

Structural insights into binding of small molecule inhibitors to Enhancer of Zeste Homolog 2

Marko Kalinić · Mire Zloh · Slavica Erić

Received: 11 March 2014 / Accepted: 7 August 2014 / Published online: 20 August 2014
© Springer International Publishing Switzerland 2014

Abstract Enhancer of Zeste Homolog 2 (EZH2) is a SET domain protein lysine methyltransferase (PKMT) which has recently emerged as a chemically tractable and therapeutically promising epigenetic target, evidenced by the discovery and characterization of potent and highly selective EZH2 inhibitors. However, no experimental structures of the inhibitors co-crystallized to EZH2 have been resolved, and the structural basis for their activity and selectivity remains unknown. Considering the need to minimize cross-reactivity between prospective PKMT inhibitors, much can be learned from understanding the molecular basis for selective inhibition of EZH2. Thus, to elucidate the binding of small-molecule inhibitors to EZH2, we have developed a model of its fully-formed cofactor binding site and used it to carry out molecular dynamics simulations of protein–ligand complexes, followed by molecular mechanics/generalized born surface area calculations. The obtained results are in good agreement with biochemical inhibition data and reflect the structure–activity relationships of known ligands. Our findings suggest that the variable and flexible post-SET domain plays an important role in inhibitor binding, allowing possibly distinct binding modes of inhibitors with

only small variations in their structure. Insights from this study present a good basis for design of novel and optimization of existing compounds targeting the cofactor binding site of EZH2.

Keywords Enhancer of Zeste Homolog 2 · Epigenetic target · Binding mode elucidation · Molecular dynamics

Introduction

Histones undergo a number of post-translational modifications (PTMs) which influence either directly, or through recruitment of downstream effector proteins, the remodeling of chromatin and transcriptional accessibility of genes [1, 2]. Introduction and removal of histone PTMs is a fundamental epigenetic regulatory mechanism underlying the maintenance of specific transcriptional programs within cells. Aberrations in normal epigenetic signaling have been associated with several human diseases, such as cancer, neuropsychiatric and autoimmune disorders [3]. Consequently, enzymes and proteins which form the intricate epigenetic regulatory network—writers, erasers and readers of the ‘histone code’ [4, 5]—have been recognized as promising drug targets [6, 7]. For example, histone deacetylase inhibitors have already gained clinical approval, whereas the therapeutic potential of targeting other histone modifying enzymes is only beginning to emerge [8–11].

One such family of enzymes are protein lysine methyltransferases (PKMTs) which catalyze, with high substrate and product specificity, the sequential methylation of lysine residues on histone tails and some other protein substrates [12]. PKMT function relies on cofactor *S*-adenosyl-L-methionine (SAM), which acts as a methyl group

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9788-1) contains supplementary material, which is available to authorized users.

M. Kalinić · S. Erić (✉)
Department of Pharmaceutical Chemistry, Faculty of Pharmacy,
University of Belgrade, Vojvode Stepe 450,
11000 Belgrade, Serbia
e-mail: seric@pharmacy.bg.ac.rs

M. Zloh
Department of Pharmacy, University of Hertfordshire,
Hatfield AL10 9AB, Hertfordshire, UK

donor in an S_N^2 type transfer reaction. With the exception of DOT1L, all human PKMTs identified to date are organized around a conserved SET domain characterized by a central pseudo-knot fold and the presence of ELx₂F/YDY and NHS/CxxPN motifs, 'x' being any amino acid [13]. The SET domain is approximately 130 residues long and contains a SAM binding pocket and a substrate binding cleft which intersect at the catalytic site. Several potent and selective chemical probes targeting the substrate binding site have been reported and analysis of resolved complex structures has provided useful guidelines for the design of novel, selective, substrate-competitive PKMT inhibitors [14].

Conversely, druggability of PKMT cofactor binding site has been far less explored. This site is only partly formed by the SET domain, as a short and variable post-SET domain contributes to its full enclosure. The post-SET domain is highly flexible, most frequently cysteine-rich and folded around a coordinated zinc ion, but is predicted to be only partly ordered in the apo form of PKMTs [15]. Analysis of the cofactor binding site in crystal structures of various PKMTs has revealed a significant degree of conservation in the intermolecular interaction patterns established by the cognate ligand SAM [16]. Although this finding highlighted the challenge of designing inhibitors with an adequate selectivity profile, diversity of side-chains forming the binding site has been recognized as a potential basis for design of selective, SAM-competitive PKMT inhibitors.

Chemical tractability of selectively targeting the cofactor binding site of SET-domain PKMTs was recently confirmed by the discovery of highly selective Enhancer of Zeste Homolog 2 (EZH2) inhibitors [17–21]. EZH2 is a SET domain PKMT which acts as the catalytic sub-unit of the Polycomb Repressive Complex 2 (PRC2). Aberrant expression of EZH2 and point-mutations in its catalytic region have been related to the development and aggressiveness of several human malignancies, including breast and prostate cancer, pediatric rhabdoid tumors and myelodysplastic syndromes [22–24]. Thus, in addition to investigating the biology of chromatin-modifying enzymes, there seems to be significant therapeutic potential in targeting EZH2/PRC2 enzymatic activity.

Discovery of EZH2 inhibitors resulted from high-throughput screening campaigns carried out independently by several groups, which, interestingly, all led to the identification of molecules based on a common scaffold containing a pyridone moiety connected to a central aromatic core by an amide linkage. Reported attempts to co-crystallize the inhibitors with EZH2/PRC2 did not lead to an experimental structure of the complex [20]. Never the less, as EZH2 inhibitors represent, to date, the only group of compounds to selectively target the cofactor binding site

of any SET-domain PKMT, there is significant information which can be obtained from systematically studying their binding. In addition to providing a rational basis for further structure-based molecular design of potential EZH2 inhibitors, understanding the mechanisms underlying molecular recognition between these highly selective compounds and EZH2 can also provide important guidelines for targeting other PKMTs. Therefore, in this study we aimed to integrate the available SAR and biochemical data in a computational study which would provide a structural basis for understanding the observed activity spectrum of EZH2 inhibitors.

A significant challenge to this end is lack of information on the fully ordered structure of the cofactor binding site. Recent efforts to characterize EZH2 structure by X-ray crystallography have resulted in two publicly available structures of its SET and CXC domains [25, 26]. However, both reported structures do not provide direct insight into binding of either SAM or the EZH2 inhibitors.

The next challenge is addressing the plasticity of the binding site resulting from post-SET motions inherently required for cofactor turnover. This plasticity implies that exploring the large configurational space of potential EZH2-inhibitor interactions cannot be carried out without considering the dynamic nature of the interaction, which has not been done so far. Recently released results on the kinetics of EZH2 inhibitor binding also suggest that minor alternations in structure lead to large differences in binding profiles [27], suggesting possibly distinct binding modes of closely related inhibitors. In this respect, molecular dynamics (MD) simulations offer the possibility to explicitly take into account protein and ligand motions as well as the opportunity to test how well the obtained binding modes can quantitatively account for biochemical activity data.

Lastly, EZH2 is known to undergo several point-mutations within the catalytic domain which affect its substrate and product specificity. Such mutations have been reported to affect EZH2 inhibitor binding affinities up to fivefold [20, 21], but the structural basis for these observations remains incompletely understood.

To address all the issues raised and gain a detailed insight into the molecular interactions between EZH2 and its selective inhibitors, we developed a model of EZH2 SET and post-SET domains, based on recently released crystallographic structures, but with the cofactor site fully formed. We refined the initial model using MD simulations guided by the conserved interaction patterns to cofactor SAM. The established model then served as basis for testing and extending an EZH2 inhibitors binding hypothesis through MD simulations and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) calculations. Finally, we studied the changes that EZH2 active site

undergoes in point-mutants and attempted to rationalize the enhanced activity of selected inhibitors against these altered forms of the enzyme.

Our findings provide a detailed insight into the interaction of selective SAM-competitive PKMT inhibitors and their target and provide a good basis for further structure-based computer-aided design of molecules inhibiting the enzymatic activity of EZH2.

Methods

EZH2 modeling

We modeled a 136 residue long EZH2 segment with a fully formed cofactor-binding site by using the SET domain from a reported EZH2 crystallographic structure [25] and a model of the post-SET domain generated using homology modeling.

Structures of SETD7 and SETD8 were used as templates for modeling the post-SET domain as these are PKMTs closely related to EZH2 that do not have a cystine-rich post-SET domain. Canonical sequences of human EZH2 segment (residues 611–746), SETD7 segment (residues 214–366), and SETD8 segment (residues 257–393) were retrieved from the UniProt Protein Knowledgebase (UniProtKB entries Q15910, Q8WTS6, and Q9NQR1, respectively) and aligned using T-Coffee Server [28, 29]. The alignment is provided as Supplementary Fig. 1. Ternary complex structure of SETD7 resolved to 1.75 Å [30] (PDB entry: 1O9S) and ternary complex structure of SETD8 resolved to 1.45 Å [31] (PDB entry: 1ZKK) were downloaded from the Protein Data Bank and imported into Chimera 1.8 [32]. Structures were processed so that single chain of each structure was kept, histone substrates were removed but the cofactor product *S*-adenosyl-*L*-homocysteine (SAH) was retained throughout the modeling process. Multiple template homology modeling was carried out using Modeller 9.11 [33] with Chimera as a graphical interface. One thousand EZH2 models were generated and 50 best models, as evaluated by the Modeller objective function, were saved.

Crystallographic structure of EZH2 CXC and SET domains resolved to a resolution of 2 Å [25] (PDB entry: 4MI0) was truncated to a segment covering residues 611–726. Missing residues Y658 and D659 were added using the Modeller loop modeling protocol. The 50 previously saved EZH2 homology models were superimposed onto the truncated crystallographic structure and best matching model was chosen, as judged by C α atoms' coordinate root mean square deviation (RMSD). The chosen model had a 3.6 Å RMSD from the crystallographic structure, with respect to SET domain residues. Segment of

this model covering residues 727–746 was then grafted to the experimentally resolved SET domain structure. Structure of SAH, which was included in the modeling from the template structures of SETD7 and SETD8, was not kept but was subsequently added by superposition from the crystallographic structure of GLP [34] (PDB entry: 3HNA). This was done for consistency, as other authors previously discussed binding site similarity between EZH2 and GLP [25]. We note, however, that due to its highly conserved conformation and contacts to the protein backbone, the position of SAH in structures of both SETD7 and SETD8 is nearly identical to that in the developed model, with heavy atom RMSD (after protein backbone alignment) of 0.38 and 0.87 Å, respectively.

The resulting 136 residue long model of EZH2 in complex with SAH was refined through classical molecular dynamics (MD) and accelerated MD (aMD) simulations. N-terminus of the EZH2 segment was capped with an acetyl group and missing hydrogens were added so that protonation of all residues corresponded to predictions obtained from the PROPKA3 [35] web server at environmental pH of 8.0. Tleap module of AmberTools 12 [36] was used to solvate the complex in a pre-equilibrated rectangular water box, leaving a 10 Å solvent padding in each direction. System was neutralized by adding a single Na⁺ ion. Amber ff99SB forcefield [37] was used for the protein and the TIP3P model [38] for water molecules. SAH was parameterized using previously reported Amber-compatible parameters [39].

System was then minimized to total potential energy convergence through 10,000 steps of conjugate gradient minimization; for the first 5,000 steps, coordinates of all protein and cofactor atoms were kept fixed, with only solvent molecules allowed to move, followed by an additional 5,000 steps of minimization with protein backbone atoms and cofactor heavy atoms harmonically restrained using a force constant of 10 kcal (mol Å²)⁻¹. These restraints were kept for the heating phase during which the system temperature was gradually raised from 0 to 298 K in 1 K per 0.3 ps increments. After heating, a 2 ns equilibration run was carried out in the isothermal-isobaric ensemble. During this simulation, post-SET domain residues were allowed to move freely, whereas cofactor and SET domain backbone were restrained using a force constant of 1 kcal (mol Å²)⁻¹. Last nanosecond of the simulation, after the 2 ns equilibration, was used in determining optimal parameters for the aMD simulation, as detailed below.

Implementation of aMD in NAMD [40, 41] was used to enhance the sampling of the post-SET domain while keeping the SET domain backbone and cofactor close to their experimentally determined coordinates with a weak harmonic restraint of 1 kcal (mol Å²)⁻¹. Dual boost mode [42] was used with acceleration parameters chosen as described

in previous studies [43]. Specifically, the parameters were calculated according to the following formulae:

$$E_{dih} = (V_{dih} + 4 \cdot n_{res}$$

$$\alpha_{dih} = \frac{4}{5} \cdot n_{res}$$

$$E_{total} = (V_{total} + \frac{n_{res}}{5}$$

$$\alpha_{total} = \frac{n_{atom}}{5}$$

where $\langle V_{dih} \rangle$ and $\langle V_{total} \rangle$ are the average dihedral and total potential energy calculated from conventional MD, n_{res} is the number of protein residues, and n_{atom} is the total number of atoms in the system. Calculation of parameters using these formulae has been proposed to work well as an initial choice for simulations of solvated proteins utilizing the Amber forcefield [44]. Accordingly, boost parameters were set to values of $E_{dih} = 1,954 \text{ kcal mol}^{-1}$ and $\alpha_{dih} = 109 \text{ kcal mol}^{-1}$, whereas total potential boost parameters were set to values of $E_{tot} = -60,509 \text{ kcal mol}^{-1}$ and $\alpha_{tot} = 5,300 \text{ kcal mol}^{-1}$. Simulation was run for a total of 25 ns, using a 2 fs time step. It should be noted that due to the modified nature of the potential, the aMD simulation time of 25 ns does not reflect ‘actual time’ required for the observed structural transitions [45]. The resulting trajectory was clustered based on protein backbone RMSD and a representative structure from the most populated cluster in which the post-SET domain was in direct contact to SAH was chosen. Care was taken to consider only the top 30 % energetically most favorable structures, based on the largest added boost.

To obtain a structure of EZH2 in apo form suitable for study of inhibitor binding, the chosen structure from the aMD run was processed by removing the cofactor product SAH, and then solvating and neutralizing the system. Upon minimization and heating, a 15 ns unrestrained simulation was carried out. A representative structure from this simulation was determined, minimized, evaluated for stereochemical correctness using PROCHECK [46] at SAVS v3 server and then subsequently used in docking studies. Choice of the representative structure was made by two passes of clustering: first, based on backbone atoms with a cutoff of 1.5 Å, and second, based on binding site residues’ heavy atoms with a cutoff of 1 Å. Based on preliminary docking studies on all cluster centers with more than 150 members, we identified the second most populated cluster as the one best suited to reproduce the experimentally proposed binding modes.

Binding site similarity assessment

Similarity of the modeled EZH2 cofactor binding site was compared to SET domain PKMTs: G9a (PDB entry: 2O8J),

GLP (PDB entry: 3HNA), MLL (PDB entry: 2W5Z), SETD7 (PDB entry: 3M53), SETD8 (PDB entry: 1ZKK), SETD2 (PDB entry: 4H12), SETMAR (PDB entry: 3BO5), and Suv39H2 (PDB entry: 2R3A). Similarity assessment was performed using FuzCav [47]. FuzCav is an alignment-free method which generates a binding site fingerprint in the form of a vector of 4,833 integers. Fingerprints are assigned to ligand binding sites by mapping pharmacophoric properties to C_α atom coordinates and then iteratively counting unique pharmacophoric triplets occurring in fixed distance bins. Binding sites were defined as in the original publication based on sc-PDB definition [48]; specifically, any histone substrates present in the retrieved PDB structures were first removed for uniformity and only residues within 6.5 Å of any heavy atom of the cofactor SAM or cofactor product SAH were retained for comparison. To evaluate the contribution of post-SET residues to binding site diversity, an additional set of files were prepared containing only binding site residues occurring N-terminal to Y726 in EZH2, conserved tyrosine being retained as well.

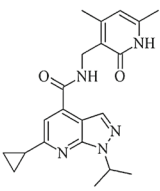
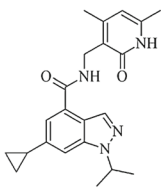
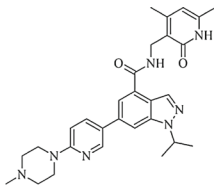
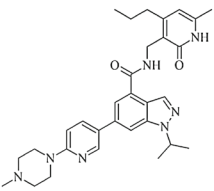
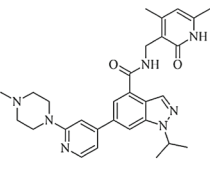
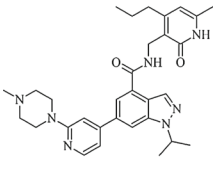
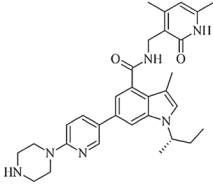
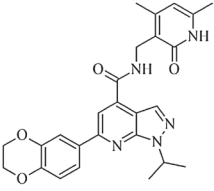
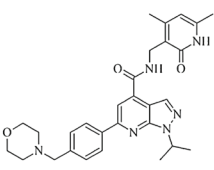
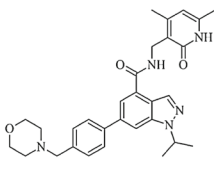
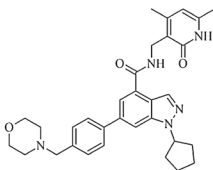
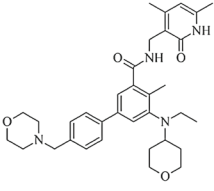
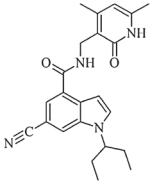
EZH2-inhibitor complex preparation and MD simulations

We studied the binding mode of 13 EZH2 inhibitors reported in the literature, whose structures and biochemical activity data are summarized in Table 1.

Initial complex configurations were prepared by docking the inhibitors to the representative apo structure of EZH2 using AutoDock Vina [49]. Ligand protonation was adjusted to reflect the major microspecies at pH 8.0, as predicted by the Calculator Plugin of Marvin 6.0.3 [50]. Specifically, all ligands were considered to be neutral except 126, which was singly protonated. Search box was centered on the putative position of the cofactor and spanned 20 Å in each direction. Search exhaustiveness was set to 16 and 15 binding modes were retained for manual inspection. Binding mode which best reflected the previously proposed orientation and hydrogen bonding pattern of the inhibitors, as discussed under [Results and discussion](#) in “[Analysis of EZH2 inhibitors’ binding mode](#)” section, and illustrated in Fig. 3, was kept for further study irrespective of Vina docking score. Vina scores for best scoring pose and the one chosen for further study, with RMSD values between the two, are reported in Supplementary Table 1.

The resulting complexes were prepared for molecular dynamics simulations observing the following procedure. The ff99SB forcefield was used for the protein. Ligands were parameterized according to the general Amber forcefield (GAFF) [51], version 1.4, using Antechamber [52] to automatically assign GAFF-compatible atom types.

Table 1 Structures of 13 EZH2 inhibitors and their biochemical activity reported in the literature, with abbreviated identifiers (IDs) under which they are referred to in this study

				
ID: G1K	ID: G2K	ID: G3K	ID: G4K	ID: G5K
$K_i^a = 149 \pm 28$ nM	$K_i = 74 \pm 11$ nM	$K_i = 7.9 \pm 3$ nM	$K_i = 0.6 \pm 0.05$ nM	$K_i = 14 \pm 5$ nM
Ref: [17]	Ref: [17]	Ref: [17]	Ref: [17]	Ref: [17]
				
ID: G6K	ID: 126	ID: E1Z	ID: E2Z	ID: E3Z
$K_i = 1.2 \pm 0.2$ nM	$K_i = 0.57$ nM	$K_i = 310$ nM	$K_i = 180$ nM	$K_i = 80$ nM
Ref: [17]	Ref: [19]	Ref: [20]	Ref: [20]	Ref: [20]
				
	ID: E4Z	ID: E5Z	ID: EL1	
	$K_i = 24$ nM	$K_i = 2.5$ nM	$K_i = 13 \pm 3$ nM	
	Ref: [20]	Ref: [21]	Ref: [18]	

^a K_i values reported in the cited references were uniformly derived by fitting inhibition data into the Cheng–Prusoff model of competitive inhibition. Assays were performed with a 5 component PRC2 complex (consisting of EZH2, SUZ12, EED, RBBP4, and AEBP2), high SAM concentration of 7.5 μ M, utilizing histone peptide substrates at pH 8.0 and at room temperature

Partial charges were derived from electrostatic potential calculated at HF/6-31G* level of theory, by performing a RESP fitting procedure [53] with R.E.D. Tools IV [54]. Three conformers of each ligand were used to derive the charges. Equilibrium geometries of all structures were optimized using ORCA 2.9 [55] (Hartree–Fock method, 6-31G* basis set) prior to submission to R.E.D. Server [56] where the final quantum–mechanical calculations were interfaced to Gaussian 09 [57].

To solvate the system, a pre-equilibrated rectangular TIP3P water box was used, large enough to leave a 10 Å solvent padding in each direction of the simulation box. Protonation of side-chains was identical to that used in

simulations of EZH2 in the apo form, reflecting predictions from the PROPKA web server as described in section “EZH2 modeling”. Specifically, all aspartate and glutamate side-chains were deprotonated, whereas all lysines and arginines were protonated. Histidines were singly protonated. Based on visual inspection of potential hydrogen bonding sites, H629 was delta protonated (Amber residue name HID) and H689 and H706 were epsilon protonated (Amber residue name HIE). A single Na^+ ion was added to neutralize the system, except for the complex of 126 and EZH2, where no neutralization was required. The final simulation box contained on average 24,500 atoms, including 7,500 water molecules and, prior

to minimization and equilibration, had the dimensions of $68 \times 60 \times 68 \text{ \AA}$.

After parametrization, systems were subjected to 5,000 steps of conjugate gradients minimization with the coordinates of all protein and ligand atoms fixed, followed by an additional 5,000 steps of conjugate gradient minimization with a harmonic restraint of $10 \text{ kcal (mol \AA}^2)^{-1}$ imposed on protein heavy atoms. Minimized systems were then gradually heated from 0 to 298 K in 1 K per 0.3 ps increments, keeping the protein backbone and ligand heavy atoms weakly restrained with a $1 \text{ kcal (mol \AA}^2)^{-1}$ harmonic restraint. Heating was followed by a short 1 ns equilibration run during which SET domain backbone atoms were still weakly restrained whereas post-SET domain and ligand were allowed to move freely. After the initial 1 ns, further 4 ns of unrestrained equilibration were carried out in the isothermal-isobaric ensemble. Complex equilibration was monitored by plotting the RMSD of protein backbone, binding site heavy atom, and ligand heavy atom coordinates. If satisfactory stability was achieved after the initial equilibration period, a 4 ns production run in the isothermal-isobaric ensemble was initiated which was used for binding free energy evaluation. Otherwise, equilibration was continued until RMSD metrics showed adequate stability prior to beginning the 4 ns production run. In order to investigate whether longer simulation times would lead to changes in inhibitor binding modes or could affect the binding free energy estimates, after the first 4 ns production runs, velocities were restarted and a further 10 ns simulation was carried out for each complex.

Simulations of EZH2 point mutants were carried out in the same way. Initial structures were prepared by mutating the appropriate residues of wild-type EZH2-containing complexes using the Mutagenesis Wizard of PyMOL [58]. The conformation of the replacement side-chain was chosen from the Dunbrack library [59] (except for glycine), visually assuring that the mutation does not cause major steric clashes (A687V) and maintains the previous orientation of the side-chain (Y641F). Changes in binding site volume were evaluated using MDpocket [60]. Structures obtained from MD trajectories were first aligned with respect to backbone atoms and 500 evenly spaced snapshots were extracted from the final 4 ns of simulations using the ptraj utility from AmberTools. The cavity detection and characterization method is detailed in the original publication of MDpocket [60]. Briefly, pocket detection relies on the Voronoi tessellation procedure of Fpocket [61] that involves placement of alpha-spheres into protein cavities. An alpha-sphere represents a sphere with a defined minimum and maximum radius whose outer surface makes contacts with four protein atoms without having any atoms in its core. Each placed alpha-sphere is

assigned to the closest intersection point of a 1 \AA spaced grid mapped to the first provided snapshot. Number of alpha-spheres assigned to each grid point is counted for every snapshot and this count is normalized by the number of snapshots in the provided trajectory to obtain a density map. Minimum sphere radius of 3 \AA , maximum of 6 \AA , and an isodensity map with an isovalue of 8 were used as parameters for mapping the cavities. Isodensity of 8 corresponds to a network of neighboring grid points where each point has at least 3 alpha-sphere centers in 8 \AA^3 of surrounding space in each evaluated snapshot. This choice of parameters allowed clear distinction of the ligand binding site from the neighboring cavities. Ligand binding site volumes of each mutant protein were normalized by the median volume of wild-type EZH2 to allow easier monitoring of changes in pocket size induced by the examined mutations.

General MD simulation protocol

All molecular dynamics simulations were carried out with NAMD 2.8b3 [62] under periodic boundary conditions in all spatial coordinates. A timestep of 2 fs was used for integrating the equations of motion and for evaluating all non-bonded interactions (van der Waals, short- and long-range electrostatics). All covalent bonds involving hydrogen atoms were restrained using the SHAKE algorithm. Real space cutoff of 12 \AA was applied to the calculation of van der Waals and short-range electrostatic forces, with an analytical tail correction used to compensate for truncation of the Lennard–Jones potential. Long-range electrostatics were treated by particle-mesh Ewald summation with a grid spacing of 1 \AA . Temperature was maintained using Langevin dynamics with a damping coefficient of 2 ps^{-1} . Nose–Hoover Langevin piston with an oscillation period of 500 fs and damping over 200 fs was enabled to maintain constant pressure conditions with a target of 1.01 bar.

Trajectory visualization and analysis was performed using VMD 1.9.1 [63] and gcrarna [64]. Charts were prepared with Gnuplot [65].

MM/GBSA calculations and alanine scanning

Binding free energies of the studied inhibitors were estimated using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method as implemented in the MMPBSA.py program [66] of AmberTools 12. The MM/GBSA method [67] provides a framework for estimating free energy changes associated with ligand binding through an end-point approach. The method has been reviewed recently and has been successfully applied to a number of systems of biological interest [68]. The calculations were performed using a single trajectory approach on 500

snapshots from the 4 ns production runs. The relative binding free energy estimates are reported as $\Delta G_{\text{MM/GBSA}}$. We note that entropic contributions to binding free energy were neglected, with the assumption that they are approximately the same for all the analyzed complexes. Further details on MM/GBSA calculations are provided in Supplementary Information.

MM/GBSA calculations on the extended 10 ns production run trajectories were performed on evenly spaced snapshots collected every 10 ps. Averages over increasingly longer periods (e.g. 0–1, 0–2, 0–3 ns...) were calculated and are reported in Supplementary Table 3.

Alanine scanning functionality of MMPBSA.py was used to evaluate the contribution of specific side chains to the binding free energy of the inhibitors by sequentially mutating chosen residues to alanine within the trajectory used for MM/GBSA calculations, as described above. In order to facilitate the use of an existing trajectory for binding free energy estimates on mutant proteins, an assumption is made that the change of a larger residue to alanine will have a negligible effect on the protein conformation [66]. This assumption has been shown as reasonable in qualitatively reproducing experimental observations in several systems [69]. The following 13 residues were mutated: V621, W624, Y661, C663, F665, N688, S690, F723, Y726, Y728, I739, E742, and M743. The change in MM/GBSA energy is reported as:

$$\Delta\Delta G_{\text{MM/GBSA}} = \Delta G_{\text{MM/GBSA}}^{\text{mutant}} - \Delta G_{\text{MM/GBSA}}^{\text{wild-type}}$$

It should be noted that the obtained $\Delta\Delta G$ values are relative and therefore primarily useful only in identifying residues with key contributions to stabilizing the ligand in the binding site.

Conformational analysis

Analyses of conformational preferences of inhibitors 126 and G3K were performed using metadynamics [70] in its well-tempered form [71], as implemented in PLUMED 1.3 [72]. Bias was added to two torsional collective variables (CVs) corresponding to rotation of two bonds connecting the central amide to the pyridone moiety. Gaussian width of 0.35 radians was used for both CVs with an initial Gaussian deposition rate of $0.3 \text{ kcal mol}^{-1} \text{ ps}^{-1}$ which was gradually decreased on basis of the adaptive bias with a factor of six ($\Delta T = 1,490 \text{ K}$).

Ligands were solvated in a pre-equilibrated TIP3P water box with a padding of at least 18 Å. Systems were heated to 298 K and equilibrated in the isothermal-isobaric ensemble for 100 ps followed by an additional 100 ps in the isothermal-isochoric ensemble. Metadynamics simulations were carried out in the isothermal-isochoric ensemble

with a patched version of NAMD 2.8b3, with the remaining parameters the same as detailed under section “[General MD simulation protocol](#)” General MD simulation protocol. Based on deposited Gaussians’ height and stable differences in the reconstructed free energy surface, it was estimated that the simulations were converged after 20 ns of sampling.

Results and discussion

Developing a model of EZH2 with a fully formed cofactor binding site

EZH2 achieves catalytic competence only upon association with Polycomb proteins SUZ12 and EED, which together with histone-binding proteins RBBP7 or RBBP4 form the four essential components of a functional PRC2. PRC2 acts as a major gene silencer and has an evolutionarily conserved role in organisms ranging from plants and insects to humans [73, 74]. It mono-, di-, and trimethylates lysine 27 on histone H3 (H3K27), H3K27me3 being an epigenetic mark associated with transcriptionally repressed chromatin. Despite its important biological role, structural information on PRC2 is, at present, restricted to an electron microscopy structure providing an insight into the architecture of the complex [75] and two X-ray crystallography structures of EZH2 truncated to CXC and SET domains. These truncated structures have been shown, however, not to bind either SAM or the potent EZH2 inhibitors [25, 26]. Given that portion of the cofactor binding site formed by the SET domain appears fully structured, it has been hypothesized that a possible explanation for this finding is in the unexpected orientation of the post-SET domain [25]. Post-SET domain structure is largely unresolved in both structures but it seems to project away from the cofactor binding site and towards the CXC domain. This is in contravention to all of the other resolved structures of PKMTs, where post-SET contributes to cofactor stabilization and forms one side of the binding site. Availability of these PKMT binary and ternary complex structures offers a number of templates which can serve as starting points in modeling the approximate position and structure of EZH2 post-SET domain in a state relevant for ligand binding. Importantly, in contrast to post-SET domain which exhibits significant flexibility, conserved part of the cofactor binding site is not expected to undergo significant rearrangement between the apo state, binary and ternary complexes [15]. This suggests that available EZH2 structures could serve as basis for models of the cofactor binding site with relatively high reliability.

To generate a model of EZH2 with the cofactor binding site fully-formed, we modeled the post-SET domain based

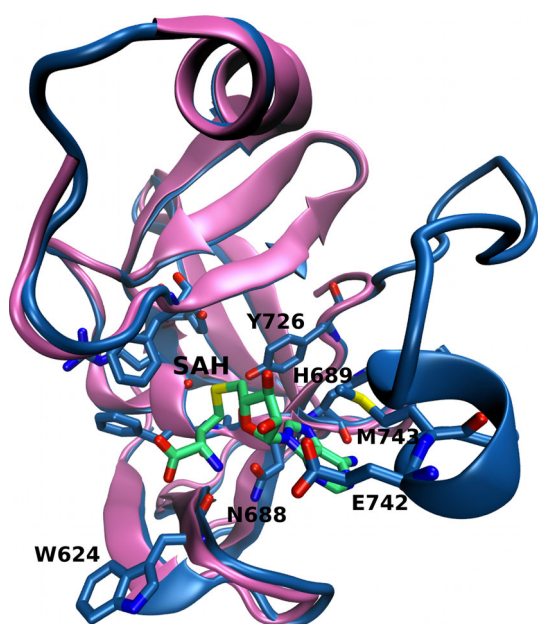


Fig. 1 Overlay of a truncated crystallographic structure of EZH2 (PDB entry: 4MI0), shown in *pink*, and a model of the EZH2-SAH binary complex, shown in *blue* with the cofactor product SAH depicted *green*. SAH is able to form the highly conserved hydrogen-bonding network with backbone atoms of residues W624 and H689, and with the sidechain of N688. Post-SET domain forms one side of the cofactor binding site, with residues E742 and M743 providing additional stabilization and shielding from the bulk solvent

on homology to the corresponding domains of SETD7 and SETD8, which like EZH2—do not have a cysteine-rich post-SET domain. Although this domain in EZH2 is only 19 residues long, due to low sequence similarity to the used templates and high expected degree of flexibility, we needed to refine the initial model by considering a greater number of physically viable structures in order to obtain a likely configuration with respect to the cofactor binding site. To achieve this, we took advantage of a well established fact that conformation of the cofactor product SAH is essentially identical in all resolved PKMT structures [16]. By restraining SAH in this conserved conformation within the formed part of the binding site, the post-SET domain was allowed to adjust to the remainder of the structure without significant overall deviation from the initial experimentally determined SET domain coordinates. Dual boost aMD was used to enhance the sampling of conformational space by simultaneously and independently improving the exploration of torsional degrees of freedom and accelerating diffusive motions. Application of aMD still allows relevant minima to be properly sampled [40].

Figure 1 shows an alignment of the truncated crystallographic structure of EZH2 and the putative binary complex that was extracted from aMD simulation. The active site is largely identical between the two structures, but in the latter, instead of projecting away from the cofactor

binding site, the post-SET domain is positioned so that it can participate in stabilization of both SAH and ligands whose binding mode was subsequently studied. However, the established model does not represent a catalytically competent structure of EZH2, given that I-SET domain remains positioned in a way that blocks access for the histone substrate. Nevertheless, since the cofactor binding site is fully formed, as evidenced by contacts made to SAH, and is unlikely to undergo significant rearrangement, this model can be seen as an appropriate starting point for the study of inhibitor binding. With this respect, biochemical evidence that the studied inhibitors do not displace or compete with histone substrates is relevant given that otherwise the structure of the lysine-binding channel would also have to be modeled in the relevant state.

It should also be noted that although the two reported EZH2 structures do not differ in the region of the cofactor binding site, a loop segment that is missing from 4MI0 (Y658, D659, and K660, which has side-chain atoms missing) is resolved in 4MI5. Based on the positioning of C β and backbone atoms of K660, it can be inferred that there are some discrepancies between the two structures in the region of this loop. We used the structure 4MI0 in this study, as 4MI5 was unavailable at the time of its conception, but retrospectively analyzed the potential impact of the noted differences. We found that models based on either structure exhibit very similar essential dynamics of the apo form (results not shown) and that this segment of the protein did not influence the dynamics of inhibitors in the binding site. None the less, for studies of substrate binding, involving the transition of I-SET to a conformation permissive of histone tail docking, this loop segment might play a more significant role and further investigations would be warranted.

Starting with the structure depicted in Fig. 1, we subsequently conducted an unrestrained classical MD run of the apo form of EZH2 to explore the stability and dynamics of the established model. This simulation was not intended to facilitate full exploration of the conformational variability of the apo form, since we wanted to keep the structure reasonably close to the one of the binary complex with SAH. Rather, it was intended to relax the structure sufficiently to allow docking of the inhibitors. During a 15 ns simulation, the apo form of the structure remained stable, with principal motions of the protein restricted to opening and closing movements at the entrance to the cofactor binding site, as illustrated in Fig. 2a. The post-SET domain was observed to be highly mobile, as were the residues forming a short loop connecting an antiparallel beta-sheet to an alpha-helix within the I-SET domain; and those at the tip of the beta-hairpin forming one side of the binding site. Movements of the latter residues (sequence numbering 620–623) were also found to be weakly

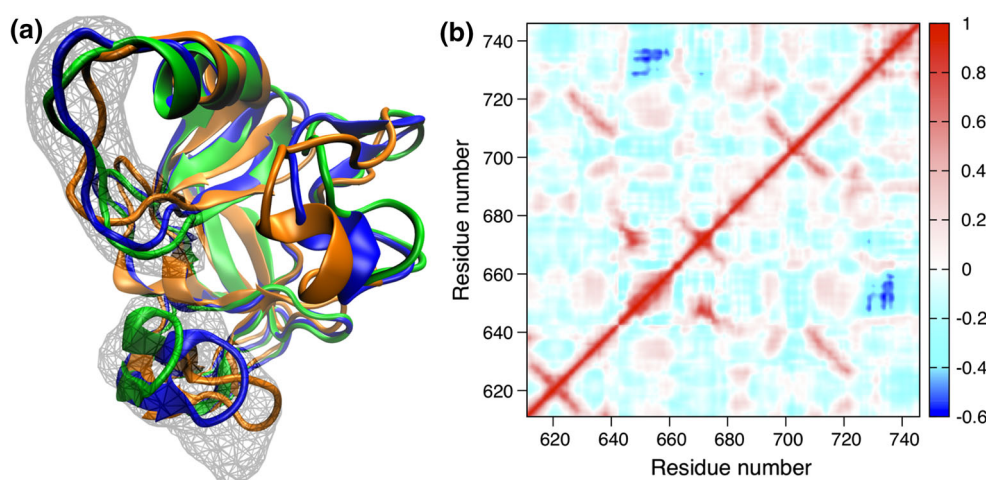


Fig. 2 Dynamics of EZH2 in the apo form, as observed during a 15 ns MD simulation. **a** *Orange ribbons* depict the cofactor-bound structure and *green* the configuration with greatest deviation from this initial structure. Structure shown in *blue* is the representative structure of the apo form, considering C_{α} atoms only. *Wire-mesh* represent volumetric occupancy maps of flexible residues at the entrance to the

cofactor-binding site. **b** A dynamical cross-correlation map of C_{α} atom fluctuations. Color range corresponds to the degree of correlation—*red* areas depict highly correlated motions, *blue* areas anticorrelated motions, and *white* indicates no correlation in movement of the corresponding residues

anticorrelated to the movements of C-terminal residues in the post-SET domain (Fig. 2b). A stronger degree of anticorrelated motions was observed between residues forming the helical part of I-SET domain and the opposing residues belonging to post-SET, which likely has implications for histone substrate binding. In overall, the observed dynamics are permissive for a binding event and indicate significant plasticity of the cofactor binding site with respect to accommodating small molecules. We note that in the limit of a much longer run, the essential dynamics of the system would likely be those of post-SET unfolding, since in absence of both cofactor and substrate it is not expected to remain ordered.

We also examined the cofactor binding site similarity between the putative EZH2 binary complex and a number of PKMTs with structures available in the Protein Data Bank. Similarity indices can provide a general insight into the likelihood of cross-reactivity of inhibitors targeting this site. Such an analysis has already been performed by Campagna-Slater et al. [16], but at that time did not include EZH2. In addition to comparing full-length structures, we truncated all of the analyzed structures at the conserved Y726 (EZH2 sequence numbering) to examine the contribution of post-SET domain residues to binding site diversity. We found that portion of the binding site formed by the SET domain is highly conserved with all pairwise similarities greatly exceeding the 0.16 threshold relevant for the FuzCav method [47], with the exception of SETD7-Suv39H2 pair. Thus, if only the SET domain residues are accounted for, achieving a desirable selectivity profile would be a considerable challenge. However, when fully

structured sites are considered, the degree of similarity is much smaller. SAM-binding site of EZH2 is structurally closest to MLL1 and SETD8 with a similarity score of 0.18 and 0.16, respectively, with all other pairwise similarities being below the threshold. A graphical representation of the similarity matrices is available as Supplementary Fig. 2. The observed similarity pattern is consistent with previous phylogenetic analyses based on multiple sequence alignments [15]. As the post-SET domain is the principal contributor to binding site diversity, it can be inferred that the binding mode of the studied inhibitors likely needs to reflect this through appropriate contacts which contribute to the observed selectivity profile. Absence of cross-reactivity between EZH2 inhibitors and structurally closest PKMTs suggests that a desirable selectivity profile can be achieved at the FuzCav similarity threshold of 0.16, though this may not be the case for PKMTs in general.

Analysis of EZH2 inhibitors' binding mode

Understanding of how the EZH2 inhibitors bind to their target is currently based primarily on biochemical studies. The reported inhibitors are active in the nanomolar range and exhibit a remarkable selectivity profile. Specifically, they are in excess of 1,000-fold selective for EZH2 against a panel of other PKMTs, with a single exception of EZH1—against which reported selectivity ranges from 25- to 142-fold. EZH1 is a close homologue of EZH2, with the two enzymes diverging in the primary sequence of SET and post-SET domains by only 11 residues. Biochemical profiling has conclusively shown that the reported

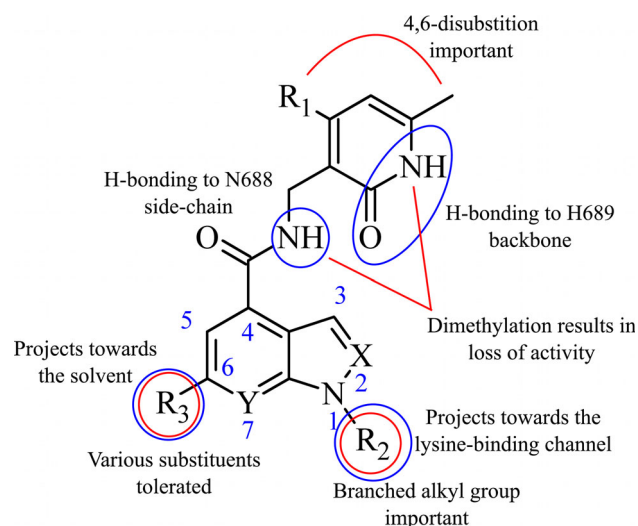


Fig. 3 A schematic representation of available SAR data and proposed interaction patterns of the studied inhibitors and EZH2. Blue markings denote interaction patterns inferred from homology models and red markings denote published SAR data. Adapted from [17, 76]. Central heterocycle numbering is used consistently throughout the text

inhibitors act in a SAM-competitive, histone substrate non-competitive manner. Based on these finding and structure–activity relationships (SAR) data, it has been suggested the inhibitors bind to the cofactor binding site in a manner which largely recapitulates the conserved hydrogen-bonding pattern of SAM [17–19]. Additional experimental evidence on the putative binding mode of EZH2 inhibitors has been the development of UNC1999 as a chemical probe targeting EZH2 and EZH1 cofactor binding sites [76]. UNC1999 was designed by varying the structure of existing inhibitors guided by a homology model of EZH2. A summary representation of SAR disclosed in previously cited studies and interaction patterns inferred from homology models is illustrated in Fig. 3.

The proposed binding hypothesis outlined in previous sections was used as a starting point for knowledge-guided docking, as it is well established in available experimental data. Starting from the initial complex geometries, we were able to obtain stable trajectories within 5–16 ns of simulation time. Consistent with tight binding of the inhibitors, dynamics of EZH2 differ when complex and apo structures are compared. Figure 4 shows a reduction in root mean square fluctuations (RMSF) of protein C α atoms which occurs with ligand binding and is pronounced within the residues forming the binding site. Fluctuations of C-terminal residues of the post-SET domain are less affected, as they remain free in solution. Residues 700–710 are usually in contact with the CXC domain so the high degree of observed fluctuations is likely a consequence of EZH2 truncation. As these residues are distant from the binding

site, the overall stability of the simulated complexes was not affected.

To evaluate whether the obtained binding modes could account for experimentally determined binding affinities, we conducted MM/GBSA calculations whose results are graphically represented in Fig. 5 and are fully summarized in Supplementary Table 2. The correlation coefficient between ΔG_{exp} and $\Delta G_{MM/GBSA}$ for all 13 inhibitors is fair with a value of $r = 0.74$ (Spearman correlation coefficient $\rho = 0.72$). A closer analysis revealed two major outliers (labeled in Fig. 5); the most potent inhibitor 126 has a $\Delta G_{MM/GBSA}$ in the rank of tenfold weaker inhibitors, whereas conversely, the least active inhibitor E1Z ranks with tenfold more active compounds. When these two results are omitted, the correlation is somewhat better with a value of $r = 0.88$ ($\rho = 0.88$). A more detailed discussion on specific binding modes presented later in this section provides some additional insight into potential reasons for the deviation of these two results from the general observed trend. For clarity of presentation, we conclude for now that the obtained correlation is comparable to results of benchmarking MM/GBSA calculations [77] (r in the range of 0.5–0.7) performed on simulations started from complexes resolved by crystallography. Thus, the fair agreement between experimental data and binding free energy estimates provide support that the binding modes we obtained likely reflect the bioactive ones and can be used as basis for a more in-depth analysis.

We note that extending simulations beyond the 4 ns production runs did not lead to significant changes in either the inhibitor binding modes or the free energy estimates. The results of 10 ns production runs are summarized in Supplementary Table 3. Supplementary Fig. 3 also illustrates that the correlation coefficient between ΔG_{exp} and $\Delta G_{MM/GBSA}$ is converged after approximately 4–5 ns and is essentially equal to that obtained using the original 4 ns runs.

Figure 6 illustrates a typical binding mode of the studied inhibitors, with G3K depicted occupying the cofactor binding site of EZH2. The pyridone moiety interacts with backbone atoms of H689 dominantly through formation of a hydrogen bond between the pyridone nitrogen and backbone oxygen. At a distance cut-off of 3 Å and an angle cut-off of 25°, the geometrical prerequisites for formation of this bond are satisfied throughout approximately 35 % of the production run trajectories. Pyridone oxygen to H689 backbone nitrogen H-bond has a negligible occupancy in the instance of indazole-based structures, where the formation of an H-bond to the nitrogen in position 2 of the indazole core seems more favored. Post-SET domain folds over the pyridone structure, enclosing this part of the inhibitor in a largely hydrophobic pocket formed by residues I739, E742, and M743, with V621 contributing to the

Fig. 4 Root mean square fluctuations (RMSF) of protein C α atoms calculated from 4 ns production simulations of inhibitor-bound EZH2 and the last 4 ns of apo-EZH2 simulation. Protein motions diminish with inhibitor binding which is particularly pronounced in the region of the cofactor binding site. Post-SET domain, however, retains a significant degree of flexibility

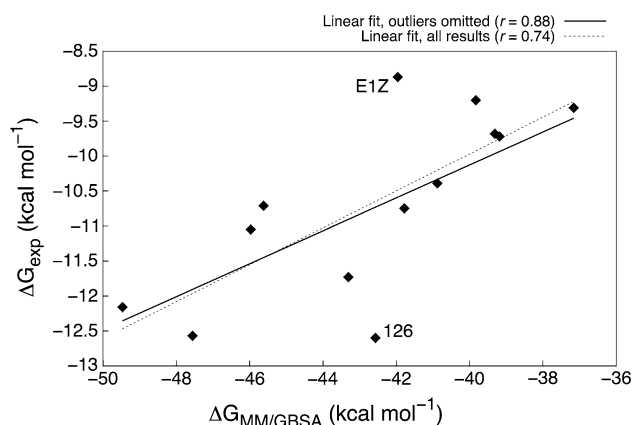
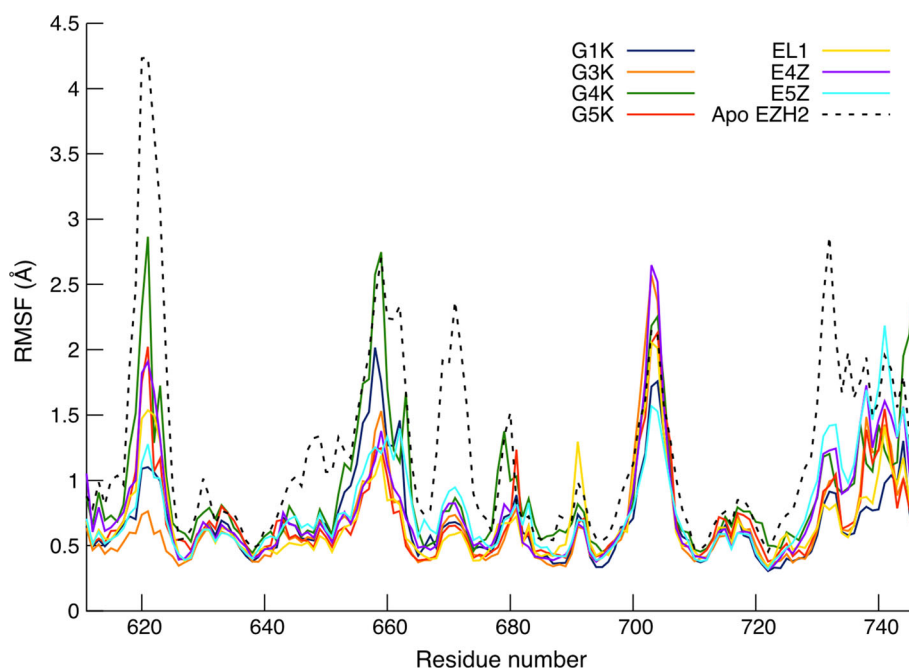


Fig. 5 Plot illustrating the agreement between binding free energy estimated using the MM/GBSA method ($\Delta G_{\text{MM/GBSA}}$) and experimentally determined values (ΔG_{exp}). Data points are depicted as *black rhomboids*. Experimental binding free energies were calculated from reported K_i values according to the relationship $\Delta G^\circ = RT \ln K_i$, taking the temperature of 25° C (298 K), at which assays were performed. Two major outliers, representing the results for inhibitors 126 and E1Z, are labeled in the graph. For a full summary of results, refer to Supplementary Table 2

lower segment of the binding site. Isopropyl group in position 1 of the indazole core projects towards the end of the lysine-binding channel and interacts favorably with Y726 and F665 through dispersion interactions. The indazole core and substituent in position 6 of G3K interact with residues C663, Y661, and F665. C663 is positioned so that the sulfur atom is approximately 3.7 Å above the face of the aromatic core, geometrically consistent with a sulfur-arene interaction [78]. Substituent in position 6 projects

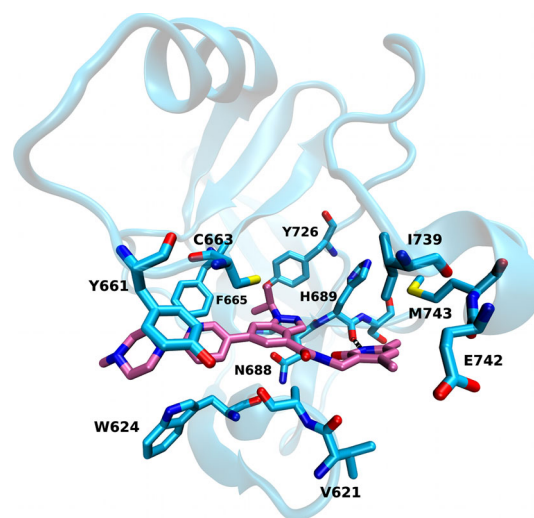


Fig. 6 Binding mode of G3K (depicted in *pink*) in the cofactor-binding site of EZH2 (depicted as a semi-transparent *cyan ribbon*), demonstrating the principal interaction features identified for most of the studied compounds, as described in the main text. Residues lining the binding site (within 3 Å of any ligand atom) are shown in *darker cyan*. F723 and Y728, as well as hydrogen atoms, have been omitted from the representation for clarity

towards the solvent and is only partly enclosed in the binding site, consistent with experimental evidence that larger alkyl substituents of the piperazine ring are tolerated [76]. The pyridine ring interacts largely with Y661 in a parallel-displaced configuration, whereas the piperazine ring retains more mobility, transiently making contacts with W624 side-chain as well. The observed pattern of interactions is quantitatively reflected in MM/GBSA results which demonstrate that the principal favorable enthalpic

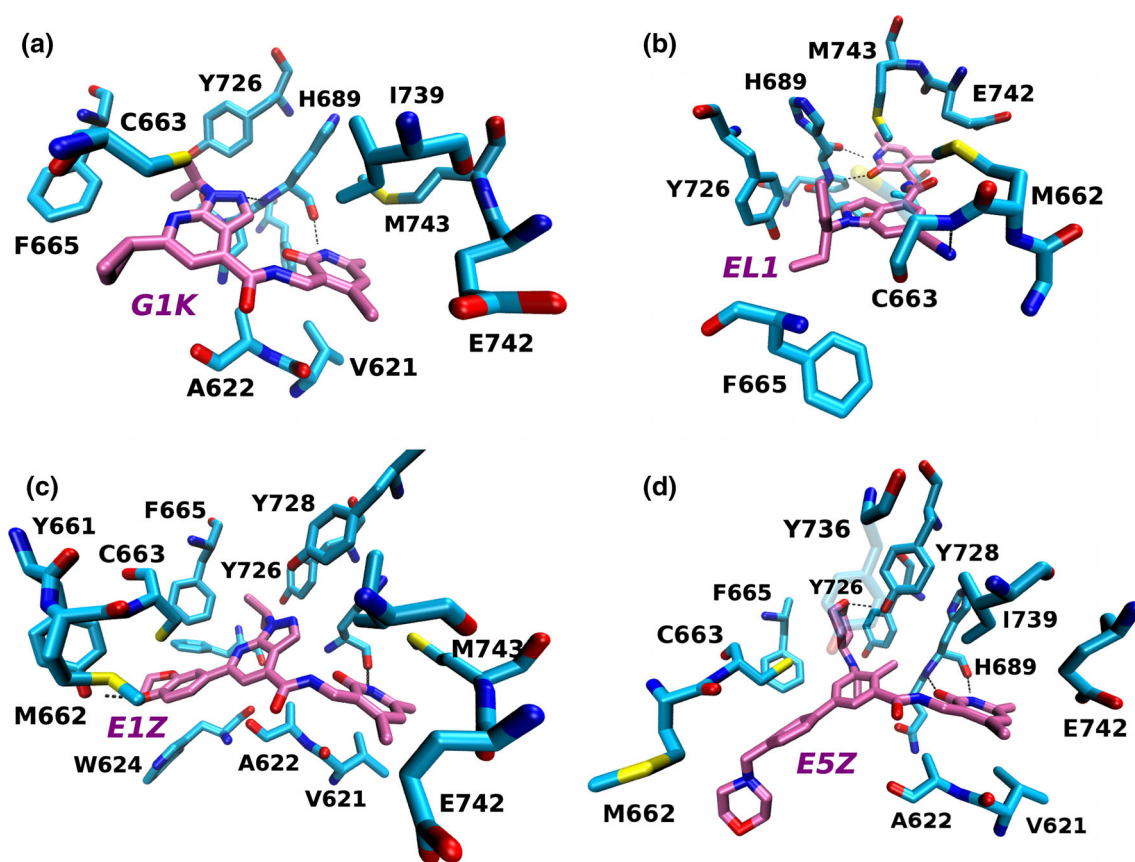


Fig. 7 Binding modes which feature additional interactions differing from the general pattern illustrated in Fig. 5. Ligands are shown in pink and binding site residues in cyan, with hydrogens omitted for

clarity. Putative hydrogen bonds are depicted as dashed black lines. Panels illustrate representative snapshots of: **a** G1K; **b** EL1; **c** E1Z and **d** E5Z

component of binding energy for G3K and closely related compounds (G5K, E2Z, E3Z, and E4Z) are van der Waals interactions.

Although the described interaction pattern is fairly conserved within many of the studied complexes, some notable differences were also observed and are illustrated in Fig. 7. Compounds G1K (Fig. 7a) and G2K do not exploit the possibility of interaction with Y661, as the cyclopropyl group in position 6 of the indazole core is capable only of limited interactions with C663 and F665—a binding mode feature clearly reflected in less negative binding free energy.

EL1 (Fig. 7b) has a unique binding mode, as compared to other of the analyzed inhibitors, characterized by additional H-bonds formed by the nitrile moiety to backbone nitrogens of C663 and M662. These two H-bonds alternate in stabilizing the inhibitor, with an occupancy of about 15 % each. There is a slight overall shift of the inhibitor structure in the direction of the I-SET domain as a consequence of this hydrogen-bonding, and a larger inclination of the indole core. This in turn permits M743 from the post-SET domain and V621 to ‘sandwich’ the pyridone

moiety. The added stabilization results in less fluctuation of the pyridone ring, also reflected in greater occupancy of the anchoring H-bond to H689 (approximately 54 %). Together with the favorable dispersion interactions of the pentane chain, these specific interaction features contribute to the large ligand efficiency of EL1.

E1Z (Fig. 7c) contains a 1,4-benzodioxan ring system connected to a pyrazolopyridine core, and was found to induce a significant strain on the flexible loop forming the upper portion of the binding site. The benzodioxan ring interacts with C663, Y661, M662, and W624 which results in enthalpically favorable binding mode, as shown by highly negative $\Delta G_{MM/GBSA}$. However, within the context of the single trajectory approach used to approximate the binding free energy, the energetic cost of achieving the unusual packing of side-chains around the benzodioxan core is not accounted for and therefore likely results in the discrepancy to experimental data, in addition to the entropic penalty which we neglected.

E5Z (Fig. 7d), the only studied structure not containing a fused ring central core, recapitulates most of the previously discussed interaction features, with principal

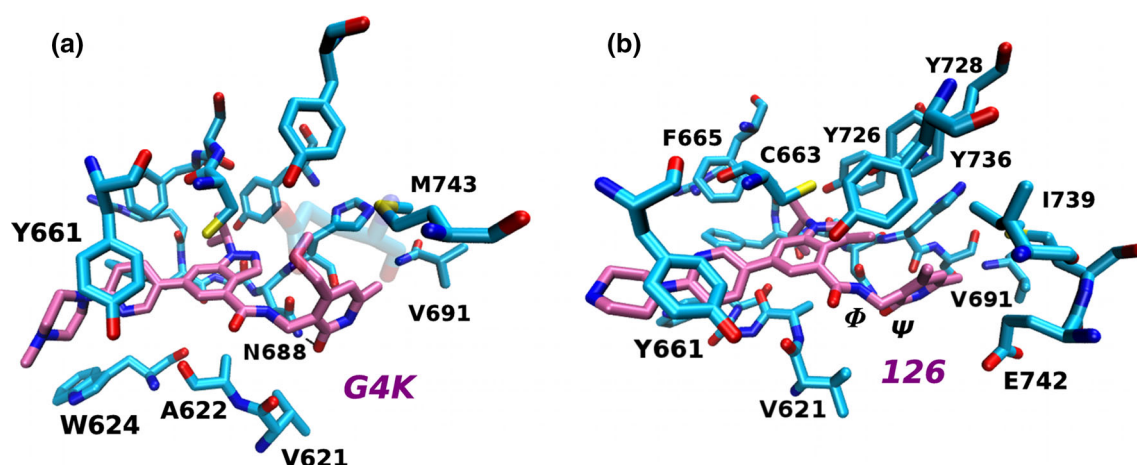


Fig. 8 Binding modes of **a** G4K and **b** 126; resulting from conformational rearrangement and a shift in N688 side-chain orientation. Ligands are shown in *pink* and binding site residues in *cyan*. Rearrangement results primarily from rotation about two torsion angles labeled Φ and Ψ in **b**

differences in the region around the site where methyl group transfer would typically occur. The oxane ring attached to a ternary amine is involved in contacts with Y726, Y728, and Y736; Y728 is centrally positioned and forms a hydrogen bond to the inhibitor with an occupancy of 50 %, whereas Y726 and Y728 flank the oxane ring from opposite sides. This positions the oxane ring largely parallel and adjacent to the putative position of the lysine-binding channel and results in a slight reorientation of the pyridone moiety. Similarly to EL1, in this orientation the pyridone ring becomes more enclosed by I739 and V621 and fluctuates less. The morpholine ring, on the other hand, is very mobile and makes contacts with M662 and W624 in an alternating manner.

Finally, G4K, G6K, and 126 were found to assume a binding mode distinct from those observed for the remaining inhibitors. We initially had difficulties in obtaining stable complex trajectories within the 4 ns equilibration period, so we extended the simulations. Starting from the initial complex geometry which was largely the same for all inhibitors, after approximately 10 ns of equilibration we observed significant torsional rearrangement of two rotatable bonds connecting the pyridone moiety to the central indazole core, and a subsequent shift in N688 rotamer conformation. The resulting complex proved stable and is illustrated in Fig. 8. As can be seen, the propyl group at position 4 of the pyridone moiety projects upward towards E742, M743, and Y736 with the post-SET domain shifted slightly outwards, whereas the methyl group at position 6 is in contact with V691. Pyridone carbonyl oxygen and nitrogen from the central amide linkage hydrogen-bond to the alternative rotamer of N688. Additionally, H689 backbone oxygen and pyridone nitrogen can interact via a water molecule or a network of two adjacent water molecules. The remainder of the structure

retains the canonical interactions with EZH2. The principal driving force behind the transition to the observed binding mode seems to be the minimization of propyl group solvent exposure, which is then followed by the rotamer shift. Chosen snapshots from the trajectory and graphs depicting RMSD evolution and number of water contacts to the propyl group are provided in Supplementary Fig. 4.

Since we were uncertain whether the alternative identified binding mode was potentially the more stable binding mode for all of the structures, we conducted additional simulations. Specifically, we extended the simulation of each complex to a cumulative time of 12–20 ns but did not observe a similar transition for any of the inhibitors except G4K, G6K and 126. We also attempted to recreate the ‘canonical’ binding mode (e.g. that of G3K) for G4K, G6K, and 126. Repeated simulations from different random seeds either did not result in a stable complex or proved to be less favorable in MM/GBSA calculations (up to 9 kcal mol⁻¹ greater $\Delta G_{\text{MM/GBSA}}$). Furthermore, we took the rearranged structure of EZH2, manually docked G5K to replicate the binding mode of G6K, and then started a simulation from this initial complex geometry. We did not observe inhibitor dissociation during 10 ns of simulation, but MM/GBSA calculations determined the alternative binding mode is approximately 7 kcal mol⁻¹ less favorable. Thus, although transiently stable, this binding mode is energetically not the most stable for G5K and related compounds. In addition, the side-chain of N688 usually forms a hydrogen bond to the backbone oxygen of W624. Simulations started with the alternative rotamer and smaller inhibitors, such as G1K, result in dissociation of the complex as the beta-hairpin is no longer stabilized by the H-bond nor by the smaller inhibitors, as they do not ‘bridge’ the full depth of the binding site.

Some additional support to the hypothesis of distinct binding modes for G4K, G6K, and 126 can be found in a study published during the preparation of this manuscript. Van Aller et al. [27] described the binding kinetics of 126, G5K, and G6K to unactivated and activated forms of the EZH2/PRC2 complex. They found that while inhibition of activated EZH2/PRC2 by 4,6-dimethylpyridone-bearing G5K was rapidly reversible, 4-propyl-6-methylpyridone-bearing G6K and in particular 126 had long residence times (43 and 62 min, respectively). This seems to suggest that small differences in inhibitors' structure result in significant differences in both potency and binding kinetics within the studied compound class, potentially reflecting the possibility of unique binding modes we observed during our simulations. We do note, however, that the published results are concerned with allosterically activated form of PRC2 and therefore should not be considered as fully transferable to our EZH2 model. In particular, conformational rearrangement of the inhibitor and the binding site upon PRC2 activation is not something that was addressed in our study.

Interestingly, with inhibitors bound in the alternative orientation, protein C_{α} atoms RMSF is somewhat increased (as can be seen in Fig. 4 for G4K), indicating greater residual mobility. This may suggest that the three potent inhibitors stably bound to EZH2 in the alternative orientation profit from a reduced entropic penalty. When considered in the context of EZH2 dynamics in the apo form, most inhibitors seem to stabilize the enzyme in a state close to that of the binary complex. Conversely, G4K, G6K, and 126 seem to favor a state in which the entrance to the binding site is more open and flexible.

While we were able to rationalize that 4-propylpyridone substituted inhibitors could undergo the transition to the alternative binding mode to minimize solvent exposure of the longer alkyl chain, for 126 this was clearly not the case. Specifically, G3K and 126 are structurally highly similar yet seem to exhibit distinct binding characteristics. We focused initially on the additional methyl group found in position 3 of the indole core in 126. Leung et al. [79] analyzed a number of examples where introduction of a methyl group into a scaffold results in significant activity improvement. They identified favorable interactions with a subpocket of the binding site and changes in the ligand conformational preference as principal factors contributing to differences in binding free energy. As the added methyl group could sterically influence rotations of the pyridone moiety in 126, we performed an analysis of the conformational preferences of unbound 126 and G3K in explicit solvent, and used the results to evaluate the strain of their conformers in the bound state. If the 'canonical' orientation of the pyridone moiety is considered (e.g. the one depicted for G3K in Fig. 6), the binding mode of 126 differs from

that of G3K due to two primary factors arising from the presence of the additional methyl group. Namely, the central amide is preferentially offset from the plane of the indole core by an additional 20° and the added methyl group is capable of favorable interactions with residues Y728, Y736, and I739. To maintain optimal contacts with the enzyme in this orientation, however, 126 becomes conformationally strained with respect to torsions Φ and Ψ , as labeled in Fig. 8b, by approximately 3 kcal mol^{-1} . In the alternative, more stable orientation depicted in Fig. 8b, the bioactive conformation is effectively not strained while favorable contacts to the enzyme are preserved. By comparison, G3K in the 'canonical' and alternative orientations experiences roughly equal strains of 1 and $0.2 \text{ kcal mol}^{-1}$, respectively, but the 'canonical' binding mode illustrated in Fig. 6 is energetically significantly more favorable overall. The reconstructed free energy surfaces with respect to torsions Φ and Ψ and details of G3K and 126 binding modes are illustrated in Supplementary Fig. 5.

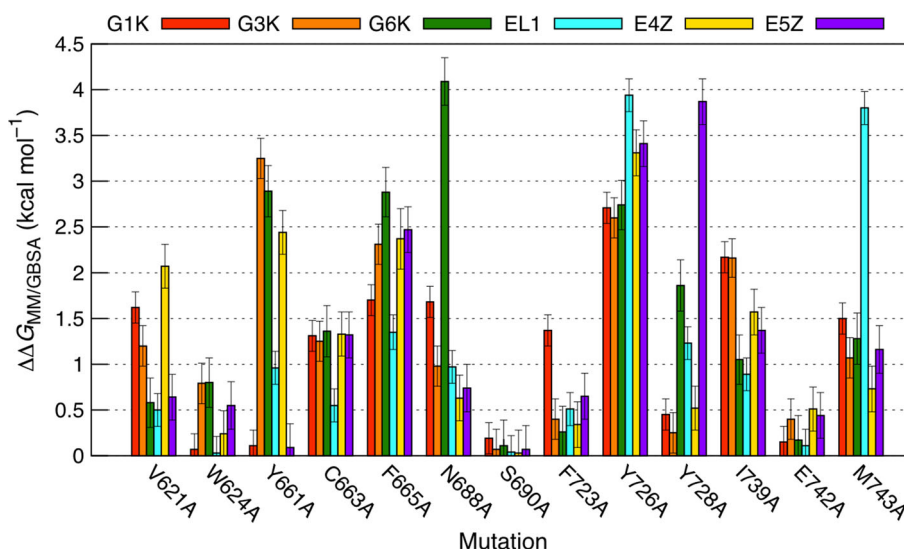
Thus, rearrangement of 126 from the 'canonical' to the alternative binding mode is driven by minimization of conformational strain whereby the rearranged conformer is $2.8 \text{ kcal mol}^{-1}$ lower in free energy in solution. The increased potency of 126 results from an interplay of several factors: (1) favorable interactions with post-SET domain residues Y728, Y736, and I739, which adjust to the additional methyl group by forming a small, hydrophobic subpocket; (2) additional branching of the alkyl chain in position 1 of the indole core; (3) reduced strain of the bioactive conformation. However, significant ligand conformational strains in the bound state are not uncommon [80, 81] and therefore it is likely the 'canonical' binding mode is at least transiently present in EZH2-126 complex.

Alanine scanning results and SAR rationalization

To quantify the contribution of individual residues lining the binding site to the binding free energy of the inhibitors, we conducted alanine scanning calculations. Detailed results for all 13 inhibitors are presented in Supplementary Table 3 and a set of representative results illustrated in Fig. 9. The representatives were chosen to reflect the influence of chemical modifications of the analyzed inhibitors to their binding modes and interactions with binding site residues. As can be seen in the plot, there are a number of common features but also clear distinctions in the relative contribution of the mutated residues to the stabilization of each of the studied inhibitors. These differences quantitatively reflect the presented analysis of the observed binding modes.

Residues participating in contacts with the pyridone moiety of the inhibitors—V621, N688, I739, E742, and M743—contribute to a similar extent to the stabilization of

Fig. 9 Results of alanine scanning calculations for a representative subset of the studied inhibitors. Height of the bars correspond to mean values from 500 evaluated snapshots, with error lines for standard error of the mean



each of the inhibitors with major exceptions of G6K (and related inhibitors G4K and 126) and EL1. As discussed previously, M743 is positioned above the plane of the pyridone moiety in EL1 and contributes significantly to its energy of binding. The large contribution of N688 to the interaction with G4K, G6K, and 126 reflects the additional hydrogen-bonding made possible by changes in ligand conformation and N688 side-chain orientation. Conversely, reduced effects of I739A mutation correspond to more extensive contacts of Y736 with the propyl substituent in position 4 of the pyridone moiety. E742A mutation has a negligible effect in all instances, consistent with the observation that it is the beta-carbon which is primarily in contact with the inhibitors.

F665 and Y726 make a large contribution to the binding of each inhibitor, consistent with numerous contacts with the hydrophobic central core and branched alkyl substituent found in position 1. For E5Z, Y728 is equally involved but through hydrogen-bonding to the oxane ring. Y661 makes a major enthalpic contribution, consistent with arene–arene interactions to aromatic parts of substituents in position 6 of the central core.

The role of C663 and M743 in ligand stabilization is particularly interesting given that these two residues are replaced in the structure of EZH1 by S664 and T744, respectively. M743-T744 switch is only the first of 4 differences occurring between EZH2 and EZH1 C-terminal segments of the post-SET domain, whereas C663-S664 substitution is fairly isolated within otherwise conserved I-SET sequence. Even if the overall fold of post-SET C-terminus is largely unaltered, replacement of methionine for a threonine would be unfavorable particularly considering that it is the sulphur atom and the attached methyl group that are principally in contact with the inhibitors.

Smaller side chain of threonine and its increased polarity likely result in under-stabilization of the pyridone moiety. The selectivity profile of EL1 in particular, viewed in context of its proposed binding mode, seems to reflect this hypothesis. EL1 is the most selective of the reported inhibitors, exhibiting 142-fold greater affinity for EZH2 compared to EZH1 (IC_{50} values of 0.009 and 1.34 μ M, respectively, which translates to a $\Delta\Delta G$ of 2.99 kcal mol^{-1}), and is also most stabilized by M743. Thus, interaction with M743 from the post-SET domain seems to be a major selectivity determinant between EZH2 and EZH1.

Replacement of cysteine for serine is, by comparison, a much smaller change. C663A mutation is energetically unfavorable which illustrates the contribution of the sulfur atom to the interaction with the aromatic system of the inhibitor core. While serine, positioned equally, could putatively be involved in hydrogen bonding to the π -system, we hypothesize this would be energetically less favorable for one of two reasons. If the orientation of C663 in the crystallographic structure of EZH2 is conserved with S664 in EZH1, this serine could effectively act a proline mimetic, hydrogen bonding to backbone carbonyl oxygen of neighboring S665 and making the needed change in rotamer orientation energetically less favored than when cysteine is involved. Alternatively, higher desolvation penalty of the more polar serine could also be seen to contribute to lower affinity for EZH1. Taken together, the observed differences in binding sites can account for the moderate selectivity profile of the studied inhibitors against EZH1. This finding further supports the validity of binding modes we observed.

Finally, results of alanine scanning and binding mode analysis allow us to rationalize the available SAR data from a structural point of view:

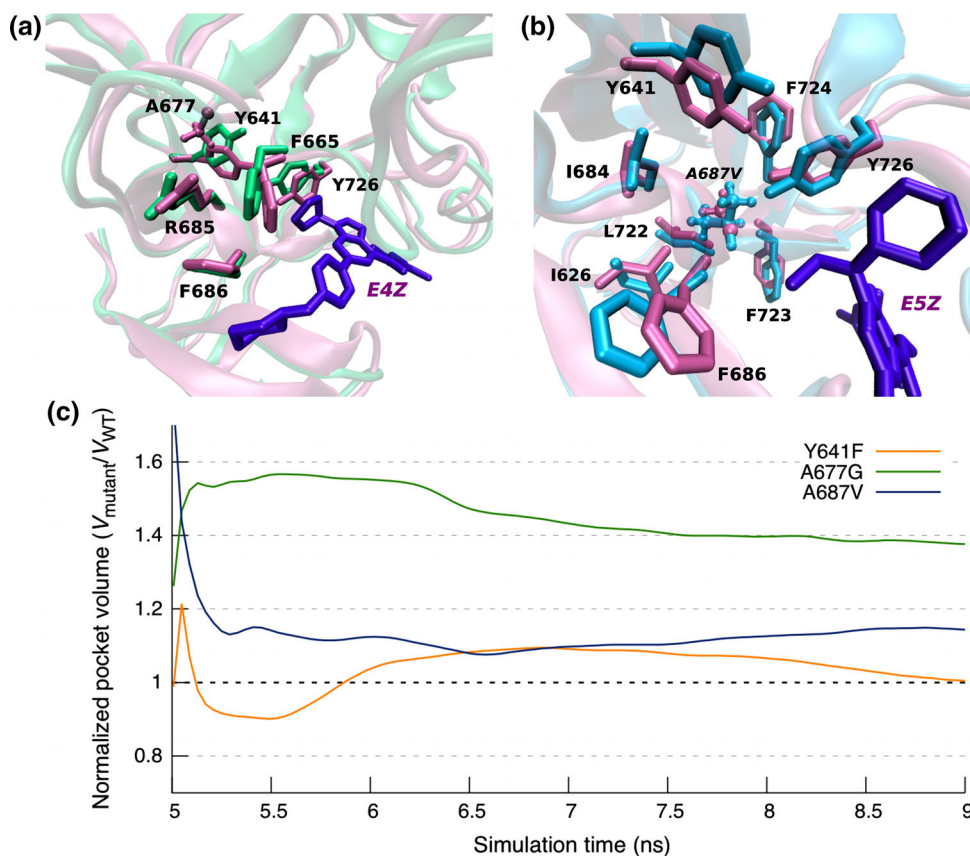
- The pyridone moiety provides a scaffold which anchors the inhibitors to the cofactor binding site primarily through hydrogen-bonding between the pyridone nitrogen and backbone carbonyl of H689. Although the carbonyl of the pyridone moiety can form a hydrogen-bond with the backbone nitrogen of H689, the formation is not essential for activity.
- Alkyl substituents in positions 4 and 6 of the pyridone moiety are essential as they make contacts with residues E742 and M743 from the post-SET domain through which they help stabilize a partially strained torsional arrangement of bonds connecting the pyridone moiety to the central indazole core. I739 contributes further by positioning above the plane of the pyridone ring. Methyl to propyl substitution at position 4 alters the orientation of the pyridone moiety, increasing the number of contacts to the post-SET domain.
- Offset planarity of the arylamide connecting pyridone and indazole cores provides optimal linker geometry while contributing electrostatically to inhibitor binding, though not consistently through directional interactions with side-chain of N688. Loss of activity in methyl disubstituted analogues (both pyridone and central amides methylated) is therefore likely a combined effect of loss of pyridone hydrogen-bonding potential and unfavorable shape changes.
- Alkyl substituents at position 1 of the central core interact through dispersion interactions with a hydrophobic subpocket formed by F665 and Y726, thus leading to activity increase with the growing number of carbons in this substituent. Optimal distance between the centre of this alkyl chain and the anchoring H-bond donor of the pyridone ring likely represents a key pharmacophoric feature of the studied compounds.
- The central aromatic core is involved in contacts with C663 and F665 and is enclosed in a prevalently hydrophobic environment. Nitrogen in position 7 of the azaindazole-based structures is not capable of any polar interactions with the binding site but adds to the desolvation penalty which accounts for activity gains when it is replaced by an indazole core. Although nitrogen in position 2 of the indazole core can hydrogen-bond to the H689 backbone, no loss in activity results from switching to an indole-based structure possibly due to its π -exceedent character.
- An aromatic ring directly attached to position 6 of the central core affords significant activity improvements as it is optimally positioned to interact with Y661 side-chain. Fused bicyclic systems in this position do not provide activity gains as they induce a significant conformational strain of the loop forming the upper portion of the binding site.

Influence of EZH2 point-mutations on inhibitor binding

Wild-type EZH2 preferentially monomethylates H3K27, with decreasing affinity for mono- and dimethylated substrates. Y641F(N/S/H/C), A677G, and A687V have been described as gain-of-function mutations which confer greater affinity of EZH2 towards higher-order methylated substrates [82–84]. It has been proposed that concerted activity of wild-type and mutated forms of EZH2 lead to the development of oncogenic phenotypes characterized by hypermethylation of H3K27 [85]. Since such phenotypes are recurrently found in a subset of non-Hodgkin lymphoma patients, adequate activity of EZH2 inhibitors against these mutated forms of the enzyme is therefore of considerable therapeutic importance. E4Z and E5Z have been reported to exhibit as large as fivefold greater affinity for A677G and A687V, with conserved or slightly diminished activity against Y641F(C/S/H/N) mutants [20, 21].

We studied the influence of these EZH2 point-mutations on inhibitor binding by simulating EZH2^{A677G}–E4Z, EZH2^{A677G}–E5Z, EZH2^{A687V}–E5Z, EZH2^{Y641F}–E4Z, and EZH2^{Y641F}–E5Z complexes. Each complex was analyzed with the wild-type counterpart as reference. The changes we observed during the simulations are largely consistent with predictions [25, 26] made on basis of the released crystal structures and changes in substrate affinity. Mutation of A677 to glycine, predicted to enable reorientation of H3K27me2 in a position suitable for transfer of a third methyl group, was found to lead to significant rearrangement of residues around the end of the putative lysine-binding channel. As illustrated in Fig. 10a, without the steric hinderance of alanine, Y641 tends to shift upwards, whereas F665 and Y726 repack inwards leading to a significant increase in cofactor pocket volume (shown in Fig. 10c). In turn, the alkyl substituent in position 1 of the central inhibitor core is allowed to gain deeper access to the enlarged subpocket. A687V mutation results in repacking of a series of residues, as shown in Fig. 9b. Specifically, this switch affects the overall position of L722, I684, I626, F724, F723, and F686, which all flank the valine enclosed in a hydrophobic core. The most pronounced effect, however, remains the shifting of Y641 upwards, consistent with its proposed gatekeeper function. These changes also lead to an overall enlargement of the cofactor binding site, but this effect is not as pronounced as with A677G (Fig. 10c). The effects on inhibitor binding are similar as described previously. With Y641F, we did not observe significant changes in the overall structure of the cofactor binding site, except for the shift of the mutated residue upwards, as seen previously. The volume of the binding site remains close to the median volume determined for the wild-type complexes (Fig. 10c).

Fig. 10 Changes in EZH2 cofactor binding site resulting from point-mutations. **a** Overlay of wild-type EZH2 (shown in *mauve*) and A677G mutant (shown in *pastel green*). The alanine affected by the mutation is depicted as a ball-and-stick model. Structure of E4Z is shown in *purple*. **b** Overlay of wild-type EZH2 (shown in *mauve*) and A687V mutant (shown in *cyan*). The alanine affected by the mutation is depicted as a ball-and-stick model. Structure of E5Z is shown in *purple*. **c** Graph illustrating the moving average of cofactor binding site volume during simulations of Y641F, A677G, and A687V mutated forms of EZH2. Pocket volume has been normalized by the median volume of the cofactor-binding site in wild-type EZH2—values greater than 1 indicate pocket enlargement and values smaller than 1 indicate reduction in pocket volume as compared to wild-type enzyme



The changes in EZH2 cofactor binding site we observed during the simulations did not lead to considerable changes in E4Z and E5Z binding modes. Using MM/GBSA calculations, we were unable to demonstrate differences in binding free energy greater than the statistical uncertainty of the method. The binding free energy, based on experimental data, should approximately be 1 kcal mol⁻¹ more negative for EZH2^{A677G}–E4Z, EZH2^{A677G}–E5Z, and EZH2^{A687V}–E5Z complexes, as opposed to wild-type counterparts. We hypothesize that a key element in understanding the enhanced affinity are differences in binding site hydration. Specifically, in crystal structures, Y641 forms a hydrogen bond to a water molecule that needs to be displaced during inhibitor binding. Shifting of Y641 leads to hydroxyl group projecting away from the cofactor binding site, and could make the displacement of water energetically more favorable. However, the same would hold true for Y641F, yet the binding free energy of both E4Z and E5Z is approximately 0.2 kcal mol⁻¹ more positive when mutants are contrasted to wild-type enzyme. Therefore, a more conclusive explanation requires further simulations to account for binding site hydration and water displacement during the binding process, which we have not conducted as part of this study.

Never the less, we can preliminarily conclude that the enhanced affinity of E4Z and E5Z for A677G and A687 V point-mutants results from the fact they occupy part of the binding site which undergoes expansion and reduction in polarity. These changes account for altered substrate preference but clearly, can also be exploited in the design of EZH2 inhibitors. In this respect, the substituent in position 1 of the central core seems to be a major contributor to the increased binding affinity.

Conclusions and outlook

To our knowledge, this is the first computational study to systematically address the binding of newly discovered EZH2 inhibitors to their target. We have attempted to integrate and rationalize available experimental data using molecular dynamics simulations and MM/GBSA calculations. Although the end-point method we used comes with some inherent limitations, the results we presented provide a useful basis for understanding the activity and selectivity of the studied scaffold. These results also present a good starting point for further studies using enhanced sampling methods which could provide additional insights into the binding process and potentially explain the unique

differences in binding kinetics between closely related structures.

Very recently, reports have been published detailing the discovery of a novel group of selective EZH2 SAM-competitive inhibitors based on a tetramethylpiperidine scaffold [86, 87]. This novel chemotype, also discovered through high-throughput screening, provides further opportunities for exploring the molecular requirements for selective EZH2 inhibition. We hope that our results will contribute a useful structural insight to this field of ongoing research and provide a sound basis for design and optimization of potential leads targeting the cofactor binding site of EZH2.

Acknowledgments This work was funded by the Ministry of Education and Science of the Republic of Serbia through Project Number 172009. Results presented in this work were obtained using the computational resources of the PARADOX cluster at the Scientific Computing Laboratory of the Institute of Physics Belgrade, Serbia, as part of the High-Performance Computing Infrastructure for South East Europe's Research Communities (HP-SEE). HP-SEE is a project co-funded by the European Commission (under Contract Number 261499) through the Seventh Framework Programme (<http://www.hp-see.eu/>). The authors gratefully acknowledge Dr. Jelena Randelović (University of Belgrade—Faculty of Pharmacy, Department of Organic Chemistry) for insightful discussions and technical assistance.

References

- Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21:381–395. doi:10.1038/cr.2011.22
- Luger K, Dechassa ML, Tremethick DJ (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* 13:436–447. doi:10.1038/nrm3382
- Arrowsmith CH, Bountra C, Fish PV et al (2012) Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* 11:384–400. doi:10.1038/nrd3674
- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403:41–45
- Gardner KE, Allis CD, Strahl BD (2011) Operating on chromatin, a colorful language where context matters. *J Mol Biol* 409:36–46. doi:10.1016/j.jmb.2011.01.040
- Helin K, Dhanak D (2013) Chromatin proteins and modifications as drug targets. *Nature* 502:480–488. doi:10.1038/nature12751
- Dawson MA, Kouzarides T (2012) Cancer epigenetics: from mechanism to therapy. *Cell* 150:12–27. doi:10.1016/j.cell.2012.06.013
- Jones P (2012) Development of second generation epigenetic agents. *MedChemComm* 3:135. doi:10.1039/c1md00199j
- Rodríguez-Paredes M, Esteller M (2011) Cancer epigenetics reaches mainstream oncology. *Nat Med* 17:330–339. doi:10.1038/nm.2305
- Copeland RA, Moyer MP, Richon VM (2013) Targeting genetic alterations in protein methyltransferases for personalized cancer therapeutics. *Oncogene* 32:939–946. doi:10.1038/onc.2012.552
- Copeland RA, Solomon ME, Richon VM (2009) Protein methyltransferases as a target class for drug discovery. *Nat Rev Drug Discov* 8:724–732. doi:10.1038/nrd2974
- Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* 6:227. doi:10.1186/gb-2005-6-8-227
- Qian C, Zhou MM (2006) SET domain protein lysine methyltransferases: structure, specificity and catalysis. *Cell Mol Life Sci* 63:2755–2763. doi:10.1007/s00018-006-6274-5
- Nguyen KT, Li F, Poda G et al (2013) Strategy to target the substrate binding site of SET domain protein methyltransferases. *J Chem Inf Model* 53:681–691. doi:10.1021/ci300596x
- Schapiro M (2011) Structural chemistry of human SET domain protein methyltransferases. *Curr Chem Genomics* 5:85–94. doi:10.2174/1875397301005010085
- Campagna-Slater V, Mok MW, Nguyen KT et al (2011) Structural chemistry of the histone methyltransferases cofactor binding site. *J Chem Inf Model* 51:612–623. doi:10.1021/ci100479z
- Verma SK, Tian X, La France LV et al (2012) Identification of potent, selective, cell-active inhibitors of the histone lysine methyltransferase EZH2. *ACS Med Chem Lett* 3:1091–1096. doi:10.1021/ml3003346
- Qi W, Chan H, Teng L et al (2012) Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc Natl Acad Sci* 109:21360–21365. doi:10.1073/pnas.1210371110
- McCabe MT, Ott HM, Ganji G et al (2012) EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 492:108–112. doi:10.1038/nature11606
- Knutson SK, Wigle TJ, Warholik NM et al (2012) A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat Chem Biol* 8:890–896. doi:10.1038/nchembio.1084
- Knutson SK, Warholik NM, Wigle TJ et al (2013) Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc Natl Acad Sci* 110:7922–7927. doi:10.1073/pnas.1303800110
- Chase A, Cross NCP (2011) Aberrations of EZH2 in cancer. *Clin Cancer Res* 17:2613–2618. doi:10.1158/1078-0432.CCR-10-2156
- Chang C-J, Hung M-C (2012) The role of EZH2 in tumour progression. *Br J Cancer* 106:243–247. doi:10.1038/bjc.2011.551
- Tan J, Yan Y, Wang X et al (2013) EZH2: biology, disease, and structure-based drug discovery. *Acta Pharmacol Sin* 35:161–174. doi:10.1038/aps.2013.161
- Wu H, Zeng H, Dong A et al (2013) Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. *PLoS ONE* 8:e83737. doi:10.1371/journal.pone.0083737
- Antonyam S, Condon B, Druzina Z et al (2013) Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. *PLoS ONE* 8:e84147. doi:10.1371/journal.pone.0084147
- Van Aller GS, Pappalardi MB, Ott HM et al (2013) Long residence time inhibition of EZH2 in activated polycomb repressive complex 2. *ACS Chem Biol*. doi:10.1021/cb4008748
- Tommaso PD, Moretti S, Xenarios I et al (2011) T-coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* 39:W13–W17. doi:10.1093/nar/gkr245
- Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217. doi:10.1006/jmbi.2000.4042

30. Xiao B, Jing C, Wilson JR et al (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* 421:652–656. doi:[10.1038/nature01378](https://doi.org/10.1038/nature01378)
31. Couture J-F, Collazo E, Brunzelle JS, Trievel RC (2005) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev* 19:1455–1465. doi:[10.1101/gad.1318405](https://doi.org/10.1101/gad.1318405)
32. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi:[10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084)
33. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815. doi:[10.1006/jmbi.1993.1626](https://doi.org/10.1006/jmbi.1993.1626)
34. Wu H, Min J, Lunin VV et al (2010) Structural biology of human H3K9 methyltransferases. *PLoS ONE* 5:e8570. doi:[10.1371/journal.pone.0008570](https://doi.org/10.1371/journal.pone.0008570)
35. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7:525–537. doi:[10.1021/ct100578z](https://doi.org/10.1021/ct100578z)
36. Case DA, Darden TA, Cheatham TE III et al (2012) AMBER 12. University of California, San Francisco
37. Hornak V, Abel R, Okur A et al (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinform* 65:712–725. doi:[10.1002/prot.21123](https://doi.org/10.1002/prot.21123)
38. Jorgensen WL, Chandrasekhar J, Madura JD et al (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926. doi:[10.1063/1.445869](https://doi.org/10.1063/1.445869)
39. Stacklies W, Xia F, Gräter F (2009) Dynamic allostery in the methionine repressor revealed by force distribution analysis. *PLoS Comput Biol* 5:e1000574. doi:[10.1371/journal.pcbi.1000574](https://doi.org/10.1371/journal.pcbi.1000574)
40. Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys* 120:11919–11929. doi:[10.1063/1.1755656](https://doi.org/10.1063/1.1755656)
41. Wang Y, Harrison CB, Schulten K, McCammon JA (2011) Implementation of accelerated molecular dynamics in NAMD. *Comput Sci Discov* 4:015002. doi:[10.1088/1749-4699/4/1/015002](https://doi.org/10.1088/1749-4699/4/1/015002)
42. Hamelberg D, de Oliveira CAF, McCammon JA (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J Chem Phys* 127:155102. doi:[10.1063/1.2789432](https://doi.org/10.1063/1.2789432)
43. De Oliveira CAF, Grant BJ, Zhou M, McCammon JA (2011) Large-scale conformational changes of *Trypanosoma cruzi* proline racemase predicted by accelerated molecular dynamics simulation. *PLoS Comput Biol* 7:e1002178. doi:[10.1371/journal.pcbi.1002178](https://doi.org/10.1371/journal.pcbi.1002178)
44. Wereszczynski J, McCammon JA (2012) Accelerated molecular dynamics in computational drug design. In: Baron R (ed) *Comput Drug Discov Des*. Springer, New York, pp 515–524
45. Mücksch C, Urbassek HM (2013) Enhancing protein adsorption simulations by using accelerated molecular dynamics. *PLoS ONE* 8:e64883. doi:[10.1371/journal.pone.0064883](https://doi.org/10.1371/journal.pone.0064883)
46. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291. doi:[10.1107/S0021889892009944](https://doi.org/10.1107/S0021889892009944)
47. Weill N, Rognan D (2010) Alignment-free ultra-high-throughput comparison of druggable protein–ligand binding sites. *J Chem Inf Model* 50:123–135. doi:[10.1021/ci900349y](https://doi.org/10.1021/ci900349y)
48. Kellenberger E, Muller P, Schalon C et al (2006) sc-PDB: an annotated database of druggable binding sites from the protein data bank. *J Chem Inf Model* 46:717–727. doi:[10.1021/ci050372x](https://doi.org/10.1021/ci050372x)
49. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. doi:[10.1002/jcc.21334](https://doi.org/10.1002/jcc.21334)
50. (2013) Marvin 6.0.3. ChemAxon
51. Wang J, Wolf RM, Caldwell JW et al (2004) Development and testing of a general Amber force field. *J Comput Chem* 25:1157–1174. doi:[10.1002/jcc.20035](https://doi.org/10.1002/jcc.20035)
52. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 25:247–260. doi:[10.1016/j.jmgm.2005.12.005](https://doi.org/10.1016/j.jmgm.2005.12.005)
53. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97:10269–10280. doi:[10.1021/j100142a004](https://doi.org/10.1021/j100142a004)
54. Dupradeau F-Y, Pigache A, Zaffran T et al (2010) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* 12:7821–7839. doi:[10.1039/C0CP00111B](https://doi.org/10.1039/C0CP00111B)
55. Neese F (2012) ORCA, version 2.9, an ab initio, density functional and semiempirical program package. Max Planck-Institute for Bioinorganic Chemistry: Mülheim a.d., Ruhr, Germany
56. Vanqualef E, Simon S, Marquant G et al (2011) R.E.D. server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* 39:W511–W517. doi:[10.1093/nar/gkr288](https://doi.org/10.1093/nar/gkr288)
57. Frisch MJ, Trucks GW, Schlegel HB et al (2009) Gaussian 09, revision D.01. Gaussian, Inc., Wallingford
58. Schrödinger LLC (2010) The PyMOL molecular graphics system, version 1.3
59. Dunbrack RL, Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 6:1661–1681. doi:[10.1002/pro.5560060807](https://doi.org/10.1002/pro.5560060807)
60. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* 27:3276–3285. doi:[10.1093/bioinformatics/btr550](https://doi.org/10.1093/bioinformatics/btr550)
61. Guilloux VL, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinform* 10:168. doi:[10.1186/1471-2105-10-168](https://doi.org/10.1186/1471-2105-10-168)
62. Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802. doi:[10.1002/jcc.20289](https://doi.org/10.1002/jcc.20289)
63. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38
64. Koukos PI, Glykos NM (2013) Grcarma: a fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J Comput Chem* 34:2310–2312. doi:[10.1002/jcc.23381](https://doi.org/10.1002/jcc.23381)
65. Williams T, Kelley C (2013) Gnuplot 4.6: an interactive plotting program
66. Miller BR, McGee TD, Swails JM et al (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput* 8:3314–3321. doi:[10.1021/ct300418h](https://doi.org/10.1021/ct300418h)
67. Kollman PA, Massova I, Reyes C et al (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897. doi:[10.1021/ar000033j](https://doi.org/10.1021/ar000033j)
68. Hayes JM, Archontis G (2012) MM-GB(PB)SA calculations of protein–ligand binding free energies. *Mol Dyn Stud Synth Biol Macromol*
69. Massova I, Kollman PA (1999) Computational alanine scanning to probe protein–protein interactions: a novel approach to

- evaluate binding free energies. *J Am Chem Soc* 121:8133–8143. doi:[10.1021/ja990935j](https://doi.org/10.1021/ja990935j)
70. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci* 99:12562–12566. doi:[10.1073/pnas.202427399](https://doi.org/10.1073/pnas.202427399)
71. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett* 100:020603. doi:[10.1103/PhysRevLett.100.020603](https://doi.org/10.1103/PhysRevLett.100.020603)
72. Bonomi M, Branduardi D, Bussi G et al (2009) PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 180:1961–1972. doi:[10.1016/j.cpc.2009.05.011](https://doi.org/10.1016/j.cpc.2009.05.011)
73. Margueron R, Reinberg D (2011) The polycomb complex PRC2 and its mark in life. *Nature* 469:343–349. doi:[10.1038/nature09784](https://doi.org/10.1038/nature09784)
74. Sparmann A, van Lohuizen M (2006) Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* 6:846–856. doi:[10.1038/nrc1991](https://doi.org/10.1038/nrc1991)
75. Ciferri C, Lander GC, Maiolica A et al (2012) Molecular architecture of human polycomb repressive complex 2. *eLife*. doi:[10.7554/eLife.00005](https://doi.org/10.7554/eLife.00005)
76. Konze KD, Ma A, Li F et al (2013) An orally bioavailable chemical probe of the lysine methyltransferases EZH2 and EZH1. *ACS Chem Biol* 8:1324–1334. doi:[10.1021/cb400133j](https://doi.org/10.1021/cb400133j)
77. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 51:69–82. doi:[10.1021/ci100275a](https://doi.org/10.1021/ci100275a)
78. Salonen LM, Ellermann M, Diederich F (2011) Aromatic rings in chemical and biological recognition: energetics and structures. *Angew Chem Int Ed* 50:4808–4842. doi:[10.1002/anie.201007560](https://doi.org/10.1002/anie.201007560)
79. Leung CS, Leung SSF, Tirado-Rives J, Jorgensen WL (2012) Methyl effects on protein–ligand binding. *J Med Chem* 55:4489–4500. doi:[10.1021/jm3003697](https://doi.org/10.1021/jm3003697)
80. Perola E, Charifson PS (2004) Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J Med Chem* 47:2499–2510. doi:[10.1021/jm030563w](https://doi.org/10.1021/jm030563w)
81. Hao M-H, Haq O, Muegge I (2007) Torsion angle preference and energetics of small-molecule ligands bound to proteins. *J Chem Inf Model* 47:2242–2252. doi:[10.1021/ci700189s](https://doi.org/10.1021/ci700189s)
82. Sneeringer CJ, Scott MP, Kuntz KW et al (2010) Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci* 107:20980–20985. doi:[10.1073/pnas.1012525107](https://doi.org/10.1073/pnas.1012525107)
83. McCabe MT, Graves AP, Ganji G et al (2012) Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci* 109:2989–2994. doi:[10.1073/pnas.1116418109](https://doi.org/10.1073/pnas.1116418109)
84. Majer CR, Jin L, Scott MP et al (2012) A687V EZH2 is a gain-of-function mutation found in lymphoma patients. *FEBS Lett* 586:3448–3451. doi:[10.1016/j.febslet.2012.07.066](https://doi.org/10.1016/j.febslet.2012.07.066)
85. Copeland RA (2013) Molecular pathways: protein methyltransferases in cancer. *Clin Cancer Res* 19:6344–6350. doi:[10.1158/1078-0432.CCR-13-0223](https://doi.org/10.1158/1078-0432.CCR-13-0223)
86. Garapaty-Rao S, Nasveschuk C, Gagnon A et al (2013) Identification of EZH2 and EZH1 small molecule inhibitors with selective impact on diffuse large B cell lymphoma cell growth. *Chem Biol*. doi:[10.1016/j.chembiol.2013.09.013](https://doi.org/10.1016/j.chembiol.2013.09.013)
87. Nasveschuk CG, Gagnon A, Garapaty-Rao S et al (2014) Discovery and optimization of tetramethylpiperidiny benzamides as inhibitors of EZH2. *ACS Med Chem Lett*. doi:[10.1021/ml400494b](https://doi.org/10.1021/ml400494b)