Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplementary Information

Conformational Selection Underpins Recognition of Multiple DNA sequences by Proteins and Consequent Functional Actions

Gitashri Naiya, Paromita Raha, Manas Kumar Mondal, Uttam Pal, Rajesh Saha, Susobhan Chaudhuri, Subrata Batabyal, Samir Kumar Pal, Dhananjay Bhattacharyya, Nakul C. Maiti and Siddhartha Roy

Table S1: Comparison of Protein-DNA interfaces of O1/LacI and O2/LacI complexes

KEI			KEJ		
Protein Atom	**DNA Atom**	**Distance**	**Protein Atom**	**DNA Atom**	**Distance**
K2:A (NZ)	**G13:D (OP2)**	**3.06**			
L6:A (N)	C14:D (OP2)	2.75	L6:A (N)	C14:D (OP2)	2.77
16S:A (N)	**G6:C (OP2)**	**3.07**	**16S:A (N)**	**G6:C (OP2)**	**2.84**
			16S:A (OG)	**G6:C (OP2)**	**2.82**
17Y:A (OH)	**G8:C (O6)**	**2.78**	**17Y:A (OH)**	**G8:C (O6)**	**3.06**
Q18:A (OE1)	**C16:D (N4)**	**3.13**	**Q18:A (OE1)**	**C16:D (N4)**	**3.35**
T19A:A (OG1)	**G6:C (OP2)**	**2.93**	**T19A:A (OG1)**	**G6:C (OP2)**	**2.59**
R22:A (NH2)	**T5:C (OP2)**	**2.66**			
N25:A (ND2)	**T15:D (OP2)**	**2.82**			
			S21:A (OG)	**T15:D (OP2)**	**2.69**
			S31:A (OG)	**T5:C (OP2)**	**2.88**
T34:A (OG1)	**T5:C (OP1)**	**2.77**	**T34:A (OG1)**	**T5C (OP1)**	**2.46**
Y47:A (OH)	**C14:D (OP1)**	**3.01**	**Y47:A (OH)**	**C14:D (OP1)**	**2.74**
N50:A (N)	C14:D (OP1)	3.05	N50:A (N)	C14:D (OP1)	3.03
N50:A (ND2)	**C14:D (OP1)**	**3.11**			
			A53:A (O)	**G13:D (N2)**	**3.41**
			K59:A (NZ)	**T15:D (OP1)**	**3.03**
			M1:B (N)	**G12:C (OP1)**	**2.68**
			L6:B (N)	**A13:C (OP2)**	**2.85**
S16:B (N)	G6:D (OP2)	2.94	S16:B (N)	G6:D (OP2)	2.81
S16:B (OG)	G6:D (OP2)	2.79	S16:B (OG)	G6:D (OP2)	2.79
T19:B (OG1)	G6:D (OP2)	2.61	T19:B (OG1)	G6:D (OP2)	2.58
			S21:B (OG)	**G14:C (OP2)**	**2.74**
R22:B (NH1)	**G6:D (O6)**	**2.90**			
R22:B (NH2)	**G6:D (N7)**	**2.90**			
N25:B (ND2)	T15:C (OP2)	2.84	N25:B (ND2)	T15:C (OP2)	2.81
			V30:B (N)	**T4:D (OP2)**	**3.15**
S31:B (N)	T5:D (OP2)	2.91	S31:B (N)	T5:D (OP2)	2.76
S31:B (OG)	T5:D (OP1)	2.92	S31:B (OG)	T5:D (OP1)	2.80
T34:B (OG1)	**T5:D (OP1)**	**2.52**			
			Y47:B (OH)	**A13:C (OP1)**	**2.54**
N50:B (N)	G13:C (OP1)	2.91	N50:B (N)	A13:C (OP1)	2.96
A53:B (O)	**G12:C (N2)**	**2.96**	**A53:B (O)**	**G12:C (N2)**	**2.79**
			K59:B (NZ)	**T15:C (OP2)**	**2.79**

Atom distances were obtained from WHATIF server.

Table S2

The Effect of Base Mutations in the Left half

Seq O_E	Seq sub.	K_d (nM)	$\Delta\Delta G^o$ Kcal. Mole^{-1}	SE	Seq O_I	Seq sub.	K_d (nM)	$\Delta\Delta G^o$ Kcal. Mole^{-1}	SE	P
G1	T	31±14	1.71	0.38	A1	T	820±77	3.18	0.21	***
	A	304±182	3.08	0.48		G	1000±77	3.30	0.21	
	C	48±21	1.99	0.38		C	906±90	3.24	0.21	***
T2	A	589±487	3.47	0.75	T2	A	1080±99	3.35	0.21	
	C	487±278	3.36	0.46		C	>2 μM	4.21		
	G	156±124	2.67	0.7		G	>2 μM	4.21		***
G3	C	>2 μM	4.21		G3	C	646±39	3.03	0.21	***
	A	437±398	3.29	0.96		A	285±32	2.55	0.21	
	T	458±318	3.32	0.57		T	307±42	2.59	0.22	
T4	A	599±255	3.48	0.37	T4	A	460±38	2.83	0.21	
	G	98±58	2.40	0.47		G	94±19	1.88	0.24	
	C	452±369	3.31	0.73		C	74±29	1.74	0.32	**
A5	T	110±53	2.47	0.40	A5	T	723±26	3.10	0.21	
	G	Not done				G	>2 μM	4.21		
	C	13.1±4.2	1.19	0.31		C	717±177	3.10	0.25	***
A6	T	483±216	3.35	0.38	A6	T	596±59	2.99	0.21	
	G	540±190	3.42	0.33		G	333±26	2.64	0.21	**
	C	92±64	2.36	0.57		C	533±50	2.92	0.21	
A7	G	36.9±17.8	1.81	0.40	C7	G	701±74	3.09	0.21	***
	C	32.5±15	1.73	0.39		A	642±45	3.03	0.21	***
	T	13.3±4.2	1.20	0.32		T	918±71	3.25	0.21	***
C8	G	>2 μM	4.21		C8	G	>2 μM	4.21		
	A	450±124	3.31	0.30		A	239±46	2.44	0.24	**
	T	Not done				T	63±11	1.64	0.23	

The standard errors of the ΔΔGs are the propagated errors from the wild-type and mutant sequences. $\Delta\Delta G^o$s and corresponding propagated errors were calculated using wild-type O_I K_d is 4.07±1.33 nM and wild-type O_E K_d of 1.8±0.7 nM. p of ** and *** indicate rejection of the null hypothesis at the level of 0.05 and 0.01 level of confidence, respectively. The standard errors used to calculate the p value are the average of the two standard errors of the corresponding positions. The K_ds that were estimated to be greater than 2 μM were derived from nearly linear binding isotherms in the measured concentration range; thus, the value reflects a lower limit. For the determination of the p value in these cases, the standard errors were assumed to be that of the other observation in the pair. The font color is color matched with the plot presented in Figures 1 and 2 of the manuscript. SE stands for the standard error of the mean.

Table S3

The Effect of Base Mutations in the Right half

Seq O_E	Seq sub.	K_d (nM)	$\Delta\Delta G^o$ Kcal. Mole^{-1}	SE	Seq O_I	Seq sub.	K_d (nM)	$\Delta\Delta G^o$ Kcal. Mole^{-1}	SE	P
G1'	C	194±76	2.80	0.35	G1'	C	101±34	1.92	0.29	
	A	62±22	2.12	0.33		A	210±84	2.36	0.33	
	T	202±58	2.83	0.31		T	128±49	2.07	0.32	*
T2'	A	206±62	2.84	0.31	T2'	A	56±15	1.57	0.26	***
	G	>2 μM	4.21			G	127±44	2.06	0.30	***
	C	278±135	3.02	0.39		C	26±22	1.11	0.78	**
G3'	C	564±225	3.45	0.36	G3'	C	29±13	1.18	0.35	***
	A	489±465	3.36	1.14		A	33±7	1.25	0.24	**
	T	464±278	3.33	0.48		T	1010±59	3.31	0.21	
G4'	C	155±64	2.67	0.36	G4'	C	452±49	2.82	0.21	
	A	106±33	2.44	0.32		A	>2 μM	4.21		***
	T	64±30	2.14	0.39		T	649±103	3.04	0.23	**
A5'	T	48±23	1.97	0.40	T5'	A	230±80	2.42	0.31	
	G	316±104	3.10	0.32		G	1220±223	3.42	0.23	
	C	>2 μM	4.21			C	393±112	2.74	0.27	***
A6'	T	>2 μM	4.21		A6'	T	783±93	3.15	0.22	***
	G	>2 μM	4.21			G	420±85	2.78	0.24	***
	C	>2 μM	4.21			C	58±20	1.59	0.30	***
T7'	C	30±7.7	1.69	0.29	G7'	C	58±24	1.59	0.33	
	A	361±130	3.18	0.34		A	81±30	1.79	0.31	***
	G	4.4±3.9	0.53	0.88		T	>2 μM	4.21		***
C8'	G	525±252	3.40	0.40	C8'	G	18±5	0.89	0.27	***
	A	31±13	1.71	0.37		A	676±152	3.06	0.25	***
	T	>2 μM	4.21			T	>2 μM	4.21		

The standard errors of the $\Delta\Delta G$s are the propagated errors from the wild-type and mutant sequences. $\Delta\Delta G^o$s and corresponding propagated errors were calculated using wild-type O_I K_d is 4.07±1.33 nM and wild-type O_E K_d of 1.8±0.7 nM. p of ** and *** indicate rejection of the null hypothesis at the level of 0.05 and 0.01 level of confidence, respectively. The standard errors used to calculate the p value are the average of the two standard errors of the corresponding positions. The K_ds that were estimated to be greater than 2 μM were derived from nearly linear binding isotherms; thus, the value reflects a lower limit. For the determination of the p value in these cases, the standard errors were assumed to be that of the other observation in the pair. The font color is color matched with the plot presented in Figures 1 and 2 of the manuscript. SE stands for the standard error of the mean.

Table S4: Comparison of Protein-DNA interfaces of O_E/GalR and O_I/GalR complexes

Protein Residue(Chain)	O_E DNA Residue (Chain)	O_E DNA Atom	O_E distance	O_E Angle	Protein Residue(Chain)	O_I Residue (Chain)	O_I DNA Atom	O_I distance	O_I angle
I5-N(A)	G10(D)	O2P	2.82	166.7	I5-N(A)	T10(D)	O2P	2.78	166.7
S15-N	G2(C)	O2P	2.95	157.3	S15-N(A)	G2(C)	O1P	3.1	120.4
					S15-N(A)	G2(C)	O2P	3.1	150.0
S15-OG(A)	G2(C)	O2P	2.55	165.9	S15-OG(A)	G2(C)	O2P	2.51	175.7
S20-OG(A)	T11(D)	O2P	2.63	162.4	S20-OG(A)	T11(D)	O2P	2.99	158.5
R21-NH1(A)	G2(C)	N7	3.11	163.7	R21-NH1(A)	G2(C)	O6	2.70	143.8
R21-NH2(A)	G2(C)	O6	2.80	170.4	R21-NH2(A)	G2(C)	N7	2.90	169.6
N24-ND2(A)	T11(D)	O1P	2.75	164.5	N24-ND2(A)	T11(D)	O1P	3.28	165.8
Y46-OH(A)	G10(D)	O1P	2.80	175.0	Y46-OH(A)	T10(D)	O1P	3.07	146.6
N49-N(A)	G10(D)	O1P	2.89	138.8	N49-N(A)	T10(D)	O1P	2.89	178.3
R53-NH1(A)	T11(D)	O1P	2.84	147.3	R53-NH1(A)	T11(D)	O1P	2.80	162.9
Q58-NE2(A)	T12(D)	O1P	2.97	168.9					
A52-O(A)	G9(D)	N2	3.46	142.2	A52-O(A)	G9(D)	N2	3.20	142.2
					K1-NZ(B)	C7(C)	O1P	2.83	147.5
I5-N(B)	C9(C)	O2P	2.94	172.1	I5-N(B)	A9(C)	O2P	2.87	174.5
S15-N(B)	G13(D)	O2P	2.77	168.2	S15-N(B)	G3(D)	O2P	2.81	173.6
S15-OG(B)	G13(D)	O2P	2.80	175.7	S15-OG(B)	G3(D)	O2P	3.07	140.3
					S15-OG(B)	G3(D)	5'	3.23	128.2
S20-OG(B)	T10(C)	O2P	2.81	171.0	S20-OG(B)	T10(C)	O2P	2.65	171.1
S20-OG(B)	T10(C)	O5'	3.39	123.5					
R21-NH1(B)	G13(D)	N7	2.99	126.3	R21-NH1(B)	G3(D)	N7	3.02	142.2
R21-NH1(B)	G13(D)	O6	2.96	143.0	R21-NH2(B)	G3(D)	O6	2.94	176.9
R21-NH2(B)	C13(C)	N4	3.44	132.0	R21-NH2(B)	C12(C)	N4	3.34	128.5
R21-NH2(B)	G13(D)	O6	2.86	147.4					
N24-ND2(B)	T10(C)	O1P	2.97	166.3	N24-ND2(B)	T10(C)	O1P	2.74	165.4
K28-NZ(B)	G1(D)	N7	2.88	170.6					
S30-N(B)	T2(D)	O2P	2.92	161.2	S30-N(B)	T2(D)	O2P	2.84	170.1
S30-OG(B)	T2(D)	O1P	2.68	170.0	S30-OG(B)	T2(D)	O1P	2.66	164.1
					S30-OG(B)	T2(D)	O2P	3.44	130.5
S33-OG(B)	T2(D)	O2P	2.83	170.0	S33-OG(B)	T2(D)	O1P	2.87	147.3
					S33-OG(B)	T2(D)	O2P	3.40	142.5
Y46-OH(B)	C9(C)	O1P	2.67	160.2	Y46-OH(B)	A9(C)	O1P	2.88	166.9
N49-N(B)	C9(C)	O1P	3.41	164.9	N49-N(B)	A9(C)	O1P	3.02	166.6
A52-O(B)	G8(C)	N2	2.82	170.7	A52-O(B)	G8(C)	N2	3.29	149.0
					R53-NE(B)	A9(C)	O3'	3.22	143.5
					R53-NE(B)	T10(C)	O1P	2.99	149.9
					R53-NH2(B)	T10(C)	O1P	2.90	157.6

Table S5

Regulatory region sequence of the reporter gene constructs used in this study

The plasmid composition	Regulatory region gene contains	Sequence of the regulatory region
O_EO_I	Operators O_E & O_I, promoters P1 & P2	5' GGATCC GTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCAT CTTGTTATGCTATGGTTATTTCATACCATAAGCCTAATGGAGCGA ATTATGA GAGTTCTGGTTACCGGTGGTAGCGGTTACATTCTAGA –3'
O_EO_E	Operators O_E & O_E, promoters P1 & P2	5'GGATCC GTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCAT CTTGTTATGCTATGGTTATTTCATACCATAAGCCTAATGGAGCGA ATTATGA GAGTTCTGGTTACCGGTGTAAACGATTCCACTCTAGA –3'
O_IO_I	Operators O_I & O_I, promoters P1 & P2	5'GGATCCGTGGTAGCGGTTACATTAATTTATTCCATGTCACACTTTTCGCAT CTTGTTATGCTATGGTTATTTCATACCATAAGCCTAATGGAGCGA ATTATGA GAGTTCTGGTTACCGGTGGTAGCGGTTACAT TCTAGA–3'

The operator O_E is highlighted yellow, O_I is highlighted Magenta and the promoters are highlighted green.

Figures

Figure S1: Sequences of Gal operators O_E and O_I. Red marked bases show the differences between the two operators. The sequences in bold represent actual operator sequences which are flanked by two GC base pairs. The numbering is shown at the top.

Figure S2: The red marked bases are at positions 7/7', the most variable position in the sequences. The logo was designed by webserver http://weblogo.berkeley.edu/logo.cgi

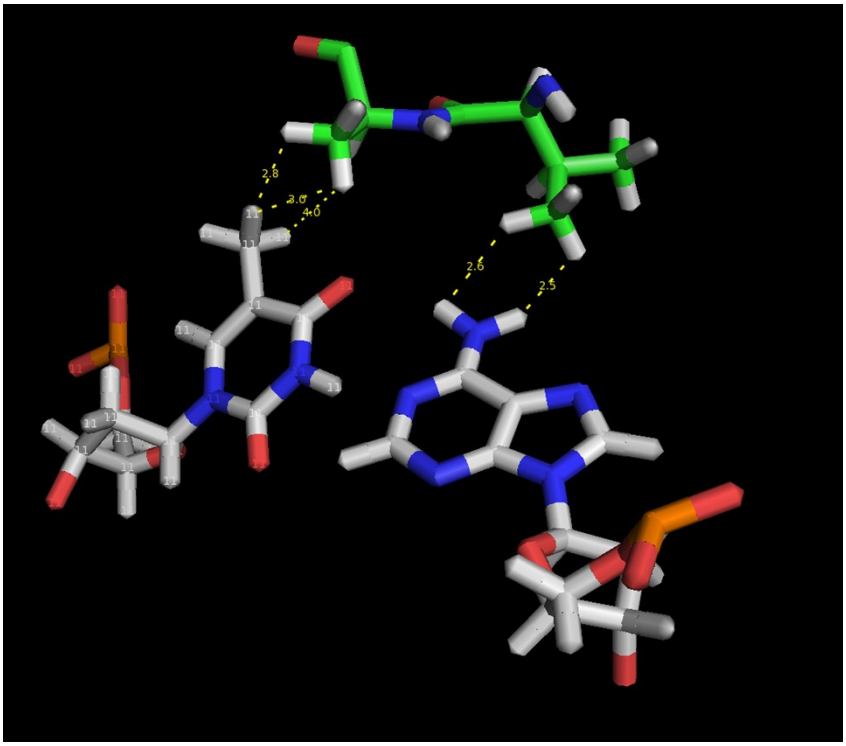
(A) Determination of consensus sequence based on four half sites of Gal operators.

G	**T**	**G**	**T**	**A**	**A**	**A**	**C**	**O_E --sequence 1**
G	**T**	**G**	**G**	**A**	**A**	**T**	**C**	**O_E--sequence 2**
G	**T**	**G**	**G**	**T**	**A**	**G**	**C**	**O_I--sequence 3**
A	**T**	**G**	**T**	**A**	**A**	**C**	**C**	**O_I--sequence 4**
G_3A	**T_4**	**G_4**	**G_2T_2**	**A_3T**	**A_4**	**ATGC**	**C_4**	**Distribution**
G	**T**	**G**	**G/T**	**A**	**A**	**N**	**C**	**Consensus**

(B) Sequence Logo

Figure S3: (A) Structure and interaction of 5'-basepair O_E (5' refers to position No.). (B) Structure and interaction of 5'-basepair O_I (5' refers to position No.)

A



B

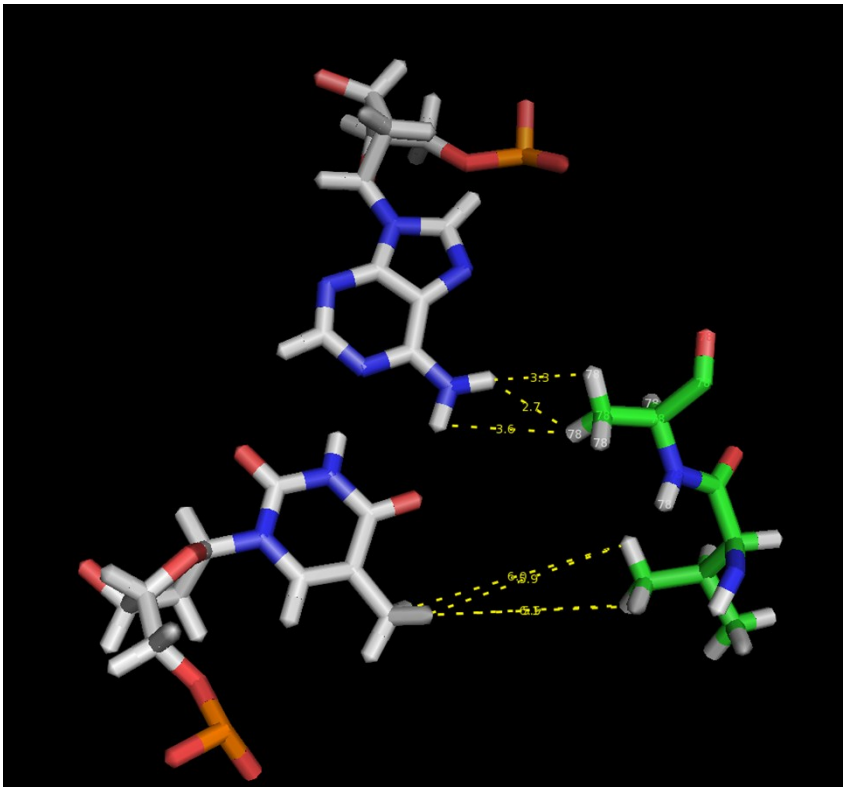
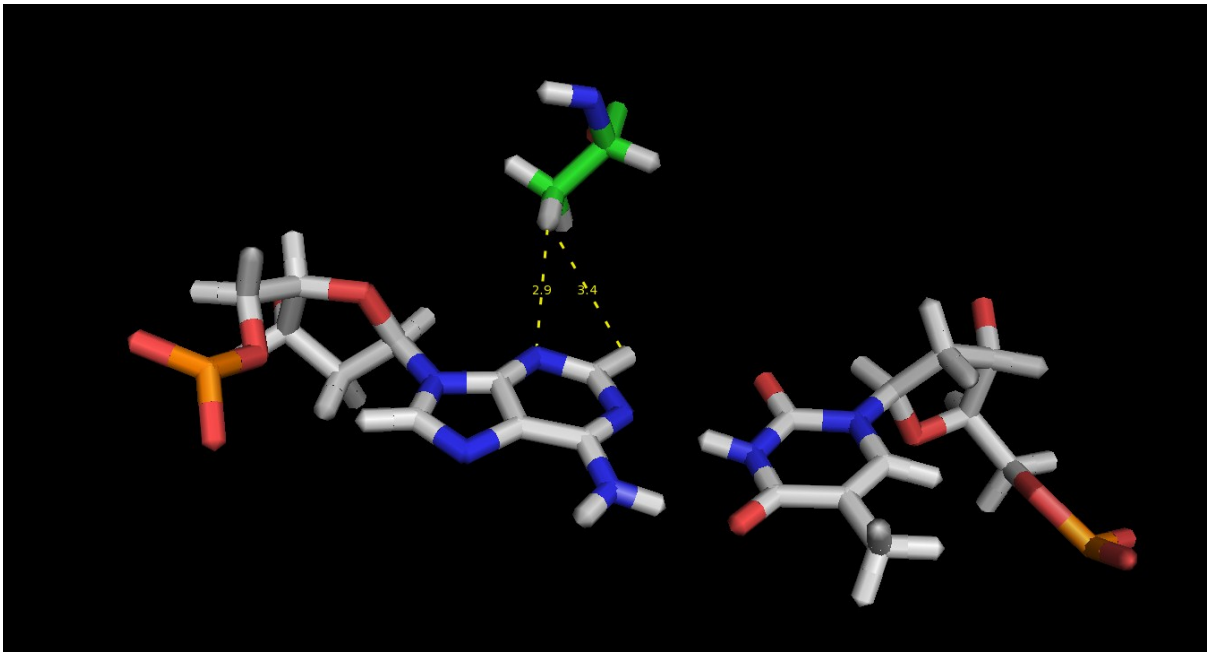


Figure S4: (A) Structure and interaction of 7'-basepair O_E. (B) Structure and interaction of 7'-basepair O_I.

A



B