P_{out}\bullet 5B_{out}$	6
	$\#(R_{out}/B_{out})=6$	$3P_{out}\bullet R_{out}\bullet 5B_{out}$	1
$\#P_{out}=4$ (254)	$\#(R_{out}/B_{out})=0$	$4P_{out}$	105
	$\#(R_{out}/B_{out})=1$	$4P_{out}\bullet R_{out}$	36
		$4P_{out}\bullet B_{out}$	76
	$\#(R_{out}/B_{out})=2$	$4P_{out}\bullet R_{out}\bullet B_{out}$	6
		$4P_{out}\bullet 2B_{out}$	26
	$\#(R_{out}/B_{out})=3$	$4P_{out}\bullet R_{out}\bullet 2B_{out}$	1
		$4P_{out}\bullet 3B_{out}$	2
	$\#(R_{out}/B_{out})=4$	$4P_{out}\bullet R_{out}\bullet 3B_{out}$	1
	$\#(R_{out}/B_{out})=5$	$4P_{out}\bullet R_{out}\bullet 4B_{out}$	1
$\#P_{out}=5$ (22)	$\#(R_{out}/B_{out})=0$	$5P_{out}$	8
	$\#(R_{out}/B_{out})=1$	$5P_{out}\bullet R_{out}$	2
		$5P_{out}\bullet B_{out}$	5
	$\#(R_{out}/B_{out})=2$	$5P_{out}\bullet R_{out}\bullet B_{out}$	1
		$5P_{out}\bullet 2B_{out}$	6

Part III

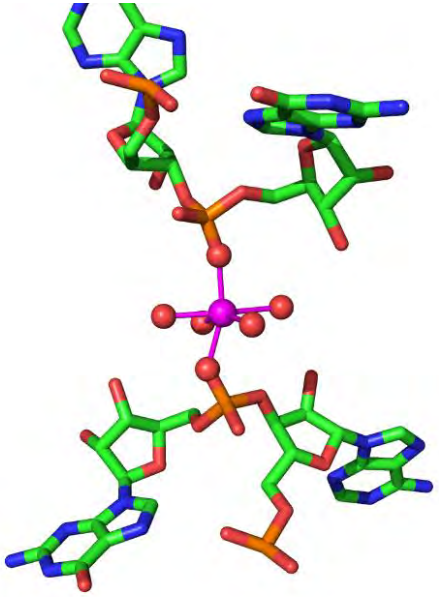
Detection of magnesium-binding motifs in RNA structures and discovery of new motifs

Site types define sites very specifically. Only a few additional definitions are needed to describe most magnesium binding motifs

Detection of previously reported motifs

"magnesium clamp"

magnesium binds two RNA strands through its inner sphere



Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann,B., Ehresmann,C. and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure,* **7,** 1439-1449.

Site types

cis-2O_P

trans-2O_P

\+

Additional criteria

The phosphates should be from different chains or from distant residues

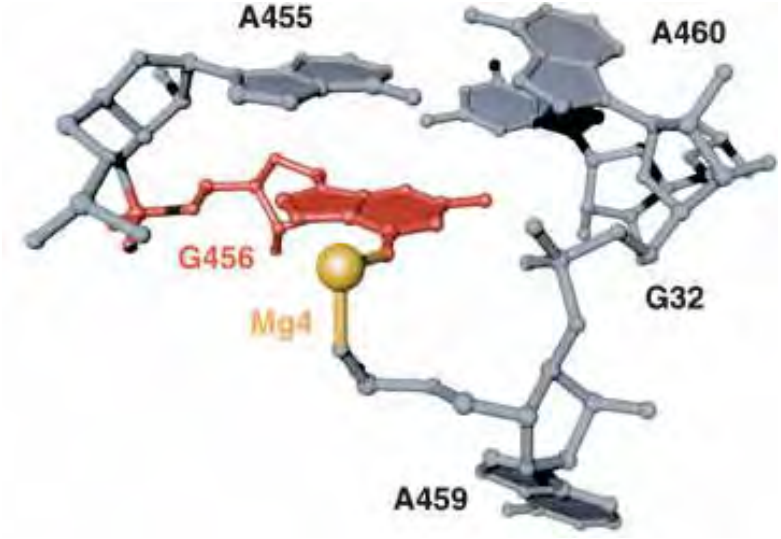
=

Sufficient motif definition

814 sites found in the benchmark dataset

Detection of previously reported motifs

G-phosphate magnesium-binding motif



Klein, D.J., Moore, P.B. and Steitz, T.A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA,* **10,** 1366-1379.

Figure reproduced from Klein et al. (2004) under the terms of a Creative Commons License

Site type
$O_P \bullet O_B$

+

Additional criteria
O_B atom should be from distant guanine

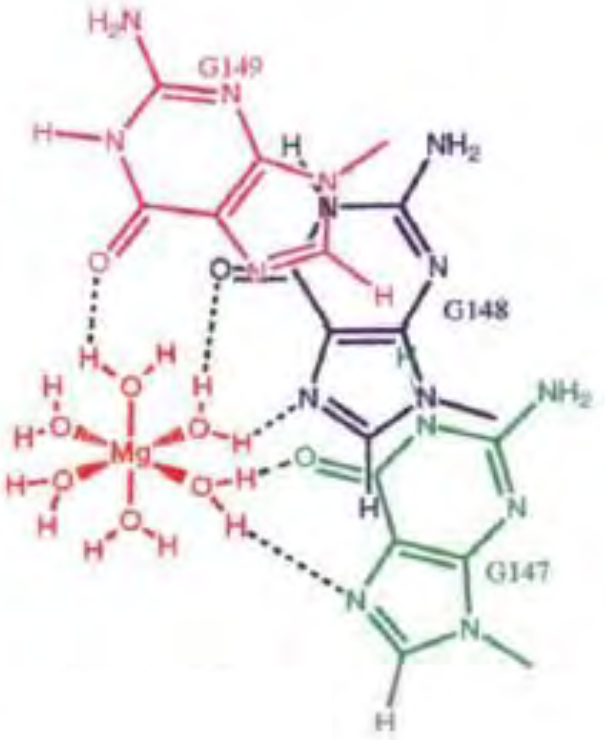
=

208 sites found in the benchmark dataset

Detection of previously reported motifs

Triple G magnesium binding motif

Tinoco, I.,Jr. and Kieft, J.S. (1997) The ion core in RNA folding. *Nat. Struct. Biol.,* **4,** 509-512.



Site type
$3B_{out}$

+

Additional criteria
$3B_{out}$ are sequentially consecutive guanines

=

23 sites found in the benchmark dataset

- “Validated motif”: a specific structural arrangement provided by RNA for magnesium binding which can be found in structures of multiple RNA molecules
- Only significantly represented motifs (>10 sites) are counted as “Validated motifs” in this study

Six “validated” motifs reported previously

id	Motif name	reference	Systematic classification		Detailed features of the motif	Number of sites
			Class	Type		
I	Magnesium clamp	Ennifar et al., 1999	RNA-inner	***cis*-2O$_P$** ***trans*-2O$_P$**	The two O$_P$ atoms are from distant phosphates	675 - cis 139 - trans
II	10-member ring	Hsiao et al., 2009	RNA-inner	***cis*-2O$_P$**	The two O$_P$ atoms are from sequentially consecutive phosphates	652
				***fac*-3O$_P$** ***mer*-3O$_P$**	The three O$_P$ atoms are from sequentially consecutive phosphates	Fac (57) Mer (32)
				***fac*-3O$_P$** ***mer*-3O$_P$**	10-member ring with additional O$_P$ atom coming from phosphate which is separated by one residue from either 10-member ring phosphates	Fac (5) Mer (71)
				4O$_P$	One magnesium coordinated by two unrelated 10-member rings	31
III	G-phosphate	Klein et al., 2004	RNA-inner	**O$_P$•O$_B$**	O$_B$ atom is from guanine	208
IV	G•G metal binding site	Correll et al., 1997	RNA-outer	**2P$_{out}$•2B$_{out}$**	Magnesium is bound through the outer-sphere by two sequentially consecutive guanines and by two phosphate moieties from residues n and n-1 where n is residue number of the lower guanine in the sequence.	61
		3ofa-A1569		**2B$_{out}$**	Magnesium is bound through the outer-sphere by two sequentially consecutive guanine moieties	268
		3v23-A3619		**P$_{out}$•2B$_{out}$**	Magnesium is bound through the outer-sphere by two consecutive guanines and a distant phosphate	38
V	Triple G motif	Tinoco & Keift, 1997	RNA-outer	**3B$_{out}$**	Magnesium is bound through the outer-sphere by three sequentially consecutive guanine bases	23
VI	Metal ion zipper	Correll et al., 1997	RNA-outer	**2P$_{out}$**	Magnesium binds two distant phosphates through the outer-sphere	74

Number of atoms		Site type	Number of sites
$\#O_P=0$ (657)	$\#(O_R/O_B/N_B)=1$	O_R	17
		O_B	300
		N_B	187
	$\#(O_R/O_B/N_B)=2$	$2O_R$	5
		$O_R \bullet O_B$	1
		$2O_B$	10
		$O_B \bullet N_B$	13
		$2N_B$	122
	$\#(O_R/O_B/N_B)=3$	$2O_B \bullet N_B$	1
	$\#(O_R/O_B/N_B)=4$	$2O_R \bullet 2O_B$	1
$\#O_P=1$ (3937)	$\#(O_R/O_B/N_B)=0$	O_P	410
		$O_P \bullet P_{out}$	913
		$O_P \bullet 2P_{out}$	1156
		$O_P \bullet 3P_{out}$	645
		$O_P \bullet 4P_{out}$	173
		$O_P \bullet 5P_{out}$	27
	$\#(O_R/O_B/N_B)=1$	$O_P \bullet O_R$	44
		$O_P \bullet O_B$	454
		$O_P \bullet N_B$	104
	$\#(O_R/O_B/N_B)=2$	$O_P \bullet 2O_R$	8
		$O_P \bullet 2O_B$	1
		$O_P \bullet O_B \bullet N_B$	1
	$\#(O_R/O_B/N_B)=3$	$O_P \bullet O_R \bullet 2O_B$	1
$\#O_P=2$ (1843)	$\#(O_R/O_B/N_B)=0$	*cis*-$2O_P$	1403
		trans-$2O_P$	193
	$\#(O_R/O_B/N_B)=1$	*cis*-$2O_P \bullet O_R$	35
		cis-$2O_P \bullet O_B$	94
		cis-$2O_P \bullet N_B$	96
	$\#(O_R/O_B/N_B)=2$	*cis*-$2O_P \bullet 2O_R$	1
		cis-$2O_P \bullet O_R \bullet O_B$	2
		cis-$2O_P \bullet 2O_B$	18
		trans-$2O_P \bullet 2O_B$	1
$\#O_P=3$ (796)	$\#(O_R/O_B/N_B)=0$	*fac*-$3O_P$	353
		mer-$3O_P$	365
	$\#(O_R/O_B/N_B)=1$	*fac*-$3O_P \bullet O_R$	3
		mer-$3O_P \bullet O_R$	19
		fac-$3O_P \bullet O_B$	55
		mer-$3O_P \bullet O_B$	1
$\#O_P=4$ (52)	$\#(O_R/O_B/N_B)=0$	$4O_P$	52

RNA-inner class

7285 sites

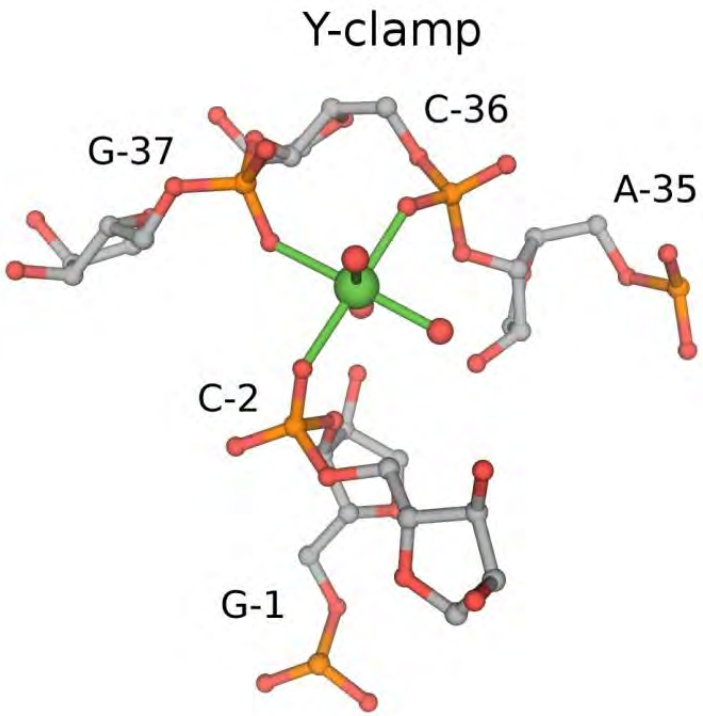
39 site types

Populous site types can represent "validated" motifs

New motifs discovery is important for annotation, understanding and prediction of RNA structure

New motifs (I-II)

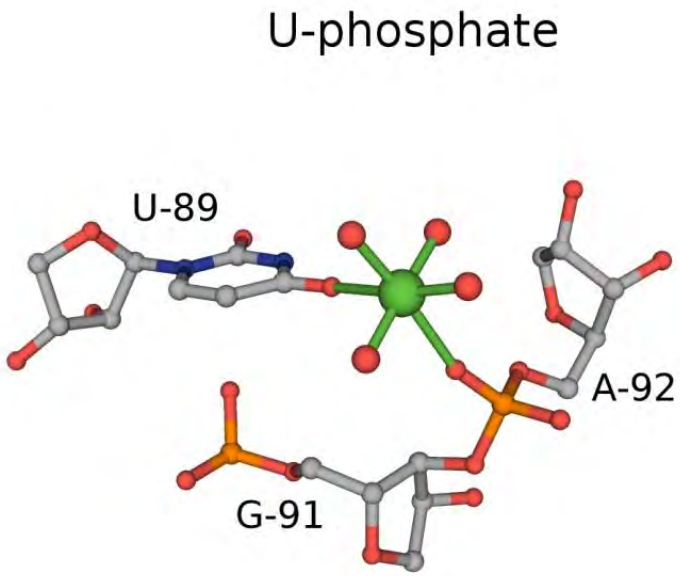
I **Site type: *mer*-3O_P**



10-member ring and one distant O_P atom in *mer*- conformation

238 sites

II **Site type: O_P•O_B**



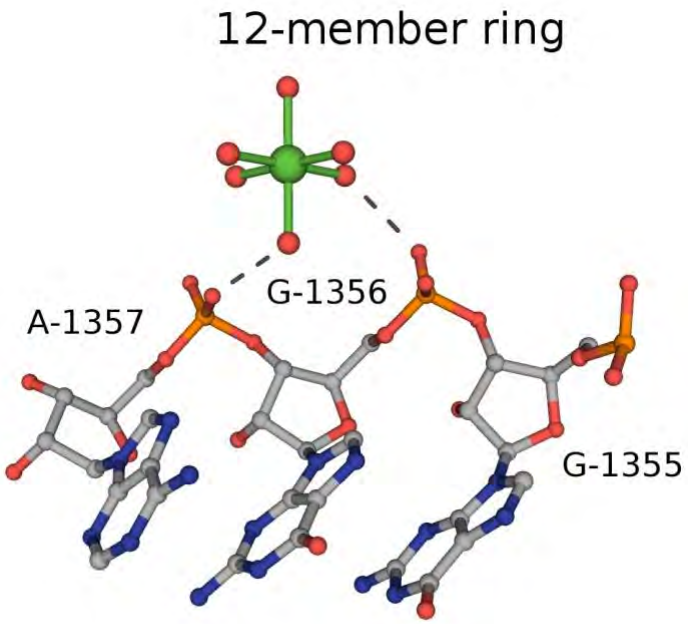
Similar to G-phosphate, but O_B atom is from uracil

165 sites

New motifs (III-IV)

III

Site type: $2P_{out}$

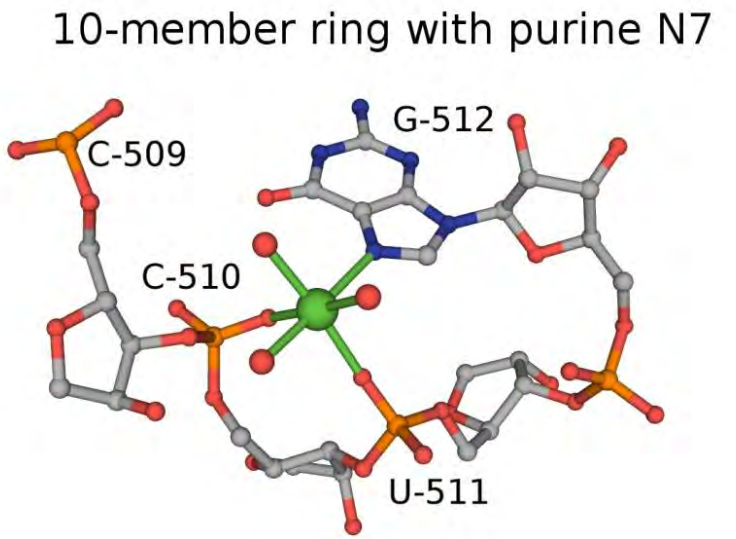


Two sequentially consecutive phosphates in the outer-sphere, forming a ring of 12 non-hydrogen atoms

56 sites

IV

Site type: $2O_P \bullet N_B$



10-member ring with N7 atom of purine base which is separated by one residue from either of the 10-member ring phosphates

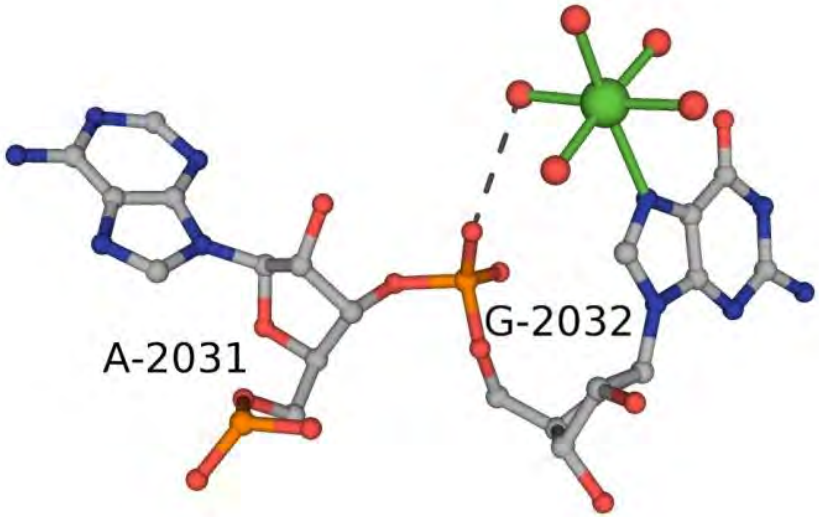
40 sites

New motifs (V-VI)

V

Site type: N_B

G-N7 macrochelate I



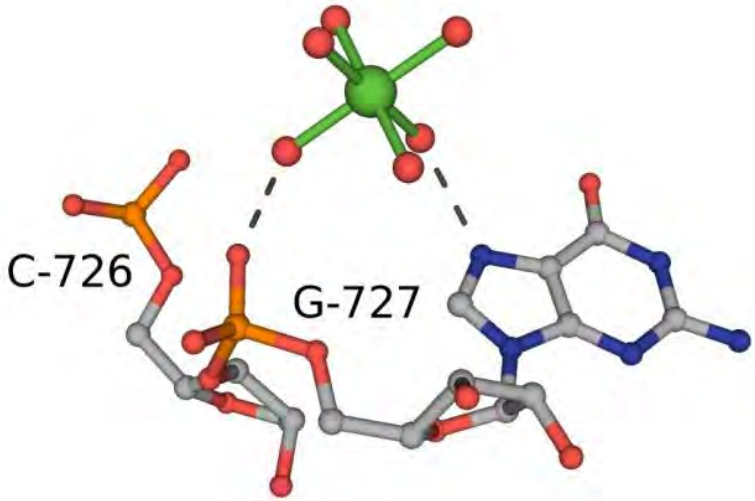
Magnesium is bound by guanine N7 atom in inner-sphere and phosphate of the same residue in the outer-sphere

238 sites

VI

Site type: P_{out}•B_{out}

G-N7 macrochelate II



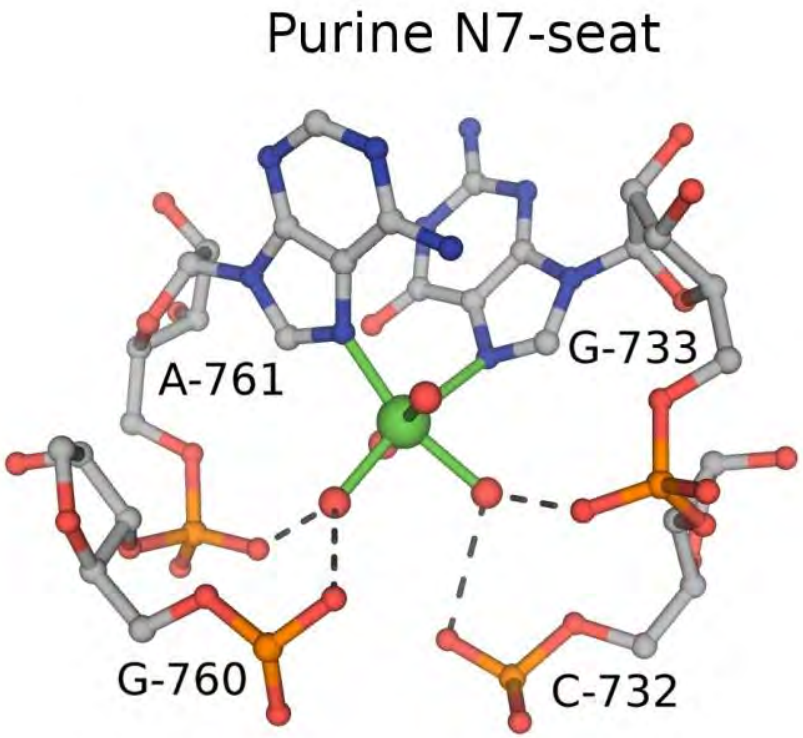
Magnesium is bound through the outer-sphere by guanine N7 atom and phosphate of the same residue

16 sites

New motifs (VII)

VII

Site type: $2N_B$



Magnesium is bound by two guanine or adenine N7 atoms in inner-sphere and capped by 3 or 4 downstream phosphates in the outer-sphere

107 sites

Summary

- Benchmark dataset of reliable sites is built
- A systematic tree-like classification of magnesium sites is developed
- The classification was used to describe previously reported magnesium-binding motifs, to discover new motifs, and to screen the whole PDB for presence of these motifs

Potential application

- Can be used to discover unique magnesium-binding patterns and new motifs
- Annotation of RNA structures
- Magnesium binding sites prediction
- Atomic preferences can be helpful during crystallographic refinement of structures
- Improvement of magnesium modelling in new structures

The dataset is available upon request by contacting
wladek@iwonka.med.virginia.edu

Acknowledgements

Heping Zheng
(equal contribution)

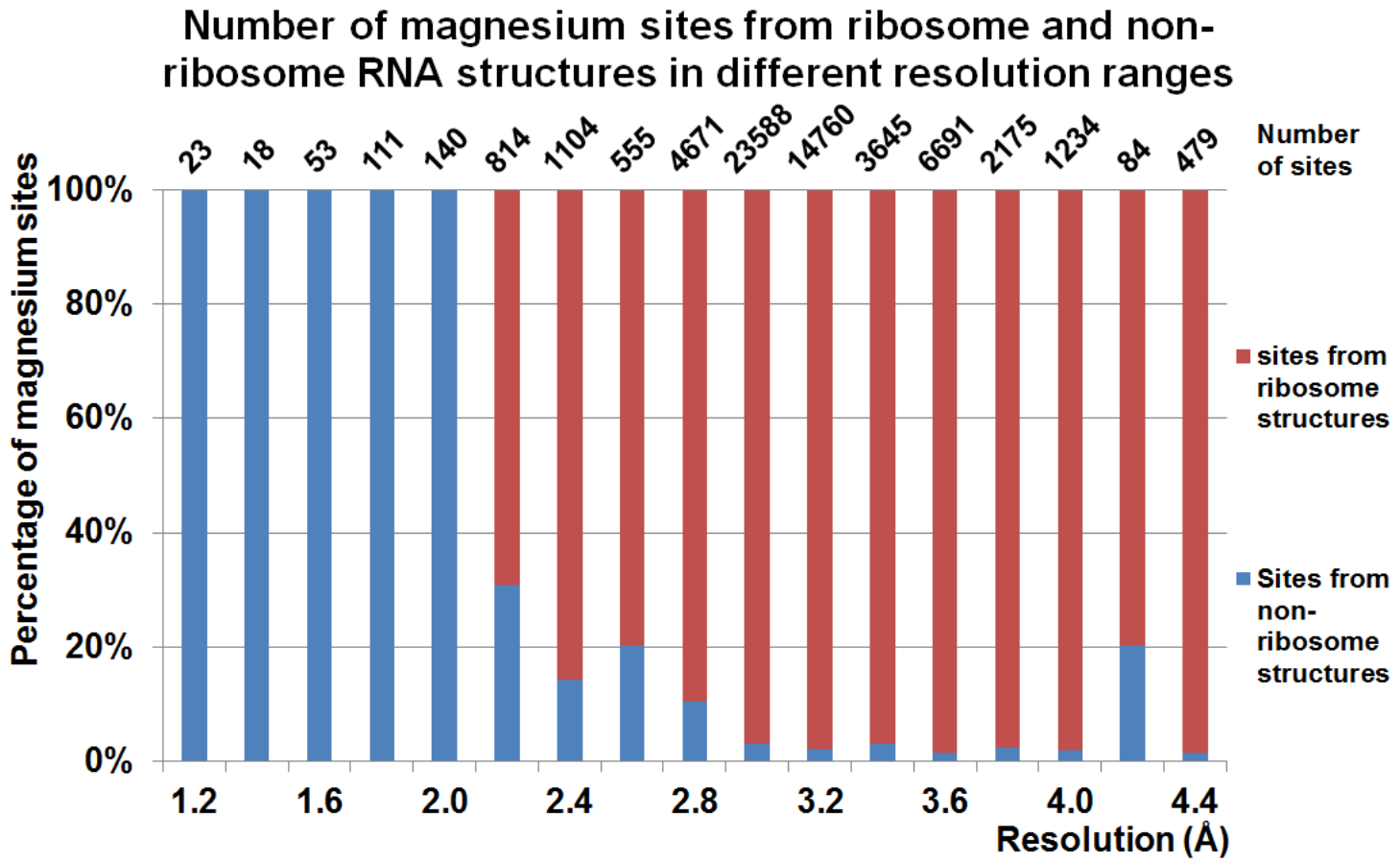
Katarzyna Mikolajczak

Prof. Wladek Minor

All the lab members

- CSGID, NIAID contract HHSN272201200026C
- MCSG U54- GM094585
- X-ray Crystallography of Macromolecules GM053163

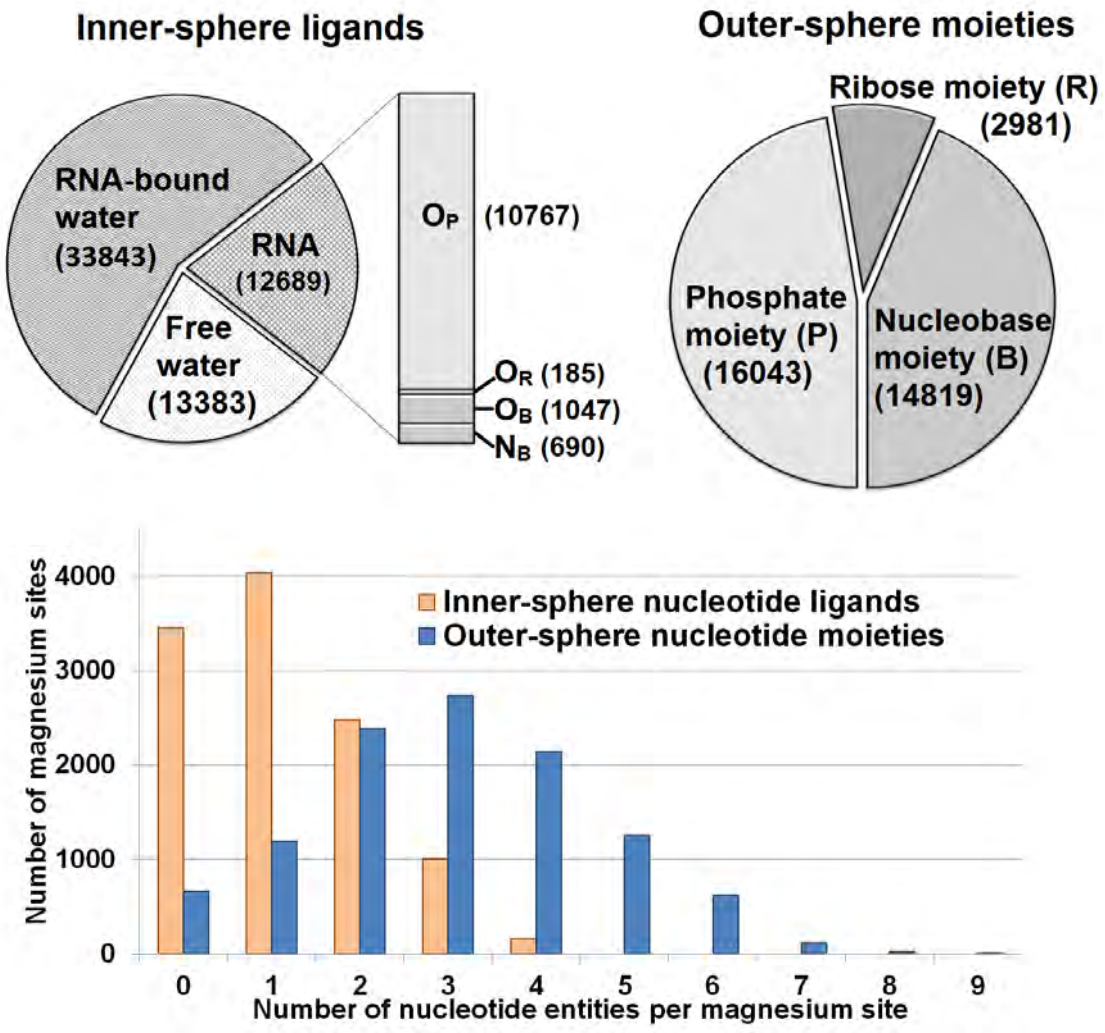
Number of sites vs. resolution



Five classes of sites in the benchmark dataset

	Total	**RNA-polynuclear**	**RNA-hybrid**	**RNA-inner**	**RNA-outer**
all data	11041 [18]	293 [7]	220 [20]	7285 [24]	3243 [56]
Ribosome	10499 [18]	259 [7]	156 [17]	7064 [24]	3020 [56]
non-ribosome	542 [20]	34 [27]	64 [34]	221 [18]	223 [59]

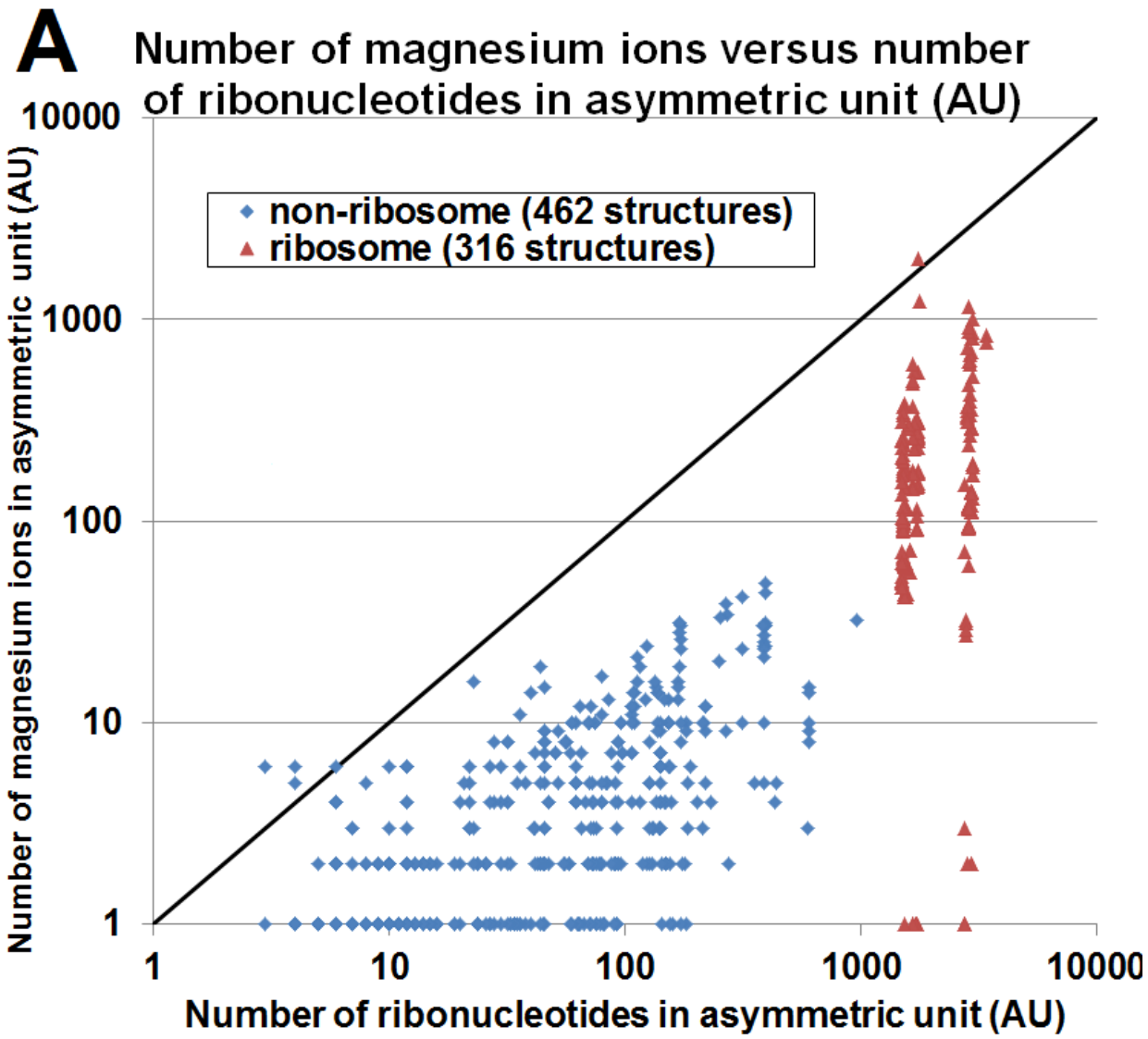
Number of inner-sphere ligands and outer-sphere moieties per site

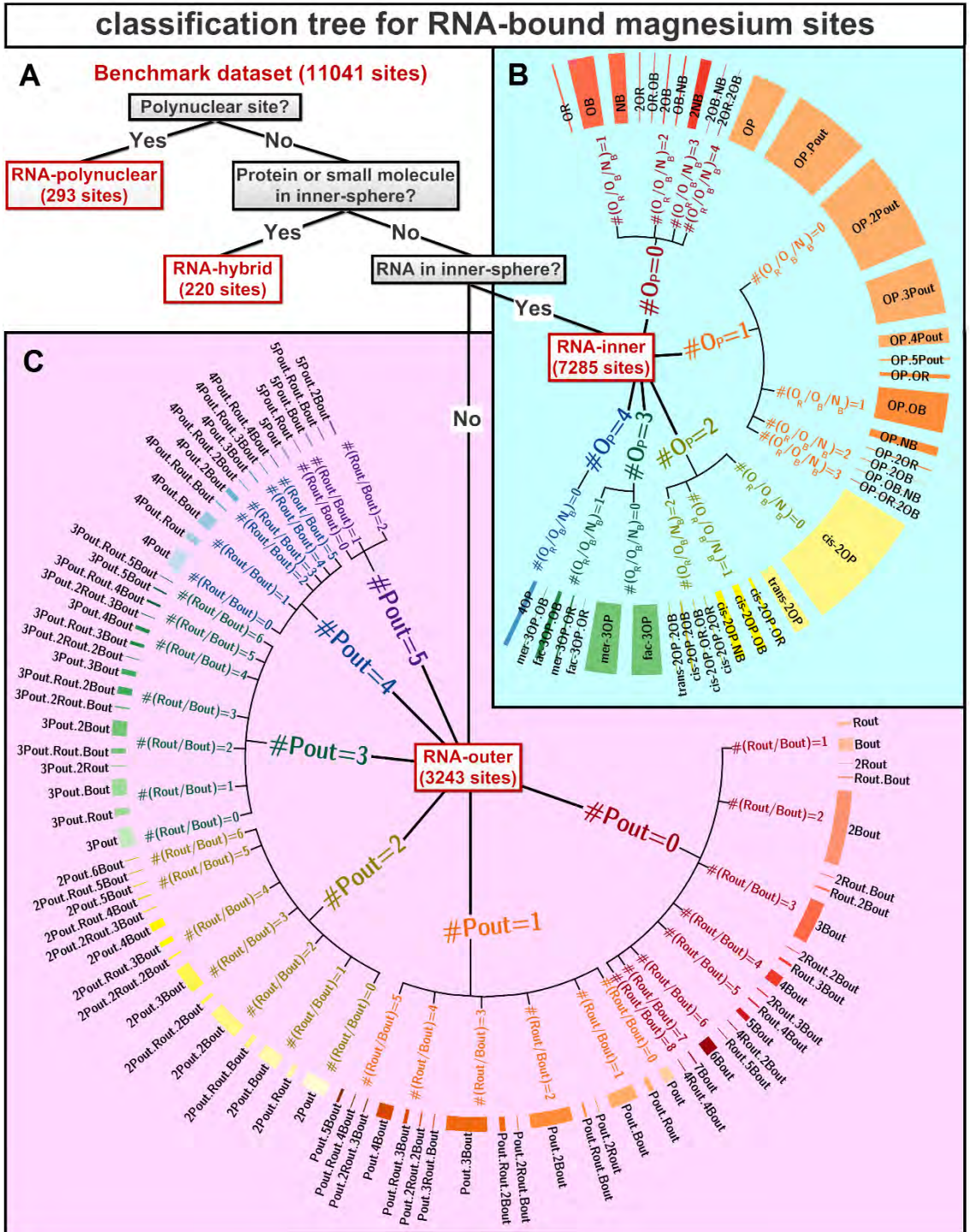


Newly defined motifs

id	Motif name	reference	Systematic classification		Detailed features of the motif	Number of sites
			Class	type		
A	Y-clamp	2z75-B301	RNA-inner	*mer*-3O_P	10-member ring and one distant O_P atom in *mer*-conformation	238
B	U-phosphate	2yie-Z1116	RNA-inner	O_P•O_B	O_B atom is from uracil	165
C	12-member ring	2avy- A1566	RNA-outer	2P_{out}	Magnesium is bound through the outer-sphere by two sequentially consecutive phosphates froming a ring with 12 non-hydrogen atoms	56
D	Purine-N7 seat	2qba-B3321	RNA-inner	2N_B	Magnesium is bound by two guanine or adenine N7 atoms in inner-sphere and capped by 3 or 4 downstream phospates through the outer-sphere	107
E	G-N7 macro-chelate I	3v29-A3359	RNA-inner	N_B	Magnesium is bound by guanine N7 atom in inner-sphere and phospate of the same residue in the outer-sphere	38
F	G-N7 macro-chelate II	2aw7-A1569	RNA-outer	P_{out}•B_{out}	Magnesium is bound through the outer-sphere by guanine N7 atom and phospate of the same residue	16
G	10-member ring with Purine-N7	3v23-A3540	RNA-inner	2O_P•N_B	10-member ring with N7 atom of purine base which is separated by one residue from either of the 10-member ring phosphates	40

Number of magnesium sites per asymmetric unit

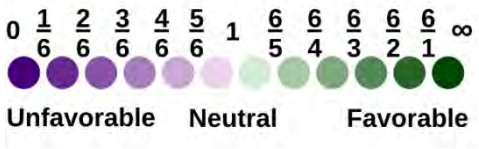



classification tree for RNA-bound magnesium sites
A
Benchmark dataset (11041 sites)
Polynuclear site?
Yes
No
RNA-polynuclear (293 sites)
Protein or small molecule in inner-sphere?
Yes
No
RNA-hybrid (220 sites)
RNA in inner-sphere?
Yes
No
RNA-inner (7285 sites)
RNA-outer (3243 sites)
B
C
#Pout=0
#Pout=1
#Pout=2
#Pout=3
#Pout=4
#Pout=5

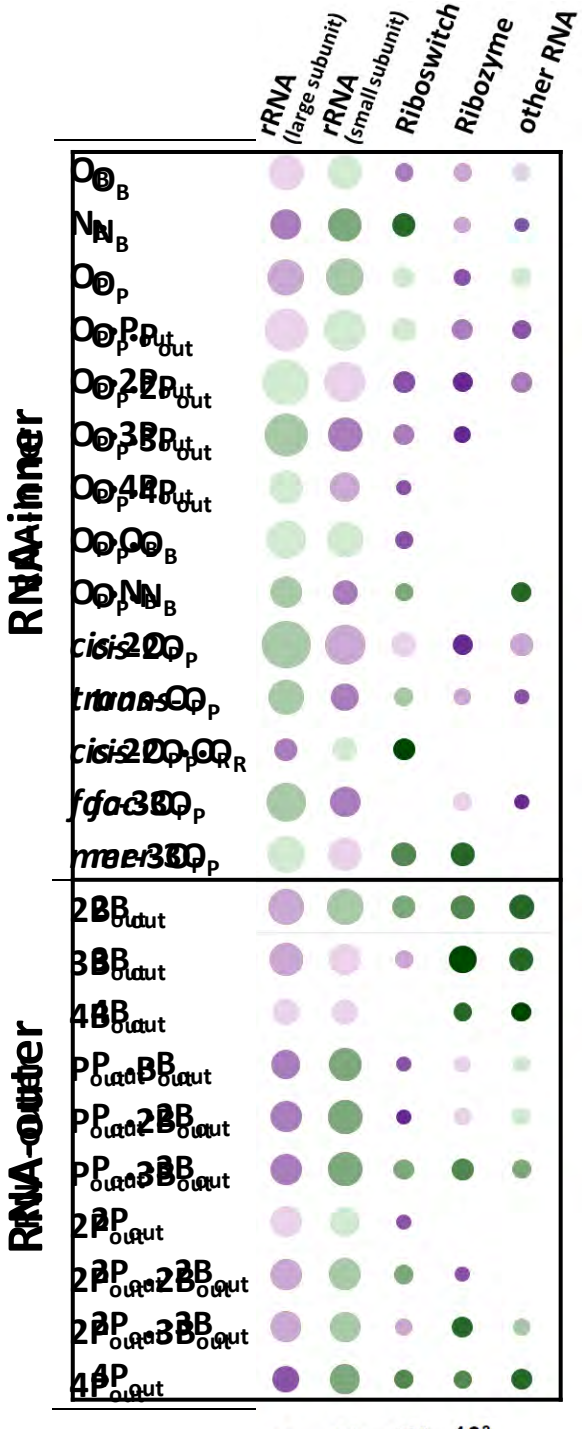
Relative abundance of magnesium site types in different RNA types

- RNA classification is based on SCOR database

$$F_{S\ in\ R}(S) = \frac{N_{sites\ of\ type\ S\ in\ RNA\ type\ R}\ /Nall_{\ sites\ of\ type\ S}}{N_{sites\ in\ RNA\ type\ R}\ /Nall_{\ sites}}$$



- RNA-inner sites are more abundant in the large ribosomal subunit
- RNA-outer sites are more abundant in the small ribosomal subunit



Atom Name		Inner-sphere interactions				Outer-sphere interactions				Number of atoms in the full dataset
		F_{atom}	Number of bench-mark sites	Number of all sites	Percent. of sites retained in bench-mark set	F_{atom}	Number of bench-mark sites	Number of all sites	Percent of sites retained in bench-mark set	
O_P	OP1	4.30	5008	16932	29.6	1.97	10326	16175	63.8	899146
	OP2	4.94	5759	20182	28.5	2.96	15496	24143	64.2	899146
O_R	O2'	0.08	95	2909	3.3	0.53	2775	4912	56.5	899146
	O3'	0.03	36	1432	2.5	0.54	2802	4388	63.9	899146
	O4'	0.005	6	519	1.2	0.06	294	486	60.5	899146
	O5'	0.04	48	605	7.9	0.54	2823	4091	69.0	899146
A	N1 (N-endo)	0.06	16	229	7.0	0.60	750	1018	73.7	214231
	N3 (N-endo)	0.004	1	246	0.4	0.18	228	557	40.9	214231
	N6 (N-amino)	0	0	31	0	0.84	1048	1814	57.8	214231
	N7 (N-endo)	0.75	208	926	22.5	2.41	3002	4055	74.0	214231
	N9 (N-ribose)	0	0	4	0	0.0008	1	10	10.0	214231
G	O6 (sugar opposite)	1.47	564	7288	7.7	3.77	6475	10271	63.0	295424
	N1 (NH-endo)	0	0	43	0	0.199	333	480	69.4	295424
	N2 (N-amino)	0.005	2	38	5.3	0.09	147	226	65.0	295424
	N3 (N-endo)	0	0	163	0	0.11	182	351	51.9	295424
	N7 (N-endo)	1.21	462	2787	16.6	3.59	6175	9736	63.4	295424
	N9 (N-ribose)	0	0	18	0	0.0054	9	18	50.0	295424
C	O2 (sugar side)	0.13	39	752	5.2	0.37	490	846	57.9	228245
	N1(N-ribose)	0	0	3	0	0.005	6	18	33.3	228245
	N3 (N-endo)	0	0	180	0	0.36	477	667	71.5	228245
	N4 (N-amino)	0.003	1	60	1.8	0.44	587	870	67.5	228245
U	O2 (sugar side)	0.01	2	457	0.4	0.25	231	358	64.5	161246
	O4 (sugar opposite)	2.12	442	3364	13.1	2.44	2291	3500	65.5	161246
	N1 (N-ribose)	0	0	6	0	0.002	2	3	66.7	161246
	N3 (NH-endo)	0	0	22	0	0.07	68	112	60.7	161246

Type (# PDB structures)		PDB ID
5s/23s rRNA (103)		1dfu, 1feu, 1ffk, 1jj2, 1k73, 1k8a, 1k9m, 1kc8, 1kd1, 1kqs, 1m1k, 1m90, 1mms, 1n8r, 1nji, 1q7y, 1q81, 1q82, 1q86, 1qvf, 1qvg, 1s72, 1vq4, 1vq5, 1vq6, 1vq7, 1vq8, 1vq9, 1vqk, 1vql, 1vqm, 1vqn, 1vqo, 1vqp, 1vs6, 1vs8, 1vt2, 1w2b, 1yhq, 1yi2, 1yij, 1yj9, 1yjn, 1yjw, 1ytu, 2aw4, 2awb, 2i2t, 2i2v, 2otj, 2otl, 2qam, 2qao, 2qba, 2qbc, 2qbe, 2qbg, 2qbi, 2qbk, 2qex, 354d, 3cc2, 3cc4, 3cc7, 3cce, 3ccj, 3ccl, 3ccm, 3ccq, 3ccr, 3ccs, 3ccu, 3ccv, 3cd6, 3cpw, 3cxc, 3g6e, 3g71, 3i1n, 3i1r, 3i20, 3i22, 3i56, 3kni, 3oas, 3oat, 3ofc, 3ofd, 3ofq, 3ofr, 3ofz, 3og0, 3orb, 3r8s, 3r8t, 3v23, 3v25, 3v27, 3v29, 3v2d, 3v2f, 4gar, 4gau
16s rRNA (86)		1vs5, 1vs7, 1zz5, 2avy, 2aw7, 2fqn, 2g5k, 2i2p, 2i2u, 2j02, 2oe5, 2qal, 2qan, 2qb9, 2qbb, 2qbd, 2qbf, 2qbh, 2qbj, 2qou, 2qov, 2qow, 2qox, 2qoy, 2qoz, 2qp0, 2qp1, 2uua, 2uuc, 2v46, 2v48, 2wdg, 2wdh, 2wdk, 2xfz, 2z4k, 2z4l, 2z4m, 2z4n, 3df1, 3df2, 3df3, 3df4, 3i1m, 3i1o, 3i1p, 3i1q, 3i1s, 3i1t, 3i1z, 3i21, 3i8g, 3i9b, 3i9d, 3oaq, 3oar, 3ofa, 3ofb, 3ofo, 3ofp, 3ofx, 3ofy, 3or9, 3ora, 3td0, 3v22, 3v24, 3v26, 3v28, 3v2c, 3v2e, 4dh9, 4dha, 4dhb, 4dhc, 4dr1, 4dr2, 4dr3, 4dr4, 4dr5, 4dr6, 4dr7, 4gaq, 4gas, 4gd1, 4gd2
Riboswitch (40+11)	Adenosine	3la5
	Biotin binding	1f27
	c-di-GMP-II	3q3z
	di-GMP-I (10)	1lng, 3mum, 3mur, 3mut, 3muv, 3mxh, 3ucu, 3ucz, 3ud3, 3ud4
	Fluoride (4)	4en5, 4ena, 4enb, 4enc
	FMN (2)	2yie, 2yif
	Glycine (10)	3owi, 3oww, 3owz, 3ox0, 3oxb, 3oxd, 3oxe, 3oxj, 3oxm, 3p49
	Lysine	3dil
	Purine	1y26
	SAM (2)	3gx5, 3v7e
	Streptomycin binding	1ntb
	Tetracycline	3egz
	TPP (4)	2gdi, 3d2g, 3d2v, 3d2x
	ykoK	2qbz
Ribozyme (20+11)	glmS (11)	2gcs, 2gcv, 2h0s, 2h0w, 2h0x, 2ho6, 2nz4, 2z75, 3b4a, 3b4b, 3b4c
	Delta virus	3nkb
	Diels-Alder	1yls
	HPRZ	4g6r
	Intron_gpI (7)	1hr2, 1l8v, 1x8w, 1zzn, 2r8s, 3bo2, 3bo3
	Intron_gpII (5)	4e8m, 4e8n, 4far, 4faw, 4fb0
	L1-RNA-ligase (3)	3hhn, 3ivk, 3r1l
	Lead dependent (2)	1nuj, 1nuv
tRNA (14)		1ehz, 1evv, 1ffy, 1j1u, 1n78, 1qu2, 1tra, 2zxu, 3l0u, 3q1q, 3q1r, 4ari, 4tra, 6tna
mRNA (4)		1l2x, 2hw8, 3mei, 437d
SRP (3)		1d4r, 1dul, 1hq1
Structural model (25)	bulge	1jzv
	duplex (17)	1dno, 1dnt, 1fuf, 1ik5, 1o3z, 1pjo, 1y99, 2ao5, 2g91, 2oiy, 2q1r, 2r1s, 2r20, 3cgs, 3jxq, 3ssf, 462d
	4-way junction	1egk
	Kinkturn	3iin
	Kissing-loops	2b8s
	Pseudoknot	2a43
	Single strand	3twh
	Stem-loop	1xjr
RNA-protein complex (38)		1tez, 1zbi, 1zbl, 2bx2, 2c0b, 2c4r, 2ez6, 2g8f, 2g8h, 2g8v, 2nuf, 2nug, 2o5i, 2ppb, 3aoh, 3dd2, 3hax, 3hif, 3hk2, 3hm9, 3ho1, 3htx, 3hvr,

The numbers of magnesium sites found for each magnesium binding motif

Magnesium binding motif	All RNA	rRNA (large subunit)	rRNA (small subunit)	Riboswitch	Ribozyme	Other RNA
Magnesium clamp	814	573	229	6	3	3
10-member ring (IIa) ***cis*-2O_P**	652	495	124	19	5	9
10-member ring (IIb) ***fac*-3O_P, *mer*-3O_P**	89	54	30	5	0	0
10-member ring (IIc) ***fac*-3O_P, *mer*-3O_P**	76	63	8	0	5	0
10-member ring (IId) **4O_P**	51	35	13	0	3	0
G-phosphate	208	136	72	0	0	0
G•G metal binding site (IVa) **2P_{out}•2B_{out}**	61	28	30	2	0	0
G•G metal binding site (IVb) **2B_{out}**	268	124	125	9	7	5
G•G metal binding site (IVc) **P_{out}•2B_{out}**	38	15	23	0	0	0
Triple G motif	23	12	9	0	1	1
Metal ion zipper	74	41	33	0	0	0
Y-clamp	238	170	57	11	11	0
U-phosphate	165	72	91	2	0	0
12-member ring	56	32	23	1	0	0
Purine-N7 seat	107	60	47	0	0	0
G-N7 macrochelate I	38	24	14	0	0	0
G-N7 macrochelate II	16	2	14	0	0	0
10-member ring with Purine-N7	40	35	5	0	0	0