

Zinc Finger Database (ZiFDB) v2.0: a comprehensive database of C₂H₂ zinc fingers and engineered zinc finger arrays

Fengli Fu^{1,2} and Daniel F. Voytas^{1,2,*}

¹Department of Genetics, Cell Biology and Development and ²Center for Genome Engineering, University of Minnesota, 6-160 Jackson Hall, 321 Church Street SE, Minneapolis, MN 55455, USA

Received September 14, 2012; Revised October 25, 2012; Accepted October 26, 2012

ABSTRACT

ZiFDB (Zinc Finger Database, <http://zifdb.msi.umn.edu>) is a web-accessible database that houses information on individual C₂H₂ zinc fingers (ZFs) and engineered zinc finger arrays (ZFAs). ZiFDB serves as a resource for biologists interested in engineering ZFAs for use as sequence-specific DNA-binding reagents. Here, we describe four new features of ZiFDB: (i) the database allows users to input new ZFs and ZFAs; (ii) a shadow database temporarily stores user-submitted data, pending approval by the database curator and subsequent loading into the persistent database; (iii) ZiFDB contains 181 Context-Dependent Assembly (CoDA) ZFAs, which were generated by this newly described ZFA engineering platform; and (iv) the database also now contains 319 F1F2 CoDA units and 334 F2F3 CoDA units that can be used to construct CoDA arrays. In total, the new release of ZiFDB contains 1226 ZFs and 1123 ZFAs.

INTRODUCTION

The C₂H₂ zinc finger (ZF) motif, which was first described in the transcription factor TFIIIA from *Xenopus laevis* (1), is one of the most abundant DNA-binding motifs in nature. Each ZF comprises about 30 amino acids that fold into a ββ α structure through hydrophobic interactions and binding of a zinc ion by two conserved cysteine and histidine residues. A ZF typically recognizes a continuous 3-bp DNA sequence. Owing to this DNA-binding capacity, ZFs serve as a framework for constructing engineered DNA-binding proteins: ZFs are linked in tandem to form zinc finger arrays (ZFAs), which can recognize extended DNA sequences (2,3).

The methods used to engineer ZFAs to recognize novel target sequences can be classified into two categories: modular assembly and selection-based methods. Modular assembly simply involves linking together ZFs that recognize known target sequences (3). However, modular assembly is not necessarily reliable. For example, in one study, more than half of the ZFAs created by modular assembly showed little or no activity (4). This is because specificity and affinity of a given finger is influenced by context—i.e. the position of a finger in an array and its neighboring fingers. Selection-based methods identify fingers that work well together and are thus more reliable for producing functional ZFAs; however, selection-based methods often require considerable time and a high level of molecular biology expertise to perform. In 2008, the Zinc Finger Consortium implemented a selection-based platform called Oligomerized Pool Engineering (OPEN) (5). Since its debut, >500 ZFAs have been generated by OPEN [558 are housed in Zinc Finger Database (ZiFDB) v2.0].

Despite the high efficacy of OPEN, the labor and expertise required by this method have prevented it from being widely used. With the goal of combining the simplicity of modular assembly with the reliability of OPEN, another publicly available platform, named Context-Dependent Assembly (CoDA), was described in 2011 by the Zinc Finger Consortium (6). To assemble CoDA ZFAs, two, two-finger units, derived through selection, are used that have a common ZF at position two (F2). For example, an F1F2 CoDA unit (recognizing 3'-GAGG GG) is fused to an F2F3 CoDA unit (recognizing 3'-GGG GTG) such that the resulting three-finger array recognizes a novel DNA sequence (3'-GAGGGGGTG). Although Moore *et al.* (7) showed that the activity of CoDA zinc finger nucleases (ZFNs) is lower than those made by OPEN, CoDA does not require selection steps and is therefore easier for most researchers to use.

*To whom correspondence should be addressed. Tel: +1 612 626 4509; Fax: +1 612 624 6264; Email: voytas@umn.edu

Regardless of the method used to construct ZFAs, these engineered DNA-binding domains have become powerful tools for both basic and applied biological research. Engineered zinc finger transcription factors, for example, can be created by fusing transcriptional repressor or activator domains to engineered ZFAs. These artificial transcription factors have been used to repress or activate genes in a variety of species with a high degree of specificity (8,9). Similarly, engineered ZFNs have proven effective as targeted mutagens in diverse eukaryotes (10,11). ZFNs are typically composed of a customized array of ZFs fused to the non-specific FokI restriction endonuclease cleavage domain (12). As FokI needs to dimerize to be functional, ZFAs are designed in pairs to recognize two unique DNA sequences separated by a short DNA spacer. Binding to the two-target sequences allows FokI to dimerize, cut the DNA and introduce a double-stranded break in the spacer. When double-stranded breaks are repaired through non-homologous end-joining or homologous recombination, targeted sequence modifications can be introduced at or near the break site.

ZiFDB is designed to serve as a resource for those interested in engineering custom ZFAs or better understanding how ZF proteins recognize target DNA (13). In addition

to housing information on ZFs and engineered ZFAs, ZiFDB is linked to the output from ZiFiT—a software package that assists biologists in finding sites within target genes for engineering ZF proteins (14). Consequently, ZiFDB is particularly valuable for determining whether a given ZFA (or portion thereof) has previously been constructed and whether it has the requisite DNA-binding activity for a given experiment. ZiFDB v2.0 houses an expanded number of ZFAs, from 652 to 1123. Likewise, the number of ZFs has expanded from 716 to 1126. The updated database also allows users to input ZFAs into a shadow database, which are then deposited into the persistent database by the curator on approval. The information in this database will continue to help molecular biologists develop ZF reagents that meet their needs for genome modification.

NEW FEATURES

Shadow database

ZiFDB v2.0 allows users to directly input into the database new information about novel ZFAs. To ensure data quality, a shadow database has been created.

ZINC FINGER
CONSORTIUM

ZiFDB version 2.0

Menu

[Introduction](#)
[Instructions](#)
[Search ZiFDB](#)
[Submit data to ZiFDB](#)

Links

[ZiFiT software](#)
[Zinc Finger Consortium](#)
[Addgene](#)

The nine nucleotide binding site you inputted is:

F1	F2	F3
GGA	GGC	GGG

The following F1F2 CoDA unit recognizes Target_F1 and Target_F2:

ID	ID_F1	ID_F2	Triplet_F1	Triplet_F2	Helix_F1	Helix_F2	Article_ID
19	<input type="text" value="898"/>	<input type="text" value="434"/>	GGA	GGC	RPSKLVL	LKEHLTR	<input type="text" value="36"/>

The following F2F3 CoDA unit recognizes Target_F2 and Target_F3:

ID	ID_F2	ID_F3	Triplet_F2	Triplet_F3	Helix_F2	Helix_F3	Article_ID
3	<input type="text" value="434"/>	<input type="text" value="1061"/>	GGC	GGG	LKEHLTR	RGDKLAL	<input type="text" value="36"/>

Note:

1. The seven amino acids are provided for each recognition helix and correspond to positions -1, 1, 2, 3, 4, 5 and 6 (numbered relative to the start of the recognition helix). Residues -1, 3 and 6 frequently make DNA contacts.
2. Input fields subsite F1 and subsite F2 to acquire only F1F2 CoDA units.
3. Input fields subsite F2 and subsite F3 to acquire only F2F3 CoDA units.

Voytas Laboratory
 University of Minnesota
 Phone: (612) 626-4509
 Email: voytas@umn.edu

Figure 1. Sample output from the CoDA unit search page when GGA is provided as triplet F1, GGC as triplet F2 and GGG as triplet F3.

The information submitted by the user is collected into the shadow database and held pending approval by the database curator. On approval, the information is then loaded into the persistent database. The information collected into the shadow database includes array name and array type (e.g. if it was derived by OPEN, CoDA or modular assembly), the sequence of the targeted DNA triplets, the sequences of the recognition helices of the ZFs, journal information if the array has been published and information about the submitter (see below for additional details).

New database content

Classes

ZiFDB stores ZF information as a set of objects defined by Java classes. The previous version of ZiFDB had four major classes: Zinc Finger, Zinc Finger Array, Article and Author. Two additional classes were added to this release of ZiFDB to accommodate the two-finger CoDA units (designated as CoDA_F1F2 and CoDA_F2F3). Similar to the Zinc Finger and Zinc Finger Array class designations, both CoDA_F1F2 and CoDA_F2F3 point to the Article class, which provides information about relevant

publications. Information about submitters of the new arrays is integrated into ZiFDB's existing Author class.

Novel three-finger arrays and CoDA units

Since the release of the previous version of ZiFDB, OPEN has been used to generate numerous additional ZFAs. More than 500 OPEN ZFAs are now housed in ZiFDB. CoDA arrays are a new array type in this version of ZiFDB, and the database currently has 181 CoDA arrays. ZiFDB v2.0 also has information about 319 F1F2 CoDA units and 334 F2F3 units.

Updated interface

Search page for CoDA units

In addition to the ability to search for individual ZFs and three-finger ZFAs, ZiFDB v2.0 allows users to search for all available CoDA units that recognize a given target sequence. By providing the DNA sequence of the F1F2 or F2F3 triplets to be targeted, all of the corresponding CoDA units matching the input sequence are returned. If the user provides the nucleotide triplets for a 9-bp target site, information is provided about both F1F2 and F2F3 units (Figure 1).

ZINC FINGER CONSORTIUM

ZiFDB version 2.0

Menu

- [Introduction](#)
- [Instructions](#)
- [Search ZiFDB](#)
- [Submit data to ZiFDB](#)

Links

- [ZiFIT software](#)
- [Zinc Finger Consortium](#)
- [Addgene](#)

Submit zinc finger array information

Array information:

Name: Type:

Comments:

Finger information:

	DNA triplet	RH AA seq	Finger AA seq
F1	<input type="text"/>	<input type="text"/>	<input type="text"/>
F2	<input type="text"/>	<input type="text"/>	<input type="text"/>
F3	<input type="text"/>	<input type="text"/>	<input type="text"/>

Journal information if the array is published (optional):

Title	<input type="text"/>	Journal	<input type="text"/>
Volume	<input type="text"/>	Page	<input type="text"/>
Year	<input type="text"/>		

Submitter information:

Submitter's name	<input type="text"/>	PI's name	<input type="text"/>
Institution	<input type="text"/>	email	<input type="text"/>

Note:

1. If the binding site is 5'-GAA-GCC-GTT-3', then the 'F1' is GTT, 'F2' is GCC and 'F3' is GAA.
2. In the comment area, please provide information about the array activity or any additional information you would like to disclose about the array.
3. Your email address will not be disclosed to anyone.

Voytas Laboratory
 University of Minnesota
 Phone: (612) 626-4509
 Email: voytas@umn.edu

Figure 2. The array submission page.

Page for submitting arrays

The array submission page collects four types of information (Figure 2):

- (1) Array information. The user can input the name given to a particular ZFA and the array type. As indicated above, the array type specifies the method used to engineer the ZFA (i.e. modular assembly, OPEN or CoDA). It is also possible to designate an array as natural (i.e. found in nature). The option remains to include engineering methods that may be developed at a later date. The user can also provide other useful information such as data concerning array activity.
- (2) Finger information. In this section, the DNA triplet, amino acid sequence of the recognition helix and the amino acid sequence of the entire ZF are provided for each finger in a ZFA.
- (3) Journal information. If the array has been published, then the citation is provided.
- (4) Submitter information. In this section, the name of the submitter and/or principal investigator, the name of the submitter's institution and an email address are provided. This information is used by the database curator to contact the submitter if there are any questions concerning the data. No information about the submitter is publicly disclosed.

Newly added features and attributes in the array query output page

In the output page of the array query, a search link for CoDA units is provided. This makes it convenient for the user to check whether CoDA units are available for assembling ZFAs that target the 9-bp DNA sequence they provided. Another newly added attribute of the array query output is the array type. This currently includes natural, modular assembly, OPEN and CoDA. As the success rate and efficacy of ZFAs engineered by different platforms differ considerably, this information may be valuable for users when choosing previously engineered arrays to use in their experiments.

CONCLUSION

One important new feature in this version of ZiFDB is the added ability of users to input ZFAs and ZFs into the database. This will ensure that the ZiFDB captures new information generated by the scientific community, including unpublished data. In addition, by now housing information about two-finger CoDA units and CoDA arrays, ZiFDB is current with the most recent ZFA engineering practices. Finally, the expanded number of ZFs and ZFAs that are now stored in ZiFDB will provide a rich

resource for users interested in either ZFA engineering or better understanding how ZFs recognize their target DNA.

FUNDING

Funding for open access charge: National Science Foundation [DBI-0923827 to D.F.V.].

Conflict of interest statement. None declared.

REFERENCES

1. Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.*, **4**, 1609–1614.
2. Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
3. Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Kokscha, B., Lund, C.V., Magnenat, L., Valente, D. *et al.* (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, **42**, 2137–2148.
4. Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
5. Maeder, M.L., Thibodeau-Beganny, S., Osiaik, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. *et al.* (2008) Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
6. Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. *et al.* (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*, **8**, 67–69.
7. Moore, F.E., Reyon, D., Sander, J.D., Martinez, S.A., Blackburn, J.S., Khayter, C., Ramirez, C.L., Joung, J.K. and Langenau, D.M. (2012) Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PLoS One*, **7**, e37877.
8. Urnov, F.D. and Rebar, E.J. (2002) Designed transcription factors as tools for therapeutics and functional genomics. *Biochem. Pharmacol.*, **64**, 919–923.
9. Blancafort, P., Segal, D.J. and Barbas, C.F. III (2004) Designing transcription factor architectures for drug discovery. *Mol. Pharmacol.*, **66**, 1361–1371.
10. Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773–782.
11. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
12. Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl Acad. Sci. USA*, **93**, 1156–1160.
13. Fu, F., Sander, J.D., Maeder, M., Thibodeau-Beganny, S., Joung, J.K., Dobbs, D., Miller, L. and Voytas, D.F. (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res.*, **37**, D279–D283.
14. Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K. and Dobbs, D. (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.*, **38**, W462–W468.