Supporting Information

IsRNA1: *de novo* prediction and blind screening of RNA 3D structures

Dong Zhang, Jun Li and Shi-Jie Chen

Department of Physics, Department of Biochemistry, and Institute of Data Science and Informatics, University of Missouri, Columbia, Missouri, 65211, USA

Section I: Energy functions in coarse-grained force field

The energy functions for three local covalent energy terms in the coarse-grained (CG) force field (Eq. 1 in the main text) have the following functional forms:

$$E_{bond}(b) = K_b(b - b_0)^2 + \frac{H_b}{\sqrt{2\pi}\sigma_b} e^{-0.5(b-b_1)^2/\sigma_b^2} \quad \text{(S1)}$$

$$E_{angle}(\theta) = K_a(\theta - \theta_0)^2 + \frac{H_a}{\sqrt{2\pi}\sigma_a} e^{-0.5(\theta-\theta_1)^2/\sigma_a^2} \quad \text{(S2)}$$

$$E_{torsion}(\phi) = \sum_{n=1}^{4} K_n[1 + \cos(n\phi - \phi_n)] \quad \text{(S3)}$$

where b, θ and ϕ denote the bond stretching length between two connected beads, the bond angle between two adjacent bonds, and the torsion angle between three successive bonds, respectively. Parameters K_b, b_0, H_b, σ_b, and b_1 in Eq. S1 describe the strength of bond stretching and the equilibrium bond length. Parameters K_a, θ_0, H_a, σ_a, and θ_1 in Eq. S2 characterize the bending energy of the bond angle and the equilibrium bond angle. K_n and ϕ_n (n = 1, 2, 3, 4) in Eq. S3 are related to the local minima on the profile of the torsional energy. Those energy parameters were determined using the iterative simulated reference state approach, and their detailed values are given in Table S3~S5, respectively. Compared to the widely used harmonic potentials for bond stretching and bond angle terms in previous CG models, extra Gaussian terms are introduced in Eqs. S1 and S2 in order to capture more precise details of the statistical potential and to enable broader sampling of the backbone conformations in simulations.

The non-local term $E_{pair}(r)$ that accounts for non-canonical base pairing, base stacking, base-backbone, and backbone-backbone interactions in the CG force field is written as:

$$E_{pair}(r) = D_0\left[e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}\right] + \frac{H_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(r-r_1)^2}{2\sigma_1^2}} + \frac{H_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(r-r_2)^2}{2\sigma_2^2}}, r < r_{cut} \quad \text{(S4)}$$

Here r is the pairwise distance between two non-neighboring beads (not involved in the same one bond, bond angle, and torsion angle). The first Morse potential term in Eq. S4 is designed to give the overall shape of the knowledge-based energies, while

the latter two Gaussian terms account for local features of the energy profiles. All the energy parameters D_0, α, r_0, H_1, σ_1, r_1, H_2, σ_2, r_2, and the cut-off distance r_{cut} for different types of pairwise interactions were determined by the iterative simulated reference state approach, and the detailed values are listed in Table S6.

The last term $E_{LJ}(r)$ in the CG force field (Eq. 1 in the main text) describes the excluded volume interaction between any two non-neighboring beads (not involved in the same bond, bond angle, and torsion angle) through the repulsive Lennard-Jones potential

$$E_{LJ}(r) = \varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right],\ r < \sigma \tag{S5}$$

Here, the energy constant is set to $\varepsilon = 0.5kcal/mol$, and $\sigma = (\sigma_i + \sigma_j)/2$ is for the interaction between the *i*-th and *j*-th types of CG beads with diameters σ_i and σ_j (see Table S1 for details), respectively.

Section II: Accounting for energy correlation between different structural elements

As shown in Fig. S1, for some structural element, such as the bond angle $x = \theta_2(\mathrm{G_O} - \mathrm{R_C} - \mathrm{G_N})$, whose conformation is fixed by the strong chain connectivity constraints (bond stretching energies), the distribution $p_{ref}(\theta_2)$ in the coarse-grained Molecular Dynamics (CGMD) simulation generated reference state would match the observed statistics $p_{obs}(\theta_2)$, and we have $E(\theta_2) \approx 0$ according to Eq S6. In this case, this structural parameter should not be included in the coarse-grained force field in order to avoid redundancy. As a result, not all the structural elements need to be selected as collective variables in the final coarse-grained force field. For other structural parameters like bond angle $x = \theta_1(\mathrm{S} - \mathrm{R_C} - \mathrm{G_O})$, the distribution $p_{ref}(\theta_1)$ in the CGMD generated reference state is partially affected by the chain connectivity (bond stretching energies) and excluded volume interactions, but the reference distribution is not close to the observed $p_{obs}(\theta_1)$. After the related statistical potential $E(\theta_1) =$

$-k_B T \ln[p_{obs}(\theta_1)/p_{ref}(\theta_1)]$ (Eq S6) is embedded in the force field, the resultant simulated distribution, $p_{sim}(\theta_1)$, matches the observed $p_{obs}(\theta_1)$ (see Fig. S1B).

Fig. S2 procedurally illustrates how the proposed iterative simulated reference state approach deals with correlation effects within three correlated torsional angles through iterative construction of the reference state and stepwise force field building. Along with embedding the energy functions of the correlated structural parameters into the force field one-by-one, the CGMD-generated reference states $p_{ref}^{(n)}(\phi)$ ($n = 0,1,2$) for illustrated torsional angle ϕ were iteratively updated to change the effective potential energy $E^{(n)}(\phi) = -k_B T \ln\left[p_{obs}(\phi)/p_{ref}^{(n)}(\phi)\right]$ at each new construction step. After the final energy function $E(\phi) = E^{(2)}(\phi)$ is embedded into the force field, the simulated distribution $p_{sim}(\phi)$ is consistent with the observed distribution $p_{obs}(\phi)$ (see Fig. S2B). Compared with previous approaches that employ heuristic parameters, such as different weight coefficients for the different energy terms, the proposed iterative simulated reference state method distinguishes itself by directly treating the correlation effects, and the number of effective collective variables in the final force field is properly reduced. More details about the proposed iterative simulated reference state approach are given in our previous study.

Section III: Extraction of probability distributions $p_{obs}(x)$, $p_{ref}(x)$, and $p_{sim}(x)$

To obtain the observed probability distributions $p_{obs}(x)$, all the 592 structures in the experimental PDB dataset were firstly transformed from the atomic structures into coarse-grained representations, and then the probability density $p_{obs}(x)$ for each concerned structural parameter x in the coarse-grained representation was calculated. The statistical bin sizes for the bond length b, bond angle θ, torsion angle ϕ, and pairwise distance r are 0.01 Å, 1°, 6°, and 0.2 Å, respectively. For the pairwise distributions $p_{obs}(r)$, only the pairwise beads within distance $r \leq 20$ Å are considered in the statistics. In total, under our coarse-grained conformational representation, we observed 12 bonds (b), 18 bond angles (θ), 22 torsion angles (ϕ), and 55 contact

pairwise types (r) (those components for canonical base pair interactions are not included).

To generate the reference state $p_{ref}(x)$ (without the related energy function $E(x)$ at work in the force field) and the simulated distribution $p_{sim}(x)$ (with the related energy function $E(x)$ embedded into force field) under a given force field, using the native structures in the simulated dataset (contains 121 RNA molecules) as the initial states, we performed a series of coarse-grained Molecular Dynamics (CGMD) simulations through implementing Langevin dynamics (NVT ensemble) in the modified open source software LAMMPS. The simulation time is $t = 60$ ns at constant temperature $T = 300$K for each simulated structure, and the time step for integration was set to $\Delta t = 1$fs. After the first 50ns (5×10^7 simulation steps) of equilibration simulation under the given force field, the structure snapshots were collected from the last 10 ns simulation trajectories in the interval of 10ps to obtain a conformational ensemble. In total, there are 121,000 structures in this CGMD generated conformation ensemble. Using the previously mentioned statistical bin sizes, the distributions $p_{ref}(x)$ and $p_{sim}(x)$ for concerned structural elements were obtained based on the CGMD generated conformation ensemble.

Section IV: Parameterization of energy functions in the force field

The selection of effective collective variables for the final force field and determination of their related contributions (energy functions) into the force field are dependent on the embedded order in the proposed iterative simulated reference state approach, but the resultant total force field should be the same since the simulated distributions $p_{sim}(x)$ guided by the final total force field remains consistent with the observed distributions $p_{obs}(x)$ for all the structural parameters. Here, as in our previous work, we used the following embedding order to parameterize the energy functions in the force field: bond stretching and excluded volume interactions → bond angle potentials→ torsion angle energies → pairwise interactions.

1. We first directly determined all 18 bond stretching energies $E_{bond}(b)$ (including 6 types for base pairing interactions) from the observed statistics $p_{obs}(b)$ through $E_{bond}(b) = -k_B T \ln[p_{obs}(b)]$ (here the reference state $p_{ref}(b)$ is assumed to be constant). From the bond length geometries, the diameters (σ) of all coarse-grained beads for the excluded volume interactions $E_{LJ}(r)$ were estimated.
2. Using the bond stretching and excluded volume interactions as the basis for the force field, we parameterized the bond angle energies one-by-one through the iterative simulated reference state approach and then embedded the obtained energy functions into the force field until the simulated distributions $p_{sim}(\theta)$ were consistent with the observed distributions $p_{obs}(\theta)$ for all the bond angles, including those for the base pairing interactions.
3. With all the extracted bond angle energies added to the bond stretching and excluded volume interactions in our force field, the torsional angle energy functions were parametrized one-by-one through the iterative simulated reference state approach and imposed on the force field until the simulated distributions $p_{sim}(\phi)$ matched the observed ones $p_{obs}(\phi)$ for all the torsional angles, including those for base pairing interactions.
4. Finally, utilizing all the extracted torsional angle energies along with the bond stretching, bond angle, and excluded volume interactions in our force field, the pairwise interactions were determined step-by-step through the iterative simulated reference state approach and then embedded into the force field until the simulated distributions $p_{sim}(r)$ were consistent with the observed statistics $p_{obs}(r)$ for all 55 types of pairwise interactions.

To optimize the convergence speed for obtaining the targeted conformational ensemble and to minimize the number of total collective variables involved in the final CG force field, within each structural parameter category (such as all the bond angles in Step 2 above, torsional angles in Step 3, or pairwise distances in Step 4), the

structural parameter x of the largest $|E(x)| = \left|-k_B T \ln[\rho_{obs}(x)/\rho_{ref}(x)]\right|$ is chosen as the newest collective variable for the force field in each embedded step.

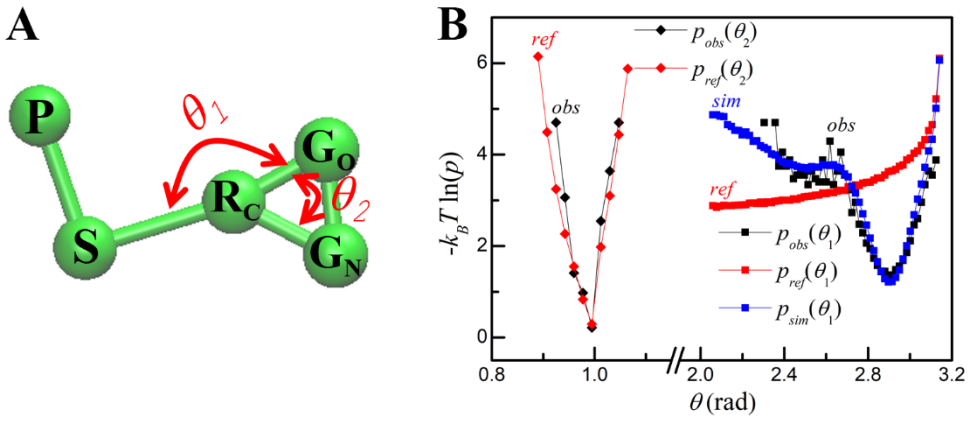

Figure S1 Effects of chain connectivity and excluded volume in construction of coarse-grained energy functions. (A) The definition of two representative bond bending angles in the IsRNA model are shown. (B) The observed probability distributions $p_{obs}(\theta)$ from the native structure dataset, the reference state probability distributions $p_{ref}(\theta)$ generated by coarse-grained Molecular Dynamics (CGMD) simulations using only the bond stretching and excluded volume interactions as the force field, and the simulated Boltzmann-like distribution $p_{sim}(\theta_1)$ from the CGMD simulation with $E(\theta_1) = -k_B T \ln [p_{obs}(\theta_1)/p_{ref}(\theta_1)]$ also embedded into the force field are shown in black, red, and blue, respectively.

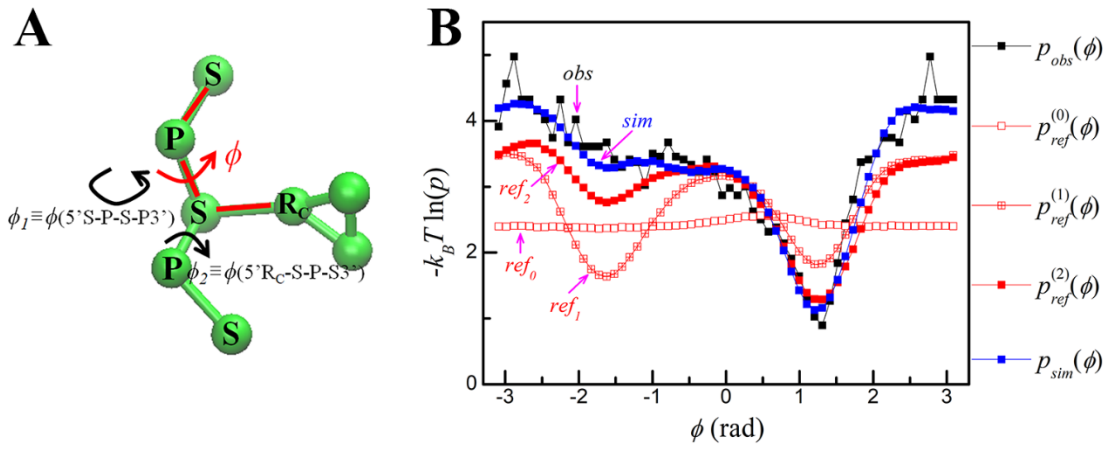

Figure S2 Energy correlations between different structural elements recognized by the proposed iterative simulated reference state approach. (A) Definitions of a torsion angle $\phi \equiv \phi(5'\mathrm{S}-\mathrm{P}-\mathrm{S}-\mathrm{R_C}3')$ and its two correlated torsion angles $\phi_1 \equiv \phi(5'\mathrm{S}-\mathrm{P}-\mathrm{S}-\mathrm{P}3')$ and $\phi_2 \equiv \phi(5'\mathrm{R_C}-\mathrm{S}-\mathrm{P}-\mathrm{S}3')$ in the IsRNA model are shown. (B) The iteratively updated reference distributions $p_{ref}(\phi)$ for the torsion angle ϕ due to the correlation effects during stepwise constructing the force field are shown in red. The $p_{ref}^{(0)}(\phi)$ distribution is generated by CGMD simulation with the bond stretching, bond angle bending, and the LJ excluded volume interactions as force field. The $p_{ref}^{(1)}(\phi)$ includes the correlated energy term $E(\phi_1)$ also embedded into the force field, and $p_{ref}^{(2)}(\phi)$ includes all the correlated energy terms $E(\phi_1)$ and $E(\phi_2)$ within the force field. The black squares denote the observed distribution $p_{obs}(\phi)$ from the known structures. The blue squares indicate the simulated distribution $p_{sim}(\phi)$ generated by the CGMD simulation with its related energy term $E(\phi) = -k_BT\ln[p_{obs}(\phi)/p_{ref}^{(2)}(\phi)]$ embedded into the force field.

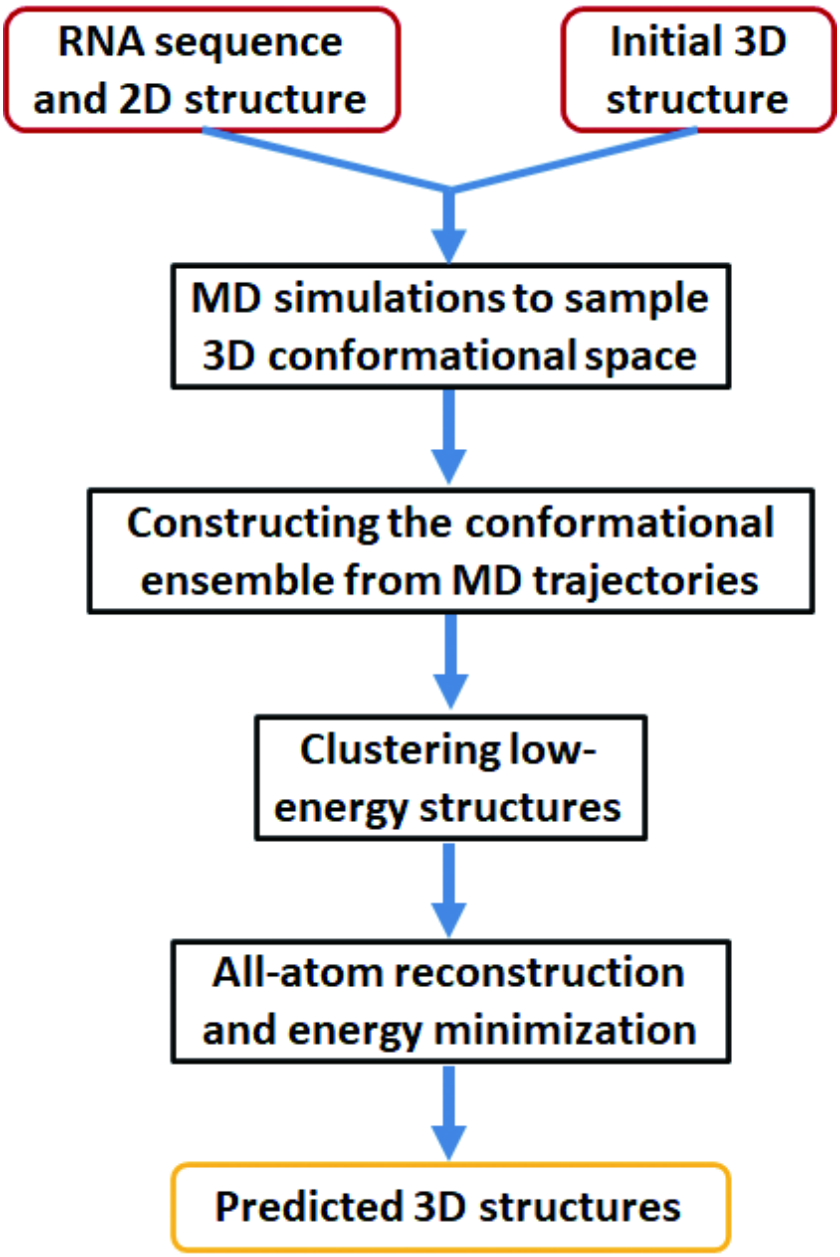

Figure S3 Illustrative flowchart for RNA 3D structure prediction by IsRNA1 model.

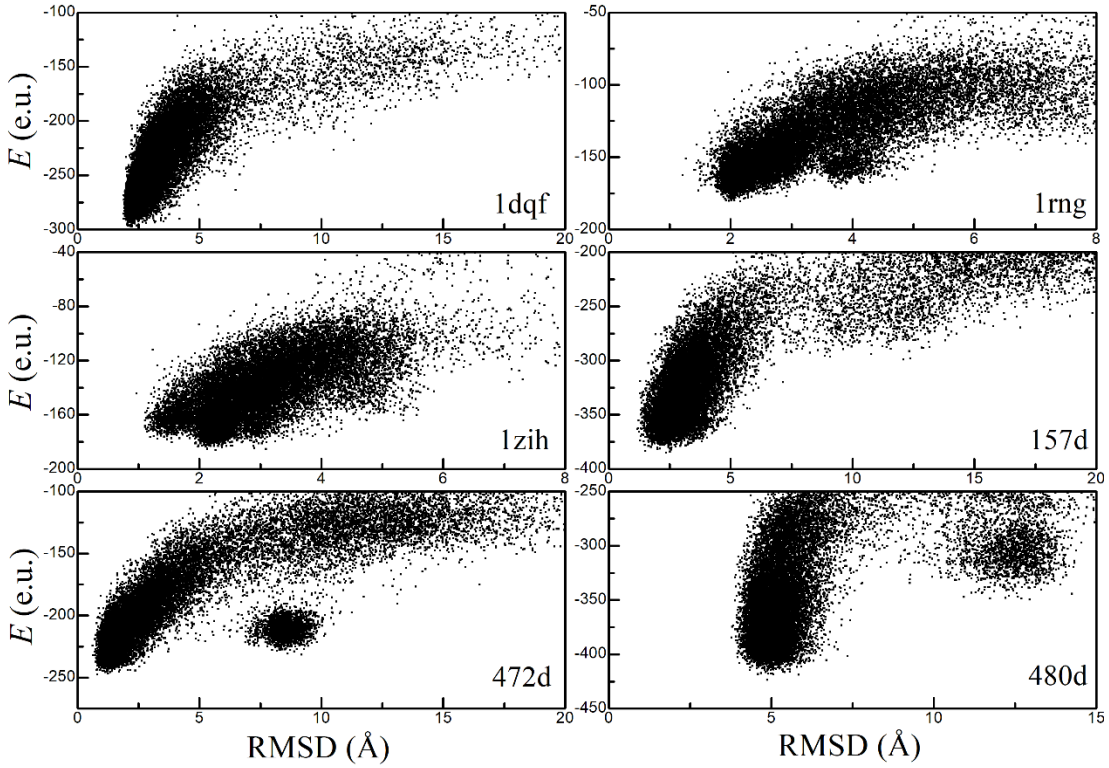

Figure S4 Plots of RMSD (in Å) with total potential energies *E* (in unit e.u.) during *de novo* folding of some typical small RNA molecules with simple topologies by the IsRNA model.

Table S1 Properties of ten types of coarse-grained beads in the IsRNA/IsRNA1 model.

CG bead	Mass (amu)	Diameter (Å)	Grouped heavy-atom
P	94.97	4.0	P, OP1, OP2, O5', O3'
S	92.05	3.8	C5', C4', O4', C3', C2', O2', C1'
R_C	78.05	3.2	N9, C8, N7, C5, C4, N3
A_C	26.02	3.0	C6, N6
A_N	26.02	3.0	C2, N1
G_O	28.01	3.0	C6, O6
G_N	40.03	3.0	C2, N2, N1
Y_C	38.03	3.4	N1, C5, C6
C_N	68.04	2.6	C2, O2, N3, C4, N4
U_O	70.03	2.6	C2, O2, N3, C4, O4

Table S2 PDB ids of 121 structures in the simulated dataset used to parameterize the IsRNA1 coarse-grained force field.

Simulated dataset											
157d	1a51	1c2x	1csl	1duq	1e95	1et4	1f5g	1f9l	1fqz	1fyp	1hwq
1k2g	1kod	1kp7	1kpz	1kxk	1l1w	1l2x	1m5l	1mis	1mnx	1mwg	1my9
1n8x	1nbk	1nlc	1ntb	1osw	1q8n	1q96	1rau	1raw	1rmn	1scl	1tjz
1uuu	1xv0	1yfv	1ylg	1ymo	1ync	1z43	1zo3	2a43	2au4	2d17	2d1a
2d1b	2euy	2gdi	2gm0	2hem	2hoj	2irn	2jlt	2jwv	2k4c	2kbp	2khy
2krl	2kzl	2l5z	2lub	2m18	2m8k	2mis	2miy	2mtj	2mxk	2oiu	2quw
2qwy	2rn1	2rqj	2tob	2xeb	2zy6	3a3a	3b31	3d2v	3e5c	3f4g	3fu2
3gx5	3ibk	3izd	3mij	3npn	3owz	3p22	3q3z	3q50	3r1c	3sj2	3sux
3t4b	480d	4aob	4b5r	4e5c	4enb	4frn	4jf2	4jrc	4k27	4k31	4l81
4lvz	4lx6	4mgm	4p5j	4plx	4pqv	4qjh	4qk8	4qk9	4qlm	4tzx	4wfl
5a17											

Table S3 Energy parameters for 18 types of bond stretching energies $E_{bond}(b)$ in the IsRNA1 force field, in which the last 6 types are for base pairing interactions. The unit for potential energy in the IsRNA1 model is e.u.

Type	K_b (e.u./Å^2)	b_0 (Å)	H_b (e.u.*Å)	b_1 (Å)	σ_b (Å)
P-S	4.249	3.577	-0.618	3.913	0.086
S-R_C	3.816	4.216	-1.667	4.477	0.226
S-Y_C	6.965	3.990	-0.499	3.820	0.145
R_C-A_C	271.887	2.967	-0.079	2.981	0.016
R_C-A_N	187.023	2.867	-0.125	2.881	0.022
R_C-G_O	128.131	2.998	-0.106	3.026	0.021
R_C-G_N	230.498	3.278	-0.068	3.273	0.015
A_C-A_N	81.904	2.333	-0.069	2.310	0.012
G_O-G_N	67.744	3.018	-0.087	2.998	0.018
Y_C-C_N	277.956	2.027	-0.114	2.048	0.020
Y_C-U_O	268.397	2.033	-0.105	2.045	0.020
S-P	3.766	3.557	-0.532	3.884	0.098
G_N-C_N	4.081	3.460	-0.312	3.487	0.169
G_O-C_N	0.818	4.350	-5.031	3.722	0.604
A_N-U_O	2.414	3.609	-1.825	3.280	0.345
A_C-U_O	1.718	3.922	-1.003	3.839	0.267
G_N-U_O	0.590	4.203	-1.435	4.633	0.336
G_O-U_O	3.008	3.530	-0.457	3.322	0.164

Table S4 Energy parameters for 18 types of bond angle potentials $E_{angle}(\theta)$ in the IsRNA1 force field, in which the last 6 types are for base pairing interactions. The unit for potential energy in the IsRNA1 model is e.u.

Type	K_a (e.u./radian^2)	θ_0 (degree)	H_a (e.u.*radian)	θ_1 (degree)	σ_a (radian)
P-S-P	2.226	117.51	-1.126	92.53	0.268
P-S-R_C	2.158	110.12	-0.867	87.15	0.185
P-S-Y_C	2.189	103.93	-1.263	74.77	0.195
S-P-S	2.220	107.20	-0.512	104.56	0.131
S-R_C-A_C	9.975	166.04	-0.158	165.70	0.064
S-R_C-A_N	12.725	116.71	4.011	120.49	0.273
S-R_C-G_O	6.595	168.97	-0.257	164.10	0.085
S-R_C-G_N	22.924	109.03	7.916	110.35	0.303
S-Y_C-C_N	1.362	109.15	-1.151	133.90	0.118
S-Y_C-U_O	17.112	99.12	-7.164	143.41	0.244
R_C-S-P	6.427	108.69	-0.666	91.85	0.130
Y_C-S-P	4.569	115.45	-0.630	94.31	0.128
G_N-C_N-Y_C	6.897	162.89	-0.202	157.51	0.109
G_O-C_N-Y_C	4.593	169.43	-0.100	152.41	0.083
A_N-U_O-Y_C	5.523	174.87	-0.250	167.42	0.124
A_C-U_O-Y_C	5.231	173.90	-0.086	155.11	0.071
G_N-U_O-Y_C	1.150	132.24	-0.455	128.97	0.113
G_O-U_O-Y_C	4.642	179.97	-0.025	166.22	0.037

Table S5 Energy parameters for 26 types of torsional angle energies $E_{torsion}(\phi)$ in the IsRNA1 force field, in which the last 6 types are for base pairing interactions. The unit for potential energy in the IsRNA1 model is e.u.

Type	K_1 (e.u.)	ϕ_1 (degree)	K_2 (e.u.)	ϕ_2 (degree)	K_3 (e.u.)	ϕ_3 (degree)	K_4 (e.u.)	ϕ_4 (degree)
P-S-P-S	0.636	39.71	0.279	-79.13	0.147	77.41	0.095	28.71
P-S-R_C-A_C	1.695	-161.46	0.347	-12.83	-0.033	-99.87	0.282	-123.99
P-S-R_C-A_N	0.849	16.73	0.383	-86.12	0.072	-127.37	0.274	141.06
P-S-R_C-G_O	-1.624	11.97	0.320	-37.70	0.135	-59.93	0.176	-86.86
P-S-R_C-G_N	0.502	7.62	-0.584	108.86	0.106	-88.58	-0.252	-6.13
P-S-Y_C-C_N	1.189	-3.50	0.557	-145.82	0.354	-74.54	-0.010	108.35
P-S-Y_C-U_O	-1.108	-164.15	-0.387	8.08	0.150	-14.44	0.239	82.22
S-P-S-P	0.563	-19.08	-0.351	5.04	0.186	-56.32	-0.087	-5.21
S-P-S-R_C	-0.613	20.40	0.231	-105.77	0.149	3.21	0.036	-99.64
S-P-S-Y_C	-0.551	9.51	0.310	-111.61	0.257	22.92	0.100	80.33
S-R_C-A_C-A_N	-0.483	-11.34	0.228	55.86	-0.290	100.38	0.196	-94.42
S-R_C-G_O-G_N	-0.495	-2.86	0.251	78.84	-0.237	91.73	0.252	-114.88
R_C-S-P-S	0.835	134.87	-0.075	49.96	0.275	39.30	0.097	30.83
A_C-R_C-S-P	0.742	26.41	0.306	25.21	0.350	-25.73	0.272	-11.34
A_N-R_C-S-P	-0.870	56.67	-0.633	131.49	0.109	51.57	0.109	64.92
G_O-R_C-S-P	0.655	36.67	0.249	64.23	0.384	-32.09	0.274	-4.47
G_N-R_C-S-P	0.750	-145.19	0.929	-39.08	0.190	29.05	-0.134	-81.13
Y_C-S-P-S	0.613	145.30	0.219	-55.23	0.380	22.98	0.056	-99.01
C_N-Y_C-S-P	0.719	-100.55	-0.360	68.24	0.319	65.66	0.058	24.92
U_O-Y_C-S-P	0.848	-101.99	0.488	-127.77	0.282	53.57	-0.050	50.53
R_C-G_O-G_N-C_N	-0.783	-12.03	0.137	-117.40	0.621	1.78	-0.826	-0.17
G_N-G_O-C_N-Y_C	0.218	-118.55	-0.199	-95.23	0.104	-40.39	-0.128	10.37

R_C-A_C-A_N-U_O	-0.674	-6.70	0.194	175.55	0.560	0.69	-0.726	0.23
A_C-A_N-U_O-Y_C	0.426	159.40	-0.111	10.03	0.133	35.07	-0.080	-49.79
R_C-G_O-G_N-U_O	-0.666	-3.38	0.041	-64.11	0.543	1.38	-0.793	-0.11
G_N-G_O-U_O-Y_C	0.119	135.22	0.373	165.99	-0.186	18.33	-0.036	6.25

Table S6 Energy parameters for 41 types of pairwise interactions $E_{pair}(r)$ in the IsRNA1 force field. The unit for potential energy in the IsRNA1 model is e.u.

Type	D_0	α	r_0	H_1	r_1	σ_1	H_2	r_2	σ_2	r_{cut}
P--P	0.034	0.369	9.061	-0.431	5.903	0.351	0.700	6.451	1.197	9.4
P--A_N	0.176	2.148	4.140	0.241	7.774	0.324	0.363	10.461	0.771	12.6
P--G_N	0.095	0.298	8.972	-0.310	6.090	0.472	0.489	10.503	0.935	12.5
P--C_N	0.067	0.326	8.095	0.183	7.765	0.589	0.363	11.200	0.964	13.5
S--S	0.044	0.531	6.401	0.513	5.382	0.549	0.055	7.218	0.132	7.7
S--R_C	0.099	0.247	8.593	0.647	7.413	0.414	0.726	10.266	0.902	12.6
S--G_N	0.455	0.507	5.180	-0.970	3.802	0.499	0.178	10.798	0.521	12.2
S--U_O	0.937	1.062	3.857	-0.909	6.951	1.323	0.145	11.291	0.333	12.2
R_C--R_C	2.915	0.975	3.837	8.378	3.904	1.253	-0.375	6.089	0.156	10.0
R_C--A_C	0.182	2.449	3.867	0.580	5.897	0.663	-0.345	6.472	0.348	8.0
R_C--A_N	0.038	0.303	9.076	-1.120	3.448	0.363	0.456	8.963	0.581	10.6
R_C--G_O	0.131	1.401	4.071	0.119	5.937	0.253	-0.246	7.097	0.353	8.2
R_C--G_N	0.336	2.098	3.376	-0.274	6.806	0.404	-0.112	10.380	0.305	11.3
R_C--Y_C	0.502	1.699	3.615	0.274	5.908	0.314	0.258	8.913	0.487	10.4
R_C--C_N	0.376	0.648	4.795	0.859	5.128	0.556	0.220	6.356	0.227	11.0
A_C--A_C	1.126	0.739	3.990	3.100	4.484	1.008	0.409	7.725	0.565	10.2
A_C--A_N	0.192	2.100	3.569	0.537	5.460	0.572	0.155	9.018	0.348	10.0
A_C--G_O	0.197	5.134	3.359	-0.239	5.739	0.128	-0.297	7.374	0.527	8.9
A_C--G_N	0.510	0.486	4.920	0.882	6.002	0.716	0.381	9.172	0.659	10.7
A_C--Y_C	0.818	0.728	4.039	3.592	5.215	1.303	-0.529	6.564	0.386	8.8
A_C--C_N	0.630	0.512	4.556	1.619	5.321	0.880	0.406	8.644	0.652	10.3
A_C--U_O	0.561	1.835	3.357	-0.451	6.590	0.529	-0.232	9.784	0.567	11.6
A_N--A_N	0.103	1.278	4.181	2.397	6.103	1.513	-1.482	6.536	0.530	10.3
A_N--G_O	0.573	2.094	3.431	0.681	4.869	0.898	-0.191	6.835	0.257	7.7
A_N--G_N	0.437	0.625	4.672	2.247	5.482	0.412	-1.972	5.466	0.299	10.0
A_N--Y_C	0.247	1.519	3.708	0.334	6.036	0.466	-0.547	6.788	0.955	9.5
A_N--C_N	0.813	0.632	4.356	2.493	5.281	1.110	0.420	8.681	0.579	10.2
A_N--U_O	0.524	1.459	3.617	0.911	5.457	0.664	-0.297	7.940	0.506	9.6
G_O--G_O	0.188	1.013	3.930	0.527	5.176	0.716	0.222	8.439	0.618	10.0
G_O--Y_C	0.785	0.670	4.133	2.730	4.778	1.195	0.390	8.686	0.598	10.5
G_O--C_N	0.012	3.637	3.534	-0.416	5.679	0.195	-0.321	7.446	0.409	8.6
G_O--U_O	0.757	1.829	3.385	-0.171	6.473	0.221	0.210	8.273	0.410	9.5
G_N--Y_C	0.508	1.636	3.657	0.309	4.553	0.268	-0.250	7.332	0.313	8.2
G_N--C_N	0.188	2.265	3.543	1.254	4.984	0.826	-0.192	7.096	0.236	7.8
G_N--U_O	0.358	1.096	4.016	0.263	4.795	0.301	-0.402	5.422	0.165	7.7
Y_C--Y_C	0.046	1.051	4.335	0.684	4.531	0.393	-0.477	6.721	0.736	9.0
Y_C--C_N	0.133	0.486	7.097	-2.494	3.531	0.401	0.435	8.614	0.599	10.1
Y_C--U_O	0.176	0.456	5.839	1.619	5.819	0.823	0.276	9.463	0.484	11.9
C_N--C_N	0.576	0.636	4.701	2.278	5.558	0.692	0.479	8.904	0.435	10.3
C_N--U_O	0.359	0.600	4.529	0.967	5.390	0.539	0.352	9.046	0.440	10.3
U_O--U_O	1.158	0.485	4.686	3.559	5.431	0.930	0.844	9.227	0.745	11.2

Table S7 Summary of simulation details for each study in this work. Simulation technique: cMD is conventional Molecular Dynamics, SA MD is simulated annealing Molecular Dynamics, and REMD is replica-exchange Molecular Dynamics.

Studies (Number of cases)	Length (nts)	Simulation length, technique, and temperature	Initial structure	Obtainment of conformation ensemble and Output	CPU time (hours)
Parameterization of energy functions in IsRNA1 (121)	12~120	60ns cMD at T=300K	Native structures	First 50ns for relaxation, collect structure snapshots every 10ps in last 10 ns for statistics	0.8~7
De novo folding of small RNA in IsRNA (15)	12~36	50ns SA MD with T=1000K to 350K, then 500ns SA MD with T=350K to 200K, final 500ns cMD at T=200K	Coil states by unfolded the native structures at T=1000K with 10ns	First 550ns for folding of structure, take snapshots every 100ps in last 500ns. the top 10% lowest potential structures of collected snapshots were clustered, the centroid structure of the largest cluster was chosen as prediction	21~68
3D structure predictions with 2D structures as constraints by IsRNA (65)	22~78	300ns REMD with 8 replicas and temperature uniformly range from T=200K to 375K	Coil states by unfolded the native structures at T=1000K with 10ns	First 200ns for folding of structure, collect structures for the conformational ensemble every 100ps in the last 100ns for each replica. Top 10% lowest potential structures make the conformation ensemble for clustering, the centroid structure of the largest cluster was chosen as the prediction	72~384
Refined 3D structure predictions by IsRNA1 (65)	22~78	10ns REMD with 11 replicas and temperature uniformly range from T=150K to 400K	Predicted structures given by IsRNA	First 50ns for relaxation, collect snapshots every 100ps in the last 50ns for each replica, top 10% lowest potential snapshots give the conformation ensemble for clustering process, the centroid of the largest cluster was chosen as the prediction	39~154

Large-scale benchmark on 3D structure predictions by IsRNA1 (130)	40~161	A predefined 30ns REMD simulation with 10 replicas and temperature T=225K to 450K to refold the tertiary base pairs; 50 ns REMD with 10 replicas and temperature uniformly range from T=200K to 420K	Top 3 predictions generated by Vfold3D/VfoldLA algorithm	First 25ns for relaxation, collect snapshots every 50ps in the last 25ns for each replica, top 10% lowest potential snapshots give the conformation ensemble for clustering process, the centroid structures of the top largest clusters were chosen as predictions	Top 3 prediction	85~200
3D structure predictions on RNA Puzzle by IsRNA1 (21)	41~188				Top 5 predictions	95~645
Structure filtration for predictions in RNA Puzzle by IsRNA1 (21)	41~188	0.5ns cMD at T=300K with integration timestep Δt =0.5fs, 5 duplicates per prediction	All predictions in RNA Puzzle	Energy minimization firstly, collect structure snapshots every 0.5ps, snapshots within RMSD = 5Å from the initial structure are used for energy average		0.3~1.5

Table S8 Tertiary folding of medium size RNA molecules with the assistance of 2D structures by IsRNA-IsRNA1 pipeline and comparison with SimRNA. The top 1 candidate structure is chosen as the predicted structure and the RMSD (in Å) is calculated over all heavy-atoms. For SimRNA, the data are collected from the original paper if available, otherwise, the top 1 prediction is given by the SimRNAweb (https://genesilico.pl/SimRNAweb/) at the default parameters is chosen. The best predictions with lowest RMSD over three models (SimRNA, IsRNA, and IsRAN1) are **boldfaced**.

PDB id	Length	Topology	SimRNA	IsRNA	IsRNA1
2g1w	22	pseudoknot	**4.64**	5.10	4.67
1k2g	22	stem-loop	**3.38**	5.68	4.51
1nc0	24	stem-loop	5.46	**5.27**	5.62
2a43	26	pseudoknot	**3.65**	4.33	4.83
2rp1	27	pseudoknot	4.13	3.78	**3.40**
1l2x	27	pseudoknot	**3.90**	4.66	4.31
1fqz	27	stem-loop	5.27	4.91	**4.88**
1ysv	27	stem-loop	**1.40**	1.51	2.27
1q9a	27	stem-loop	**4.49**	4.60	5.13
1kpz	28	pseudoknot	**4.01**	4.24	4.72
2ap5	28	pseudoknot	**4.14**	6.22	4.95
1yg4	28	pseudoknot	**4.27**	6.14	5.07
2ldz	30	stem-loop	4.86	3.86	**3.42**
1kp7	30	stem-loop	5.86	**4.73**	4.82
1na2	30	stem-loop	**3.17**	4.50	4.07
1rfr	30	stem-loop	**2.63**	2.64	2.87
2n6q	31	pseudoknot	**3.03**	3.61	3.15
1ylg	31	stem-loop	3.23	**2.74**	3.92
1kpd	32	pseudoknot	6.89	6.34	**6.05**
1kaj	32	pseudoknot	9.05	**8.86**	9.04
2kyd	32	stem-loop	2.22	2.01	**1.71**
2jxv	33	stem-loop	3.50	**2.57**	2.89
1rnk	34	pseudoknot	6.59	6.11	**5.31**
2kpv	34	stem-loop	2.85	2.90	**2.84**
1p5n	34	stem-loop	4.45	**4.20**	4.29
1nbk	34	stem-loop	**4.50**	6.27	4.66
2jtp	34	stem-loop	3.54	**3.38**	3.94
1e95	36	pseudoknot	**3.29**	4.64	3.55
2tpk	36	pseudoknot	**3.04**	3.28	3.57
3gm7	36	stem-loop	2.19	5.97	**2.06**
1b36	38	stem-loop	8.76	**7.96**	8.17
2a9l	38	stem-loop	11.55	10.80	**10.54**
4e5c	38	stem-loop	2.82	3.19	**2.41**
3szx	38	stem-loop	2.18	3.39	**1.93**
1txs	38	stem-loop	**3.14**	4.31	3.83
299d	41	3-way junction	6.96	**6.83**	10.73

1zc5	41	stem-loop	2.85	2.50	**2.01**
2l2j	42	stem-loop	3.54	**2.52**	2.57
2fey	43	stem-loop	7.18	6.04	**4.74**
1cq5	43	stem-loop	**3.97**	4.66	4.54
1a60	44	pseudoknot	6.76	**5.85**	6.23
1z2j	45	stem-loop	4.11	4.13	**3.77**
1xjr	46	stem-loop	**9.38**	10.02	10.30
1ymo	47	triplex	**5.74**	6.54	5.83
2p89	48	3-way junction	9.65	**6.58**	7.27
2ke6	48	stem-loop	4.42	**3.20**	4.68
2kur	48	stem-loop	4.13	**3.56**	4.73
2m8k	48	triplex	4.54	4.45	**3.80**
2pcw	49	3-way junction	14.87	11.46	**7.46**
1rmn	49	3-way junction	**12.73**	15.54	14.69
2lu0	49	3-way junction	**7.40**	7.46	8.46
3e5f	52	3-way junction	12.47	13.68	**10.96**
2mhi	53	3-way junction	15.66	17.25	**14.85**
4kz2	54	3-way junction	10.85	16.34	**9.07**
1p5m	55	stem-loop	**4.69**	7.18	6.66
3r4f	66	3-way junction	15.35	17.41	**7.89**
3slm	66	tertiary	**6.24**	8.18	7.17
2qus	68	3-way junction	**5.68**	7.72	9.09
2gm0	70	stem-loop	6.97	**5.84**	6.32
2n8v	70	tertiary	14.03	10.32	**7.16**
1y26	71	tertiary	10.07	**6.74**	7.06
3l0u	73	tRNA	12.97	**12.48**	14.50
1zo3	76	tRNA	17.90	**11.23**	14.19
1p5p	77	stem-loop	12.12	8.82	**8.16**
2d1a	78	stem-loop	10.51	**5.40**	7.16
Average (Small size group)			4.35±2.11	4.70±1.91	4.38±1.88
Average (Large size group)			8.79±4.27	8.33±4.36	7.74±3.50
Average (Stem-loop group)			4.86±2.73	4.74±2.21	4.60±2.23
Average (Pseudoknot group)			4.81±1.80	5.22±1.48	4.92±1.51
Average (Complicated group)			10.77±4.13	10.60±4.23	9.42±3.37
Average (Total)			6.40±3.95	6.38±3.73	5.93±3.21

Table S11 Blind screening of good predictions in RNA-Puzzles challenges by the IsRNA1 protocol without knowledge of native structures given in prior: challenge # and PDB id, RNA molecule size (column "Length"), number of submitted models by various groups (column "Number of models"), the range of RMSDs over all the submitted models (column "Predicted RMSD"), the range of Interaction Network Fidelity for canonical base pairing interaction INF_{WC} over all the predictions (column "Predicted INF_{wc}"), the Pearson correlation coefficient between potential energies *E* evaluated by IsRNA1 protocol and RMSDs (column "*R*(*E*, RMSD)"), "), the Pearson correlation coefficient between potential energies *E* and values of INF_{WC} (column "*R*(*E*, INF_{wc})"), sensitivity TPR for the fraction of actual good predictions caught by the filtration protocol (column "TPR"), precision PPV for the fraction of corrected predictions over the guessed good predictions by the filtration protocol (column "PPV"), fraction S that is classified as actual good predictions over all the submissions (column "S"), and fraction P that is filtered as good predictions over all the submissions (column "P"). Values in *parentheses* in columns "TPR", "PPV", "S", and "P" are the results that adopt 2D structure filtration before screening by removing the predictions with $INF_{wc} < 0.8$ before the blind 3D structure screening by the present IsRNA1 protocol.

Puzzle (PDB id)	Length (nt)	Number of models	Predicted RMSD (Å)	Predicted INF_{wc}	*R*(*E*, RMSD)	*R*(*E*, INF_{wc})	TPR	PPV	S	P
1 (3mei)	46	14	3.4~7.2	0.73~0.97	0.14	-0.82	0.57 (0.62)	1.0 (1.0)	1.0 (1.0)	0.57 (0.62)
3 (3owz)	84	12	7.7~24.7	0.39~0.94	0.60	-0.86	0.80 (0.80)	0.67 (0.67)	0.42 (0.50)	0.50 (0.60)
5 (4p8z)	188	25	9.0~36.2	0.66~0.94	0.72	-0.67	1.0 (1.0)	0.18 (0.18)	0.08 (0.10)	0.44 (0.52)
6 (4gxy)	168	34	11.6~37.4	0.64~0.90	-0.01	-0.23	0.57 (0.57)	0.44 (0.57)	0.21 (0.28)	0.26 (0.28)
7 (4r4p)	185	52	20.3~60.6	0.45~0.93	0.43	-0.80	0.60 (0.60)	0.60 (0.60)	0.19 (0.29)	0.19 (0.29)
8 (4l81)	96	42	4.8~31.3	0.30~0.98	0.60	-0.63	0.75 (0.75)	0.20 (0.20)	0.10 (0.11)	0.36 (0.42)
9 (5kpy)	71	32	6.1~25.4	0.73~0.92	-0.18	0.03	0.28 (0.29)	0.64 (0.78)	0.78 (0.86)	0.34 (0.32)
10 tBox (4lck)	96	26	6.0~18.5	0.78~0.94	0.37	-0.42	1.0 (1.0)	0.40 (0.43)	0.23 (0.25)	0.58 (0.58)
11 X-ray (5lys)	57	54	5.0~16.1	0.60~1.0	0.18	-0.34	0.56 (0.68)	0.96 (0.88)	0.83 (0.81)	0.48 (0.62)
12 (4qlm)	125	51	10.1~36.6	0.20~0.89	0.80	-0.81	0.40 (0.56)	0.77 (0.67)	0.49 (0.51)	0.25 (0.43)
13	71	55	5.4~31.0	0.0~0.95	0.24	-0.56	0.20 (0.20)	0.18 (0.22)	0.18 (0.28)	0.20 (0.25)
14 Bound (5ddp)	61	64	5.1~17.5	0.78~0.94	-0.05	0.19	0.08 (0.08)	0.10 (0.10)	0.20 (0.21)	0.16 (0.16)
14 Free (5ddo)	61	52	6.9~23.1	0.75~0.91	0.34	0.28	0.40 (0.32)	0.67 (0.70)	0.48 (0.51)	0.29 (0.23)
15 (5di4)	68	64	7.1~24.8	0.05~0.85	0.55	-0.21	0.56 (0.63)	0.38 (0.36)	0.28 (0.47)	0.41 (0.82)
17 (5k7c)	62	68	7.2~21.8	0.10~0.86	0.22	-0.62	0.54 (0.60)	0.33 (0.60)	0.19 (0.48)	0.31 (0.48)
18 (5tpy)	71	57	3.2~24.9	0.29~0.98	-0.18	-0.04	0.0 (0.91)	0.0 (0.77)	0.19 (0.65)	0.35 (0.76)
19 (5t5a)	62	54	5.5~22.4	0.27~0.98	0.20	-0.16	0.38 (1.0)	0.33 (0.44)	0.15 (0.31)	0.17 (0.69)
20 (5y85)	68	55	4.7~25.4	0.0~1.0	0.27	-0.48	0.50 (0.50)	0.19 (0.21)	0.18 (0.29)	0.47 (0.71)

21 (5nz6)	41	46	3.8~18.1	0.39~0.91	0.33	-0.30	0.44 (0.44)	0.89 (1.0)	0.39 (0.82)	0.20 (0.36)
Average					0.29	-0.39	0.51 (0.61)	0.47 (0.55)	0.35 (0.46)	0.34 (0.48)