**Supporting Information**

**emPDBA: protein-DNA binding affinity prediction by combining features from binding partners and interface learned with ensemble regression model**

Shuang Yang, Weikang Gong, Tong Zhou, Xiaohan Sun, Lei Chen, Wenxue Zhou, Chunhua Li\*

*Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China*

\*All correspondence should be addressed to Chunhua Li (Email: [chunhuali@bjut.edu.cn](mailto:chunhuali@bjut.edu.cn)).

**Table S1**. Detailed information of the training dataset PD304

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PDB ID** | **Protein name** | **DNA name** | ***T* (K)** | ***Kd* (mol)** | **△*G* (kcal/mol)** |
| **Double I type** | | | | | |
| 1BF5 | SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 1-ALPHA/BETA | DNA (5'-D(\*AP\*CP\*AP\*GP\*TP\*TP\*TP\*CP\*CP\*CP\*GP\*TP\*AP\*AP\*AP\*TP\*G P\*C)-3') | 298 | 2.60E-10 | -9.80 |
| 1CLQ | PROTEIN (DNA POLYMERASE) | DNA (5'-D(\*GP\*CP\*GP\*GP\*AP\*AP\*CP\*TP\*AP\*CP\*T)-3') | 298 | 7.00E-05 | -9.47 |
| 1EBM | 8-OXOGUANINE DNA GLYCOSYLASE | DNA (5'-D(\*GP\*CP\*GP\*TP\*CP\*CP\*AP\*(8OG)P\*GP\*TP\*CP\*TP\*AP\*CP\*C)-3') | 310 | 8.20E-09 | -10.25 |
| 1EFA | LAC REPRESSOR | DNA (5'-D(\*GP\*AP\*AP\*T\*TP\*GP\*TP\*GP\*AP\*GP\*CP\*GP\*CP\*TP\*CP\*AP\*CP\*AP\*AP\*TP\*T)-3') | 310 | 1.50E-11 | -15.30 |
| 1FZP | Transcriptional regulator SarA | fnbB | 298 | 2.69E-10 | -13.05 |
| 1GJI | Proto-oncogene c-Rel | IL-2 CD28RE | 310 | 2.50E-08 | -10.78 |
| 1JEY | X-ray repair cross-complementing protein 6 | Nicked DNA | 298 | 1.50E-09 | -12.03 |
| 1LEI | Nuclear factor NF-kappa-B p105 subunit | HIV-Sp1 | 298 | 4.00E-10 | -12.81 |
| 1N6Q | Reverse Transcriptase | 5'-D(\*AP\*T\*GP\*CP\*AP\*TP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3' | 310 | 1.98E-08 | -10.93 |
| 1NNE | DNA Mismatch Repair protein MutS | 5'-D(\*GP\*CP\*GP\*AP\*CP\*GP\*CP\*TP\*AP\*GP\*CP\*GP\*TP\*GP\*CP\*GP\*GP\*CP\*TP\*CP\*GP\*TP\*C)-3' | 277 | 2.28E-08 | -9.65 |
| 1OWR | Nuclear factor of activated T-cells, cytoplasmic 2 | dsDNA | 298 | 2.70E-09 | -11.68 |
| 1Q0T | DNA adenine methylase | DNA | 273 | 3.50E-08 | -9.31 |
| 1RUN | cAMP-activated global transcriptional regulator CRP | consensus DNA site for CAP (ICAP) | 296 | 2.38E-11 | -14.39 |
| 1W7A | DNA MISMATCH REPAIR PROTEIN MUTS | 5'-D(\*AP\*GP\*CP\*TP\*GP\*CP\*CP\*AP\*GP\*GP \*CP\*AP\*CP\*CP\*AP\*GP\*TP\*GP\*TP\*CP\*AP\*GP\*CP\*GP\*TP\*CP\*CP\*TP\* AP\*T)-3' | 298 | 3.30E-08 | -10.16 |
| 1Z63 | Helicase of the snf2/rad54 family | DNA duplex | 295 | 1.00E-07 | -9.45 |
| 2BSQ | Antitoxin FitA | IR36 | 296 | 4.50E-09 | -11.30 |
| 2NTZ | ParB | 5'-D(\*CP\*GP\*TP\*GP\*AP\*AP\*AP\*TP\*CP\*GP\*CP\*CP\*AP\*CP\*GP\*A)-3' | 298 | 4.90E-08 | -9.97 |
| 2QBY | Cell division control protein 6 homolog 1 | DNA (33-MER) | 310 | 2.70E-08 | -9.09 |
| 2W42 | Piwi protein AF\_1318 | dsDNA | 293 | 2.92E-08 | -10.10 |
| 2WWY | ATP-DEPENDENT DNA HELICASE Q1 | DNA OLIGO (27BP) | 310 | 6.25E-09 | -11.60 |
| 2XRO | HTH-type transcriptional regulator TtgV | ttgG | 303 | 1.90E-08 | -10.71 |
| 2XRZ | DNA photolyase | UV-damaged Specific | 298 | 4.40E-08 | -10.03 |
| 3A46 | Probable formamidopyrimidine-DNA glycosylase | 23-mer THF:C | 291 | 4.70E-11 | -13.75 |
| 3AU6 | DNA polymerase beta family (X family) | 5'-D(\*CP\*AP\*GP\*TP\*AP\*TP\*(DDG))-3' | 298 | 6.60E-09 | -11.11 |
| 3BEP | Beta sliding clamp | 18/28 mer duplex | 295 | 1.20E-07 | -9.34 |
| 3BRD | Protein lin-12 | 19-mer DNA | 298 | 3.20E-08 | -10.22 |
| 3CVV | RE11660p | DNA | 298 | 8.60E-06 | -6.91 |
| 3GFI | Putative MarR family transcriptional regulator | ST1 DNA | 298 | 3.34E-07 | -8.83 |
| 3GLF | DNA polymerase III subunit tau | DNA | 298 | 1.80E-05 | -6.47 |
| 3K4X | Proliferating cell nuclear antigen | DNA (5'-D(\*CP\*CP\*CP\*AP\*TP\*CP\*GP\*TP\*AP\*T)-3') | 298 | 2.70E-09 | -11.64 |
| 3MKW | Protein SopB | C39 | 298 | 1.40E-08 | -10.71 |
| 3OQM | Phosphocarrier protein HPr | syn Cre | 298 | 1.60E-09 | -11.99 |
| 3POV | Shutoff alkaline exonuclease | dsDNA | 298 | 6.00E-06 | -7.12 |
| 3Q05 | Cellular tumor antigen p53 | specific | 298 | 7.70E-09 | -11.06 |
| 3RNU | Gamma-interferon-inducible protein 16 | ODN 787/788 | 298 | 5.51E-07 | -8.53 |
| 3SLP | Exonuclease | 5'-D(\*GP\*CP\*GP\*AP\*CP\*TP\*AP\*GP\*TP\*CP\*GP\*C)-3' | 298 | 2.10E-09 | -11.79 |
| 3U3W | Transcriptional activator PlcR protein | 5'-D(P\*CP\*TP\*AP\*TP\*GP\*CP\*AP\*AP\*TP\*AP\*TP\*TP\*TP\*CP\*AP\*TP\*AP\*T)-3' | 310 | 8.62E-07 | -8.57 |
| 3ZDA | Flap endonuclease Xni | Dup1 | 298 | 8.00E-07 | -8.31 |
| 3ZVK | Ribonuclease VapC2 | cognate DNA | 298 | 1.02E-07 | -9.53 |
| 4BXO | Fanconi anemia group M protein | dsDNA | 298 | 1.13E-07 | -9.47 |
| 4C2T | DNA helicase | 3' dsDNA | 298 | 3.60E-07 | -8.79 |
| 4GCL | Nucleoid occlusion factor SlmA | 12 mer SBS DNA | 298 | 1.57E-07 | -9.28 |
| 4IQR | Hepatocyte nuclear factor 4-alpha | DNA (5'-D(\*GP\*GP\*AP\*AP\*CP\*TP\*AP\*GP\*GP\*TP\*CP\*AP\*AP\*AP\*GP\*GP\*TP\*CP\*AP\*G)-3') | 296 | 8.89E-08 | -9.60 |
| 4O3M | Bloom syndrome protein | 16/24-nucleotide 3'-overhang duplex DNA | 298 | 1.28E-08 | -10.76 |
| 4PU3 | Serine/threonine-protein kinase toxin HipA | hipBA SO operator DNA | 298 | 5.50E-09 | -11.26 |
| 4PXI | CprB | DNA | 298 | 2.00E-07 | -9.13 |
| 4RIB | Fanconi-associated nuclease 1 | DNA - T1 | 277 | 1.03E-08 | -10.12 |
| 4WZW | Pumilio homolog 3 | dsDNA | 298 | 8.41E-08 | -9.65 |
| 4ZPK | Aryl hydrocarbon receptor nuclear translocator | dsDNA | 298 | 4.50E-08 | -10.02 |
| 5C51 | DNA polymerase subunit gamma-1 | (-)FTC-TP | 310 | 6.30E-05 | -5.96 |
| 5D39 | Signal transducer and activator of transcription 6 | DNA (5'-D(P\*TP\*CP\*TP\*GP\*TP\*CP\*TP\*TP\*CP\*CP\*AP\*GP\*GP\*AP\*AP\*AP\*TP\*CP\*CP\*AP\*T)-3') | 298 | 1.80E-06 | -7.83 |
| 5D9Y | Methylcytosine dioxygenase TET2 | FAM-18-bp-dsDNA(5fC) | 298 | 9.00E-07 | -8.24 |
| 5DFF | DNA- | nicked product | 293 | 2.70E-09 | -11.49 |
| 5E24 | Maltose/maltodextrin-binding periplasmic protein | dsDNA | 298 | 2.00E-09 | -11.86 |
| 5GPC | Transcriptional regulator | fadRpromoter (IR30) | 298 | 1.00E-07 | -9.54 |
| 5IUD | DNA polymerase alpha catalytic subunit | dsDNA | 296 | 4.29E-07 | -8.62 |
| 5MHK | Major viral transcription factor ICP4 | IE3\_19mer | 298 | 1.60E-09 | -11.99 |
| 5NJ8 | Aryl hydrocarbon receptor nuclear translocator | 12 mer | 293 | 9.00E-08 | -9.45 |
| 5SY7 | Aryl hydrocarbon receptor nuclear translocator | 21-mer | 298 | 1.75E-08 | -10.58 |
| 5U1J | AAA\_31 domain-containing protein | dsDNA | 298 | 1.16E-07 | -9.46 |
| 5UAN | Retinoic acid receptor beta | DR1 DNA | 298 | 3.00E-08 | -10.26 |
| 5W34 | DNA primase | dsDNA | 295 | 5.00E-07 | -8.51 |
| 6GIS | Proliferating cell nuclear antigen | dsDNA | 308 | 7.00E-04 | -4.45 |
| 6H8Q | Cohesin subunit SCC3 | 32 - bp dsDNA | 298 | 2.20E-06 | -7.71 |
| **Double II type** | | | | | |
| 10MH | PROTEIN (CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI) | DNA (5'-D(P\*CP\*CP\*AP\*TP\*GP\*(5CM)P\*GP\*CP\*TP\*GP\*AP\*C)-3') | 310 | 1.00E-10 | -14.13 |
| 1B94 | RESTRICTION ENDONUCLEASE ECORV | DNA (5'-D(\*AP\*AP\*AP\*GP\*AP\*TP\*AP\*TP\*CP\*TP\*T)-3') | 298 | 1.32E-13 | -17.50 |
| 1BHM | Type-2 restriction enzyme BamHI | 23bp oligomer | 294 | 7.69E-10 | -12.26 |
| 1BPX | PROTEIN (DNA POLYMERASE BETA) | DNA (5'-D(\*CP\*CP\*GP\*AP\*CP\*GP\*GP\*CP\*GP\*CP\*AP\*TP\*CP\*AP\*GP\*C)-3') | 310 | 7.10E-06 | -7.28 |
| 1CMA | PROTEIN (MET REPRESSOR) | DNA (5'-D(\*TP\*TP\*AP\*GP\*AP\*CP\*GP\*TP\*CP\*T)-3') | 326.2 | 2.66E-09 | -10.80 |
| 1D02 | Type-2 restriction enzyme MunI | SP(specific site) | 298 | 1.30E-06 | -8.03 |
| 1FOS | Transcription factor AP-1 | TRE duplex | 298 | 2.00E-07 | -9.13 |
| 1H88 | CCAAT/ENHANCER BINDING PROTEIN BETA | DNA(5'-(\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*GP\*CP\*AP\* AP\*TP\*CP\*CP\*TP\*TP\*AP\*AP\*CP\*GP\*GP\*AP\*CP\*TP\*G)-3') | 298 | 3.53E-08 | -10.16 |
| 1H9T | Fatty acid metabolism regulator protein | fadB operator | 298 | 3.00E-10 | -12.98 |
| 1K82 | formamidopyrimidine-DNA glycosylase | 5'-D(\*GP\*GP\*CP\*TP\*TP\*CP\*CP\*TP\*CP\*CP\*TP\*GP\*G)-3' | 310 | 3.00E-09 | -8.66 |
| 1LRR | Negative modulator of initiation of replication | hemimethlyated DNA | 277 | 7.00E-06 | -6.53 |
| 1NH2 | Transcription initiation factor IIA large subunit | ds DNA | 298 | 5.73E-09 | -11.24 |
| 1ORN | Endonuclease III | 5'-D(\*AP\*AP\*GP\*AP\*CP\*GP\*TP\*GP\*GP\*AP\*C)-3' | 310 | 8.95E-09 | -11.38 |
| 1R8D | HTH-type transcriptional activator mta | mtaT1T2 | 298 | 9.60E-08 | -9.57 |
| 1T8E | DNA polymerase | 5'-D(P\*CP\*GP\*AP\*AP\*AP\*AP\*CP\*GP\*AP\*CP\*GP\*GP\*CP\*CP\*AP\*GP\*TP\*GP\*CP\*CP\*AP\*(2DT))-3' | 277 | 2.80E-05 | -5.75 |
| 2BNW | Omega transcriptional repressor | dsDNA | 310 | 2.00E-08 | -10.92 |
| 2FQZ | R.Ecl18kI | DNA STRAND 1 | 310 | 3.50E-09 | -11.95 |
| 2G1P | DNA adenine methylase | 5'-D(\*TP\*CP\*TP\*AP\*GP\*AP\*TP\*CP\*TP\*AP\*GP\*A)-3' | 295 | 8.40E-08 | -9.44 |
| 2IT0 | Iron-dependent repressor IdeR | fxbA | 298 | 2.47E-05 | -6.28 |
| 2IVK | ENDONUCLEASE I | 5'-D(\*GP\*AP\*AP\*TP\*TP\*CP\*GP\*AP\*TP\*CP \*GP\*AP\*AP\*TP\*T)-3' | 310 | 1.00E-08 | -11.31 |
| 2KY8 | Methyl-CpG-binding domain protein 2 | dsDNA | 298 | 2.10E-06 | -7.74 |
| 2MOE | Methyl-CpG-binding domain protein 4 | DNA (5'-D(\*GP\*GP\*AP\*TP\*(5CM)P\*GP\*GP\*CP\*TP\*C)-3') | 298 | 1.42E-05 | -6.74 |
| 2NTC | Large T antigen | MCV-P4-P5-P6 | 293 | 2.05E-07 | -8.97 |
| 2VLA | RESTRICTION ENDONUCLEASE R.BPUJI | 5'-D(\*GP\*GP\*TP\*AP\*CP\*CP\*CP\*GP\*TP\*GP \*GP\*A)-3' | 298 | 7.30E-09 | -11.09 |
| 2W8K | DNA polymerase IV | 24-mer/36-mer (N2-Naph) | 298 | 2.90E-05 | -6.19 |
| 2YPF | Hax2 | BS3 | 298 | 8.80E-08 | -9.62 |
| 3AAF | Werner syndrome ATP-dependent helicase | blunt-ended duplex | 298 | 2.53E-07 | -9.00 |
| 3COQ | Regulatory protein GAL4 | dsDNA | 298 | 2.46E-08 | -10.38 |
| 3GQC | DNA repair protein REV1 | 5'-D(\*AP\*TP\*CP\*CP\*TP\*CP\*CP\*CP\*CP\*TP\*AP\*(DOC))-3' | 310 | 3.90E-08 | -10.07 |
| 3H0D | Transcriptional regulator CtsR | 26-bp DNA duplex containing thec ctsr box | 303 | 2.22E-08 | -10.61 |
| 3IV5 | DNA-binding protein fis | DNA (27-MER) | 100 | 2.00E-10 | -12.29 |
| 3JSP | LexA repressor | AT-GC spacer | 298 | 1.59E-09 | -12.00 |
| 3MFK | Protein C-ets-1 | stromelysin-1 promoter(wt+4) | 298 | 6.10E-08 | -9.84 |
| 3MQ6 | SgraIR restriction enzyme | 18-1 DNA | 298 | 6.00E-10 | -12.57 |
| 3ODH | OkrAI | cognate DNA | 298 | 1.29E-08 | -10.76 |
| 3Q8L | Flap endonuclease 1 | DNA (5'-D(\*AP\*CP\*TP\*CP\*TP\*GP\*CP\*CP\*TP\*CP\*AP\*AP\*GP\*AP\*CP\*GP\*GP\*T)-3') | 330 | 1.00E-12 | -18.05 |
| 3QI5 | DNA-3-methyladenine glycosylase | DNA (5'-D(\*GP\*AP\*CP\*AP\*TP\*GP\*(EDC)P\*TP\*TP\*GP\*CP\*CP\*T)-3') | 310 | 1.30E-08 | -9.73 |
| 3S6I | DNA-3-methyladenine glycosylase 1 | (5'-D(\*TP\*GP\*TP\*CP\*CP\*AP\*(3DR)P\*GP\*TP\*CP\*T)-3') | 277 | 8.00E-07 | -7.70 |
| 3SPD | Aprataxin-like protein | DNA | 298 | 4.26E-06 | -7.32 |
| 3W6V | AdpA | DNA | 277 | 2.00E-06 | -7.22 |
| 3WTS | Protein C-ets-1 | dsDNA | 277 | 4.15E-10 | -11.89 |
| 4AIJ | Transcriptional regulator SlyA | upstream of promoter P2 | 310 | 4.60E-08 | -10.41 |
| 4B5F | Exodeoxyribonuclease III | Nape DNA with orphan cytosine base | 298 | 2.29E-07 | -9.05 |
| 4CJA | BurrH | BurrH | 298 | 2.50E-08 | -10.37 |
| 4EJY | 3-Methyladenine DNA glycosylase | dsDNA | 298 | 8.60E-10 | -12.36 |
| 4EOT | Transcription factor MafA | 19 bp DNA | 298 | 4.60E-10 | -12.73 |
| 4FLW | DNA polymerase 1 | Primer strand | 303 | 2.51E-07 | -9.00 |
| 4L0Z | Runt-related transcription factor 1 | TCR alpha enhancer | 298 | 3.30E-09 | -11.56 |
| 4LLL | MepR | Palindromized mepR operator sequence | 298 | 3.98E-08 | -10.36 |
| 4QTK | White-opaque regulator 1 | 20-bp DNA sequence | 277 | 1.33E-07 | -8.72 |
| 4S04 | DNA-binding transcriptional regulator BasR | DNA (25-MER) | 298 | 2.01E-07 | -9.15 |
| 4UX5 | Transcription factor MBP1 | dsDNA | 298 | 8.10E-07 | -8.31 |
| 4YIR | DNA repair protein RAD4 | undamaged DNA | 298 | 4.30E-07 | -8.68 |
| 4ZQ9 | Protein Rep68 | 41-mer DNA | 298 | 1.28E-07 | -9.40 |
| 5CO8 | Nuclease-like protein | dsDNA | 293 | 7.70E-09 | -10.88 |
| 5D8C | Uncharacterized HTH-type transcriptional regulator HI\_0186 | PadhC-estD/nmlR IR operator | 298 | 2.52E-08 | -10.36 |
| 5DWB | Type-2 restriction enzyme AgeI | NC | 298 | 5.00E-08 | -9.96 |
| 5ED4 | Possible two component system response transcriptional positive regulator PhoP | Pho-DNA | 298 | 1.92E-08 | -10.52 |
| 5GKG | Endonuclease NucS | GG mismatch DNA1 | 310 | 2.80E-09 | -12.13 |
| 5GNJ | Transcription factor MYC2 | G-box DNA | 293 | 8.00E-07 | -8.17 |
| 5HRL | DNA polymerase beta-like protein | DNA (5'-D(\*CP\*GP\*TP\*TP\*CP\*TP\*AP\*TP\*GP\*TP\*GP\*TP\*AP\*CP\*TP\*CP\*AP\*C)-3') | 298 | 3.70E-07 | -8.77 |
| 5JK0 | Tyrosine recombinase XerH | difH | 303 | 1.50E-07 | -9.46 |
| 5VFX | HTH\_56 domain-containing protein | oriT1 | 293 | 5.21E-07 | -8.42 |
| 5X11 | Transcriptional regulator | operator dsDNA | 298 | 8.30E-09 | -11.02 |
| 5YI2 | Zinc transport transcriptional regulator | DNA | 298 | 7.20E-08 | -9.74 |
| 5YZZ | B3 domain-containing transcription repressor VAL1 | 13-bp DNA | 298 | 6.13E-06 | -7.11 |
| 5ZKI | Nuclease EXOG, mitochondrial | DNA/DNA | 298 | 1.32E-06 | -8.02 |
| 5ZMO | HNHc domain-containing protein | PT-DNA | 298 | 1.46E-06 | -7.96 |
| 6ASB | F-box/LRR-repeat protein 19 | CpG | 298 | 4.30E-06 | -7.32 |
| 6C31 | Probable transcriptional regulatory protein | 13/8 DNA | 298 | 3.57E-07 | -8.79 |
| 6P0G | GTPase subunit of restriction endonuclease | m5C dsDNA mm | 303 | 6.16E-07 | -8.61 |
| 6R64 | 5-methylcytosine-specific restriction enzyme A | C5mCGG | 298 | 2.20E-09 | -11.81 |
| 6U7T | Gs MutY | OG:1N | 298 | 5.00E-10 | -12.68 |
| **Double III type** | | | | | |
| 1AZP | DNA-binding protein 7d | poly(dGdC)poly(dGdC) | 298 | 9.35E-07 | -8.22 |
| 1B3T | Epstein-Barr nuclear antigen 1 | EBNA1 recognition site 1 | 298 | 1.00E-09 | -12.27 |
| 1B72 | Homeobox protein Hox-B1 | ds DNA(TTAC) | 298 | 4.36E-06 | -7.31 |
| 1CKT | High mobility group protein B1 | Probe A | 281 | 2.27E-07 | -8.54 |
| 1CW0 | Very short patch repair protein | 35-mers (Dcm methylation site) | 310 | 2.30E-09 | -12.25 |
| 1DH3 | Cyclic AMP-responsive element-binding protein 1 | DNA | 298 | 2.00E-09 | -11.86 |
| 1E7J | High mobility group protein D | dsDNA | 303 | 1.00E-06 | -8.32 |
| 1F4K | Replication termination protein | DNA | 277 | 2.20E-11 | -13.51 |
| 1G9Z | DNA ENDONUCLEASE I-CREI | 5'-D(\*GP\*CP\*AP\*AP\*AP\*AP\*CP\*GP\*TP\*CP\*GP\*TP\*GP\*A)-3' | 310 | 2.00E-10 | -13.76 |
| 1GCC | Ethylene-responsive transcription factor 1A | GCC box(16 bp) | 298 | 1.67E-10 | -13.33 |
| 1GXP | PHOSPHATE REGULON TRANSCRIPTIONAL REGULATORY PROTEIN | 5'-D(\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*TP\* AP\*AP\*AP\*GP\*TP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*G)-3' | 310 | 1.50E-06 | -8.23 |
| 1HCR | DNA-invertase hin | Hin recombination site 26-bp(hixL) | 298 | 1.62E-08 | -10.62 |
| 1IG4 | Methyl-CpG-binding domain protein 1 | Methylated DNA | 298 | 4.75E-08 | -9.99 |
| 1IGN | DNA-binding protein RAP1 | DNA | 298 | 1.30E-11 | -14.84 |
| 1J5N | Non-histone chromosomal protein 6A | 98 bp DNA | 295 | 1.00E-08 | -10.80 |
| 1JE8 | Nitrate/nitrite response regulator protein NarL | 20-mer oligonucleotide | 298 | 1.50E-10 | -13.40 |
| 1LE8 | Mating-type protein ALPHA2 | 23-bp DNA duplex | 293 | 4.49E-08 | -9.85 |
| 1LO1 | Steroid hormone receptor ERR2 | DNA | 298 | 7.10E-09 | -11.11 |
| 1MDM | Protein C-ets-1 | SC1-27 | 298 | 1.10E-11 | -14.94 |
| 1MDY | Myoblast determination protein 1 | enhancer | 310 | 8.80E-09 | -11.43 |
| 1NK2 | Homeobox protein vnd | oligo 1 | 277 | 1.90E-10 | -12.32 |
| 1OCT | POU domain, class 2, transcription factor 1 | ds DNA | 273 | 3.50E-09 | -10.56 |
| 1OUQ | Cre recombinase | loxP DNA | 303 | 5.80E-08 | -10.00 |
| 1OWG | Integration Host Factor Alpha-subunit | Phage lambda H' site | 298 | 2.00E-09 | -9.13 |
| 1P51 | DNA-binding protein HU | TF1 protein site(hmU content) | 298 | 2.60E-08 | -10.34 |
| 1PP8 | 39 kDa initiator binding protein | 12 bp -SCS | 298 | 1.90E-07 | -9.16 |
| 1PUE | Transcription factor PU.1 | CRYSK | 295 | 2.52E-08 | -10.26 |
| 1PVI | PROTEIN (PVUII (E.C.3.1.21.4)) | DNA (5'-D(\*TP\*GP\*AP\*CP\*CP\*AP\*GP\*CP\*TP\*GP\*GP\*TP\*C)-3') | 310 | 1.80E-08 | -9.33 |
| 1R7M | Intron-encoded endonuclease I-SceI | 5'-D(\*CP\*AP\*CP\*GP\*CP\*TP\*AP\*GP\*GP\*GP\*AP\*TP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*TP\*AP\*AP\*TP\*AP\*C)-3' | 298 | 2.30E-09 | -11.74 |
| 1T2T | Intron-associated endonuclease 1 | Operator Site | 298 | 5.30E-09 | -11.28 |
| 1TC3 | Transposable element Tc3 transposase | TC3 | 313 | 1.75E-07 | -9.68 |
| 1TF3 | Transcription factor IIIA | 15 bp DNA | 318 | 2.00E-07 | -9.75 |
| 1TN9 | Transposase from transposon Tn916 | 13 bp duplex DNA | 303 | 3.06E-07 | -9.03 |
| 1TRO | Trp operon repressor | Operator DNA | 296 | 1.49E-10 | -13.31 |
| 1TW8 | Hinc II endonuclease | 5'-D(\*GP\*CP\*CP\*GP\*GP\*TP\*CP\*GP\*AP\*CP\*CP\*GP\*G)-3' | 298 | 3.35E-07 | -8.80 |
| 1VFC | Telomeric repeat-binding factor 2 | DNA duplex | 298 | 7.48E-07 | -8.35 |
| 1YUI | Transcription factor GAGA | 20-bp oligonucleotide h3/h4 GAGA | 277 | 5.30E-09 | -10.49 |
| 1ZQ3 | Homeotic protein bicoid | dsDNA | 298 | 4.28E-10 | -12.77 |
| 2ER8 | Regulatory protein LEU3 | 5'-D(\*CP\*CP\*CP\*GP\*GP\*TP\*AP\*CP\*CP\*GP\*GP\*G)-3' | 293 | 3.00E-08 | -10.05 |
| 2GLO | Brinker | oligos omb12 | 277 | 3.00E-08 | -9.54 |
| 2H3A | Antitoxin CcdA | 12bp | 298 | 1.49E-04 | -5.22 |
| 2KAE | MED-1 | dsDNA | 298 | 8.00E-09 | -11.04 |
| 2KDZ | MYB24 | dupex DNA | 298 | 1.24E-09 | -12.14 |
| 2KEI | Lactose operon repressor | O1 | 298 | 5.00E-11 | -14.05 |
| 2KKF | Histone-lysine N-methyltransferase HRX | 5'-D(\*CP\*CP\*CP\*TP\*GP\*CP\*GP\*CP\*AP\*GP\*GP\*G)-3' | 298 | 4.40E-05 | -5.94 |
| 2KMK | Zinc finger protein Gfi-1 | 16-mer | 298 | 2.60E-10 | -13.07 |
| 2KO0 | THAP domain-containing protein 1 | 16-bp rrm1 oligonucleotide | 298 | 4.80E-07 | -8.62 |
| 2LD5 | Homeobox protein Hox-A13 | DNA | 288 | 7.50E-09 | -10.71 |
| 2LEV | Ler | DNA (5'-D(\*GP\*CP\*GP\*AP\*TP\*AP\*AP\*TP\*TP\*GP\*AP\*TP\*AP\*GP\*G)-3') | 298 | 1.48E-07 | -8.12 |
| 2LKX | Pituitary homeobox 3 | DNA (5'-D(\*GP\*CP\*TP\*CP\*TP\*AP\*AP\*TP\*CP\*CP\*CP\*CP\*G)-3') | 298 | 6.60E-08 | -9.79 |
| 2MRU | Antitoxin MazE | DNA'a' | 298 | 5.60E-06 | -7.16 |
| 2MXF | MvaT | 5'-D(\*CP\*GP\*CP\*AP\*TP\*AP\*TP\*AP\*TP\*GP\*CP\*G)-3' | 277 | 1.07E-05 | -4.46 |
| 2NP2 | DNA-binding protein HU | WT Hbb | 277 | 5.50E-08 | -9.20 |
| 2O8K | RNA polymerase sigma factor RpoN | Non-template/template | 298 | 1.14E-07 | -9.47 |
| 2QL2 | Transcription factor E2-alpha | E47-NeuroD1-specific E2-box | 277 | 3.70E-09 | -10.69 |
| 2VS7 | Homing endonuclease I-DmoI | DNA | 298 | 1.09E-07 | -9.49 |
| 2W7N | TrfB transcriptional repressor protein | 18-bp DNA | 310 | 2.33E-08 | -10.83 |
| 3BS1 | Accessory gene regulator protein A | DNA (5'-D(\*DTP\*DTP\*DTP\*DAP\*DAP\*DCP\*DAP\*DGP\*DTP\*DTP\*DAP\*DAP\*DGP\*(BRU)P\*DAP\*DT)-3') | 298 | 8.00E-08 | -9.68 |
| 3CLC | Regulatory protein | 35-MER | 353 | 5.10E-10 | -14.96 |
| 3CO6 | Forkhead box protein O1 | DBE1 DNA | 277 | 6.00E-08 | -9.15 |
| 3CRO | Regulatory protein cro | 14mer | 298 | 2.00E-07 | -9.13 |
| 3EH8 | Intron-encoded DNA endonuclease I-AniI | 20 base pair | 298 | 1.00E-06 | -8.18 |
| 3FD2 | Site-specific DNA endonuclease I-MsoI | dsDNA | 303 | 4.02E-08 | -10.25 |
| 3GX4 | Alkyltransferase-like protein 1 | O6-mG | 298 | 3.50E-10 | -12.89 |
| 3IMB | R.BcnI | 5'-D(\*GP\*TP\*CP\*CP\*GP\*GP\*GP\*CP\*G)-3' | 298 | 5.00E-10 | -12.68 |
| 3KDE | Transposable element P transposase | 5'-D(\*GP\*TP\*TP\*AP\*AP\*GP\*(BRU)P\*GP\*GP\*A)-3' | 310 | 1.60E-07 | -9.27 |
| 3LWH | Chromatin protein Cren7 | DNA (5'-D(\*GP\*TP\*AP\*AP\*TP\*TP\*AP\*C)-3') | 298 | 1.34E-07 | -7.76 |
| 3O9X | Antitoxin MqsA | mqsRA DNA | 298 | 8.00E-10 | -12.40 |
| 3PVP | Chromosomal replication initiator protein DnaA | DnaA-box DNA | 298 | 6.00E-08 | -9.85 |
| 3ZI5 | Restriction endonuclease | dna | 298 | 2.00E-09 | -11.86 |
| 3ZKC | HTH-type transcriptional regulator SinR | Inverted DNA | 298 | 3.60E-04 | -4.70 |
| 4ATI | Microphthalmia-associated transcription factor | M-Box | 298 | 2.50E-09 | -11.73 |
| 4EVV | Methyl-CpG-binding domain protein 4 | DNA (5'-D(\*TP\*CP\*AP\*GP\*CP\*GP\*CP\*AP\*TP\*GP\*G)-3') | 298 | 8.30E-06 | -6.90 |
| 4FTH | Transcriptional regulator | DNA | 298 | 2.00E-07 | -9.13 |
| 4G92 | HapE | cycAp30 | 298 | 3.30E-09 | -11.56 |
| 4H10 | Aryl hydrocarbon receptor nuclear translocator-like protein 1 | 16-bp DNA | 298 | 1.52E-06 | -7.93 |
| 4HF1 | HTH-type transcriptional regulator IscR | hya | 298 | 1.70E-08 | -10.59 |
| 4KIS | Putative integrase [Bacteriophage A118] | DNA (26-MER) | 294 | 1.90E-05 | -6.33 |
| 4M9V | Zinc finger protein 57 | DNA (5'-D(\*TP\*AP\*TP\*TP\*GP\*CP\*(5CM)P\*GP\*CP\*AP\*G)-3') | 298 | 8.00E-08 | -11.04 |
| 4NHJ | Response regulator | DNA-22 | 298 | 5.15E-06 | -7.21 |
| 4NNU | Transcription factor A, mitochondrial | Non-specific DNA | 298 | 7.40E-09 | -11.09 |
| 4QLC | Histone H5 | 167bpDNA | 298 | 3.50E-07 | -8.80 |
| 4RKH | E3 ubiquitin-protein ligase msl-2 | DNA (5'-D(\*AP\*TP\*GP\*AP\*GP\*CP\*GP\*AP\*GP\*AP\*TP\*GP\*GP\*AP\*T)-3') | 298 | 3.64E-06 | -7.58 |
| 4XR0 | DNA replication terminus site-binding protein | forked TerB | 293 | 2.12E-10 | -12.97 |
| 5A77 | DNA endonuclease I-CvuI | Sro1.3 | 298 | 1.96E-08 | -10.51 |
| 5D4S | Arabinose metabolism transcriptional repressor | ORX1 | 298 | 2.77E-07 | -8.94 |
| 5E5O | Uncharacterized protein | dna | 298 | 1.50E-07 | -9.30 |
| 5FD3 | Protein lin-54 homolog | CHR13 | 298 | 2.75E-06 | -7.58 |
| 5HLG | Transcriptional regulator, MarR family | DNA | 295 | 3.00E-07 | -8.80 |
| 5JH0 | ARS-binding factor 2, mitochondrial | AF2\_22 | 298 | 1.66E-07 | -9.24 |
| 5VMU | Transcriptional regulator Kaiso | MeKBS | 298 | 2.00E-10 | -13.23 |
| 5X5L | AdeR | DNA | 298 | 2.20E-06 | -7.71 |
| 5YJ3 | Telomere zinc finger-associated protein | DNA | 293 | 1.70E-07 | -9.08 |
| 5ZUX | Repressor Rok | Seq1 DNA | 277 | 7.90E-09 | -10.27 |
| 6AKO | Forkhead box protein C2 | DBE2 | 298 | 7.90E-07 | -8.32 |
| 6JNL | Lysine-specific demethylase REF6 | NAc004P | 298 | 7.35E-08 | -9.73 |
| 7GAT | AREA DBD | dsDNA | 298 | 3.10E-06 | -7.51 |
| **MISC type** | | | | | |
| 1EYG | Single-stranded DNA-binding protein | dC(pC)34 | 298 | 1.05E-08 | -10.88 |
| 1QBJ | PROTEIN (DOUBLE-STRANDED RNA SPECIFIC ADENOSINE DEAMINASE (ADAR1)) | DNA (5'-D(\*TP\*CP\*GP\*CP\*GP\*CP\*G)-3') | 298 | 1.86E-07 | -9.18 |
| 1SFU | 34L protein | dsDNA | 298 | 8.72E-08 | -9.63 |
| 1XSL | DNA polymerase lambda | 5'-D(\*CP\*GP\*GP\*CP\*AP\*GP\*CP\*GP\*CP\*AP\*C)-3' | 310 | 3.30E-06 | -7.75 |
| 2LTT | PC4 domain-containing protein | ssDNA | 298 | 1.60E-08 | -10.63 |
| 3EYI | Z-DNA-binding protein 1 | dsDNA | 298 | 4.94E-07 | -8.60 |
| 4B20 | Endonuclease V | Undamaged DNA | 298 | 1.10E-07 | -9.49 |
| 4L5S | Interferon-activable protein 202 | 12-mer dsDNA | 298 | 7.10E-06 | -7.02 |
| 4NI7 | Interleukin-6 | SL1025 | 298 | 2.00E-10 | -13.23 |
| 4WCG | Protein ORF112 | T(CG)3 | 298 | 3.93E-07 | -8.73 |
| 5ZMD | Alpha-ketoglutarate-dependent dioxygenase FTO | FAM-6mA-DNA | 310 | 3.69E-06 | -7.71 |
| 1CKQ | PROTEIN (ENDONUCLEASE) | DNA (5'-D(\*TP\*CP\*GP\*CP\*GP\*AP\*AP\*TP\*TP\*CP\*GP\*CP\*G)-3') | 294 | 1.04E-10 | -15.71 |
| 1DP7 | MHC class II regulatory factor RFX1 | EF-C/MDBP target site(13bp biotinylated DNA) | 298 | 1.53E-07 | -9.29 |
| 1J5K | Heterogeneous nuclear ribonucleoprotein K | DNA | 308 | 3.00E-06 | -7.78 |
| 1J75 | Z-DNA-binding protein 1 | DNA | 283 | 3.50E-08 | -9.65 |
| 1KFS | PROTEIN (DNA POLYMERASE I KLENOW FRAGMENT (E.C.2.7.7.7)) | DNA (5'-D(\*GP\*CP\*TP\*TP\*AP\*CP\*G)-3') | 310 | 2.10E-09 | -12.27 |
| 1LAU | Uracil-DNA glycosylase | HAP-G DNA | 298 | 8.10E-06 | -6.94 |
| 1MWJ | G/U mismatch-specific DNA glycosylase | 5'-D(\*CP\*GP\*CP\*GP\*A\*GP\*(DU)P\*TP\*CP\*GP\*CP\*G)-3' | 25 | 1.86E-07 | -0.77 |
| 1OMH | TrwC | dsDNA | 293 | 7.00E-08 | -9.59 |
| 1OTC | Telomere-binding protein subunit alpha | ssDNA | 293 | 9.00E-08 | -9.45 |
| 1PO6 | Heterogeneous nuclear ribonucleoprotein A1 | TR2-6F | 298 | 1.57E-07 | -9.28 |
| 1QPZ | HTH-type transcriptional repressor PurR | F-ApC | 298 | 2.60E-09 | -11.71 |
| 1S40 | Cell division control protein 13 | DNA | 298 | 3.40E-10 | -12.91 |
| 1XJV | Protection of telomeres protein 1 | ssDNA | 298 | 9.20E-09 | -10.96 |
| 2A0I | TraI protein | F plasmid single-stranded oriT DNA | 100 | 4.30E-10 | -12.77 |
| 2BOP | Regulatory protein E2 | ssDNA | 296 | 1.69E-07 | -9.17 |
| 2CCZ | PRIMOSOMAL REPLICATION PROTEIN N | 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*TP \*TP\*TP\*TP\*TP\*T)-3' | 298 | 1.00E-07 | -9.54 |
| 2FD8 | Alkylated DNA repair protein alkB | 5'-D(P\*TP\*(MA7)P\*T)-3' | 297 | 4.10E-09 | -11.36 |
| 2HAX | Cold shock protein cspB | 5'-D(\*TP\*TP\*TP\*TP\*TP\*T)-3' | 298 | 4.00E-08 | -10.05 |
| 2L4L | Gag-Pol polyprotein | mini-cTAR | 293 | 1.60E-06 | -7.77 |
| 2MNA | Single-stranded DNA binding protein (SSB) | DNA | 298 | 1.70E-07 | -9.20 |
| 2N8A | Poly ADP-ribose polymerase 1 | gapped DNA | 298 | 9.50E-08 | -9.58 |
| 2Q2K | Conserved domain protein | 20 bp DNA | 298 | 1.75E-08 | -10.58 |
| 2VY1 | PROTEIN LEAFY | 5'-D(\*TP\*TP\*AP\*CP\*GP\*GP\*AP\*CP\*CP\*AP \*CP\*TP\*GP\*GP\*TP\*CP\*CP\*TP\*TP\*CP)-3' | 277 | 9.50E-08 | -8.90 |
| 2VYE | REPLICATIVE DNA HELICASE | 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*TP)-3' | 293 | 4.69E-08 | -9.83 |
| 3D2W | TAR DNA-binding protein 43 | TAR32 | 298 | 1.38E-07 | -9.35 |
| 3D6Y | Multidrug-efflux transporter 1 regulator | 22-bp DNA | 308 | 1.03E-05 | -7.03 |
| 3H15 | Protein MCM10 homolog | ssDNA | 298 | 3.40E-06 | -7.46 |
| 3HXQ | von Willebrand factor | ARC1172 | 298 | 6.60E-10 | -12.52 |
| 3HZI | Antitoxin HipB | DNA | 298 | 1.50E-05 | -6.58 |
| 3LSP | DesT | DNA (5'-D(\*TP\*TP\*AP\*CP\*AP\*TP\*CP\*AP\*GP\*TP\*GP\*AP\*AP\*CP\*GP\*CP\*TP\*TP\*GP\*TP\*TP\*GP\*AP\*CP\*TP\*CP\*GP\*AP\*TP\*TP\*G)-3') | 295 | 9.30E-07 | -8.11 |
| 3PIH | UvrABC system protein A | Unmodified Palindromic 32-mer DNA | 298 | 1.90E-09 | -11.89 |
| 3R8F | Chromosomal replication initiator protein DnaA | Duplex DNA | 298 | 1.10E-07 | -9.49 |
| 3THW | DNA mismatch repair protein Msh3 | Loop4 | 298 | 5.00E-11 | -14.05 |
| 3UPU | ATP-dependent DNA helicase dda | 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*T)-3' | 298 | 3.40E-08 | -10.18 |
| 3VOK | Transcriptional regulator | 19-bp DNA | 298 | 2.00E-10 | -13.23 |
| 3WPD | Toll-like receptor 9 | DNA (5'-D(\*CP\*CP\*TP\*GP\*GP\*AP\*TP\*GP\*GP\*G)-3') | 298 | 3.00E-09 | -11.58 |
| 4A15 | ATP-DEPENDENT DNA HELICASE TA0057 | 5'-D(\*DTP\*AP\*CP\*GP)-3' | 310 | 6.10E-08 | -9.84 |
| 4A75 | Uncharacterized protein | dT6 | 293 | 1.69E-07 | -9.08 |
| 4CH1 | PROTEIN SUP-12, ISOFORM B | GGTGTGC | 298 | 2.49E-07 | -9.52 |
| 4HIK | Protection of telomeres protein 1 | 9mer | 298 | 2.40E-08 | -10.39 |
| 4HQB | Single-stranded DNA-binding protein DdrB | 20bp dT | 298 | 3.60E-06 | -7.42 |
| 4HQU | Platelet-derived growth factor subunit B | SL1 | 310 | 2.00E-11 | -15.18 |
| 4HT4 | Nicking enzyme | DNA | 298 | 1.82E-08 | -10.55 |
| 4J1J | Nucleoprotein | ssDNA | 293 | 5.05E-07 | -8.44 |
| 4KB1 | Ribonuclease T | STEM-LOOP DNA | 298 | 1.61E-07 | -9.26 |
| 4KMF | Interferon-inducible and double-stranded-dependent eIF-2kinase | DNA (5'-D(\*TP\*CP\*GP\*CP\*GP\*CP\*G)-3') | 298 | 8.51E-07 | -8.28 |
| 4LJR | DNA processing chain A | DNA | 298 | 3.15E-08 | -10.19 |
| 4OWX | Integrator complex subunit 3 | dT12 | 298 | 8.20E-06 | -6.94 |
| 4R22 | HTH-type transcriptional regulator TnrA | 21-mer | 298 | 1.16E-08 | -10.82 |
| 4S0N | Helicase-like transcription factor | 40-mer ssDNA | 298 | 1.30E-08 | -10.75 |
| 4YGI | Histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH5 | Fully 5hmCG | 298 | 1.50E-06 | -7.94 |
| 4ZSF | BsaWI endonuclease | DNA | 298 | 2.90E-06 | -11.64 |
| 5ITH | Nucleolysin TIA-1 isoform p40 | DNA | 298 | 3.00E-08 | -10.26 |
| 5UC6 | Interleukin-1 alpha | DNA | 298 | 7.30E-09 | -11.09 |
| 6BWY | DNA dC->dU-editing enzyme APOBEC-3G | oligoDNA | 310 | 2.60E-05 | -6.50 |
| 6CRM | DNA helicase | G4 binding | 298 | 4.70E-06 | -7.26 |
| 6EO7 | Prothrombin | Thrombin-binding aptamer | 298 | 2.80E-09 | -11.66 |
| 6FWR | ATP-dependent DNA helicase DinG | ssDNA | 298 | 3.10E-07 | -8.87 |
| 6KBS | Abasic site processing protein | ssDNA | 298 | 2.20E-06 | -7.71 |

**Table S2**. Detailed information of the independent testing dataset PD36

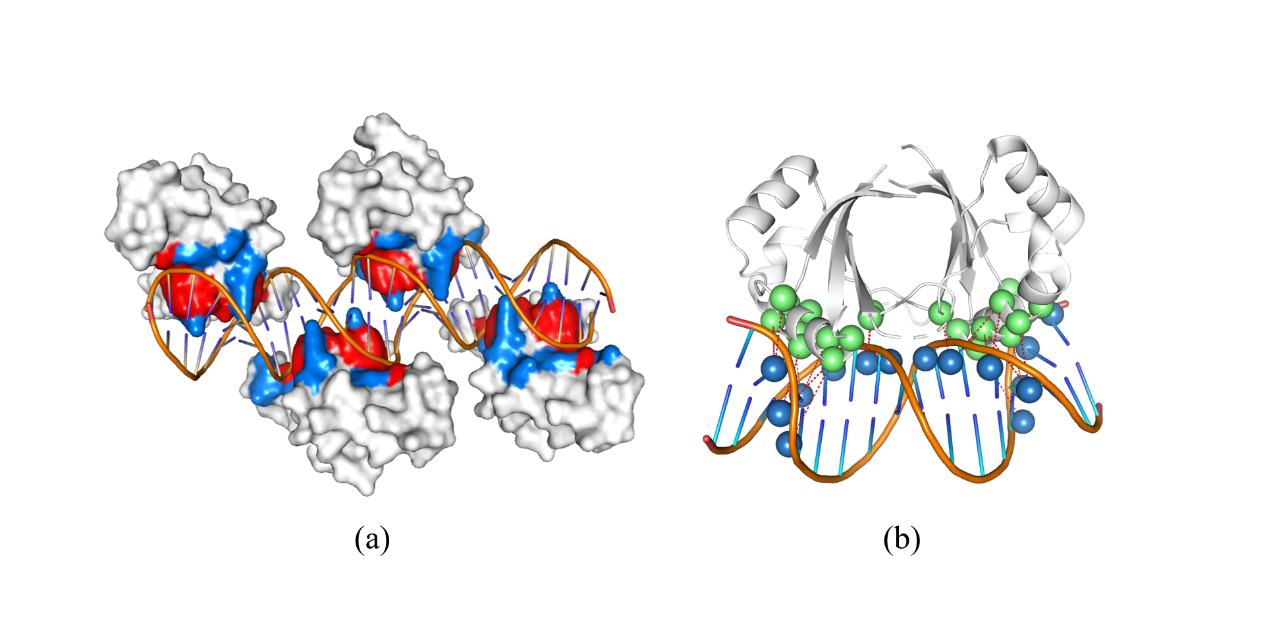
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PDB ID** | **Protein name** | **DNA name** | ***T* (K)** | ***Kd* (mol)** | **△*G* (kcal/mol)** |
| **Double I type** | | | | | |
| 5MPF | Grainyhead-like protein 1 homolog, Grhl1 DBD | 12-mer | 277 | 9.00E-08 | -8.93 |
| 5ZD4 | mMBP-A-BIL1/BZR1 DBD | dsDNA containing a G-box motif (CACGTG) | 298 | 1.20E-08 | -10.80 |
| **Double II type** | | | | | |
| 1MHD | SMAD3 MH1 DOMAIN | DNA | 277 | 1.14E-07 | -8.80 |
| 3C58 | wt-LlFpg | N7-Benzyl-cFapydG-DNA | 277 | 9.00E-10 | -11.46 |
| 3W3C | Virulence regulon transcriptional activator VirB, VirB core domain (129-250) | cis-acting site upstream icsb DNA promoter | 288 | 1.60E-08 | -10.27 |
| 3ZPL | Sco3205, a MarR family transcriptional regulator from Streptomyces coelicolor | 22-mer dsDNA | 293 | 2.40E-09 | -11.56 |
| 4KUD | N-terminal acetylated Sir3 BAH domain D205N mutant | yeast nucleosome core particle | 298 | 4.00E-08 | -10.09 |
| 4M8B | WOPR domain of Wor1 (white-opaque regulator 1) | 20-mer dsDNA containing the cis-regulatory element | 298 | 2.20E-06 | -7.71 |
| 4QPQ | Relaxosome protein TraM | 28-mer sbmA dsDNA | 277 | 2.00E-07 | -8.49 |
| 4UZB | KSHV LANA (ORF73) C-terminal domain mutant (R1039Q, R1040Q, K1055E, K1109A, D1110A, A1121E, K1138S, K1140D, K1141D) | LBS1 DNA | 298 | 7.20E-08 | -9.74 |
| 5I50 | his-tagged OmoMYC | E-box oligonucleotide | 277 | 3.91E-07 | -8.12 |
| 5T1J | Tbox DNA binding domain of the transcription factor T-bet | 24-bp ds TAMRA-DNA | 298 | 1.90E-08 | -10.53 |
| 5UI5 | Aquifex aeolicus sigmaN (Aae Delta-N sigma N) | DNA fragment containing the Aae dhsU promoter | 295 | 1.03E-07 | -9.43 |
| 5YEJ | TetR family transcriptional regulator, BioQ | naturel double-stranded DNA operator | 298 | 1.38E-08 | -10.72 |
| **Double III type** | | | | | |
| 1BDT | WILD TYPE GENE-REGULATING PROTEIN ARC | DNA | 298 | 1.20E-11 | -14.89 |
| 1JJ4 | Regulatory protein E2 | E2 binding site DNA | 298 | 1.60E-09 | -11.99 |
| 2AOQ | DNA mismatch repair protein mutH | DNA sequence gCATGATCATGc | 277 | 3.50E-06 | -6.92 |
| 2IRF | INTERFERON REGULATORY FACTOR 2 | DNA | 277 | 1.70E-07 | -8.58 |
| 2RBF | Bifunctional protein putA(1-52) | operator DNA O2(operator 2, 21-bp) | 298 | 2.10E-07 | -9.11 |
| 3GXQ | Putative regulator of transfer genes ArtA | 11-mer | 298 | 1.27E-07 | -9.40 |
| 4ON0 | SeMet NolR from Sinorhizobium fredii | 22-bp oligo AA DNA duplexes | 277 | 3.60E-07 | -8.17 |
| 4RBO | Nanog homeobox (NANOG) from Homo sapiens | 12-mer OCT4 Promoter DNA | 298 | 5.90E-06 | -7.13 |
| 4XIC | Homeotic protein antennapedia, ANTPHD | 15bp di-thioate modified DNA duplex | 298 | 2.90E-09 | -11.64 |
| 4Y60 | Transcription factor SOX18-HMG box | PROX1-DNA | 298 | 6.10E-09 | -11.20 |
| 5EGB | hPRDM9A (hPRDM9, allele A) ZnF8-12 | THE1B recombination hot spot sequence II | 298 | 3.40E-08 | -10.18 |
| 5GZB | Transcription Factor TEAD4 | M-CAT DNA | 298 | 2.30E-07 | -9.05 |
| 5I44 | Bacillus subtilis RacA DBD (1-70) | 14-mer palindromic centromere DNA | 298 | 9.20E-06 | -6.87 |
| 5KKQ | Homo sapiens CCCTC-binding factor (CTCF) ZnF3-7 | oligo DNA containing the CORE sequence | 298 | 9.00E-09 | -10.97 |
| 5Z6Z | human DUX4 homeodomains, DUX4-DH | DNA1 | 298 | 4.52E-08 | -10.01 |
| 6AMK | Streptomyces venezuelae BldC | whiI opt DNA | 298 | 1.65E-08 | -10.61 |
| 6E93 | Zinc finger and BTB domain-containing protein 38, ZBTB38 ZF 6-9 | mCZ38BS DNA | 298 | 5.00E-09 | -11.32 |
| **MISC type** | | | | | |
| 1EVW | I-PPOI HOMING ENDONUCLEASE | DNA | 298 | 1.00E-11 | -15.00 |
| 2NQ9 | Endo IV Y72A mutant | AP DNA substrate | 298 | 9.70E-08 | -9.56 |
| 3ZH2 | Plasmodium falciparum lactate dehydrogenase | a 35-mer DNA aptamer | 298 | 4.20E-08 | -10.06 |
| 4ZBN | Beta-nerve growth factor, NGF | 28-mer SOMAmer DNA SLN-1049 | 310 | 2.10E-10 | -13.73 |
| 5HRT | mouse autotaxin (ATX) | anti-ATX aptamer RB011 | 310 | 1.30E-09 | -12.60 |

**Table S3**. Features selected for each type of protein-DNA complexes by Sequential Forward Selection algorithm

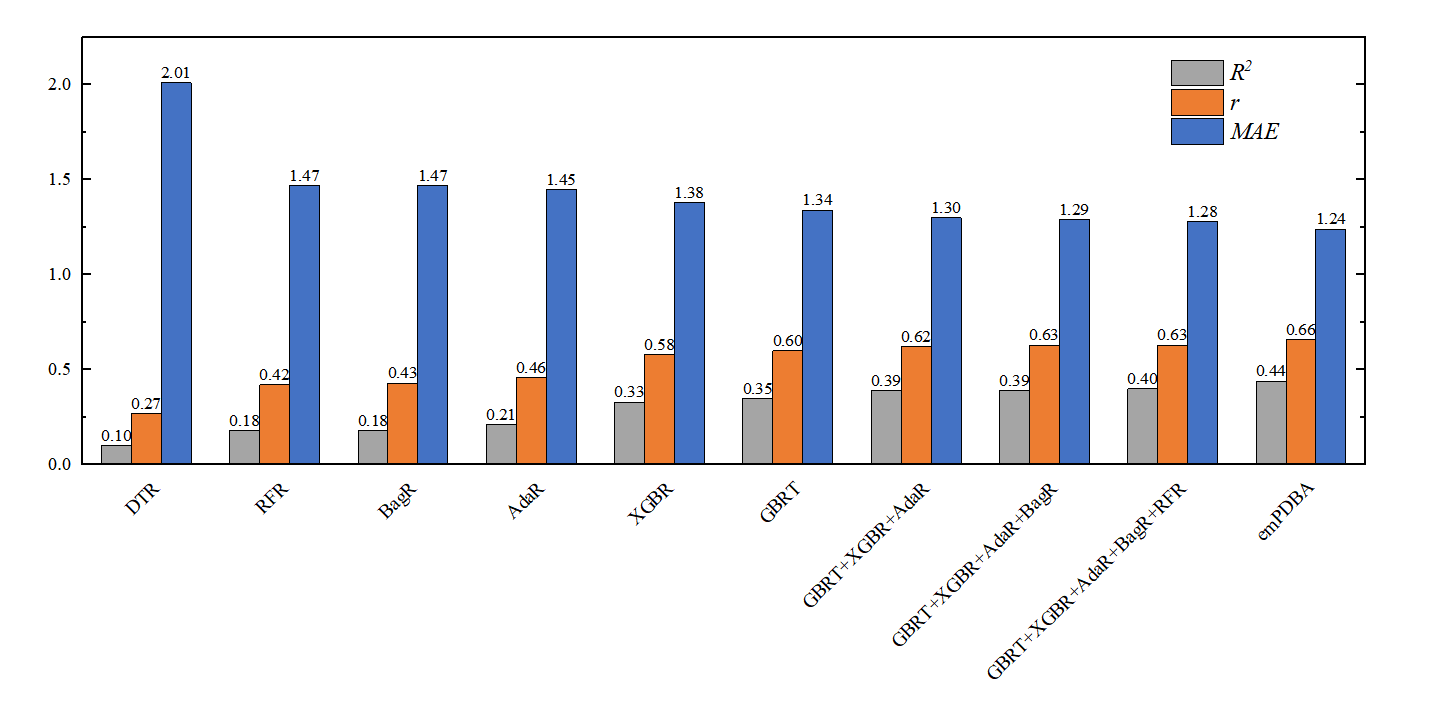
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **Double I** | **Double II** | **Double III** | **MISC** |
| % of the negatively charged residues in the protein | √ |  |  |  |
| % of the β-bridge residues in the protein | √ |  |  |  |
| the amount of π-helix residues in the protein | √ |  |  |  |
| % of π-helix residues in the protein | √ |  |  |  |
| the amount of turn residues in the protein | √ |  |  |  |
| the amount of bend residues in the protein | √ |  |  |  |
| % of positively charged residues in the binding sites | √ |  |  |  |
| interface area | √ | √ |  |  |
| vdW repulsive energy | √ | √ | √ | √ |
| average degree centrality | √ |  |  |  |
| the amount of positively charged residues in the protein |  | √ |  |  |
| % of polar uncharged residues in the binding sites |  | √ |  |  |
| the amount of T in the binding sites |  | √ |  |  |
| electrostatic short-range attractive energy |  | √ |  |  |
| the volume of DNA |  | √ |  |  |
| the reaction field energy of DNA |  | √ | √ |  |
| % of polar uncharged residues in the protein |  |  | √ |  |
| the amount of α-helix residues in the protein |  |  | √ |  |
| the amount of α-helices in the protein |  |  | √ |  |
| the mass of α-helices in the protein |  |  | √ |  |
| the mass of DNA |  |  | √ |  |
| the amount of interface hydrogen bonds |  |  | √ |  |
| residue-nucleotide pairwise potential |  |  | √ |  |
| the change of reaction field energy of binding monomers and complex |  |  | √ |  |
| the amount of nonpolar residues in the protein |  |  |  | √ |
| % of 310-helix residues in the protein |  |  |  | √ |
| % of nonpolar residues in the binding sites |  |  |  | √ |
| the amount of negatively charged residues in the binding sites |  |  |  | √ |
| % of A in the binding sites |  |  |  | √ |
| % of C in the binding sites |  |  |  | √ |
| electrostatic short-range repulsive energy |  |  |  | √ |
| average closeness centrality |  |  |  | √ |
| the volume of complex |  |  |  | √ |
| the amount of dinucleotide CA | √ |  |  |  |
| the amount of dinucleotide CG |  | √ |  |  |
| the amount of dinucleotide AC |  |  |  | √ |
| the amount of dinucleotide AA |  | √ |  |  |
| the amount of dinucleotide AT |  | √ |  |  |
| % of dinucleotide AT |  | √ | √ |  |
| the amount of dinucleotide TT |  | √ |  |  |
| the amount of dinucleotide TC |  |  | √ |  |
| % of dinucleotide TA |  |  | √ | √ |
| the amount of dinucleotide TA | √ | √ |  |  |
| the amount of dinucleotide GA |  |  | √ |  |
| % of dinucleotide GA |  |  |  | √ |
| the amount of dinucleotide TG |  |  |  | √ |
| % of dinucleotide CA |  |  | √ | √ |
| % of dinucleotide GC |  |  | √ |  |
| the amount of dinucleotide GC | √ |  |  |  |

\*Only features selected out by the feature selection procedure are displayed in the table.

**Figure captions**



**Figure S1**. Representative structures for the Double II and Double III types of protein-DNA complexes from the independent testing dataset. (a) Complex (PDB ID 4QPQ) from the Double II type with the percentage of polar uncharged amino acids in the binding interface reaches 44.16%. The polar uncharged residues in the binding area are shown in red, and the other binding residues are in blue. The experimental and predicted Δ*G* are -8.49 kcal/mol and -8.59 kcal/mol, respectively; (b) Complex (PDB ID 1JJ4) from the Double III type with protein helical motifs recognizing major grooves to form extensive hydrogen bonds with nucleotide bases. The interfacial hydrogen bonds are shown in red lines, and the residues and nucleotides involved in the hydrogen bonds are shown in green and blue spheres respectively. The experimental and predicted Δ*G* are -11.99 kcal/mol and -9.84 kcal/mol, respectively.



**Figure S2**. Prediction performance comparison of emPDBA with single regression models and their combinations on the training dataset via five-fold cross-validation.