

Supplementary Figures

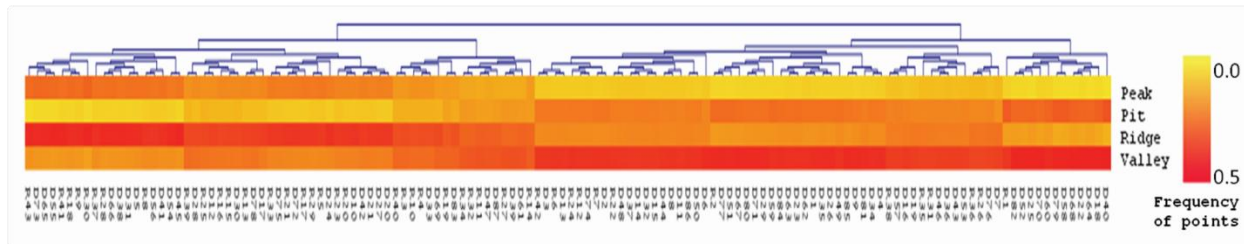


Figure S1 Hierarchical clustering of the redundant set of dsDNA and ssRNA interfaces

Hierarchical clustering of the interface vectors extracted from the “NAbind-130” dataset. Interfaces are labeled according to the NA-binding protein label (see supplementary data file S1), R and D, denote RNA-binding and DNA-binding interfaces, respectively. Vectors represent the frequency of the binding interface points related to the local geometric shapes Peak, Pit, Ridge and Valley. High frequency is colored red and low frequency colored yellow. Color bar is shown.

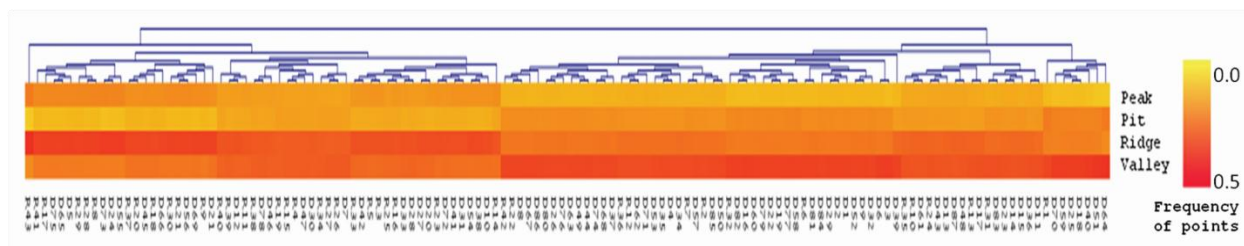


Figure S2 Hierarchical clustering of the dsDNA and ssRNA electrostatic patches

Hierarchical clustering of the vectors representing the largest positive patches, extracted from the “NAbind-77nr” dataset. Interfaces are labeled according to the NA-binding protein label (see supplementary data file S1), R and D, denote RNA-binding and DNA-binding interfaces, respectively. Vectors represent the frequency of the patch points related to the local geometric shapes Peak, Pit, Ridge and Valley. High frequency is colored red and low frequency colored yellow. Color bar is shown.

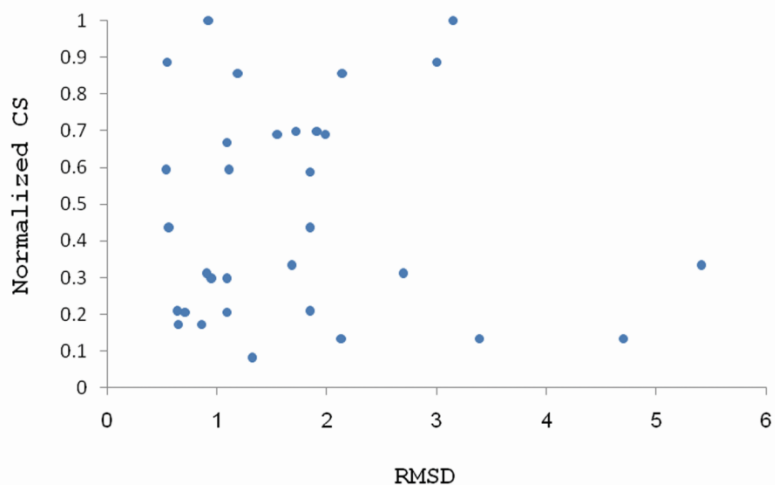


Figure S3: Correlation between the RMSD and Normalized Correlation Score. RMSD was calculated between the pairs of bound and unbound proteins. Correlation Score (CS) was calculated between the DG vectors of the bound and unbound interfaces. As shown no correlation was found between the RMSD and the normalized CS (Pearson correlation = -0.037)

Supplementary Tables

Table S1: The fundamental surface shapes

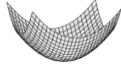
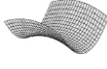
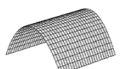
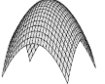
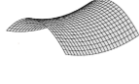
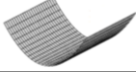
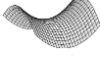
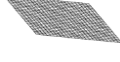
	$K > 0$	$K < 0$	$K = 0$
$H > 0$	 Pit	 Saddle Valley	 Ridge
$H < 0$	 Peak	 Saddle Ridge	 Valley
$H = 0$	Other	 Minimal Surface	 Flat

Table S2: Summary of results

	Training *	Testing *	False DNA	False RNA	True DNA	True RNA	Training Accuracy	Sensitivity	Specificity	MCC ^s	Testing accuracy
Unsupervised Clustering											
NAbind-130	I	I	31	13	56	30	66	81	49	0.33	
NAbind-77nr	I	I	11	6	41	19	78	87	63	0.53	
NAbind-130	P	P	28	14	59	29	67	80	50	0.33	
Decision Tree											
NAbind-130	I	I	23	6	64	37	78	91	62	0.57	
NAbind-77nr	I	I	11	4	41	21	81	91	66	0.6	
NAbind-130	P	I	9	19	78	24	79	80	73	0.49	
SVM **											
NAbind-130 E	I	I	14	36	51	29	62	59	67	0.24	0.63
NAbind-77nr E	I	I	10	19	33	15	62	64	60	0.22	0.62
NAbind-130 EG	I	I	12	24	63	31	72	72	72	0.42	0.78
NAbind-77nr EG	I	I	4	9	43	21	83	83	83	0.64	0.82

^s MCC was calculated as described in Material and Methods

* I=real interface, P=predicted patch

**All SVM test were performed using hold-one-out cross validation , E =SVM was trained on electrostatic features only, EG=SVM was trained on Electrostatic+Geometric features

Table S3: Statistical analysis of individual parameters selected for the SVM

Feature	T-test *	F-test*	Mann whitney test*
Frequency of interface points belong to the Peak	4.7e-7	5.8e-1	4.9 e-6
Frequency of interface points belong to the Pit	9.8e-6	7.8e-1	3.1e-5
Frequency of interface points belong to the Ridge	5.6e-6	9.9e-1	2.6e-5
Frequency of interface points belong to the Valley	7.6e-7	8.8e-1	1.9e-6
Number of atoms in the third patch [§]	8.3e-4	2.9e-4	4.5e-3
Number of atoms in the negative patch [§]	5.3e-3	4.0e-7	2.0e-2
Molecular weights per residue [§]	2.1e-2	6.5e-1	4.6e-2
Frequency of alpha helix in positive patch [§]	8.1e-2	3.3e-1	8.7e-2
Frequency of beta sheet in positive patch [§]	2.2e-1	8.9e-1	1.3e-1
Number of polar residues in positive patch [§]	1.0e-1	2.6e-1	1.3e-1

*exact p-values are given

[§]Described in details in Shazman, S. and Mandel-Gutfreund, Y. (2008) Classifying RNA-binding proteins based on electrostatic properties. *PLoS Comput Biol*, **4**, e1000146.

Table S4: Distinguishing DNA- from RNA-binding Zinc Finger domains

PDB code	Protein name	SCOP family	Nucleic-acid	Prediction
1ubdC	Ying-Yang 1 protein ZINC FINGER DOMAIN	C2H2 and C2HC zinc fingers	DNA	RNA binding
1a1hA	QGSR ZINC FINGER PEPTIDE	C2H2 and C2HC zinc fingers	DNA	DNA binding
1zaaC	ZIF268	C2H2 and C2HC zinc fingers	DNA	RNA binding
1yujA	GAGA FACTOR	C2H2 and C2HC zinc fingers	DNA	DNA binding
2gliA	PROTEIN (FIVE- FINGER GLI)	C2H2 and C2HC zinc fingers	DNA	DNA binding
1a1fA	THREE-FINGER ZIF268 PEPTIDE	C2H2 and C2HC zinc fingers	DNA	DNA binding
1f4sP	TRANSCRIPTIONAL FACTOR ALCR	Zn2/Cys6 DNA- binding domain	DNA	DNA binding
1rgoA	Butyrate response factor 2	CCCH zinc finger	RNA	RNA binding
1a1tA	NUCLEOCAPSID	retrovirus zinc finger like domain	RNA	DNA binding
1u6pA	Gag polyprotein	retrovirus zinc finger like domain	RNA	RNA binding
2hghA	FINGERS 4,5 AND 6 OF TFIIIA	C2H2 and C2HC zinc fingers	RNA	RNA binding
1tf3A	FINGERS 1-3 OF TFIIIA	C2H2 and C2HC zinc fingers	DNA	DNA binding