# A Comprehensive *In Silico* Perspective For Discovery of Novel Inhibitory Candidates Targeting Versatile Transcriptional Repressor MBD2

**Zihni Onur Çalışkaner** ( zihnionur@gmail.com )
Üsküdar University

**Research Article**

**Keywords:** DNA methylation, epigenetics, cancer, iPSC reprogramming, MD simulation

**Posted Date:** June 9th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-578382/v1

# Abstract

DNA methylation is a key epigenetic mechanism in various biological events such as development, cellular differentiation, cancer progression, aging, and iPSC reprogramming. Crosstalk between DNA methylation and regulation in gene expression is employed through MBD2, known as reader of DNA methylation and suggested as a drug target. Despite its magnitude of significance and rationale of nomination, scarcely limited number of druggable ligands has been detected so far. Hence, we screened a comprehensive compound library and then certain of them were subjected to computational docking analysis by targeting methylated DNA-binding domain of human MBD2. We could detect reasonable binding energies and docking residues presumably located in druggable pockets. Docking results were also validated via MD simulation and per-residue energy decomposition calculation. Drug-likeness of tested ligands was assessed through ADMET prediction in order to foresee off-target side effects for future studies. Herein, on the basis of collaborating approaches such as molecular docking, MD simulation, energy decomposition and ADMET prediction, notably two compounds named CID3100583 and 8,8-Ethylenebistheophylline, have become prominent as novel candidates, possibly disrupting $MBD2_{MBD}$–DNA interaction. Hereby, these compounds exhibit a promising usage potential in a wide range of implementation from cancer treatment to somatic cell reprogramming protocols.


# Introduction

Epigenetic regulation is a key process in which involves various mechanisms, dynamically altering and specifying gene expression and cell fate [1]. As a predominant epigenetic mechanism, DNA methylation involves covalent attachment of methyl groups ($-CH_3$) from S-adenosyl-L-methionine as the methyl donor to the carbon-5 positions (C5) of the cytosine residues within a context of cytosine-guanine (CpG) dinucleotides resulting in 5-methylcytosine (5-mC) analogue [2]. Principally, three catalytically active DNA Methyltransferases in mammals, DNMT1, DNMT3A and DNMT3B, maintain an overall DNA methylation pattern specific to either cell type, developmental period and differentiation status [3,4]. Methylation on CpG dinucleotides near and/or inside promoter sequences reveals a well-known hallmark for DNA-binding proteins, and thus gene transcription is repressed due to avoidance of transcription factors and machinery [5,6]. Besides its importance in normal development and cellular function [7], methylation status of related genes also constitutes a significant diagnostic marker, causation and therapeutic target in a variety of human diseases such as cancer [8], birth defects and syndromes [9], metabolic disorders [10], autoimmune diseases [11], muscular dystrophies [12], cardiovascular diseases [13], aging [14], neurological dysfunctions [15].

Correspondingly, miscellaneous epigenetic/genetic arrangements, mostly deacetylation of histones along related regions by Histone Deacetylases (HDACs), are accompanied and coordinated with DNA methylation in order to furtherly cement electrostatic interactions between histone core and DNA strands, resulting in compacted chromatin, which downregulates transcription [16,17]. Herein, Methyl-CpG Binding Domain (MBD) family members, also known as “readers of DNA methylation” [5], are outstanding

transcriptional repressors, which facilitate crosstalk between DNA methylation and histone deacetylation [18]. All members of MBD-containing protein family in mammals share a highly conserved methylated DNA-binding domain, consists of 70–85 amino acids [5,19].

As one of the family member, human *MBD2* gene encodes three isoforms, namely MBD2a, MBD2b and MBD2c (or MBD2t) with 411 amino acids (~ 43 kDa), 262 amino acids (~ 29 kDa), and 301 amino acids (~ 32 kDa), respectively [20–23]. MBD2a is a full-length isoform including all of four distinctive domains within the protein. An N-terminal arginine-glycine (G/R) repeat region has been shown to be related with post-translational modifications such as methylation which may alter the interaction with its binding protein partners [24–26]. An approximately 80-amino acid MBD corresponding to residues between 145–225, exists in all isoforms and binds 5-mC nucleotides in CpG islands [19,23,27]. MBD is followed by a C-terminal Transcription Repression Domain (TRD) which is crucial for recruitment of essential co-repressor complex such as Mi-2/NuRD (Nucleosome Remodeling and Deacetylase) complex onto 5-mC hallmarks for expressional silencing [18,20,28,29]. Another C-terminal domain, coiled-coil (CC), corresponds to amino acids 361–393 and its ability has been revealed to critically bind p66α for globin gene silencing [21,29,30]. Besides, it has been found that CC domain is present in MBD2, MBD3 and MBD3L2, predicting a potential heterodimerization of these closely related proteins via this domain [21,29,31]. MBD2b isoform arises from an alternative translation start codon by excluding the first 149 amino acids and therefore it lacks N-terminal RG-rich domain [23,32]. The third isoform, MBD2c or MBD2t, encodes a truncated variant as a consequence of alternatively spliced region spanning between exon 2–3, in which C-terminal TRD and CC domain is lacking [32–34]. In fact, MBD2a is the most predominant variant that ubiquitously expressed in several tissues while expression of MBD2c, the second prevalent isoform, is especially restricted to testes, embryonic stem cells (ESCs) and human pluripotent stem cells (PSCs), suggesting a functional discrepancy in accordance with significant structural (domain) and expressional inequality [18,23,35]. Along with isotype-specific expression in a various tissues, MBD2–mediated epigenetic regulation has been implicated in important biological functions such as neuronal activity and development, proliferation and maturation of various immune cells, hematopoiesis, pluripotency, and cancer initiation and progression [32,36,37]. Therefore, MBD2 stands out as potential targets in especially treatment of cancer [37] and regulation of pluripotency-differentiation balance in iPSC reprogramming [34].

Although structure of MBD2 and its function in several biological processes and disease progression have been well-characterized suggesting as key targets, number of available inhibitors against this epigenetic regulator is still barely limited. Consequently, we accomplished a comprehensive screening to reveal small molecules as potential inhibitors in this study. Our results have pointed these small molecules prominently docked significant residues such as Val177, Arg188, Lys190, Phe206, Phe208, and Arg209 in DNA–binding domain, predicting an interrupting impact on MBD2-methylated DNA recognition and interaction. Furthermore, Molecular Dynamics (MD) simulation validated $MBD2_{MBD}$–ligand interaction by meaningful RMSD, RMSF, B-Factor values, and energetics. Besides, it's been noticed that predicted 3-dimensional binding pockets on $MBD2_{MBD}$ and residues in which ligands have occupied, were

considerably overlapped. Taken together, extracted small molecules have been emerged as potential inhibitor/drug candidates with pretty good affinity and inhibition constants as shown by computationally. Hereby, current study has provided a substantial data, particularly for compounds CID3100583 and 8,8-Ethylenebistheophylline which might lead to *in vitro* and/or *in vivo* studies testing its applicability in cancer treatment and somatic reprogramming protocols.

# Results

## Docking Based Virtual Screening To Elicit Potential Binders

A screening library was prepared from ZINC15 database compounds. We regarded to certain criteria such as logP value, MW, and charge for selection of molecule tranches. Wyhs et al. had demonstrated Mitoxantrone, Idarubicin, NF449 and Aurintricarboxylic Acid impaired $MBD2_{MBD}$ binding to methylated DNA in TR-FRET assay however mechanism of action remained unclear [38]. Since these compounds predominantly have charges, logP and MW ranging from − 3 to + 2, 0 to 3, 400 to 500 Da, respectively, we exploited related characteristics emerging as screening strategy to start out. Besides, logP, MW and charge were retained moderate in order to explore ligands with favorable solubility in aqueous vehicles, convenience for passing through plasma membrane and intracellular delivery for potential use in prospective in vitro/in vivo studies. 15,893 different compounds yielded numerous docking poses including 120,310 available conformations with reasonable binding affinities. Among these, certain of hit compounds with docking score over − 5.0 kcal/mol which was pointed according to structure and/or biological activity (if available), were listed in Supplementary Table S1. Although most of the compounds have not been annotated, assayed and/or functionally identified, already identified/annotated hits were also recovered by the virtue of screening. Most of the hits have been attained to inhold heterocyclic groups, especially carboxamide, quinoxaline, imide, benzene, thiazole, piperazine, naphthalene, pyridazine, isoquinoline, pyrrole, pyrimidine/pyridine, furan, aniline, imidazole (see Supplementary Fig. S1). Moreover, ligands composed of xanthine, guanine, adenine or their derivatives and/or methylated groups, were also quite notable in terms of implying a plausible compatibility between potential binders and methylated-DNA binding domain.

## Elaboration of Target Sites of Selected Leading Compounds Throughout $MBD2_{MBD}$

Existence of significant structures and groups within hit compounds (see Supplementary Fig. S1) led us to determine binding regions in more comprehensively by computational docking analysis. Thereby, selected ligands with higher screening scores were docked against refined receptor corresponding to human $MBD2_{MBD}$ (PDB: 7AO8, Chain C). Among 10 of unidentified and 5 of annotated compounds, CID3100583 exhibited the best docking pose with − 11.42 kcal/mol, 275.20 Å and 4.22 nM as binding energy, RMSD and estimated inhibition constant ($K_i$), respectively (Table 1). It was followed by CID343482, CID136748749 as proximately. Nevertheless, estimated $K_i$ of CID3100583 (4.22 nM) was well

ahead closest candidates as roundly 8-fold (Table 1), suggesting less amount could be sufficient for a half-maximal inhibition on $MBD2_{MBD}$–DNA binding in actual fact. Correspondingly, each of CID3100583 and CID343482 constituted 9 H-bonds (Table 2) in company with 2 hydrophobic interactions and 3 hydrophobic interactions, respectively (Table 3). Albeit both ligands established the same number of H-bonds, more diverse amino acids contributed interaction with CID3100583 atoms.

Table 1
Calculated binding energies, RMSD values and inhibition constants ($K_i$) of the best enzyme-drug docking poses.

| PubChem ID or Annotation | Binding Energy (kcal/mol) | RMSD (A°) | Estimated $K_i$ (nM) |
|---|---|---|---|
| CID3100583 | −11.42 | 275.20 | 4.22 |
| CID343482 | −10.14 | 277.80 | 37.09 |
| CID136748749 | −10.01 | 274.46 | 45.59 |
| CID46959745 | −9.88 | 281.55 | 57.48 |
| CID126001284 | −9.67 | 277.00 | 80.96 |
| CID125615373 | −9.41 | 274.33 | 127.17 |
| CID28689118 | −9.26 | 274.75 | 162.21 |
| CID25826758 | −9.20 | 279.74 | 180.51 |
| CID323153 | −8.34 | 276.53 | 766.71 |
| CID24241 | −8.29 | 277.420 | 833.56 |
| 8,8'-Ethylenebis-Theophylline | −10.32 | 278.82 | 27.30 |
| Regadenoson | −9.83 | 285.73 | 62.67 |
| Sanguinarine | −8.76 | 277.98 | 308.44 |
| GHMCytosine | −8.18 | 277.95 | 1001.00 |
| Golgicide A | −7.81 | 278.05 | 1880.00 |

Table 2
Number of hydrogen bonds (H-bond) and the list of interacted amino acids with inhibitor (Certain residues have been colored in association with druggable pockets).

| *Compound* | *Inhibition Constant* | *Number of H-bonds* | *Amino acids involved in interaction* |
|---|---|---|---|
| CID3100583 | 4.22 nM | 9 | **Val177**, **Arg188**, Lys190, Ser204, **Phe206**, Asp207, **Phe208**, **Arg209** |
| CID343482 | 37.04 nM | 9 | **Val177**, Lys190, **Phe206**, Asp207, **Phe208**, **Arg209** |
| CID136748749 | 45.59 nM | 6 | Lys190, **Phe206**, **Phe208**, **Arg209** |
| CID46959745 | 57.48 nM | 5 | **Asp151**, **Cys152**, **Leu155**, Gly158 |
| CID126001284 | 80.96 nM | 2 | **Ser205**, **Phe206** |
| CID125615373 | 127.17 nM | 1 | **Ala154** |
| CID28689118 | 162.21 nM | 2 | **Asp176**, **Phe208** |
| CID25826758 | 180.51 nM | 5 | **Val177**, **Arg188**, Lys190, **Arg209** |
| CID323153 | 766.71 nM | 2 | **Arg209** |
| CID24241 | 833.56 nM | 4 | **Phe208**, **Arg209** |
| 8,8'-Ethylenebis-Theophylline | 27.30 nM | 5 | **Val177**, **Phe208**, **Arg209** |
| Regadenoson | 62.67 nM | 9 | Trp159, Ser183, Tyr196, Thr200 |
| Sanguinarine | 308.44 nM | – | – |
| GHMCytosine | 1.01 μM | 9 | **Val177**, **Arg188**, Lys190, **Phe206**, **Phe208**, **Arg209** |
| Golgicide A | 1.88 μM | 2 | **Phe208**, **Arg209** |

Table 3
Number of hydrophobic and other interactions and the list of interacted amino acids with inhibitor (Certain residues have been colored in association with druggable pockets).

| *Compound* | *# of Hydrophobic Interactions* | *# of π-Stacking/Cation* | *# of Salt Bridges* | *# of Halogen Bonds* | *Interacting Amino Acids* |
|---|---|---|---|---|---|
| CID3100583 | 2 | – | – | – | **Phe208**, **Arg209** |
| CID343482 | – | 2 | 1 | – | Asp207, **Phe208** |
| CID136748749 | 2 | – | – | – | **Asp176**, **Arg209** |
| CID46959745 | 6 | – | – | – | **Leu155**, **Pro156**, Pro157, **Val201**, **Leu203** |
| CID126001284 | 6 | – | – | – | Leu155, **Pro156**, **Val201**, **Leu203** |
| CID125615373 | 5 | – | – | – | **Ala154**, **Val201**, **Leu203**, **Phe206**, **Lys212** |
| CID28689118 | 4 | 2 | – | – | Lys190, Asp207, **Phe208** |
| CID25826758 | 4 | – | – | – | Lys190, **Phe208** |
| CID323153 | 6 | – | – | – | **Val177**, Lys190, Asp207, **Phe208** |
| CID24241 | 6 | – | – | – | **Asp176**, Lys190, **Phe208** |
| Theophylline, 8,8'-Ethylenebis- | 3 | – | – | – | **Val177**, **Phe208**, **Arg209** |
| Regadeneson | – | 1 | – | – | Trp159 |
| Sanguinarine | 4 | – | – | – | **Val177**, **Phe208** |
| GHMCytosine | 1 | – | – | – | **Arg209** |
| Golgicide A | 3 | 1 | – | 1 | **Asp176**, **Val177**, Lys190, **Arg209** |

8,8'-Ethylenebistheophylline (Ebis-Theophylline) committed the strongest binding energy (-10.32 kcal/mol) with the lowest $K_i$ (27.30 nM) and the second rank among annotated drugs and all compounds, respectively (Table 1). Although Regadenoson and Glucosylhydroxymethylcytosine (GHMCytosine) formed 9 H-bonds (Table 2), their binding affinities and estimated $K_i$ were relatively lower and besides less number of divergent amino acids participated in binding sites in a similar way with others (Tables 2 and 3).

Together, it's been realized that all compounds occupied overlapping regions throughout $MBD2_{MBD}$, containing mutual residues (see Supplementary Fig. S2). Hereby, residues Asp176, Val177, Arg188, Lys190, Val201, Phe206, Asp207, Phe208 and Arg209 were the main interactional contributors (Fig. 1), commonly encountered within docking sites of ligands. Therefore, we rendered intermolecular interactions specific to $MBD2_{MBD}$–CID3100583 and $MBD2_{MBD}$–Ebis-Theophylline complexes, appeared as the most promising ones. As shown in Fig. 2, CID3100583 affiliated not only hydrogen bonds with moderately close distances but also various interactions such as van der Waals, π-anion and π-alkyl (Fig. 2a and 2c). Furthermore, it's been noticed that interacted residues were aligned around by configuring a cavity-like surface, in which hypothetically facilitates settlement of CID3100583 in an easier and trapped manner (Fig. 2b). On the other part, Ebis-Theophylline built close and distinct interactions with the $MBD2_{MBD}$ receptor (Fig. 2d-e). Five H-bonds were bridged between Val177, Phe209 and Arg209 (Fig. 2d and Table 2) while amino acids Arg166, Asp176, Tyr178, Arg188, Ser189, Lys190, Pro191, and Asp207 fell within distinct interactive relations such as van der Waals, π-anion and π-alkyl forces (Fig. 2e and Table 3).

# Determination of Potential Binding Pockets Over $MBD2_{MBD}$

Next, we sought out presumptive druggable pockets on the surface or interior of $MBD2_{MBD}$ in order to conclude whether ligand-docking sites represented any significancy in this sense. Three different pockets, P_1, P_2 and P_3, were predicted with DS 0.24, 0.18 and 0.16, respectively. Actually, each druggable pockets has contained considerable residues which also coincided with binding regions of certain ligands. Conformations and amino acid descriptors of druggable pockets have been depicted in Supplementary Fig. S3 and prominent residues in docking sites of ligands were colored based on the context of pocket in Tables 2 and 3. For instance, it has been observed that CID3100583–docking site was broadly overlapped with pocket P_3, including Val177, Arg188, Phe208 and Arg209 while Phe206 was the only amino acid resided in pocket P_2 (Supplementary Fig. S3, Tables 2 and 3). Other ligands such as CID343482, CID136748749, CID25826758, CID323153, CID24241, GHMCytosine, Golgicide A, Sanguinarine showed a comparable coincidence widely with pocket P_3 and then with pocket P_2 in minority (Tables 2 and 3). Otherwise, interacted residues with CID46959745 and CID125615373 were also detected in druggable pockets P_1 and P_2 while CID126001284 was seen to fit within pocket P_2 only. In terms of Ebis-Theophylline, pocket P_1 was predominantly appeared as potential residential district. Apart from amino acids directly connected with ligands, adjacent residues could contribute convergence with predicted druggable pockets, as well. Interestingly, there was no harmony between Regadenoson binding environment and any druggable cleft (see Supplementary Fig. S3, Tables 2 and 3).

# Crosscheck of Dynamics and Flexibility of $MBD2_{MBD}$–CID3100583 and $MBD2_{MBD}$–Ebis-Theophylline Complexes via MD Simulation

Because CID3100583 was emerged as the most promising druggable ligand in consequence of either virtual screening and docking analysis, we facilitated MD simulation in order to pursue dynamic behavior and to prove flexibility/rigidity of protein-ligand complex binding along given time. Along with CID3100583–$MBD2_{MBD}$ complex fitted as superior (Table 1), MD simulations of other ligand-protein complexes with lower docking scores were also conducted as comparison (Fig. 3a). Apparently, $MBD2_{MBD}$ conformations with ligands CID3100583, CID343482, CID136748749, CID46959745, CID126001284, CID125615373, CID25826758, CID323153, CID24241 established mean RMSD values 0.89 Å, 0.87 Å, 0.94 Å, 1.04 Å, 1.14 Å, 1.14 Å, 1.27 Å, 0.97 Å, and 0.94 Å, respectively, against the average structure along the trajectory. Likewise, CID3100583-protein complex fluctuated slightly through snapshots and followed by CID343482- and surprisingly by CID24241-protein complexes. Despite these complexes were fluctuating proximately, CID3100583 could be distinguished owing to an earlier equilibrium state and a lesser chequered progress during simulation. The RMSD fluctuation suggested that the MD trajectories except CID12561537, have been overall stable for ligand-protein structures during simulation (Fig. 3a). It was noted that both protein backbone and CID3100583 ligand in the complex have been reached equilibrium in a shorter time (in approximately 200 frames) and remained stable except for negligible wobbles in a wider time span (Fig. 3b). Afterward, RMSF and B-Factors were interpreted in order to detect atomic fluctuations and x-ray scattering caused by thermal motion in specific to per residue for $C_\alpha$ atoms. Importantly, we apparently detected sharp declines in fluctuation especially over three regions built by Val177, Arg188, Lys190, Ser204, Phe206, Asp207, Phe208, Arg209 and their adjacencies (Fig. 3c, orange line), demonstrating residues have rigidly docked to CID3100583 atoms and therefore been constrained less flexible. Correspondingly, the regions in which B-factor was lower, account for regular secondary structures (Fig. 3c, blue line). For instance, Val177 and Arg188 accommodates in strands and Lys190 is included in helix while residues Ser204–Arg209 locates in a loop (see Supplementary Fig. S4). In other respects, RoG has been moderately altering over time in parallel with RMSD chart. As plotted in Fig. 3d (blue line) during initial 200 frames, here detected an unsteadiness in which also corresponded to protein-ligand equilibrium phase (Fig. 3b). However beyond this point, RoG decreased upon probably binding with CID3100583, in concordance with equilibrium of the system (Fig. 3b and 3d). Fraction of native contacts (Qx) increases based upon more binding residues by determining transition in folding state, that is, the more the folding (lesser RoG), the higher the Q(x) is assumed [39,40]. Notably, Q(x) has been conducted verifying RoG data (Fig. 3d).

Despite the fact that Ebis-Theophylline, as the most probable repositioning drug (Table 1), showed the closest RMSD profile among the annotated drugs, Ebis-Thophylline–$MBD2_{MBD}$ system could catch the equilibrium a bit later and also fluctuated much more during simulation trajectory compared to CID3100583–$MBD2_{MBD}$ system (Fig. 4a). Average RMSD along MD trajectory was 0.89 Å, 0.90 Å, 0.98 Å, 0.94 Å, and 1.05 Å for CID3100583, Ebis-Theophylline, Golgicide A Sanguinarine, and GHMCytosine respectively. It's been noticed that Golgicide A–$MBD2_{MBD}$ complex trajected more stably although it was the fifth rank in the sense of docking (Table 1) and reached steady state at the latest, flashing as another alternative repurposing drug candidate (Fig. 4a). In more detail, both protein backbone and Ebis-Theophylline in the complex could get the equilibrium by the 450 frames but later on they were seemed to

considerably fit in parallel until the end of time course (Fig. 4b). Similarly, atomic fluctuations and B-factor was diminished in line with interacted residues such as Asp176, Val177, Arg188, Lys190, Asp207, Phe208, Arg209 (Fig. 4c). In case of RoG and Q(x), it could be estimated that there has been a transition in folding state at around 300 frames and continued with almost none of remarkable alteration in contrast to CID3100583 system, implying Ebis-Theophylline could detain receptor in more stable as soon as it has interacted with $MBD2_{MBD}$ (Fig. 4d). In conclusion, dynamic behaviors and flexibilities of crystal structure of both $MBD2_{MBD}$–CID3100583 and $MBD2_{MBD}$–Ebis-Theophylline complexes during MD trajectory has been consistent, proving strict interactions constructed in computational docking approach.

# Calculation of MM-PB(GB)SA Binding Free Energy and $E_{Dec}$ Per Residue

After MD simulation, we employed MM-PBSA calculation and per-residue energy decomposition, aiming to elucidate putative binding free energy ($\Delta G_{Bind}$) between $MBD2_{MBD}$ and CID3100583 or Ebis-Theophylline, and the energetic contribution of each amino acid involved in ligand docking, respectively. As expected, compound-ligand complexes with lower docking scores and unfine fluctuations established relatively lower binding free energy. As expected, 8,8-Ethylenebistheophylline had the strongest binding affinity (-15.03 kcal/mol) among annotated library and second rank within all ligands. However, CID3100583 was ranked as the third (-12.26 kcal/mol) in the list, and interestingly CID126001284 (-18.23 kcal/mol) and CID136748749 (-13.88 kcal/mol) demonstrated the strongest binding affinity (see Supplementary Table S2). In $MBD2_{MBD}$–CID3100583 trajectory, van der Waals interactions (-34.49 kcal/mol) and polar solvation energy (-26.10 kcal/mol) mainly contributed to ligand binding while non-polar solvation energy (12.26 kcal/mol) conduced to unfavorable contributions (see Supplementary Table S2). From this point of view, we concluded van der Waals forces and polar interactions should have an indispensable role in CID3100583 docking on $MBD2_{MBD}$.

We next examined contribution per residue to binding free energy. Amino acids with energy decompose ($E_{Dec}$) contribution ranked in top ten, were Asp207, Arg209, Phe208, Val177, Phe206, Lys190, Asp176, Ser205, Tyr178, Thr210 in protein-ligand system (Fig. 5a, top panel). It should be noticed that top energetic contributors to binding free energy have consistently been amino acids that were diversely interacted with CID3100583 within binding pocket (Fig. 2c). As illustrated in Fig. 5a (bottom panels), the other residues except top ten contributors had a less impact on binding free energy but certain neighboring residues around top ten could faintly contribute, as well (Fig. 5a, bottom panels). Herein, van der Waals forces have been detected as main $E_{Dec}$ contribution however electrostatic energetics have constitutively been provided by Val177, Arg188, Ser204, Asp207 and Phe208 (Fig. 5a, bottom). In other respects, Ebis-Theophylline showed the top van der Waals (-37.55 kcal/mol) and polar solvation (-28.21 kcal/mol) contributions among all (see Supplementary Table S2). Accordingly, Arg188, Arg209, Phe208, Tyr178, Asp176, Val177, Val164, Asp207, Lys186, Lys190 were orderly top ten energetic contributors resulted from MD simulation (Fig. 5b, top panel), demonstrating an harmony with the amino acids

achieved in docking analysis (Fig. 2d–e). As plotted in Fig. 5b (bottom) Asp176, Val177, Tyr178, Arg188, Phe208 and Arg209 have been essential for van der Waals contribution and also principle contributors to overall binding free energy. Though Val177, Tyr178 and Lys186 provided electrostatic forces (Fig. 5b), such forces have been accounted weakly in total energy decompose (see Supplementary Table S2). Thus, these results demonstrated docking might have been enabled by favor of such energetics contributed by especially amino acids enclosing CID3100583 or 8,8-Ethylenebistheophylline binding pockets.

# ADMET Property Prediction and Interpretation For Assessment of Drug-Likeness

Concurrently, we performed a comprehensive ADMET prediction in order to gain insight about undesirable pharmacokinetics and putative adverse side-effects of candidate compounds that are main conflicts in drug development [41]. Indeed, CID3100583, Ebis-Theophylline, Golgicide A, and Regadenoson have relatively been appreciated with fine drug-likeness and less toxicity at overall while others established poorer ADMET properties (see Supplementary Table 3). Particularly, CID3100583 was once again the leading compound in this analysis. QED (Quantitative estimate of drug-likeness) score for CID3100583 was lower, indicating as an unattractive compound since it is too complex [42]. Even though it was estimated poor Caco-2 (human colon adenocarcinoma cell line) permeability, an in vitro parameter for intestinal uptake of especially orally administered drugs (-6.209 log.cm/s; see Supplementary Table 3), MDCK (Madin – Darby Canine Kidney cells) permeability value (permeability coefficient, $P_{app}$), another widely used in vitro standard to evaluate uptake of drugs, indicate high passive permeability ($4x10^{-6}$ cm/s; data not shown). In case of PPB (plasma protein binding), it has been convenient, suggesting free and efficient distribution without pharmacodynamically influenced by plasma proteins [43]. Importantly, BBB (Blood-Brain barrier) penetration was the lowest, meaning the least impact on Central Nervous System. Besides, CID3100583 demonstrate pretty lower probability to inhibit and/or to be catalyzed by metabolic enzymes such as CYP1A2, CYP2C19 (see Supplementary Table 3) and CYP2C9, CYP2D6, CYP3A4, avoiding a possible disruption on metabolic reactions. Along with high clearance, carcinogenic effect and unfavorable toxicity has not been foreseen over most of the vital organs such as heart, liver, respiratory system, skin, eye (see Supplementary Table 3). Moreover, there has been no reported adverse outcome for Tox21 pathway components some like Aryl Hydrocarbon Receptor (0.043), Androgen Receptor (0.15), Estrogen Synthetase (0.007), Estrogen Receptor (0.043), PPAR-gamma (0.001), Antioxidant response element (0.162), Heat shock factor response element (0.002) in which between range of excellent as 0-0.3. Similarly, Ebis-Theophylline has been interpreted having negligible undesirable pharmacokinetic behavior and adverse toxic effects, as well. Briefly, these results led us to assess those inhibitor candidates in regards to not only docking approach but also drug-likeness, deciding on the most promising compound for future *in vitro* or *in vivo* studies.

# Discussion

Since MBD2 is a miscellaneous participant in various biological phenomena and disorders, that links epigenetic methylation marks and gene expression regulation, it has arisen as a potent target for multi-

purpose. In current study, we aimed to inspect novel drug candidates targeting $MBD2_{MBD}$ through docking–based virtual screening and drug repurposing approaches. In this respect, we generated a comprehensive library including over 15,000 non-annotated compounds and biologically identified drugs, from ZINC15 database. Certain of ligands have been listed in Supplementary Table S1 as hitter ligands by filtering according to resulted binding affinity and physicochemical properties such as complexity, molecular weight, tPSA and also convenience to be purchasable or synthesizable for *de facto* attempts. Here, we have realized that numerous hit compounds (see Supplementary Table S1) and also unlisted ones consists of remarkable groups such as pyridine, quinoxaline, imidazole, isoquinoline, which are incorporated either in nucleosides and nucleotide analogs, or at least related structures to directly nucleotides or DNA-related constructs [44–46]. For instance, quinoxaline and isoquinoline derivatives have been demonstrated to bind and act as antagonists of cyclic nucleotide-related enzymes such as cyclic nucleotide phosphodiesterase [47] and P2X7 nucleotide receptor [48,49], suggesting these groups might be attractive for DNA– or nucleotide–recognizing domains. Furthermore, some of screened chemicals were purine– or pyrimidine–containing molecules (ex. CID3100583, CID136707272, CID135623722, CID124903520, CID136732748, TB6:RiboPyrimidoP) or straightly nucleotide/nucleoside derivatives (ex. Regadenoson, Wyosine, 8-Methylamino-Guanosine, Thiosangivamycin, Adenosine 5'-monophosphoramide, 5,6-Trimethyleneuridine, Fludarabine Base, 8-Aminoguanosine, Tubercidin, Hypoxanthosine, GHMcytosine, 6-Mercapto-7-Methylguanosine, 8-Chloroguanosine, Isovaleriansaures Coffein, 2-Methylformycin, 5'-N-Methylcarboxamidoadenosine, etc.), documented in Supplementary Table S1. Additionally, these compounds highly contain methyl groups. Indeed, it has been quite interesting that a considerable number of particularly adenosine and guanosine derivatives have become prominent as a consequence of drug screening. It was reported that residues in MBD region of MeCP2 and MBD2 exhibited binding affinity to mCAT, mCAC, mCAT, mCGG patterns and surprisingly unmethylated TG-dinucleotides [31,50]. Thereby, such interplay preferences of MBD might reflect why these compounds were recorded as hit molecules amongst the whole library. Then top ten of non-annotated and top five annotated molecules were proceeded for further docking analysis. Principally, CID3100583 (ZINC3109386) and 8,8-Ethylenebistheophylline (ZINC8612354) urged us for the exhaustive investigation due to the highest binding energy and the lowest $K_i$ (Table 1). At this point, chemical contexts of both substances have been intriguing since CID3100583 contains two pyrimidine–based rings and Ebis-Theophylline is composed of two interconnected druggable Theophylline (Dimethylxanthine) molecules, a methylated xanthine analog which antagonizes human A1, A2a, A2b Adenosine Receptors as in the treatment of airways diseases and vasodilation [51–53]. Together, similarities between actual recognition target of $MBD2_{MBD}$ (various nucleotide patterns above) and top hitters in the context of chemical and structural features have been speculated as possible reason underlying complementarity and fine docking.

Actually, all ligands have been demonstrated interacting with similar residues (Fig. 1). CID3100583 was prone to anchor $MBD2_{MBD}$ via 9 hydrogen bonds with amino acids Val177, Arg188, Lys190, Ser204, Phe206, Asp207, Phe208, and Arg209 (Table 2) and besides through van der Waals forces (Asp176, Tyr178, Ser189 and Phe208), pi-Alkyl (Val177), pi-Anion (Asp207) interactions (Table 3 and Fig. 2a-c). In

case of Ebis-Theophylline, residues Val177, Phe208, Arg209 have participated in total 5 hydrogen bonds (Table 2) and van der Waals interactions between Arg166, Ser189, Pro191; pi-Alkyl interactions with Tyr178, Arg188, Lys190, Arg209; and pi-Anion interactions with Asp176 (Table 3 and Fig. 2d-e), signifying both ligands have been stabilized by cementing highly reactive surroundings. In this respect, Liu et al. reported MBD2 bound to DNA via constitution of H-bonding between Arg166, Asp176, Tyr178, Arg188 within $MBD2_{MBD}$ and methylated-Cytosine (mC) along with adjacent Thymine or Guanine bases in the context of mCGG, mCAG, mCAT, mCC, mCT di-/tri-nucleotide patterns [31]. Along with pi-Anion interactions between referred amino acids and aromatic bases above, alternate residues such as Lys167, Ser168, Ser171, Lys186, Ser189, Lys190, Arg209, interacted with sugar-phosphate backbone of DNA. Buchmuller et al. also attributed regions 164–177 and 186–195 in vicinity to DNA duplex [27]. In other studies, Lys174 and Tyr178 have been reported as critical amino acids to selectively recognize and bind methylated CpG probes [50,54]. Favorably, all of these stated amino acids have been detected highly conserved between MBD1, MBD3, MBD4 and MeCP2 in human (see Supplementary Fig. S4) and MBD2 in different organisms via multiple sequence alignment. Residual counterparts in both MBD1 (Val20, Arg22, Tyr34, Arg44 and Ser45) [55] and MeCP2 (Arg106, Arg133, Trp155, Thr158) [31,56] have been involved in methylated-DNA recognition and interaction. These highlight the fact that ligands achieved in our study, especially CID3100583 and 8,8-Ethylenebistheophylline could actually interfere DNA–binding by selectively invading of those indispensable residues (Fig. 2) for the usual function of related domain. Moreover, our *in silico* prediction also yielded in a convergence between putative druggable pockets and ligand–anchored pockets (see Supplementary Fig. S3 and Table 2–3), again supporting the importance of docking sites of screened compounds, all together (Fig. 2 and see Supplementary Fig. S2). In this regard, CID3136570 (ZINC3151175) as one of the patented compound by Nelson et al., was quite outstanding for our consideration. Therein, it was invented to inhibit DNA-binding activity of both $MBD2_{MBD}$ and $MeCP2_{MBD}$ with relatively better $IC_{50}$ of 0.67 μM and 0.61 μM, respectively [57]. Because binding action remained unclear, it led us to survey its docking residues and energy on $MBD2_{MBD}$ in a similar manner. As a reference, binding energy resulted in our docking analysis was − 7.61 kcal/mol, even lower than assumptions for suggested binders in present study (Table 1). More interestingly, we detected Val177, Arg188, Arg209 in H-bonding while Arg166, Tyr178, Lys190, Phe208, Arg209 facilitated in hydrophobic interactions (see Supplementary Fig. S5), thus compatibly appreciating the value and expected potential of our outcome in a prospective real assay.

Next, we simulated each ligand–protein complex under a force–field in order to test their structural dynamics and stability in a time course. Initially, RMSD along the whole trajectory for each $MBD2_{MBD}$–ligand system was compared. As plotted in Fig. 3a and 4a, CID3100583, CID343482 and Ebis-Theophylline established a more stable dimensionality effects or conformational distributions over time, corresponding docking affinities (Table 1). Herein, both ligands (CID3100583 and Ebis-Theophylline) along with protein backbone, showed amenable fluctuations and equilibration timing concurrently with $MBD2_{MBD}$ receptor (Fig. 3b and 4b). As known, RMSD is not useful in practice to deeply analyze and distinguish local regions in more flexible or rigid [58]. Consistently, amino acids participated in particularly

H-bonding with both CID3100583 and Ebis-Theophylline, such as Val177, Arg188, Lys190, Ser204, Phe206, Asp207, Phe208, Arg209, remained non-fluctuating as revealed by RMSF and B-Factor graphics (Fig. 3c and 4c). Actually, regions between 160–166, 175–780, 186–188, coincide with strand and residues 190–198 corresponds to a helix[31] (also see Supplementary Fig. S4). As we expected, diminished B-factor along these districts were corroborated with the consent that increased fluctuations and thus higher B-factor is predicted for irregular loop regions [59]. More in particular, secondary structures including Val177, Arg188, Lys190 displayed more regularity than part of 160–166, suggesting durability contributed by ligand binding. Surprisingly, even though amino acids between Ser204 and Arg209 have been involved within an irregular loop region (see Supplementary Fig. S4), they possessed lower B-factor values (Fig. 3c and 4c). The magnitude of that reduction may presumably be due to immobilization of the loop region by extensive interactions with either CID3100583 or Ebis-Theophylline over there. As a matter of fact that interactions with CID3100583 were potently concentrated on residues between Ser204 and Arg209, hence ensuring a further decline even rather than strands or helices and corresponding zone in $MBD2_{MBD}$–Ebis-Theophylline complex (Fig. 3c and 4c). Of note, RoG has moderately diminished at around frames 250 and 400 for CID3100583 and Ebis-Theophylline (Fig. 3d and 4d), respectively, postulating a compaction in both protein–ligand systems [39,40]. Within this context, spectroscopic analysis revealed that upon methylated CpG binding, dynamic mobility and conformation of complex has been shifted under favor of DNA bending by $MBD2_{MBD}$ [60]. Similarly, Liu et al. reported a compaction in methylated DNA–bound $MeCP2_{MBD}$ detected by atomic force microscopy (AFM) in parallel with decrease in RoG correlation [61]. That may point the hypothesis that reduction in RoG and a parallel increment in Q(x) could be resulted from a further compaction and folding due to a configuring action of docked CID3100583 or Ebis-Theophylline by mimicking an ordinary MBD–DNA binding. In correlation, acquired energetics data also corroborated our docking results and conclusions so far. Both CID3100583 and Ebis-Theophylline have been ranked amongst the top ligands with the highest binding free energy (see Supplementary Table S2). Energy decomposition for CID3100583 showed Asp207, Arg209, Phe208 and Phe206 appeared in the top five in terms of contribution to overall $\Delta G_{Bind}$ (Fig. 5a), supporting rigidity in the loop region discussed above. Besides, per residue energy decomposition (Fig. 5) and interactions in docking analysis (Fig. 2) established a fine correspondence for both ligands. For instance, Phe208, one of the top van der Waals energy–decomposing residue for CID3100583 was also constituted in van der Waals interaction as a result of docking analysis while Val177 incorporated with Ebis-Theophylline by electrostatic forces via both approaches (Fig. 5b).

Consequently, current computational study has established prosperous and crosschecking evidence for certain ligands in a comprehensive manner. CID3100583 and 8,8-Ethylenebistheophylline have successfully been invented as novel and most likely MBD2 inhibitors in the light of effective binding affinity and negligible off-target probability. Though *in silico* approach was utilized within the scope of this paper, we have provided a consolidated basis for the next experimental studies testing their actual druggability. Likewise, a wide compound repertoire has also been gained to give insight into drug development or repurposing against multi-functional MBD2 target and still required to be evaluated. In

conclusion, we rationally propose that CID3100583 and 8,8-Ethylenebistheophylline are highly versatile with capability to be notably used as anti-cancer drugs and agents in iPSC reprogramming cocktails by impeding MBD2 function in related pathways.

## Methods

# Data Sources and Acquisition of 3D Protein Model and Ligand Structure Files

3D model file of DNA–binding domain of human MBD2 ($MBD2_{MBD}$) was obtained from RCSB PDB Protein Databank (https://www.rcsb.org/). The experimentally resolved structure of Chromatin-binding domains in the NuRD deacetylase complex (PDB: 7AO8) [62] was downloaded. Residues between 149–212 (Chain C) corresponding to $MBD2_{MBD}$ receptor was extracted from the whole assembly by removal of other chains in AutoDock tools.

Total 1293 tranches in ZINC15 database [63] were filtered by molecule charge (between 0 and + 2), molecular weight (MW, between 200 and 400 Da), and logP value (between − 1 and + 4.5) and thereby over 15,000 ligands could be extracted in .sdf format. Afterwards, ligand structure files were converted into .pdbqt file format via Open Babel toolbox [64] for the further docking analysis.

# High-throughput Virtual Screening

Structure–based virtual screening was accomplished by Vina wizard in PyRx program [65] in order to discover possible binders for related domain of MBD2 after ligand libraries including 15,893 compounds and $MBD2_{MBD}$ (7AO8; C149–C212) was pre-processed. Grid box was adjusted with centering coordinates 121.02, 178.69, 185.85 and sizes with 38.57 Å x 35.46 Å x 31.60 Å for X–, Y–, Z–dimensions, respectively, in order to involve entire structure. Before starting screening, all compounds in the library were energy minimized. Results were sorted in descending according to affinity and top 10 compounds with the best scores were evaluated for in silico docking approach.

# In Silico Molecular Docking

Molecular docking analysis was performed by AutoDock 4.2 (version 1.5.6), a widely used automated program for prediction of molecular interactions between protein and ligand structure [66]. Prior to docking run, both protein model (PDB: 7AO8, Chain C) and each ligand were structured and prepared using the mentioned software. Accordingly, removal of any other existed ligands/ions/heteroatoms and water molecules to obtain a crude protein structure, addition of polar hydrogens and missing atoms, calculation of Kollman charges, assignment of search parameter (Genetic Algorithm) and output format (Lamarckian GA) were all implemented as previously reported[67]. As distinct from, we have played 150 runs (conformations) with population size of 300 as rigid docking for each ligand-receptor complex. Besides,

grid box was adjusted in same coordinates and sizes as used in virtual screening with 0.375 Å spacing for blind docking since any possible binding residues in Methyl-DNA Binding Domain (MBD) of human MBD2 ($MBD2_{MBD}$) were not missed out. The first ranked conformation with the best binding energy and inhibition constant [68] was proceeded for further examination.

# Mapping Molecular Interactions and Distances

Intermolecular interactions such as hydrogen bonds (H-bonds), hydrophobic interactions, salt bridges, halogen bonds, and etc. in the best docked ligand-protein pose with calculated distances, were extracted and visually rendered using PLIP-Protein Ligand Interaction Profiler tool [69] and PyMOL Molecular Graphics System, Version 2.4.1 Schrödinger, LLC. Additionally, 2D interaction diagrams were illustrated by BIOVIA Discovery Studio Visualizer, v17.2.0.16349, Dassault Systèmes.

# Prediction of Druggable Pockets

Druggable binding pockets over $MBD2_{MBD}$ (7AO8, Chain C) were predicted via DoGSiteScorer tool [70] in Protein Plus Server designed by Universitat Hamburg ZBH Center For Bioinformatics (www.proteins.plus/), enabling us to widely deduce significancy of amino acids interacted with inhibitor candidates. A grid-based method with Difference of Gaussian filter is utilized to distinguish potential drug-binding sites regarding to various descriptors such as volume, hydrophobicity, chemical features, shape and enclosure on sole 3D structure of protein independent from ligands. Consequently, calculated Druggability Scores (DS) were exploited for each pocket. Herein, the higher the DS (ranges between 0 and 1) the more druggable the pocket is supposed, as indicated before[67,70].

# Molecular Dynamics (MD) Simulation

Dynamic behaviors of docked protein-ligand complexes were computed through VMD ver1.9.4a51 software [71] and a parallel molecular dynamics code NAMD [72] and CHARMM22 force field [73]. The parameters and protein topology files were generated automatic PSF builder inside VMD while ligand structures were utilized using CHARMM-GUI Ligand Reader and Modeler tool [74]. The systems were solvated with an octahedron box of TIP3P water environment [75] extended 10 Å in each direction and $Na^+/Cl^-$ counterions were added to neutralize the system. Upon energy minimization in 3000 steps, whole system was heated to 300K, in 30 ps, at 1 atm pressure under periodic boundary conditions to avoid edge effects. Finally, a production run of MD simulation was carried out for 20 ns duration. Flexibility and dynamics of each trajectory were evaluated by the query of Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (RoG), B-factor (Debye–Waller factor), fraction of native contacts (Qx), and estimated MM-PB(GB)SA binding free energy (deltaPB/ deltaGB). Resulted values were plotted and rendered by GraphPad Prism 6 graphic and statistics software.

# Calculation of MM-PB(GB)SA Binding Free Energy and Energy Decomposition

The Poisson–Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA) approaches are utilized in certain drug design strategies in order to estimate the free energy binding of small ligands to receptors [76,77]. First, a total of 200 snapshots from simulated MD trajectory for 20 ns were extracted to explore MM-PB(GB)SA free energy and energy decomposition ($E_{DEC}$) per residue. Then, van der Waals force ($\Delta E_{VDW}$), electrostatic interaction ($\Delta E_{ELE}$), total (polar + nonpolar) solvation free energy ($\Delta E_{SOL}$) contributions, and total estimated binding free energy ($\Delta G_{Bind}$) were employed as post-processing approach, using two different web servers AMMIS [78] and farPPI [79], by following recommended tool instructions.

# ADMET Property Prediction

Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) of chemicals have comprehensively been predicted by using upgraded ADMETLab 2.0 server [41]. After SMILES for each compound were prepared, they were submitted to server. Certain predicted properties were received for consideration in terms of Medicinal Chemistry (QED, Pfizer Rule, PAINS Alert), Absorption (Caco-2 Permeability, Pgp-inhibitor, HIA), Distribution (PPB, BBB Penetration), Metabolism (CYP1A2 inhibitor, CYP1A2 substrate, CYP2C19 inhibitor, CYP2C19 substrate), Excretion (CL, $T_{1/2}$), Toxicity (hERG, H-HT, LI, Sk, Re, Ca, Ey). Herein, we utilized ADMET-scores and reference ranges reported for each selected descriptor by server, in order to interpret chemical drug-likeness [41,80].

# Multiple Sequence Alignment

$MBD2_{MBD}$ protein (UniProt: Q9UBB, 145–213 aa) was subjected to homology-based multiple alignment with the related domains of human MeCP2 (UniProt: P516, 90–162 aa), MBD1 (UniProt: Q9UIS, 1–69 aa), MBD3 (UniProt: O9598, 1–72 aa) and MBD4 (UniProt: O9524, 76–148 aa), using PRALINE server [81] in order to discover outstanding residues compared to interacted amino acids in docking. Besides amino acid conservation, various features such as secondary structure prediction based on PSIPRED [82,83], hydrophobicity scale [84] and residue type adapted from CLUSTALX [85] was also compared by integrated modules.

# Declarations

### Author Contributions

Z.O.Ç. solely employed throughout the workflow of manuscript including conceptualization, investigation, data acquisition and interpretation, figure formatting, writing original draft, and proofreading manuscript.

### Competing Interests

The author declare no competing interests.

# References

1. Aloia, L. Epigenetic Regulation of Cell-Fate Changes That Determine Adult Liver Regeneration After Injury. *Frontiers in Cell and Developmental Biology* vol. 9 (2021).
2. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* vol. 38 (2013).
3. Gowher, H. & Jeltsch, A. Erratum: Mammalian DNA methyltransferases: New discoveries and open questions (Biochemical Society Transactions (2018)46 (1191-1202)Doi: 10.1042/BST20170574). *Biochemical Society Transactions* vol. 47 (2019).
4. Jurkowska, R. Z. & Jeltsch, A. Enzymology of mammalian DNA methyltransferases. *Adv. Exp. Med. Biol.* **945**, (2016).
5. Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, (2016).
6. Kribelbauer, J. F., Lu, X. J., Rohs, R., Mann, R. S. & Bussemaker, H. J. Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *Journal of Molecular Biology* vol. 432 (2020).
7. Vogt, G. Evolution of epigenetic mechanisms in animals and their role in speciation. in *Handbook of Epigenetics: The New Molecular and Medical Genetics* (2017). doi:10.1016/B978-0-12-805388-1.00027-4.
8. Locke, W. J. *et al.* DNA Methylation Cancer Biomarkers: Translation to the Clinic. *Frontiers in Genetics* vol. 10 (2019).
9. Dean, W., Lucifero, D. & Santos, F. DNA methylation in mammalian development and disease. *Birth Defects Research Part C - Embryo Today: Reviews* vol. 75 (2005).
10. Samblas, M., Milagro, F. I. & Martínez, A. DNA methylation markers in obesity, metabolic syndrome, and weight loss. *Epigenetics* vol. 14 (2019).
11. Brandt, B., Rashidiani, S., Bán, Á. & Rauch, T. A. DNA methylation-governed gene expression in autoimmune arthritis. *International Journal of Molecular Sciences* vol. 20 (2019).
12. Rugowska, A., Starosta, A. & Konieczny, P. Epigenetic modifications in muscle regeneration and progression of Duchenne muscular dystrophy. *Clinical Epigenetics* vol. 13 (2021).
13. Zhong, J., Agha, G. & Baccarelli, A. A. The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. *Circulation Research* vol. 118 (2016).
14. Bell, C. G. *et al.* DNA methylation aging clocks: Challenges and recommendations. *Genome Biology* vol. 20 (2019).
15. Stricker, S. H. & Götz, M. DNA-methylation: Master or slave of neural fate decisions? *Frontiers in Neuroscience* vol. 12 (2018).
16. Irvine, R. A., Lin, I. G. & Hsieh, C.-L. DNA Methylation Has a Local Effect on Transcription and Histone Acetylation. *Mol. Cell. Biol.* **22**, (2002).

17. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: Patterns and paradigms. *Nature Reviews Genetics* vol. 10 (2009).
18. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the mbd protein family. *Cell* **153**, (2013).
19. Stirzaker, C. *et al.* Methyl-CpG-binding protein MBD2 plays a key role in maintenance and spread of DNA methylation at CpG islands and shores in cancer. *Oncogene* **36**, 1328–1338 (2017).
20. Ng, H. H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* **23**, 58–61 (1999).
21. Jin, S. G., Jiang, C. L., Rauch, T., Li, H. & Pfeifer, G. P. MBD3L2 interacts with MBD3 and components of the NuRD complex and can oppose MBD2-MeCP1-mediated methylation silencing. *J. Biol. Chem.* **280**, 12700–12709 (2005).
22. Hendrich, B. & Bird, A. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
23. Wood, K. H. *et al.* Tagging methyl-CpG-binding domain proteins reveals different spatiotemporal expression and supports distinct functions. *Epigenomics* **8**, 455–473 (2016).
24. Becker, A. *et al.* Direct Homo- and Hetero-Interactions of MeCP2 and MBD2. *PLoS One* **8**, e53730 (2013).
25. Le Guezennec, X. *et al.* MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different Biochemical and Functional Properties. *Mol. Cell. Biol.* **26**, 843–851 (2006).
26. Tan, C. P. & Nakielny, S. Control of the DNA Methylation System Component MBD2 by Protein Arginine Methylation. *Mol. Cell. Biol.* **26**, 7224–7235 (2006).
27. Buchmuller, B. C., Kosel, B. & Summerer, D. Complete Profiling of Methyl-CpG-Binding Domains for Combinations of Cytosine Modifications at CpG Dinucleotides Reveals Differential Read-out in Normal and Rett-Associated States. *Sci. Rep.* **10**, 1–9 (2020).
28. Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U. & Renkawitz, R. The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. *J. Biol. Chem.* **275**, 34963–34967 (2000).
29. Desai, M. A. *et al.* An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.* **43**, 3100–3113 (2015).
30. Gnanapragasam, M. N. *et al.* p66α - MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2 - NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7487–7492 (2011).
31. Liu, K. *et al.* Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. *J. Biol. Chem.* **293**, 7344–7354 (2018).
32. Wood, K. H. & Zhou, Z. Emerging molecular and biological functions of MBD2, a reader of DNA methylation. *Frontiers in Genetics* vol. 7 (2016).

33. Hendrich, B. *et al.* Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. *Mamm. Genome* **10**, 906–912 (1999).
34. Lu, Y. *et al.* Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell Stem Cell* **15**, 92–101 (2014).
35. Du, Q., Luu, P. L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: Readers of the epigenome. *Epigenomics* vol. 7 1051–1073 (2015).
36. Menafra, R. & Stunnenberg, H. G. MBD2 and MBD3: Elusive functions and mechanisms. *Frontiers in Genetics* vol. 5 (2014).
37. Ginder, G. D. & Williams, D. C. Readers of DNA methylation, the MBD family as potential therapeutic targets. *Pharmacology and Therapeutics* vol. 184 98–111 (2018).
38. Wyhs, N., Walker, D., Giovinazzo, H., Yegnasubramanian, S. & Nelson, W. G. Time-resolved fluorescence resonance energy transfer assay for discovery of small-molecule inhibitors of methyl-CpG binding domain protein 2. *J. Biomol. Screen.* **19**, 1060–1069 (2014).
39. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17874–17879 (2013).
40. Brylinski, M. & Skolnick, J. Q-dock: Low-resolution flexible ligand docking with pocket-specific threading restraints. *J. Comput. Chem.* **29**, 1574–1588 (2008).
41. Xiong, G. *et al.* ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* (2021) doi:10.1093/nar/gkab255.
42. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **4**, 90–98 (2012).
43. Buscher, B. *et al.* Bioanalysis for plasma protein binding studies in drug discovery and drug development: Views and recommendations of the European Bioanalysis Forum. *Bioanalysis* **6**, 673–682 (2014).
44. García, B. *et al.* Interaction of the DNA bases and their mononucleotides with pyridine-2-carbaldehyde thiosemicarbazonecopper(II) complexes. Structure of the cytosine derivative. *J. Inorg. Biochem.* **102**, 1892–1900 (2008).
45. Ali, I. A. I., Al-Masoudi, I. A., Aziz, N. M. & Al-Masoudi, N. A. New acyclic quinoxaline nucleosides. Synthesis and anti-HIV activity. *Nucleosides, Nucleotides and Nucleic Acids* **27**, 146–156 (2008).
46. Röthlisberger, P. *et al.* On the enzymatic incorporation of an imidazole nucleotide into DNA. *Org. Biomol. Chem.* **15**, 4449–4455 (2017).
47. Parra, S. *et al.* Imidazo[1,2-a]quinoxalines: Synthesis and cyclic nucleotide phosphodiesterase inhibitory activity. *Eur. J. Med. Chem.* **36**, 255–264 (2001).
48. Humphreys, B. D., Virginio, C., Surprenant, A., Rice, J. & Dubyak, G. R. Isoquinolines as antagonists of the P2X7 nucleotide receptor: High selectivity for the human versus rat receptor homologues. *Mol. Pharmacol.* **54**, 22–32 (1998).

49. Watano, T., Matsuoka, I. & Kimura, J. Characteristics of ATP-induced current through P2X7 receptor in NG108-15 cells: Unique antagonist sensitivity and lack of pore formation. *Jpn. J. Pharmacol.* **88**, 428–435 (2002).
50. Fraga, M. F. *et al.* The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Research* vol. 31 1765–1774 (2003).
51. Landells, L. J. *et al.* The role of adenosine receptors in the action of theophylline on human peripheral blood mononuclear cells from healthy and asthmatic subjects. *Br. J. Pharmacol.* **129**, 1140–1144 (2000).
52. Cheng, R. K. Y. *et al.* Structures of Human A1 and A2A Adenosine Receptors with Xanthines Reveal Determinants of Selectivity. *Structure* **25**, 1275-1285.e4 (2017).
53. Fieger, S. M. & Wong, B. J. Adenosine receptor inhibition with theophylline attenuates the skin blood flow response to local heating in humans. *Exp. Physiol.* **95**, 946–954 (2010).
54. Cramer, J. M. *et al.* Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *J. Biol. Chem.* **289**, 1294–1302 (2014).
55. Ohki, I. *et al.* Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105**, 487–497 (2001).
56. Ballestar, E., Yusufzai, T. M. & Wolffe, A. P. Effects of rett syndrome mutations of the Methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry* **39**, 7100–7106 (2000).
57. Nelson, W. G., Yegnasubramanian, S., Lin, X., Speed, T. J. & Reichert, Z. Agents for Reversing Epigenetic Silencing of Genes. (2010).
58. Sargsyan, K., Grauffel, C. & Lim, C. How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **13**, 1518–1524 (2017).
59. Craveur, P. *et al.* Protein flexibility in the light of structural alphabets. *Frontiers in Molecular Biosciences* vol. 2 (2015).
60. Pan, H. *et al.* CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level. *Nucleic Acids Res.* **45**, 9164–9177 (2017).
61. Liu, M. *et al.* DNA looping by two 5-methylcytosine-binding proteins quantified using nanofluidic devices. *Epigenetics and Chromatin* **13**, 1–10 (2020).
62. Millard, C. J., Fairall, L., Ragan, T. J., Savva, C. G. & Schwabe, J. W. R. The topology of chromatin-binding domains in the NuRD deacetylase complex. *Nucleic Acids Res.* **48**, 12972–12982 (2020).
63. Sterling, T. & Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
64. O'Boyle, N. M. *et al.* Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
65. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **1263**, 243–250 (2015).

66. Morris, G. M. *et al.* Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
67. Çalışkaner, Z. O. Determination of binding potential of HCV protease inhibitors against SARS-CoV-2 Papain-like protease with computational docking. *Letters in Drug Design & Discovery*. **18**, 1 (2021).
68. Pantsar, T. & Poso, A. Binding affinity via docking: Fact and fiction. *Molecules* vol. 23 1DUMMY (2018).
69. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–W447 (2015).
70. Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**, 2074–2075 (2012).
71. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
72. Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
73. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
74. Kim, S. *et al.* CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* **38**, 1879–1886 (2017).
75. Price, D. J. & Brooks, C. L. A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **121**, 10096–10103 (2004).
76. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* vol. 10 449–461 (2015).
77. Yang, J. F., Wang, F., Chen, Y. Z., Hao, G. F. & Yang, G. F. LARMD: Integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. *Brief. Bioinform.* **21**, 2206–2218 (2020).
78. Wu, F. X. *et al.* AIMMS suite: A web server dedicated for prediction of drug resistance on protein mutation. *Brief. Bioinform.* **21**, 318–328 (2018).
79. Wang, Z. *et al.* FarPPI: A webserver for accurate prediction of protein-ligand binding structures for small-molecule PPI inhibitors by MM/PB(GB)SA methods. *Bioinformatics* **35**, 1777–1779 (2019).
80. Dong, J. *et al.* Admetlab: A platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform.* **10**, 29 (2018).
81. Simossis, V. A. & Heringa, J. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* **33**, W289–W294 (2005).
82. Pirovano, W. & Heringa, J. Protein secondary structure prediction. *Methods in molecular biology (Clifton, N.J.)* vol. 609 327–348 (2010).

83. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
84. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142 (1984).
85. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).

# Figures

**Figure 1**

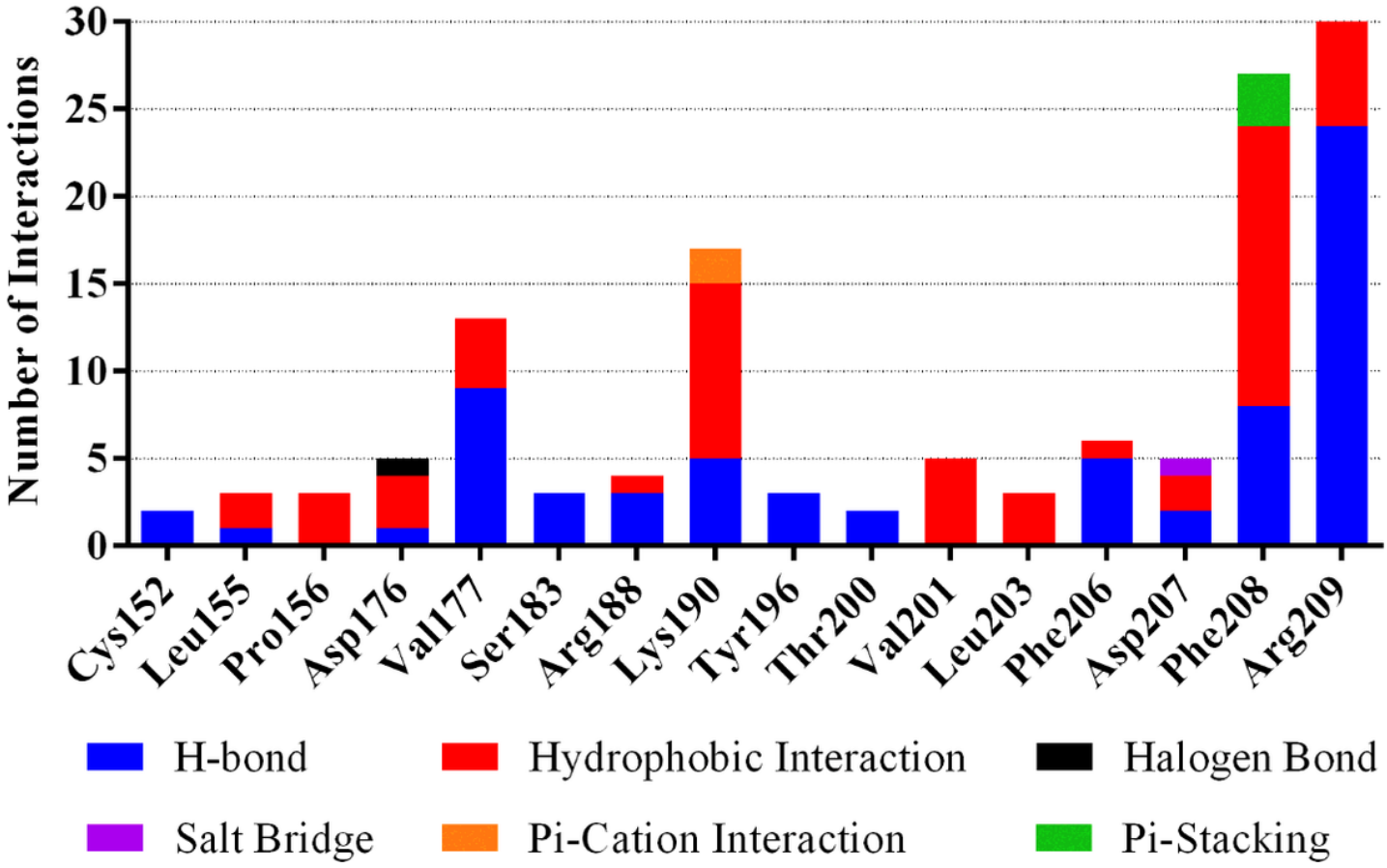
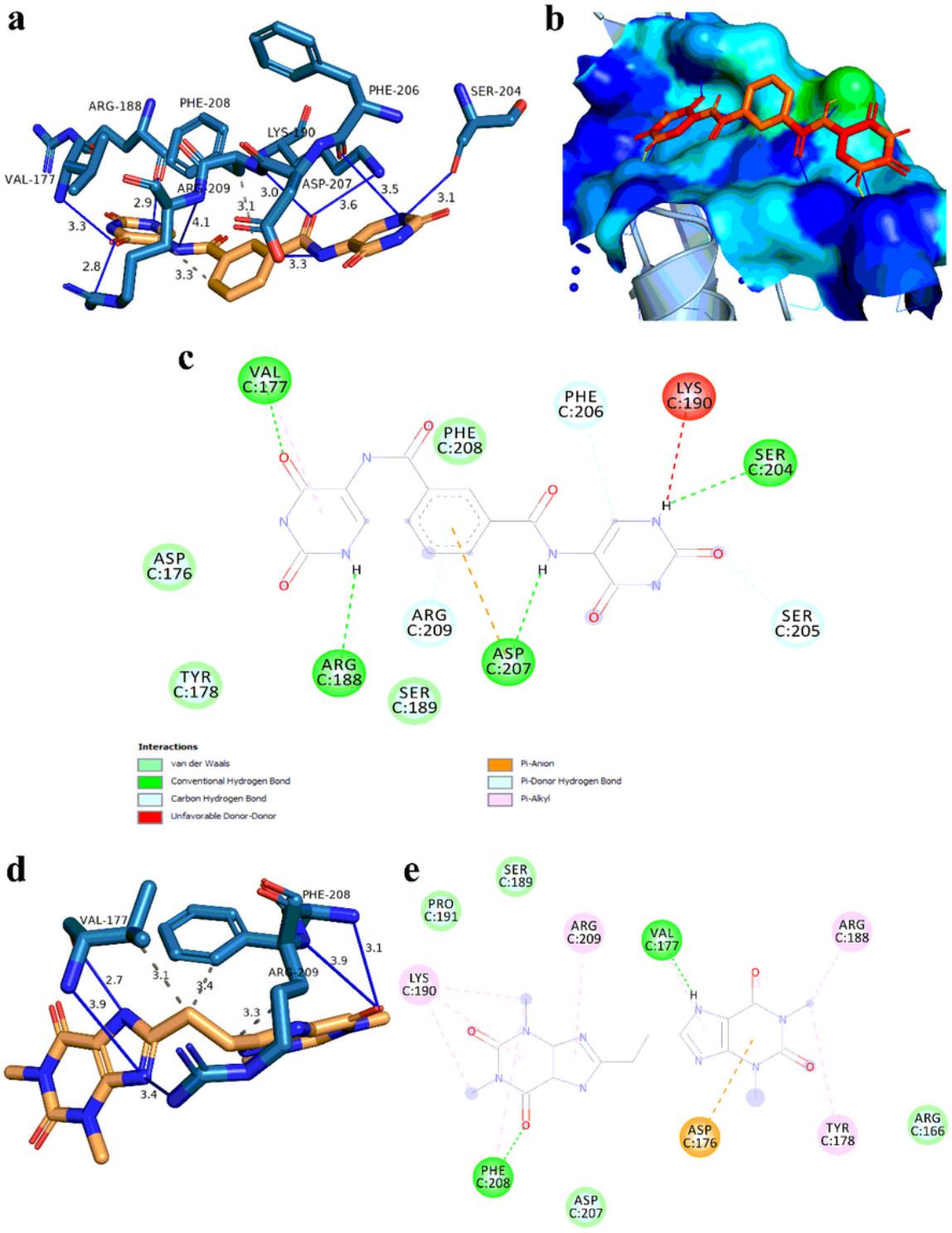



**Figure 1**

Number of interactions contributed by common amino acids which is totally distributed within top ten non-annotated and five annotated compounds from virtual screening.

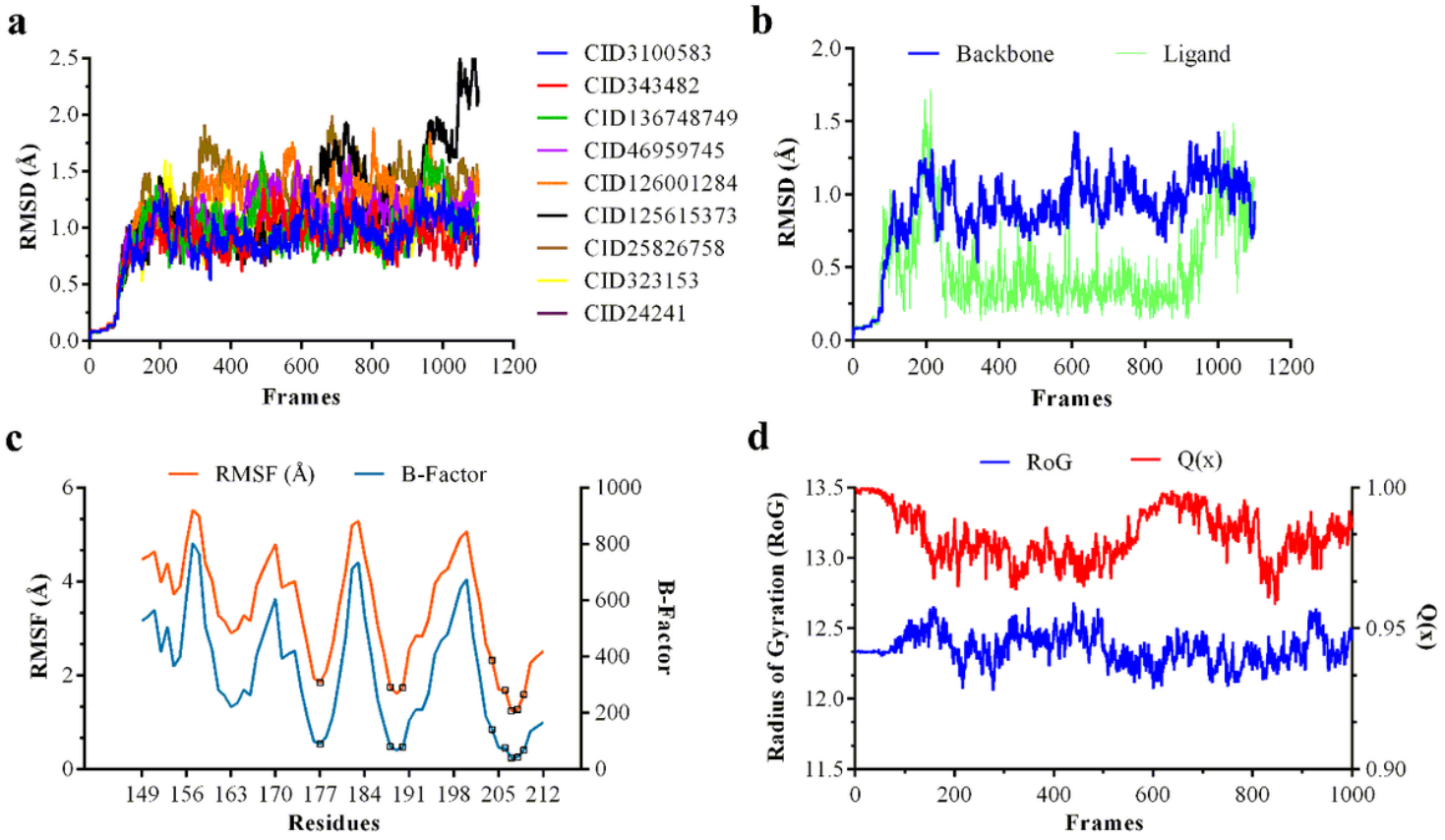
# Figure 2




## Figure 2

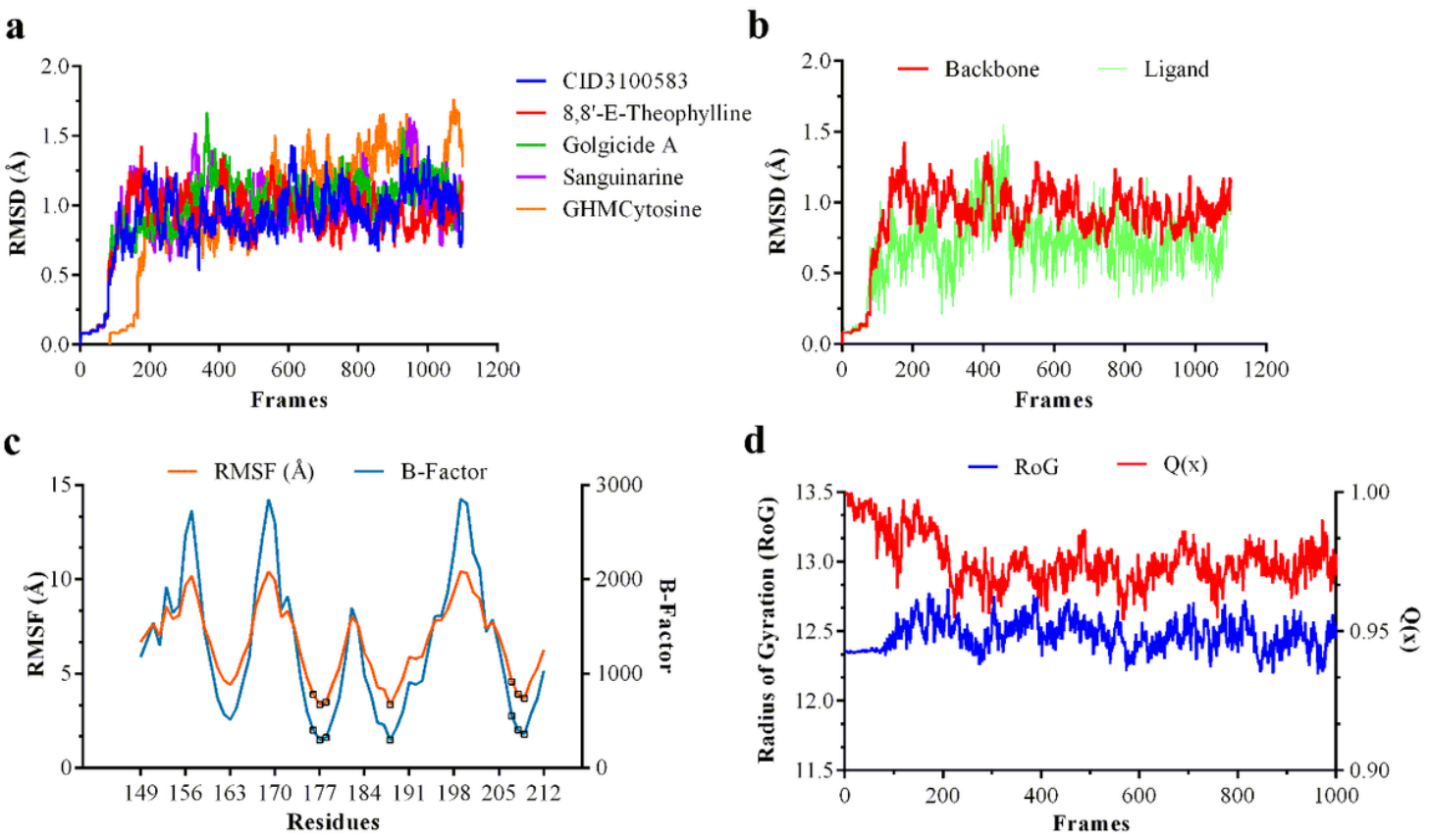
Illustration of docked MBD2MBD–Ligand complexes. The best docking pose with interactions between atoms of MBD2MBD (blue) and CID3100583 (yellow) (a), visual binding surface (b), and depiction in 2D diagram (c). Docking attitude of MBD2MBD–Ebis-Theophylline complex in spatial (d) and 2D representation (e). Hydrogen bonds and hydrophobic interactions are indicated by solid blue and dashed grey lines, respectively, with the distances between the atoms in angstrom.

Figure 3




## Figure 3

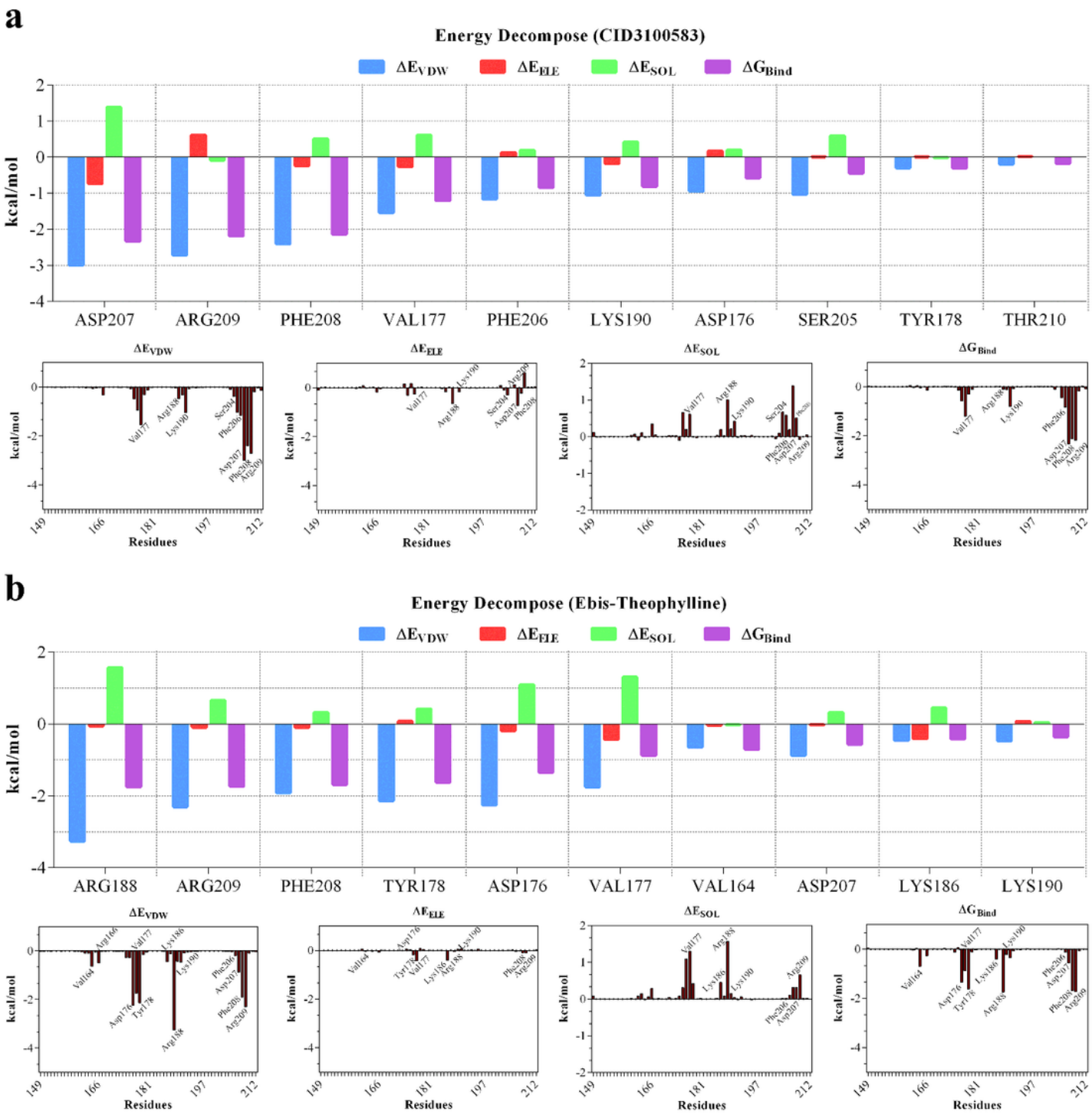
MD simulation output of MBD2MBD–CID3100583 system. RMSD values of all top non-annotated ligands along the whole trajectory (a). Overall fluctuation of ligand and protein backbone (b), atomic fluctuations and B-factor per residue (c), and RoG with Q(x) (d) are plotted for solely CID3100583. Tiny squares on RMSF and B-factor curves, corresponds to Val177, Arg188, Lys190, residues between Ser204–Arg209, respectively.




## Figure 4

MD simulation output of MBD2MBD–Ebis-Theophylline system. RMSD values of certain top annotated ligands along the whole trajectory (a). Overall fluctuation of ligand and protein backbone (b), atomic fluctuations and B-factor per residue (c), and RoG with Q(x) (d) are plotted for solely 8,8-Ethylenebistheophylline. Tiny squares on RMSF and B-factor curves, corresponds to residues between Asp176–Tyr178, Arg188, residues between Asp207–Arg209, respectively.

# Figure 5




## Figure 5

Detailed binding free energetics of both MD simulation output of MBD2MBD–Ligand systems. Graphics of top ten residues (top panels) with the highest energy decomposition per residue for MBD2MBD–CID3100583 (a) and MBD2MBD–Ebis-Theophylline (b) complexes. Bottom panels (a, b) infer different energetic contributions of all residues within each system.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf