nature portfolio

Peer Review File

So close yet so far apart: Distinct flanking sequence recognition by DNMT3A and DNMT3B

Corresponding Author: Dr Ezgi Karaca

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The substrate specificity of DNMT3A and DNMT3B has been clarified by various Deep Enzymology studies where DNMT3A was shown to preferentially methylate CGC/T sequences whereas DNMT3B prefers CGG/A. The molecular mechanism behind this discrimination at the CGX position was not fully understood. In the manuscript “A direct readout mechanism regulates the sequence selectivity of human de novo DNA methyltransferases” the authors use Molecular Dynamics (MD) simulations to investigate the divergent DNA substrate specificities of DNMT3A and DNMT3B. This was accomplished by complexing eight possible CGX[C/G/T/A] 10 bp long DNA substrates to DNMT3A or DNMT3B monomers. Subsequent analysis of hydrogen bonds, hydrophobic and ionic interactions were used to rationalize the divergent substrate specificities. Here they found that R836 in DNMT3A reads the guanine complementary to C+2 in its cognate CGC motif, while K777 in DNMT3B recognizes two consecutive guanines in its cognate CGG.
The manuscript is well written in general, and the techniques used in the work demonstrate a good understanding of the bioinformatic tools to tackle such a difficult question. Whereas the logic in the study itself is stringent, major concerns refer to the overall approach and setup of the experiments, the discussion of previous data, the reliability of the data for biological claims and comparisons.

Major Comments
1) The authors use PDB 6F57 for DNMT3A, PDB 6KDA for DNMT3B as templates to generate the starting structures. Why did the authors discard the other 3A or 3B subunit and the adjacent 3L subunits. Simulating the complete DNMT3A/B-3L tetramer is feasible and would give results, which could be used with a higher reliability. This is especially of relevance as it was shown, that 3A monomers are inactive, but must form a homotetramer (3A-3A-3A-3A-) or heterotetramer (3L-3A-3A-3L). This is even more relevant as several studies showed mutations in the interface of two 3A subunits, disrupting the interaction can lead to inactivity or altered catalytic properties. Simulating DNMT3A/3B monomers represents a state, not occurring in vivo or in in vitro experiments.

2) By using only one single DNMT3A or DNMT3B subunit, the DNA substrate length is automatically shortened to 10 bp, as also done in this study. This raises two concerns about the findings in this study. Firstly, the highly negatively charged phosphates of the short DNA are close (10-11 A) to important DNMT3A/3B residues e.g., K777, but also to many other residues. This could heavily influence their behavior. 10 bp long DNA substrates are no natural methylation target and the negatively charged phosphate does not occur in nature at this position. Secondly, the authors make claims about the DNA geometry differently affected by DNMT3A or 3B. This behavior could be influenced by the short DNA strand since long distance effects like twisting are different for short and long DNA molecules. These major comments could be ruled out by simulating a tetrameric complex with a longer DNA substrate as shown in e.g., PDB 6KDA for DNMT3B and PDB 6W8B for DNMT3A.

3) DNMTs have been shown to dynamically flip out the target cytosine. The study starts its MD simulation experiments from structures with an already flipped out cytosine. The process of base-flipping has been demonstrated to be the rate determining step for DNMT1 (PMID: 32709850). The authors should clarify this in the discussion. Another approach, which would be highly interesting, would be to extend the simulation experiments by employing advanced sampling techniques (steered MD simulation, Metadynamics, Replica Exchange MD) to model the base flipping process. This could be highly

valuable as flanking sequence dependent base flipping mechanisms could influence the substrate specificity and should therefore be considered.

4) Novelty of the found results: All discoveries reported here were already described in structural and biochemical papers (refs. 14, 18, 45, 46). This needs to be described more clearly in particular in the abstract and discussion, the references must be presented at a more prominent place. Two publications (ref. 45 and 46) contain detailed biochemical data addressing the roles of the amino acids investigated here and the conclusion made there are very similar with what has been deduced here. These papers and their data need to be discussed and considered. At other places, the description of previous work is inaccurate. CCC/T and CGG are not attributed as “cognate” methylation sites, but just preferences were reported. It should be mentioned and discussed that further preferences were observed (beyond the CGX position studied here). Hence important questions that are ideally targeted by MD simulations were not addressed.

5) The authors provide a mechanism, in which R836 or K777 hydrogen bonding is responsible for the substrate specificity. However, in the crystal structure PDB 6F57, which was used as a template for this study, the interaction of R836 with O6 and N7 of G+2’ can already be seen (as mentioned in the text). The same goes for the interaction of K777 with O6 of G+2 in DNMT3B, which can be observed in PDB 6KDA. The MD simulation experiments only confirm these findings. The authors should provide new findings for DNA substrates without crystal structure (e.g., interactions with DNA substrates that have not been crystalized). Likewise, providing movies of simulations showing novel interactions like DNMT3A with as GCA/T/C substrates and likewise for DNMT3B would be interesting.

6) Unfortunately, the entire simulation focuses on binding strength, because the particular conformations leading to transition states, which are intimately connected with catalysis are ignored. Of note, it is entirely unclear, if the contacts identified by the authors stabilize a conformation that is supportive for catalysis or if they hinder catalysis by stabilizing an off-pathway conformation. In this context, it is not clear why IDF also discarded interactions observed for less than 10% of the simulation time, considering them to be unimportant to the downstream analysis. These interactions could lead to a specific transition state, hence they could be of eminent importance.

7) An additional analysis could strengthen the approach of this study: Since biochemical methylation activity experiments for DNMT3A and DNMT3B have been done, one could try to correlate in vitro data with e.g., the number of hydrogens bonds found in the various complexes that were simulated. A connection between the activity and hydrogen bonds could elevate this study (e.g., does the number of hydrogen bonds for CGA match the methylation activity?) and help to decide if an observed interaction has a positive or negative role in catalysis.

Minor Comments
• It would be helpful for readers if the authors use the numbering scheme established in the field, in which the CGX position is considered the first outside of the CpG site, hence designated as +1 (and not +2).
• The TIP4P water model could be used instead of TIP3P as TIP4P more accurately represents a correct charge distribution. At least comparison runs should be done, showing no effect of the water model.
• Typo in line 132: DNMT3A-CGC ‘was’ modified.
• Table 1 and 2 could be more organized and clearer.
• Figure 1A: In the sequence alignment, marking D and N as yellow and therefore similar physiochemical properties seems not to be correct.
• Numbering in the provided starting structure PDB files does not match the manuscript.
• For base-specific hydrogen bonds, interactions of backbone atoms of amino acids and nucleotides were also filtered out from the interaction data frame. However, e.g, K777 could interact with DC`300/OP2 and R836 with DG`297/OP2. This information could be important.

Reviewer #2

(Remarks to the Author)
This is an interesting study of the interacttons of two classes DNA methyltransferases with four different DNA sequences in which the methylation target CpG is followed by one of A,C,G or T. From simulations of these systems the authors identify particular interactions that explain the sequence preferences of the two enzymes. For each complex 4 independent simulations were run, and it would be interesting to know how consistent the results are between the replicates.

Some points that need to be addressed.
Line 83. Reference for Mallona et al is missing
Line 97. What is an Interaction Dynamics Framework?
Line 112. Coordinate system? It looks like a numbering scheme.
Line 147. Unit is missing for the positional restraints.
Line 171. Why were the interaction between protein and methylated cytosine removed?
Line 172. Why were protein backbone atoms not included in the analysis of base-specifc contacts?
Line 194. The "interaction densities" seem to simply be the average number of hydrogen bonds, and could be described as such to make the text easier to read.
Line 236. Same as above, this is the average number of hydrogen bonds.
Lines 481-482. References missing

Reviewer #3

(Remarks to the Author)
Despite high levels of homology (91%) the two de novo methyltransferases (DNMT3A and DNMT3B) are able to selectively methylate CG steps that have different flanking sequences. Barlas and Karaka perform molecular dynamics simulations on different flanking sequences around a central CG to unveil the selectivity of different methyltrasferases.Key aminoacid differences Arg/Lys are found to be responsible for the different interaction patterns. They then identify differences in interaction patterns (e.g. hydrogen bonds) to conclude that direct interactions are responsible for selectivity.

Understanding the role of direct/inderect (base/shape) readout in protein nucleic acids has been a longstanding goal in the field. Indirect or shape readout relates to the ability of a DNA sequence to adopt conformations found in the bound state, where as the direct or base readout relates to the specfic interactions made in a complex. Understanding the key differences amongst a series of sequences requires to know what the binding free energy is for each one, and how it is distributed amongst direct and indirect readout: DG_bind = DG_base + DG_shape

And this should be done for different sequences.

However, in this study, we just see the results of very detailed analysis of interaction patterns amongst residues in the DNA and protein. This makes the article very technical and hard to read, but also provides little insight about the difference in relative importance of each readout upon changing the sequence. Is this not done because for this system its already known that direct readout is key -- e.g., from the experimental structures, and if so, what new knowledge does this work generate?

Reading in the methods about the unbound DNA simulations i thought we were going to see how sequence effects changed the conformational preferences from unbound to bound DNA. And yet, i did not see any mention of the unbound states after methods.

"We also showed that this direct readout mechanism is dictated
by the DNA groove geometry". -- Wouldn't this suggest that an indirect readout is important?

When thinking about flanking regions: why is the effect only looked at the 3' end of the CG, and why is the perturbation only up to two residues? In general, local effects can extend to the hexamer and 8-mer level.

I think a closer look at the work from Richard Lavery and from Remo Rohs at attempts to separate between direct and indirect readout for relating to differences in binding affinity would benefit this work.

And i think the authors should consider distilling their message. Currently, the technical details and how the paper is written make it very difficult to find the relevance in the work.

Technical:

Remo Rohs introduced the terms base and shape readout as more representative of protein nucleic acids than the previous direct/indirect readout mechanisms.

For future works, the Berendsen barostat is no longer recommended as it does not yield correct ensembles.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
While my first reading of the manuscript was rather optimistic, reading of the explanations in the rebuttal letter makes me afraid that principal processes are not appropriately implemented. My impression is that the oversimplifications introduced make the entire analysis unreliable and shed doubts on the usefulness of the data. The key concerns are explained in the following part:
1) DNA length: The authors used 10 bp DNA for the MD, e.g. the sequence CATGCGCTCT in which the TC step was investigates. This raises two concerns:
a) Due to the higher flexibility, the ends of DNA molecules are known the behave abnormally in MD simulations. Given that the T studied here stacks on the second to last bp, end effects are expected to have a massive impact on the conformational dynamics in this MD simulation.
b) The authors now included base shape parameters in their analysis. These parameters are known to be affected by sequences at least 3 bp in either direction. Hence, they conceptually are not suited to be used in the current investigation.
2) In their rebuttal letter the authors explain that simulations were done using only one subunit of the DNMT3 heterotetrameric complexes. This must be clearly indicated in the results and Methods sections. This appears highly questionable as both interfaces to other subunit are allosterically coupled to the DNA binding sites. It is self evident that movements in the interfaces will change the structures and dynamics of the DNA interacting residues. The authors argue to

have evidence that omission of the additional subunits does not change the results. Instead of being supportive, this argument raises concerns. If both interfaces (one of them is highly hydrophobic) artificially are hydrated, massive structural changes are expected. If these do not change the DNA interaction, the overall method is not reliable.
3) The authors base all their arguments on the contact maps observed in the simlations which are closely based on the solved structures. They argue what makes this strucuter more favorable should also stimulate catalysis. This is a massive oversimplification because enzyme do not stabilize the ground state of the enzyme substrate complex but the transition state of the chemical reaction. Transition state structures and transition state stabilization needs to be considered.
4) The authors now argue that repeated simulations were conducted. However, statistical analysis on the reproducibility and significance of observed differences were not conducted. Without this, no statements can be made about the results.
Given that fundamental technical concerns could not be resolved, I cannot recommend this manuscript for publication.

Reviewer #2

(Remarks to the Author)
My concerns ave been adequately addressed.

Reviewer #3

(Remarks to the Author)
My initial concerns have been addressed, and i feel this is a more focused and strengthened paper.

Our Point-by-point Responses to the Reviewers' Comments:

We sincerely thank the editor and reviewers for their time and thoughtful feedback, which has been invaluable in substantially improving our manuscript. In response to the comments, we undertook major revisions. Specifically, we reorganized the overall structure and added two new sections dedicated to DNA shape readout and its interplay with base readout. We also significantly expanded our analysis—nearly doubling its content—to provide deeper quantitative and mechanistic insight into DNMT3 recognition specificity.

We believe these changes have greatly enhanced the clarity, rigor, and scientific contribution of our work. Below, we provide detailed, point-by-point responses to all reviewer comments. Reviewer remarks are shown in black, and our responses are shown in red. As this version of the manuscript reflects a substantial rewrite and restructuring, we did not mark changes in red within the document. We hope this approach is acceptable to both the editor and reviewers.

All major revisions, including new sections, expanded analyses, and reorganization of the Results and Discussion, are clearly explained in our responses below. We sincerely hope that these revisions improve the biological relevance and interpretability of our study, and that our responses fully address all concerns raised.

Reviewer #1:

The substrate specificity of DNMT3A and DNMT3B has been clarified by various Deep Enzymology studies where DNMT3A was shown to preferentially methylate CGC/T sequences whereas DNMT3B prefers CGG/A. The molecular mechanism behind this discrimination at the CGX position was not fully understood. In the manuscript "A direct readout mechanism regulates the sequence selectivity of human de novo DNA methyltransferases" the authors use Molecular Dynamics (MD) simulations to investigate the divergent DNA substrate specificities of DNMT3A and DNMT3B. This was accomplished by complexing eight possible CGX[C/G/T/A] 10 bp long DNA substrates to DNMT3A or DNMT3B monomers. Subsequent analysis of hydrogen bonds, hydrophobic and ionic interactions were used to rationalize the divergent substrate specificities. Here they found that R836 in DNMT3A reads the guanine complementary to C+2 in its cognate CGC motif, while K777 in DNMT3B recognizes two consecutive guanines in its cognate CGG.

1) The manuscript is well written in general, and the techniques used in the work demonstrate a good understanding of the bioinformatic tools to tackle such a difficult question. Whereas the logic in the study itself is stringent, major concerns refer to the overall approach and setup of the experiments, the discussion of previous data, the reliability of the data for biological claims and comparisons.

We thank the reviewer for the positive evaluation of our manuscript and for acknowledging the quality of the writing. We particularly appreciate the recognition of the complexity of the problem and our efforts to tackle it with state-of-the-art computational tools. In response to the reviewer's broader concerns, we:

- Reorganized the manuscript to enhance clarity and better highlight the logic and flow of the work.
- Doubled the content of our analysis, expanding the quantitative treatment of both hydrogen bonding and shape-based recognition.
- Added a new section dedicated to the interplay between base and shape readout, which we believe provides crucial insight into the substrate preferences of DNMT3A and DNMT3B.
- Clarified our methodological choices and now provide additional justification for the system setup and interpretation of the MD results.
- Strengthened the contextualization of our work by incorporating deeper discussion of key experimental studies and explicitly linking our computational findings to previously reported biochemical observations.

2) The authors use PDB 6F57 for DNMT3A, PDB 6KDA for DNMT3B as templates to generate the starting structures. Why did the authors discard the other 3A or 3B subunit and the adjacent 3L subunits. Simulating the complete DNMT3A/B-3L tetramer is feasible and would give results, which could be used with a higher reliability.

We appreciate the reviewer's suggestion and agree that simulating the full tetrameric DNMT3A/B–3L complex is a valuable direction for future studies. However, we intentionally chose to model the catalytic monomer–DNA complexes for the following reasons:

There is no direct evidence that using a larger system would necessarily provide more reliable results. On the contrary, force fields for large heterogeneous systems, such as protein–DNA–ligand assemblies, are known to be susceptible to noise and artifacts, especially when longer DNA sequences are involved. To ensure clarity and robustness in our residue-level interaction analyses, we deliberately minimized system complexity. This approach aligns with the principle of Occam's Razor, which advocates for simplicity by positing that "entities should not be multiplied without necessity" .

In support of this approach, we refer to preliminary comparative MD data from dimeric and tetrameric 3A–3L-DNA complexes, presented in a publicly available Master's thesis from our group (Savaş, 2023). These data show that the protein–DNA interaction profiles are well-conserved between the dimer and tetramer contexts (Figure 1-left, please see below). Notably, the DNA ends in the tetrameric system exhibit substantial fluctuations, likely driven by the presence of DNMT3L at the termini, resulting in large-scale conformational oscillations (Figure 1-right, please see below). We aimed to avoid the potential confounding influence of these end effects when analyzing sequence-specific readout.

Also, importantly, Mallona et al. (2021) previously demonstrated that MD simulations of the minimal DNMT3A catalytic domain bound to SAH were sufficient to reproduce and even predict macroscopic biological outcomes, including gene knockout phenotypes. Our study builds upon and substantially extends this validated minimal system to extract new mechanistic insight into base and shape readout.

We have added this rationale to the revised Conclusions section to clarify our modeling choice.

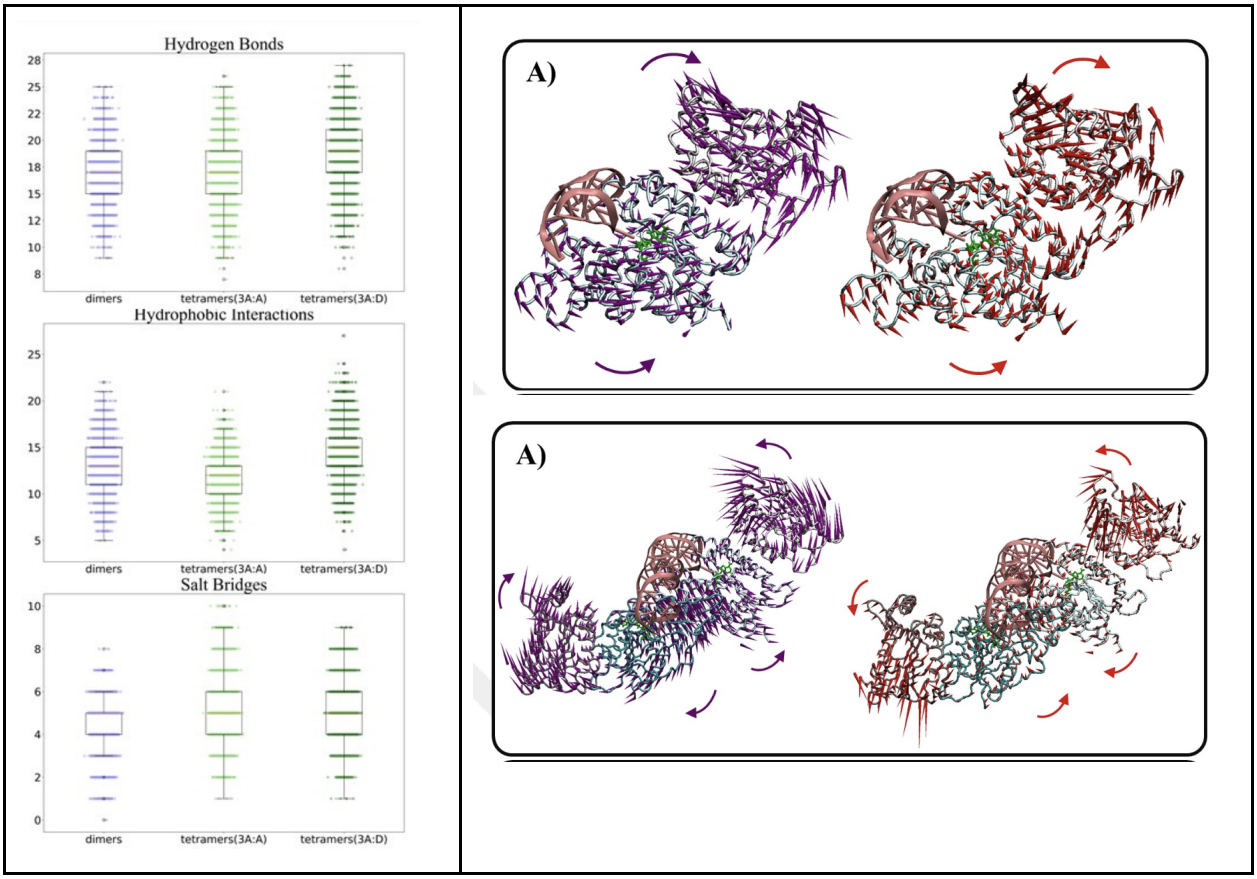

Figure 1. Preliminary data on DNMT3A–DNMT3L dimer and tetramer MD systems. Left column: Comparison of protein–DNA interaction distributions across the DNMT3A–DNMT3L dimer (PDB: 6F57) and tetramer (PDB: 5YX2) systems. Green distributions represent the protein–DNA interactions at the respective dimer interfaces. **Right column:** Amplitude of dominant fluctuations observed during MD simulations, highlighting that the tetramer system exhibits compounded effects, particularly at the DNA termini.

This is especially of relevance as it was shown, that 3A monomers are inactive, but must form a homotetramer (3A-3A-3A-3A-) or heterotetramer (3L-3A-3A-3L). This is even more relevant as several studies showed mutations in the interface of two 3A subunits, disrupting the interaction can lead to inactivity or altered catalytic properties. Simulating DNMT3A/3B monomers represents a state, not occurring in vivo or in in vitro experiments.

In addition to what we have introduced above, we would like to further explain our perspective in designing our experiments. While our simulations focused on monomeric DNMT3A and DNMT3B complexes, this approach was chosen to isolate and analyze specific base and shape readout mechanisms at an atomistic level. We recognize that this does not fully capture in vivo tetrameric state. However, our findings reveal that even within this minimal system, we can recapitulate key aspects of DNMT3A's high CpG specificity and cooperative behavior, as well as DNMT3B's

processive methylation pattern. These insights are detailed in the final section of our Results and Discussion, titled *"Distinct Methylation Strategies Are Grounded in Electrostatics and Flexibility."*

We also respectfully note that no in silico system can fully replicate in vivo or in vitro conditions. Yet this is precisely the strength of MD simulations: they allow us to understand molecular recognition principles in atomistic detail and build a mechanistic bridge between microscopic features and macroscopic behavior (which cannot be performed with any other computational or experimental method). In this context, a simple model should not be seen as inadequate, but rather as a carefully controlled system designed to isolate and illuminate the essential biophysical principles underlying sequence readout.

3) By using only one single DNMT3A or DNMT3B subunit, the DNA substrate length is automatically shortened to 10 bp, as also done in this study. This raises two concerns about the findings in this study. Firstly, the highly negatively charged phosphates of the short DNA are close (10-11 A) to important DNMT3A/3B residues e.g., K777, but also to many other residues. This could heavily influence their behavior. 10 bp long DNA substrates are no natural methylation target and the negatively charged phosphate does not occur in nature at this position.

We thank the reviewer for this important observation. Indeed, as a result of modeling a minimal monomer–DNA complex, the DNA substrate is shorter (10 bp), which could result in end-proximal phosphates being located near key protein residues. To mitigate potential artifacts arising from such proximity and charge clustering, we neutralized the terminal phosphate charges by capping the DNA ends during model preparation. This step effectively reduces artificial electrostatic interactions at the DNA termini. We have now added this information to the revised Methods section, which we overlooked to mention before.

Secondly, the authors make claims about the DNA geometry differently affected by DNMT3A or 3B. This behavior could be influenced by the short DNA strand since long distance effects like twisting are different for short and long DNA molecules. These major comments could be ruled out by simulating a tetrameric complex with a longer DNA substrate as shown in e.g., PDB 6KDA for DNMT3B and PDB 6W8B for DNMT3A.

We thank the reviewer for raising this important point. We fully agree that longer DNA fragments and multimeric complexes can influence higher-order conformational properties, such as global DNA twist, bend, or curvature. However, the primary objective of this study was to understand local sequence readout mechanisms, including base-specific hydrogen bonding, electrostatic complementarity, and groove geometry, which underlie the sequence selectivity of DNMT3 enzymes. For this purpose, short DNA fragments are not only sufficient but often preferred, as they allow for focused, high-resolution analysis of the core recognition interface without the added complexity of long-range structural fluctuations. Indeed, numerous high-impact molecular dynamics and structural studies of protein–DNA recognition have successfully employed short duplexes (typically 10–15 bp) to investigate base and shape readout. For example:

- Paillard *et al.* used 9–10 bp DNA to analyze sequence-dependent deformation and protein–DNA energy profiles in their ADAPT framework (doi: 10.1093/nar/gkh1003).

- Etheve *et al.* used MD simulations to study the interaction of Caenorhabditis elegans transcription factor SKN-1 with DNA. For ADAPT analysis, the nine central base pairs of the DNA oligomer were scanned together with base pairs added at both ends and sequence-independent binding energies were calculated for a total region of 13 bp. (doi: 10.1093/nar/gkv1511)
- Abe *et al.* performed structural and mutational studies examining the DNA recognition of Hox proteins bound to 12 bp DNA sequences (doi: 10.1016/j.cell.2015.02.008).
- Machado *et al.* based their SELEX-seq-informed modeling of DNA shape readout and TF specificity on 10-mer cores (e.g., MEF2B), where properties like minor groove width, roll, twist, and electrostatic potential were accurately captured (doi: 10.1093/nar/gkaa642).
- We recently demonstrated that the 12 bp DNA conformation and its interactions with Sox are preserved in the nucleosome-bound state, as revealed by MDsimulations and validated by experimental data (Ozden et al., 2023; doi: 10.1021/acs.jcim.2c01520). This finding supports the validity of using short, focused DNA segments to gain mechanistic insights into protein–DNA recognition, even for the proteins active in complex chromatin environments. We have now incorporated this observation into the Conclusion section of the revised manuscript.

Moreover, the DNMT3A–CGC crystal structure (PDB: 6F57), which we used as the starting point for our simulations, also features a 10 bp DNA substrate. Our modeling choices are thus consistent with both experimental structural biology and state-of-the-art computational practice in the field of sequence-specific DNA recognition.

4) DNMTs have been shown to dynamically flip out the target cytosine. The study starts its MD simulation experiments from structures with an already flipped out cytosine. The process of base-flipping has been demonstrated to be the rate determining step for DNMT1 (PMID: 32709850). The authors should clarify this in the discussion. Another approach, which would be highly interesting, would be to extend the simulation experiments by employing advanced sampling techniques (steered MD simulation, Metadynamics, Replica Exchange MD) to model the base flipping process. This could be highly valuable as flanking sequence dependent base flipping mechanisms could influence the substrate specificity and should therefore be considered.

We thank the reviewer for emphasizing the importance of base-flipping dynamics in DNA methyltransferases. We fully agree that understanding the energetics and sequence dependence of the flipping process would offer deeper mechanistic insight. However, our study was intentionally designed to isolate and examine the pre-reaction phase, focusing specifically on how flanking bases influence the geometry and stability of the recognition interface. To this end, we modeled pre-reactive complexes containing a flipped-out cytosine, in line with available crystal structures (e.g., PDB: 6F57 and 6KDA). This choice enabled us to analyze the side-chain–DNA interactions and local DNA shape features without the added complexity of the chemical reaction step. Our assumption was that flanking sequence preferences could be traced independently of the flipping or catalytic events and our results confirm that this was a valid and productive assumption.

Additionally, from a technical perspective, modeling base-flipping or methyl transfer reactions would require a hybrid quantum mechanics/molecular mechanics (QM/MM) approach to account for electronic rearrangements, proton transfer pathways, and transition state stabilization. Such simulations on full DNMT–DNA–SAM systems in explicit solvent are currently not tractable for broad conformational sampling across multiple sequence variants and were thus beyond the scope of this work. Moreover, standard biomolecular force fields are not equipped to handle the energetics or coordinate-dependent protonation shifts associated with cytosine flipping. Without custom-developed collective variables or force field corrections tailored to the active site environment, enhanced-sampling simulations (e.g., steered MD, metadynamics) would carry significant uncertainties.

We now clarify these points in the Conclusion section of the revised manuscript.

5) Novelty of the found results: All discoveries reported here were already described in structural and biochemical papers (refs. 14, 18, 45, 46). This needs to be described more clearly in particular in the abstract and discussion, the references must be presented at a more prominent place. Two publications (ref. 45 and 46) contain detailed biochemical data addressing the roles of the amino acids investigated here and the conclusion made there are very similar with what has been deduced here. These papers and their data need to be discussed and considered.

We thank the reviewer for this important comment and fully agree on the need to clearly situate our work within the context of existing structural and biochemical literature, including references 14, 18, 45, and 46. In particular, the studies by Dukatz et al. (2020) and Dukatz et al. (2022) have made substantial contributions by identifying key amino acid residues (e.g., Arg836, Lys777, Asn838, Asn779) involved in the sequence preferences of DNMT3A and DNMT3B, based on crystallographic, enzymological, and deep methylation profiling data. We now discuss these works in more detail and have moved the references to more relevant positions in the Results and Discussion sections. That said, the novelty of our study lies not in the identification of these residues, but in revealing how they function dynamically in different sequence contexts. Our Comparative Dynamics Analysis (CDA) framework introduces a time-resolved, mechanistic perspective on DNMT3–DNA recognition that goes beyond what is accessible through static crystal structures or endpoint biochemical assays.

To be more specific:

- While previous studies identified key interactions with flanking bases, our 2 μs MD simulations per system, analyzed through the CDA framework, reveal bidentate hydrogen bonding geometries, interaction frequencies, and alternative donor–acceptor chemistries that vary in a sequence-specific and time-dependent manner. These features explain how flanking nucleotides modulate specificity in ways that are not captured by static structures alone.
- To our knowledge, this is the first study to systematically investigate shape readout mechanisms in DNMT3 enzymes. We quantified DNA backbone deformation (via P-RMSD), electrostatic interaction energies, and minor groove geometry across different

flanking sequences. These analyses revealed distinct strategies employed by DNMT3A and DNMT3B to recognize DNA shape—an aspect not addressed in the cited works.

- Additionally, we analyzed residue-level interactions, including shape-dependent hydrogen bonds and salt bridge networks, shedding light on the structural basis of enzyme flexibility and anchoring mechanisms that underlie differential methylation behavior.

At other places, the description of previous work is inaccurate. CCC/T and CGG are not attributed as “cognate” methylation sites, but just preferences were reported.

Regarding terminology, we appreciate the reviewer’s concern that the term “cognate” may imply exclusivity. In response, we have replaced “cognate” with “preferred” where appropriate. However, we retained its use in select instances, only for DNMT3A–CGC and DNMT3B–CGG sequences, to reflect their natural correspondence and to maintain consistency and readability.

It should be mentioned and discussed that further preferences were observed (beyond the CGX position studied here). Hence important questions that are ideally targeted by MD simulations were not addressed.

We agree that DNMT3 enzymes exhibit preferences that extend beyond the +2 base, and that comprehensive sequence context (e.g., NNCGNN) plays a role in vivo. Our decision to focus on CGX motifs was a deliberate one, aimed at resolving +2 base effects in high resolution. We now emphasize in the Conclusion that our CDA framework is modular and extensible and can be applied in future studies to longer flanking contexts.

6) The authors provide a mechanism, in which R836 or K777 hydrogen bonding is responsible for the substrate specificity. However, in the crystal structure PDB 6F57, which was used as a template for this study, the interaction of R836 with O6 and N7 of G+2’ can already be seen (as mentioned in the text). The same goes for the interaction of K777 with O6 of G+2 in DNMT3B, which can be observed in PDB 6KDA. The MD simulation experiments only confirm these findings. The authors should provide new findings for DNA substrates without crystal structure (e.g., interactions with DNA substrates that have not been crystalized). Likewise, providing movies of simulations showing novel interactions like DNMT3A with as GCA/T/C substrates and likewise for DNMT3B would be interesting.

We thank the reviewer for this helpful comment and for the opportunity to clarify the contributions of our study beyond the insights available from crystallographic structures such as PDB 6F57 (DNMT3A) and PDB 6KDA (DNMT3B). We view the fact that our minimal models recapitulate key interactions observed experimentally as a strength and a proof of concept, validating our simulation setup and the interpretability of our results. However, our findings extend beyond confirmation of crystal structures in several important ways:

- As explained above, while crystal structures like PDB 6F57 position Arg836 near G+1, they do not resolve or describe the dominant, stable bidentate hydrogen bond that we observe between Arg836 and the O6 and N7 atoms of G+2′ in the CGC context. This

interaction, which we quantify as highly frequent and stable across multi-microsecond simulations, is a new insight into how DNMT3A stabilizes its preferred substrate. Similarly, in DNMT3B, while Lys777 is seen interacting with G+2 in 6KDA, the cooperative role of Asn779, particularly in CGT and CGA contexts, is not resolved in crystal structures and is characterized here for the first time at high temporal and atomic resolution. These findings are summarized in our new Figures 4 and 5, and supported by Figure 1, which maps the structural basis used to build the models.

- We emphasized that two of the CGX motifs we analyzed (CGC for DNMT3B; CGG for DNMT3A) do not have available crystallographic complexes. Our simulations provide the first mechanistic descriptions of how DNMT3A and DNMT3B engage these sequences, revealing sequence-dependent switching of base-readout modes, engagement of secondary guanine readers, and rewiring of hydrogen bond networks that are not inferable from known structural data.
- Beyond base-specific hydrogen bonding, our study provides the first systematic analysis of shape readout in DNMT3 enzymes. By quantifying DNA backbone deformation, minor groove width, and electrostatic complementarity, we show how DNMT3A and DNMT3B differentially exploit DNA shape across CGX motifs. These analyses uncover a second layer of specificity, complementing base readout, that has not been addressed in prior structural or biochemical studies.

Finally, we agree that movies can enhance the accessibility of dynamic molecular interactions. However, for the systems involving DNMT3A–CGT, DNMT3A–CGA, DNMT3A–CGG, DNMT3B–CGT, DNMT3B–CGA, and DNMT3B–CGC we observed multiple switching interaction modes over the course of the simulation. Given the complex and state-dependent nature of these transitions, we concluded that a movie would not adequately capture the full extent of these dynamics and might even risk oversimplifying the interpretation. Instead, to ensure full transparency and reproducibility, we have made the complete MD trajectories and coordinate files for these systems publicly available via Zenodo (doi: 10.5281/zenodo.12545440). This allows interested readers and reviewers to explore the interaction dynamics in detail, using their preferred analysis tools and at their own pace. The Zenodo link is also provided in the revised manuscript.

7) Unfortunately, the entire simulation focuses on binding strength, because the particular conformations leading to transition states, which are intimately connected with catalysis are ignored. Of note, it is entirely unclear, if the contacts identified by the authors stabilize a conformation that is supportive for catalysis or if they hinder catalysis by stabilizing an off-pathway conformation. In this context, it is not clear why IDF also discarded interactions observed for less than 10% of the simulation time, considering them to be unimportant to the downstream analysis. These interactions could lead to a specific transition state, hence they could be of eminent importance.

We fully agree that transition-state stabilization is a central feature of enzyme catalysis, and that understanding whether specific contacts support or inhibit catalysis would require targeted investigation. However, we would like to highlight once more that the primary objective of our study was not to model catalysis or transition-state formation, but rather to characterize how DNA

sequence context, specifically variation at the +2 position, affects the recognition phase preceding catalysis.

Regarding the 10% interaction frequency threshold, we applied this cutoff to filter out highly transient or noise-level contacts that are not consistently sampled across the 2 μs trajectories per system. Our goal was to highlight robust, reproducible patterns of interaction that reflect the steady-state binding preferences of DNMT3A and DNMT3B across CGX motifs. We acknowledge the reviewer's valid point that rare or low-frequency contacts could contribute to transition-state stabilization. However, such phenomena would need to be explored through a distinct methodological framework, such as reaction-pathway analysis, enhanced sampling, or hybrid QM/MM simulations, which are beyond the scope of this current study.

In response to internal review and development, we have also transitioned from our earlier Interaction Dynamics Framework (IDF) to the current Comparative Dynamics Analysis (CDA). CDA provides a more comprehensive analysis of recognition features by integrating shape-dependent and base-specific readout, offering a systems-level view of how sequence preferences are encoded in the DNMT–DNA interface. While it does not probe transition states directly, CDA helps resolve how sequence influences the structural preconditions necessary for catalysis, namely, stable binding, proper alignment, and interface adaptability.

8) An additional analysis could strengthen the approach of this study: Since biochemical methylation activity experiments for DNMT3A and DNMT3B have been done, one could try to correlate in vitro data with e.g., the number of hydrogens bonds found in the various complexes that were simulated. A connection between the activity and hydrogen bonds could elevate this study (e.g., does the number of hydrogen bonds for CGA match the methylation activity?) and help to decide if an observed interaction has a positive or negative role in catalysis.

Indeed, correlating simulation-derived interaction metrics with experimental methylation activity data offers a valuable means of assessing the biological relevance of our findings. To address this, we compared our calculated hydrogen bond interaction intensities across CGX contexts (as presented in Figures 4 and 5) with in vivo methylation activity data reported in Mallona et al. (2021), which includes CpG methylation scores derived from whole-genome bisulfite sequencing (WGBS).

Notably, in our experiments, DNMT3A exhibited both a broader and higher range of hydrogen bond intensities compared to DNMT3B, consistent with its higher sequence specificity and methylation activity reported in both Mallona *et al.* and prior deep enzymology studies (e.g., Gao et al.). In particular, the maximum hydrogen bond intensity observed in the DNMT3A–CGC complex (1.9) was approximately threefold higher than that observed in the most stable DNMT3B interaction (0.66). We have now added this comparative analysis to the Results and Discussion section of the manuscript, and we are grateful to the reviewer for this suggestion, which helped us strengthen the functional interpretation of our findings.

Figure Redacted

Taken from Mallona et al., 2021

Minor Comments

It would be helpful for readers if the authors use the numbering scheme established in the field, in which the CGX position is considered the first outside of the CpG site, hence designated as +1 (and not +2).

We thank the reviewer for raising this point. We acknowledge that multiple numbering conventions exist in the literature, and some studies indeed designate the base immediately following the CpG site as +1. However, our numbering scheme follows the nucleotide-centered convention, where the cytosine of the CpG site is designated as position 0, consistent with prior studies, including those by Mao *et al.* and Mallona *et al.*, which have used this same framework.

The TIP4P water model could be used instead of TIP3P as TIP4P more accurately represents a correct charge distribution. At least comparison runs should be done, showing no effect of the water model.

We appreciate the reviewer's suggestion regarding the use of the TIP4P water model. While TIP4P offers improved electrostatic properties for pure water simulations, our study focuses on protein–DNA interactions, where the TIP3P water model is commonly employed due to its compatibility with standard force fields and its computational efficiency. Our simulations, conducted using TIP3P, have yielded results consistent with experimental observations, supporting the validity of our approach.

We acknowledge that different water models can influence simulation outcomes, and we agree that exploring the effects of alternative models like TIP4P could provide additional insights. However, given the scope of our current study and the consistency of our results with experimental data, we have chosen to continue with the TIP3P model. We will consider incorporating comparative analyses with other water models in future research to further validate and enhance our findings.

Typo in line 132: DNMT3A-CGC ‘was’ modified.

We have a new text now. So this should have been addressed in the new version.

Table 1 and 2 could be more organized and clearer.

We have a whole new Table 1 now. We also cited Table 2 and Table 3 (interaction statistics) clearly in the next for not causing any confusion.

Figure 1A: In the sequence alignment, marking D and N as yellow and therefore similar physiochemical properties seems not to be correct.

The reviewer is right. We updated this Figure accordingly.

Numbering in the provided starting structure PDB files does not match the manuscript.

We thank the reviewer for this helpful observation and for carefully checking the residue numbering. At the beginning of the study, we reset residue numbering to start from 1 for both DNMT3A and DNMT3B constructs to allow for direct, position-by-position sequence-based comparisons between the two enzymes during model building and analysis. For clarity and consistency with external references and UniProt annotations, we later adjusted the residue numbering in the manuscript by adding 629 for DNMT3A and 570 for DNMT3B, corresponding to the full-length human sequence numbering. These UniProt-based residue numbers are used consistently throughout the manuscript, figures, and analysis. We have now clarified this numbering convention in the Methods section and provided a note on our GitHub repository accordingly.

For base-specific hydrogen bonds, interactions of backbone atoms of amino acids and nucleotides were also filtered out from the interaction data frame. However, e.g, K777 could interact with DC`300/OP2 and R836 with DG`297/OP2. This information could be important.

We thank the reviewer for this thoughtful comment. In this study, as outlined, we focused on base-specific hydrogen bonds to capture interactions that directly reflect sequence-dependent recognition. To that end, we excluded contacts involving backbone atoms of amino acids and nucleotides, as these are generally non-specific and structurally conserved across DNA–protein complexes. We now clarify this filtering criterion in the first section of the CDA framework.

Reviewer #2:

This is an interesting study of the interactions of two classes DNA methyltransferases with four different DNA sequences in which the methylation target CpG is followed by one of A,C,G or T. From simulations of these systems the authors identify particular interactions that explain the sequence preferences of the two enzymes. For each complex 4 independent simulations were run, and it would be interesting to know how consistent the results are between the replicates.

We thank the reviewer for this supportive comment and for emphasizing the importance of replicate consistency. To assess the reproducibility of our findings, we analyzed the four independent simulation runs performed for each DNMT–DNA complex, focusing on base-specific hydrogen bonding patterns. As shown in Supplementary Figures 1 and 2, several key interactions, particularly those involving high-frequency, base-specific hydrogen bonds, were consistently reproduced across replicates, demonstrating stable and sequence-dependent recognition behavior.

However, not all interactions were identically observed in every run. Lower-frequency or transient contacts showed some variability, which likely reflects the inherent conformational flexibility and multi-state nature of the protein–DNA interface. Given this variability, we concluded that pooling data across replicates would provide the most comprehensive and statistically robust view of the interaction landscape. This strategy allowed us to emphasize consistent, functionally relevant features while minimizing noise arising from rare or simulation-specific events. We have included this clarification in the Conclusions section of the revised manuscript.

Some points that need to be addressed.
Line 83. Reference for Mallona et al is missing

Included.

Line 97. What is an Interaction Dynamics Framework?

The Interaction Dynamics Framework (IDF) was our initial approach for analyzing base-specific hydrogen bonding in protein–DNA complexes. In response to other reviewers' feedback, we have now expanded and renamed this approach as Comparative Dynamics Analysis (CDA). The CDA combines base readout, defined as hydrogen bonding between protein side chains and DNA bases in the major groove, with shape readout, which includes DNA groove geometry and electrostatic potential. This integrated framework allows us to assess how both readout mechanisms contribute to sequence preference. We now comprehensively explain CDA in the main text (Figure 3) and provide additional details in the Methods and Results and Discussion sections.

Line 112. Coordinate system? It looks like a numbering scheme.

The reviewer is right. We changed the "coordinate system" to "Target DNA site".
Line 147. Unit is missing for the positional restraints.

The reviewer is right. The unit of positional restraints is kJ mol^{-1} nm^{-2}. This was added in the Methods.

Line 171. Why were the interaction between protein and methylated cytosine removed?

To isolate sequence-dependent effects and minimize confounding by conserved catalytic features, we excluded interactions involving the SAM cofactor and the flipped cytosine, which are

structurally invariant across complexes. This explanation is now included in the Results and Discussion section.

Line 172. Why were protein backbone atoms not included in the analysis of base-specifc contacts?

We thank the reviewer for this thoughtful comment. In this study, we focused on base-specific hydrogen bonds to capture interactions that directly reflect sequence-dependent recognition. To that end, we excluded contacts involving backbone atoms of amino acids and nucleotides, as these are generally non-specific and structurally conserved across DNA–protein complexes. We now clarify this filtering criterion in the first section of the CDA framework.

Line 194. The "interaction densities" seem to simply be the average number of hydrogen bonds, and could be described as such to make the text easier to read. Line 236. Same as above, this is the average number of hydrogen bonds.

We thank the reviewer for the suggestion. In this version of the manuscript, we use “interaction intensity” as a unified metric to quantify normalized frequencies of noncovalent contacts. To ensure consistency and avoid introducing redundant terminology, we apply this term across all interaction types. This definition is now clearly stated in the Methods section.

Lines 481-482. References missing

Added.

Reviewer #3:

Despite high levels of homology (91%) the two de novo methyltransferases (DNMT3A and DNMT3B) are able to selectively methylate CG steps that have different flanking sequences. Barlas and Karaka perform molecular dynamics simulations on different flanking sequences around a central CG to unveil the selectivity of different methyltrasferases. Key aminoacid differences Arg/Lys are found to be responsible for the different interaction patterns. They then identify differences in interaction patterns (e.g. hydrogen bonds) to conclude that direct interactions are responsible for selectivity. Understanding the role of direct/inderect (base/shape) readout in protein nucleic acids has been a longstanding goal in the field. Indirect or shape readout relates to the ability of a DNA sequence to adopt conformations found in the bound state, where as the direct or base readout relates to the specfic interactions made in a complex. Understanding the key differences amongst a series of sequences requires to know what the binding free energy is for each one, and how it is distributed amongst direct and indirect readout: DG_bind = DG_base + DG_shape. And this should be done for different sequences.

We thank the reviewer for this very insightful comment. In response, we have added a dedicated section on shape readout, where we systematically analyze DNA backbone deformation (via P-RMSD), minor groove geometry, and interfacial electrostatics. These features serve as structural correlates of ΔG_shape, revealing distinct shape readout strategies adopted by DNMT3A and DNMT3B across CGX contexts. Rather than directly computing ΔG_bind or its components, which would require enhanced sampling or QM/MM techniques, we pursued a different and more accessible strategy: using interface-based and shape-based metrics to quantify sequence-specific recognition. This direct approach provides interpretable structural observables and can be readily applied by researchers without specialized expertise in free energy methods. We have previously demonstrated its utility in related systems (e.g., Karakulak et al., https://doi.org/10.3389/fmolb.2021.658906; Ozden et al., https://doi.org/10.1021/acs.jcim.2c01520), and have now formalized it into the Comparative Dynamics Analysis (CDA) framework introduced in this study.

As highlighted in the Conclusion, CDA integrates base and shape readout to explain recognition specificity using standard molecular dynamics simulations. We fully agree that future integration with rigorous free energy decomposition (ΔG_bind = ΔG_base + ΔG_shape) would be highly valuable for further deepening the thermodynamic understanding of these interactions. We are grateful to the reviewer for raising this point, which directly contributed to expanding the conceptual scope and practical relevance of our work.

However, in this study, we just see the results of very detailed analysis of interaction patterns amongst residues in the DNA and protein. This makes the article very technical and hard to read, but also provides little insight about the difference in relative importance of each readout upon changing the sequence. Is this not done because for this system its already known that direct readout is key -- e.g., from the experimental structures, and if so, what new knowledge does this work generate?

It is true that prior structural and biochemical studies have established the significance of direct (base) readout, particularly via Arg836 in DNMT3A and Lys777/Asn779 in DNMT3B, in mediating flanking sequence preferences. These findings formed the structural basis and motivation for our work, and much of the early literature indeed focused on base readout as the primary driver of specificity. However, thanks to this comment, we re-examined the system from a broader perspective and systematically quantified shape readout across all CGX sequence contexts. We computed DNA deformation metrics (P-RMSD), minor groove geometry, and interfacial electrostatics, and found that shape readout features differ substantially between sequences. These findings add a new dimension to our understanding of DNMT3 recognition strategies. In response, we added a new section that discusses the complementarity of base and shape readout mechanisms, showing that DNMT3A exhibits tighter shape coupling and electrostatic anchoring, while DNMT3B maintains flexibility across varied sequences. This is, to our knowledge, the first study to systematically examine shape readout in DNMT3 enzymes, and it expands the known recognition model beyond what is captured in static structures.

Furthermore, our MD simulations reveal dynamic interaction patterns, such as a bidentate hydrogen bond between Arg836 and G+2′ in DNMT3A–CGC, which are not resolved in crystal structures but are dominant in the simulations. These time-resolved features offer a richer and more nuanced understanding of how specificity is achieved. To improve accessibility, we also revised the manuscript to reduce technical jargon where possible and added additional discussion to better highlight the conceptual and methodological advances.

Reading in the methods about the unbound DNA simulations i thought we were going to see how sequence effects changed the conformational preferences from unbound to bound DNA. And yet, i did not see any mention of the unbound states after methods.

The reviewer is right. In the revised manuscript, we commented on free DNMT3 simulations at the last part of Results and Discussion:

“RMSF analysis further revealed that DNMT3A’s catalytic loop is inherently more rigid in the unbound state and becomes further stabilized upon DNA binding, consistent with a pre-organized, site-specific recognition mechanism. In contrast, DNMT3B’s catalytic loop (648–662) remains comparatively dynamic even in the bound state, supporting a scanning, processive engagement strategy (Supplementary Figure S6-7). Notably, the TRD loop is more flexible in DNMT3A than in DNMT3B, indicating that while DNMT3A depends on rigid catalytic loop anchoring, it may use TRD flexibility to fine-tune sequence discrimination (Supplementary Figure S8).”

"We also showed that this direct readout mechanism is dictated by the DNA groove geometry". -- Wouldn't this suggest that an indirect readout is important?

We thank the reviewer for this insightful comment. The reviewer is absolutely right. Our previous phrasing did not clearly distinguish between the mechanism of interaction (direct vs. indirect readout) and the underlying structural features influencing it, which obviously have caused confusion. To clarify this in the revised manuscript, we have developed and articulated a comprehensive framework that separates and analyzes both readout modes:

- Base readout (direct): Captures explicit, base-specific hydrogen bonding between protein side chains and DNA bases, primarily in the major groove.
- Shape readout (indirect): Quantifies sequence-dependent DNA deformation (P-RMSD), groove geometry, electrostatics, and hydrogen bonding or hydrophobic contacts with the minor groove.

This framework explicitly highlights that direct readout is influenced and constrained by shape readout, and that both mechanisms synergistically determine DNMT3 specificity. To reflect this more clearly, we have also added a dedicated section in the revised manuscript titled “Distinct Methylation Strategies Are Grounded in Electrostatics and Flexibility”, where we directly compare the two mechanisms and explore their sequence-dependent contributions. We thank the reviewer again for helping us sharpen this conceptual distinction in the manuscript.

When thinking about flanking regions: why is the effect only looked at the 3' end of the CG, and why is the perturbation only up to two residues? In general, local effects can extend to the hexamer and 8-mer level.

We thank the reviewer for this to-the-point comment. Our initial focus on the +2 position at the 3′ end was based on strong structural and biochemical evidence indicating that this position plays a dominant role in DNMT3A and DNMT3B sequence preferences (e.g., Gao et al., 2020; Anteneh et al., 2020). However, in response to this helpful comment, we extended our analysis to include upstream (5′) flanking positions—specifically at –1, –2, and –3. In this region, we evaluated hydrogen bonds and hydrophobic contacts formed between DNA bases and residues within both the catalytic and TRD loops of DNMT3A and DNMT3B. Our results revealed that residues such as Arg720, Val716, Ile715, and Lys844 in DNMT3A, and their DNMT3B counterparts Arg661, Val657, and Asn656, form minor groove interactions with the 5′ flank. These interactions contribute to DNA shape deformation and electrostatic complementarity and vary in a sequence-dependent manner. These findings are presented in Figures 6–7 and Supplementary Figures X–Y. We also have clarified this point in the revised Conclusions section:

> *“We intentionally excluded water-mediated contacts and non-CpG methylation contexts, as the complexity of the core CpG systems alone demanded a focused, high-resolution analysis. However, recent experimental studies suggest that scanning selective and differential interactions across a broader flanking nucleotide range (e.g., positions -3 to +3) is essential for a comprehensive understanding of DNMT3 selectivity. Such analyses could also be extended to non-CpG DNA sequences. Additionally, the proposed allosteric influence introduced by DNMT3L should be considered when characterizing all factors contributing to selective de novo DNA methylation”*

I think a closer look at the work from Richard Lavery and from Remo Rohs at attempts to separate between direct and indirect readout for relating to differences in binding affinity would benefit this work.

We thank the reviewer for this great suggestion. We fully agree that the pioneering work of Richard Lavery and Remo Rohs on separating base and shape readout mechanisms in protein–DNA recognition has been foundational in the field, and we appreciate the opportunity to integrate their contributions into our study. In this revised version:

- Inspired by Lavery's energetic decomposition strategies, we applied residue-level analyses of hydrogen bonding and side-chain contacts alongside DNA deformation metrics to systematically separate base and shape readout contributions in a sequence-dependent context.

- Building on the work of the Rohs group, we incorporated quantitative measures of minor groove width, electrostatic complementarity, and backbone deformation, which are critical indicators of DNA shape recognition. By correlating these structural features with interfacial electrostatics across different CGX motifs, we introduce a dynamics-informed approach to estimate the role of shape readout, as captured over long-timescale MD simulations.

Together, these additions strengthen the conceptual foundation of our study and link our Comparative Dynamics Analysis (CDA) framework more directly to established principles in the literature. We thank the reviewer again for encouraging us to more clearly connect our work with these important prior contributions.

And i think the authors should consider distilling their message. Currently, the technical details and how the paper is written make it very difficult to find the relevance in the work.

We did our best to fix this problem and we are open for further critics, if still needed.

Technical:

Remo Rohs introduced the terms base and shape readout as more representative of protein nucleic acids than the previous direct/indirect readout mechanisms.

We changed the direct/indirect terminology to the base/shape readout mechanisms.

For future works, the Berendsen barostat is no longer recommended as it does not yield correct ensembles.

Noted. Thank you for the recommendation.