SUPPLEMENTARY DATA

Site-Directed Spin Labeling of Long Natural RNAs Exemplified with Hepatitis C Virus RNA Internal Ribosome Entry Site

Elena S. Babaylova,[1,5] Alexey A. Malygin,[1,5] Alexander A. Lomzov,[1,5] Dmitrii V. Pyshnyi,[1,5] Maxim Yulikov,[2] Gunnar Jeschke,[2] Olesya A. Krumkacheva,[3,5] Matvey V. Fedin,[3,5] Galina G. Karpova,[1,5] Elena G. Bagryanskaya[4,5]

[1]Institute of Chemical Biology and Fundamental Medicine SB RAS, Lavrentiev ave. 8, Novosibirsk, 630090, Russia

[2]Laboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

[3]International Tomography Center SB RAS, Institutskaya str. 3a, Novosibirsk, 630090, Russia

[4]N. N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS, Lavrentiev ave. 9, Novosibirsk, 630090, Russia

[5]Novosibirsk State University, Pirogova str. 2, Novosibirsk, 630090, Russia

I. Site-directed introduction of amino linkers into HCV RNA IRES

After site-specific alkylation of HCV RNA IRES by oligodeoxyribonucleotide derivatives, covalent adducts formed were subjected to purification by 8% PAGE (illustrated in Figure S1).

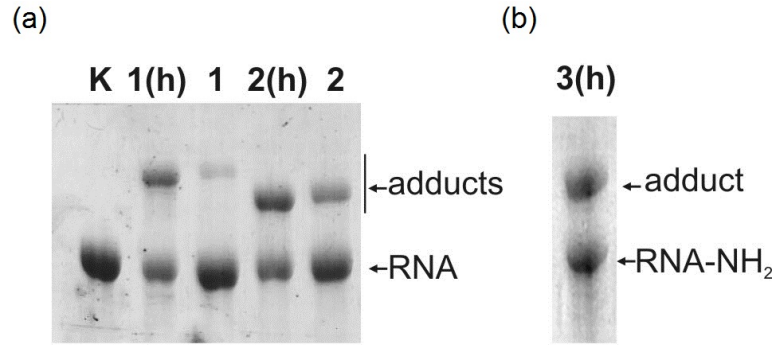

Figure S1. Purification of covalent adducts of HCV RNA IRES with oligodeoxyribonucleotide derivatives by 8% denaturing PAGE. (a) Separation of covalent adducts formed after complementary-addressed alkylation of HCV RNA IRES by DNA-$NHCH_2RCl$ and unmodified RNA (lanes 1 and 2, the alkylation of RNA by DNA1-$NHCH_2RCl$ and DNA2-$NHCH_2RCl$, respectively, without helper; lanes 1(h) and 2(h), the alkylation of RNA by the same DNA derivatives in the presence of helper; lane k corresponds to unmodified HCV RNA IRES. (b) Separation of double covalent adduct of HCV RNA IRES with derivatives of DNA1 and DNA2 (adduct) and singly- or non-alkylated RNAs, which migrate as one band (RNA-NH_2), 3(h) designates that the alkylation of RNA by both DNA1-$NHCH_2RCl$ and DNA2-$NHCH_2RCl$ occurred in the presence of helper.

II. Identification of the modified sites in HCV RNA IRES

Identification of the modified nucleotides in HCV RNA IRES after complementary-addressed treatment by alkylating oligodeoxyribonucleotide derivative, DNA2-$NHCH_2RCl$, and hydrolysis of phosphoramide bond in the covalent adduct formed was performed by means of reverse transcription (illustrated in Figure S2).

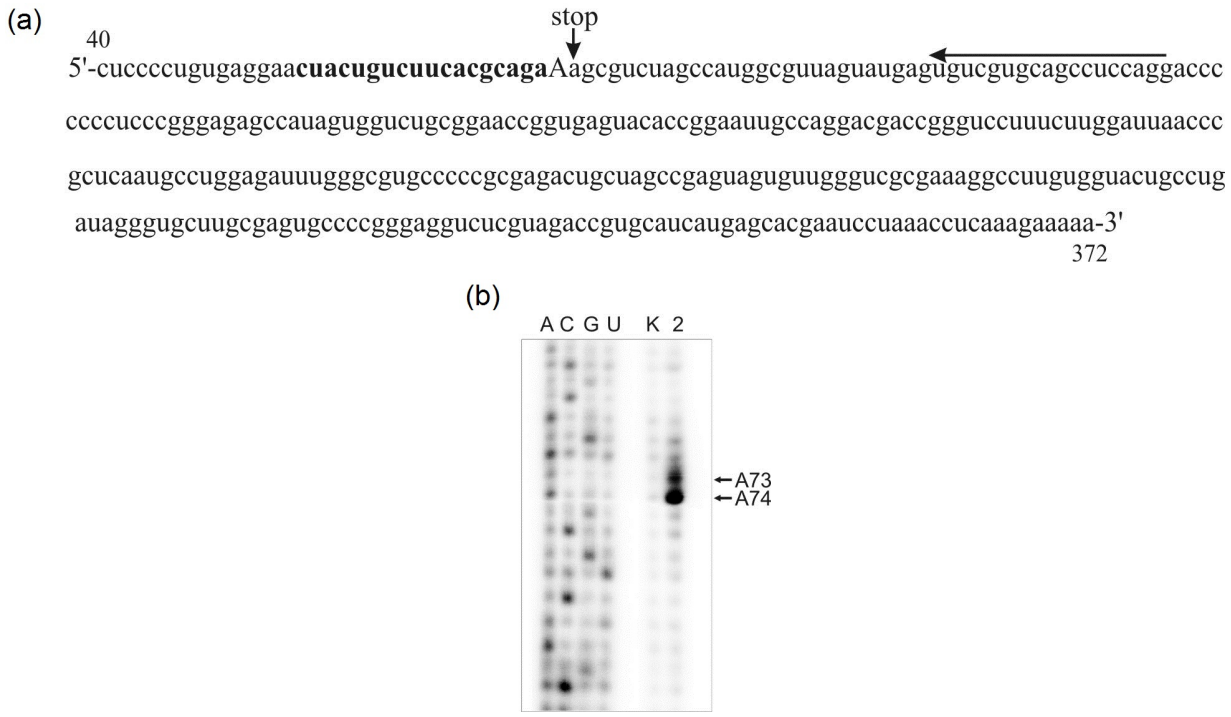

Figure S2. Identification of HCV RNA IRES nucleotides alkylated by DNA2-$NHCH_2RCl$. (a) Schematic representation of the reverse transcription on the modified RNA. On the primary structure of the HCV IRES, the region complementary to the alkylating derivative of oligodeoxyribonucleotide is shown in bold font, and the binding site of primer for reverse transcription is marked (arrow indicates the direction of primer elongation by reverse transcriptase). The designation "stop" indicates a nucleotide, on which reverse transcriptase makes a stop. Modified nucleotide is marked by large letter. (b) Autoradiographs of gels after electrophoretic separation of the products of reverse transcription. A, C, G and U – sequencing lanes. K – lane of unmodified RNA. Lane 3 corresponds to RNA modified with DNA2-$NHCH_2RCl$. The nucleotides, at which the stops of reverse transcriptase occurred, are indicated on the right.

Two stops of reverse transcription observed at A74 and A73 may suggest that modification of HCV RNA IRES occurred at two nucleotides, A73 and A72, respectively. However, N1 atom in A72 is involved in the Watson-Crick base-pairing with the respective atom of the 5'-terminal thymidine bearing the alkylating group in the complementary oligodeoxyribonucleotide and, therefore, it was inaccessible for the modification. Furthermore, there are observations that reverse transcriptase is often prone to some kind of stuttering (doubling of the band on the gel) at modified residues of RNA (see

e.g. (S1,S2,S3)). The reasons for such a phenomenon are not yet clear. Usually, a stop of reverse transcription at the nucleotide 3' to the modified one is the strongest in the stuttering and it is considered as a signal from the modified nucleotide. This is why we believe that in the case of modification of HCV RNA IRES with DNA2-$NHCH_2RCl$ only A73 was modified. Nevertheless, a possibility of modification of A72 cannot be completely excluded. It could happen if the 5'-terminus of DNA2-$NHCH_2RCl$ bound to HCV RNA IRES was partially melted, making N1 atom in A72 accessible for the modification. However, the expected melting temperature of the heteroduplex formed is about 30 ^oC higher than that at which the modification of HCV RNA IRES by DNA2-$NHCH_2RCl$ occurred and, therefore, the extent of melting of the 5'-terminus of DNA2-$NHCH_2RCl$ in the heteroduplex at the ambient temperature would be negligible.

III. Functional assay of doubly spin-labelled HCV RNA IRES

To validate the functional activity of the doubly spin-labeled HCV RNA IRES, the capability of the 3'-[^{32}P]pCp-labelled dslRNA of the 48S pre-initiation complex formation was examined. To assemble this complex, a mammalian cell-free protein synthesis system based on rabbit reticulocyte lysate (RRL) was utilized, and unhydrolyzed GTP analog, GMPPNP, which stalls translation at the stage of this kind complex formation, was used instead of GTP. In the control experiment, unmodified 3'-[^{32}P]pCp-labelled HCV RNA IRES was applied. When the complexes assembled on these RNAs were isolated by ultracentrifugation in sucrose gradient, the radioactive label was found in 40S fractions of the gradient for both dslRNA and unmodified RNA (Figure S3a), indicating the binding of these RNAs to the 40S subunits. In order to prove that these complexes were actually the 48S pre-initiation ones, they were tested for the presence of Met-tRNA$_i^{Met}$ by 3'-[^{32}P]pCp-post-labelling of total RNA isolated from the complexes. The analysis of labelled RNAs by denaturing PAGE showed occurrence of the bands corresponding to tRNA and dslRNA (or HCV RNA IRES) with both complexes, and the intensities of tRNA bands were similar for dslRNA and unmodified HCV RNA IRES samples (Figure S3b). All this implies that functional activity of dslRNA sample remains practically the same as that of unmodified RNA.

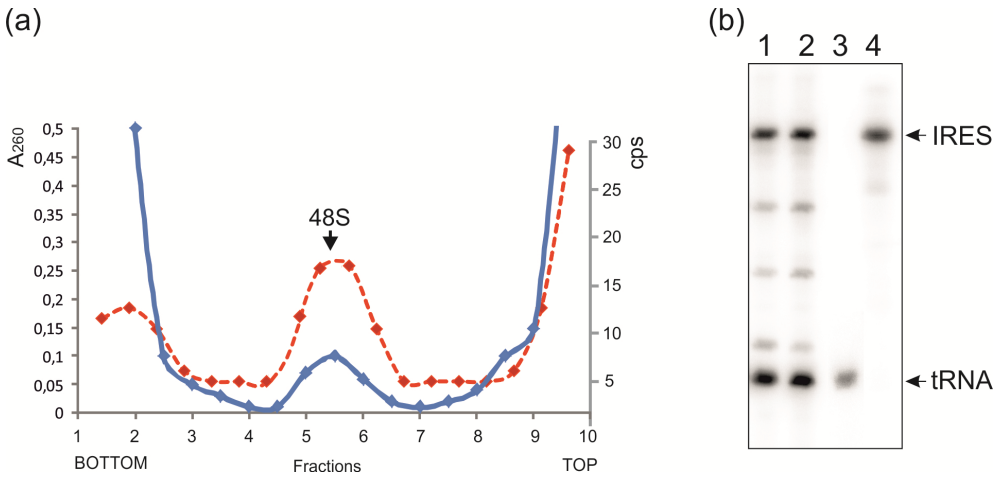

Figure S3. Analysis of functional activity of dslRNA. (a) Typical sucrose density gradient sedimentation profile of ribosomal complex assembled on ^{32}P-labelled dslRNA or unmodified HCV RNA IRES in RRL in the presence of GMPPNP. Blue line – optical density at 260 nm, red line – radioactive label in cps. Position of the 48S pre-initiation complex is indicated by arrow. (b) Denaturing PAGE analysis of post-labelling products of total RNA isolated from 48S pre-initiation complexes purified by sucrose gradient. Lane 1, RNA isolated from the 48S complex formed with dslRNA and incubated under conditions of aminoacyl-tRNA deacylation with the subsequent labelling with [5′-^{32}P] pCp; lane 2, the same but with HCV RNA IRES used instead of dslRNA; lanes 3 and 4, [^{32}P]pCp-labelled tRNA^{Phe} and HCV RNA IRES, respectively, as markers. Positions of dslRNA and HCV RNA IRES are indicated as "IRES" (since electrophoretic mobilities of these RNA species is indistinguishable), and positions of tRNA$_i^{Met}$ and tRNA^{Phe} are indicated as "tRNA" (see the above comment on "IRES").

IV. Details of CW EPR simulations

Table S1 summarizes the parameters used in simulations of the experimental X-band (ν_{mw}≈9.67 GHz) CW EPR data. The room-temperature EPR spectra of dslRNA can only be simulated assuming anisotropic motion of spin label, which is typical for labels attached to biomolecules. Figure S4a clearly demonstrates that the model of isotropic diffusional motion cannot satisfactorily describe the experimental spectrum.

Table S1. The parameters used in simulation of experimental CW EPR spectra shown in Figure 3 (main text) and Figure S4 (Supplementary Data). Γ refers to the Lorentzian peak-to-peak linewidth.

	shown in:	**g-tensor**	**A-tensor / MHz**	**τ_{corr} / ns**	**Γ / mT**	**[$\lambda_{2,0}$ $\lambda_{2,2}$]**
NHS-M	Fig.3a	[2.0091 2.0061 2.0022]	[14 14 107.1]	0.18	0.1194	[0 0]
NHS-M + RNA	Fig.3b					
dslRNA	Fig.S3a		[14 14 107.5]	1.02	0.0415	[0 0]
dslRNA	Fig.S3b					[-0.60 0.00]
dslRNA	Fig.3c, Fig.S3c					[0.53 0.55]

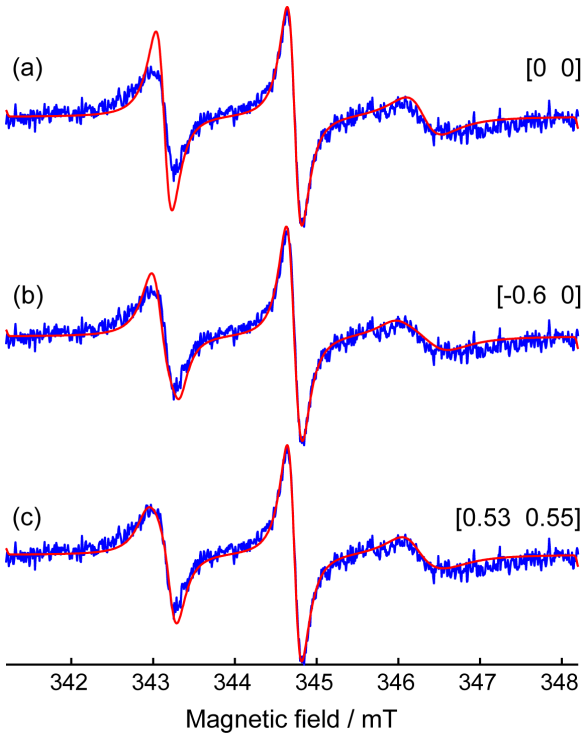

Figure S4. Simulation of room-temperature X-band CW EPR spectrum of dslRNA in different models of nitroxide motion. The numbers in brackets indicate the coefficients $\lambda_{2,0}$ and $\lambda_{2,2}$ of the orienting potential in MOMD approach. Upper trace (a) corresponds to the isotropic motion, middle one (b) to the motion of nitroxide in XY molecular plane, and bottom trace (c) to the motion in XZ plane (X axis is directed along NO and Z axis perpendicular to the radical plane).

At the same time, two models of anisotropic motion (Figure S2b and S2c) provide much better quality of fitting. Based on RMSD value, the model with $[\lambda_{2,0}\ \lambda_{2,2}]=[0.53\ 0.55]$ has been chosen and presented in Figure 3c of the main text. However, within experimental accuracy these two models are hardly distinguishable. Since our main focus here was the verification of spin label attachment, we did not investigate the motion of the attached label in more detail

V. Original DEER data and background correction

Figure S5 shows the original DEER/PELDOR data before the intermolecular background subtraction and the background fit function itself. The background-corrected data are shown in Figure 4a of the main text.

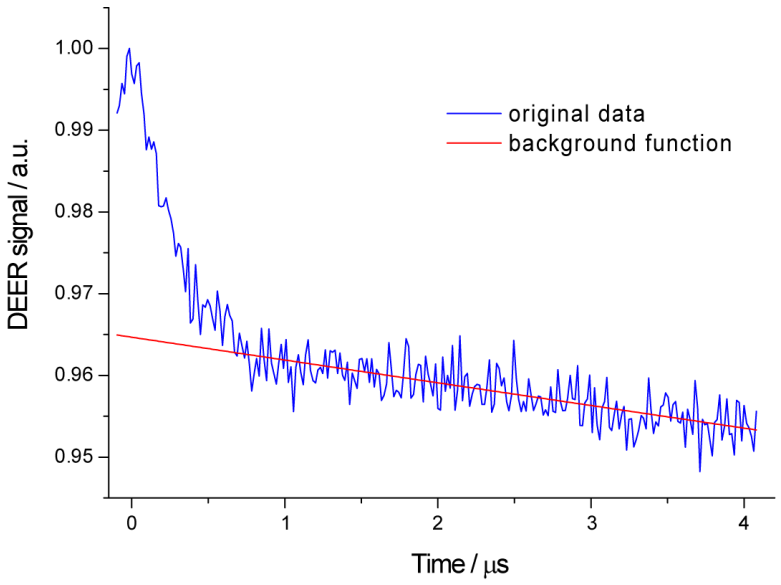

Figure S5. DEER/PELDOR time trace of dslRNA before the intermolecular background subtraction (blue) and the background fit function used (red).

VI. Estimation of spin labeling efficiency

Assume that the overall number of molecules is N. Assume also that spin labeling efficiency/probability is the same for two positions of RNA and is denoted as x (x<=1). Then the probability of double spin-labeling of the same molecule will be x^2, whereas the probability of single spin-labeling will be $x(1-x)$ and the probability of no spin labeling is $(1-x)^2$. For comparison with experimental data we will need the following values expressed using x:

- the number of spins localized in doubly-labeled molecules is $N_2=2x^2N$ (here coefficient 2 comes from two spins per molecule)
- the number of spins localized in singly-labeled molecules is $N_1=2x(1-x)N$ (here coefficient 2 comes from two possibilities to label first or second alkyl group)
- the ratio of spins residing in pairs to the overall amount of spins is $R_2=N_2/(N_1+N_2)=x$. This is reasonable, because if spin labeling efficiency x=1 we find all spins residing in spin pairs, i.e. all RNA molecules will be doubly spin-labeled.
- the ratio of all spins to the overall amount of molecules is $R=(N_2+N_1)/N=2x$. This is also reasonable, because if spin labeling efficiency x=1 we find twice more spins than the molecules.

The first experimental approach employs comparison of the DEER modulation depth in dslRNA with reference samples having 100% of spin pairs (model biradicals), as is briefly described in the Section “Measurement of interspin distances in the spin-labelled domain II of HCV RNA IRES” of the manuscript. For similar experimental settings, it turned out to be that the amount of spin labels in pairs was approximately 16% of total amount of spin labels (obtained as the ratio of ~3.5% modulation depth in dslRNA sample relative to ~20-25% in model biradicals), i.e. R_2=0.16 and labeling efficiency x=0.16.

The second experimental approach uses the total amount of spin labels estimated from room-temperature CW EPR spectrum of dslRNA sample compared with the reference one (solution of free label) with known spin concentration. The spin concentration in dslRNA sample was found to be 12 μM. At the same time, the concentration of IRES RNA in the sample was approximately 25 μM, resulted from initial concentration of 40 μM diluted with buffer solution (10%) and glycerol (40-50%). Thus, the ratio of number of spins to the number of molecules R=12/25~0.5. This means that x=R/2=0.25.

In our opinion, taking into account that the experimental estimations are rather crude, the agreement between two numbers (x=0.16 vs. 0.25) is very reasonable. Therefore, we estimate the labeling efficiency in our experiments as approximately 0.2.

This number is not very high; however, it is still sufficient for the EPR studies, as is demonstrated in our work. One of the ways of increasing of spin-labelling efficiency in future could be repeated treatment of RNA with fresh portion of N-hydroxysuccinimide ester of spin label including isolation of the RNA sample after each round of the reaction (though it should be taken into account that this must inevitably lead to a decrease in the yield of the final sample).

Thus, we have estimated efficiency of spin labeling as 0.2, spin concentration in the sample used as 12 μM and absolute concentration of doubly spin-labeled RNA as 0.2×12 μM = 2.4 μM.

VII. Details of MD simulations

The MD simulation included six steps: (1) adding 31 sodium ions to neutralize the net charge of the system and adding a water shell (ions, TIP3P water, cuboid periodic box, edge 12 Å); (2) minimization of all systems with RNA fully fixed (max. 10000 steps, constant restraint force of 500 cal/mol/Å^2); (3) heating of all the system with fixed RNA for 2.5 ns with time step of 0.0005 ps (500 cal/mol/Å^2); (4) equilibration of the system density at the constant volume for 500 ps; (5) equilibration of the system density at the constant pressure (1 bar) for 500 ps; (6) MD productive trajectory simulation, 10 series of 100 ns trajectories with different initial atoms' speed. At steps (4)-(6) harmonic potential for heavy atoms in MD simulation (1 kcal/mol/Å^2) was applied for the following nucleotides G67-G71, G75-A85, C88-U91 and U94-C100 to prevent changes in the unmodified part of RNA. Coordinates of each atom of the system (snapshots) were saved every 1 ps. The interspin distance was measured between oxygen atoms of nitroxide labels. The distances were measured using cpptraj program (AmberTools 12). The analysis of the MD trajectories and molecular graphics was performed using the UCSF Chimera package.

The typical time evolutions of interspin distances are presented in Figure S6.

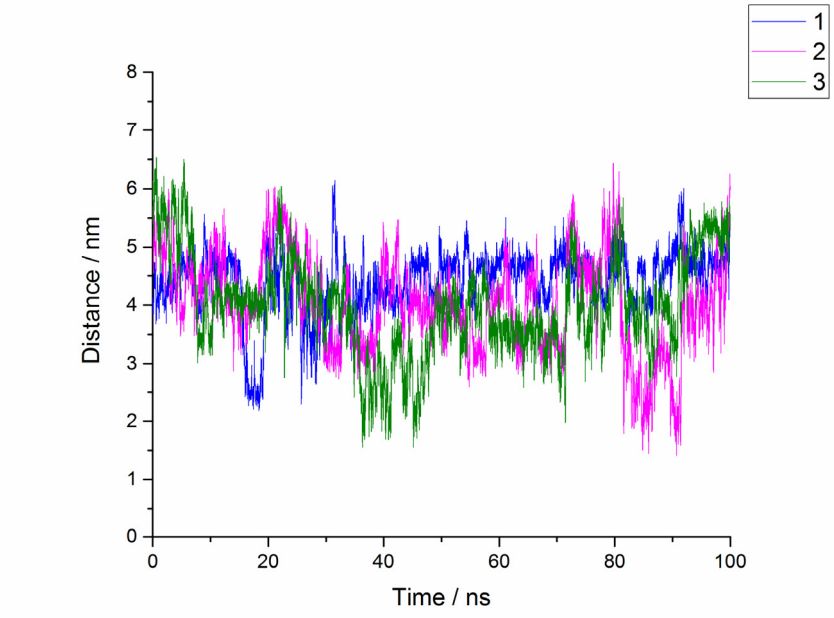

Figure S6. Typical time dependence of interspin distances obtained for independent molecular dynamics trajectories of dslRNA with different initial atoms' speed.

Figure S7 compares the calculated distance distribution between spin labels ("spin-spin distance") with calculated distance distribution between corresponding label attachment sites ("site-site

distance"), i.e. nitrogen atoms of the bases to which the aliphatic linkers are attached (C83 (N3) and A73 (N1)). Obviously, site-site distance is shorter and the distribution width (σ≈0.3 nm) characterizes the intrinsic mobility of the corresponding region of HCV RNA IRES in our simulations. Spin-spin distance distribution is noticeably broader, and is characterized by σ≈0.8 nm. Therefore, we can estimate the contribution to the overall distribution width from intrinsic mobility of spin label as σ≈0.5 nm. This value nicely agrees with our previously published data obtained on short model 10-mer RNA duplex using the same spin-labeling strategy.[S4] In that case spin labels were attached to the terminal base pairs of the duplex, and the distances obtained by DEER were $\langle r_{DEER} \rangle$ = 3.48 nm, σ = 0.54 nm. Since the intrinsic mobility of the termini of such short RNA duplex can be neglected, this value of σ = 0.54 nm is a good measure of intrinsic distribution width introduced by the label.

Thus, the present MD calculation and previous experimental data nicely correlate and yield the estimate of label-induced broadening as σ≈0.5 nm. Note that in fact both considered situations should be viewed as "unfortunate" examples. When labels are attached to the termini of RNA duplex, evidently their mobility is potentially higher compared to the midst of the duplex. Similarly, the attachment of one of the labels to C83 allows its virtually unrestricted motion in a huge solid angle, as illustrated in Figure 4c. In this sense, it is well possible that the complementary-addressed spin-labeling into the less "open" sites will lead to the narrower distributions, and therefore the value σ≈0.5 nm should be considered as more or less the worst situation possible with complementary-addressed SDSL.

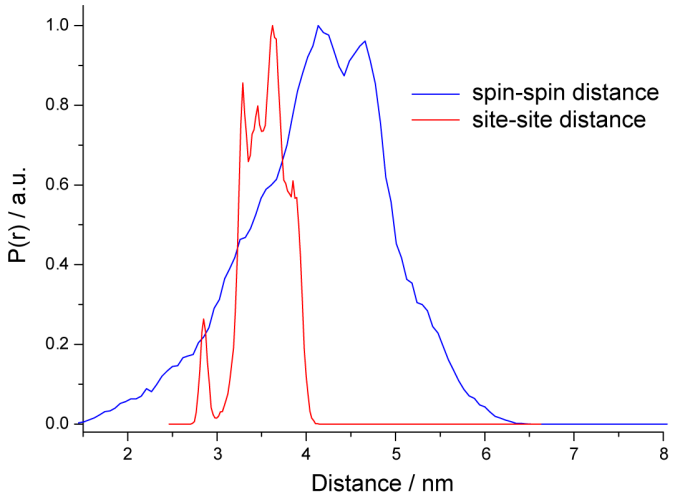

Figure S7. MD calculated distance distribution between spin labels of dslRNA ("spin-spin distance") and MD calculated distance distribution between corresponding label attachment sites ("site-site distance"), i.e. nitrogen atoms of the bases to which the aliphatic linkers are attached.

REFERENCES

S1. Bakin,A. and Ofengand,J. (1993) Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754-9762.

S2. Gustafsson,C. and Persson,B.C. (1998) Identification of the rrmA gene encoding the 23S rRNA m1G745 methyltransferase in Escherichia coli and characterization of an m1G745-deficient mutant. *J Bacteriol*., **180**, 359-365.

S3. Xaplanteri,M.A., Andreou,A., Dinos,G.P. and Kalpaxis,D.L. (2003) Effect of polyamines on the inhibition of peptidyltransferase by antibiotics: revisiting the mechanism of chloramphenicol action. *Nucleic Acids Res*., **31**, 5074-5083.

S4. Babaylova,E.S., Ivanov,A.V., Malygin,A.A., Vorobjeva,M.A., Venyaminova,A.G., Polienko,Y.F., Kirilyuk,I.A., Krumkacheva,O.A., Fedin,M.V., Karpova,G.G. and Bagryanskaya,E.G. (2014) A versatile approach for site-directed spin labeling and structural EPR studies of RNAs. *Org. Biomol. Chem*., **12**, 3129-3136.