ISSN 2075-3810 (Print), ISSN 2075-3829 (Online)

БІОФІЗИЧНИЙ ВІСНИК, 2022, Вип. 48 BIOPHYSICAL BULLETIN, 2022, Iss. 48

MOLECULAR BIOPHYSICS

Original article
https://doi.org/10.26565/2075-3810-2022-48-02

UDC 577.21:577.29:577.323

PROTNA-ASA DATA BASE: NEW VERSION INCLUDING INFORMATION ABOUT ELECTROSTATIC POTENTIAL OF DNA MINOR GROOVE

M. Yu. Zhytnikova, A. V. Shestopalova
Usikov Institute for Radiophysics and Electronics, NAS of Ukraine, 12 Acad. Proskura str., Kharkiv, 61085, Ukraine
e-mail: avshestopalova1@gmail.com
Submitted November 15, 2022;
Accepted December 13, 2022

Background: In the past decades, the rapid development of molecular biology has led to a generation of an unprecedented amount of biological data obtained by the scientific community. Therefore, there is a significant and unmet need to store, process, and make sense of such a vast amount of data. There are currently available a number of databases, that cover different fields of molecular biology.
Objectives: In this paper, we describe Protein-Nucleic Acid Structural Database with Information on Accessible Surface Area, ProtNA-ASA, http://www.ire.kharkov.ua/ProtNA-ASA/index.php. The main aim of ProtNA-ASA is to provide quick and convenient access to structural information about DNA and protein-DNA complexes, that can be used for comprehensive study of protein-DNA recognition.
Materials and Methods: ProtNA-ASA database comprise information based on X-ray or NMR structures derived from Nucleic Acids Data Bank: 973 structures of protein-DNA complexes, 129 structures of naked A- and 403 of B-DNA ones; following structural parameters for each structure: conformational DNA parameters calculated with the 3DNA/CompDNA analyzer; DNA accessible surface area calculated using the modified algorithm of Higo and Go; DNA electrostatic potential calculated with DelPhi package.
Results: The recent update of ProtNA-ASA includes the electrostatic potential of the DNA minor groove since it plays an essential role in the indirect protein-DNA recognition process. The update also includes an advanced search, which serves to ease the use of the database and contribute to a more accurate structure selection. Advanced search allows finding structures by PDB/NDB ID, citation, length and sequence of a protein or DNA chain, type of structure, method of structure obtaining and resolution. All these queries can be used in different combinations with and/or statements.
Conclusion: The combination of structural information and physical characteristics from the ProtNA-ASA database is particularly useful to scientists studying the indirect readout, that based on DNA deformability. The detail analyzes of protein-DNA complexes and mechanisms of protein-DNA recognition is essential for implications in understanding cellular processes, DNA metabolism, transcriptional regulation, and developing therapeutic drugs.
KEY WORDS: structural database ProtNA-ASA; protein-DNA complexes; DNA structure; protein-DNA recognition.

The DNA functions (reproduction and regulation of genetic information) require the specific binding of proteins to the target DNA sequences. Proteins need to find the appropriate target rapidly, distinguishing between many similar and competing DNA sequences, and tightly bind to their “own” sites [1, 2]. But despite the extensive study of protein-DNA recognition, it is still impossible to formulate universal rules of its mechanisms similar to the principle of complementarity [3, 4]. A successful approach for understanding the mechanisms of protein-

In cites: Zhytnikova MYu, Shestopalova AV. ProtNA-ASA data base: new version including information about electrostatic potential of DNA minor groove. Biophysical Bulletin. 2022;48:18–24. https://doi.org/10.26565/2075-3810-2022-48-02

nucleic acid recognition, along with the results from experimental investigations, is the data analysis using molecular biological databases.

The rapid development of molecular biology in recent decades has led to the accumulation of scientific data and the creation of various bioinformatics databases [5]. The databases contain information received from scientific experiments, computational analyses, and published literature. It allows scientists to access a wide range of biologically relevant data. Structural databases [6–9] are informative for investigating protein-DNA recognition mechanisms. Exactly, polymorphism and the ability to conformational rearrangements are properties of the DNA double helix essential for successful recognition by proteins of their binding sites on DNA and the formation of specific protein-nucleic acid complexes [10–13].

For example, there are several tools for extracting DNA structure parameters (3DNA [14] and Curves+ [15]) and databases that hold protein-DNA interaction data: 1) 3d-footprint [16], which provide structure-based binding specificities and sequence logos; 2) DP-Bind [17], that takes a user-supplied sequence of a DNA-binding protein and predicts residue positions involved in interactions with DNA; 3) ProNIT [18] provides experimentally determined thermodynamic interaction data between proteins and nucleic acids; and 4) hPDI [19] holds experimental protein-DNA interaction data for humans identified by protein microarray assays; 5) PDIdb [20], a protein-DNA interaction interface database; 6) DNAproDB [21] automatically lays out nucleotide and residue interactions maps; ProDFace [22] is a web tool that characterizes the binding region of a protein-DNA complex based on amino acid propensity, hydrogen bond (HB) donor capacity (number of solvent accessible HB donor groups), sequence conservation at the interface core and rim region, and geometry.

To the best of the authors' knowledge, among existing structural databases, most of them collect or generate information about protein-DNA interactions but not presents sufficient information on the accessible surface area (ASA) or electrostatic potential of protein-nucleic acid complexes. However, the dependence of the availability of the nucleic acid surface for contacts with solvent or/and protein, on DNA sequence is of great importance for understanding the indirect readout mechanism.

Here we present an updated version of the Protein-Nucleic Acid Structural Database with Information on the Accessible Surface Area (ProtNA-ASA) Data Base [23]. The latest update includes information on minor groove electrostatic potential. In our previous research [24], we demonstrated that conformation changes in the sugar-phosphate DNA backbone alter the electrostatic potential of the DNA minor groove and are dependent on DNA sequences. Thus, the data stored in the updated ProtNA-ASA Data Base can facilitate the investigation of protein-DNA recognition mechanisms and allows the determination of several structural and physical properties of the DNA that affect protein-DNA binding affinity.

MATERIALS AND METHODS

The presented ProtNA-ASA Data Base was built using the *PHP 8.0.7* and the MySQL database management system.

The ProtNA-ASA is based on the structural data of free DNA and protein-DNA complexes extracted from Nucleic Acids Bank, NDB [6] according to the following criteria: X-ray or NMR structures; resolution better than 2 Å; double stranded DNA; non-modified base pairs; more than 4 nucleotides in each chain.

Conformational parameters of DNA were calculated with the 3DNA/CompDNA analyzer [25, 26].

Accessible surface area (ASA) of each DNA atom in minor and major groves was calculated using the modified algorithm of Higo and Go [27], see ref. [28] for detail.

The minor-groove electrostatic potential of DNA was obtained by solving the non-linear Poisson-Boltzmann equation at physiological ionic strength using the software package DelPhi [29]. Many studies showed that DelPhi provide an accurate calculation of DNA electrostatic potential distribution, which is crucial for revealing the mechanisms of indirect recognition [24, 30, 31]. According to a method proposed [11], an electrostatic potential for each DNA base pair was defined at the reference point located approximately in the center of minor groove and in plane of the base pair. Such definition of reference point allows to measure of electrostatic potential as a function of DNA sequence and conformation (see ref. [24] for detail of calculation).

Each entry of the ProtNA-ASA database and calculated parameters are presented in separate files available for viewing and downloading.

RESULTS AND DISCUSSION

1. Content of ProtNA-ASA database

After the update, the content of the ProtNA-ASA database increased to 1505 structures, of which: 973 structures of protein-DNA complexes; 129 structures of free A-DNA and 403 structures of free B-DNA. All structures were extracted from the Nucleic acids Data Bank [6].

The user interface is presented at Fig 1 a. Each entry of ProtNA-ASA Data Base is referred by PDB/NDB ID and contains the following information (Fig. 1 b–g):

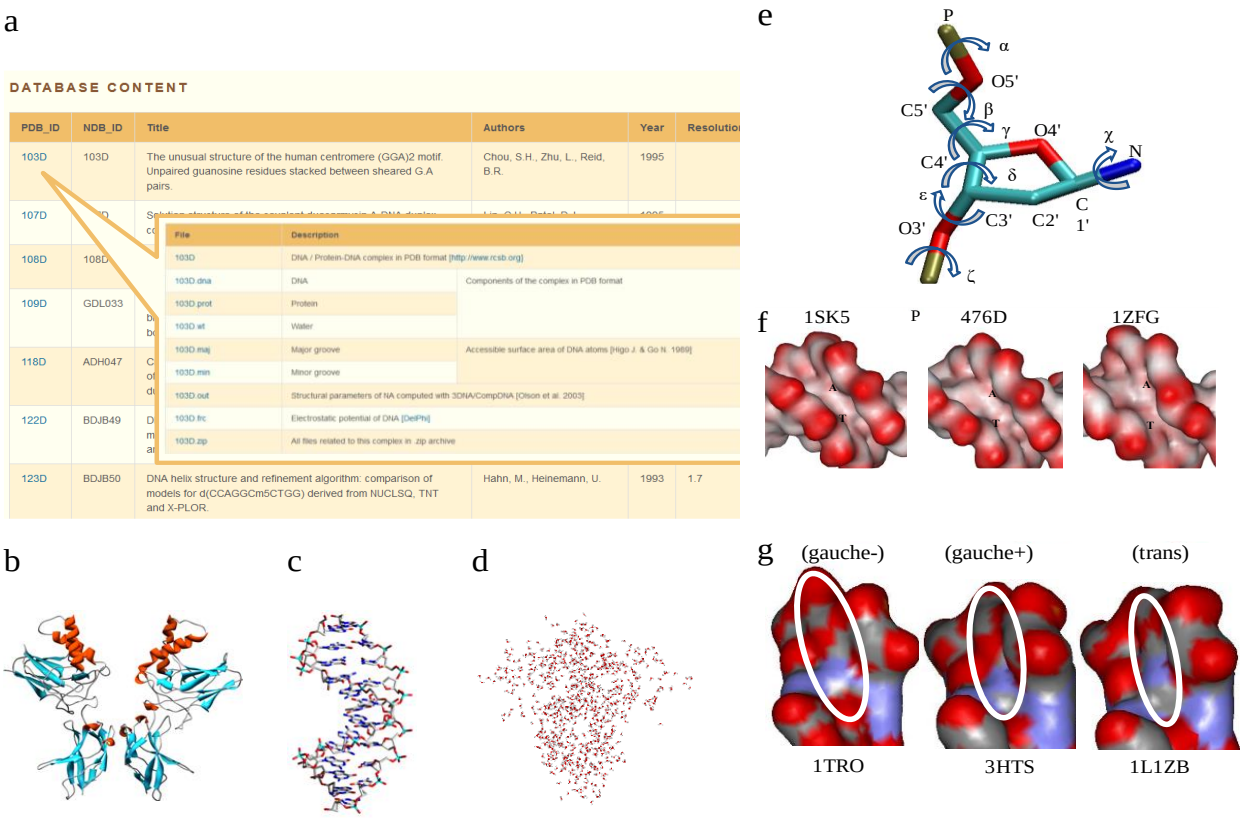

Fig. 1. Content of ProtNA-ASA Data Base. (a) an interface of ProtNA-ASA Data Base; components of protein-DNA complex from PDB file: protein (b), DNA (d) and water (d); (e) nucleotide torsion angles calculated with 3DNA/CompDNA program and available in the database (base pair and step parameters not shown); (f) visualization of AT base pair electrostatic potential with different γ angle conformations in minor groove, electrostatic distribution colors (f): red — negative values, white — positive values; (g) the values of calculated ASA for Adenine in different γ angle conformations in minor groove; ASA colors: red — O atoms, blue — N atoms, gray — C atoms.

- crystallographic coordinates of the structure in pdb format and their separate components: protein, DNA, water (if available);

- conformational parameters of DNA for each nucleotide;
- ASA of each DNA atom in minor and major groves separately;
- electrostatic potential of DNA minor groove.

2. Minor groove electrostatic potential.

The new version of the ProtNA-ASA Data Base is supplemented with information about minor groove electrostatic potential. Electrostatic potential plays an essential role in the protein-DNA interaction process since it is the most common and long-ranged type of interaction [32]. The influence of DNA conformation on its electrostatic potential only recently has been discovered [30, 31, 33]. It was shown that the regions with a narrow minor groove are often associated with enhanced negative electrostatic potential [11]. Moreover, the direct calculations of electrostatic potential are sensitive to chemical signatures [34]. Despite similar minor groove width, the electrostatic potential depends on functional groups because A-T and C-G base pairs carry different partial charge distributions.

Early [24] we demonstrated that the structural rearrangements of a double helix (minor groove sizes and the conformations of the sugar-phosphate backbone) induce changes in the distribution of minor groove electrostatic potential and are dependent on DNA sequence. Therefore, the minor groove sizes and distribution of its electrostatic potential can be used as a signal for proteins' recognition of their binding sites on the target DNA in the realization of the indirect or shape readout.

The examples of the distribution of minor grooves' electrostatic potentials are presented in Fig. 1 f. The visualization of AT base pairs' electrostatic potential with different γ angle conformations in the minor groove is shown. One can see the significant difference in the structural state and minor groove's sizes and the distribution of their electrostatic potentials for the AT base pairs in the three different DNA structures. Thus, the information available in the ProtNA-ASA Data Base make it possible to obtain a combination of data for the conformation of local DNA fragments, the spatial and physical characteristics corresponding to these fragments — the areas of the accessible surface of atoms, the changes in the polarity of these surfaces, and the distribution of the electrostatic potential.

3. Web user interface

The database can be accessed through Internet at http://www.ire.kharkov.ua/ProtNA-ASA/index.php. The web server has a simple user interface and include two search strategies: basic and advanced. In a basic search, you have one search field for entering one PDB ID. In contrast, an advanced search contains multiple fields which allow you to make more sophisticated queries Fig 2. The simplest search by PDB or NDB ID list of identifiers will return all matching entries in our database. Additional features can be specified as a filter, and only entities matched this criterion will be returned. It includes search by structure summary, DNA form or protein and their sequence, as well as experimental details. Description of each field is available as hints. Typically, the result of a submitted query will return list of entries available in the database with short description.

CONCLUSIONS

The increasing interest in the protein-DNA recognition process results in a growing number of relevant data, as well as existing resources and databases that facilitate the study of protein-DNA complexes. The information from ProtNA-ASA Data Base provides a possibility to investigate how structural and physical characteristics of the DNA affect the protein binding affinity. The ProtNA-ASA database represents a combination of data on DNA conformation, its accessible surface area, and electrostatic potential. Consequently, the ProtNA-ASA Data

Base can be useful for (i) discovering the general rules of the DNA conformation variability and its specificity, (ii) study of the influence of double-strand conformation on the availability of atoms for contact in DNA grooves and on physical characteristics such as electrostatic potential (Fig. 1 f) or DNA accessible surface area (Fig. 1 g), (iii) for prediction of protein-DNA binding sites. Predictions based on DNA sequence and structural information comprise major computational strategies commonly used to identify DNA-binding residues in a protein [35]. Therefore, using structural databases, in particular, the ProtNA-ASA Data Base [36] can improve the accuracy of such predictions.

ProtNA-ASA Data Base is freely accessible for any academic and educational purposes.

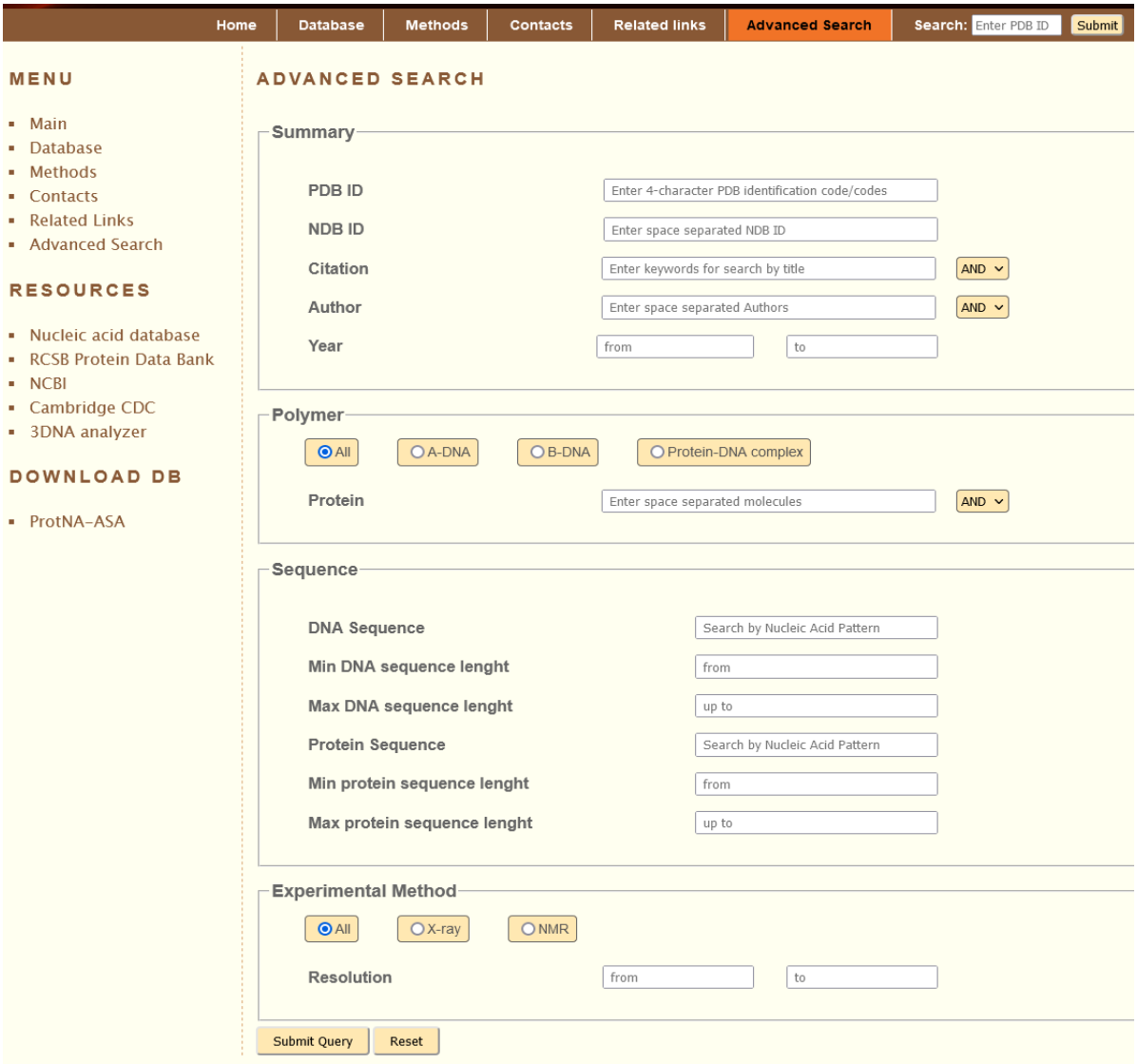

Fig. 2. Advanced search appearance.

ACKNOWLEDGEMENTS

The authors thank Dr. Gubin A. I. and Dr. Volovichev I. N. for technical support and consultation.

CONFLICT OF INTEREST

The authors report that there is no conflict of interest.

Authors' ORCID ID

M. Zhytnikova https://orcid.org/0000-0003-0003-7375
A. Shestopalova https://orcid.org/0000-0001-7613-7212

REFERENCES

1. Seeman NC, Rosenberg JM, Rich A. Sequence specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci USA. 1976;73:804–8. https://doi.org/10.1073/pnas.73.3.804
2. Rhodes D, Schwabe JW, Chapman L, Fairall L. Towards an understanding of protein-DNA recognition**.** Philos Trans R Soc Lond B Biol Sci. 1996;351:501–9. https://doi.org/10.1098/rstb.1996.0048
3. Matthews BW. No code for recognition. Nature. 1988;335:294-295. https://doi.org/10.1038/335294a0
4. Wetzel JL, Zhang K, Singh M. Learning probabilistic protein-DNA recognition codes from DNA-binding specificities using structural mappings. 2022. Preprint bioRxiv. https://doi.org/10.1101/2022.01.31.477772
5. Burks C. Molecular Biology Database List. Nucleic Acids Res. 1999;27:1–9. https://doi.org/10.1093/nar/27.1.1
6. Berman HM, Olson WK, Beveridge DL, Gelbin A, Demeny T, Hsieh SH, et al. The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J. 1992;63(3):751–9. https://doi.org/10.1016/S0006-3495(92)81649-1
7. Bairoch A, Boeckmann B. The SWISS-PROT protein sequence data bank, recent developments. Nucleic Acids Res. 1993;21(13):3093–6. https://doi.org/10.1093/nar/21.13.3093
8. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The Protein Data Bank. Acta Cryst Sec D Biol Cryst. 2002;58(1):899–907. https://doi.org/10.1107/S0907444902003451
9. Kinjo AR, Bekker GJ, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, et al. Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures. Nucleic Acids Res. 2017;45(D1):D282–D288. https://doi.org/10.1093/nar/gkw962
10. Svozil D, Kalina J, Omelka M, Schneider B. DNA conformations and their sequence preferences. Nucleic Acids Res. 2008;36:3690–706. https://doi.org/10.1093/nar/gkn260
11. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. The role of DNA shape in protein–DNA recognition. Nature. 2009;461:1248–53. https://doi.org/10.1038/nature08473
12. Rohs R, West SM, Liu P, Honig B. Nuance in the double-helix and its role in protein-DNA recognition. Curr Opin Struct Biol. 2009;19:171–7. https://doi.org/10.1016/j.sbi.2009.03.002
13. Abe N, Dror I, Yang L, Slattery M, Zhou T, Bussemaker HJ, et al. Deconvolving the recognition of DNA shape from sequence. Cell. 2015;161(2):307–18. https://doi.org/10.1016/j.cell.2015.02.008
14. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures [Internet]. [cited 2022 Nov 7]. Available from: http://web.x3dna.org/
15. Curves+ [Internet]. [cited 2022 Nov 7]. Available from: http://gbio-pbil.ibcp.fr/Curves_plus
16. 3D-footprint is a database of DNA-binding protein structures [Internet]. [cited 2022 Nov 7]. Available from: http://floresta.eead.csic.es/3dfootprint
17. DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins [Internet]. [cited 2022 Nov 7]. Available from: http://lcg.rit.albany.edu/dp-bind/
18. ProNIT database provides experimentally determined thermodynamic interaction data between proteins and nucleic acids [Internet]. [cited 2022 Nov 7]. Available from: http://dna00.bio.kyutech.ac.jp/pronit/
19. Human Protein-DNA Interactome (hPDI) [Internet]. [cited 2022 Nov 7]. Available from: http://bioinfo.wilmer.jhu.edu/PDI/
20. PDIdb Protein-DNA Interface Database [Internet]. [cited 2022 Nov 7]. Available from: http://melolab.org/pdidb/web/content/home
21. DNAproDB [Internet]. [cited 2022 Nov 7]. Available from: https://dnaprodb.usc.edu/
22. ProDFace: Server for the Analysis of Protein-DNA Interface [Internet]. [cited 2022 Nov 7]. Available from: http://structbioinfo.iitj.ac.in/resources/bioinfo/pd_interface/
23. Tkachenko MYu, Boryskina OP, Shestopalova AV, Tolstorukov MY. ProtNA-ASA: Protein-Nucleic Acid Structural Database with Information on Accessible Surface Area. Int J Quant Chem. 2010;110:230–2. https://doi.org/10.1002/qua.22067
24. Zhitnikova MY, Shestopalova AV. DNA minor groove electrostatic potential: influence of sequence-specific transitions of the torsion angle gamma and deoxyribose conformations. J Biomol Struct Dyn. 2017;35(15):3384–97. https://doi.org/10.1080/07391102.2016.1255259
25. Gorin AA, Zhurkin VB, Olson WK. B-DNA twisting correlates with base-pair morphology. J Mol Biol. 1995;247(1):34–48. https://doi.org/10.1006/jmbi.1994.0120
26. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 2003;31(17):5108–21. https://doi.org/10.1093/nar/gkg680

27. Higo J, Gō N. Algorithm for rapid calculation of excluded volume of large molecules. J Comput Chem 1989;10:376–9. https://doi.org/10.1002/jcc.540100311
28. Zhitnikova MYu, Boryskina OP, Shestopalova AV. Sequence-specific transitions of the torsion angle gamma change the polar-hydrophobic profile of the DNA grooves: implication for indirect protein-DNA recognition. J Biomol Struct Dyn. 2014;32(10):1670–85. https://doi.org/10.1080/07391102.2013.830579
29. Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, et al. DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys. 2012;5:9. https://doi.org/10.1186/2046-1682-5-9
30. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of specificity in protein-DNA recognition. Annu Rev Biochem. 2010;79:233–69. https://doi.org/10.1146/annurev-biochem-060408-091030
31. West SM, Rohs R, Mann RS, Honig B. Electrostatic interactions between arginines and the minor groove in the nucleosome. J Biomol Struct Dyn. 2010;27(6):861–6. https://doi.org/10.1080/07391102.2010.10508587
32. Chen X, Tsai MY, Wolynes PG. The Role of Charge Density Coupled DNA Bending in Transcription Factor Sequence Binding Specificity: A Generic Mechanism for Indirect Readout. J Am Chem Soc. 2022;144(4):1835–45. https://doi.org/10.1021/jacs.1c11911
33. Tullius T. Structural biology: DNA binding shapes up. Nature. 2009; 461:1225-1226. https://doi.org/10.1038/4611225a
34. Chiu TP, Rao S, Mann RS, Honig B, Rohs R. Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein–DNA binding. Nucleic Acids Res. 2017;45(21):12565–76. https://doi.org/10.1093/nar/gkx915
35. Si J, Zhao, R, Wu R. An Overview of the Prediction of Protein DNA-Binding Sites. Int J Mol Sci. 2015;16:5194–215. https://doi.org/10.3390/ijms16035194
36. Zhitnikova MYu, Shestopalova AV DNA-protein complexation: contact profiles in DNA grooves. Biophys Bull. 2017;38(2):54–65. https://doi.org/10.26565/2075-3810-2017-38

БАЗА ДАНИХ PROTNA-ASA: НОВА ВЕРСІЯ, ЩО ВКЛЮЧАЄ ІНФОРМАЦІЮ ПРО ЕЛЕКТРОСТАТИЧНИЙ ПОТЕНЦІАЛ МАЛОГО ЖОЛОБКА ДНК

М. Ю. Житнікова, Г. В. Шестопалова

Інститут радіофізики та електроніки ім. О. Я. Усікова НАН України, вул. Ак. Проскури, 12, Харків, Україна, 61085

Надійшла до редакції 15 листопада 2022 р. Прийнята до друку 13 грудня 2022 р.

Актуальність: В останні десятиліття швидкий розвиток молекулярної біології призвів до створення безпрецедентної кількості біологічних даних, отриманих науковою спільнотою. Тому існує значна потреба у зберіганні, обробці та розумінні великого обсягу даних. На даний час існує велика кількість специфічних баз даних, які охоплюють різні галузі молекулярної біології.

Мета роботи: Стаття присвячена опису оновленої бази даних ProtNA-ASA (http://www.ire.kharkov.ua/ProtNA-ASA/index.php). Основна мета бази даних ProtNA-ASA — забезпечити швидкий і зручний доступ до інформації про ДНК та білково-нуклеїнові комплекси, яка може бути використана для всебічного дослідження механізмів білково-нуклеїнового впізнавання.

Матеріали та методи: База даних ProtNA-ASA містить інформацію про структури з Банку даних нуклеїнових кислот, отриманих методами рентгеноструктурного аналізу та ЯМР: 973 структури білково-нуклеїнових комплексів, 129 структур вільної A- та 403 структури вільної B-ДНК; наступні параметри для кожної структури: конформаційні параметри ДНК, розраховані за допомогою програми 3DNA/CompDNA; площу доступної поверхні ДНК, розраховану за допомогою модифікованого алгоритму Higo і Go; електростатичний потенціал ДНК, розрахований з використанням пакету DelPhi.

Результати: Оновлена база даних ProtNA-ASA включає значення електростатичного потенціалу ДНК, розподіл якого у малому жолобку відіграє важливу роль у непрямому білково-нуклеїновому впізнаванні. Останнє оновлення також дозволяє проводити розширений пошук інформації з використанням PDB/NDB ID; цитування; довжини та послідовності білка і ДНК; типу структури; методу отримання структури та роздільної здатності. Усі запити можна застосовувати в різних комбінаціях з операторами та/або. Це суттєво полегшує використання бази даних і сприяє більш точному відбору структур.

Висновки: Об’єднання структурної інформації та фізичних характеристик з бази даних ProtNA-ASA є важливим при дослідженні непрямого механізму впізнавання, заснованого на здатності подвійної спіралі ДНК до конформаційних перебудов. Детальний аналіз структур білково-нуклеїнових комплексів і механізмів впізнавання необхідний для розуміння метаболізму ДНК в клітині, регуляції транскрипції, розробки терапевтичних препаратів.

КЛЮЧОВІ СЛОВА: структурна база даних ProtNA-ASA; білково-нуклеїнові комплекси; структура ДНК; білково-нуклеїнове впізнавання.