Recent Publications

Kel, A., Kel, O., Ischenko, I., Kolchanov, N., Karas, H., Wingender, E. and Sklenar, H. (1996). TRRD and COMPEL databases on transcription linked to TRANSFAC as tools for analysis and recognition of regulatory sequences. Computer Science and Biology. Proceedings of the German Conference on Bioinformatics (GCB'96), R. Hofestadt, T. Lengauer, M. Löffler, D. Schomburg (eds.). University of Leipzig, Leipzig 1996, pp. 113-117.

Wingender, E., Kel, A. E., Kel, O. V., Karas, H., Heinemeyer, T., Dietze, P., Knueppel, R., Romaschenko, A. G. and Kolchanov, N. A. (1997). TRANSFAC, TRRD and COMPEL: Towards a federated database system on transcriptional regulation. Nucleic Acids Res., in press.

Ananko E.A., Ignatieva E.V., Kel A.E., Kolchanov N.A (1996). WWWTRRD: Hypertext information system on transcription regulation. Computer Science and Biology. Proceedings of the German Conference on Bioinformatics (GCB'96), R. Hofestadt, T. Lengauer, M. Löffler, D. Schomburg (eds.). University of Leipzig, Leipzig 1996, pp. 153-155.

A.E. Kel, O.V. Kel, O.V. Vishnevsky, M.P. Ponomarenko, I.V. Ischenko, H. Karas, N.A. Kolchanov, H. Sklenar, E. Wingender (1997). TRRD and COMPEL databases on transcription linked to TRANSFAC as tools for analysis and recognition of regulatory sequences. (1997) LECTURE NOTES IN COMPUTER SCIENCE, in press.

Holger Karas, Alexander Kel, Olga Kel, Nikolay Kolchanov, and Edgar Wingender (1997). Integrating knowledge on gene regulation by a federated database approach: TRANSFAC, TRRD and COMPEL. Jurnal Molekularnoy Biologii (Russian), in press.

Kel A.E., Kolchanov N.A., Kel O.V., Romaschenko A.G., Ananko E.A., Ignatyeva E.V., Merkulova T.I., Podkolodnaya O.A., Stepanenko I.L., Kochetov A.V., Kolpakov F.A., Podkolodniy N.L., Naumochkin A.A. (1997). TRRD: A database on transcription regulatory regions of eukaryotic genes. Jurnal Molekularnoy Biologii (Russian) in press.

O.V. Kel, A.E. Kel, A.G. Romaschenko, E. Wingender, N.A. Kolchanov (1997). Composite regulatory elements: classification and description in the COMPEL data base. Jurnal Molekularnoy Biologii (Russian), in press.

Data-Management Tools for Genomic Databases

Victor M. Markowitz and **I-Min A. Chen**
Information and Computing Sciences Division; Lawrence Berkeley National Laboratory; Berkeley, CA 94720
510/486-6835, Fax: -4004, *vmmarkowitz@lbl.gov*
http://gizmo.lbl.gov/opm.html

The Object-Protocol Model (OPM) data management tools provide facilities for constructing, maintaining, and exploring efficiently molecular biology databases. Molecular biology data are currently maintained in numerous molecular biology databases (MBDs), including large archival MBDs such as the Genome Database (GDB) at Johns Hopkins School of Medicine, the Genome Sequence Data Base (GSDB) at the National Center for Genome Resources, and the Protein Data Bank (PDB) at Brookhaven National Laboratory. Constructing, maintaining, and exploring MBDs entail complex and time-consuming processes.

The goal of the Object-Protocol Model (OPM) data management tools is to provide facilities for efficiently constructing, maintaining, and exploring MBDs, using application-specific constructs on top of commercial database management systems (DBMSs). The OPM tools will also provide facilities for reorganizing MBDs and for exploring seamlessly heterogenous MBDs. The OPM tools and documentation are available on the Web and are developed in close collaboration with groups maintaining MBDs, such as GDB, GSDB, and PDB.

Current work focuses on providing new facilities for constructing and exploring MBDs. The specific aims of this work are:

(1) Extend the OPM query language with additional constructs for expressing complex conditions, and enhance the OPM query optimizer for generating more efficient query plans.

(2) Develop enhanced OPM query interfaces supporting MBD-specific data types (e.g., protein data type) and operations (e.g., protein data display and 3D search), and assisting users in specifying and interpreting query results.

(3) Provide support for customizing MBD interfaces.

(4) Extend the OPM tools with facilities for managing permissions (object ownership) in MBDs, and for physical database design of relational MBDs, including specification of indexes, allocation of segments, and handling of redundant (denormalized) data.

(5) Develop OPM tools for constructing and maintaining multiple OPM views for both relational and non-relational (e.g., ASN.1, AceDB) MBDs. For a given MBD, these tools will allow customizing different OPM views for different groups of scientists. For heterogeneous MBDs, this tool will allow exploring them using common OPM interfaces.

(6) Develop tools for constructing OPM based multidatabase systems of heterogeneous MBDs and for exploring and manipulating data in these MBDs via OPM interfaces. As part of this effort, the OPM-based multidatabase system which consists currently of GDB 6.0 and GSDB 2.0, will be extended to include additional MBDs, primarily GSDB 2.2 (when it becomes available), PDB, and Genbank.

(7) Develop facilities for reorganizing OPM-based MBDs.The database reorganization tools will support automatic generation of procedures for reorganizing MBDs following restructuring (revision) of MBD schemas.

In the past year, the OPM data management tools have been extended in order to address specific requirements of developing MBDs such as GDB 6 and the new version of PDB.

The current version of the OPM data management tools (4.1) was released in June 1996 for Sun/OS, Sun/Solaris and SGI. The following OPM tools are available on the Web at http://gizmo.lbl.gov/opm.html:

(1) an editor for specifying OPM schemas;