Nucleic Acids Research, 1993, Vol. 21, No. 21 4857–4959

Nucleosomal DNA sequence database

Ilya Ioshikhes[1,2,*] and Edward N.Trifonov[2]
[1]Research Institute, The College of Judea and Samaria, Ariel 44820 and [2]Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

Received June 2, 1993; Revised and Accepted August 13, 1993

The database of the nucleosomal DNA sequences presented below is result of screening of large amount of literature since 1978. Each documented case of experimental mapping of the nucleosomes is scrutinized, to extract necessary information on the sequence, nucleosome location and experimental uncertainty of the mapping.

The nucleosomal DNA sequences are sorted below in the descending order of experimental mapping accuracy. The database contains 143 nucleosomal DNA sequences from 14 eukaryotes and SV40 virus. Prokaryotic sequences are not included. The references correspond to the papers with original mapping results. The EMBL Nucleotide Sequence Database accession numbers are indicated in Table 1.

The nucleosomal center positions as they are claimed in the original papers or deduced from the published data, are given in the Table 1. The actual nucleosomal centers are located within the experimental error bars. In cases when in the original article the accuracy was not indicated, we assigned the error bars on the basis of typical error values for the mapping techniques used. There are six major techniques used for the nucleosomal mapping: 1) DNaseI digestion of chromatin or reconstituted nucleosomes. Its mapping accuracy is about ±10 bases (15–19). 2) DNaseI digestion in combination with the sequencing gel, with highest possible accuracy ±1 base (1, 2). 3) Exonuclease III with nuclease S1 digestion, ±1 base (3, 4). 4) OH·-radical mapping, ±5 bases or better (6, 9). 5) MNase digestion of chromatin. Its mapping uncertainty is about 20 bases or more. 6) MNase digestion in combination with the cloning and sequencing of nucleosomal DNA sequences, with error of about ±5 bases (7). If several techniques are combined, the experimental error is correspondingly reduced (see, for example, legend to Figure 1).

In few cases authors described the nucleosomes only as tentative ones, which is also indicated in the database. In cases when several overlapping alternative positions are detected, the most prominent position is presented in the database. Inevitably, the choice of the representative position as well as error assignments are somewhat subjective.

In this printed form of the database the data for only 44 nucleosomal DNA sequences from 11 eukaryotes and SV40 virus are presented, with highest experimental accuracy of their mapping. Data for other sequences are available in the EMBL Nucleotide Sequence Database, accompanied with similar descriptions.

ACKNOWLEDGEMENTS

This work has been supported by Ministry of Absorption, Ministry of Science and Technology and partly by NIH grant # GM47853-01. Authors appreciate discussions with Dr A.Bolshoy.

REFERENCES

1. Drew,H.R. and Calladine,C.R. (1987) *J. Mol. Biol.* **195**, 143–173.
2. Rhodes,D. (1985) *EMBO J.* **4**, 3473–3482.
3. Zhang,X.-Y., Fittler,F. and Horz,W. (1983) *Nucleic Acids Res.* **11**, 4287–4306.
4. Zhang,X-Y. and Horz,W. (1984) *J. Mol. Biol.* **176**, 105–129.

```
990
|
CGAGATGATCCCGCATTTTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGAGGCAAGAATGATCATCACCGTAGTGAG
          ∇
AGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTGACACC
                                                                                              ∇
GATTATTTAAAGCTGCAGCATACGATATATATACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTATACGAACAGTATGATACTGAAGATG
ACAAGGTAATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTTTTGCTTTTTCTTTTTTTTTCTCTTGAACTCGAGAAAAA
```

Figure 1. The nucleosome positioning (No. 32) in the pet56-his3-ded1 region (11). The claimed nucleosomal center is placed in the middle of the 400 bp long sequence so that up to 130 bp of upstream and downstream sequences are included as well. The nucleosome is located between two nuclease-sensitive sites mapped precisely by combination of DNAseI, MNase and restriction enzymes. The first site is at 1108, the second one — at 1270. The distance between them is 162 which is larger than typical nucleosome size 145 bp. The error assigned is ± 9 bp. The arrowheads indicate approximate ends of the sequence involved in the nucleosome.

*To whom correspondence should be addressed at: Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

Table 1.

	Acc. #	Species	Gene (region)	Position	Error (bases)	Ref.
1	V01426	Frog *X.borealis*	5S RNA gene 5′ flanking region	71	1	1
2	V01426	Frog *X.borealis*	5S RNA gene 5′ flanking & coding region	590	1	2
3	X04339	African green monkey	α-satellite	133	1	3
4	V00846	Mouse	Satellite	120	1	4
5	M10468	Sea urchin *L.variegatus*	5S rRNA gene segment	88	2	5
6	V01380	SV40		384	2	6
7	V01380	SV40		547	3	7
8	V01380	SV40		1542	3	7
9		Chicken	Fragment 223	74[a]	4[b]	8
10	V01380	SV40		1087	4	7
11	V01380	SV40		1793	4	7
12	V01380	SV40		4877	4	7
13	V01380	SV40		1362	4	7
14	V01380	SV40		3175	4	7
15	V01380	SV40		3466	4	7
16	V01380	SV40		4037	4	7
17	X01175	Frog *X.laevis*	Vitellogenin B1 promoter	724	4	9
18	J00784	Rat	Satellite 1	347	5	10
19	J00784	Rat	Satellite 1	163	5	10
20	V01380	SV40		220	5	7
21	V01380	SV40		4374	5	7
22	V01380	SV40		4705	5	7
23	X03245	Yeast *S.cerevisiae*	Pet56-his3-ded1 gene	530	5[b]	11
24		Chicken	Fragment 213	72[a]	6[b]	8
25	X03245	Yeast *S.cerevisiae*	Pet56-his3-ded1 gene	1662	6[b]	11
26	V01380	SV40		5118	6	12
27	V01380	SV40		792	7	7
28	V01380	SV40		2274	7	7
29	X12505	*Drosophila melanogaster*	Hsp26 gene promoter	181	8[b]	13
30	X60225	*Drosophila melanogaster*	H1−H3 intergenic spacer	1004	8[b]	14
31	V01380	SV40		5049	9	12
32	X03245	Yeast *S.cerevisiae*	Pet56-his3-ded1 gene	1189	9[b]	11
33	X02382	*Zea mays*	Shrunken gene 5′ region	496[c]	10[b]	15
34	K02115	Yeast *S.cerevisiae*	GAL10 transcription region	71	10[b]	16
35	K02115	Yeast *S.cerevisiae*	GAL1−GAL10 promoter	257	10[b]	16
36	K02115	Yeast *S.cerevisiae*	GAL1−GAL10 promoter	604	10[b]	16
37	K02115	Yeast *S.cerevisiae*	GALT 5′ & transcription region	773	10[b]	16
38	V01380	SV40		3	10[b]	17
39	X02364	*Tetrahymena thermophila*	rDNA 5′ NTS region	15	10	18
40	X02364	*Tetrahymena thermophila*	rDNA 5′ NTS region	215	10	18
41	X02364	*Tetrahymena thermophila*	rDNA 5′ NTS region	415	10	18
42	X02364	*Tetrahymena thermophila*	rDNA central region	615	10	18
43	Z00030	*Drosophila melanogaster*	Adh-s gene 5′ region	4653[c]	10[b]	19
44	Z00030	*Drosophila melanogaster*	Adh-s gene 5′ region	4873[c]	10[b]	19

[a]Position as in the reference.
[b]Evaluated error.
[c]Tentative nucleosome.

5. Simpson,R.T. and Stafford,D.W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 51–55.
6. Powers,J.H. and Bina,M. (1991) *J. Mol. Biol.* **221**, 795–803.
7. Ambrose,C., Lowman,H., Rajadhyaksha,A., Blasquez,V. and Bina,M. (1990) *J. Mol. Biol.* **214**, 875–884.
8. Pennings,S., Muyldermans,S., Meersseman,G. and Wyns,L. (1989) *J. Mol. Biol.* **207**, 183–192.
9. Schild,C., Claret,F.-X., Wahli,W. and Wolffe,A.P. (1993) *EMBO J.* **12**, 423–433.
10. Igo-Kemenes,T., Omori,A. and Zachau,H.G. (1980) *Nucleic Acids Res.* **8**, 5377–5390.
11. Losa,R., Omari,S. and Thoma,F. (1990) *Nucleic Acids Res.* **18**, 3495–3502.
12. Ambrose,C., Rajadhyaksha,A., Lowman,H. and Bina,M. (1989) *J. Mol. Biol.* **210**, 255–263.
13. Thomas,G.H. and Elgin,S.C.R. (1988) *EMBO J.* **7**, 2191–2201.
14. Kas,E. and Laemmli,U.K. (1992) *EMBO J.* **11**, 705–716.
15. Frommer,W.-B. and Starlinger,P. (1988) *Mol. Gen. Genet.* **212**, 351–359.
16. Rainbow,M., Lopez,J. and Lohr,D. (1989) *Biochemistry* **28**, 7486–7490.
17. Nobile,C., Nickol,J. and Martin,R.G. (1986) *Mol. Cell. Biol.* **6**, 2916–2922.
18. Palen,T.E. and Cech,T.R. (1984) *Cell* **36**, 933–942.
19. Ewel,A., Jackson,J.R. and Benyajati,C. (1990) *Nucleic Acids Res.* **18**, 1771–1781.