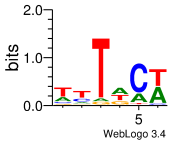
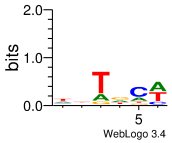
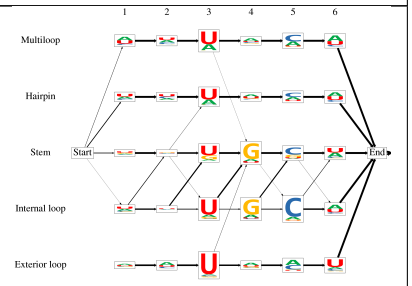
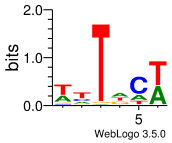
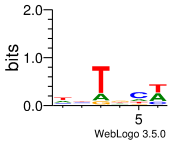
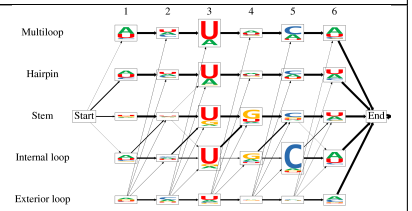
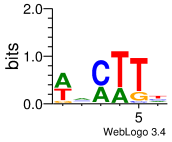
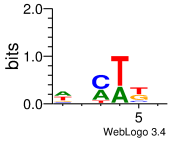
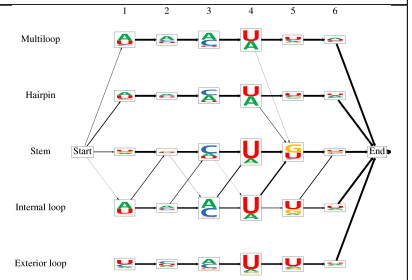
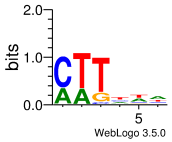
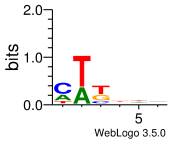
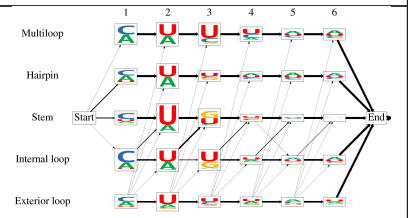
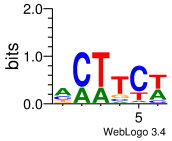
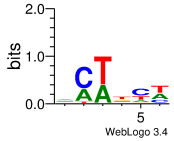
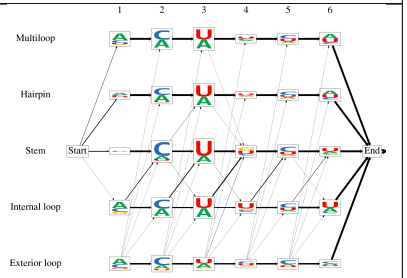
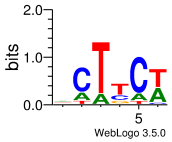
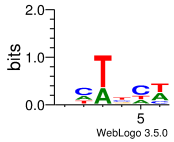
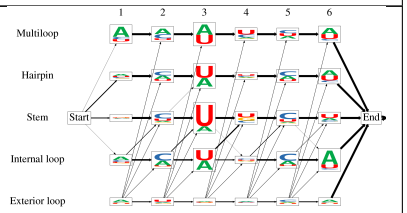
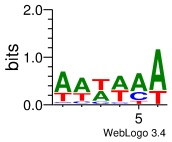
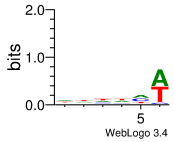
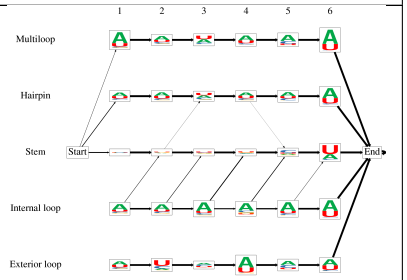
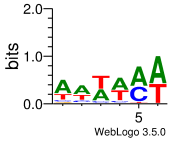
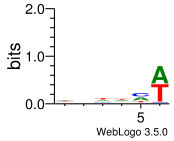
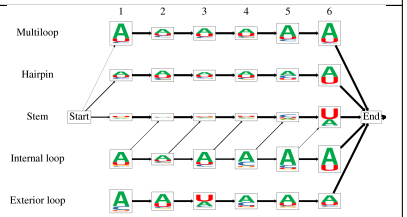
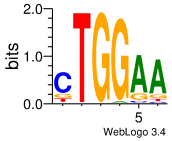
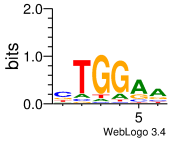
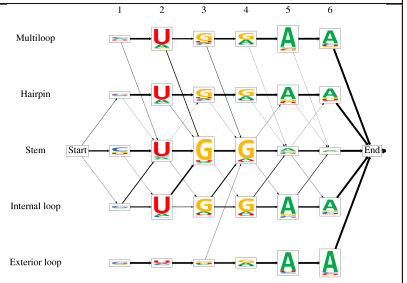
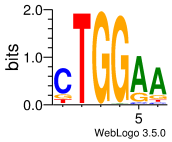
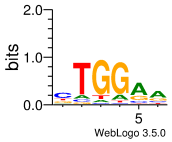
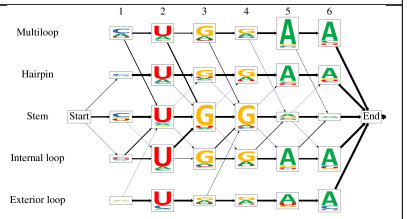
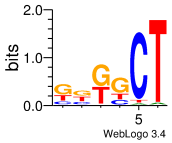
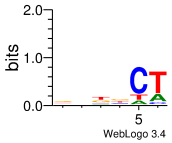
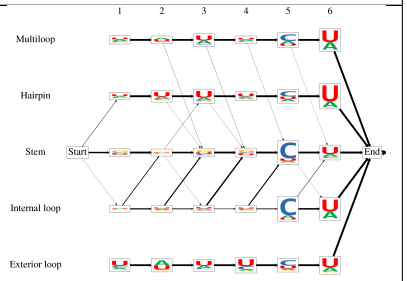
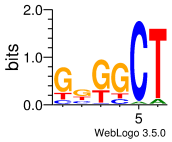
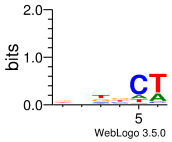
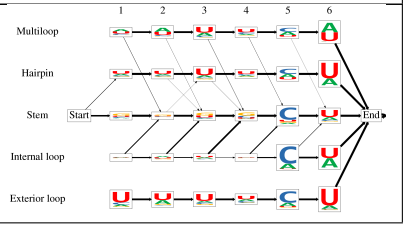
Additional File 3: ssHMM motifs for RNAshapes and RNAstructure

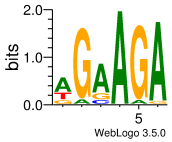
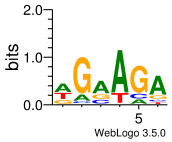
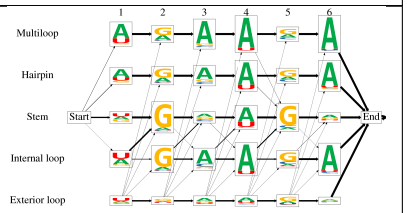
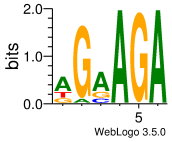
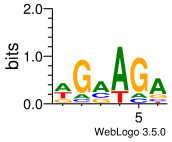
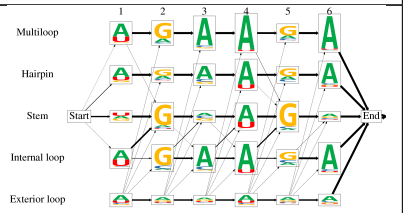
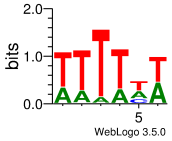
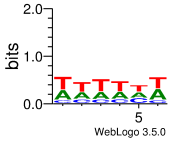
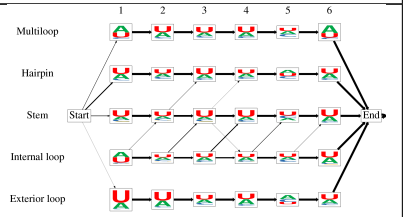
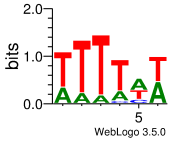
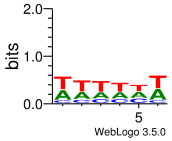
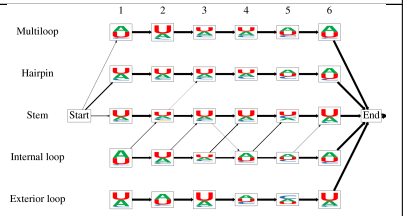
On the following pages, two motifs that were recovered by ssHMM from each of the 25 CLIP-Seq datasets are shown. For each protein, one motif was produced using structure predictions from *RNAshapes*, another using structure predictions from *RNAstructure*. The table has the following eight columns:

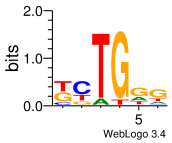
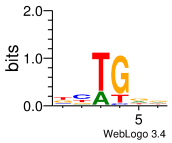
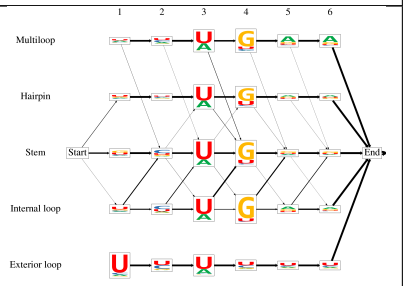
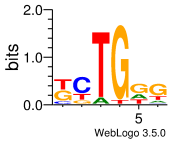
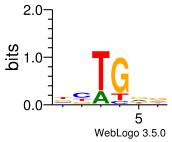
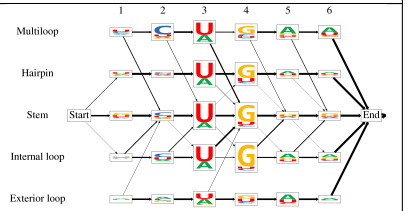
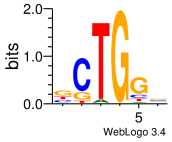
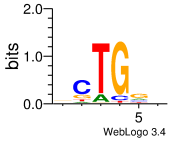
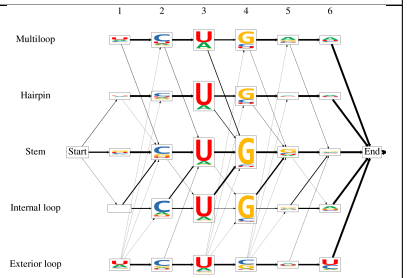
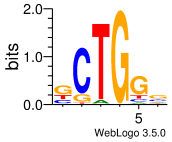
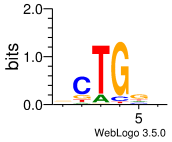
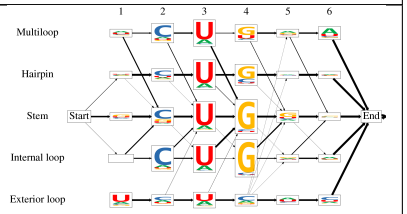
1. Protein - Name of the RBP (with protocol in parentheses if multiple CLIP-Seq datasets exist for the protein)
2. Structure - Tool used to predict secondary structures: *RNAshapes* or *RNAstructure*
3. Sequence-structure likelihood - Likelihood of all sequences and structures given the trained motif model
4. Information content (from the top sequences) - Information content of the motif derived from the top 1000 sequences
5. Information content (directly from the model) - Information content of the motif derived directly from the ssHMM
6. Sequence logo (from the top sequences) - Sequence logo derived from the top 1000 sequences
7. Sequence logo (directly from the model) - Sequence logo derived from the ssHMM
8. ssHMM - Visualization of the full ssHMM

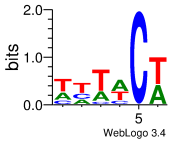
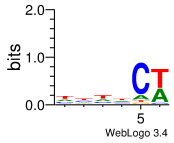
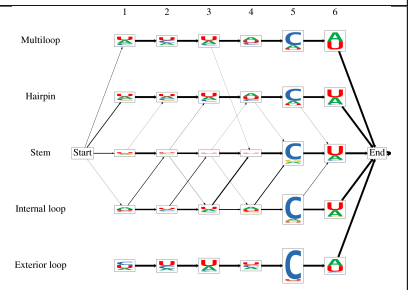
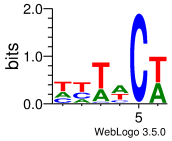
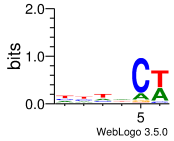
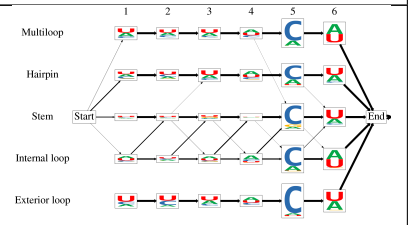
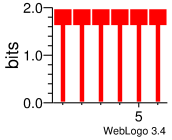
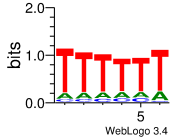
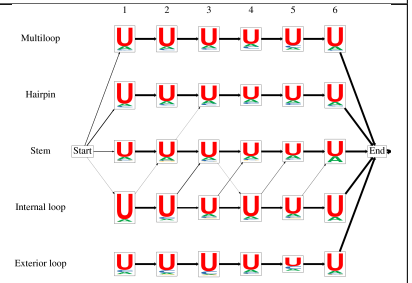
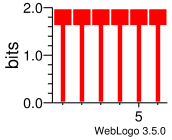
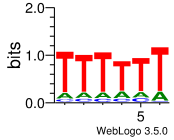
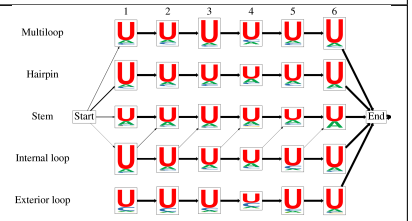
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
AGO1234	shapes	-42752.0	0.69/0.97/1.93	0.33/0.59/1.08			
AGO1234	structures	-41387.0	0.66/1.38/2.31	0.31/0.84/1.29			
AGO2 (HITSCLIP)	shapes	-41335.0	0.66/0.77/1.73	0.38/0.58/1.09			
AGO2 (HITSCLIP)	structures	-40596.0	0.65/1.31/2.2	0.35/0.8/1.26			

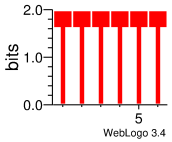
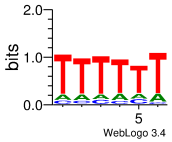
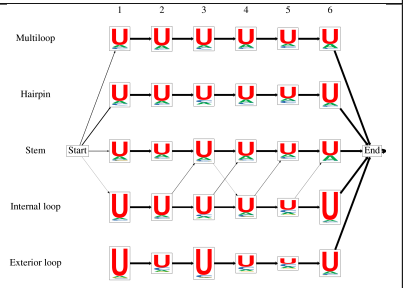
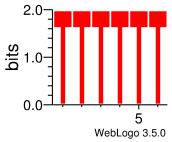
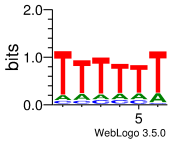
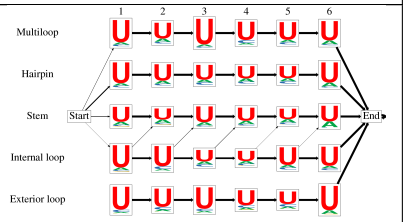
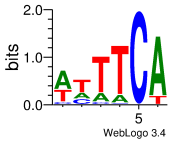
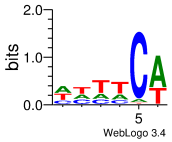
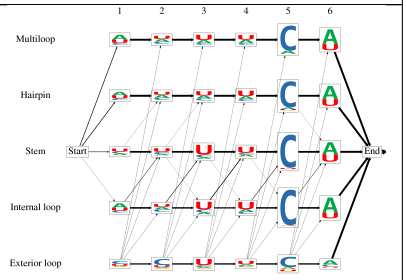
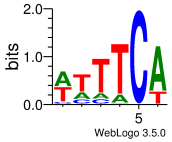
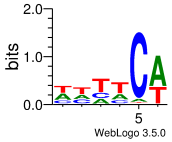
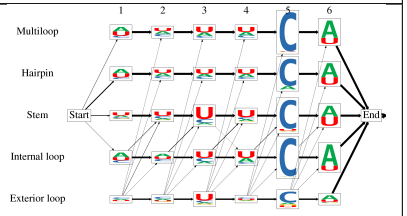
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
AGO2 (PARCLIP)	shapes	-44522.0	0.76/0.83/1.79	0.45/0.54/1.07			
AGO2 (PARCLIP)	structures	-42720.0	0.78/1.24/2.2	0.42/0.75/1.25			
CAPRIN1	shapes	-45640.0	0.72/0.91/1.87	0.23/0.62/1.0			
CAPRIN1	structures	-44551.0	0.69/1.2/2.16	0.23/0.78/1.14			

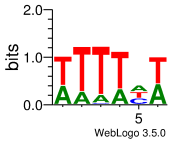
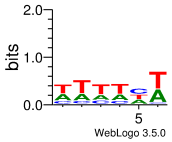
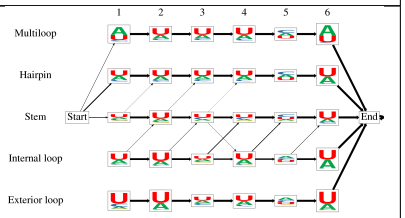
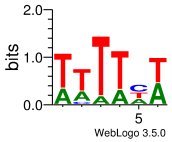
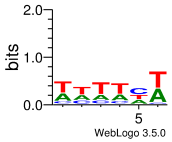
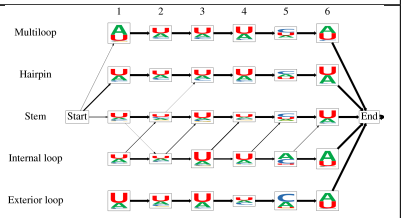
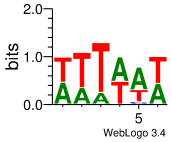
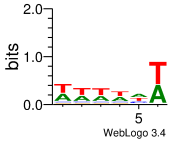
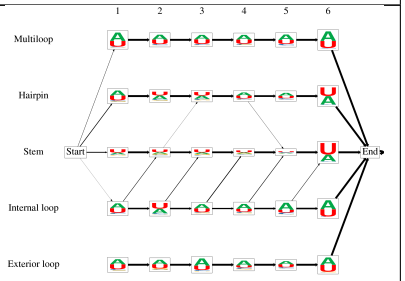
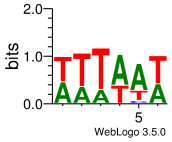
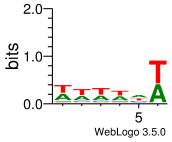
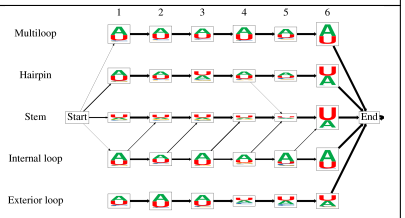
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
DGCR8	shapes	-47319.0	1.39/1.17/2.67	0.65/0.77/1.48			
DGCR8	structures	-45383.0	1.38/1.36/2.86	0.63/0.95/1.65			
DICER	shapes	-102384.0	0.93/1.65/2.82	0.33/1.0/1.43			
DICER	structures	-97813.0	1.03/1.79/3.02	0.29/1.21/1.6			

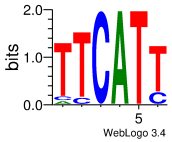
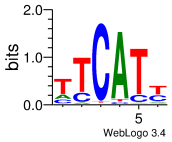
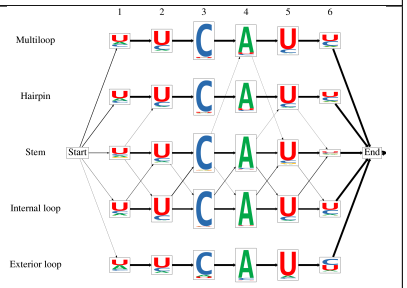
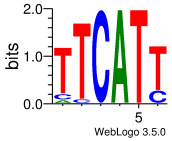
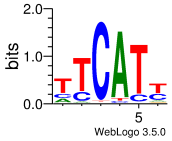
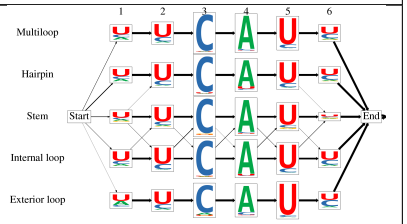
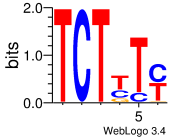
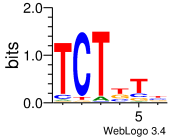
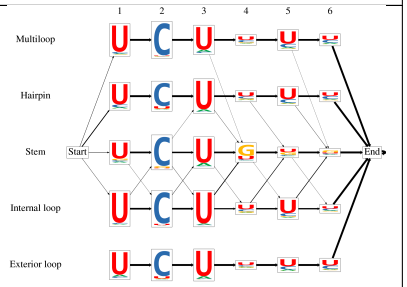
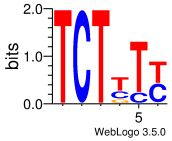
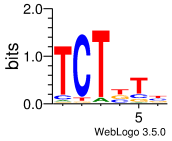
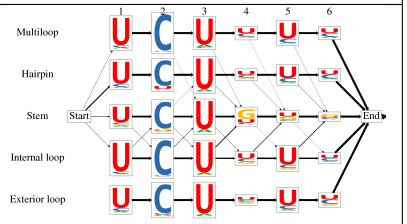
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
EIF4A3	shapes	-49247.0	1.43/0.96/2.53	0.75/0.73/1.57			
EIF4A3	structures	-47503.0	1.48/1.15/2.75	0.76/0.9/1.74			
EWSR1	shapes	-67540.0	1.08/1.17/2.31	0.49/0.44/0.96			
EWSR1	structures	-66054.0	1.06/1.65/2.76	0.49/0.57/1.08			

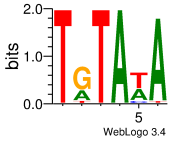
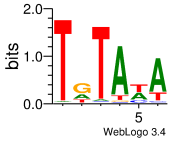
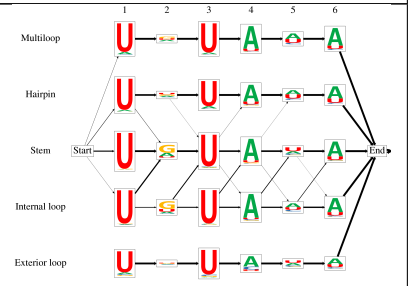
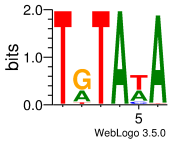
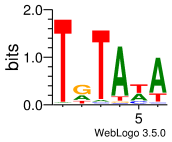
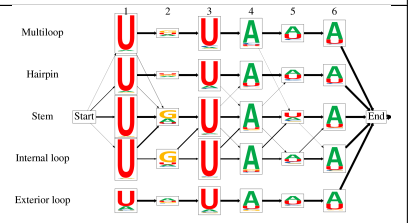
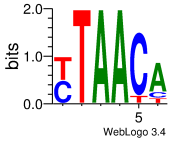
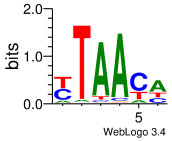
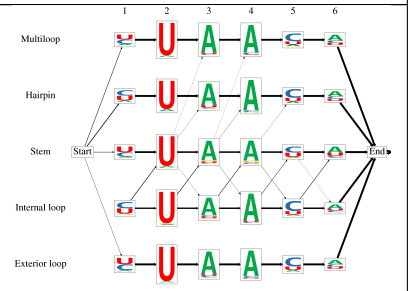
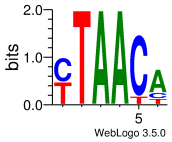
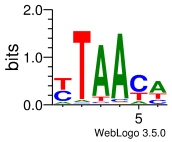
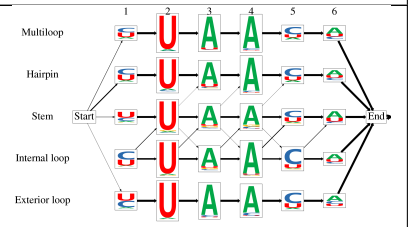
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
EZH2	shapes	-60837.0	0.82/1.83/2.69	0.47/0.85/1.37			
EZH2	structures	-59057.0	0.87/2.03/2.92	0.48/1.04/1.55			
FMRP	shapes	-42690.0	0.94/2.17/3.13	0.57/1.02/1.64			
FMRP	structures	-41180.0	1.02/2.32/3.34	0.59/1.17/1.8			

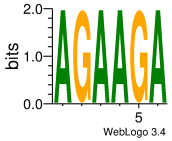
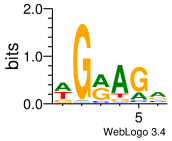
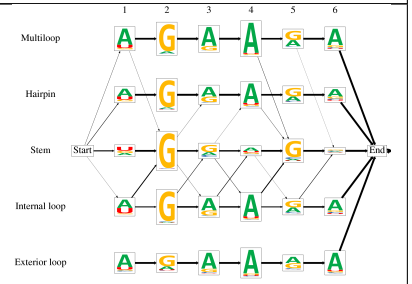
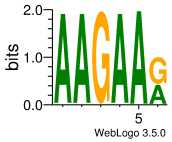
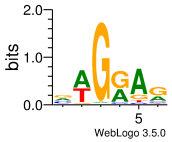
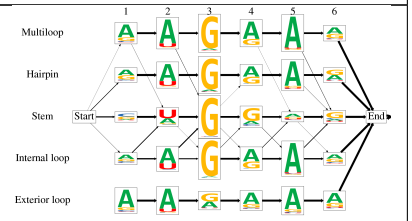
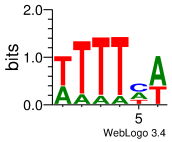
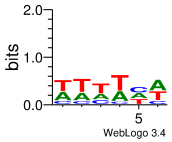
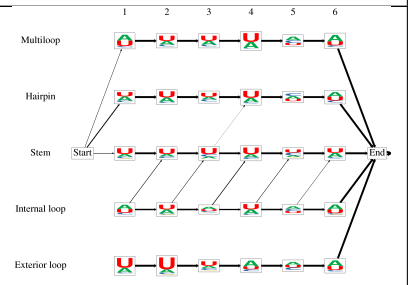
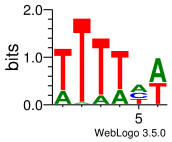
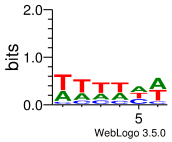
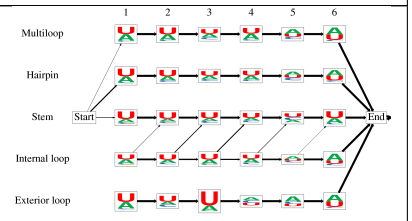
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
FXR2	shapes	-49360.0	0.89/1.49/2.43	0.38/0.46/0.93			
FXR2	structures	-48724.0	0.86/1.99/2.89	0.38/0.58/1.05			
HuR (HITSCLIP)	shapes	-31420.0	2.0/1.33/3.32	1.03/0.39/1.44			
HuR (HITSCLIP)	structures	-30157.0	2.0/1.43/3.43	1.04/0.48/1.54			

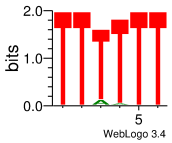
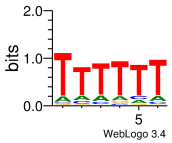
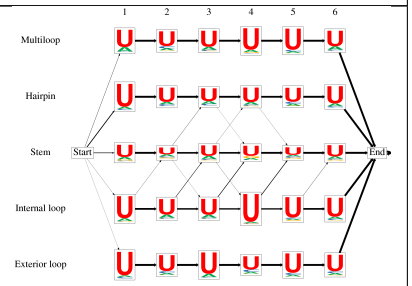
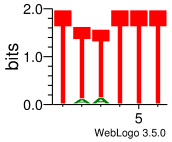
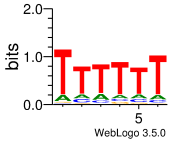
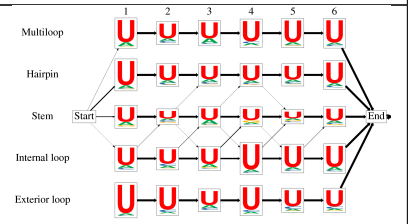
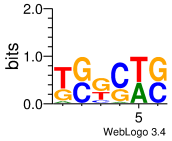
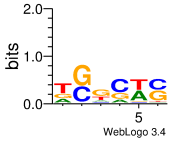
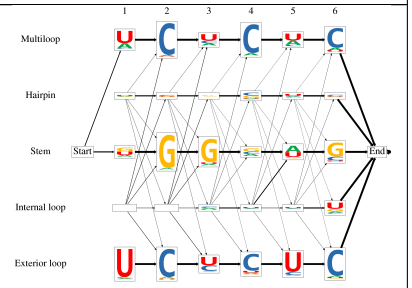
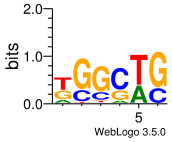
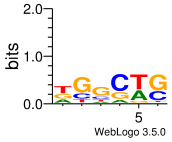
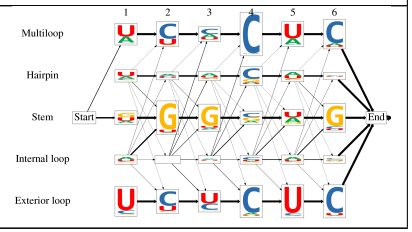
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
HuR (PARCLIP)	shapes	-29022.0	2.0/0.89/2.88	0.99/0.4/1.4			
HuR (PARCLIP)	structures	-28136.0	2.0/1.01/3.0	0.99/0.5/1.5			
IGF2BP123	shapes	-38436.0	1.16/0.87/2.1	0.72/0.5/1.25			
IGF2BP123	structures	-37401.0	1.12/1.18/2.41	0.66/0.61/1.32			

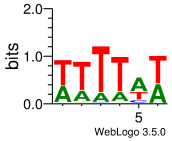
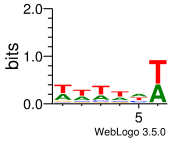
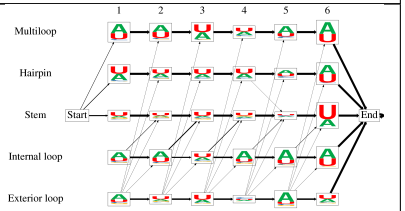
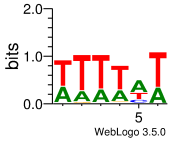
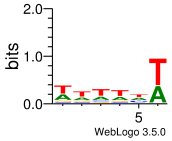
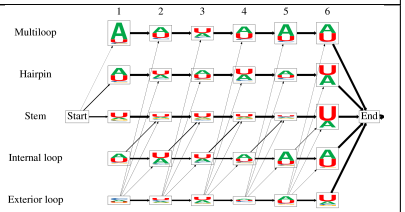
Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
LIN28B	shapes	-42080.0	1.01/1.42/2.48	0.47/0.54/1.05			
LIN28B	structures	-41553.0	0.98/1.99/2.99	0.47/0.66/1.17			
MOV10	shapes	-146964.0	1.03/0.81/2.23	0.42/0.55/1.02			
MOV10	structures	-144431.0	1.02/1.4/2.81	0.4/0.74/1.19			

Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
NOVA	shapes	-37197.0	1.68/1.08/2.8	1.06/0.17/1.26			
NOVA	structures	-35978.0	1.69/1.18/2.89	1.08/0.22/1.33			
PTBP1	shapes	-45769.0	1.46/1.35/2.88	0.86/0.31/1.27			
PTBP1	structures	-43512.0	1.46/1.85/3.38	0.85/0.48/1.43			

Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
PUM2	shapes	-35589.0	1.55/1.23/2.86	1.06/0.62/1.71			
PUM2	structures	-34360.0	1.52/1.72/3.32	1.07/0.74/1.84			
QKI	shapes	-28418.0	1.56/1.14/2.74	1.02/0.23/1.27			
QKI	structures	-26657.0	1.53/1.7/3.25	1.01/0.35/1.38			

Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
SFRS1	shapes	-170516.0	2.0/1.32/3.32	0.76/0.57/1.46			
SFRS1	structures	-165584.0	1.85/2.32/4.17	0.76/0.68/1.56			
TAF2N	shapes	-70599.0	1.13/1.36/2.56	0.5/0.52/1.05			
TAF2N	structures	-68007.0	1.14/1.81/3.0	0.51/0.67/1.2			

Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
TIA1	shapes	-35300.0	1.91/0.69/2.61	0.93/0.38/1.34			
TIA1	structures	-35297.0	1.85/1.29/3.15	0.88/0.55/1.47			
YY1	shapes	-10655.0	0.87/1.32/2.85	0.52/1.13/2.13			
YY1	structures	-11437.0	0.86/1.51/2.89	0.52/1.22/2.13			

Protein	Structure	Seq.-str. ll	IC (1000)	IC (model)	Motif (1000)	Motif (model)	ssHMM
ZC3H7B	shapes	-45515.0	0.94/0.79/1.98	0.41/0.69/1.18			
ZC3H7B	structures	-44709.0	0.9/1.32/2.49	0.4/0.94/1.43