Influence of local sequence context on damaged base conformation in human DNA polymerase ι: Molecular dynamics studies of nucleotide incorporation opposite a benzo[a]pyrene-derived adenine lesion

Supplementary material

Kerry Donny-Clark and Suse Broyde*
Department of Biology, New York University, New York, NY, 10003, USA
*Corresponding author: Suse Broyde, Tel. (212) 998-8231, Fax (212) 995-4015
Email: broyde@nyu.edu
Clustering
Due to large changes in the structures we examined over the 20 ns trajectory we elected to cluster the individual frames in order to reveal common, persistent structural motifs in the trajectory. We chose to implement a clustering method that would identify the number of substates that a trajectory sampled, as well as their duration and flexibility. This eliminated a large number of clustering algorithms that require the researcher to provide the number of clusters before the clustering takes place, such as hierarchical, linkage, k-means, and centripetal algorithms (1). In analysis of a dynamics trajectory the number of clusters, each representing a semi-stable substate, can vary widely depending on length of the trajectory, temperature, flexibility of the structure, initial conformation, and so on. Thus when examining a large number of trajectories it is beneficial to have an algorithm that can determine the appropriate number of clusters without intervention by the researcher. Another requirement was that not all structures needed to be included in a cluster. A molecular dynamics trajectory represents the exploration of an energy landscape. We wanted to capture the predominant substates in our clusters, and exclude the less prevalent and therefore less preferred intermediate structures. Finally, we wanted a method that would perform well with simulations of more and less flexible entities with equal facility. If a given protein is more flexible then the clusters would be expected to include a more diverse range of structures in a single cluster, while a less flexible protein might have only subtle differences in distinct substates.

We settled upon a quality threshold clustering algorithm, also known as a nearest neighbor algorithm. This algorithm has been used in clustering gene expression data (2), as well as in selecting models for ab initio protein structure prediction (3). The algorithm works by creating temporary clusters that contain all structures that are sufficiently similar to a given frame, i.e., with an RMSD that is below a given cutoff value (see below for details about the generation of cutoff values). The largest of these clusters is accepted as permanent, and the structures (trajectory frames) contained in that cluster are withdrawn from the trajectory. This process is repeated until there are no temporary clusters with more than 500 members. We chose this algorithm because it can produce an appropriate number of clusters without input from the researcher, and it does not force all frames into clusters. We implemented this algorithm within the open-source program ptraj, part of the Ambertools package.

The form of the algorithm is as follows:
While there are clusters with more than 500 members
For (I) each structure (frame) of the trajectory
    Start a new cluster with this frame as the initial member
    For (II) each remaining structure in the trajectory
        If the RMSD between the initial member and the structure being examined is less than a cutoff value, then add the structure being examined to the cluster
    End for (II)
End for (I)
Select the cluster with the most members.
Remove all members of this cluster from the trajectory
End while
The cutoff value for these experiments was determined automatically by taking an average of RMSD values in two adjacent sliding windows of 50 values each. If the
difference between the averages of these two windows was greater than 0.3 Å then said difference was averaged with all other differences greater than 0.3 Å. This average of differences was then used as the cutoff value.
| Structure           | χ (°)  | α (°)  | β (°)  |
|---------------------|--------|--------|--------|
| Hoogsteen SeqI torsion 1 | 21.60  | 0.17   | -179.94 |
| Watson-Crick SeqI torsion 1 | -158.39 | 166.28 | -114.12 |
| Hoogsteen SeqI torsion 2  | 21.60  | 65.47  | 30.46  |
| Watson-Crick SeqI torsion 2 | -158.39 | 108.44 | 171.90 |
| Hoogsteen SeqII torsion 1 | 21.60  | 65.47  | 30.46  |
| Watson-Crick SeqII torsion 1 | -158.38 | 108.44 | 171.90 |
| Hoogsteen SeqII torsion 2  | 21.60  | 0.17   | -179.94 |
| Watson-Crick SeqII torsion 2 | -158.39 | 166.28 | -114.12 |
| Binary complex SeqI torsion 1 | -158.39 | 166.28 | -114.12 |
| Binary complex SeqI torsion 2  | -158.39 | 108.44 | 171.90 |
| Binary complex SeqII torsion 1 | -158.39 | 108.44 | 171.90 |
| Binary complex SeqII torsion 2  | -158.39 | 166.28 | -114.12 |
Table S2. Initial glycosidic bond conformation of templating BP-dA and incoming dNTPs for all ternary BP-dA simulations

|                 | Seq I dATP | Seq I dCTP | Seq I dGTP | Seq I dTTP | Seq II dATP | Seq II dCTP | Seq II dGTP | Seq II dTTP |
|-----------------|------------|------------|------------|------------|------------|------------|------------|------------|
| **Templating BP-dA** | anti       | anti       | anti       | anti       | syn        | syn        | syn        | syn        |
| **Incoming dNTP** | syn        | anti       | syn        | anti       | anti       | anti       | anti       | anti       |
| Atom Name | Atom Type | Topology type | Partial Charge |
|-----------|-----------|---------------|----------------|
| P         | P         | M             | 1.131218       |
| O1P       | O2        | E             | -0.803405      |
| O2P       | O2        | E             | -0.803405      |
| O5'       | OS        | M             | -0.445756      |
| C5'       | CT        | M             | 0.063485       |
| H5'1      | H1        | E             | 0.063539       |
| H5'2      | H1        | E             | 0.063539       |
| C4'       | CT        | M             | 0.117748       |
| H4'       | H1        | E             | 0.069453       |
| O4'       | OS        | E             | -0.363278      |
| C3'       | CT        | M             | 0.274079       |
| H3'       | H1        | E             | 0.032242       |
| C2'       | CT        | 3             | -0.143767      |
| H2'1      | HC        | E             | 0.049317       |
| H2'2      | HC        | E             | 0.075334       |
| C1'       | CT        | B             | 0.222074       |
| H1'       | H2        | E             | 0.082067       |
| N9        | N*        | B             | -0.196515      |
| C8        | CK        | B             | 0.237235       |
| H8        | H5        | E             | 0.128072       |
| N7        | NB        | S             | -0.556831      |
| C5        | CB        | E             | -0.087619      |
| C4        | CB        | S             | 0.543716       |
| N3        | NC        | S             | -0.683713      |
| C2        | CQ        | B             | 0.525806       |
| H2        | H5        | E             | 0.065395       |
| N1        | NC        | S             | -0.742266      |
| C6        | CA        | S             | 0.750177       |
| N6        | N*        | B             | -0.691924      |
| H6        | H         | E             | 0.353175       |
| C10       | CT        | 3             | -0.030365      |
| H10       | H1        | E             | 0.121855       |
| C1A       | CA        | S             | -0.007134      |
| C1B       | CA        | S             | 0.002375       |
| C16       | CA        | B             | -0.131161      |
| H16       | H4        | E             | 0.142308       |
| C15       | CA        | S             | -0.251353      |
| H15       | H4        | E             | 0.164741       |
| C9        | CT        | 3             | 0.178241       |
| H9        | H1        | E             | 0.082457       |
| O9        | OH        | S             | -0.617190      |
| HO9       | HO     | E | 0.406802 |
|-----------|--------|---|----------|
| C         | CT     | 3 | 0.057393 |
| HC        | H1     | E | 0.152618 |
| O8        | OH     | S | -0.588034|
| HO8       | HO     | E | 0.370703 |
| C7        | CT     | 3 | 0.261105 |
| H7        | H1     | E | 0.072959 |
| O7        | OH     | S | -0.597298|
| HO7       | HO     | E | 0.374368 |
| C6A       | CA     | M | 1.131218 |
| C1        | CA     | E | -0.803405|
| H1        | HA     | E | -0.803405|
| C5A       | CA     | M | -0.445756|
| C3        | CA     | M | 0.063485 |
| H3        | HA     | E | 0.063539 |
| C11       | CA     | E | 0.063539 |
| H11       | HA     | M | 0.117748 |
| C3A       | CA     | E | 0.069453 |
| C12       | CA     | E | -0.363278|
| H12       | HA     | M | 0.274079 |
| C13       | CA     | E | 0.032242 |
| H13       | HA     | 3 | -0.143767|
| C14       | CA     | E | 0.049317 |
| H14       | HA     | E | 0.075334 |
| C2A       | CA     | B | 0.222074 |
| C2B       | CA     | E | 0.082067 |
| C2C       | CA     | B | -0.196515|
| O3'       | OS     | B | 0.237235 |
Table S4. Pα-O3` distances and angles and Mg²⁺ to Mg²⁺ distance

| Structure          | Pα-O3` distance (Å) | Pα-O3` angle (°) | Mg²⁺ to Mg²⁺ distance (Å) |
|--------------------|---------------------|-----------------|---------------------------|
| BP-dA SeqI dATP    | 3.28±0.14           | 171.72±3.95     | 3.85±0.10                 |
| Syn Control SeqI dATP | 3.21±0.19           | 168.86±5.70     | 3.92±0.13                 |
| BP-dA SeqII dATP   | 3.22±0.16           | 168.66±5.38     | 3.92±0.11                 |
| Syn Control SeqII dATP | 3.27±0.12          | 169.76±4.50     | 3.88±0.11                 |
| BP-dA SeqI dCTP    | 3.39±0.17           | 166.18±5.74     | 3.88±0.12                 |
| Syn Control SeqI dCTP | 3.33±0.14           | 165.22±5.25     | 3.90±0.12                 |
| BP-dA SeqII dCTP   | 3.14±0.11           | 170.77±4.60     | 3.84±0.12                 |
| Syn Control SeqII dCTP | 3.37±0.18           | 163.32±6.06     | 3.93±0.12                 |
| BP-dA SeqI dGTP    | 3.21±0.17           | 170.19±4.79     | 3.88±0.13                 |
| Syn Control SeqI dGTP | 3.08±0.10           | 171.76±4.16     | 3.80±0.13                 |
| BP-dA SeqII dGTP   | 3.27±0.15           | 169.26±4.91     | 3.86±0.13                 |
| Syn Control SeqII dGTP | 3.48±0.24           | 166.31±5.85     | 3.93±0.14                 |
| BP-dA SeqI dTTP    | 3.37±0.15           | 166.28±5.84     | 3.95±0.13                 |
| Syn Control SeqI dTTP | 3.37±0.14           | 165.84±5.40     | 3.92±0.12                 |
| Anti Control SeqI dTTP | 3.39±0.14           | 165.19±5.38     | 3.96±0.12                 |
| BP-dA SeqII dTTP   | 3.22±0.14           | 165.19±6.11     | 3.95±0.12                 |
| Syn Control SeqII dTTP | 3.41±0.21           | 161.65±6.43     | 4.04±0.13                 |
| Anti Control SeqII dTTP | 3.32±0.12           | 168.64±5.30     | 3.88±0.12                 |
Table S5. Mg\textsubscript{\textsuperscript{2+}} coordination distances and standard deviations

| Structure       | Mg\textsubscript{\textsuperscript{2+}} to Asp126O\textsubscript{δ1} (Å) | Mg\textsubscript{\textsuperscript{2+}} to dC O\textsubscript{3} (Å) | Mg\textsubscript{\textsuperscript{2+}} to Glu127 O\textsubscript{ε2} (Å) | Mg\textsubscript{\textsuperscript{2+}} to dNTP O\textsubscript{1α} (Å) | Mg\textsubscript{\textsuperscript{2+}} to Asp34 O\textsubscript{δ2} (Å) |
|-----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| BP-dA SeqI dATP | 1.92±0.05                   | 2.08±0.08                   | 1.86±0.04                   | 2.12±0.20                   | 1.90±0.05                   |
| Syn Control SeqI dATP | 1.92±0.05             | 2.14±0.10                   | 1.86±0.04                   | 1.99±0.08                   | 1.91±0.05                   |
| BP-dA SeqII dATP | 1.94±0.06                   | 2.14±0.10                   | 1.86±0.04                   | 1.98±0.08                   | 1.91±0.05                   |
| Syn Control SeqII dATP | 1.94±0.06             | 2.09±0.08                   | 1.87±0.04                   | 1.97±0.07                   | 3.65±0.11                   |
| BP-dA SeqI dCTP | 1.92±0.05                   | 2.12±0.10                   | 1.86±0.04                   | 1.99±0.15                   | 1.91±0.05                   |
| Syn Control SeqI dCTP | 1.94±0.06             | 2.08±0.08                   | 1.87±0.04                   | 1.93±0.06                   | 1.92±0.05                   |
| BP-dA SeqII dCTP | 1.93±0.05                   | 2.12±0.09                   | 1.86±0.04                   | 1.97±0.09                   | 1.91±0.05                   |
| Syn Control SeqII dCTP | 1.93±0.06             | 2.10±0.08                   | 1.87±0.04                   | 1.94±0.07                   | 1.92±0.05                   |
| BP-dA SeqI dGTP | 1.92±0.05                   | 2.16±0.11                   | 1.86±0.04                   | 1.99±0.12                   | 1.90±0.05                   |
| Syn Control SeqI dGTP | 1.94±0.06             | 2.15±0.11                   | 1.86±0.04                   | 1.98±0.10                   | 1.90±0.05                   |
| BP-dA SeqII dGTP | 1.93±0.05                   | 2.10±0.08                   | 1.87±0.04                   | 1.97±0.10                   | 1.91±0.05                   |
| Syn Control SeqII dGTP | 1.92±0.05             | 2.11±0.09                   | 1.87±0.04                   | 2.05±0.25                   | 1.90±0.05                   |
| BP-dA SeqI dTTP | 1.92±0.05                   | 2.10±0.09                   | 1.86±0.04                   | 1.99±0.14                   | 1.91±0.05                   |
| Syn Control SeqI dTTP | 1.93±0.05             | 2.10±0.09                   | 1.87±0.04                   | 1.96±0.08                   | 1.92±0.05                   |
| Anti Control SeqI dTTP | 1.92±0.05            | 2.10±0.09                   | 1.87±0.04                   | 1.97±0.10                   | 1.91±0.05                   |
| BP-dA SeqII dTTP | 1.91±0.05                   | 2.12±0.09                   | 1.86±0.04                   | 1.96±0.07                   | 1.91±0.05                   |
| Syn Control SeqII dTTP | 1.92±0.05             | 2.11±0.09                   | 1.87±0.04                   | 1.95±0.07                   | 1.91±0.05                   |
| Anti Control SeqII dTTP | 1.93±0.05            | 2.09±0.08                   | 1.86±0.04                   | 2.01±0.14                   | 1.92±0.05                   |
| Structure       | Mg$^{2+}$ to Asp34 Oδ1 (Å) | Mg$^{2+}$ to dNTP O1γ (Å) | Mg$^{2+}$ to Leu35 O (Å) | Mg$^{2+}$ to dNTP O2β (Å) | Mg$^{2+}$ to Asp126 Oδ2 (Å) | Mg$^{2+}$ to dNTP O1α (Å) |
|-----------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| BP-dA SeqI dATP | 1.89 ±0.05                | 1.82±0.04                 | 1.94±0.06                 | 1.85±0.04                 | 1.92±0.05                 | 2.50±0.26                 |
| Syn             |                           |                           |                           |                           |                           |                           |
| Control SeqI dATP | 1.88±0.05                | 1.82±0.03                 | 1.92±0.05                 | 1.85±0.04                 | 1.91±0.05                 | 2.86±0.24                 |
| BP-dA SeqII dATP | 1.88±0.04                | 1.81±0.03                 | 1.92±0.05                 | 1.85±0.04                 | 1.91±0.05                 | 2.88±0.20                 |
| Syn             |                           |                           |                           |                           |                           |                           |
| Control SeqII dATP | 3.21±0.13                | 1.81±0.03                 | 1.92±0.05                 | 1.85±0.04                 | 1.91±0.05                 | 2.82±0.20                 |
| BP-dA SeqI dCTP | 1.87±0.04                | 1.82±0.04                 | 1.92±0.05                 | 1.91±0.05                 | 1.90±0.05                 | 2.84±0.29                 |
| Syn             |                           |                           |                           |                           |                           |                           |
| Control SeqI dCTP | 1.87±0.04                | 1.82±0.03                 | 1.91±0.05                 | 1.91±0.05                 | 1.90±0.05                 | 2.94±0.20                 |
| BP-dA SeqII dCTP | 1.88±0.04                | 1.82±0.04                 | 1.92±0.05                 | 1.92±0.05                 | 1.90±0.05                 | 2.74±0.23                 |
| Syn             |                           |                           |                           |                           |                           |                           |
| Control SeqII dGTP | 1.87±0.04                | 1.82±0.03                 | 1.91±0.05                 | 1.89±0.05                 | 2.97±0.22                 |
| BP-dA SeqI dGTP | 1.88±0.04                | 1.85±0.04                 | 1.94±0.06                 | 1.85±0.04                 | 1.91±0.05                 | 2.69±0.29                 |
| Syn             |                           |                           |                           |                           |                           |                           |
| Control SeqI dGTP | 1.90±0.05                | 1.84±0.04                 | 1.93±0.06                 | 1.85±0.04                 | 1.91±0.05                 | 2.64±0.25                 |
| BP-dA SeqII dGTP | 1.90±0.05                | 1.84±0.04                 | 1.93±0.06                 | 1.85±0.04                 | 1.90±0.05                 | 2.74±0.27                 |
| Syn             |                           |                           |                           |                           |                           |                           |
| Control SeqII dGTP | 1.89±0.05                | 1.84±0.04                 | 1.94±0.06                 | 1.85±0.04                 | 1.91±0.05                 | 2.75±0.39                 |
|                | BP-dA | SeqI  | dTTP | Syn Control | SeqI  | dTTP | Anti Control | SeqI  | dTTP |
|----------------|-------|-------|------|-------------|-------|------|--------------|-------|------|
|                |       |       |      |             |       |      |              |       |      |
| dA             | 1.87±0.04 | 1.82±0.04 | 1.92±0.05 | 1.91±0.05 | 1.90±0.05 | 2.87±0.28 |
| SeqI           | 1.87±0.04 | 1.82±0.03 | 1.92±0.05 | 1.91±0.05 | 1.89±0.05 | 2.91±0.22 |
| dTTP           | 1.87±0.04 | 1.82±0.04 | 1.92±0.05 | 1.91±0.05 | 1.90±0.05 | 2.92±0.25 |
| SeqII          | 1.86±0.04 | 1.82±0.04 | 1.92±0.05 | 1.91±0.05 | 1.91±0.05 | 3.01±0.23 |
| dTTP           | 1.86±0.04 | 1.82±0.04 | 1.91±0.05 | 1.91±0.05 | 1.90±0.05 | 3.12±0.25 |
| SeqII          | 1.86±0.04 | 1.83±0.04 | 1.93±0.05 | 1.91±0.05 | 1.90±0.05 | 2.71±0.27 |
Figures S1-S22 are presented in the following uniform format:

**A.**

This is a listing of all hydrogen bonds involving either the templating base (BPA or dN394) or the incoming dNTP (dNTP421) with an occupancy > 10%.

HG = Hoogsteen, i.e. templating base syn.
WC = Watson-Crick, i.e. templating base anti.

HG SEQUENCE 1 INCOMING dGTP

These graphs show occupancy of hydrogen bonds and cluster membership, respectively. If a given frame has a hydrogen bond or is a member of a cluster then there is a mark at that timepoint.

A list of the clusters obtained from this trajectory. Numbering is arbitrary, but Cluster 0 always represents the structures excluded from all other clusters due to large RMSD values.

Mean and standard deviation of each value on the graph

All graphs share the same timescale.

**B. and C.**
The most representative structures from the last cluster of each trajectory have been selected for illustrative purposes (see Methods for details). **B.** Shows the active site, while **C.** shows the whole ternary complex. Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.

**D.**
Active site RMSD time course analysis. The active site is defined as all residues having atoms within 8.0Å of any atom in the nascent base pair.
Figure S1: Sequence I unmodified binary complex
A. Hydrogen bonding, clustering, and time course analysis

Binary control sequence 1

![Diagram showing binary control sequence 1 with clusters and time course analysis for hydrogen bonding, clustering, and time course analysis.

- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4

Templating base...-12.00 ± 1.27 kcal/mol

Base stacking energy (kcal/mol)

Torsion angle (degrees)
Figure S1: Sequence I unmodified binary complex continued.  
B. Active site view  C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

![Graph showing RMSD over time for binary complex control sequence 1.](image)
Figure S2: Sequence II unmodified binary complex
A. Hydrogen bonding, clustering, and time course analysis

Binary control sequence 2

No Hydrogen bonds with occupancy > 10%

Cluster 4
Cluster 3
Cluster 2
Cluster 1
Cluster 0

Templating base...-12.74 ± 1.29 kcal/mol

Base Stacking Energy (kcal/mol)

Torsion Angle (degree)

15
Figure S2: Sequence II unmodified binary complex continued.

B. Active site view 

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S3: Sequence I BP-dA binary complex
A. Hydrogen bonding, clustering, and time course analysis
Figure S3: Sequence I BP-dA binary complex continued.

B. Active site view  
C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

BP-DA BINARY COMPLEX SEQUENCE 1
Figure S4- Sequence II BP-dA binary complex
A. Hydrogen bonding, clustering, and time course analysis
Figure S4- Sequence II BP-dA binary complex continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S5: Sequence I unmodified syn ternary complex with incoming dTTP
A. Hydrogen bonding, clustering, and time course analysis

CONTROL SEQUENCE 1 INCOMING dTTP

Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

C1 - C1: 9.06 ± 0.37 Å

Templating base... -11.47 ± 1.58 kcal/mol
incoming dNTP... -10.59 ± 1.74 kcal/mol

Stacking Energy [kcal/mol]
Figure S5: Sequence I unmodified syn ternary complex with incoming dTTP continued.

B. Active site view
C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

CONTROL SEQUENCE 1 INCOMING dTTP
Figure S6: Sequence II unmodified syn ternary complex with incoming dTTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S6: Sequence II unmodified syn ternary complex with incoming dTTP continued.

B. Active site view  C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

CONTROL SEQUENCE 2 INCOMING dTTP
Figure S7: Sequence I unmodified anti ternary complex with incoming dTTP

A. Hydrogen bonding, clustering, and time course analysis
Figure S7: Sequence I unmodified *anti* ternary complex with incoming dTTP continued.

**B. Active site view**  
**C. Whole enzyme view**

**D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.**

![Graph showing RMSD over time for active site of enzyme with incoming dTTP. The x-axis represents time in ps (picoseconds) ranging from 0 to 20,000, and the y-axis represents RMSD in Å (angstroms) ranging from 0 to 5. The graph shows a steady RMSD value with fluctuations around the 2 Å mark.]
Figure S8: Sequence II unmodified \textit{anti} ternary complex with incoming dTTP

A. Hydrogen bonding, clustering, and time course analysis
Figure S8: Sequence II unmodified anti ternary complex with incoming dTTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S9: Sequence I unmodified syn ternary complex with incoming dCTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S9: Sequence I unmodified syn ternary complex with incoming dCTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S10: Sequence II unmodified syn ternary complex with incoming dCTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S10: Sequence II unmodified *syn* ternary complex with incoming dCTP continued.

**B. Active site view**

**C. Whole enzyme view**

**D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.**
Figure S11: Sequence I unmodified syn ternary complex with incoming dATP
A. Hydrogen bonding, clustering, and time course analysis
Figure S11: Sequence I unmodified syn ternary complex with incoming dATP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

CONTROL SEQUENCE 1 INCOMING dATP
Figure S12: Sequence II unmodified syn ternary complex with incoming dATP

A. Hydrogen bonding, clustering, and time course analysis

CONTROL SEQUENCE 2 INCOMING dATP

Cluster 5
Cluster 4
Cluster 3
Cluster 2
Cluster 1
Cluster 0

C1-C1...14.69 ± 1.33 Å

Templating base...-9.66 ± 1.21 kcal/mol
incoming dNTP...-7.94 ± 1.44 kcal/mol

Stacking Energy (kcal/mol)

Tension (deg)
Figure S12: Sequence II unmodified syn ternary complex with incoming dATP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

CONTROL SEQUENCE 2 INCOMING dATP
Figure S13: Sequence I unmodified syn ternary complex with incoming dGTP

A. Hydrogen bonding, clustering, and time course analysis
Figure S13: Sequence I unmodified syn ternary complex with incoming dGTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S14: Sequence II unmodified syn ternary complex with incoming dGTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S14: Sequence II unmodified syn ternary complex with incoming dGTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

CONTROL SEQUENCE 2 INCOMING dGTP
Figure S15: Sequence I BP-dA anti ternary complex with incoming dTTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S15: Sequence I BP-dA *anti* ternary complex with incoming dTTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S16: Sequence II BP-dA syn ternary complex with incoming dTTP
A. Hydrogen bonding, clustering, and time course analysis

HG SEQUENCE 2 INCOMING dTTP

Cluster 7
Cluster 6
Cluster 5
Cluster 4
Cluster 3
Cluster 2
Cluster 1
Cluster 0

C1-C1 Distance (Å)

Tempating base...-9.46 ± 1.29 kcal/mol
incoming dNTP...-6.24 ± 1.45 kcal/mol

Temperature (deg)

-33.29 ± 14.65°  0.64 ± 17.93°  88.33 ± 12.08°
Figure S16: Sequence II BP-dA syn ternary complex with incoming dTTP continued.

B. Active site view  C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S17: Sequence I BP-dA anti ternary complex with incoming dCTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S17: Sequence I BP-dA *anti* ternary complex with incoming dCTP continued.

**B. Active site view**

**C. Whole enzyme view**

**D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.**

![Active site RMSD graph](image)
Figure S18: Sequence II BP-dA syn ternary complex with incoming dCTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S18: Sequence II BP-dA syn ternary complex with incoming dCTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

HG SEQUENCE 2 INCOMING dCTP TORSION 2
Figure S19: Sequence I BP-dA Ianti ternary complex with incoming dATP
A. Hydrogen bonding, clustering, and time course analysis
Figure S19: Sequence I BP-dA \textit{anti} ternary complex with incoming dATP continued.

B. Active site view 

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

WC SEQUENCE 1 INCOMING dATP TORSION 2
Figure S20: Sequence II BP-dA syn ternary complex with incoming dATP
A. Hydrogen bonding, clustering, and time course analysis
Figure S20: Sequence II BP-dA syn ternary complex with incoming dATP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

HG SEQUENCE 2 INCOMING dATP TORSION 1
Figure S21: Sequence I BP-dA anti ternary complex with incoming dGTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S21: Sequence I BP-dA anti ternary complex with incoming dGTP continued.

B. Active site view  

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.

WC SEQUENCE 1 INCOMING dGTP TORSION 2
Figure S22: Sequence II BP-dA syn ternary complex with incoming dGTP
A. Hydrogen bonding, clustering, and time course analysis
Figure S22: Sequence II BP-dA syn ternary complex with incoming dGTP continued.

B. Active site view

C. Whole enzyme view

D. Active site RMSD for all 20ns of the trajectory. RMSD is measured against the first frame following equilibration.
Figure S23: Binary complex SeqI torsion 1

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S24: Binary complex SeqII torsion 1

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S25: Binary complex SeqI torsion 2 (stereo)

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S26: Binary complex SeqII torsion 2 (stereo)

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg^{2+}, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S27: Ternary complex SeqI incoming dTTP (stereo)

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S28: Ternary complex SeqII incoming dTTP (stereo)

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S29: Ternary complex SeqI incoming dGTP

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg\(^{2+}\), purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S30: Ternary complex SeqII incoming dGTP

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S31: Ternary complex with *anti* BP-dA and incoming dATP in SeqII

Note the lack of hydrogen bonding between the incoming dATP and the templating BP-dA. The templating BP-dA is forced into the major groove in order to minimize the C1′-C1′ distance at 9.7Å.

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
In this structure the templating BP-dA moves towards the major groove, losing any contacts with the incoming dCTP. The C1′-C1′ distance is 10.1Å.

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg^{2+}, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
In this structure the templating BP-dA moves towards the major groove. No hydrogen bonds are formed with the incoming dGTP, and the C1′-C1′ distance is 10.0Å. Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
In this structure the lesion on the template stacks with the dT 5' to the templating BP-dA, pulling it towards the major groove and preventing hydrogen bond formation in the nascent base pair. The C1'-C1' distance is 10.7 Å.

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg^{2+}, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S35: Ternary complex with syn BP-dA and incoming dATP in SeqI

In this structure the templating BP-dA moves towards the major groove. There is no hydrogen bonding with the incoming dATP, and the C1′-C1′ distance is 11.2Å. Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg²⁺, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
Figure S36: Ternary complex with syn BP-dA and incoming dCTP in SeqI

In this structure the incoming dCTP is forced out of the active site towards the minor groove by the insertion of the BP rings stacking with the primer terminus. The C1’-C1’ distance is 12.8Å.

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
In this structure the incoming dGTP forms two strong hydrogen bonds with the templating BP-dA. The C1′-C1′ distance is 10.1Å. However, the binary complex simulation results suggest that this structure would not be obtained, and dG is not incorporated well in this sequence according to the experimental results (4). Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
In this structure the template twists and moves towards the major groove. No hydrogen bonds are formed with the incoming dTTP, and the C1’-C1’ distance is 8.9Å. However, the binary complex simulation results suggest that this structure would not be obtained, and dT is incorporated more readily than any other nucleotide in this sequence according to the experimental results (4).

Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green; Mg$^{2+}$, purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is grey.
1. Shao, J., Tanner, S.W., Thompson, N. and Cheatham, T.E., 3rd. (2007) Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation*, 3, 2312-2334.

2. Heyer, L.J., Kruglyak, S. and Yooseph, S. (1999) Exploring expression data: Identification and analysis of coexpressed genes. *Genome Research*, 9, 1106-1115.

3. Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E.M. and Baker, D. (2001) Rosetta in CASP4: Progress in *ab initio* protein structure prediction. *Proteins: Structure, Function, and Genetics*, 45, 119-126.

4. Frank, E.G., Sayer, J.M., Kroth, H., Ohashi, E., Ohmori, H., Jerina, D.M. and Woodgate, R. (2002) Translesion replication of benzo[a]pyrene and benzo[c]phenanthrene diol epoxide adducts of deoxyadenosine and deoxyguanosine by human DNA polymerase ι. *Nucleic Acids Res*, 30, 5284-5292.