Supplementary Figures and Text for Kepple, Patel, Salamon and Segall,

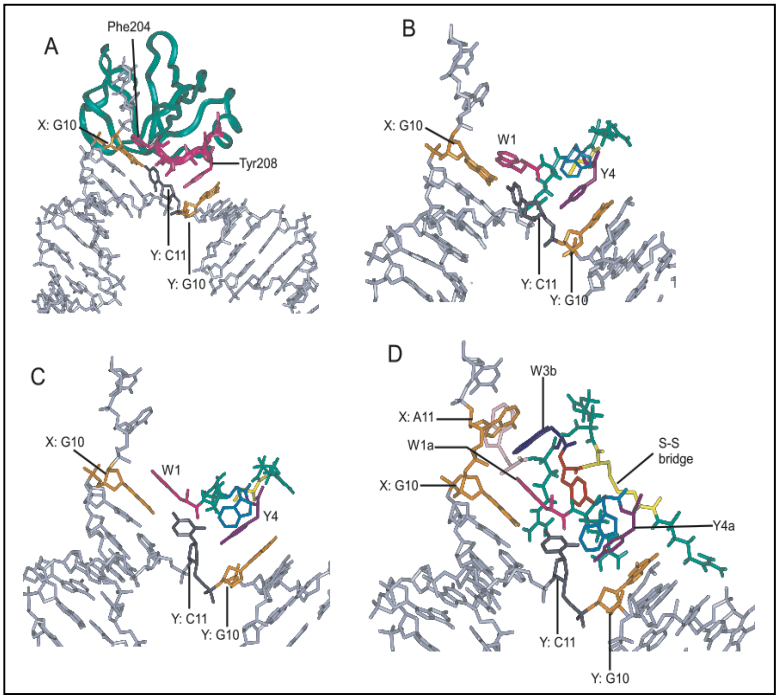
Interactions between branched DNAs and peptide inhibitors of DNA repair

Table of Contents

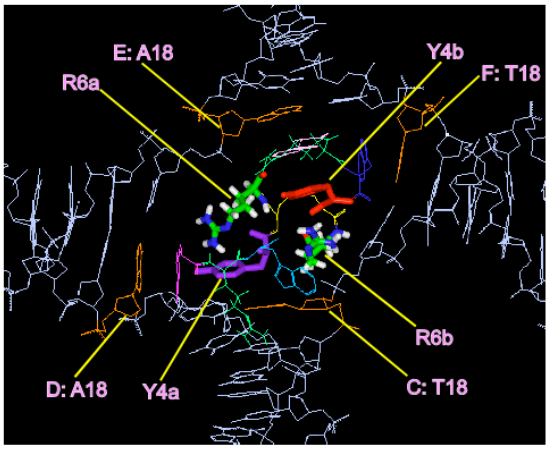
Supplementary Table 1. Sequences of oligonucleotides used in this study. The position of the 2-AP residue, when it appears, is underlined.

Oligo	Description[a]	Sequence (5' to 3')
1		CGCGGTACGGCCGGAAGCCTGCTTTTTTATACAAGC TTAACTCCC
2	Oligos for peptide/RecG binding assays	GGGAGTTAAGCTTGTATAAATGAGGTACTGTTAGGA ATTCTCCAAGTGACTACGG
3		CCGTAGTCACTTGGAGAATTCCTAACAGTACCTCAT TTATACCTGCAGCCGACGC
4		GCGTCGGCTGCAGGTATAAAAAAGCAGGCTTCCGG CCGTACCGCG
5	Strand 1	CGCGGTACGGCCGGAAGCCTGGTTTACAATGCAAGC TTGACTGGCAGCCC
6	Strand 1 AP1	CGCGGTACGGCCGGAAGCCTGGTTT<u>A</u>CAATGCAAGC TTGACTGGCAGCCC
7	Strand 2	GGGCTGCCAGTCAAGCTTGCATTGTTGGAGGTACTG TTAGGAATTCTCCAAGTGACTCGG
8	Strand 3	CCGAGTCACTTGGAGAATTCCTAACAGTACCTCCAT ATACCTCTGCAGCCGACTCGACGC
9	Strand 3 AP1	CCGAGTCACTTGGAGAATTCCTAACAGTACCTCC<u>A</u>T ATACCTCTGCAGCCGACTCGACGC
10	Strand 3 AP2	CCGAGTCACTTGGAGAATTCCTAACAGTACCTCCAT <u>A</u>TACCTCTGCAGCCGACTCGACGC
11	Strand 4	GCGTCGAGTCGGCTGCAGAGGTATAAAACCAGGCTT CCGGCCGTACCGCG
12	Strand 4 AP1	GCGTCGAGTCGGCTGCAGAGGTAT<u>A</u>AAACCAGGCTT CCGGCCGTACCGCG
13	Strand 4 AP2	GCGTCGAGTCGGCTGCAGAGGTATA<u>A</u>AACCAGGCTT CCGGCCGTACCGCG
14	Strand 4 AP3	GCGTCGAGTCGGCTGCAGAGGTATAA<u>A</u>ACCAGGCTT CCGGCCGTACCGCG
15	Strand 4 AP4	GCGTCGAGTCGGCTGCAGAGGTATAAA<u>A</u>CCAGGCTT CCGGCCGTACCGCG
16	Leading strand	GGGCTGCCAGTCAAGCTTGCATTGT
17	Lagging strand	TATACCTCTGCAGCCGACTCGACGC
18	Lagging strand AP2	T<u>A</u>TACCTCTGCAGCCGACTCGACGC

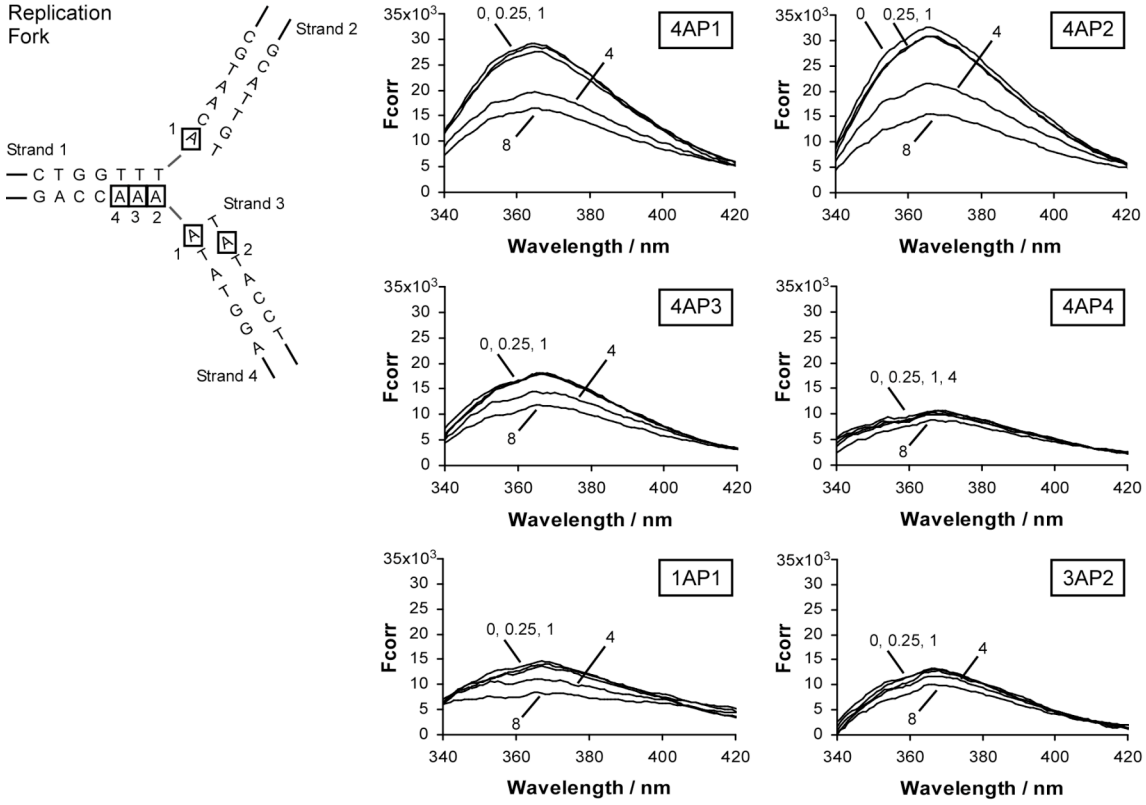
[a]Oligonucleotides 1-4 were used to assemble the HJ substrate for band shift assays. Oligonucleotide 1 was labeled with ^{32}P in these assays. The remaining oligonucleotides were used in 2-AP assays.



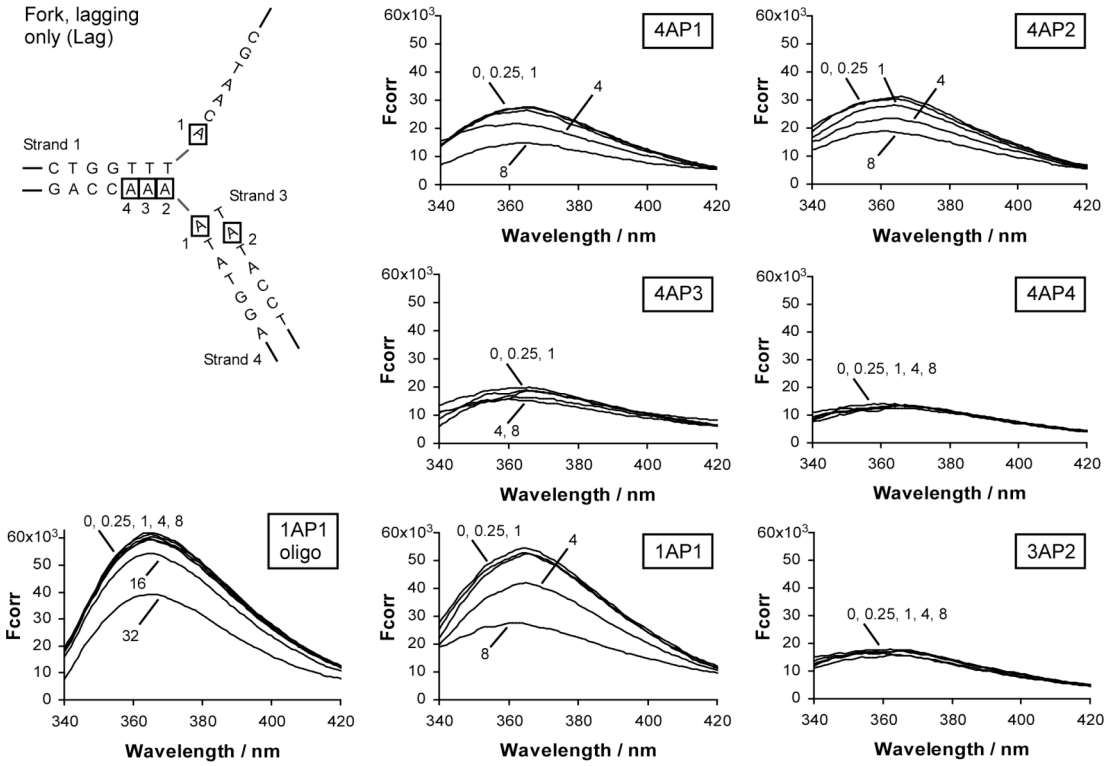
Supplementary Figure 1. Modeling of $(WRWYCR)_2$ on a lagging strand replication fork. **(A)** Co-crystal structure of the 98-amino acid *T. maritima* RecG wedge domain (green) and a lagging strand replication fork (light gray) (Singleton et al., 2001). Amino acids 204-210 (magenta) are located near the branch point of the fork. Phe204 stacks with G10 on the X or leading strand template (X: G10, orange) and Tyr208 stacks with G10 on the Y or lagging strand template (Y: G10, orange). One central base in the lagging strand template (Y: C11, dark gray) is flipped outward as a consequence of RecG binding. **(B)** Six amino acids (position 205-210, wild type NQDYLQ) were mutated to WRWYCR (see Materials and Methods). The first tryptophan (W1) is colored magenta and arginine residues (R2 and R6) are in green. The tryptophan at the third position (W3) is turquoise, tyrosine (Y4) is purple, and cysteine (C5) is yellow. The remaining wedge domain structure is not shown. A stacking interaction between Y4 and the Y: G10 base is observed but W1 does not contact the X: G10 base. **(C)** Interactions between the WRWYCR monomer and replication fork after manual rotation of W1 and energy minimization. **(D)** Model of $(WRWYCR)_2$ bound to a replication fork. The letters "a" and "b" are used to differentiate each peptide monomer. Arginine residues (R2a, R6a, R2b, R6b) are again colored green and the cysteine residues are in yellow. The four tryptophan residues are W1a (magenta), W1b (pink), W3a (turquoise), W3b (blue) and the two tyrosine residues are Y4a (purple) and Y4b (red). The WRWYCR monomer in part C was dimerized through a disulfide bridge and W3b was manually rotated to stack with the X: A11 base.



Supplementary Figure 2. π – cationic intra-peptide interaction between R6A and Y4a (purple).



Supplementary Figure 3. Peptide $(WRWYCR)_2$ quenches 2-AP fluorescence at the branch point of a replication fork. A diagram of the central region of the fork substrate is shown. The sequence of each arm is identical to the HJ (Fig. 3), except that the right arm of the junction is missing. The aminopurines are again named first by strand number and then by sequential location from the 5' end of the oligonucleotide. The exception is 3AP2 in which the name is unchanged for consistency. Fluorescence emission spectra were recorded with 2-AP at different positions of the replication fork substrate using an excitation wavelength of 315 nm. Reactions (125 µl) contained 0.2 µM fork substrate, 20 mM Tris-HCl (pH 7.8), 100 mM NaCl, 1 mM EDTA, 0.032% DMSO, and the indicated amount of $(WRWYCR)_2$. The numbers pointing to each trace again represent the ratio of peptide dimer concentration to substrate concentration. $(WRWYCR)_2$ concentrations used in each reaction were 0.05 µM (0.25:1 peptide to substrate ratio), 0.2 µM (1:1 peptide to substrate ratio), 0.8 µM (4:1 peptide to substrate ratio), and 1.6 µM (8:1 peptide to substrate ratio). F_{corr} is the corrected fluorescence intensity, in arbitrary units.



Supplementary Figure 4. 2-AP fluorescence is quenched at the branch point of a replication fork with only the lagging strand. A diagram of the central region of the substrate is shown. The substrate is identical to that of the replication fork (Fig. 3 above), except that the "leading strand" oligonucleotide is missing. The aminopurines are named by strand number and by sequential location from the 5' end of the oligonucleotide. The name of 3AP2 is again unchanged for consistency. Fluorescence emission spectra of junctions with 2-AP at different positions of the lagging strand-only fork substrate were recorded at an excitation wavelength of 315 nm. Reactions (125 µl) contained 0.2 µM substrate, 20 mM Tris-HCl (pH 7.8), 100 mM NaCl, 1 mM EDTA, 0.032% DMSO, and the indicated amount of $(WRWYCR)_2$. Peptide concentrations used in each reaction were 0.05 µM (0.25:1 peptide to substrate ratio), 0.2 µM (1:1 peptide to substrate ratio), 0.8 µM (4:1 peptide to substrate ratio), and 1.6 µM (8:1 peptide to substrate ratio). Fluorescence emission spectra of the 1AP1 oligonucleotide alone (bottom left graph) were recorded during titration of $(WRWYCR)_2$ in a separate experiment. Reactions (2 ml) contained 0.2 µM 1AP1 oligonucleotide, 20 mM Tris-HCl (pH 7.8), 100 mM NaCl, 1 mM EDTA, 1.6% DMSO, and the indicated amount of peptide. A >4-fold higher concentration of $(WRWYCR)_2$ (more than 32-fold excess) is required to significantly quench 2-AP fluorescence of the single oligonucleotide. F_{corr} is the corrected fluorescence intensity, in arbitrary units.

Error Analysis

Background

The standard model underlying any curve fitting effort is that the value of one variable determines the value of a second variable up to a random term often associated with measurement errors (1). The standard deviation of the error term is referred to as the *standard error* of the prediction and plotted as error bars in a graphical representation of the fitted model.

Standard Estimates of Error

In our manuscript, the first variable is the peptide concentration, L_t, which is fitted to the second variable, the fraction complexed, F_c, through the formula in the section entitled *Model Fitting of Fluorescence Data*. For convenience it is repeated here with an explicit error term

$$F_c = \frac{L_t + R_t + K_d - \sqrt{\left(L_t + R_t + K_d\right)^2 - 4 L_t R_t}}{2 R_t} + \text{Error}. \qquad \text{(S1)}$$

Substituting the fitted value of the estimated parameter K_d, and the concentration of the substrate, R_t, used in the preparation, results in a value of the Error for each of our measurements of L_t, and F_c. In this context, the Error is usually called the residual.

Since only two or three measurements were taken at each peptide concentration for a given substrate, direct estimation of the standard estimates of error for each substrate is not practical. From a direct examination of the residual values, however, it is clear that these residuals depend rather strongly on the peptide concentration. This is to be expected; at a peptide concentration L_t=0, all of the predictions for F_c must be zero. Similarly, at sufficiently high peptide concentrations, F_c must reach one. Thus at these extremes, the standard error is extremely small or zero. At intermediate peptide concentrations we see sizeable residuals. Our approach was to pool the data into two groups: extremes (F_c<0.2 or F_c >0.9) and intermediates (0.2< F_c <0.9). The pooled residuals fit a normal distribution well and the standard deviations of these pools were used for the error bars shown in Figure 6 of the main text (2).

Confidence Intervals for K_d

The normal model of the error term described in the previous paragraph was used to find confidence intervals for the K_d. With the normal approximation, we are able to simulate many runs of the experiment according to the above equation for F_c by choosing a normally distributed random number corresponding to each measurement. We can then fit K_d values to the simulated data. Doing this many times is called a *parametric bootstrap* (3) and can reveal confidence intervals for K_d. We ran 1000 simulations for the experiment and discarded the highest and lowest 50 values of K_d for the 90% confidence intervals shown in Supplementary Figure 5. The standard deviation of the K_d values found was used for the ± terms in Table 4 of the main text.

A More Complicated Model

As mentioned in the text, we also explored an alternative hypothesis allowing the number of binding sites for the peptide elsewhere on the substrate. Such additional sites can offer alternate modes of decay for the excited fluorophore and thus contribute to the quenching. We thus fitted the data to (44)

$$F_c = \frac{L_b}{n R_t} = \frac{L_t + n R_t + K_d - \sqrt{\left(L_t + n R_t + K_d\right)^2 - 4 L_t n R_t}}{2 n R_t} \quad \text{(S2)}$$

where the symbols are as defined in the main manuscript with the addition of the new quantity n which represents the number of receptor sites per substrate molecule. Note that this formula reduces to the one in the main manuscript when n=1. The resulting fits are shown in Supplementary Figure 6.

A comparison of the models is shown in Supplementary Table 2. The two models are hierarchical, i.e., the simpler model can be obtained as a specific case of the more complicated model by setting n to one. This allows a simple test (1) for the comparison of the two models: an adjusted sum squared error which divides the sum squared error by the number of data points minus twice the number of parameters. For all 6 substrates, the more complicated model gave a smaller adjusted sum squared error. Thus the test says that adding the variable number of sites is warranted and gives a better model. The model with n=1 is simpler and more standard. Thus we have stuck to this simpler model in the main manuscript. The qualitative conclusions are unchanged: the substrates can be divided into three groups according to strength of binding with the strongest binding for HJ, medium strength binding for Lead, Lag, and Fork, and weak binding for Flayed and Oligo.

As indicated in Supplementary Table 2, the affinity of the peptide dimer for the substrates that mimic replication forks (complete, lagging or leading) is less than for the HJ, but the fold reduction in affinity depends on how many peptide dimers are assumed to bind to each substrate, and the two models (single binding site vs. multiple binding sites) are roughly equal in the goodness of fit. For the simpler model (n = 1), the peptide affinity for the replication fork substrates ranges from 4.4X to 9.2X lower than for the HJ substrates, and 44X and 51X lower for the flayed fork and oligo substrates, respectively. However, in the case of the flayed and oligo substrates, the model permitting n to increase to 18 peptide dimers per substrate gives a much better fit, and lower K_ds: 14 nM for the oligo substrate and 18 nM for the flayed substrate. Note that we have seen no binding of peptide to oligo substrates using band shift assays below peptide concentrations at least 400X in excess of the DNA, suggesting that low affinity of binding of the peptide to these substrates (M. Rideout and A. Segall, unpublished results). These substrates share with each other the most single-stranded character of all the substrates tested, and the high number of peptide dimers predicted by the second model suggests that the binding is in fact non-specific, coating and potentially bridging two single strands. Experiments directly measuring peptide stoichiometry will be necessary to distinguish between high K_d values and high numbers of binding sites.

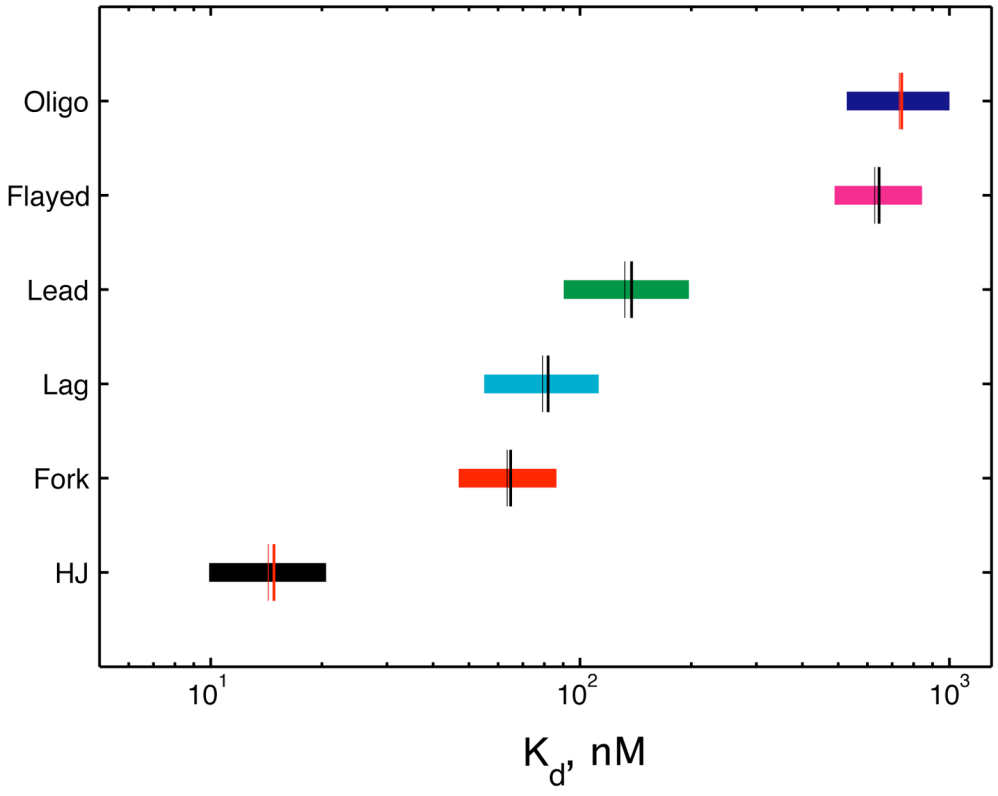
References

1. Draper, N.R. and Smith, H. (1998) Applied Regression Analysis Wiley Series in Probability and Statistics.

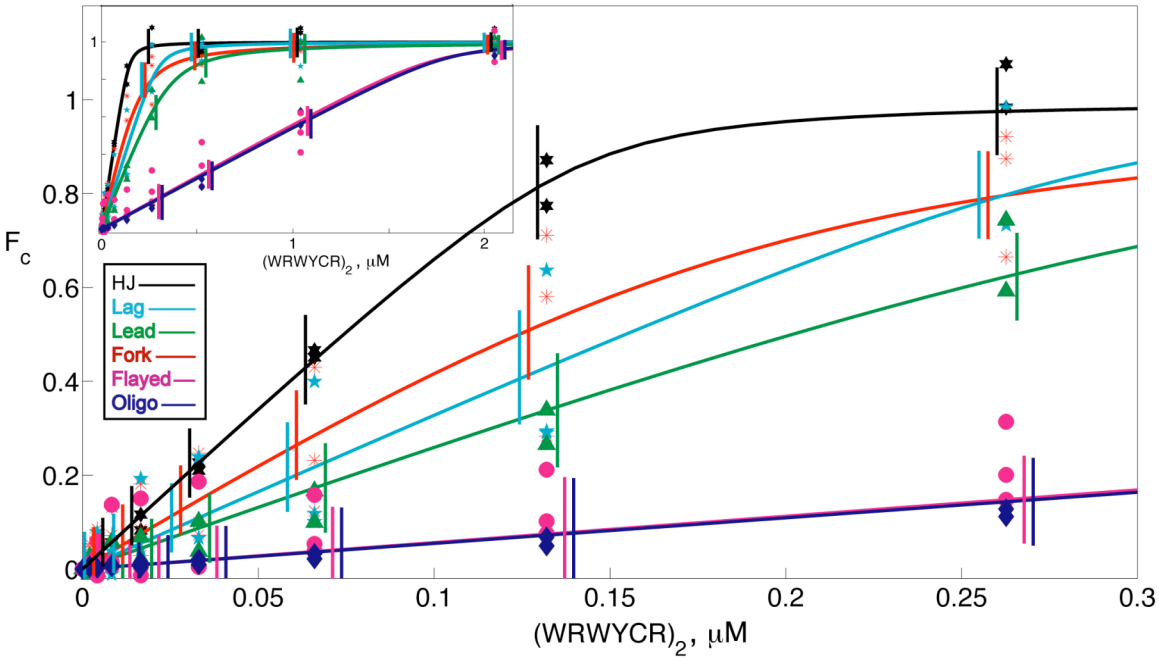
2. Chambers, John; William Cleveland, Beat Kleiner, and Paul Tukey (1983). Graphical Methods for Data Analysis. Wadsworth.

3. Efron, B. and Tibshirani, R. (1993) An Introduction to the Bootstrap, Chapman & Hall / CRC Press.

Supplementary Figure 5. Ninety percent confidence for the calculated values of Kd obtained from a parametric bootstrap. The thin vertical lines indicate our fitted values shown in Table 3 of the main manuscript. The thicker vertical lines indicate the mean values of the data obtained from the bootstrap.



Supplementary Figure 6. The fraction of substrate complexed plotted as a function of $(WRWYCR)_2$ concentration. The inset graph shows the binding curves over almost the entire range of concentrations tested. The curves are obtained using a two-parameter fit to the equation S2. Comparing to Figure 6 of the main text that fit only K_d values to the data, the curves here also adjusted n, the number of binding sites on the substrate. Note that in distinction to the parametric form in the simpler model, these curves can (and do) cross.



Supplementary Table 2. Comparison of calculated K_d values (and n peptide dimer binding sites) for DNA substrates.

Substrate	**n = 1**			**n ≥ 1**			
	K_d (nM)	**R^2**	**Fold reduced vs. HJ**	**K_d (nM) (varied n)**	**n sites for peptide dimer**	**R^2**	**Fold reduced vs. HJ**
Holliday junction	14.3	0.992	1	3.1	1.4	0.997	1
Complete fork	63.5	0.958	4.4	28	1.9	0.963	1.4
Lagging strand fork	79.2	0.942	5.5	7.0	3.0	0.961	2.1
Leading strand fork	132	0.963	9.2	27	3.5	0.981	2.5
Flayed fork	628	0.913	43.9	18	18	0.953	13
Oligo	731	0.923	51.1	14	18	0.996	13