Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

for

Packing energetics determine the folding routes of the RNase-H proteins

Shilpa Yadahalli[1,2], Shachi Gosavi[1,*]

[1]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bellary Road, Bangalore-560065, India.

[2]Manipal University, Madhav Nagar, Manipal-576104, India.

*Correspondence to:

Shachi Gosavi, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bellary Road, Bangalore-560065, India. Tel: +91-80-23666105, Fax: +91-80-23636662, Email: shachi@ncbs.res.in.

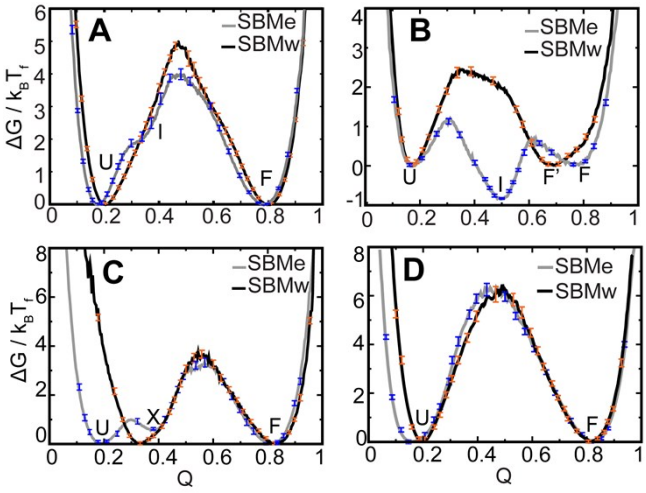

Figure S1. Free energy profiles (FEPs) of (A) eco, (B) hiv, (C) lc11 and (D) bh RNase-H show the scaled free energies (ΔG/kBT) as a function of the fraction of native contacts formed (Q) (see Methods). FEPs calculated from SBMe are in grey and from SBMw are in black. Error bars on the FEPs are shown in a different color for ease of visualization. U and F correspond to the unfolded and folded ensembles. I, F'and X are partially folded ensembles that are discussed in main text.

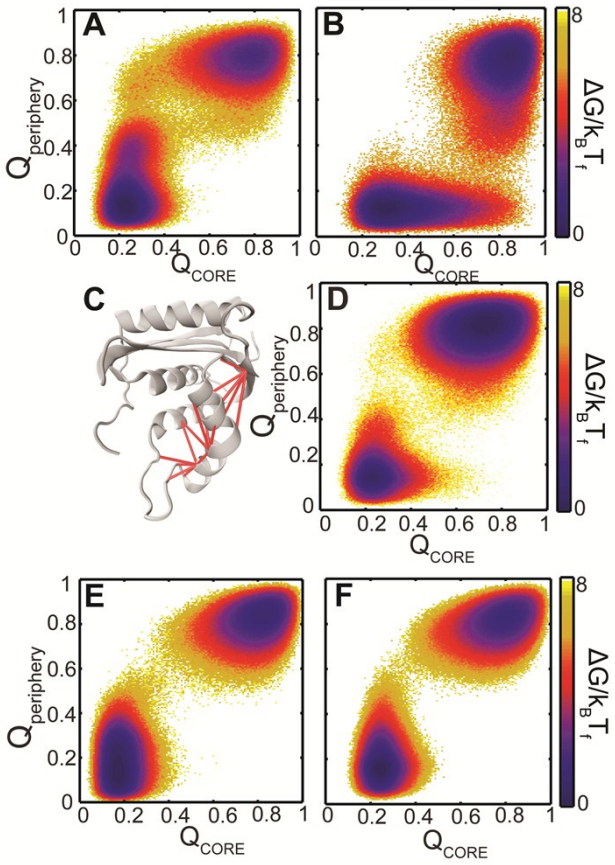

Figure S2. Flat-weighting CORE contacts modulates the CORE-periphery packing energy distribution and can cause a change in the folding routes of ecoRNase-H. 2DFES's of ecoRNase-H calculated from (A) SBMe and (B) SBMw are reproduced from Fig. 2A and 3A for ease of comparison. When the contacts shown in red on (C) the ecoRNase-H structure are flatweighted, the folding route is altered from (B) to that shown in the 2DFES of (D). (The origin of these contacts is discussed in S. Yadahalli and S. Gosavi, J. Mol. Biol. 428, 509-521, 2016.) In contrast, the 2DFES's of bhRNase-H calculated from (E) SBMe and (F) SBMw are very similar. These are reproduced from Fig. 2D and 3D for ease of comparison.

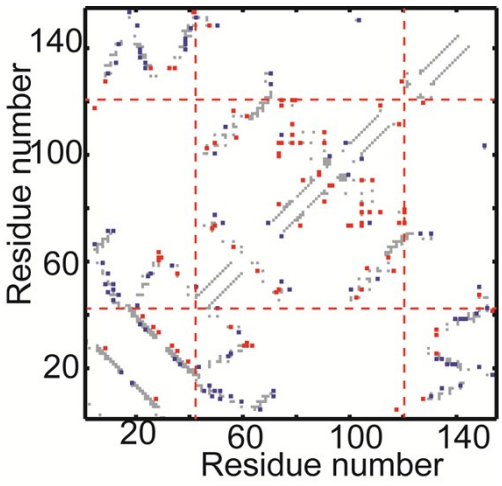

Figure S3. An example difference contact map used for calculating the number of lost and gained contacts in the RP method. A coloured square on the contact map at (x, y) implies that a contact is present between residues x and y in the ecoRNase-H structure. Contacts not present in the average RP but present in WT ecoRNase-H are shown in red and contacts not present in the WT ecoRNase-H but present in the RP are shown in blue. Contacts present in both the average RP and the WT are shown in grey. The red and blue contacts are further partitioned into CORE and periphery contacts for plotting Fig. 5. The average RP is created as follows: The frequency of every contact in the 200 RPs is calculated. The most frequently occurring N contacts are chosen to be in the average RP, where N is the number of contacts in WT ecoRNase-H. This average RP is not directly used in the analysis. However, it can be used for understanding which regions of the protein lose (or gain) the most contacts upon random permutation. In ecoRNase-H, CORE (the central red dashed square; residues 43 to 122) loses the most contacts.

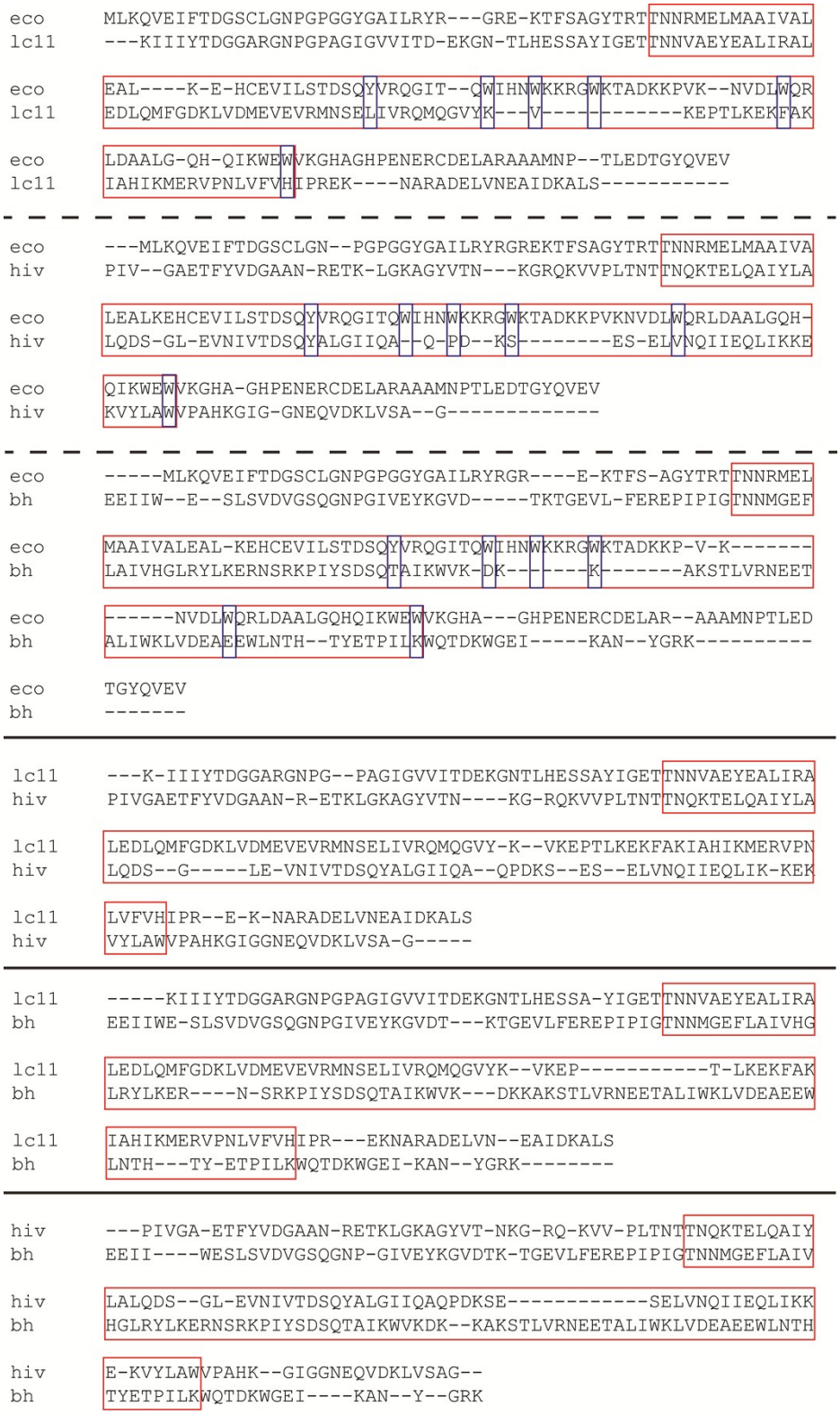

Figure S4. A pairwise structure-based sequence alignment for all pairs of sequences is calculated using the STAMP algorithm of VMD [62,63]. CORE regions of the proteins are enclosed in red boxes. The six aromatic residues of ecoRNase-H CORE whose contacts are responsible for the change in the folding route (Fig. S2) are marked by blue rectangles on the eco RNase-H alignment profiles. The

residues that they align with are also shown. No more than two of the aligned residues are aromatic residues in the other proteins.

	eco	lc11	hiv	bh
eco	-	19.08 (14)	19.13 (21.5)	10.16 (11.7)
lc11	19.08 (14)	-	15.89 (15.3)	12.27 (11.2)
hiv	19.13 (21.5)	15.89 (15.3)	-	15.79 (20.2)
bh	10.16 (11.7)	12.27 (11.2)	15.79 (20.2)	-

Table S1a. Percent sequence identities obtained from the pairwise structure-based sequence alignments shown in Fig. S4. Values shown in brackets are percent sequence identities calculated only from COREs of the RNase-Hs. These are the number of identical amino acids in CORE in an alignment/alignment length of CORE X 100.

	eco	lc11	hiv	bh
eco	-	2.06	1.98	3.01
lc11	2.06	-	2.13	3.34
hiv	1.98	2.13	-	3.23
bh	3.01	3.34	3.23	-

Table S1b. Pairwise Cα RMSD values in Å calculated from the structural alignment.

Protein	$M_{C\alpha}/M_{AA}$	CORE $M_{C\alpha}/M_{AA}$	periphery $M_{C\alpha}/M_{AA}$	T_f (ε units) SBMe	SBMw
ecoRNase-H	352/2068	142/954	210/1114	1.04	1.02
hivRNase-H	248/1570	95/631	153/939	0.96	0.96
lc11RNase-H	310/1809	124/818	186/991	1.01	0.96
bhRNase-H	315/2020	140/924	175/1096	1.05	1.00

Table S2. Protein and simulation details.

$M_{C\alpha}$=Number of Cα-Cα native contacts.

M_{AA}=Number of heavy atom native contacts.

T_f=folding temperature; ε, the basic energy scale in our simulations is set to 1kJ/mol

Protein	$M_{C\alpha}/M_{AA}$	CORE $M_{C\alpha}/M_{AA}$	periphery $M_{C\alpha}/M_{AA}$	T_f
ecoRNase-H flat-weighted mutant	352/1890	142/776	210/1114	1.03
hivRNase-H	248/1355	95/520	153/835	0.98
flat-weighted mutant				
only CORE flat	248/1435	95/520	153/915	0.96
only periphery flat	248/1466	95/631	153/835	0.97
lc11RNase-H	310/1176	124/492	186/684	1.02
flat-weighted mutant				
only CORE flat	310/1483	124/492	153/991	0.98
only periphery flat	310/1502	124/881	153/684	0.95

Table S3a. Flat-weighted mutants and their simulation details.

$M_{C\alpha}$=Number of Cα-Cα native contacts.

M_{AA}=Number of heavy atom native contacts.

T_f=folding temperature; ε is basic energy scale in our simulations set to 1kJ/mol

ecoRNase-H flat-weighted mutant	hivRNase-H flat-weighted mutant	lc11RNase-H flat-weighted mutant
81-77, 81-76, 81-85, 98-85, 90-85, 120-75, 91-85, 89-85, 104-85, 81-104, 81-73, 120-79, 120-74, 120-98, 120-69	From CORE: 87-81, 75-71, 92-54, 98-60, 97-59, 55-51 From periphery: 31-27, 39-19, 33-25, 112-9, 100-1	From CORE: 63-59, 108-59, 57-53, 81-77, 87-81, 114-83, 86-81, 109-105, 89-82, 114-72, 87-82, 79-48, 82-78, 84-80, 117-74, 56-52, 107-103, 108-104, 116-77, 117-75, 55-51, 95-89 From periphery: 36-19, 40-15, 22-5, 34-21, 128-124, 24-3, 29-25, 37-17, 73-5, 26-1, 43-13, 20-7, 119-75, 16-11, 32-23, 40-16, 124-33, 130-126, 31-23, 38-17, 70-2

Table S3b. List of contacts edited in the flat-weighted mutants

RNase-H	average contact strength in CORE	average contact strength in periphery	$r_{C:P}$
ecoRNase-H	1.14	0.91	1.26
hivRNase-H	1.05	0.97	1.08
lc11RNase-H	1.12	0.91	1.23
BhRNase-H	1.03	0.98	1.05

Table S4. Energetic heterogeneity in the RNase-Hs. Average contact strengths can be calculated from Table S2 as $(M_{AA}^{X}/M_{C\alpha}^{X})\times(M_{C\alpha}/M_{AA})$ where X is the region for which the contact strength is being calculated, i.e., CORE or periphery. $r_{C:P}$ is the ratio of the average contact strength of CORE to that of the periphery.