STRUCTURAL
BIOLOGY

Volume 73 (2017)

Supporting information for article:

Mismodeled purines: implicit alternates and hidden Hoogsteens

Bradley J. Hintze, Jane S. Richardson and David C. Richardson

Hidden Hoogsteens in the Data

The Supplement

Bradley J. Hintze, Jane S. Richardson, and David C. Richardson

Contents

1 Potential purine *anti/syn* decoys identified by *find_purine_decoys*

A list of all base pairs identified as *strong* and *very strong* purine decoys by *find_purine_decoys*. The 'Evidence' column gives a qualitative score reflecting the strength of electron density evidence present to justify a flip:

- **N**: The modeled conformation is correct.
- **Y**: Strong evidence exists for a flip.
- **n**: The electron density is ambiguous but the modeled conformation is more probable to be correct.
- **y**: The electron density is ambiguous but a flip may be appropriate.
- **?**: The electron density is too ambiguous to tell if either conformation is correct.

The 'Confirm' column indicates that rebuilding and refinement were performed. A check mark confirms the flipped conformation to be correct. An **N** indicates that a purine flip is NOT appropriate but some other diagnosis explains the difference density; this is further explained in the 'Comments' column.

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
1JKR	2.27	B	16	A	DA	Y	✓	symmetry pair, sticky end
1S97	2.40	J	7		DA	Y	✓	n-2 position of the DNA polymerase
2WQ6	2.30	D	8		DG	y	✓?	no pair, (6-4)DNA photolyase
2WQ7	2.00	D	8		DG	Y	✓	
2XM3	2.30	Q	12		DG	Y	✓	symmetry pair, sticky end, DNA transposase
2XO6	1.90	E	12		DG	Y	✓	

Continued on next page

Table S1 – continued from previous page

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
3GX4	2.70	Z	219		DA	Y	✓	modeled HG, should be WC
3HXO	2.40	B	23		DG	Y	✓	DNA aptamer, neither HG or WC
3BRF	2.47	C	2		DA	Y	✓	symmetry pair, sticky end, C 1 modeled as HG
3ODH	2.30	G	12		DA	Y	✓	modeled HG, should be WC
3V6J	2.30	B	7		DA	Y	✓	n-2 position of DNA polymerase with adjacent modified residues.
3V6T	1.85	H	2		DA	Y	✓	terminal blunt end, DNA-bound dHax3
3V9W	1.70	G	7		DA	Y	✓	single strand, no pair
4DTN	1.96	T	3		DA	Y	✓	no pair, reading strand in DNA polymerase0
4I2O	1.77	X	5		DA	Y	✓	modeled HG, should be WC, unmodeled alternates or up-side-down duplex
1C0W	3.18	F	501		DA	n		
1D8X	1.04	B	14		DA	N		
1EN9	0.98	A	6	A	DG	N		
1F4K	2.30	E	21		DG	N		

Continued on next page

Table S1 – continued from previous page

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
1F5T	3.00	F	406		DA	N		
1GU4	1.70	C	15		DA	N	**N**	Unmodeled symmetry alternates
1GU5	2.08	D	101		DA	N		Unmodeled symmetry alternates
1JT0	2.87	F	30		DA	N		
1K7A	2.79	C	2		DA	N		
1L3T	1.70	C	12		DG	N		
1NVP	2.10	E	6		DG	N		
1ORN	1.70	B	1		DA	?		alternate?, no pair, terminal end
1OUQ	3.20	X	117		DG	?		
1OWF	1.94	D	22		DA	N		
1PP8	3.00	K	29		DA	y		alternate?
1Q3U	2.90	C	101		DG	y		
1Q9X	2.70	F	928		DG	N		
1SXQ	1.80	D	1		DA	N		
1T2K	3.00	E	23		DA	N		
1U0C	2.50	D	568		DG	N		
1U78	2.69	B	5		DG	N		
1WTE	1.90	Y	7		DG	N		
1XSL	2.28	N	6		DG	N		
1ZJM	2.10	D	5		DG	N		
2A07	1.90	A	1		DA	?		
2AS5	2.70	D	5001		DA	y		sticky end
2AXY	1.70	E	499		DA	y		no pair, alternate?
2B9S	2.27	E	105		DA	N		

Continued on next page

Table S1 – continued from previous page

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
2DP6	1.80	C	12		DG	N		
2E42	1.80	C	4		DG	N		Unmodeled symmetry alternates
2EFW	2.20	D	12		DA	N	**N**	the duplex was fit in the ED up side down.
2GII	2.10	F	5		DG	N		
2IIE	2.41	C	-39		DA	N		
2NQB	2.30	J	246		DG	N		
2OFI	1.85	C	7		DG	N		
2PYJ	2.03	R	5		DA	N		
2VY2	2.30	W	5	A	DA	N	**N**	half site, deceptive symmetry
2W8L	2.90	T	3	B	DA	?		
2YPA	2.60	F	23		DA	N		
3A4K	2.17	N	6		DG	?		
3A5T	2.60	C	6		DG	N		
3AAF	1.90	C	7		DA	N		
3AV1	2.50	I	145		DA	N	**N**	nucleosome, sequence misalignment
3AZG	2.40	I	58		DG	N		ambiguous density
3D1N	2.51	D	6		DA	n		
3DPG	1.91	C	3	B	DG	?		no pair, modeled syn/anit alts
3EEO	1.94	C	411		DG	N		wrong sequnce

Continued on next page

Table S1 – continued from previous page

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
3G73	2.21	C	1		DA	N		alternates symmetrical DNA binding, unmodeled alternate strand
3G99	1.80	C	1		DA	Y	**N**	unmodeled alternate strand
3G9I	1.85	D	8		DA	Y	**N**	unmodeled alternate strand
3G9O	1.65	D	1		DA	Y	**N**	unmodeled alternate strand
3G9P	1.65	D	1		DA	Y	**N**	unmodeled alternate strand
3H8X	1.95	C	279		DA	N		
3HP6	1.81	F	12		DG	N		
3HT3	1.70	C	8		DG	N		
3JXC	1.90	A	40		DG	N		
3MKW	2.99	T	4		DG	N		
3MKY	2.86	T	4		DG	N		
3MQ6	2.00	R	2		DA	n		
3NDH	1.30	C	11	B	DG	N		
3OD8	2.40	I	1	B	DG	N		
3ODC	2.80	E	5		DG	N		
3P57	2.19	G	1		DA	Y	**N**	symmetry pair, sticky end, symmetrical DNA binding, unmodeled alternate strand

Continued on next page

Table S1 – continued from previous page

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
3PX0	1.73	F	12		DG	N		
3Q8P	1.95	P	868		DG	N		
3QX3	2.16	D	10		DG	N		
3S3N	2.49	C	1	B	DA	N		
3SI6	1.85	P	105		DG	N		
3V4I	2.80	T	705		DA	n		
3V6D	2.71	T	707		DG	?		
3WAA	3.19	J	186		DG	N		nucleosome, sequence misalignment
4B9V	1.97	B	11		DA	y		alternate?
4BDP	1.80	T	3		DA	N		
4BXO	2.15	H	11		DA	y		alternate?
4E0G	2.20	D	16		DA	N		
4EA4	2.00	C	11		DG	N		
4ELV	1.90	B	102		DA	N		
4EZ2	1.60	A	9		DG	N		
4F2S	1.65	F	4		DG	N		
4G4R	1.89	C	9		DA	n		
4HC9	1.60	Z	2		DA	n		
4HF2	2.20	D	5		DA	N		
4HIO	1.75	B	5		DA	n		no pair
4IBU	1.70	G	12	B	DG	N		
4K8X	2.28	T	2	A	DA	N		ambiguous density
4L0Z	2.70	C	1		DG	Y	?	symmetry pair, sticky end, density too ambiguous to tell the correct conformation

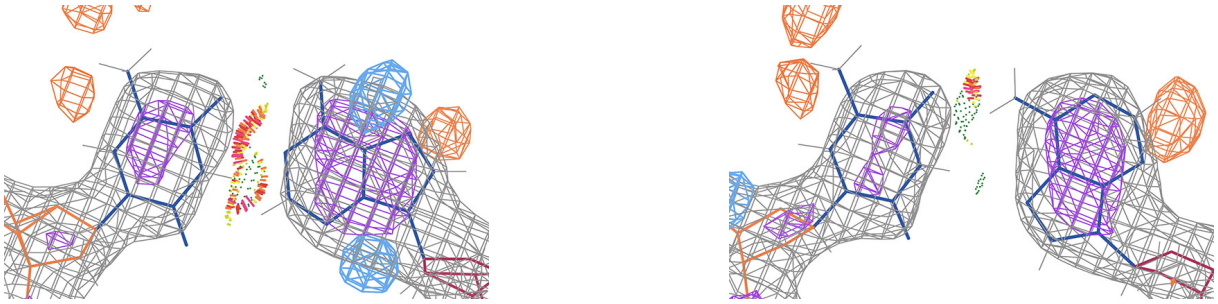
Continued on next page

Table S1 – continued from previous page

PDB	Resolution	Chain	#	Alt	Type	Evidence	Confirm	Comments
4L62	2.90	R	4		DA	N		the duplex was fit in the ED up side down.
4LNQ	2.00	C	20		DG	?		
4NNU	2.81	C	2		DA	n		incorrect sequence?
4O6A	1.86	C	16		DA	?		

2 Before and after figures for confirmed flips

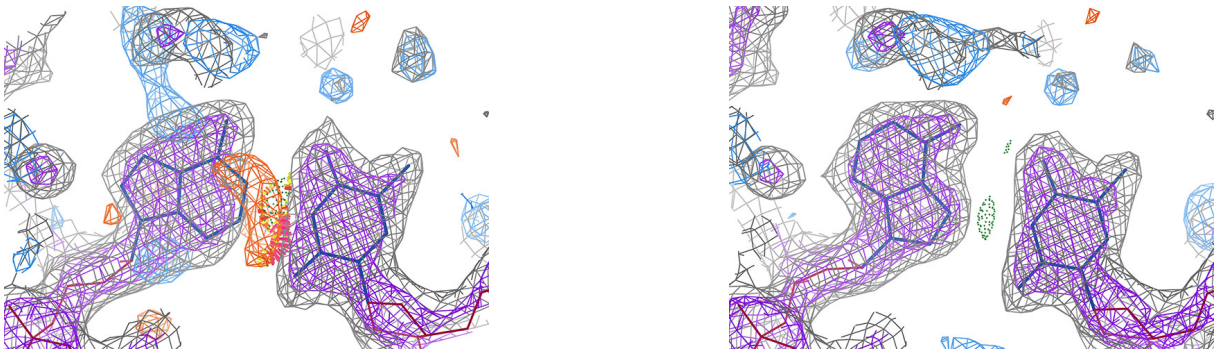
2.1 1JKR



(a) Original (b) Flipped

Figure S1: dT (B 17) : dA (B 16). Sticky end. This is the small HHIN recombinase DNA binding domain bound to DNA

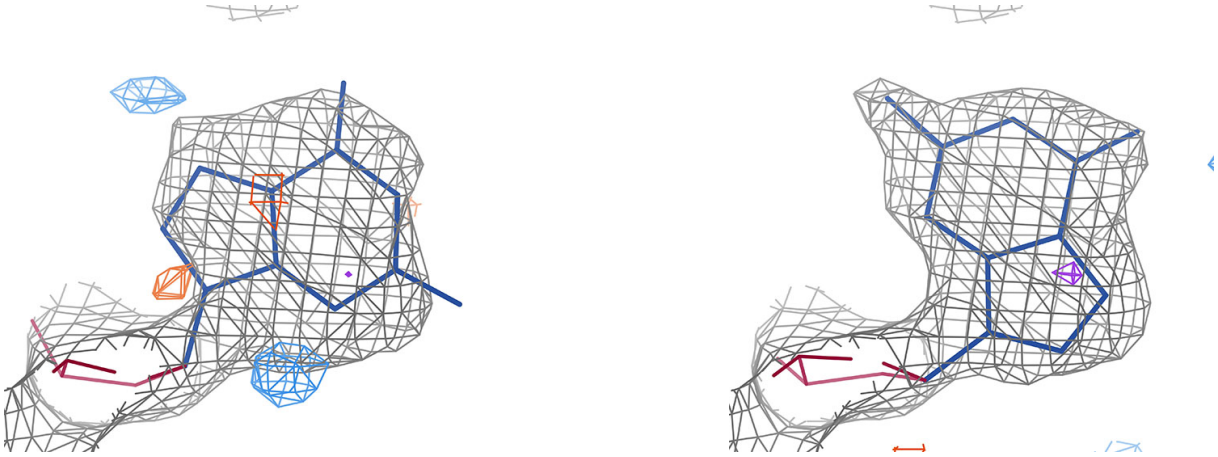
2.2 1S97



(a) Original (b) Flipped

Figure S2: dA (J 7) : dT (F 12). This base pair is in the n-2 position of the DNA polymerase Dpo4 from *Sulfolobus solfataricus*. Adjacent, in the n-1 position, is a G•T mismatch in a reverse wobble conformation. There are 3 other molecules in the asymmetric unit; all seem to require a flip of this dA.

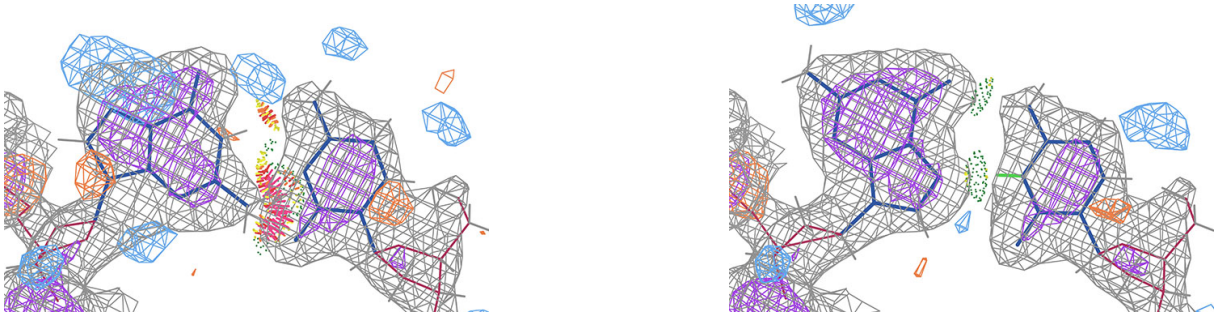
2.3 2WQ7



(a) Original (b) Flipped

Figure S3: dG (D 8). This base is part of a duplex but its intended pair is a lesion bound to (6-4)DNA photolyase. While not a HG example, the guanine clearly should be modeled in the *syn* conformation.

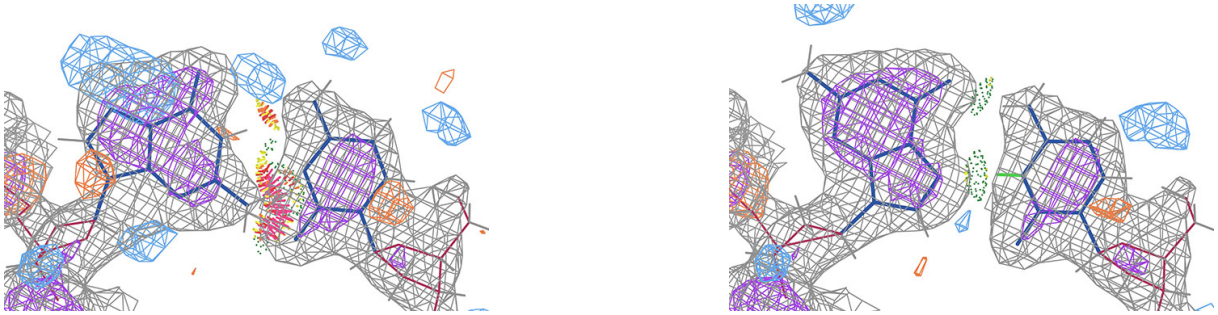
2.4 2XM3



(a) Original (b) Flipped

Figure S4: dG (Q 12) : dC (K 11). This is a symmetry-related base pair, i.e. the two bases are in separate asymmetric units. The structure is of DNA transposase and there are 3 dimers in the asymmetric unit. Along with Q 12, I 12 shows strong evidence for a flip and M 12 may need to flip but the density is less clear.

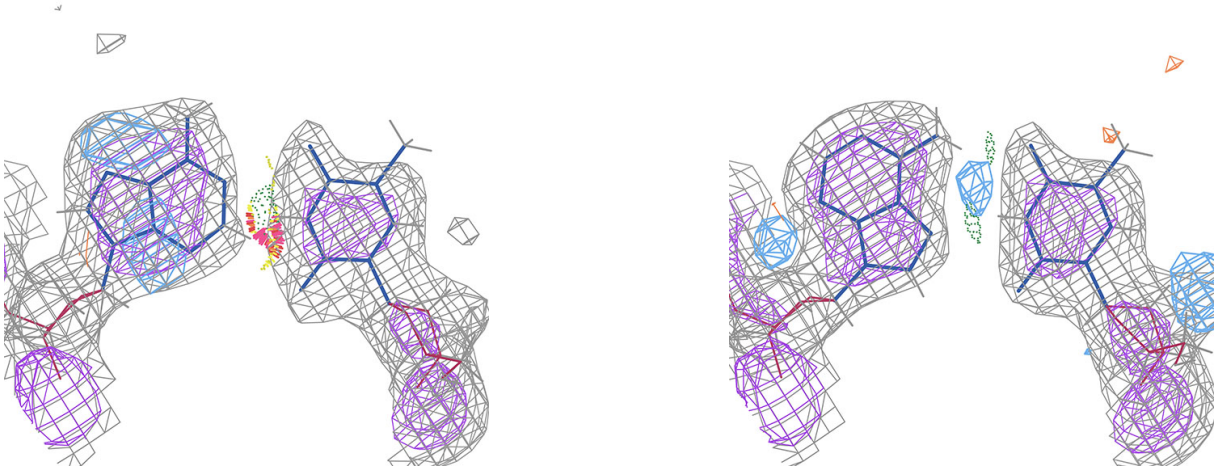
2.5 2XO6



(a) Original (b) Flipped

Figure S5: dG (E 12) : dC (B 11). This is a symmetry-related base pair, i.e. the two bases are in separate asymmetric units. This comes from the same study as 2XM3 and is very similar except there is only one dimer in the asymmetric unit.

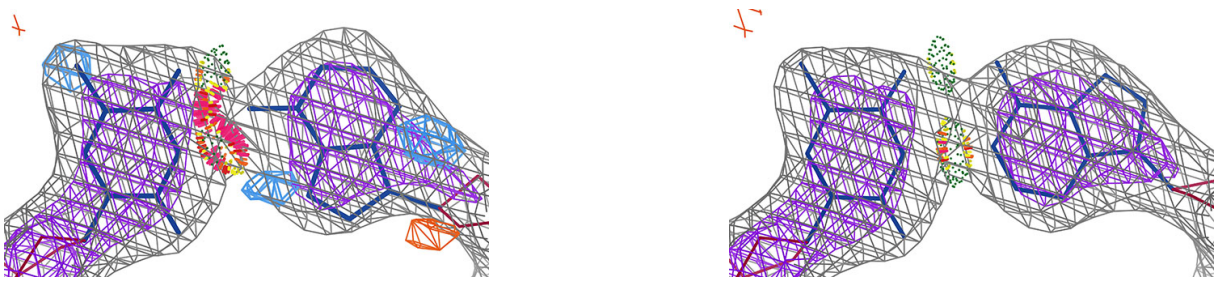
2.6 3BRF



(a) Original (b) Flipped

Figure S6: dG (C 2) : dC (B 1). This is a symmetry-related base pair, i.e. the two bases are in separate asymmetric units. The duplex is bound to Lag-1 (CSL).

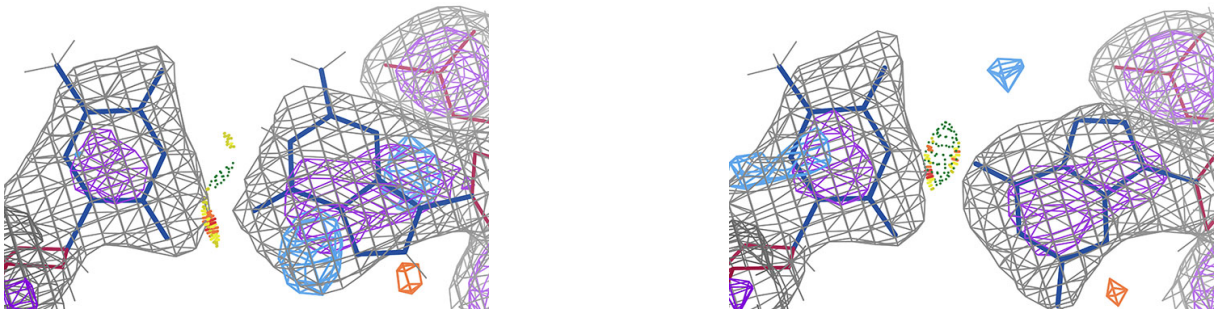
2.7 3GX4



(a) Original (b) Flipped

Figure S7: dT (Y 209) : dA (Z219). Originally modeled as a Hoogsteen, this is really a canonical Watson-Crick A•T.

2.8 3HXO



(a) Original (b) Flipped

Figure S8: dT (B 35) : dG (B 23). This base pair interaction is not in a canonical DNA helix but rather in a DNA aptamer and should be in the *anti* conformation.

2.9 3ODH

(a) Original (b) Flipped

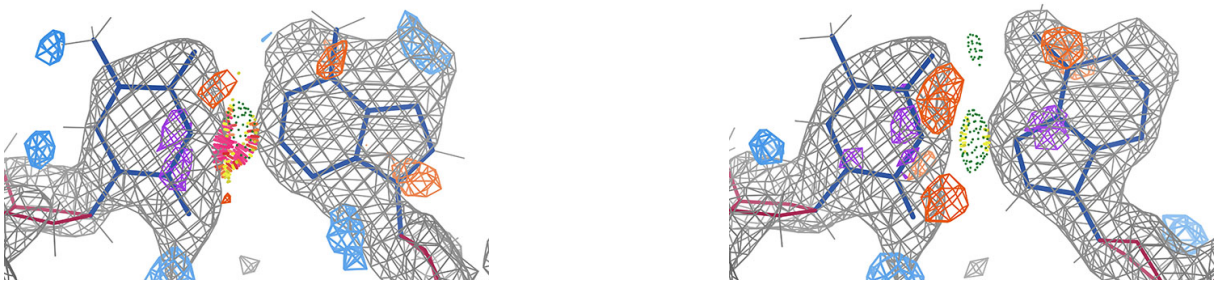
Figure S9: dT (H 1) : dG (G 12). Originally modeled as a Hoogsteen, this terminal base pair is really a canonical Watson-Crick A•T.

2.10 3V6J

(a) Original (b) Flipped

Figure S10: dT (P 12) : dA (B 7). This base pair is in the n-2 position of DNA polymerase Dpo4 adjacent to several modified bases. the n-1 position is correctly modeled as HG. There is a bulky modified base (N^2,3-Ethenoguanine) in the reading frame at the insertion site that may be contributing to the HG conformations seen in positions n-1 and n-2.

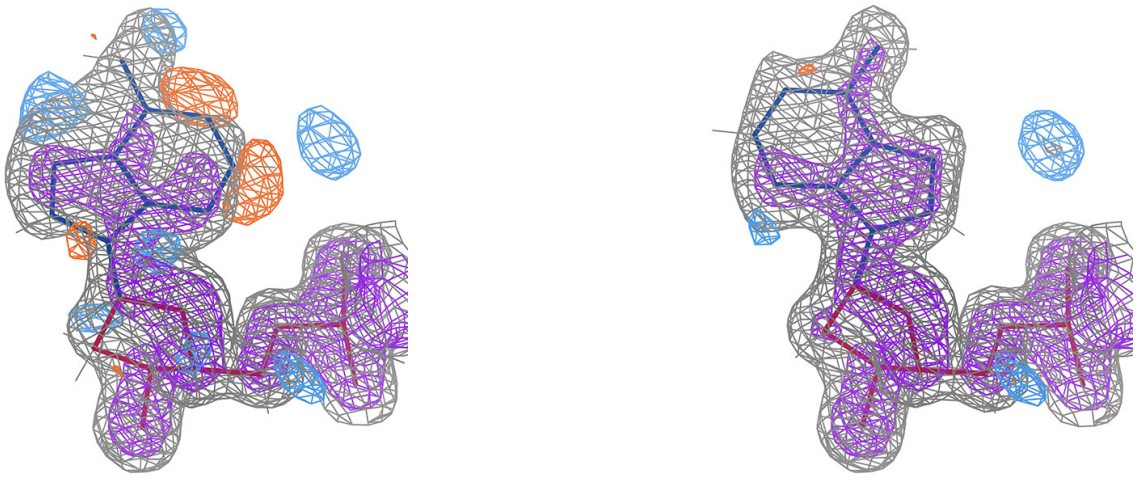
2.11 3V6T



(a) Original (b) Flipped

Figure S11: dT (G -2) : dA (H 2). This base pair is at the blunt end of a DNA duplex bound to dHax3.

2.12 3V9W



(a) Original (b) Flipped

Figure S12: dA (G 7). This base is at the terminal end of a 3 residue single stranded DNA bound to RNase T from *E. coli* and should be in the *anti* conformation.

2.13 4DTN

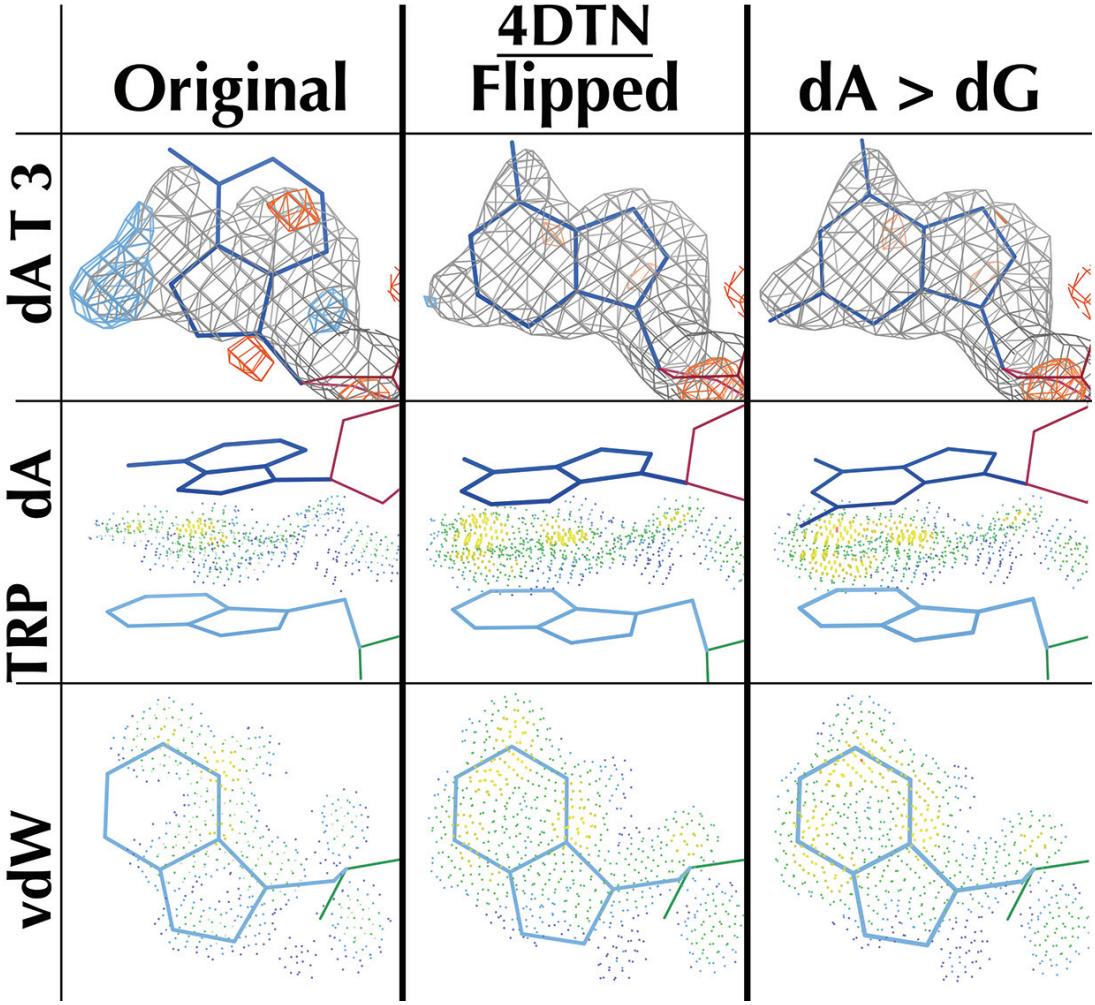
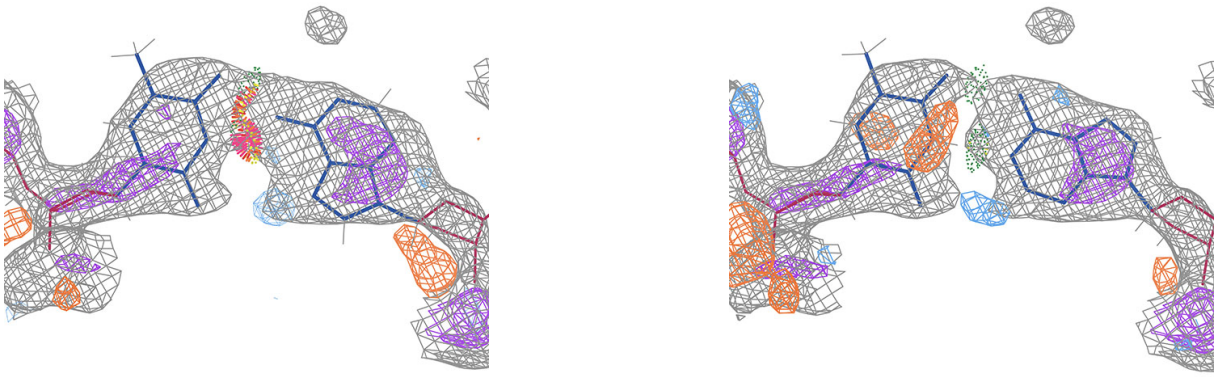


Figure S13: dA (T 3). This base is in the single-stranded region of the reading strand of RB69 DNA Polymerase, i.e. before incorporation and thus lacking a pair. The purine is stacking with a tryptophan of the polymerase. Originally modeled as an adenine and reported to be such in the publication, the purine is very likely a guanine because the high resolution density suggests it and six other structures from the same study models a guanine in the same position (with unambiguous, high resolution density). The reason that the purine prefers the *anti* conformation is likely due to vdW packing with the TRP, which is more extensive in the *anti* conformation relative to *syn*.

2.14 4I2O



(a) Original (b) Flipped

Figure S14: dA (X 5). This purine is part of a base pair in the interior of a helix bound to transcription factor FixK2 from *Bradyrhizobium japonicum*. Originally modeled as HG, this really should be WC. Further visual investigation of the density revealed that this structure contains an unmodeled alternate helix that runs in the opposite direction, like the glucocorticoid receptor dicussed in the Results and in Section S3. The base pair shown here has an alternate pair wherein the dA alternates with a dC and the dT alternates with a dA (not shown.

3 Glucocorticoid Receptor: purine flip NOT appropriate

3.1 3G9P

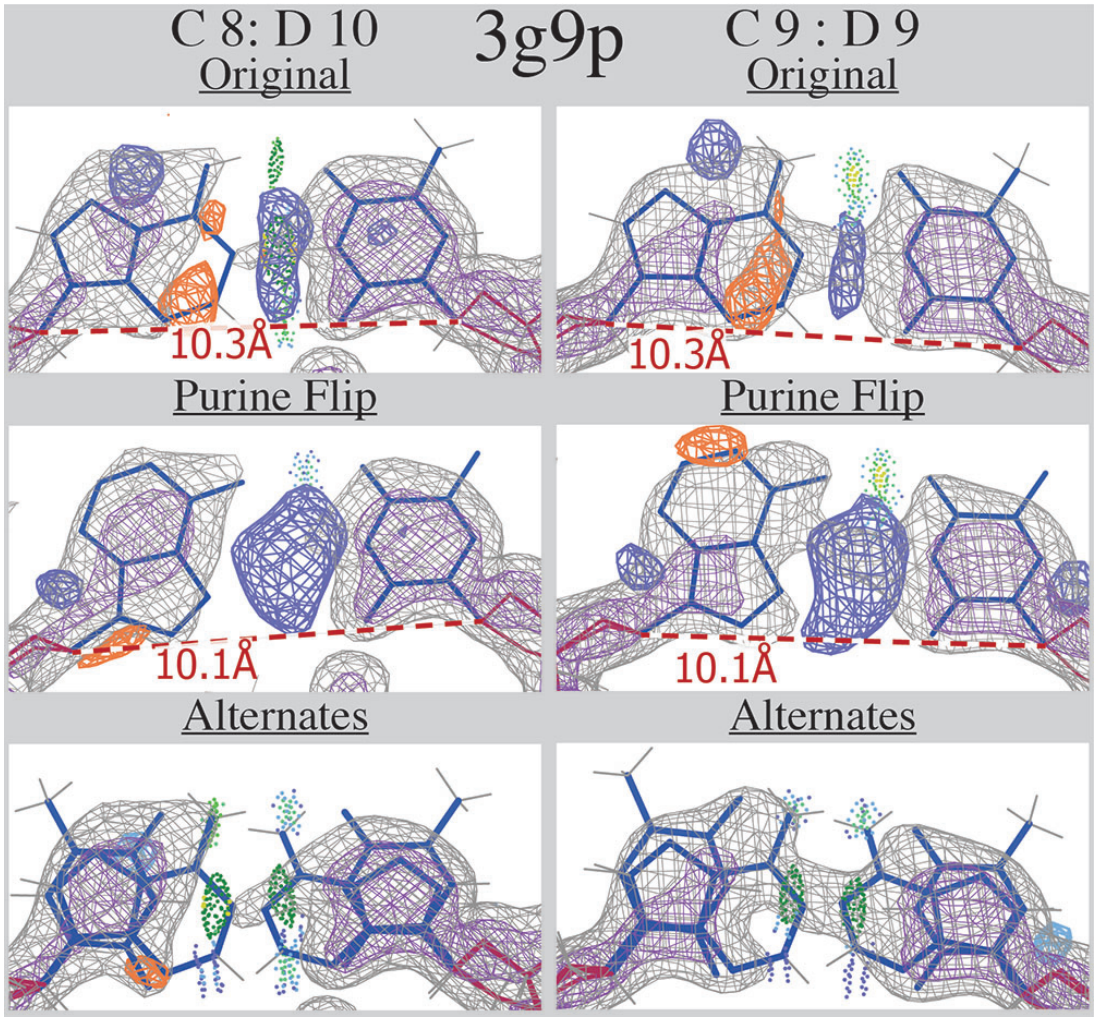


Figure S15: Shown are two of the A•T pairs in 3G9P, their proposed flips to HG, and a the correct alternate model. While the flipped purine did fit the electron density better, the large positive difference peak between the bases suggested that the HG model was not correct. We tried HG/WC alternates. While that did get rid of a good amount of difference density, the HG pairs were still too far apart. The strong density of the ribose sugars and their distance from each other leaves little doubt that a HG base pair cannot be modeled here. Rather it is an unmodeled duplex alternate running in the opposite direction, as shown in the bottom pane and futher detailed in the manuscript.

4 Transcription Factor FixK2

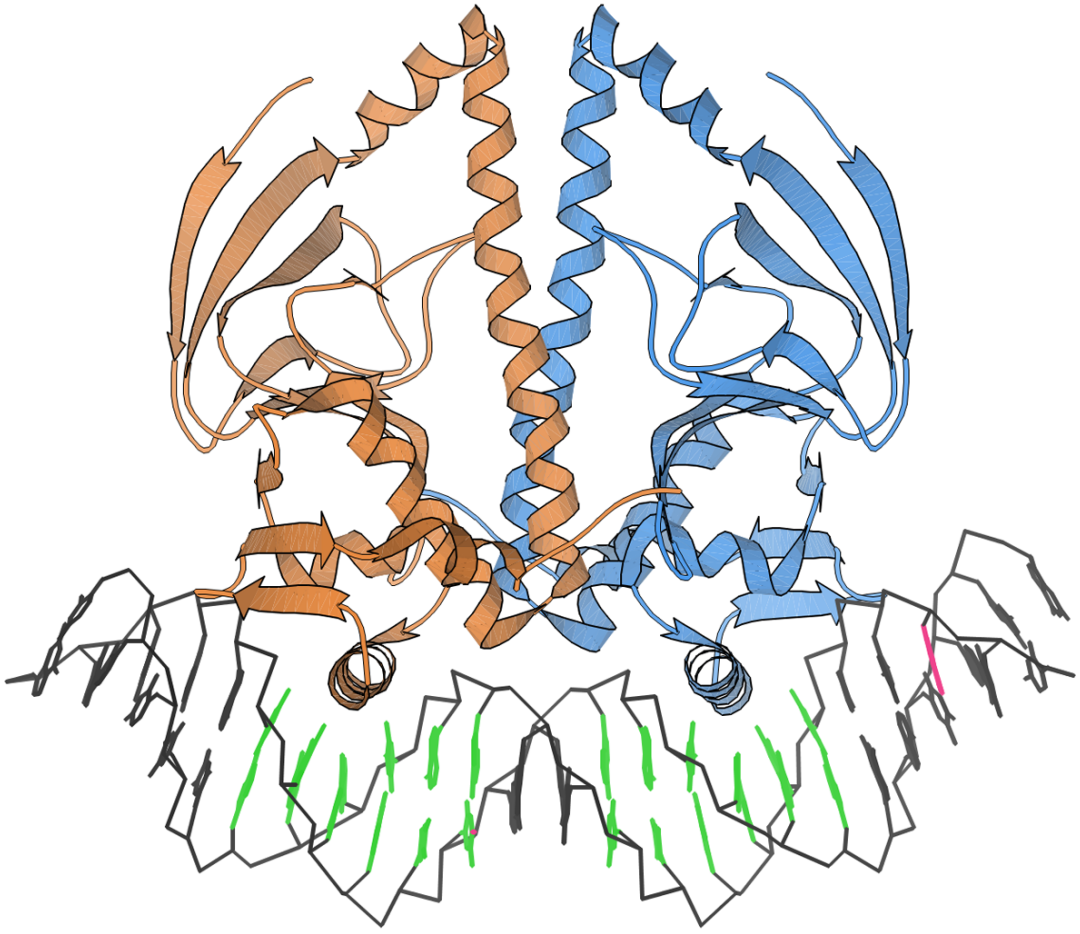


Figure S16: Homodimer ribbons for transcription factor FixK2 bound to DNA (PDB 4i2o). The adenine X 5 base needing a flip to *anti* conformation is highlighted in pink. The 6 base pairs at identical binding sites to each protein monomer are in green, while other bases are non-palindromic.

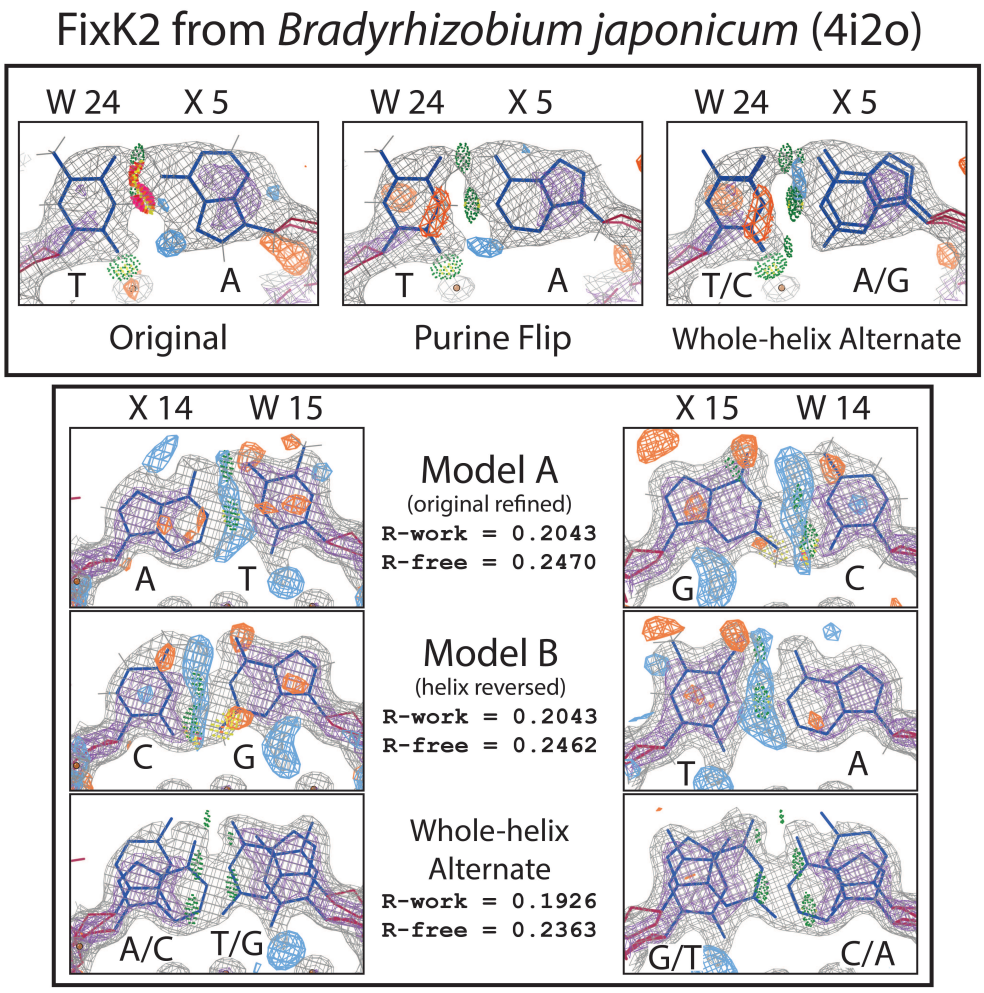


Figure S17: Several different ways to model base pairs from the transcription factor FixK2/DNA complex are shown (PDB 4i2o). Adenine X 5 was flagged by *find_purine_decoys* as a potential purine decoy. As shown in the top panels, the original density and model sterics support a *syn*-to-*anti* flip, and dA X 5 is included in Table 1. However, the flipped density and difference density suggest that a C•G pair might also be present, which would happen if this semi-palindromic DNA bound to the Fix2 homodimer in both directions. Further investigation revealed density anomalies around base pairs X 14:W 15 and X 15:W 14 that strongly suggest purine/pyrimidine alternates. As shown in the bottom panels, a whole-helix alternate model fits both density and sterics significantly better, improving both R and R-free by more than 1%.

References

[1] Sebastiaan H Meijsing, Miles a Pufall, Alex Y So, Darren L Bates, Lin Chen, and Keith R Yamamoto. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science (New York, N.Y.)*, 324(5925):407–410, April 2009.

[2] U Strähle, G Klock, and G Schütz. A dna sequence of 15 base pairs is sufficient to mediate both glucocorticoid and progesterone induction of gene expression. *Proceedings of the National Academy of Sciences*, 84(22):7871–7875, 1987.