CSL-DNA interactions

CSL achieves DNA site specificity through a combination of major and minor groove interactions in which major groove contacts are primarily contributed by the NTD and minor groove contacts by the BTD. Four strictly conserved residues (Arg234, Gln226 and Lys368 from NTD and Gln401 from BTD), plus the backbone carbonyl of Ser400, function in sequence-specific DNA recognition.

The guanadium group of Arg234 from the DNA recognition loop makes a bidentate, almost planar, interaction with Gua 8 in the major groove. Sidechain atoms NH1 and NH2 hydrogen bond with atoms O6 and N7 of the base (Figure 7B). The Arg234 sidechain is positioned by the interactions of its NE and NH1 atoms with the carboxylic group of Glu232. Only guanine can accept the two hydrogen bonds in this major groove interaction with an arginine sidechain . This is consistent with *in vivo* and *in vitro* data showing an overwhelming preference for guanosine at this position (Tun *et al.*, 1994). The CSL recognition-loop conformation and the arginine-DNA interaction are strikingly similar to those in DNA complexes with Rel proteins, and in arginine-guanine contacts as commonly observed in protein-DNA complexes (Ghosh *et al.*, 1995; Luscombe *et al.*, 2001; Muller *et al.*, 1995).

The side chain of Gln226 makes a water-mediated bifurcated interaction with the N7 and O6 of Gua 9 (Figure 7B). The NH2 nitrogen of Arg234 also contributes a hydrogen bond to this water. The hydrogen-bond distances from the water to the Gua base are not ideal, approximately 3.6 Å; however, the conservation of this residue, the geometry of the interactions, and the similar B-factors of associating atoms suggest that this interaction is relevant to DNA site specificity by CSL. This hydrogen bonding

stereochemistry only specifies a purine contact, whereas *in vitro* data show an approximate 78% preference for Gua and only a 13% preference for Ade at position 9. The preference for guanine over adenine at this site is not clear, as it seems that an Ade - water - Gln association could make quasi-equivalent interactions, but it is noteworthy that the aforementioned specific interaction of Gua8 with Arg 234 is supported by greater than 96% selectivity for a guanine base at position 8 (Tun *et al.*, 1994).

Lys368 of CSL, which is a part of the domain interlinker region, also makes specific major groove contacts with DNA but in a more complex manner that involves diagonal interactions with sequential base steps Gua 10 and Thy 7′ (Figure 7C). Lys368 is positioned such that its ε-amino group interacts simultaneously with the O4 atom of Thy 7′ and the O6 atom of Gua 10. This type of diagonal interaction was also seen in the structure of the Rel protein P50 bound to DNA (Ghosh *et al.*, 1995; Muller *et al.*, 1995). This interaction requires guanosine and thymine at the respective positions as only these bases have the required acceptors for hydrogen bonds from the lysine side chain. Similarly, *in vitro* data show a very high preferences for a Gua at position 10 and a Thy at position 7′ (Tun *et al.*, 1994).

CSL makes additional specific DNA contacts in the minor groove through the novel utilization of a beta-trefoil domain. The extended loop between strands βA1 and βA2 inserts into the minor groove and positions the sidechain of Gln401 for its NE2 amido group to donate a hydrogen bond to N3 of Ade 13′ (Figure 7D). NE2 of Gln401 also hydrogen bonds to the O4′ ribose oxygen of Gua 14′ to further orient the interaction with Ade 13′. In addition, the backbone carbonyl of Ser400 interacts with the N2 nitrogen atom of Gua 6 (Figure 7D). Only a guanine base has the requisite amino group to interact

with this carbonyl group; however, either adenine or guanine could make the interaction seen with the sidechain of Gln401. In fact, a purine base is exactly what is observed in *in vitro* selection experiments and in *in vivo* CSL promoter binding sites (Tun *et al.*, 1994).

CSL also makes numerous nonspecific interactions with DNA, the importance of which is illustrated by the strict conservation of the residues that participate in these interactions (Figures 1A and 7A). These nonspecific interactions involve primarily basic and polar amino acid sidechains that interact with phosphate groups on DNA. In addition, nonspecific van der Waals contacts are made by the sidechains of Phe235 and Tyr229 with the ribose moieties of Thy 7′ and Gua 6. These non-specific contacts are consistent with previous CSL mutagenesis and DNA binding studies. In particular, the mutation of Arg397 to histidine and Arg399 to glycine resulted in less than 10% of wild-type binding, and the mutation of Lys476 decreased binding to 32% to that of wild-type protein (Chung *et al.*, 1994).

References

Chung CN, Hamaguchi Y, Honjo T, Kawaichi M (1994) Site-directed mutagenesis study on DNA binding regions of the mouse homologue of Suppressor of Hairless, RBP-J kappa. *Nucleic Acids Res* **22**: 2938-2944

Ghosh G, van Duyne G, Ghosh S, Sigler PB (1995) Structure of NF-kappa B p50 homodimer bound to a kappa B site. *Nature* **373**: 303-310

Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **29**: 2860-2874

Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**: 5108-5121

Muller CW, Rey FA, Sodeoka M, Verdine GL, Harrison SC (1995) Structure of the NF-kappa B p50 homodimer bound to DNA. *Nature* **373**: 311-317

Tun T, Hamaguchi Y, Matsunami N, Furukawa T, Honjo T, Kawaichi M (1994) Recognition sequence of a highly conserved DNA binding protein RBP-J kappa. *Nucleic Acids Res* **22**: 965-971

Table 2. DNA sequence and conformation*

Basepair	Conformation Analysis					Site Preference# (%G, %A, %T, %C)
	Incline Angle (deg)	Propeller Twist (deg)	Basepair Buckle (deg)	Basepair Opening (deg)	Helical Twist (deg)	
3-A:T-15′	5.04	-7.67	2.20	-5.85	31.28	n/a
4-C:G-14′	4.79	-8.14	0.94	-3.33	30.85	n/a
5-T:A-13′	6.23	-6.40	-8.02	-7.33	39.45	(3.3, 3.3, 16.7, 76.7)
6-G:C-12′	3.02	-5.72	5.07	-5.49	36.77	(90.6, 6.3, 3.1, 0)
7-T:A-11′	4.21	-9.63	-4.29	-2.05	32.29	(0,0,100,0)
8-G:C-10′	7.56	5.91	7.79	0.18	36.56	(96.9, 3.1, 0, 0)
9-G:C-9′	3.58	-3.71	10.26	-1.14	27.56	(78.1, 12.5, 9.4, 0)
10-G:C-8′	3.26	-1.01	6.60	5.23	41.53	(96.9, 3.1, 0, 0)
11-A:T-7′	0.01	-11.49	2.62	8.82	31.53	(0, 93.8, 6.2, 0)
12-A:T-6′	0.78	-10.14	1.65	4.22	37.75	(9.7, 90.3, 0, 0)
13-A:T-5′	-2.14	-12.84	-5.00	6.64	38.89	(0, 42.9, 17.8, 39.3)
14-G:C-4′	-0.93	-9.98	-7.21	4.51	34.14	n/a
15-A:T-3′	1.00	-5.16	-9.14	-3.26	n/a	n/a

* DNA analysis was performed by 3DNA (Lu & Olson, 2003)
# Nucleotide site preferences are derived from (Tun *et al.*, 1994)