Supporting Information for

Unlocking Diverse Asymmetric Three-Component Reactions via Crowding Effects within DNA Scaffolds

Yang Zheng[1], Jingyi Liu[2], Rao Dai[1], Zheng-Yang Gu[1], Lin-Wei Wang[1], Xinglong Zhang[2]*, Ru-Yi Zhu[1]*

[1]*Department of Chemistry, National University of Singapore, Singapore 117544, Singapore.*

[2]*Department of Chemistry, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China.*

Email: xinglong.zhang@cuhk.edu.hk; rzhu@nus.edu.sg

Table of Contents

1. General information 3
   Substrates used in this study 3
   Synthesis and Chemicals 4
   Analytical Methods 4
2. List of oligonucleotides and MALDI-TOF analysis 5
   List of oligonucleotides used in this study 5
   MALDI-TOF Analysis 6
3. Optimization of reaction conditions 8
   Table S1. Buffer and pH evaluation 8
   Table S2. Buffer and pH evaluation 8
   Table S3. Evaluation of opposite bases 9
   Table S4. Evaluation of substrate ratios 9
   Table S5. Evaluation of copper sources 10
   Table S6. Evaluation of neighboring sequences 10
   Table S7. Evaluation of neighboring sequences for fluoroamination 11
   Table S8. Buffer pH evaluation 12
4. Preparation of materials 13
   General procedure for the preparation of DNA AP site 13

General procedure for the preparation of DNA hybrid catalysts 13
5. Reaction procedures 14
Racemic version 14
Asymmetric version 16
6. Control experiments (corresponding to Figure 2c, main text) 18
Without DNA, using the [Cu]/dmbpy as the catalyst 18
Using [Cu]/dmbpy & AP-DNA-U-GG as the catalyst 20
Using DIAD instead of CbzN=NCbz as the electrophile 21
Using DEAD instead of CbzN=NCbz as the electrophile 24
Using (*rac*)-**3a'** as the substrate instead of **1a** and **2a** 27
Using (*R*)-**3a** as the substrate instead of **1a** and **2a** 28
7. Absolute configuration determination 30
8. Large scale 38
Large scale for the three-component reaction with azodicarboxylate and indole 38
Large scale for the three-component reaction with Selectfluor and H_2O. 40
9. Mechanistic studies 42
Substituent effect of azodicarboxylate 42
Hammett plot 46
UV-Vis titration 54
^{1}H NMR studies on the interaction between DNA and azo-compounds 56
10. Computational section 58
11. Characterization data 69
12. References 216
13. Copies of NMR spectra 219
14. Copies of high-resolution MS (HRMS) spectra 375

Computational section

Computational methods

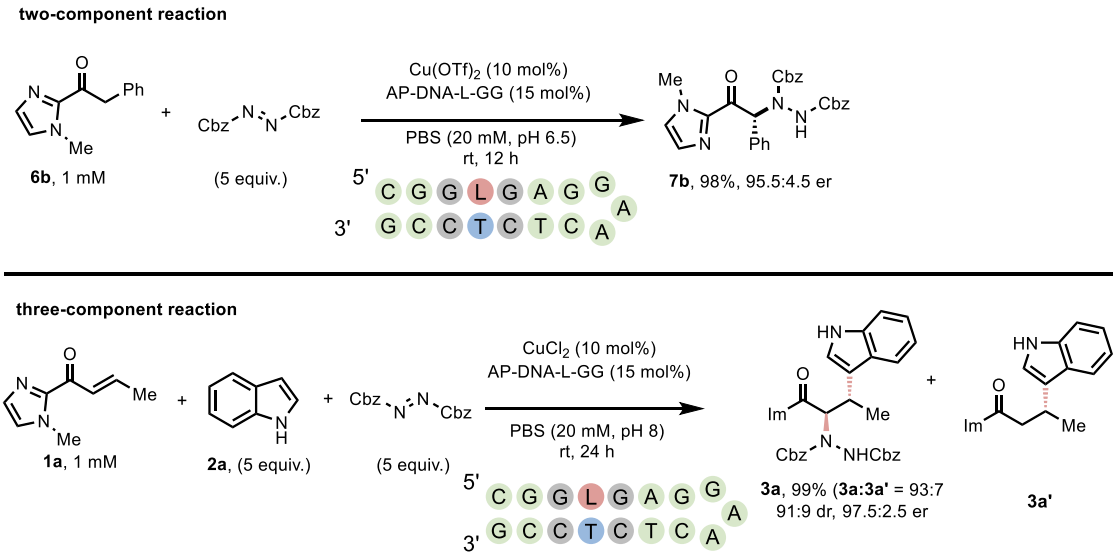
Geometries of DNA-catalyst systems were optimized at GFN2-xTB[10-12] level of theory with *crude* optimization settings in analytical linearized Poisson-Boltzmann (ALPB) implicit solvation model for water solvent. We note that geometry optimizations did not always converge for other optimization settings of *vtight*, *tight, normal, lax, loose* and *sloppy*.

The xTB-optimized structures were subject to single-point energy calculations using density function theory (DFT), which were performed using the *Gaussian 16* rev. B.01 program.[13] As with our previous work,[4b] we attempted single point (SP) correction calculations using BP86 functional[14,15] with Grimme's D3 dispersion correction[16] with Becke-Johnson damping[17-20] (hereafter denoted BP86-D3BJ) with 6-311+G(d) basis set[21] for all atoms. However, one of the SP calculations did not converge properly. We then switched to using the global-hybrid meta-NGA (nonseparable gradient approximation) MN15 functional[22] and the def2-SVP[23,24] basis set for all atoms for a balance of computational cost and accuracy, as our DNA systems are very large in size. We have also adopted the MN15 functional for its robust performance for transition metal-containing systems[25-32] and the more modern Karlsruhe def2 basis set for a more balanced description across all elements in the periodic table, where it uses all-electron basis sets for lighter elements (H to Kr), i.e., all atoms in our system.[23,24]

The implicit SMD continuum solvation model[33] was used to account for the solvent effect of water that was used in the experimental reactions. The single point energies calculated at SMD(water)-MN15/def2-SVP level of theory are given in kcal mol^{-1} and used for discussion throughout. All molecular structures are visualized using *PyMOL* software.[34] All DFT calculations were submitted and PyMOL .pse files generated using our CHEMSMART[35] codes.

Model reactions

The model reactions used for the computational study of two- and three-component reactions, with the (truncated) optimized DNA sequences, are shown as below.



Naming conventions

The starting structures for the models are taken from our previous work.[4b] We follow the naming conventions used in that work, namely that the structures are named according to **Figure S4** below. Structures

where the imidazole group and the 4-methylpyridine are in *cis* relationship will be either **A** or **B**, depending on which DNA groove it faces: if the bpy ligand N atoms face major groove, this will be **A** and if minor groove, **B**. Similarly, structures where the imidazole group and the 4-methylpyridine are in *trans* relationship will be either **C** or **D**, where in **C**, the bpy ligand N atoms face major groove and in **D**, the bpy ligand N atoms face minor groove.

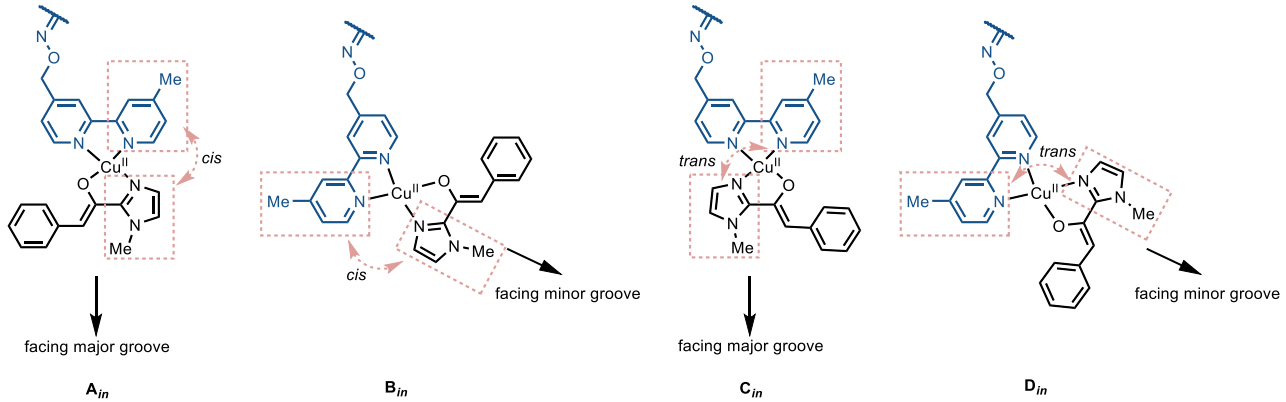
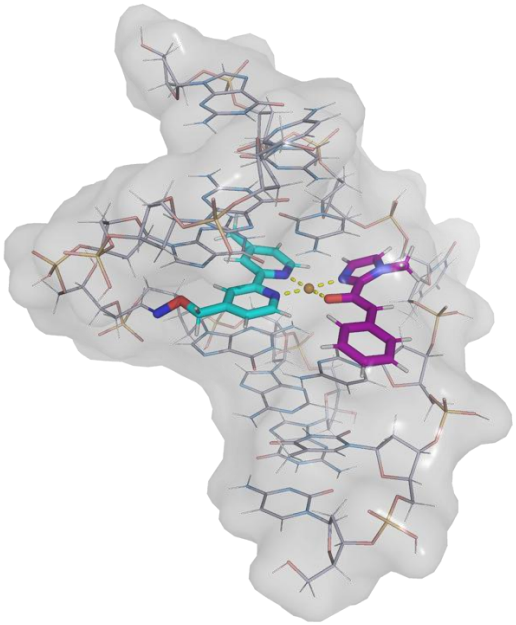
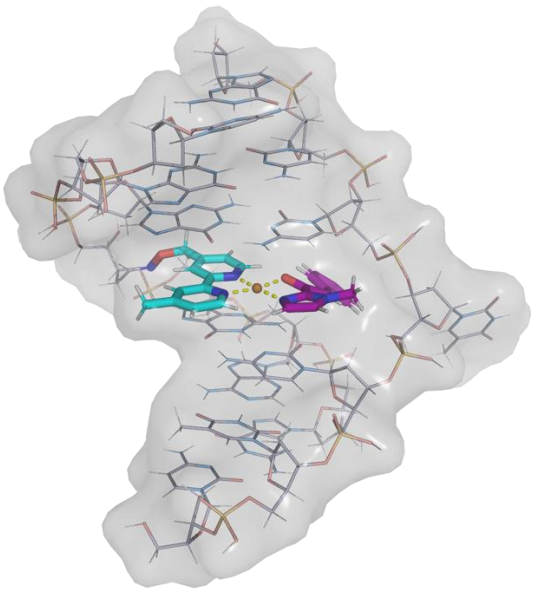


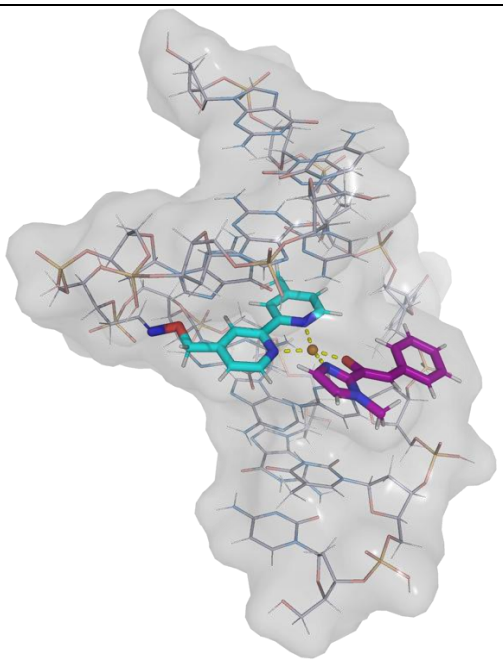
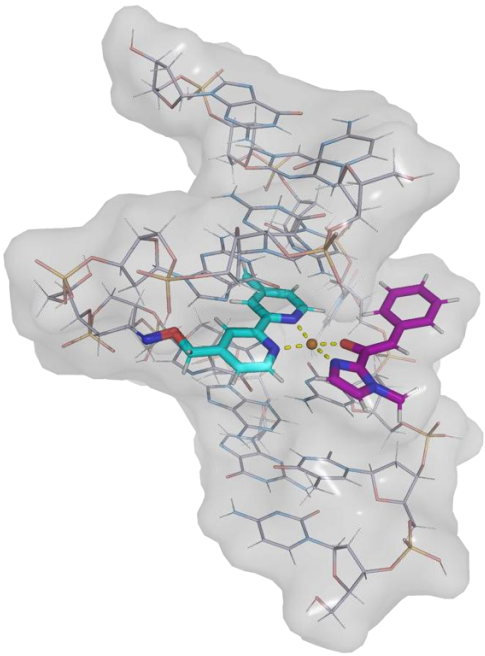
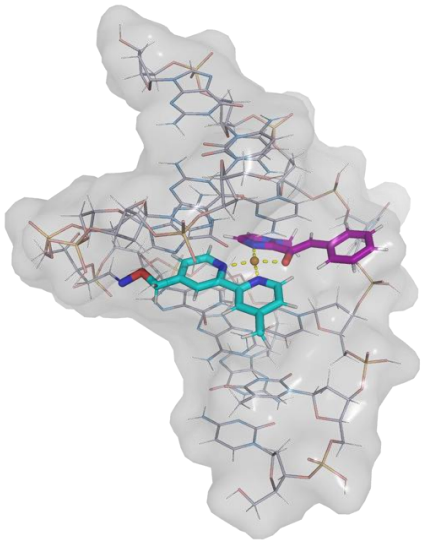
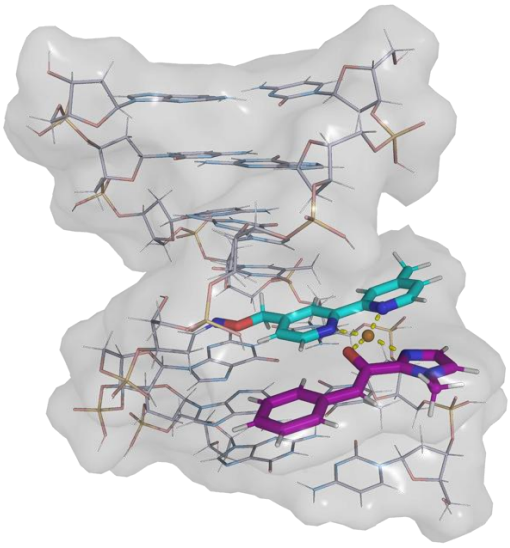
Figure S4. Naming conventions used in computational modelling.

Binding energies

Unbound states

The GFN2-xTB-optimized structures of the DNA model system used for the amination reaction are shown in **Figure S5**. The energies of the structures with the catalyst system in the "out states", where the opposite base remain inside the duplex, and in the "in states", where the catalyst system tucks in between DNA base pairs and the opposite base is flipped out, were compared. It was found that the "in states" are lower in energy than the "out states", thus more thermodynamically stable. This may be due to the more favorable π-π stacking interactions in the "in states", in agreement with what we found in our earlier study.[4b] In particular, structure **A**$_{in}$ is the most stable and is taken as the zero-energy reference.

A$_{in}$	**B**$_{in}$
ΔΔE = 0.0 kcal mol^{-1}	ΔΔE = 7.7 kcal mol^{-1}
	
C$_{in}$	**C**$_{in}$-c2

ΔΔE = 12.6 kcal mol^{-1}	ΔΔE = 13.5 kcal mol^{-1}
	
D$_{in}$	**A**$_{out}$
ΔΔE = 11.5 kcal mol^{-1}	ΔΔE = 20.1 kcal mol^{-1}
	
C$_{out}$	
ΔΔE = 14.2 kcal mol^{-1}	

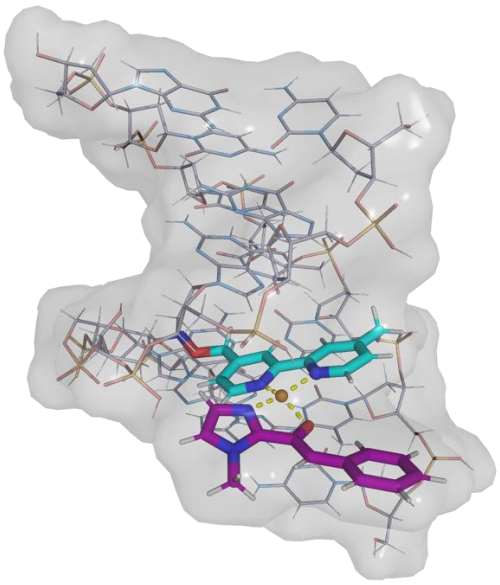


Figure S5. GFN2-xTB-optimized structures of the DNA model system used for the amination reaction. Energy values are given in kcal mol^{-1} and taken relative to the most stable complex, $\mathbf{A}_{in}$.

Bound states

Two-component reactions

The GFN2-xTB-optimized structures of the complexes where the dibenzyl (*E*)-diazene-1,2-dicarboxylate binds are shown in **Figure S6**. We have explored different binding states and herein shown are the most stable one for each case. For conformers in each case with higher energies, we deposited the GFN2-xTB-optimized structures in .xyz format in the zenodo.org weblink under open access. These additional higher energy conformers are named with added suffix c2, c3, …, to denote increasing energy.

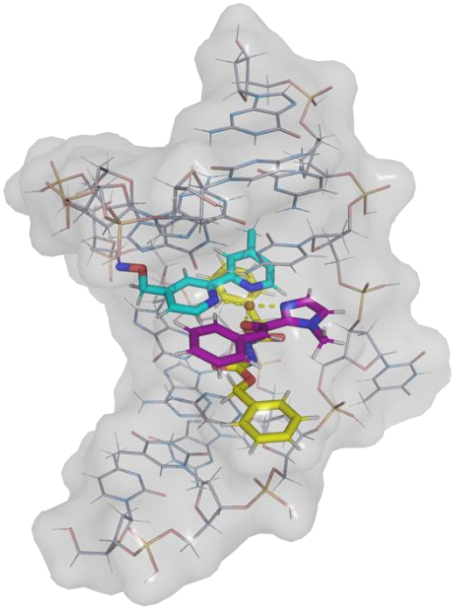
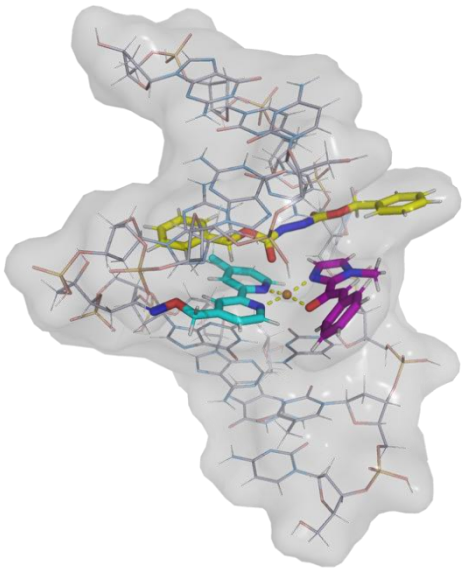
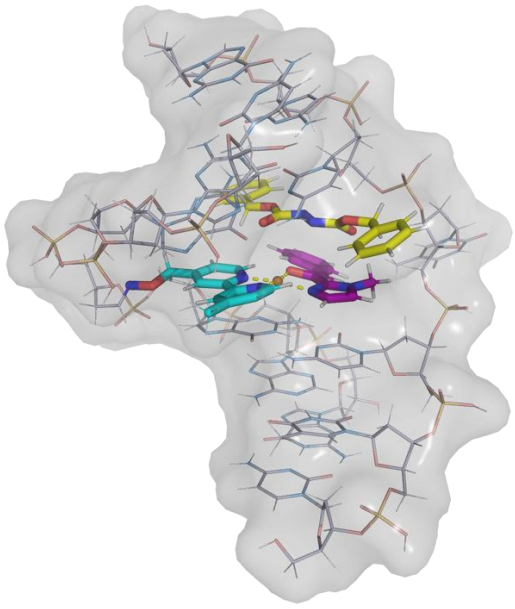
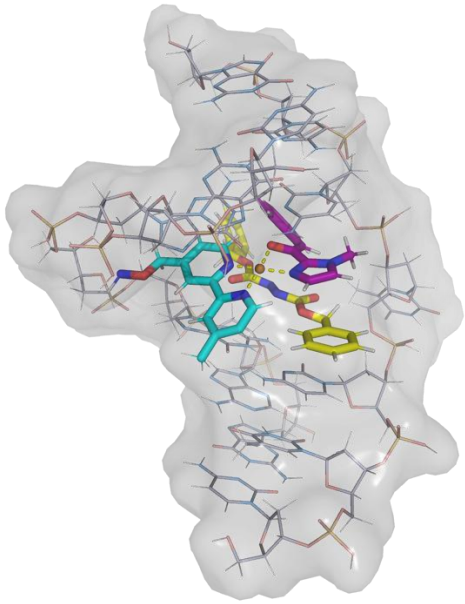
The binding energy is calculated as:

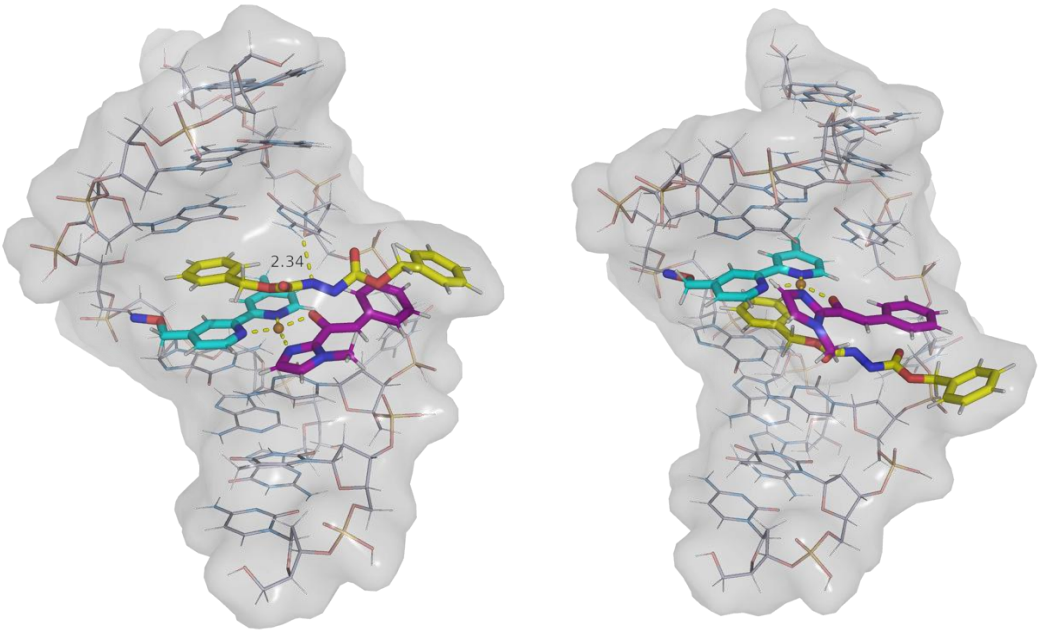
$$\Delta E = E(\text{bound complex}) - E(\text{unbound DNA}) - E(\text{diazene dicarboxylate})$$

where *E(unbound DNA)* is the energy of the most stable unbound DNA state, i.e., common reference energy of species $\mathbf{A}_{in}$ is used for all ΔE calculations.

We observed that out of all possibilities, $\mathbf{A}_{in}\boldsymbol{R}$ has the most negative binding energy (ΔE = -4.9 kcal mol^{-1}), suggesting that the formation of this complex may be the most thermodynamically favorable. For the opposite stereoselectivity, $\mathbf{B}_{in}\boldsymbol{S}$ has the most favorable binding energy, at ΔE = 3.7 kcal mol^{-1}. Due to the various approximations used (model truncation and unfeasible full DFT geometry optimization and frequency calculations of the structure), we are unable to predict the experimentally observed enantiomeric excess (ee) value exactly. Nevertheless, we note that of all the cases, **A**, **B**, **C**, **D**, the bound state leading to the *R* configuration is always favored over the *S* configuration.

We also observe the potential formation of hydrogen bonding, e.g., in the case of $\mathbf{C}_{in}\boldsymbol{R}$ where the diazene dicarboxylate has N atom forming H-bonding with the amino-group of the adjacent base (bond length of 2.34 Å, Figure S4), such interactions may assist in the stabilization of the bound complex; however, in other cases, we note that the formation of H-bond, while the diazene dicarboxylate substrate is intercalated, may distort the adjacent base-pair and make the energy less favorable than without H-bonding.

$\mathbf{A}_{in}\boldsymbol{R}$	$\mathbf{A}_{in}\boldsymbol{S}$
$\Delta E = -4.9$ kcal mol^{-1}	$\Delta E = 15.0$ kcal mol^{-1}
	
$\mathbf{B}_{in}\boldsymbol{R}$	$\mathbf{B}_{in}\boldsymbol{S}$
$\Delta E = 2.5$ kcal mol^{-1}	$\Delta E = 3.7$ kcal mol^{-1}
	
$\mathbf{C}_{in}\boldsymbol{R}$	$\mathbf{C}_{in}\boldsymbol{S}$
$\Delta E = -1.7$ kcal mol^{-1}	$\Delta E = 9.8$ kcal mol^{-1}



$\mathbf{D}_{in}\boldsymbol{R}$	$\mathbf{D}_{in}\boldsymbol{S}$
$\Delta E = 6.7$ kcal mol^{-1}	$\Delta E = 15.2$ kcal mol^{-1}

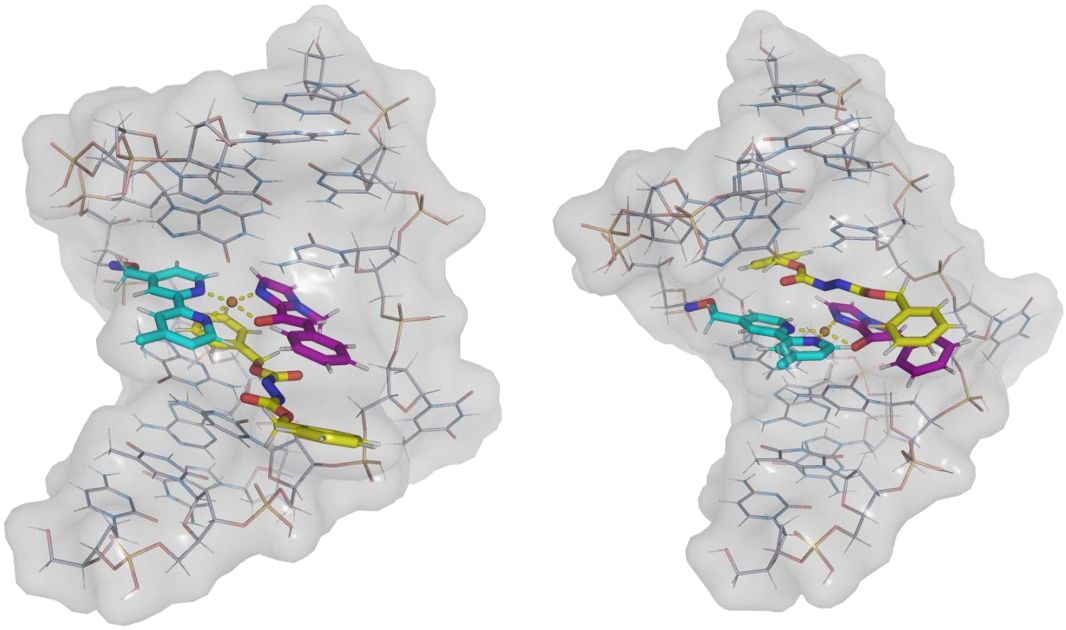
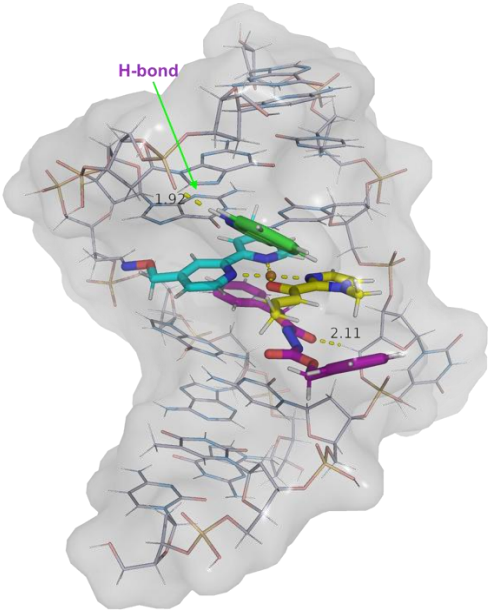
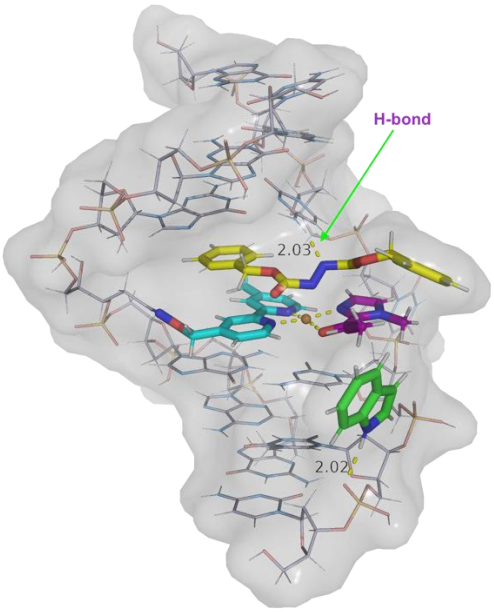
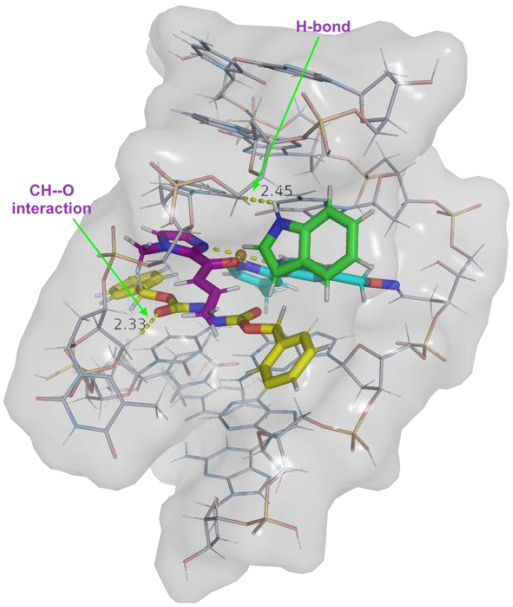
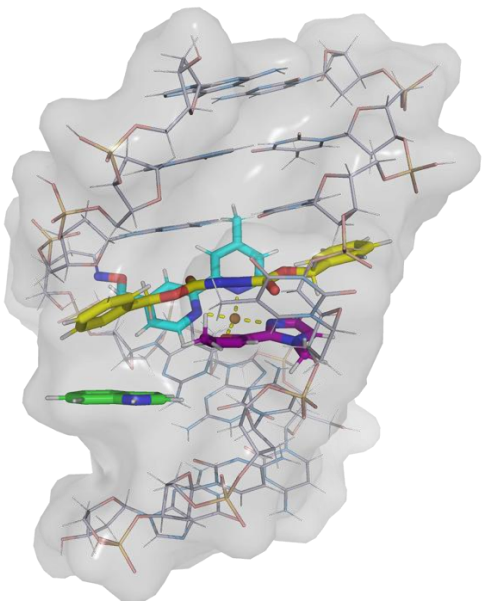


Figure S6. GFN2-xTB-optimized complexes with bound dibenzyl (*E*)-diazene-1,2-dicarboxylate. Energy values are given in kcal mol^{-1}.

Three-component reactions

We adopt the configurations in the two-component reaction shown above, where the azo substrate is intercalated between the base pairs; we replace the enolate species by the α,β-unsaturated ketone substrate and add in the indole substrate. We deprotonate two most proximal phosphate backbones (one on each DNA strand) to maintain overall charge neutrality. The GFN2-xTB-optimized structures of the complexes where the dibenzyl (*E*)-diazene-1,2-dicarboxylate binds are shown in **Figure S7** and the energies relative to the most stable complex are given. The structures are named with suffix "***3c***" to denote that these result from the three-component reactions.

A$_{in}$*R*_*3c*	**A$_{in}$*S*_*3c***
ΔΔE = 16.0 kcal mol^{-1}	ΔΔE = 17.7 kcal mol^{-1}
	
B$_{in}$*R*_*3c*	**B$_{in}$*S*_*3c***
ΔΔE = 0.0 kcal mol^{-1}	ΔΔE = 11.0 kcal mol^{-1}
	

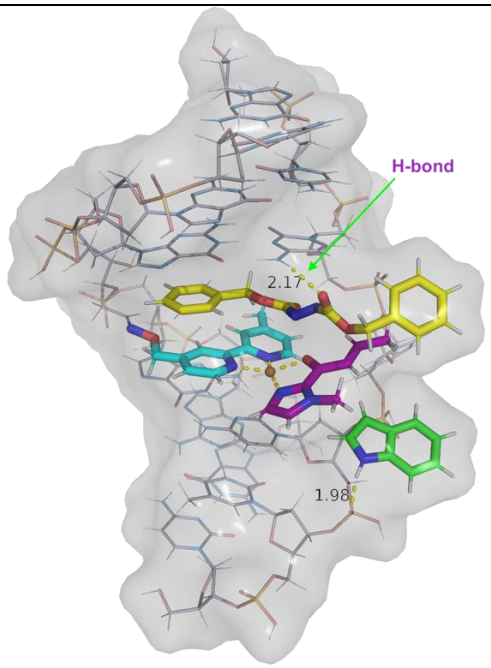
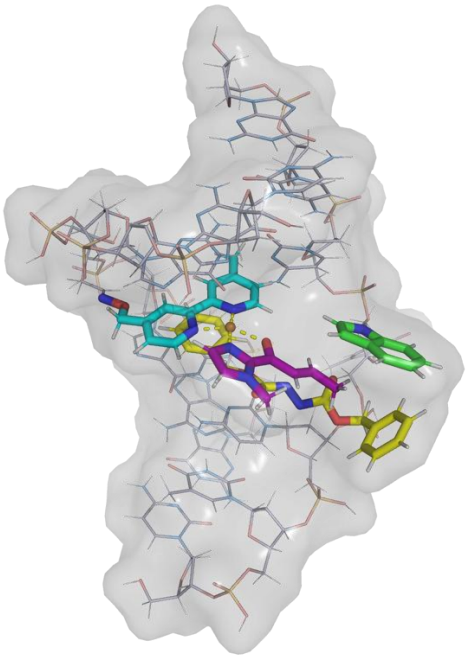
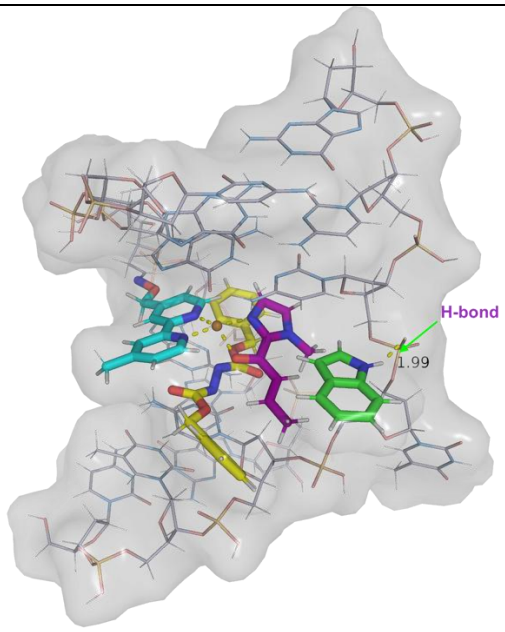
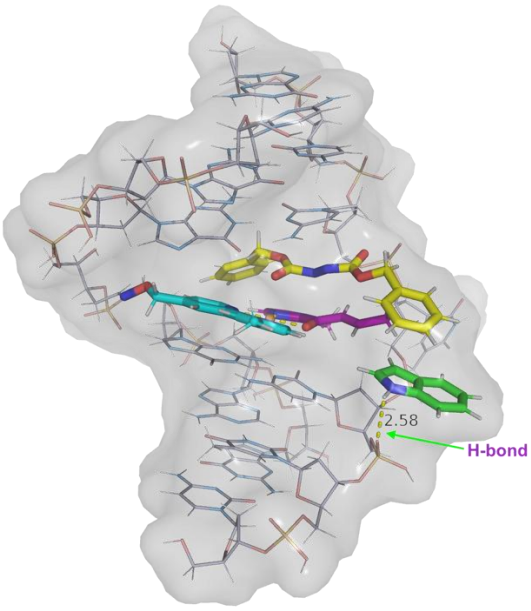
C_{in}***R_3c***	C_{in}***S_3c***
ΔΔE = 15.2 kcal mol^{-1}	ΔΔE = 31.5 kcal mol^{-1}
	
D$_{in}$*R_3c*	**D$_{in}$*S_3c***
ΔΔE = 6.5 kcal mol^{-1}	ΔΔE = 25.8 kcal mol^{-1}
	

Figure S7. GFN2-xTB-optimized complexes with bound dibenzyl (*E*)-diazene-1,2-dicarboxylate and indole substrates. The suffix ***3c*** indicates 3-component reactions. Relative energy values are given in kcal mol^{-1}.

Herein, **B$_{in}$*R_3c*** is the most stable complex, which will lead to the *S*-configuration at the β-carbon of the enone substrate/ketone product and *R*-configuration at the α-carbon of the enone substrate/ketone product, observed experimentally (**3a**, Model reaction). whereas **B$_{in}$*S_3c*** is the most stable complex leading to the product formation with opposite chirality.

To examine the binding of *N*-methylated indole **3s** (**Figure 3**), which cannot form the N–H···phosphate hydrogen bond identified for the unsubstituted indole, we constructed additional models by replacing the indole substrate in **A$_{in}$*R_3c*** with **3s**. Multiple conformations were examined using the same computational protocol. The lowest-energy structure among those sampled is shown in **Figure S8**. In this structure, substrate binding is supported by an alternative network of noncovalent interactions, including a C–H···N contact with an adjacent guanine base (H···N distance: 2.72 Å), a C–H···O contact, and π–π stacking. This model illustrates how the adaptable network of noncovalent interactions within the DNA scaffold may accommodate the N-methylated substrate in the absence of the indole N–H···phosphate hydrogen bond.

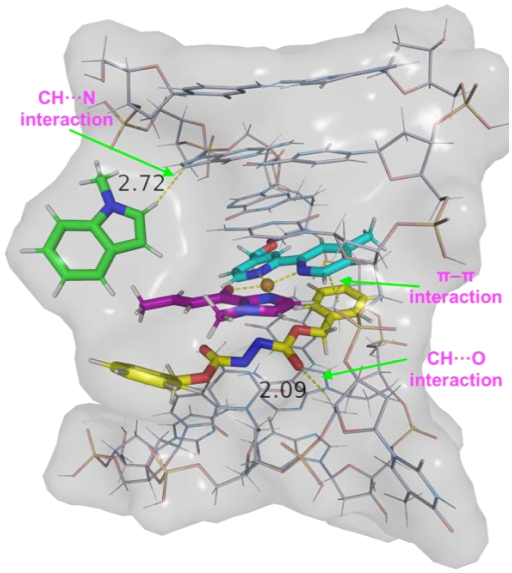
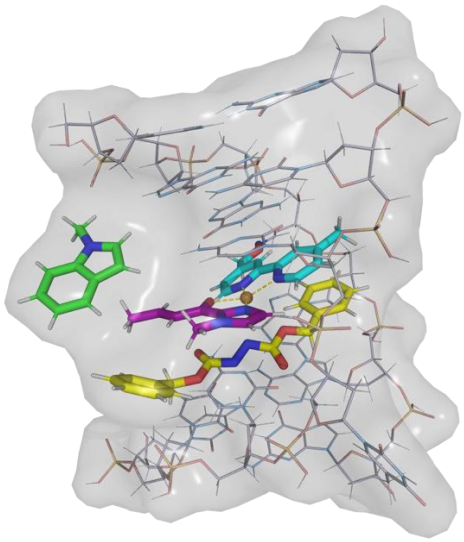
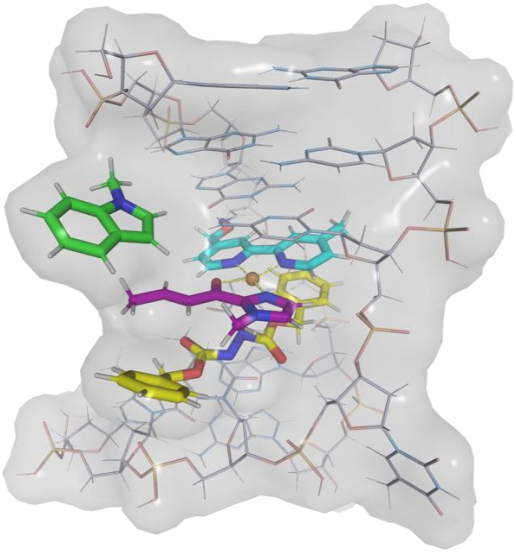
A$_{in}$*R_3c_3s_c1* (=**A$_{in}$*R_3c_3s***)	**A$_{in}$*R_3c_3s_c2***
ΔΔE = 0.0 kcal mol^{-1}	ΔΔE = 0.4 kcal mol^{-1}
	
A$_{in}$*R_3c_3s_c3*	
ΔΔE = 15.3 kcal mol^{-1}	
	

Figure S8. GFN2-xTB-optimized structures of the DNA-bound model containing compound **3s**. Relative energy values are given in kcal mol^{-1}.

GFN2-xTB-optimized structures and raw energy values

Geometries of all xTB-optimized structures (in .xyz format with their associated single point solvent-phase corrected DFT energy in Hartrees) are included in a separate folder named *xTB_optimized_structures*. All these data have been uploaded to zenodo.org with DOI: 10.5281/zenodo.23114583 (https://doi.org/10.5281/zenodo.23114583).