

Prediction of the bacterial promoter regions by DNA electrostatic properties

Temlyakova E.A., Sorokin A.A.

Institute of Cell Biophysics RAS, Pushchino, Russia, evgenia.teml@gmail.com

Electrostatic potential is the only property of a macromolecule that can influence other molecules over distances above 5Å. We studied a possibility to use electrostatic potential distribution around DNA molecule for identification of promoter regions in bacteria. For this purpose we created a number of PLS-DA [1] models based purely on characteristics of the electrostatic potential profile and trained them to discriminate promoters from non-promoter sequences. We observed the accuracy for models ranged in 80-85%. The VIP analysis of trained models shown that electrostatic potential values at as least as 15 positions relative to transcription start site is enough for 75% accuracy.

The database of promoter electrostatic profiles for fast comparison of the DNA sequences to experimentally identified promoters by similarity of their electrostatic profiles has been created. The database utilize the SAX-based indexing scheme along special PostgreSQL object type to compare the profiles with each other and properties of a query sequence [2]. The database contains the information about 1227 known promoters of *E.coli* K12. Each promoter is characterized by a nucleotide sequence, an electrostatic potential profile and the list of controlled genes. The information about transcription factors binding sites that can influence promoter activity is linked to the promoter record also.

References

1. Sarker M., Rayens W. (2003) Partial least squares for discrimination, *J.Chemom.*, **V.17**: 166
2. A. Sorokin, G. Selkov, I. Goryanin. (2012) A user-defined data type for the storage of time series data allowing efficient similarity screening, *European Journal of Pharmaceutical Sciences*, **46**:272–274.