# SEQUENCE SPECIFICITY DICTATED BY THE INTERPLAY BETWEEN DIRECT READOUT AND DNA FLEXIBILITY AT THE TATA BOX-BINDING PROTEIN - TATA BOX INTERFACE.

Leonardo Pardo[1], David Bosch[1], Mercedes Campillo[1], Nina Pastor[2], and Harel Weinstein[3]

[1]Unidad de Bioestadistica, Facultad de Medicina, Universidad Autonoma de Barcelona, 08193 Bellaterra, Spain; [2]Facultad de Ciencias, UAEM, Av. Universidad 1001, Col. Chamilpa, 62210 Cuernavaca, Morelos, México;
[3]Department of Physiology and Biophysics, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York NY 10029, U.S.A.

## ABSTRACT:

A common mechanism of DNA bending by minor groove-binding proteins is the insertion of protein side chains between basepair steps, exemplified in TBP/DNA complexes. At the first and last basepair steps of the TATA box, TBP kinks the DNA by inserting pairs of Phe side chains between the steps, and placing Leu and Pro side chains near the rim of the bases. QM calculations indicate that these side chains cannot discriminate between AT and TA basepairs. The sequence selectivity is due to the differential DNA flexibility of the basepair steps, as revealed by MD/PMF calculations, and to the ability of these steps to form H-bonds in the major groove. At the central basepair step of the TATA box, TBP markedly untwists this step, while engaging in hydrogen bonds with the bases and sugars. The H-bonds drive the conformational transition at this step, but are not capable of discriminating between AA and AT steps, as their strength is the same for both sequences. The calculated free energy cost for an equivalent conformational transition is found to be sequence dependent, being higher for AA steps than for AT steps. Consequently, AA steps have a smaller distortion in TBP/DNA complexes than AT steps.

## INTRODUCTION:

•The TATA box-binding protein (TBP) binds specifically to 8 basepairs, using the minor groove surface of DNA. This mode of interaction is seen in all TBP-DNA complexes reported to date [1].
•The TATA box consensus sequence is TATA@A@N, where @ is A or T.
•The minor groove of DNA is considered poor in information content, given the very similar placement of H-bond acceptors (T-O2 and A-N3) and hydrophobic sites (A-C2) in A•T and T•A basepairs.
•The TBP-TATA box interface is mostly hydrophobic, with Leu, Pro, and Val side chains close to A-C2.
•H-bonds are only found at the central basepair step of the TATA element, between Asn and Thr residues and the H-bond acceptors in the minor groove.
•TBP bends and untwists the TATA box drastically. There are 45º kinks at the first and last basepair steps, and a 20º unwinding at the central basepair step.
•The energetic cost of the various components of DNA distortions involved in the specific binding of TBP can be used to reveal the mechanisms underlying sequence specificity [2].

## APPROACH:

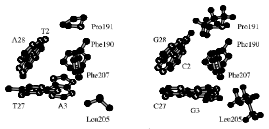
•We use quantum mechanical calculations to examine the interactions between TBP side chains and the basepair steps located at the most sequence conserved kink site (the 5' kink: the first TA step, at the MP2 level), and at the only step recognized through the formation of H-bonds(the central basepair step, using DFT/BLYP3), to determine the role of direct readout in sequence discrimination.
•We use Molecular Dynamics/Potential of Mean Force calculations with the AMBER 4.1 potential 43to estimate the free energy cost of transforming B- and A-DNA double stranded tetramers into the conformations found in high resolution crystal structures of TBP-DNA complexes, to determine the role of DNA bendability in TATA box selection by TBP.

## ENERGETICS OF DIRECT READOUT

### 5' KINK:

KINK bps: TA, AT, TT, AA, CG
TBP side chains: PHE, LEU AND PRO

Forces: STACKING
VAN DER WAALS




SCE complex
(Kim Y. et al., 1993)

| Basepair | F190 | F207 | P191 | L205 |
|---|---|---|---|---|
| **T2:A28** | **-4.4** | **-0.7** | **-1.6** | |
| A2:T28 | -3.4 | -0.7 | -1.5 | |
| C2:G28 | -3.0 | -0.6 | +1.5 | |
| **A3:T27** | **-7.0** | **-3.1** | | **-1.6** |
| T3:A27 | -5.8 | -3.5 | | -0.8 |
| G3:C27 | -7.1 | -3.3 | | +32.9 |

### FINDINGS

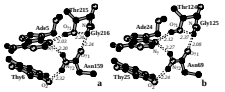
•The interaction of Phe side chains cannot discriminate among these four A•T basepair steps, or between A•T and C•G basepairs.
•Leu and Pro side chains clash against the N2 amino group in C•G basepairs, but cannot distinguish interactions with A•T from T•A basepairs.

REFERENCES:
[1]. Kim, Y. et al. (1993) Nature 365:520; Kim, J.L. et al. (1994) Nature Struct. Biol. 1:638; Nikolov, D.B. et al (1996) PNAS, USA 93:4862; Ju, Z.S. et al. (1996) J. Mol. Biol. 261:239; Patikoglou, G.A. et al. (1999) Genes Dev. 13:3217
[2]. Pastor, N. et al. (1997) Biophys. J.73:640; Pastor, N. et al. (1997) in Molecular Modeling of Nucleic Acids (ACS, Leontis, N.B. and SantaLucia Jr.,J., eds.) ) 268:329; Pardo, L., et al. (1998) Biophys. J. 74:2191; Pardo, L., et al. (2000) Biophys. J. in press; Pastor, N. and Weinstein, H. (2000) in Theoretical Biochemistry (Elsevier, Eriksson, L. ed.) in press.

### CENTRAL BP STEP:

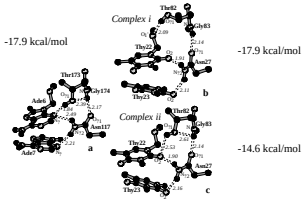
CENTRAL bps: AA, AT
TBP side chains: ASN, THR AND GLY

Forces: H-BONDS
-17.9 kcal/mol -17.8 kcal/mol




SCE complex: AT step
(Kim Y. et al., 1993)

-17.9 kcal/mol -17.9 kcal/mol




-14.6 kcal/mol

ATH complex: AA step
(Kim J.L et al., 1994)

### FINDINGS

•The strength of the H-bonds made from Asn and Thr side chains to AA or AT basepair steps is practically the same.
•Complex ii compensates the poorer interaction with DNA by improving the interaction within TBP.
•Gly is important to stabilize the conformation of the Asn and Thr side chains.

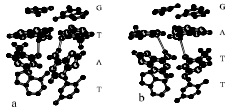
## SEQUENCE DEPENDENT DNA FLEXIBILITY

### 5' KINK:

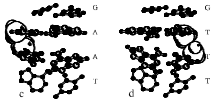
Calculation of free energy differences for the B ? TA transition of various bps

bps: ??F
GTAT: 0.0
GATT: 2.3
GTTT: 4.7
GAAT: 6.9

**?F(TA) < ?F(AT) due to better H-bonds**




**TT/AA unfavorable due to steric clash in the major groove**

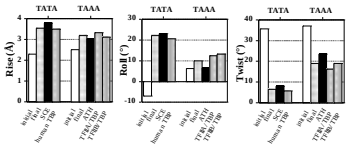



### CENTRAL BP STEP:

Free energy calculation of the A ? TA (ATH VS. SCE) transition for TAAA VS. TATA

transition ?F
TATA ? SCE: 11.8
TAAA ? ATH: 8.1

**From crystal structures and calculations: AT is always more distorted than AA**




Calculated ??F for fitting TAAA into the TATA structure of SCE:
14.4 - 11.8 = 2.6 kcal/mol

## CONCLUSIONS:

### 1. DIRECT READOUT:

•Direct readout is not responsible for the selection of TA basepair steps at the **5' kink**.
•TBP tolerates equally well AA and AT basepair steps at the **central basepair step** because the strength of the direct interactions to these two sequences is practically the same.

### 2. DNA DISTORTION:

**5' KINK:**
•TA steps are the easiest to bend into the TA-DNA conformation, because of the interactions in the major groove in the final conformation: two good intra-strand H-bonds can be made and there are no clashes.

**CENTRAL BASEPAIR STEP:**
•AT steps are more distorted than AA steps in TBP-DNA complexes, because AT steps are easier to unwind and bend than AA steps.