**Table S1. List of ids of crystal structures of B-DNA only (without any protein/ligand) obtained from NDB**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No. | NDB ids | Sr No. | NDB ids |
|  | 111d.pdb | 38. | 250d.pdb |
|  | 112d.pdb | 39. | 251d.pdb |
|  | 113d.pdb | 40. | 252d.pdb |
|  | 122d.pdb | 41. | 272d.pdb |
|  | 123d.pdb | 42. | 287d.pdb |
|  | 126d.pdb | 43. | 2af1.pdb |
|  | 167d.pdb | 44. | 2bna.pdb |
|  | 194d.pdb | 45. | 2g1z.pdb |
|  | 1bna.pdb | 46. | 2got.pdb |
|  | 1cgc.pdb | 47. | 2org.pdb |
|  | 1d28.pdb | 48. | 2qef.pdb |
|  | 1d29.pdb | 49. | 2qeg.pdb |
|  | 1d31.pdb | 50. | 307d.pdb |
|  | 1d65.pdb | 51. | 309d.pdb |
|  | 1d89.pdb | 52. | 329d.pdb |
|  | 1d98.pdb | 53. | 330d.pdb |
|  | 1d99.pdb | 54. | 3ajj.pdb |
|  | 1dcv.pdb | 55. | 3bna.pdb |
|  | 1dn9.pdb | 56. | 3gbi.pdb |
|  | 1dnm.pdb | 57. | 3l1q.pdb |
|  | 1ehv.pdb | 58. | 3n4n.pdb |
|  | 1hq7.pdb | 59. | 3nao.pdb |
|  | 1ilc.pdb | 60. | 3q5c.pdb |
|  | 1n4e.pdb | 61. | 3ubi.pdb |
|  | 1ndn.pdb | 62. | 424d.pdb |
|  | 1vtd.pdb | 63. | 425d.pdb |
|  | 1vte.pdb | 64. | 458d.pdb |
|  | 1wqy.pdb | 65. | 4b8d.pdb |
|  | 1wqz.pdb | 66. | 4bna.pdb |
|  | 1zf5.pdb | 67. | 4glc.pdb |
|  | 1zfc.pdb | 68. | 4glg.pdb |
|  | 1zfe.pdb | 69. | 4gs2.pdb |
|  | 1zff.pdb | 70. | 4gsg.pdb |
|  | 1zfg.pdb | 71. | 4gsi.pdb |
|  | 1zfh.pdb | 72. | 4j2i.pdb |
|  | 1zfm.pdb | 73. | 7bna.pdb |
|  | 237d.pdb | 74. | 9bna.pdb |

**Table S2: Values for structural and energy parameters for the unique dinucleotide steps**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AA/TT | AT | AC/GT | AG/CT | TA | CG | CC/GG | GA/TC | CA/TG | GC |
| X Disp. | -0.246 | -0.178 | -0.031 | 0.486 | 0.102 | 0.705 | 0.585 | -0.158 | 0.312 | 0.593 |
| Y Disp. | 0.004 | -0.026 | -0.051 | -0.029 | -0.286 | 0.060 | -0.078 | 0.080 | -0.068 | 0.083 |
| Inclination | -0.614 | -0.260 | 0.680 | 1.825 | 0.547 | 2.911 | 9.909 | 0.526 | 3.005 | 0.365 |
| Tip | 0.597 | 0.461 | 2.510 | 3.723 | -0.766 | 0.303 | 3.128 | 0.752 | 0.175 | 2.507 |
| Ax-Bend | 1.422 | 1.213 | 1.250 | 1.652 | 1.471 | 1.489 | 2.586 | 1.178 | 1.055 | 1.244 |
| Shear | 0.128 | 0.014 | 0.055 | -0.064 | 0.026 | 0.355 | 1.224 | 0.050 | -0.002 | 0.021 |
| Stretch | -0.180 | -0.134 | -0.125 | -0.168 | -0.177 | -0.358 | -1.123 | -0.199 | -0.149 | -0.137 |
| Stagger | 0.121 | 0.147 | 0.573 | -0.146 | 0.085 | 0.561 | 1.670 | 0.215 | 0.156 | 0.284 |
| Buckle | -1.482 | 0.279 | -5.579 | -0.850 | -0.326 | 2.413 | 8.842 | -1.169 | 0.076 | -0.983 |
| Propel | -15.308 | -16.536 | -9.838 | -11.734 | -14.037 | -3.280 | 5.707 | -13.135 | -10.824 | -10.443 |
| Opening | 3.159 | 4.110 | 2.110 | -3.668 | 3.305 | -0.849 | 5.719 | 0.943 | 2.073 | -1.116 |
| Shift | -0.110 | -0.060 | -0.120 | -0.270 | -0.050 | 0.240 | 0.380 | 0.040 | -0.070 | -0.280 |
| Slide | -0.200 | -0.440 | 0.730 | 0.210 | 0.370 | 0.680 | 0.710 | -0.200 | 0.520 | 0.170 |
| Rise | 3.250 | 3.310 | 3.660 | 3.020 | 3.390 | 3.330 | 2.960 | 3.320 | 3.290 | 3.350 |
| Tilt | 0.630 | -0.900 | -4.750 | 3.690 | -2.710 | 1.750 | -1.620 | 2.210 | -3.270 | -2.680 |
| Roll | -0.080 | -2.620 | -0.600 | -7.270 | 1.740 | 4.290 | -2.550 | 2.690 | 1.540 | -5.280 |
| Twist | 35.670 | 33.420 | 34.790 | 31.000 | 32.050 | 37.380 | 29.490 | 37.740 | 36.740 | 31.090 |
| H-Rise | 3.270 | 3.330 | 2.970 | 3.060 | 3.410 | 3.330 | 2.730 | 3.330 | 3.210 | 3.290 |
| H-Twist | 35.560 | 33.680 | 37.600 | 19.060 | 32.000 | 39.100 | 31.570 | 37.810 | 37.250 | 30.560 |
| Alpha | -55.550 | -51.913 | -45.474 | -36.214 | -50.820 | -43.178 | -38.322 | -42.611 | -52.023 | -48.267 |
| Beta | 51.044 | 59.981 | 42.118 | 36.155 | 29.230 | 57.555 | 39.807 | 54.835 | 41.346 | 36.968 |
| Gamma | 50.772 | 47.175 | 51.861 | 43.870 | 53.740 | 51.502 | 44.997 | 53.374 | 35.389 | 36.646 |
| Delta | 128.408 | 125.847 | 126.878 | 129.955 | 127.846 | 132.616 | 129.998 | 131.432 | 131.452 | 129.951 |
| Epsilon | -5.924 | -31.230 | -34.321 | -43.587 | -37.772 | -67.013 | -56.716 | -8.081 | -75.781 | -70.502 |
| Zeta | -96.656 | -93.609 | -55.767 | -34.570 | -61.970 | -62.643 | -61.650 | -94.004 | -60.380 | -63.978 |
| Chi | -110.254 | -113.738 | -116.547 | -88.341 | -106.740 | -106.586 | -103.913 | -110.983 | -91.503 | -109.979 |
| Phase | 118.784 | 107.855 | 106.334 | 73.080 | 119.920 | 103.300 | 96.172 | 118.188 | 73.933 | 84.877 |
| Amplitude | 38.996 | 40.016 | 38.650 | 38.941 | 37.824 | 37.951 | 37.361 | 37.945 | 38.744 | 38.594 |
| Hydrogen Bonding Energy | -5.440 | -5.350 | -7.140 | -6.270 | -5.830 | -8.050 | -8.480 | -7.800 | -7.010 | -8.720 |
| Stacking Energy | -26.710 | -27.200 | -27.730 | -26.890 | -26.900 | -27.930 | -26.280 | -26.780 | -27.150 | -28.130 |
| Solvation Energy | -171.840 | -173.700 | -171.110 | -174.930 | -174.350 | -176.880 | -166.760 | -167.600 | -179.010 | -165.580 |

**Supplementary Table S3: Present as separate file in excel format**

**Supplementary Table S4: Parameter index**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No. | Poperty Name | unit | Class | Index |
| 1 | X Disp. | Å | BP AXIS | a |
| 2 | Y Disp. | Å | BP AXIS | b |
| 3 | Inclination | ° | BP AXIS | c |
| 4 | Tip | ° | BP AXIS | d |
| 5 | Ax-Bend | ° | BP AXIS | e |
| 6 | Shear | Å | INTRA | f |
| 7 | Stretch | Å | INTRA | g |
| 8 | Stagger | Å | INTRA | h |
| 9 | Buckle | ° | INTRA | i |
| 10 | Propel | ° | INTRA | j |
| 11 | Opening | ° | INTRA | k |
| 12 | Shift | Å | INTER | l |
| 13 | Slide | Å | INTER | m |
| 14 | Rise | Å | INTER | n |
| 15 | Tilt | ° | INTER | o |
| 16 | Roll | ° | INTER | p |
| 17 | Twist | ° | INTER | q |
| 18 | H-Rise | Å | INTER | r |
| 19 | H-Twist | ° | INTER | s |
| 20 | Alpha | ° | BACKBONE | t |
| 21 | Beta | ° | BACKBONE | u |
| 22 | Gamma | ° | BACKBONE | v |
| 23 | Delta | ° | BACKBONE | w |
| 24 | Epsilon | ° | BACKBONE | x |
| 25 | Zeta | ° | BACKBONE | y |
| 26 | Chi | ° | BACKBONE | z |
| 27 | Phase | Å | BACKBONE | aa |
| 28 | Amplitude | Å | BACKBONE | ab |
| 29 | Hydrogen bond | Kcal/mol | SEQUENCE E. | ac |
| 30 | Stacking energy | Kcal/mol | SEQUENCE E. | ad |
| 31 | Solvation | Kcal/mol | SEQUENCE E. | ae |
|  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Phylum** | **Microorganism Name** | **No. of sequences** | **TP** | **FP** | **TN** | **FN** | **Sensitivity**  **(%)** | **Specificity**  **(%)** |
| Euryarchaeota | *Methanolobus psychrophilus* | 1254 | 1034 | 251 | 1003 | 220 | 82.46 | 79.98 |
| *Thermococcus kodakarensis* | 964 | 819 | 202 | 762 | 145 | 84.96 | 79.05 |
| *Halofrex volcanii* | 1113 | 853 | 143 | 970 | 260 | 76.64 | 87.15 |
| Actinobacteria | *Mycobacterium tuberculosis* H37Rv | 1084 | 800 | 240 | 844 | 284 | 73.80 | 77.86 |
| *Streptomyces coelicolor* A3 | 1845 | 1478 | 236 | 1609 | 367 | 80.11 | 87.21 |
| Proteobacteria | *Helicobacter pylori* | 746 | 697 | 227 | 519 | 49 | 93.43 | 69.57 |
| *Salmonella enteric* serovar *Typhimurium* | 1571 | 1212 | 515 | 1056 | 359 | 77.15 | 67.22 |
| *Escherichia coli* | 1140 | 1019 | 129 | 1011 | 121 | 89.39 | 88.68 |
| *Pseudomonas aeruginosa* PA14 | 1612 | 1260 | 199 | 1413 | 352 | 78.16 | 87.66 |
| Firmicutes | *Bacillus amyloliquefaciens* | 926 | 846 | 187 | 739 | 80 | 91.36 | 79.81 |
| Chlamydiae | *Chlamydia pneumonia CWL029* | 329 | 304 | 59 | 270 | 25 | 92.40 | 82.07 |
| Cyanobacteria | *Synechocystis* sp. PCC6803 | 296 | 239 | 71 | 225 | 57 | 80.74 | 76.01 |
|  | Overall | 12880 | 10561 | 2459 | 10421 | 2319 | **83.38** | **80.19** |

**Supplementary Table S5: Performance of SEProm in promoter identification of twelve different organisms**

**Supplementary Table S6:** Precision and F-values obtained in promoter prediction of twelve organisms using SEProm and five other tools

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name of**  **organism** |  | **F-value** | | | | | |  | | **Precision** | | | | |
| **SEProm** | | **PromPredict** | **Pepper** | **BProm** | **bTSSFinder** | **CNNPromoter\_b** | **SEProm** | **PromPredict** | | **Pepper** | **BProm** | **bTSSFinder** | **CNNPromoter\_b** |
| *Methanolobus psychrophilus* | 81.45 | | 75.24 | 68.81 | 67.32 | 66.65 | 67.14 | 80.47 | 71.73 | | 82.30 | 50.76 | 50 | 50.78 |
| *Thermococcuskodakarensis* | 82.52 | | 75.18 | 47.42 | 69.99 | 66.68 | 67.78 | 80.22 | 73.47 | | 74.10 | 53.89 | 50.04 | 53.67 |
| *Halofrex volcanii* | 80.89 | | 75.42 | 62.50 | 71.95 | 68.43 | 42.16 | 85.64 | 75.72 | | 70.19 | 74.76 | 52.05 | 55.85 |
| *Mycobacterium tuberculosis* | 75.33 | | 64.62 | 61.19 | 69.67 | 67.97 | 13.39 | 76.92 | 65.23 | | 73.71 | 68.67 | 51.50 | 26.34 |
| *Streptomyces coelicolor*A3 | 83.06 | | 70.09 | 65.02 | 55.75 | 68.71 | 6.32 | 86.23 | 73.10 | | 79.26 | 84.25 | 52.50 | 22.25 |
| *Helicobacter pylori* | 83.47 | | 71.35 | 65.56 | 66.80 | 66.63 | 66.66 | 75.43 | 71.93 | | 79.30 | 50.18 | 50 | 50.03 |
| *Salmonella typhimurium* | 83.50 | | 73.50 | 61.10 | 68.07 | 66.71 | 66.99 | 70.18 | 70.18 | | 79.70 | 51.64 | 50.06 | 52.10 |
| *Escherichia coli* | 89.07 | | 83.96 | 77.57 | 68.88 | 66.72 | 70.50 | 88.76 | 85.33 | | 79.16 | 52.56 | 50.08 | 54.66 |
| *Pseudomonas aeruginosa* | 82.06 | | 68.60 | 60.25 | 80.45 | 68.28 | 40.72 | 86.36 | 71.33 | | 88.65 | 78.43 | 51.93 | 51.63 |
| *Bacillus amyl-oliquefaciens* | 86.37 | | 74.32 | 66.53 | 67.36 | 66.64 | 67.40 | 81.90 | 71.92 | | 74.67 | 50.81 | 50 | 50.91 |
| *Chlamydia pneumonia* | 87.86 | | 68.60 | 58.06 | 66.79 | 66.60 | 66.72 | 83.75 | 65.74 | | 80 | 50.21 | 50 | 50.21 |
| *Synechocystis*sp. PCC6803 | 78.88 | | 64.77 | 57.92 | 67.48 | 66.60 | 67.85 | 77.10 | 68.42 | | 83.74 | 51.06 | 50 | 51.55 |
| Overall | **82.04** | | 72.14 | 62.66 | 68.37 | 67.21 | 53.36 | **81.08** | 72.01 | | 75.39 | 59.76 | 50.68 | 47.49 |

**S1: LOD score calculation**

1. **Regression Equations for continuous stretch just upstream of TSS**

**For ‘40’ basepair long window**

Log\_odds=8.77328685786625+(3.1911154492\*PCA1)+(1.934304789\*PCA2)+(2.3544014743\*PCA3)+(2.9467690235\*PCA4)+(0.83974823728\*PCA5)

**For ‘80’ basepair long window**

Log\_odds=14.7381717527129+(5.3235778785\*PCA1)+(2.3059575636\*PCA2)+(4.1930340959\*PCA3)+(2.5317627809\*PCA4)+(0.5109895382\*PCA5)

1. **Regression Equations for important motifs present upstream of TSS**

The 31 parameters were grouped in four vectors depending on their behavior at TSS

**Structure Increasing (SI)**: Stretch, Opening, Rise, Roll, Twist, H-Rise, H-Twist, Beta, Gamma, Epsilon, Phase, Amplitude

**Structure Decreasing (SD)**: X Disp., Y Disp., Inclination, Tip, Ax-Bend, Shear, Stagger, Buckle, Propel, Shift, Slide, Tilt, Alpha, Delta, Zeta, Chi

**Energy Increasing (EI):** Solvation energy

**Energy Decreasing (ED)**: Hydrogen bond, Stacking energy

The above four vectors were subjected to PCA analysis and top four PCAs were selected for LOD calculation

**For Combination 1**

Motif 1

Log\_odds=2.5204290379+(0.2661486777\*PCA1)+(0.3398276828\*PCA2)+(1.9261091436\*PCA3)+(0.0466519918\*PCA4)

Motif 2

Log\_odds=-0.720805026291914+(-0.0634232556\*PCA1)+(-0.3849557773\*PCA2)+(2.428200063\*PCA3)+(-4.2406272919\*PCA4)

Motif 3

Log\_odds=-0.822678079983201+(-0.0780278346\*PCA1)+(0.2439218536\*PCA2)+(1.2378985344\*PCA3)+(0.1179656519\*PCA4)

**For combination 2**

Motif 1

Log\_odds=2.2584742555357+(0.2383258115\*PCA1)+(0.3269695499\*PCA2)+(1.8774382674\*PCA3)+(0.2465648791\*PCA4)

Motif 2

Log\_odds=-0.696832650856326+(-0.0608023405\*PCA1)+(-0.3808452971\*PCA2)+(2.3931747078\*PCA3)+(-4.3013409124\*PCA4)

Motif 3

Log\_odds=-1.97517077096845+(-0.1921041494\*PCA1)+(-0.3382743213\*PCA2)+(2.1045012628\*PCA3)+(-2.0852215826\*PCA4)

Motif 4

Log\_odds=-1.61207138442233+(-0.1562536649\*PCA1)+(-0.2998186918\*PCA2)+(1.768353137\*PCA3)+(-0.7250029738\*PCA4)

**For combination 3**

Motif 1

Log\_odds=2.85544653158039+(0.3006680956\*PCA1)+(0.3519805607\*PCA2)+(2.0003000218\*PCA3)+(-0.0586177277\*PCA4)

Motif 2

Log\_odds=-0.720805026291914+(-0.0634232556\*PCA1)+(-0.3849557773\*PCA2)+(2.428200063\*PCA3)+(-4.2406272919\*PCA4)

Motif 3

Log\_odds=-1.86488556813325+(-0.1811002108\*PCA1)+(-0.3389217212\*PCA2)+(2.0601741494\*PCA3)+(-1.6716634385\*PCA4)

Motif 4

Log\_odds=-1.14946871659729+(-0.1103568656\*PCA1)+(-0.2511591797\*PCA2)+(1.3806061104\*PCA3)+(0.1176659993\*PCA4)

**For combination 4**

Motif 1

Log\_odds=2.06133681065907+(0.2164176028\*PCA1)+(0.3124004808\*PCA2)+(1.8232533658\*PCA3)+(0.5654863848\*PCA4)

Motif 2

Log\_odds=-0.720805026291914+(-0.0634232556\*PCA1)+(-0.3849557773\*PCA2)+(2.428200063\*PCA3)+(-4.2406272919\*PCA4)

Motif 3

Log\_odds=-1.86488556813325+(-0.1811002108\*PCA1)+(-0.3389217212\*PCA2)+(2.0601741494\*PCA3)+(-1.6716634385\*PCA4)

Motif 4

Log\_odds=-1.12276821942032+(-0.1077417726\*PCA1)+(-0.254950043\*PCA2)+(1.3831742083\*PCA3)+(0.1415839923\*PCA4)

**S2: Statistical measures used for evaluation of software**

|  |  |
| --- | --- |
| True Positive = Correctly identified (TP)  False Positive = Incorrectly identified (FP)  True negative = Correctly rejected (TN)  False negative = Incorrectly rejected (FN) | Sensitivity (Sens) = (TP/ (TP+FN)) \*100  Specificity (Spec) = (TN/ (TN+FP)) \* 100  Recall = TP/(TP+FN)\*100  Precision = TP/(TP+FP) \* 100  F\_Score = 2\*((Precision\*Recall)/(Precision+Recall))  Balanced Accuracy = Sensitivity\*Specificity/2 |