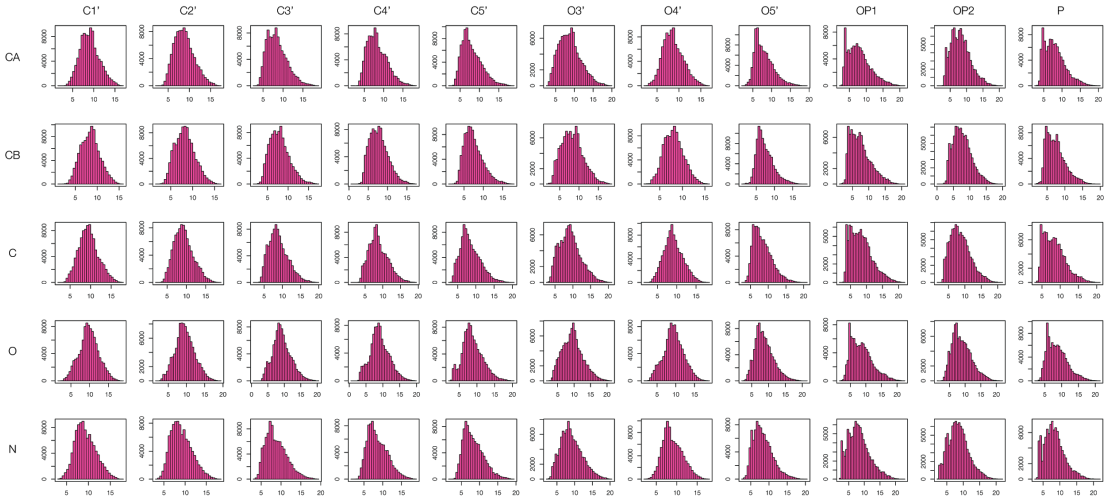
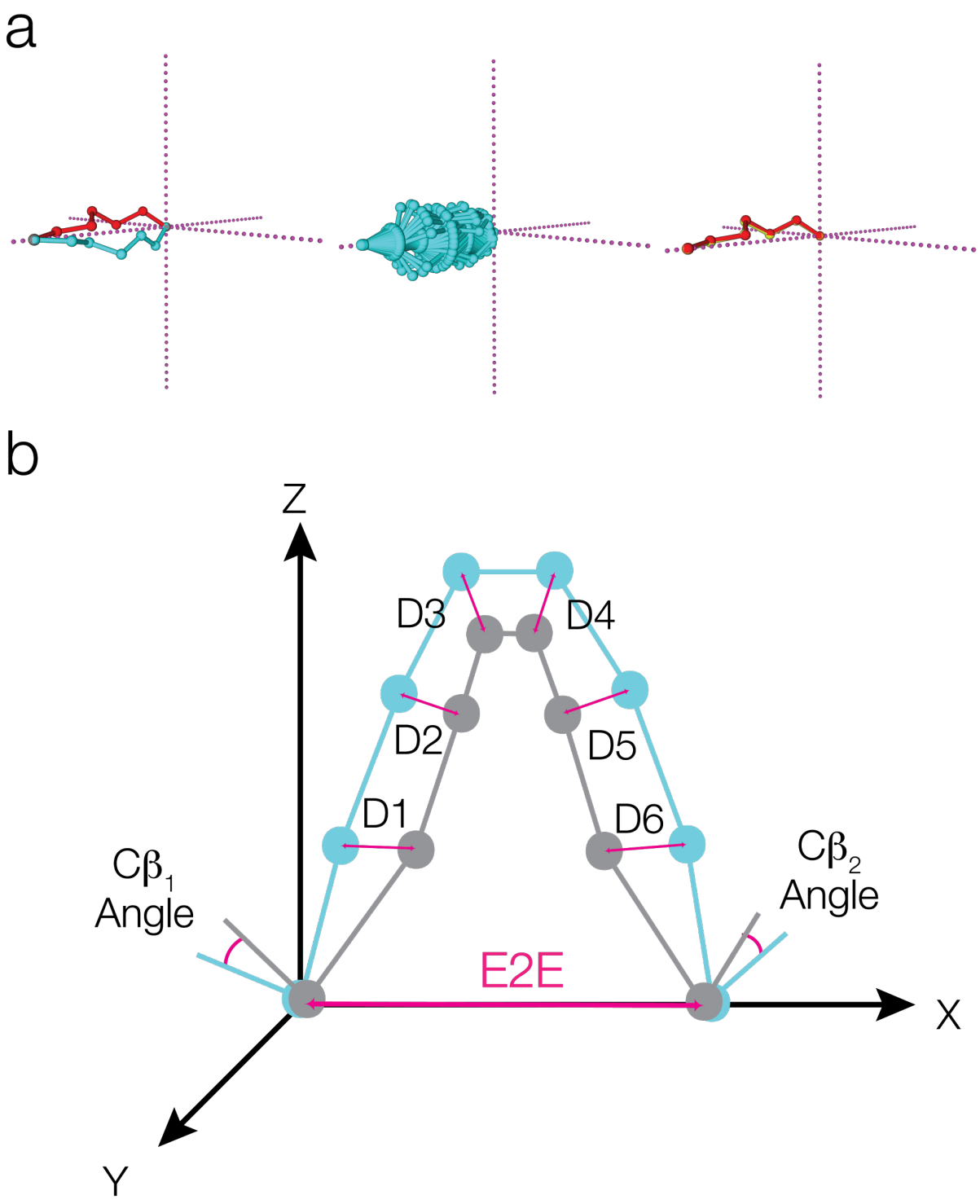
Supplementary Information

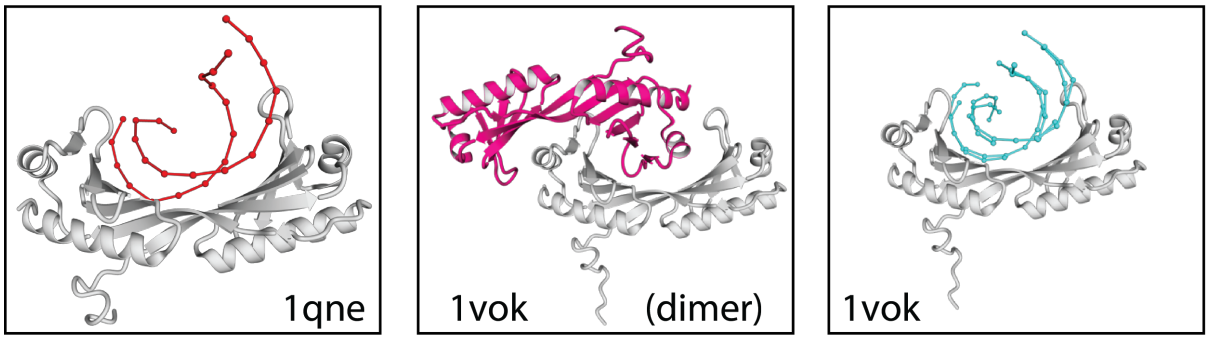


Supplementary Figure 1: Normal distance distributions for atomic distances between pepXs backbone atoms (Cα, Cβ, C, O, N) and dnaXs backbone atoms (C1', C2', C3', C4', C5', O3', O4', O5', OP1, OP2, P).

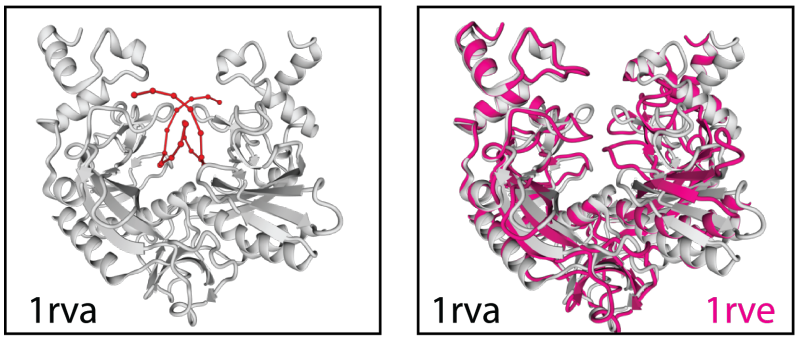


Supplementary Figure 2: (a) pinwheel rotation of pepX (blue) around x axis to find the best superimposition. (b) $C\beta_1$ and $C\beta_2$ angles are restricted by a default threshold of 5°. Schema for the different fit levels: fit level=2 considers D1 and D6 distances while fit level=3 considers D1,D2,D5,D6 distances, Fit level=4 considers D1,D2,D3,D4,D5,D6.

a



b

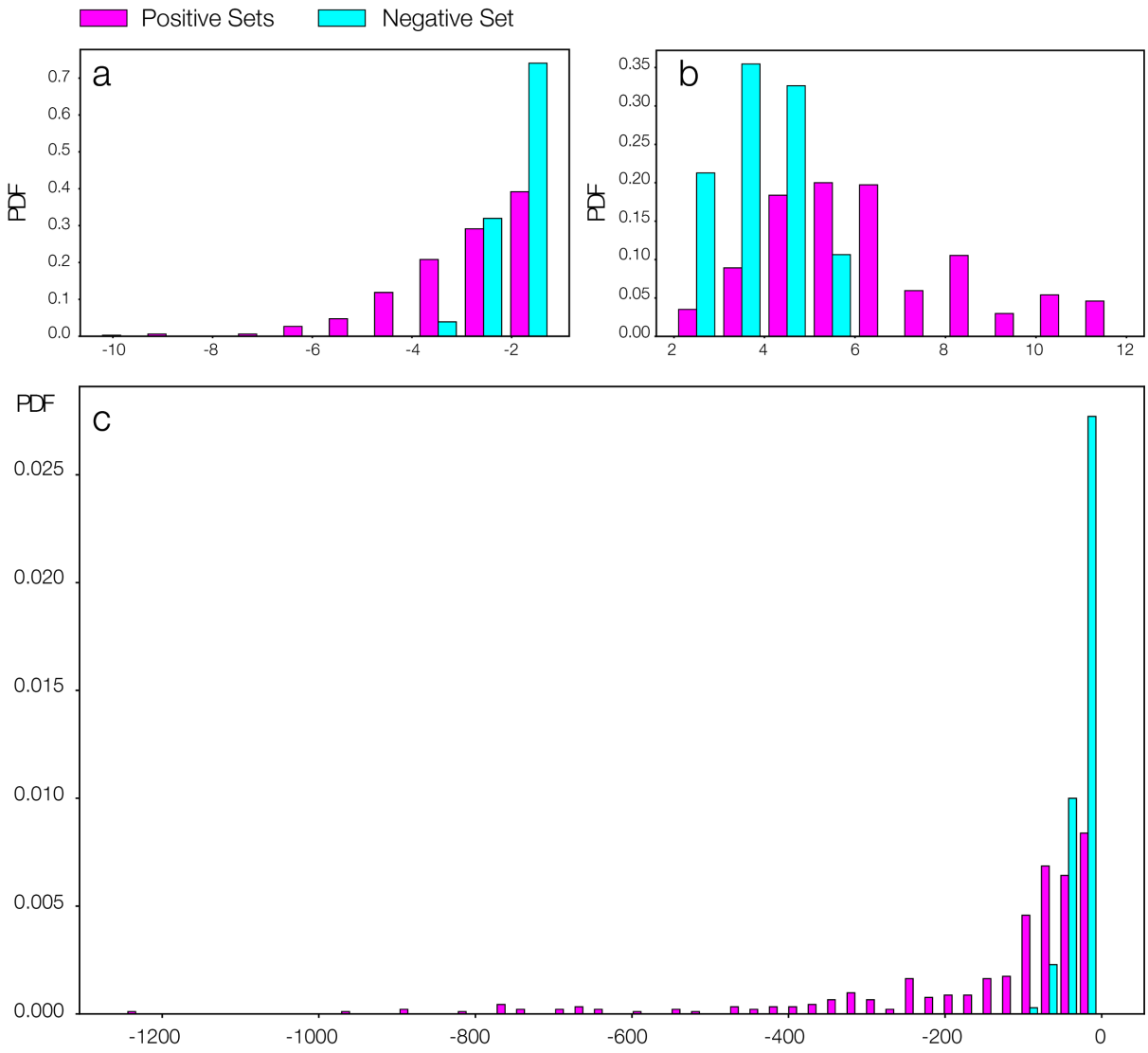


c

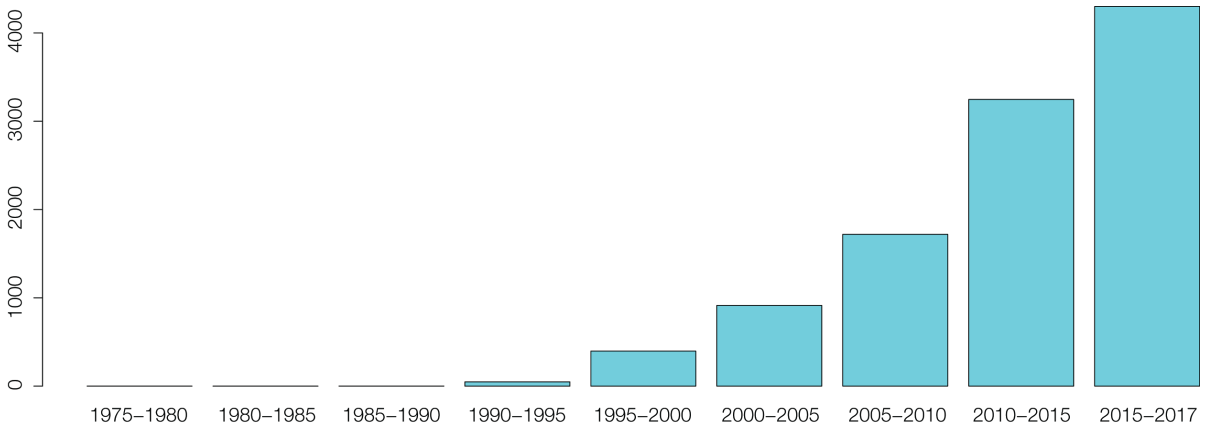
Dataset ID	TPR	FRP
DB179	98.94%	1.06%
PDNA-316	99.63%	0.37%
PD138	98.95%	1.05%
PDNA62	96.17%	3.83%
BIND54	98.23%	1.77%

Dataset ID	Docks (FP)
NB250	0

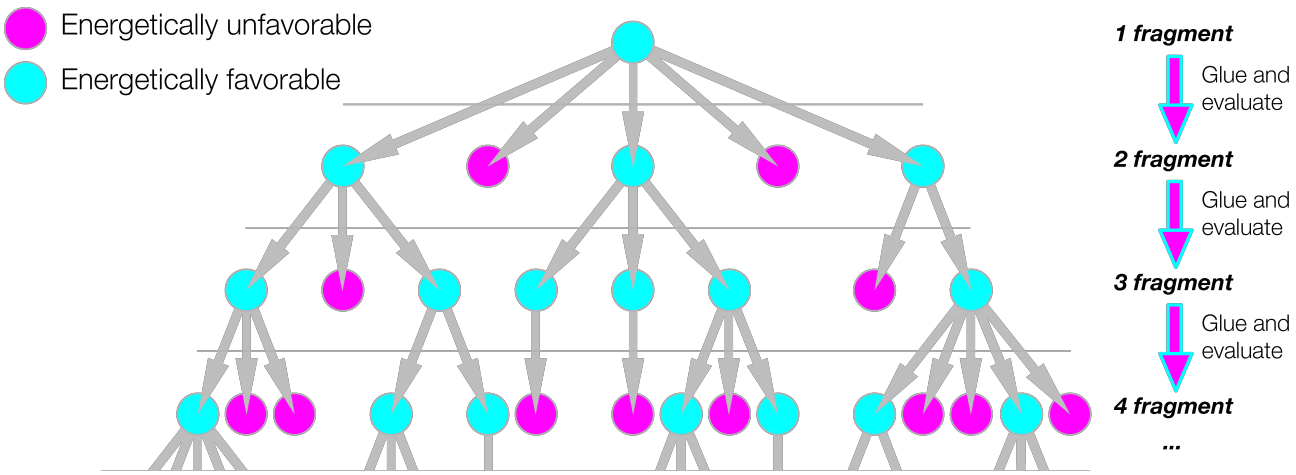
Supplementary Figure 3: (a) Incorrect predicted DNA binding site for 1vok unbound configuration is correctly predicted after removing one of the monomers blocking the binding site, (b) the unbounded structure 1rve presents significant differences in the DNA binding site making the docking implausible. (c) Summarized results for validation sets of binding and nonbinding proteins, a dock is considered as true positive if it has an RMSD < 1.8Å with the crystal one. The dockings were performed excluding from the analysis all those crystal representing the same UniProt ID of the target structure.



Supplementary Figure 4: Binding protein datasets (BIND54. DB179, PD138, PDBNA62 and PDNA-316) joined and summarized in light blue, nonbinding proteins dataset (NB250) in purple. (a) Histogram of energies ($DG_{Docking}$) for accepted dockings using the exhaustive mode over the validation datasets. (b) Histogram of the accepted dockings' number of contacting nucleotides (CN) to the target proteins over the validation dataset. (c) Histogram of filtering score defined as $DG_{Docking} * CN^2$.



Supplementary Figure 5: PDB dsDP complexes deposited records.



Supplementary Figure 6: Branch and Bound strategy of the *GlueDocks* command: each docked fragment is glued with its neighbor docks and energetically evaluated, those energetically favorables are recursively glued until reaching the maximal possible length. The solution trees are explored with a Deep-first search.

PDB	UniProt	PFam	Total Docks	Proportion of contacting nucleotides well predicted
5hbu	P0C093	PF00440	432	0.93
5hlh	Q5HKZ1	PF01047	23	0.86
5hro	P03366	PF06817	190	0.95
5i44	P45870	PF13411	142	0.74
5ink	Q9JJX7	PF03372	5	0.43
5iwm	P66937	PF00204	32	0.65
5j2q	P04585	PF00078	87	0.92
5jgh	Q02486	PF00505	225	0.92
5jk0	O25386	PF00589	111	0.74
5k58	A7ZTJ2	PF11799	81	0.89
5lgy	P04637	PF00870	985	0.94
5t14	Q9UBT6	PF00817	21	0.96
5tw1	A0QZ11	PF13397	14	0.84
Total			**2348**	**0.88**

Supplementary Table 1: Detail of the predicted best matching sequences for a set of 13 crystals belonging to different proteins (UniProt) of different protein families (PFam). The nucleotides which are not in contact (nor its pair in the dsDNA) with the protein were not considered for the analysis. The evaluation to calculate the best sequences in this table was performed with the PADA1 all-atom statistical force field.

DataSet	PDB Codes List
PADA1 Validation Data Set	5d23, 5ei9, 5exh, 5gke, 5gkf, 5gkg, 5gkh, 5gki, 5gkp, 5hbu, 5hdn, 5hf7, 5hhh, 5hhi, 5hlf, 5hlg, 5hlh, 5hlk, 5hnk, 5hod, 5hoo, 5hp1, 5hq2, 5hr4, 5hr9, 5hrb, 5hrd, 5hrf, 5hrg, 5hrh, 5hro, 5hrt, 5hso, 5i3u, 5i42, 5i44, 5i50, 5iii, 5iij, 5iik, 5iil, 5iim, 5iin, 5iio, 5ink, 5inl, 5ino, 5inq, 5ipl, 5ivw, 5iwi, 5iwm, 5j0n, 5j0o, 5j0p, 5j0q, 5j0r, 5j0s, 5j0t, 5j0u, 5j0w, 5j0x, 5j0y, 5j29, 5j2a, 5j2b, 5j2c, 5j2d, 5j2e, 5j2f, 5j2g, 5j2h, 5j2i, 5j2j, 5j2k, 5j2m, 5j2n, 5j2p, 5j2q, 5j3e, 5jgh, 5jk0, 5jlw, 5jlx, 5jub, 5jum, 5jvt, 5jxy, 5k1y, 5k58, 5k5o, 5k5q, 5k5r, 5k7z, 5k98, 5kbd, 5kbj, 5ke6, 5ke7, 5ke8, 5ke9, 5kea, 5keb, 5kfa, 5kfb, 5kfc, 5kfd, 5kfe, 5kff, 5kfg, 5kfh, 5kfi, 5kfj, 5kfk, 5kfl, 5kfm, 5kfn, 5kfo, 5kfp, 5kfq, 5kfr, 5kfs, 5kft, 5kfu, 5kfv, 5kfw, 5kfx, 5kfy, 5kfz, 5kg0, 5kg1, 5kg2, 5kg3, 5kg4, 5kg5, 5kg6, 5kg7, 5kk1, 5kl2, 5kl3, 5kl4, 5kl5, 5kl6, 5kl7, 5kn8, 5kn9, 5krb, 5kt2, 5kt3, 5kt4, 5kt5, 5kt6, 5kt7, 5kub, 5l0m, 5l1i, 5l1j, 5l1k, 5l1l, 5l2x, 5l6l, 5l7c, 5l9x, 5lej, 5lek, 5lgy, 5lrs, 5m0q, 5m0r, 5swm, 5sy7, 5szt, 5t14, 5t1j, 5t2w, 5tb8, 5tb9, 5tba, 5tbb, 5tbc, 5tct, 5tgx, 5th3, 5trd, 5tzs, 5tzv, 5u9h, 5fyw, 5fz5, 5fur, 5i2d, 5ipl, 5inp, 5iud, 5iy6, 5iy7, 5iy8, 5iy9, 5iya, 5iyb, 5iyc, 5iyd, 5sva, 5tw1, 5b0z, 5b2i, 5b2j, 5b31, 5b32, 5b33, 5b40, 5kgf

Supplementary Table 2: Validation dataset of PADA1, containing dsDP and released in the PDB after 2016. The validation set was divided into three subsets: small proteins, huge proteins and histones.

DataSet	PDB Codes List
Haddock Benchmark Data Set	1a74, 1azp, 1b3t, 1bdt, 1by4, 1cma, 1ddn, 1dfm, 1diz, 1ea4, 1emh, 1eyu, 1f4k, 1fok, 1g9z, 1h9t, 1hjc, 1jj4, 1jt0, 1k79, 1kc6, 1ksy, 1mnn, 1o3t, 1pt3, 1qne, 1qrv, 1r4o, 1rpe, 1rva, 1tro, 1vas, 1vrr, 1w0t, 1z63, 1z9c, 1zme, 1zs4, 2c5r, 2fio, 2fl3, 2irf, 2oaa, 3bam, 3cro, 4ktq, 7mht

Supplementary Table 3: Validation dataset of the Haddock algorithm used for benchmarking, consisting of 74 DNA binding proteins provided in its bounded and unbounded form.