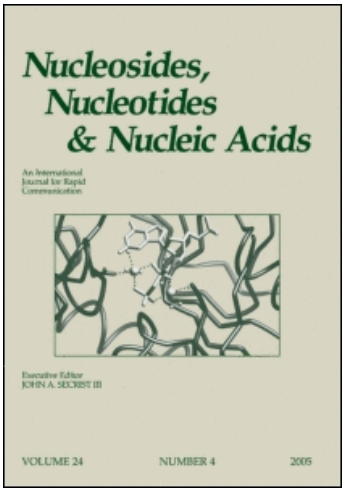

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Sequence Analysis of Fis Binding Sites Obtained by in Vitro Selection

Karsten Theis[a]; Cornelia Bartsch[a]; Wolfram Saenger[a]
[a] Freie Universisät Berlin, Institut für Kristallographie, Berlin

To cite this Article Theis, Karsten , Bartsch, Cornelia and Saenger, Wolfram(1997) 'Sequence Analysis of Fis Binding Sites Obtained by in Vitro Selection', Nucleosides, Nucleotides and Nucleic Acids, 16: 5, 579 — 584
To link to this Article: DOI: 10.1080/07328319708002919
URL: http://dx.doi.org/10.1080/07328319708002919

NUCLEOSIDES & NUCLEOTIDES, 16(5&6), 579-584 (1997)

SEQUENCE ANALYSIS OF FIS BINDING SITES OBTAINED BY IN VITRO SELECTION

Karsten Theis*, Cornelia Bartsch and Wolfram Saenger
Institut für Kristallographie, Freie Universität Berlin, Takustr. 6, 14195 Berlin

ABSTRACT To study the specificity of the FIS protein binding to DNA, we have isolated new binding sites by in vitro selection, analyzed their binding and bending by gel electrophoresis and compared the results to predictions based on sequence analysis of the known natural sites.

The DNA-binding protein FIS (factor for inversion stimulation) of *E. coli* binds and bends DNA segments which vary widely in sequence. The proposed consensus sequences fail to discriminate between binding and non-binding sequences[1]. In this study, we explore the reliability of a quantitative description of the specifity of FIS by sequence analysis using the selection theory of Hippel and Berg[2].

The selection theory describes the similarity of putative binding sites through a scoring matrix derived from nucleotide frequencies in a set of known and aligned binding sites. We have chosen to include all but 8 of the 35 published FIS binding sites given in the review of Finkel and Johnson[1] as our set of known binding sites. The distal and proximal binding sites of the enhancers hin, gin, cin, and pin have been excluded from the set because they have a double function[3]; they thus violate one assumption of the selection theory, namely that the only selective pressure on the sites is the requirement to interact with the protein studied.

In view of the dyad symmetry of the FIS protein[4] and the palindromic nature of the FIS consensus sequence, it is reasonable to assume that the interactions between FIS and DNA are also symmetrical. Each site in the set of binding sites was therefore included twice, as is and as reverse complement. Table 1 shows the nucleotide frequencies for the resulting 54 sites. From these frequencies, a scoring matrix (bottom of table 1) was calculated according to the selection theory. The score (λE in the terminology of Hippel and Berg) of a putative FIS binding site is the sum of the scores for each nucleotide at its respective position given by the scoring matrix. Sequences with high similarity to the set of known sites have a low λE and are expected to bind more strongly than sequences with a high λE. The λE values of the known sites used in this analysis range from 3.3 to 10.8, with an average value of 6.9.

The ability of selection theory to locate natural FIS binding sites is illustrated for the *proP* promoter region in table 2. In this and all additional cases tested thus far, selection theory describes the specificity of FIS better than consensus sequences do, and correctly predicts the location of known strong FIS binding sites (K.T., unpublished results).

TABLE 1: Nucleotide frequencies and scoring matrix for one half-site. The values for positions for the other half site (positions +1 to +10) follow from the palindromic symmetry.

nucleotide	position p										
b	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
	nucleotide frequencies										
A	22	14	10	2	13	15	2	39	34	27	24
C	5	9	17	0	15	11	20	1	4	3	3
G	13	8	7	52	14	8	5	12	5	5	3
T	14	23	20	0	12	20	27	2	11	19	24
	scoring matrix										
A	0.00	0.48	0.67	3.04	0.14	0.28	2.40	0.00	0.00	0.00	0.00
C	1.41	0.91	0.16	4.65	0.00	0.58	0.29	3.27	2.04	2.06	1.95
G	0.51	1.02	1.00	0.00	0.07	0.88	1.61	1.15	1.84	1.61	1.95
T	0.44	0.00	0.00	4.65	0.21	0.00	0.00	2.76	1.10	0.34	0.00

TABLE 2: Analysis of the *proP* promoter region. The four sites with lowest λE values are shown. pos, center of the site (numbering according to Xu and Johnson[5]); Δcons, number of mismatches with respect to the consensus sequence[3]; footprint, experimentally determined binding sites[5]. Additionally, a site which perfectly matches the consensus sequence but does not bind to FIS is shown in parentheses.

pos	sequence	λE	Δcons	footprint
-176	AAAGGT CATTAACTG CCCAAT	5.8	0	site II
-136	ATTGTA CATTCCTTA ACCGGA	6.9	1	site I
-262	TTGGCG TAAATAATC AGTTAC	9.5	2	
-208	GTTGAT CACAAATTT AAACAC	9.8	1	
(-85	AGAGAT TGCATCCTG CAATTC	13.9	0)	

It is not clear, however, if the λE value is a criterion for strong binding, or additionally reflects bending properties of the sites required for function. Also, some of the sequence preferences seen in the natural sites may be due to evolutionary and not functional relatedness. To test if the λE value is a general description of the specificity of FIS that may be applied to arbitrary sites and not only to natural functional sites, we studied the interaction of FIS with novel artificial binding sites.

We have obtained new FIS binding sites from two mixtures, A and B, of oligonucleotides closely resembling the known binding sites. Mixture A is derived from the "parent sequence" GGGGGTCAAAATTTGACCAAAA. The mixture was synthesized such that each position deviates from the nucleotide in the parent sequence with a probability of 1/8. Thus, sequences in the mixture are expected to deviate at 22/8=2.75 positions from the parent sequence on average. The mixture was separated on immobilized FIS (see methods). Four fractions

containing either weakly or strongly bound sequences were collected, the DNA was cloned and single clones were analyzed for binding and bending. Table 3 (top) shows the 18 sequences obtained by this procedure, ordered according to their λE values.

A comparison of the binding strength measured by gel electrophoresis with that predicted by selection theory shows a satisfactory correlation. All sequences with a λE below 6.1 are moderately strong (+) or strong (++) FIS binding sites. Sequences with a λE above 6.2 bind either weakly (-) or, in case of B06 and B09, moderately (+). The most frequent deviation from the parent sequence is T instead of G at position -8 (four occurrences); selection theory predicts that this change, along with -9 G to T, should give the greatest improvement of the parent sequence.

Sequences B18, B03 and B09 show that strong binding does not imply strong bending and vice versa, i.e. requirements for binding and bending are distinct. Sequence B18 deviates at three positions from the parent sequence. The deviations at positions -8 and -6 are observed in other sequences as well (compare B23 and B10), where they do not influence bending. Thus, the reduced bending of sequence B18 might be caused by the deviation at position +10, which interrupts the run of four adenines (+8 through +11) in the parent sequence. It is noteworthy that sequence B03, another sequence with reduced bending, has a deviation from the parent sequence at position +9, which also interrupts the adenine run. Furthermore, a sequence identical to B14 except for an A to C exchange at position +10 obtained from another selection experiment (data not shown) binds strongly but bends weakly. Thus, single nucleotide changes in the weakly conserved region +8 to +11 influence bending significantly.

Mixture B, `GGTGNNCAWWWWWTGNNCAAAA (W: A or T)`, is degenerate at the weakly conserved positions ±6 and ±5 and at the positions with A/T preference, ±2, ±1, and 0. FIS binding sequences were selected from this mixture by gel electrophoresis. Table 3 (bottom) shows the 11 sequences obtained after cloning. All sequences bind to FIS (+ or ++); there are no strong sequence preferences at the varied positions. In particular, no requirement for a certain pattern of adenines and thymines in position -2 through +2 emerges. Thus, these sequences suggest that one of the assumptions of selection theory, namely that each position contributes independently to the binding energy irrespective of the sequence context, is valid to a large degree in the region -2 through +2 and for the pairs +6,+7 and -6,-7. Small deviations from this assumption will result in the observed small variation in binding strength; these are not described well by selection theory.

The bending properties of sequences from mixture B are evidently not influenced by the sequence variations introduced, i.e. all sequences show the same degree of bending. This is in agreement with the assumption that the bending properties are determined by the flanks (positions outside of the core region -7 through +7) of the binding sites. In structural terms, this suggests that the interaction of the two helix-turn-helix motifs with two consecutive major grooves of the DNA requires the DNA to be bent in the core region. If bending the core region would require to much energy, the DNA will not bind at all. Bending of the flanking regions, however, would occur only if the energy required for this additional bending were compensated by protein DNA interactions made possible by the stronger wrapping of the DNA around the protein.

To summarize, the studies with artificial FIS binding sites show that the selection theory of Hippel and Berg is useful in describing the specificity of FIS. The λE values based

TABLE 3: Sequences obtained by in vitro selection. λE, prediction of strength by selection theory; gel retardation, relative binding strengths (–, +, ++) and bending (B, full; R, reduced).

clone	sequence	gel retardation	λE
	-7 -4 0 +4 +7		
mixture A	GGGGGT CAAAATTTG ACCAAAA		
B23	--T--- --------- -------	++B	2.7
B22	--T--- --------- -------	++B	2.7
B20	-T--T- --------- ------T	++B	2.8
B14	--T--- --------- -A-----	++B	2.8
B21	--C--- --------- -------	++B	2.8
B19	------ T-------- -A-----	++B	3.5
B10	----T- --------- -------	++B	3.8
B18	--T-T- --------- ----C--	++R	3.8
B03	------ -----G--- -----T-	++R	5.3
B05	------ ------G-- -T-----	+B	5.8
B08	------ --CT----- -------	+B	6.1
B15	-A---- ---T---A- -------	–	6.3
B12	------ -T------- ------C	–	6.4
B06	---A-- --------- -------	+R	6.7
B11	-----C -------G- ------C	–	7.5
B04	------ A---G---- -------	–	7.7
B09	---C-- T-------- -------	+B	8.0
B07	------ ------CGA T-A----	–	13.4
	-7 -4 0 +4 +7		
mixture B	ggtgNN caWWWWWtg NNcaaaa		
E17	----AT --TTTTT-- AG-----	+B	4.1
E14	----GG --AATTT-- GC-----	+B	4.1
F06	----AT --TATTT-- TT-----	++B	4.2
E08	----GT --TTATT-- GC-----	++B	4.7
E13	----TT --TTATT-- GG-----	+B	4.8
F16	----GG --ATTAT-- GC-----	+B	4.8
F03	----TT --ATTAA-- TG-----	++B	4.8
E02	----AT --TTTTT-- GT-----	++B	4.8
F01	----CG --TTATT-- AG-----	++B	4.9
F02	----TA --TTTAT-- AC-----	++B	4.9
F13	----TG --TTTTT-- GC-----	++B	5.7

on the sequence preferences of natural binding sites correlate with the strength of binding, but not with the degree of bending.

METHODS

The following primers were used to synthesize and amplify sequences with high similarity to FIS binding sites.

Sel53-A: GCTTGATATCGAATTC gggggtcaaaatttgaccaaaa TAGTTCTAGAGCGGC;
Sel53-B: GCTTGATATCGAATTC GGTGNNCAWWWWWTGNNCAAAA TAGTTCTAGAGCGGC;
SelEco: GCTTGATATCGAATTC; SelXba: GCCGCTCTAGAACTA; SelM13: GTAAAACGACGGCCAGT; SelM13Rev: AACAGCTATGACCATG; (N denotes A, C, G or T; W denotes A or T. Small letters denote a mixture of 7/8 of the given nucleotide and 1/8 of the remaining 3 nucleotides). Starting mixtures for in vitro selection were synthesized using primers SelM13Rev and either Sel53-A or Sel53-B (100 pmol each per 100 μl reaction) with plasmid pBluescript (SKII+)[6] as template. After 25 cycles of PCR, 1/5 of the reaction mixture and 4/5 of a mixture with fresh reagents were subjected to a further PCR cycle to ensure a high proportion of correctly paired double stranded DNA as opposed to re-annealed mismatched DNA. The products were 130 bp in length and carried the variable sequence at one end.

After selection (see below), the selected DNA was amplified using primers SelM13rev and SelEco. To analyze single sequences, PCR products were cloned into plasmid pBlueScript (SKII+) using the EcoRI and XbaI restriction common to both. From the clones, 220 bp segments that carry the inserted sequence in the center (as required for bending analyses) were amplified with primers SelM13Rev and SelM13.

Selection using immobilized FIS: FIS was coupled to NHS-activated Sepharose (Pharmacia) according to procedures given by the manufacturer (0.1 mg FIS, 1 ml column). To select for FIS-binding sequences, 50 μl of PCR product were diluted in 500 μl buffer (10 mM TrisCl, pH 7.5, 10 mM NaCl), bound to the column which had been equilibrated with the same buffer, and eluted with a NaCl gradient in the same buffer. For subsequent amplification, the eluted fractions were either de-salted on NAP-10 columns (Pharmacia) or by dilution.

Selection using gel retardation: A mixture of 2 μl PCR product and 8 μl FIS solution (5 ul/ml FIS, 20 % glycerol, 250 mM NaCl, 25 mM $MgSO_4$, 10 mM TrisCl, pH 8.6) were incubated for 15 min. and separated on 7.5 % polyacrylamide gels, both at room temperature. To visualize the DNA after the separation, the gels were stained with Ethidiumbromide. Bands containing DNA complexed by FIS were excised and incubated with 20 ul of a solution containing 500 mM NaCl to destroy the complex and elute the DNA from the gel. Diluted samples of the solution were used for the next round of PCR.

Analysis of bending and binding: PCR products 220 bp in length (see above) were analyzed simultaneously for binding and bending using the gel retardation method described above. The binding strength was evaluated from the proportion of free to bound DNA, while the degree of bending was evaluated from the mobility of the bound DNA. DNA amplified from a plasmid carrying the B22 sequence (table 2) was used as standard for strong binding and bending, DNA amplified from the original pBlueScriptSKII+ plasmid as one for weak binding and bending.

Sequencing of plasmids was performed using the T7 sequencing kit from Pharmacia.

ACKNOWLEDGEMENTS This study was supported by a fellowship from Studienstiftung des deutschen Volkes and by DFG-Sonderforschungsbereich 344.

REFERENCES

1. S. E. Finkel and R. C. Johnson. *Mol. Microbiol.*, **6**:3257–3265, 1992.
2. O. G. Berg and P. H. von Hippel. *J. Mol. Biol.*, **200**:709–723, 1988.
3. P. Hübner and W. Arber. *EMBO J.*, **8**:577–585, 1989.
4. D. Kostrewa, J. Granzin, C. Koch, H.-W. Choe, S. Raghunathan, Wolf, J. Labahn, R. Kahmann, and W. Saenger. *Nature*, **349**:178–180, 1991.
5. J. Xu and R. C. Johnson. *J. Bacteriol.*, **177**:5222–5231, 1995.
6. M. Alting-Mees and J. Short. *Nucleic Acids Res.*, **17**:9494–9494, 1989.