BIOINFORMATICS

Vol. 17 no. 11 2001
Pages 1027–1034

Thermodynamic database for protein–nucleic acid interactions (ProNIT)

Ponraj Prabakaran [1], Jianghong An [1], M. Michael Gromiha [1], Samuel Selvaraj [1], Hatsuho Uedaira [1], Hidetoshi Kono [2] and Akinori Sarai [1,]*

[1]RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan and [2]Department of Chemistry, University of Pennsylvania, 231 South 34 Street, Philadelphia, PA 19104-6323, USA

Received on December 8, 2000; revised and accepted on April 20, 2001

ABSTRACT

Motivation: Protein–nucleic acid interactions are fundamental to the regulation of gene expression. In order to elucidate the molecular mechanism of protein–nucleic acid recognition and analyze the gene regulation network, not only structural data but also quantitative binding data are necessary. Although there are structural databases for proteins and nucleic acids, there exists no database for their experimental binding data. Thus, we have developed a Thermodynamic Database for Protein–Nucleic Acid Interactions (ProNIT).

Results: We have collected experimentally observed binding data from the literature. ProNIT contains several important thermodynamic data for protein–nucleic acid binding, such as dissociation constant (K_d), association constant (K_a), Gibbs free energy change (ΔG), enthalpy change (ΔH), heat capacity change (ΔC_p), experimental conditions, structural information of proteins, nucleic acids and the complex, and literature information. These data are integrated into a relational database system together with structural and functional information to provide flexible searching facilities by using combinations of various terms and parameters. A www interface allows users to search for data based on various conditions, with different display and sorting options, and to visualize molecular structures and their interactions.

Availability: ProNIT is freely accessible at the URL http://www.rtc.riken.go.jp/jouhou/pronit/pronit.html.

Contact: sarai@rtc.riken.go.jp

INTRODUCTION

Protein–nucleic acid interactions play a critical role in the regulation of gene expression. Advances in genome analysis have brought an explosive amount of sequence information on nucleic acids and proteins. Also, the knowledge of complete genomes of many different organisms has now opened up the possibility of examining the complex molecular network of gene regulation at the genome level. In order to understand the complex system of a network, we first need to elucidate the interactions among proteins and nucleic acids. On the other hand, a large number of protein–nucleic acid systems have been structurally characterized and this trend will be further accelerated by structural genome projects (Domingues *et al.*, 2000). However, the mechanism of specific sequence recognition of nucleic acids by proteins is still poorly understood. In order to understand the recognition mechanism, we need to know energetics or thermodynamics of the interactions as well as the structural information (Jen-Jacobson *et al.*, 2000; Oda and Nakumara, 2000; Sarai and Kono, 2001). The integration of the structural and thermodynamic knowledge of protein–nucleic acid recognition will help us to delineate the molecular mechanism of affinity and specificity of interactions involved in the protein–nucleic acid complex (Ladbury, 1995). It will also lead to a wide spectrum of applications such as the design of novel nucleic acid binding proteins, predictive methods for the target sites, and the quantitative simulation of gene regulation network.

In recent years, several attempts have been made to catalog and systematize the accumulation of experimental data on regulatory elements of genes from different organisms through databases and archives such as TRANSFAC (Wingender *et al.*, 2000) and TRRD (Kolchanov *et al.*, 2000). Also, rapidly increasing structural data of nucleic acid binding proteins have been archived in the Protein Data Bank (PDB; Berman *et al.*, 2000) and the Nucleic Acid Database (NDB; Berman *et al.*, 1998). However, there is no on-line database describing the affinity, specificity and energetic components of protein–nucleic acid interactions. Such data are sorely needed if anyone wants to compare the structural features of a transcription

*To whom correspondence should be addressed.

factor with its thermodynamic properties of binding with DNA. It is also very difficult to systematically investigate the relationship between the structure and thermodynamics of protein–nucleic acid interactions without such a proper database. In order to relieve this situation, we have constructed the thermodynamic database for Protein–Nucleic Acid Interactions (ProNIT). It contains numerical data for several thermodynamic parameters, namely, dissociation and association constants, changes in free energy, enthalpy and heat capacity, activity etc. along with experimental methods and conditions, details about secondary structure and solvent accessibility for wild-type residues corresponding to mutation sites, and the conformational changes in protein and nucleic acid upon binding. We have integrated this database with the structural information of protein–nucleic acid complexes and the information of specific base–amino acid interactions, through 3DinSight, an integrated relational database and search tool for the structure, function and properties of biomolecules developed previously (An *et al.*, 1998). Researchers can make flexible searches on thermodynamic properties of protein–nucleic acid interactions, and examine the corresponding molecular structures, conformational properties of protein–nucleic acid complexes and specific base–amino acid interactions. The database is linked to the major biological databases. We have prepared a www interface to help users searching for the data with required conditions and sorting formats, and visualizing the results through the Internet.

CONTENTS OF DATABASE

We have been collecting thermodynamic data of protein–nucleic acid interactions from the literature to provide a unique source of protein–nucleic acid binding to the scientific community. We used a set of keywords combining thermodynamic terms, binding parameters and names of nucleic acid binding proteins to extract the relevant papers from PUBMED. The protein–nucleic acid binding data from scientific articles so retrieved have been screened and manually entered into the database. ProNIT currently contains more than 1000 entries from 28 different protein–nucleic acid complexes, for which the structures are known in most cases, and they include some mutations in proteins and/or nucleic acids. There is a variety of nucleic acid sequences involving double-stranded as well as single-stranded DNA bound to proteins. ProNIT also contains the supplementary information such as secondary structure along with the solvent accessible surface area at the mutation sites of proteins, conformational changes in proteins and nucleic acids, the details of base–amino acid interactions and visualization of molecular structures and properties.

The information about all the protein–nucleic acid complexes present in the database along with the types of mutations, experimental methods and literature are available in 'Database statistics' at the home page of ProNIT. Recently, we have made the second update of nearly 1000 entries, which gives rise to the total number ProNIT entries as more than 2000 from 40 different nucleic acid binding proteins. We are continuously updating the database so that it will cover up to the latest thermodynamic data of protein–nucleic acid interactions in the literature.

Each entry of the database is referred to by a serial number. The schema of the database is best described by the following six distinct sections. The data items available in ProNIT are explained with an example database entry (Table 1).

(1) *Protein information*: name, source, fragment and sequence of the protein, enzyme code (EC; http://www.expasy.ch/sprot/enzyme.html), Protein Information Resource (PIR; McGarvey *et al.*, 2000) code, PDB codes for wild and mutant structures, information about monomeric or oligomeric state, ProTherm number, details of mutation with mutant residue, number, secondary structure and accessibility at the mutant sites.

(2) *Nucleic acid information*: name, source and sequence of the nucleic acid, information on mutation and sequence of mutant nucleic acid, GenBank (Benson *et al.*, 1997) accession number and NDB code.

(3) *Complex information*: codes for PDB and NDB, link to protein–nucleic acid complex database (see later), details of ligand molecules, accessibility of relevant mutant residues in the complex and conformational changes of protein and nucleic acid upon binding.

(4) *Experimental condition*: temperature, pH, details about buffers, ions, additives and experimental method.

(5) *Binding data*: dissociation constant, K_d (M), association constant, K_a (M^{-1}), changes in Gibbs free energy, ΔG (kcal mol^{-1}), enthalpy, ΔH (kcal mol^{-1}) and heat capacity, ΔC_p (kcal mol^{-1} K^{-1}) for wild and mutant entities, stoichiometry of binding and activity (K_m and k_{cat}).

(6) *Literature*: reference, authors, keywords and remarks.

IMPLEMENTATION

The ProNIT data are stored as tables in the relational database system (SYBASE) at UNIX workstation, as implemented in 3DinSight (http://www.rtc.riken.go.jp/jouhou/3dinsight/3dinsight.html; An *et al.*, 1998). The database system also incorporates ProTherm: thermodynamic database for proteins and mutants

Table 1. The description of data items in ProNIT with an example entry showing the experimental data of λ repressor binding to operator DNA (Sarai and Takeda, 1989) along with the other information. Underlines indicate hyperlinks to other databases

Description of data items	Example of a ProNIT entry
Protein	
Name	Lambda repressor
Source	Bacteriophage
Sequence	STKKKPLTQEQLEDARRLKAIYEKKKNELG LSQESVADKMGMGQSGVGALFNGINALNAY NAALLAKILKVSVEEFSPSIAREIYEMYEA VS
Protein Information Resource code	RPBPL
Protein Data Bank code	1LRP
Number of biological units	2 (dimer)
Nucleic acid	
Name	OR1 operator
Source	Bacteriophage lambda DNA
Sequence	tacctctggcggtgata
Gene bank accession number	M25081
Complex	
Protein Data Bank code	1LMB
Nucleic Acid Data Bank code	PDR010
Base step parameters and flexibility	Examine sequence-dependent DNA conformational parameters
Specific base-amino acid interactions	Examine specific base-amino acid interactions Image
Conformational change of protein	The N-terminal arm wraps around operator
Conformational change of nucleic acid	DNA is bent and major groove is widened
Experimental conditions	
Temperature	0°C
pH value	7.4
Name of the buffer	Tris HCl
Concentration of the buffer	10 mM
Name of the ion	potassium chloride (KCl)
Concentration of the ion	50 mM
Method of measurement	Filter binding assay (FBA)
Thermodynamic data	
Dissociation constant (K_d)	1.00e-09 M
Association constant (K_a)	1.00e+09 1/M
Binding free energy change (ΔG)	-1.12e+01 kcal/mol
Stoichiometry of binding	2
Literature	
Reference	Proc Natl Acad Sci U S A 86, 6513-6517 (1989)
Names of authors	Sarai A, Takeda Y
Keywords	sequence-specific, asymmetric recognition

(Gromiha *et al.*, 1999a, 2000a; http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html), which is a collection of thermodynamic parameters of proteins and mutants, protein–nucleic acid complex database (http://www.rtc.riken.go.jp/jouhou/3dinsight/complexdb.html), in which the complex structures are classified according to the recognition motif and other characteristics, and base–amino acid interaction database (http://www.rtc.riken.go.jp/jouhou/3dinsight/distance.html), in which specific pairs of base–amino acid interaction can be searched in detail. The relational database enables one to make a flexible search by combining various terms and conditions using the Structured Query Language (SQL). ProNIT is also cross-linked with the major biological databases such as PDB, NDB, EC, PIR, ProTherm and NCBI PUBMED (http://www.ncbi.nlm.nih.gov/pubmed).

INTERFACE

The users can search ProNIT through a form-based www interface at the URL http://www.rtc.riken.go.jp/jouhou/pronit/pronit_search.html. This can be used to search for data under various conditions with different display and sorting options for output according to the user's purpose and convenience, and to visualize molecular structures and properties. The search interface is user-friendly and having fill-in boxes, select buttons and pull-down menus for using several conditions to search the database. ProNIT can provide various searching ways as follows:

(i) retrieving data for a particular protein by its name, source, sequence or PDB code along with the specification of mutation type, secondary structure and the solvent accessible surface area (in % or $Å^2$) range;

(ii) retrieving data for a nucleic acid by mentioning the name, source, sequence or NDB code;

(iii) the queries based on the subsequence of protein or nucleic acid can be performed;

(iv) extracting data based on various experimental conditions, namely, method, T, pH; users can also specify a particular range of values of the parameters (T, pH);

(v) selecting data based on various binding parameters such as K_d, K_a, ΔG, ΔH and ΔC_p with a preferred range of the values;

(vi) searching by author name, keywords and year of publication. The keyword search covers all the text fields, e.g. conformational changes of proteins and nucleic acids may be searched by using keywords such as bending, partial folding, etc.

Then, the output format can be specified by selecting various display options and by sorting with solvent accessible surface area, T, pH, K_d and year of publication. For example, one can search for data based on the values of dissociation constants (K_d) within a range of values and sort the output by K_d values. A few more examples illustrating the usage of ProNIT are shown as 'Tutorial' in the homepage. The on-line help for the explanation of terms and the input format of searching items is hyperlinked to each term in the search interface as well as in data entry (http://www.rtc.riken.go.jp/jouhou/pronit/binding_help.html).
The search results are linked to their relevant sequence, structure and literature databases. The structure of the protein–nucleic acid complex corresponding to the thermodynamic data can be visualized through the database of protein–nucleic acid complex structures (http://www.rtc.riken.go.jp/jouhou/3dinsight/complexdb.html) in the same database system. Here, users can also examine the conformational properties of DNA such as roll, tilt, slide, twist, rise, propeller twist of base pairs and dihedral angles of backbones (Dickerson *et al.*, 1989), and sequence-dependent flexibility (Sarai *et al.*, 1989) in the form of graphical plots, as shown in Figures 1a and b respectively. These parameters are helpful to understand sequence-dependent variations of local DNA geometry and conformational flexibility. Then, if users are interested in the specific base–amino acid interactions involved in the complex for comparison with the binding thermodynamic data, they can search for the pairs by specifying atom, residue and distance criteria. The specific base–amino acid pairs are automatically highlighted in the complex and visualized by 3D viewers, RasMol (Sayle and Milner-White, 1995) or VRML (see Figure 2). For this purpose, we provide the modified RasMol code. The complex structure can also be linked to 3DinSight, where sequence motifs and mutation sites are automatically mapped and visualized on the structure, and those sites are hyperlinked to the corresponding document information by using the modified RasMol.

APPLICATIONS

We have designed and implemented ProNIT to be a unique resource for understanding the mechanism of protein–nucleic acid recognition. Recently, we have developed the thermodynamic database for proteins and mutants (ProTherm), which is a collection of more than 9000 data entries of important thermodynamic parameters along with structural, functional and literature information. Using this database, several analyses have been carried out to understand the mechanism of protein stability upon mutation (Gromiha *et al.*, 1999b). Furthermore, models have been proposed to predict the stability of proteins and mutants (Gromiha *et al.*, 1999c, 2000b; Ooi, 2000). Using the ProNIT database, it would be possible to establish a relationship between thermodynamics and structural changes upon the formation of protein–DNA complexes, e.g. the correlation between changes in heat capacity and accessible surface area upon complex formation. It was found that most of the site-specific protein–DNA associations are accompanied by large negative heat capacity changes (Spolar and Record, 1994). Quantitative analysis of these data has led to a suggestion that the conformational changes or partial folding of proteins upon binding contribute to the negative heat capacity changes (Spolar and Record, 1994). It has also been shown that some site-specific protein–DNA complexes have a striking correspondence between the DNA deformity and the thermodynamic parameters (Jen-Jacobson *et al.*, 2000). Transcription factors usually bind to the regulatory region of DNA in combination with other factors, and the cooperativity of binding should provide clues for understanding the mechanism of protein–DNA

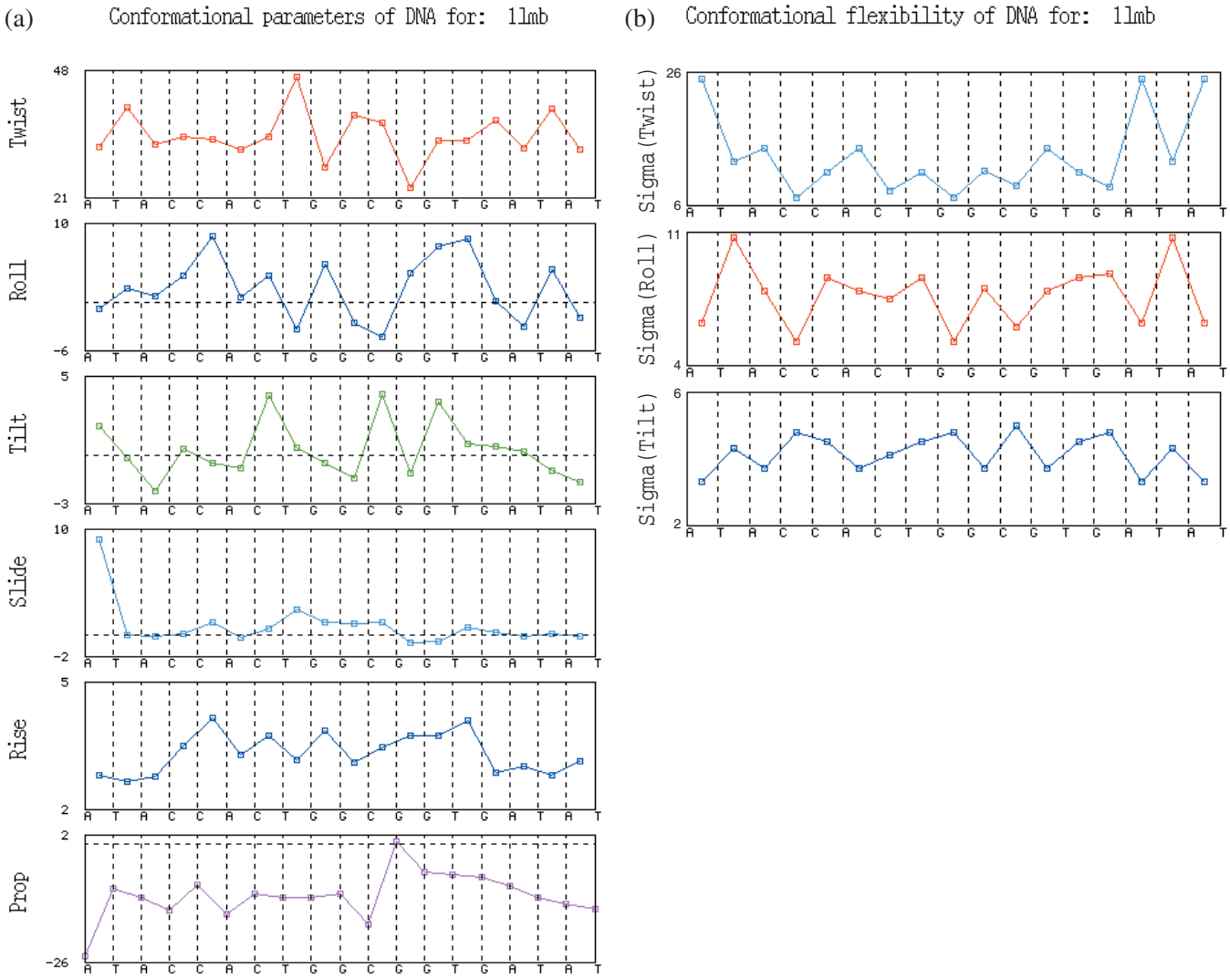

Fig. 1. Examining the standard conformational parameters and sequence-dependent fluctuations for operator DNA in the λ repressor–operator complex (PDB code: 1LMB). (a) Graphical plots of major sequence-dependent conformational parameters. The conformational parameters were calculated by using NEWHEL93 program (Dickerson, 1993) following the definition by Dickerson *et al.* (1989). Twist, Roll, Tilt and Propeller twist (Prop) are in degrees, and Slide and Rise are in Angstroms. (b) Graphical plots of conformational flexibility values for Twist, Roll and Tilt. In these plots, 'Sigma' (degrees) in y-axis shows the conformational fluctuation along each conformational coordinate, which represents the conformational flexibility of the variable, with respect to the sequence step (x-axis) of DNA base in the complex structure. The conformational fluctuations were obtained from the conformational free-energy profiles of each variable calculated by the method by Sarai *et al.* (1989) (based on Table 1 in this reference), and defined by the following equation:

$$\text{Sigma} = \sqrt{\langle (X - \langle X \rangle)^2 \rangle}$$

where $\langle X \rangle$ denotes the Boltzmann average over the variable X. Figure 1b indicates that the base steps near both ends of the DNA in this complex are subject to larger fluctuations (more flexible) along Twist coordinates than those in the interior part of the DNA.

recognition in living systems. All of these studies will require systematic quantitative analysis of structural and thermodynamic data of binding. ProNIT provides a good resource for such purposes.

The prediction of targets of transcription factors is quite important for the analysis of gene regulation network at

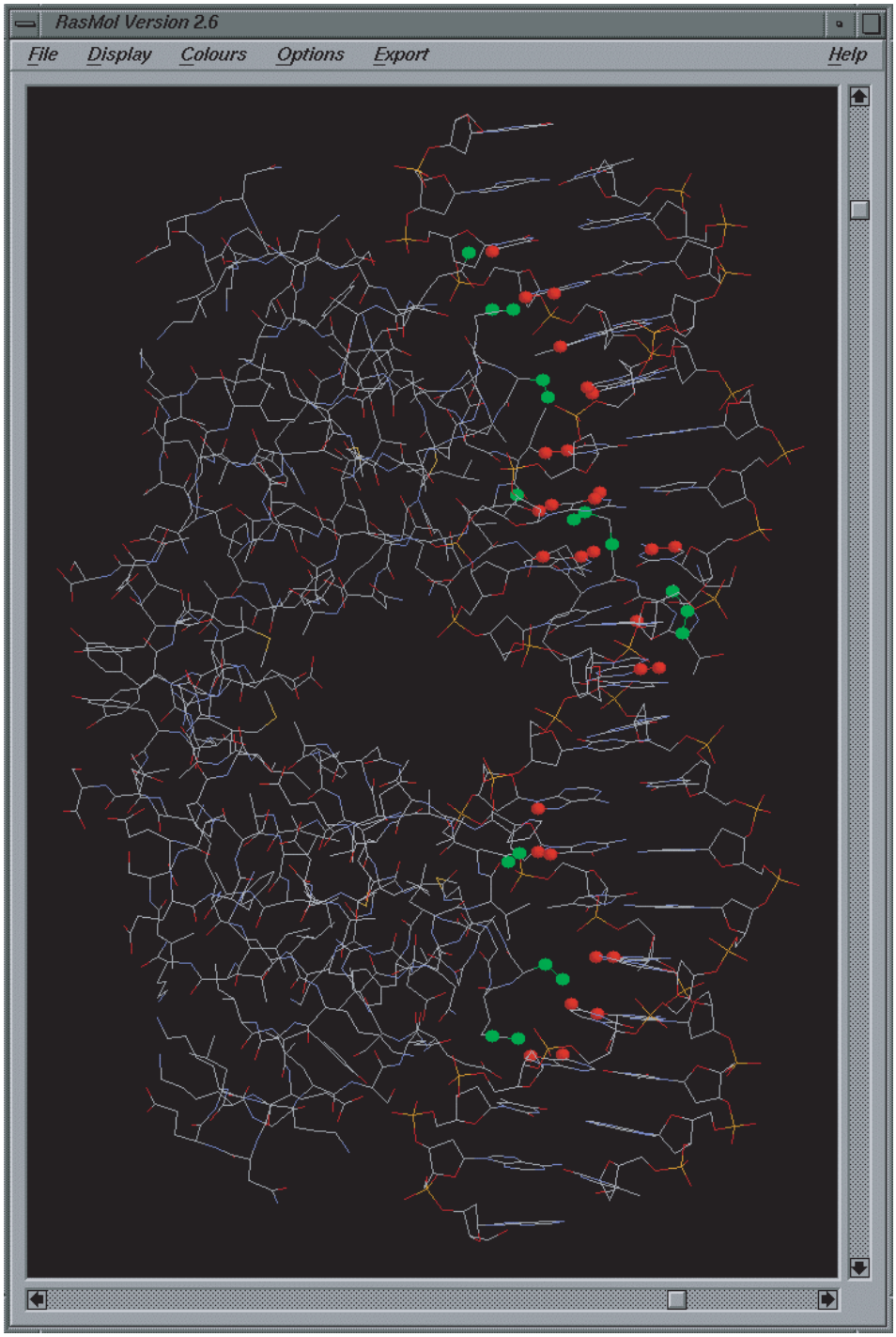

Fig. 2. The 3D structure of a protein–DNA complex with automatically highlighted specific interactions between bases and amino acids. This example shows λ repressor–operator complex (1l mb; Beamer and Pabo, 1992). RasMol view of this structure shows automatically highlighted contacting atoms of base (red dots) and amino acids (green dots) within 3.5 Å distance.

the genome level. Quantitative experimental data for the binding between proteins and DNA and their mutants can be used for the target prediction (Deng *et al.*, 1996; Sarai and Kono, 2001; Ponomarenko *et al.*, 2001a). There are many entries in ProNIT from binding experiments of DNA sequence variants, which can be analyzed together with

structural data to discern affinities at specific recognition sites.

Recent progress in genome analyses has disclosed complete genomes of more than 40 different organisms, which enables us to examine gene regulation networks at the genome level and compare them among different organisms (Tsoka and Ouzounis, 2000). Also, microarray technology makes it possible to extract information about gene expression pattern and identify a cluster of genes expressed together (Zweiger, 1999). The structural and thermodynamic data of protein–nucleic acid interactions should be compared with those data to get the total picture of the gene regulation network. The quantitative binding data of protein–nucleic acid interactions will be useful for the simulation of the network. Many of the disease-related mutations and single nucleotide polymorphisms will fall into coding region of transcription factors and regulatory DNA regions. These mutations may turn out to modulate the interactions between proteins and nucleic acids (Ponomarenko *et al.*, 2001b). The interactions can also be modulated by ligand binding and the analyses of such interactions help in designing suitable targets for the development of drugs which can alter the gene expression pattern.

So far, we have collected thermodynamic data of protein–nucleic acid interactions and combined them with structural information. In the future, we plan to collect functional information of genome, such as target sites and genes of transcription factors, functional annotation and evolutionary information of genes. By integrating such information into our database ProNIT can bridge between molecular properties and phenotype, and serve as a useful tool for functional analysis in the post-genome era.

ACKNOWLEDGEMENTS

P.P. thanks RIKEN for a postdoctoral fellowship award. M.M.G. is a recipient of RIKEN Research Fellowship.

REFERENCES

An,J., Nakama,T., Kubota,Y. and Sarai,A. (1998) 3DinSight: an integrated database and search tool for the structure, function and properties of biomolecules. *Bioinformatics*, **14**, 188–195. http://www.rtc.riken.go.jp/jouhou/3dinsight/3dinsight.html

Beamer,L.J. and Pabo,C.O. (1992) Refined 1.8 Å crystal structure of the λ repressor–operator complex. *J. Mol. Biol.*, **227**, 177–196.

Benson,D.A., Boguski,M.S., Lipman,D. and Ostell,J. (1997) GenBank. *Nucleic Acids Res.*, **25**, 1–6. http://www.ncbi.nlm.nih.gov

Berman,H.M., Zardecki,C. and Westbrook,J. (1998) The nucleic acid database: a resource for nucleic acid science. *Acta Crystallogr.*, D**54**, 1095–1104. http://ndbserver.rutgers.edu/NDB

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Ilya,N., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242. http://www.rcsb.org/pdb

Dickerson,R.E. (1993) NEWHEL93. University of California, Los Angeles, CA, USA, http://www.doe-mbi.ucla.edu/~thang/thang.html

Dickerson,R.E., Bansal,M., Calladine,C.R., Diekmann,S., Hunter,W.N., Kennard,O., von Kitzing,E., Lavery,R., Nelson,H.C.M., Olson,W.K., Saenger,W., Shakked,Z., Sklenar,H., Soumpasis,D.M., Tung,C-S., Wang,A.H.-J. and Zhurkin,V.B. (1989) Definitions and nomenclatures of nucleic acid structure components. *Nucleic Acids Res.*, **17**, 1797–1803.

Deng,Q-L., Ishii,S. and Sarai,A. (1996) Binding-site analysis of c-Myb: screening of potential binding sites by the mutational matrix derived from systematic binding affinity measurements. *Nucleic Acids Res.*, **24**, 766–774.

Domingues,F.S., Koppensteiner,W.A. and Sippl,M.J. (2000) The role of protein structure in genomics. *FEBS Lett.*, **476**, 98–102.

Gromiha,M.M., An,J., Kono,H., Oobatake,M., Uedaira,H. and Sarai,A. (1999a) ProTherm: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **27**, 286–288. http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html

Gromiha,M.M., Oobatake,M., Kono,H., Uedaira,H. and Sarai,A. (1999b) Role of structural and sequence information in the prediction of protein stability changes: comparison of buried and partially buried mutations. *Protein Eng.*, **12**, 549–555.

Gromiha,M.M., Oobatake,M., Kono,H., Uedaira,H. and Sarai,A. (1999c) Relationship between amino acid properties and protein stability: buried mutations. *J. Protein Chem.*, **18**, 565–578.

Gromiha,M.M., An,J., Kono,H., Oobatake,M., Uedaira,H., Prabakaran,P. and Sarai,A. (2000a) ProTherm, Version 2.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **28**, 283–285. http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html

Gromiha,M.M., Oobatake,M., Kono,H., Uedaira,H. and Sarai,A. (2000b) Importance of surrounding residues for protein stability of partially buried mutations. *J. Biomol. Struct. Dyn.*, **18**, 281–295.

Jen-Jacobson,L., Engler,L.E. and Jacobson,L.A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure*, **8**, 1015–1023.

Kolchanov,N.A., Podkolodnaya,O.A., Ananko,E.A., Ignatieva,E.V., Stepanenko,I.L., Kel-Margoulis,O.V., Kel,A.E., Merkulova,T.I., Goryachkovskaya,T.N., Busygina,T.V., *et al.* (2000) Transcription Regulatory Regions Database (TRRD): its status in 2000. *Nucleic Acids Res.*, **28**, 298–301. http://dragon.bionet.nsc.ru/trrd

Ladbury,J.E. (1995) Counting the calories to stay in the groove. *Structure*, **3**, 635–639.

McGarvey,P.B., Huang,H., Barker,W.C., Orcutt,B.C., Garavelli,J.S., Srinivasarao,G.Y., Yeh,L.S., Xiao,C. and Wu,C.H. (2000) PIR: a new resource for bioinformatics. *Bioinformatics*, **16**, 290–291. http://www-nbrf.georgetown.edu/pirwww

Oda,M. and Nakumara,H. (2000) Thermodynamic and kinetic analyses for understanding sequence-specific DNA recognition. *Genes to Cells*, **5**, 319–326.

Ooi,T. (2000) Thermal stability of protein predicted from its three dimensional structure. *Proc. Japan Acad.* B-*Phys.*, **7**, 97–102.

Ponomarenko,J.V., Furman,D.P., Frolov,A.S., Podkolodny,N.L., Orlova,G.V., Ponomarenko,M.P., Kolchanov,N.A. and Sarai,A. (2001a) ACTIVITY: a database on DNA/RNA sites activity

adapted to apply sequence-activity relationships from one system to another. *Nucleic Acids Res.*, **29**, 284–287. http://www.mgs.bionet.nsc.ru/mgs/systems/activity

Ponomarenko,J.V., Merkulova,T.I., Vasiliev,G.V., Levashova,Z.B., Orlova,G.V., Lavryushev,S.V., Fokin,O.A., Ponomarenko,M.P., Frolov,A.S. and Sarai,A. (2001b) rSNP_Guide, a database system for analysis of transcription factor binding to target sequences: application to SNPs and site-directed mutations. *Nucleic Acids Res.*, **29**, 312–316. http://wwwmgs.bionet.nsc.ru/mgs/programs/rsnp/

Sarai,A. and Takeda,Y. (1989) Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proc. Natl Acad. Sci. USA*, **86**, 6513–6517.

Sarai,A. and Kono,H. (2001) DNA-protein interactions: target predictions. In Crabbe, Drew,A. and Konopka (eds), *Handbook of Computational Biology*. Dekker, New York, in press.

Sarai,A., Mazur,J., Nussinov,R. and Jernigan,R.L. (1989) Sequence dependence of conformational flexibility of DNA. *Biochemistry*, **28**, 7842–7849.

Sayle,R. and Milner-White,E.J. (1995) RasMol: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374–376.

Spolar,R.S. and Record,Jr,M.T. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.

Tsoka,S. and Ouzounis,C.A. (2000) Recent developments and future directions in computational genomics. *FEBS Lett.*, **476**, 42–48.

Wingender,E., Chen,X., Hehl,R., Karas,H., Liebich,I., Matys,V., Meinhardt,T., Pruss,M., Reuter,I. and Schacherer,F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.*, **28**, 316–319. http://transfac.gbf.de/TRANSFAC

Zweiger,G. (1999) Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends Biotech.*, **17**, 429–436.