Small-molecule modulators of TRMT2A decrease PolyQ aggregation and PolyQ-induced cell death

Michael A Margreiter[1,10,†], Monika Witzenberger[2,†], Yasmine Wasser[3,†], Elena Davydova[2†], Robert Janowski[2], Carina Sobisch[3], Jonas Metz[1], Benedetta Poma[3], Oscar Paolomino-Hernandez[1,10,], Jon N Shah[5,6], Jörg B Schulz[3,6,*], Dierk Niessing[2,7,*], Aaron Voigt[3,*], Giulia Rossetti[1,3,8,9*]

1. Institute of Neuroscience and Medicine (INM-9), Forschungszentrum Jülich GmbH, Germany
2. Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Structural Biology, 85764 Neuherberg, Germany
3. Department of Neurology, RWTH University Aachen, Faculty of Medicine, RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany
5. Institute of Neuroscience and Medicine (INM-4), Forschungszentrum Jülich GmbH, Germany
6. Institute of Neuroscience and Medicine (INM-11), Forschungszentrum Jülich GmbH, Germany
7. Institute of Pharmaceutical Biotechnology, Ulm University, 89081 Ulm, Germany
8. Jülich Supercomputing Center (JSC), Forschungszentrum Jülich GmbH, Germany
9. Department of Hematology, Oncology, Hemostaseology, and Stem Cell Transplantation, Faculty of Medicine, RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany
10. Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, 52425 Aachen, Germany

† = shared first author
*= shared senior authorship

Abstract

Polyglutamine (polyQ) diseases, including Huntington's disease, are characterized by an expansion of cytosine-adenine-guanine (CAG) trinucleotide repeats encoding for an uninterrupted prolonged polyQ tract. We previously identified TRMT2A as a strong modifier of polyQ-induced toxicity in an unbiased large-scale screen in *Drosophila melanogaster*. RNAi-mediated silencing of *TRMT2A* ameliorated polyQ-induced toxicity and polyQ aggregation in flies. This work aimed at identifying and validating pharmacological TRMT2A inhibitors as treatment opportunities for polyQ diseases. An *in silico* structure- and ligand-based lead discovery approach and computer-aided drug discovery (CADD) was used to identify TRMT2A inhibitors. Additionally, the crystal structure of one protein domain, the RNA recognition motif (RRM), was determined and Biacore experiments with the RRM were performed. The identified inhibitors were functionally validated for their potency to reduce polyQ aggregation and polyQ-induced cell death in human HEK293T cells and patient derived fibroblasts. Several candidate molecules were able to decrease cell death and ameliorate the aggregation of polyQ peptides in cultured cells comparable to the TRMT2A knockdown experiments. Among these, spermidine was identified as able to cause a decrease in the abundance of polyQ aggregates in SCA3-patient derived fibroblasts. Our work provides a first step towards a pharmacological inhibition of this enzyme and indicates TRMT2A as a viable drug target for polyQ diseases.

Introduction

Polyglutamine (PolyQ) diseases are autosomal dominant inherited genetic disorders, except SMAX1/SBMA, which is recessive and linked to a mutation in the androgen receptor gene located on the X chromosome. PolyQ diseases are characterized by progressive neurodegeneration, leading to behavioral and physical impairments [1]. This disease family includes nine members: Spinobulbar muscular atrophy (SBMA or Kennedy's disease) [2], Dentatorubral-pallidoluysian atrophy (Haw River syndrome) [3], Spinocerebellar ataxias 1, 2, 3 (Machado-Joseph disease), 6, 7, 17 [4] and Huntington's disease (HD) [5]. Epidemiological data on HD, the most prevalent polyQ disease, reveals an incidence of 5-6 in 100.000 in the best-ascertained populations of Western Europe and North America [6]. All polyQ diseases feature an abnormally elongated CAG repeat expansion in the protein-coding region of the disease-linked gene [7]. As CAG encodes glutamine, the translation of mutant alleles results in a protein with an elongated, undisrupted polyQ stretch, eponymous for these diseases. The polyQ region itself is considered to be causative for disease, since they usually form cytoplasmic and/or nuclear aggregates in affected neurons. Whether these aggregates contribute to neurodegeneration remains elusive, however, the appearance of polyQ aggregates coincides with neuronal dysfunction/loss [7]. The age of onset is inversely correlated with CAG tract length, while gene-specific differences between the various polyQ diseases were reported [6]. Administration of antisense oligonucleotides (ASOs) to disease model animals suppressed expression of the disease-causing protein [8, 9]. Clinical trials with ASOs designed to degrade wild-type and mutant *Huntingtin* mRNA were started three years ago[10], but they were currently blocked based on the drug's potential benefit/risk profile for study participants. Thus, identification and targeting of novel CAG-disease modifiers can still represent a feasible therapeutic strategy.

In a high-throughput RNAi knock-down screen, tRNA methyltransferase 2 homolog A (TRMT2A) was identified as a novel modifier of polyQ-induced toxicity [11]. This protein predominantly localizes to the nucleus and converts uridine to 5-methyl uridine (m^5U) at position 54 in tRNAs [12, 13]. The latter methylation is conserved and found from yeast to man suggesting a pivotal role for this modification [14]. In *Escherichia coli*, the enzyme TrmA catalyzes the formation of m^5U_{54} [15], while in *Saccharomyces cerevisiae*, the enzyme responsible for this modification is Trm2p [15]. TRMT2A is considered to represent the ortholog of *Saccharomyces cerevisiae* Trm2p [16] in mammals. Indeed, similar to Trm2p, TRMT2A was predicted to harbor an RNA recognition motif (RRM) at the N-terminus and a methyltransferase catalytic domain (CD) at the C-terminus [12]. In vertebrates, there is a second ortholog of Trm2p, TRMT2B. TRMT2A and TRMT2B are paralogs and exert similar functions. Whereas TRMT2A is responsible for the m^5U_{54} methylation in the nucleus and cytoplasm, TRMT2B is specific for mitochondrial tRNAs [17]. Previous studies on *Escherichia coli* have shown that the presence of this methylation ensures the fidelity and efficiency of protein synthesis by stabilizing the three-dimensional structure of tRNA *in vitro* [18, 19]. So far, however, the biological function of the enzyme responsible for U_{54} methylation in mammalian species remains largely unexplored [12].

In this study, inhibitors of human TRMT2A were identified by *in silico* predictions. Two strategies of TRMT2A inhibition were followed, targeting the RRMs or the CD. In a second step, potential inhibitors were functionally validated in polyQ-disease models.

Results and Discussion

Silencing the *Drosophila* homolog of TRMT2A strongly suppressed polyQ-induced toxicity and polyQ aggregation in flies [11]. As TRMT2A is well conserved, we asked whether pharmacological

inhibition of human TRMT2A would also exert effects on polyQ-affected proteins as observed in flies.

To develop inhibitors against human TRMT2A, two parallel strategies were pursued. On the one hand, the structural and dynamic properties of the N-terminal RNA recognition motif (RRM) of TRMT2A were analyzed by solving the RRM crystal structure and leveraging this information with computational approaches. The TRMT2A RRM structure was analyzed to identify putative druggable binding pockets and subsequently enabled a structure-based virtual screening campaign. On the other hand, a ligand-based pharmacophore model across known tRNA methyltransferase inhibitors was built to find possible molecules targeting the methyltransferase catalytic domain (CD) of TRMT2A. The set of ligands identified from these two procedures was tested in cellular experiments. A selection of those targeting the TRMT2A RRM were also tested in biophysical experiments.

Structure Determination of the TRMT2A RNA Recognition Motif.

A variety of sub-fragments of TRMT2A were designed, cloned and expressed in *E. coli*. Amongst them, only a fragment from amino acid 69 to 147 that essentially included the RRM could be expressed and purified to near homogeneity. Robot-assisted crystallization screens yielded initial hits that were subsequently optimized for larger, well-diffracting crystals. An initial structure was solved from a dataset collected to 2.0 Å resolution (R_{free} = 25.14%, German Electron Synchrotron DESY; Table S1) and an improved structure was determined at 1.23 Å resolution with favorable statistics and stereochemistry (R_{free} = 17.15%, Swiss Light Source (SLS); Table S1). The structure revealed the typical arrangement of four anti-parallel β-strands and two α-helices that are packed against each other. They are arranged in a β1-α1-β2-β3-α2-β4 fold topology (Figure 1A), as already reported for other RRMs [20]. The root mean square deviation of its C_α atoms to its closest structural homolog (PDB-ID: 3vf0) is 1.5 Å. Loops that connect β1/α1 (referred to as loop 1), β2/β3 (loop 3), and α2/β4 (loop5) are located on the side featuring the β-sheet, whereas loop 2 (α1/β2) and 4 (β3/ α2) are closer to the α-helices as outlined in Figure 1A.

Typically, the β-sheet of an RRM interacts with a variable number of two to eight unpaired RNA nucleotides via two consensus regions located in the center of the two neighboring β-strands. These sub-motifs are referred to as ribonucleoprotein (RNP) 1 and 2 [21]. In the structure of this RRM known determinants of RNA-RRM molecular recognition can be distinguished, such as positively charged residues and aromatic amino acids, which form π-stacking interactions with the RNA nucleobases [22].

RNP1 was shown to be composed of K⁄R-G- F⁄Y- G⁄**A112**-**F113**⁄Y-**V114**⁄I⁄L-X-**F116**⁄Y (boldface amino acids refer to positions found in TRMT2A RNP1/RNP2; X refers to an arbitrary amino acid) [23]. RNP1 is present in the RRM β3 strand of TRMT2A and spans from aa 109 to 116. RNP2 features I⁄V⁄**L75**-F⁄Y-I⁄V⁄**L77**-X-**N79**-L; RRM RNP2 is part of the adjacent (aa 75-80) β1 strand. RRMs often display two conserved aromatic residues on RNP1 and RNP2, typically phenylalanines or tyrosines. One aromatic RRM residue is frequently π-stacking with the 3' end of the RNA, and another is inserted between two RNA ribose rings [22]. The positioning of the aromatic side chains is crucial for RNA recognition [24]. In our structure of the TRMT2A RRM, two phenylalanines are present on RNP1, but only one exposes its sidechain to the solvent (F113).

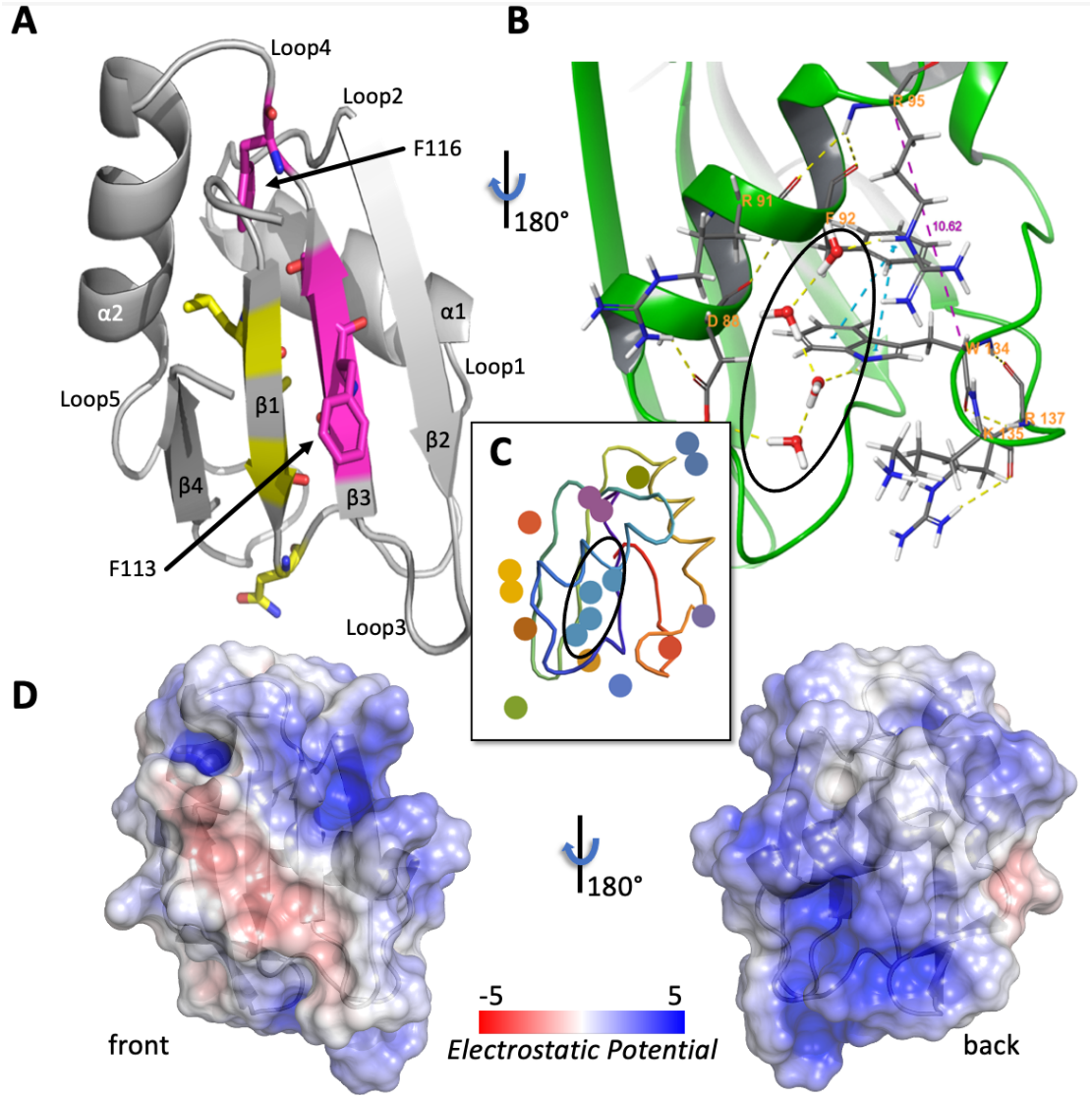


Figure 1: *Structural features of TRMT2A RRM (resolution: 2.0 Å).* ***A*** *X-ray structure of the RRM in cartoon representation. The two central β-strands feature the submotifs RNP1 (β3 strand) and RNP2 (β1 strand). Canonical RNP1/2 residues found in TRMT2A RRM are highlighted (RNP2 as yellow sticks on the β1-strand and RNP1 as magenta sticks on the β3-strand). The sidechain of F113, part of RNP1, is solvent-exposed. In contrast, the sidechain of F116 (part of RNP1 as well) is buried and part of loop 4.* ***B*** *View of the lower helical site of the RRM and the water network. The sidechain of W134 π-stacks with F92 (blue dashed lines).* C_α *distance between R95 and W134 is highlighted as a purple dashed line (see subsequent sections).* ***C*** *Agglomerative Euclidean-distance-based clustering. Discs with the same color belong to the same water cluster. The largest cluster is circled as in B.* ***D*** *Electrostatic surface obtained with the Adaptive Poisson-Boltzmann Solver (APBS)* [25] *Electrostatics PyMOL plugin and mapped onto the solvent accessible surface in units of kT/e.*

Structure-based virtual screening on RNA Recognition Motif of TRMT2A.

Neither of our determined crystal structures of TRMT2A RRM reveal distinct binding cavities (e.g., starting points for structure-based ligand discovery). Therefore, traditional druggability detection methods were augmented with alternative approaches, which incorporate protein conservation, solvation, and local flexibility considerations. With this approach a potential cryptic site was uncovered. Cryptic sites are not evident in the unbound protein surface, but

they can extend into underlying cavities upon ligand binding, providing tractable drug target sites, thus expanding the druggable proteome considerably [26].

Bioinformatics Analysis. Binding hotspots have been relatively well conserved during evolution, unlike the remainder of the protein surface [27]. Analysis of conservation patterns in the amino acid sequence, therefore, is a useful tool to detect binding sites.

To this aim, RRMdb, an evolutionary-oriented database for RNA recognition motif sequences [28] was queried, where the RRM of TRMT2A is classified as belonging to the family 262 with 25 known members from eukaryotes. The sequence conservation of the RRM residues was calculated over the multiple sequence alignment of these 25 sequences (see Methods and Figure S1 A-B for details). Highly conserved residues are not only found on the face of the β-sheet (the canonical RNA binding site), but also on the helical back (See Figure S1 C). All aligned sequences contain a lysine at position 135 (human TRMT2A numbering). A tryptophan at position 134 is present in all 25 but three sequences (88% of conservation), where an aromatic residue, tyrosine or phenylalanine takes its place. Interestingly, R95 is also highly conserved (77%). Such high sequence conservation, far from the putative tRNA binding site, suggests a functionally important role of these residues. Specifically tryptophan, given its bulky and rather hydrophobic sidechain, is indeed rarely found in loops, unless these loops are involved in protein-protein interactions [29].

Crystallographic Water Clustering and 3D-RISM Analysis. Since water clusters can mediate protein-ligand binding, these were searched at the surface of the protein [30]. An agglomeration-based clustering of water molecules resolved in the 2.0 Å crystal structure was performed and four non-singleton clusters were identified (see Figure 1B-C). The largest crystallographic cluster contains four water molecules, and is located between α1 and loop 5, in the so-called *minor/small groove*. In the 1.23 Å structure, a sulfate ion takes the place of this cluster. This motivated us to investigate solvation effects in more detail with the three-dimensional reference interaction site model (3D-RISM) theory [31]. The latter can estimate water distribution propensities around a solute (see Figure S2). This analysis confirmed the hydration site centers we found in the X-ray structure with 2.0 Å resolution. Notably, two unfavorable water propensities buried 2.5 Å beneath the sidechain of W134 were identified as well: the favorable release of corresponding water molecules to the bulk may drive local structural rearrangements.

Drug Binding Site Prediction on the Protein Crystal Structure. Next, a druggability assessment was performed, i.e. the size, shape, and physicochemical features of all RRM cavities were analyzed with DoGSiteScorer [32] (v2.0.0_15.01.2019). Five candidate binding sites were identified (see Figure S3) in the 2.0 Å structure, amongst which the largest one coincides with the above-mentioned *minor/small groove*. A corresponding, but equally inaccessible site was also identified in the 1.23 Å structure (see Figure S4A). All hereby detected sites were found to be too shallow to accommodate a small molecule ligand.

The possibility of cryptic site formation was therefore investigated. Cryptic or transient sites can be defined as only opening-up when a binder is present [26]. In a recent attempt to expand the druggable proteome, these sites were suggested as an interesting alternative for hit discovery, including proteins that lack conventional binding pockets [33].

Transient Site Discovery on TRMT2A. CryptoSite [33], which allows the rapid identification of residues contributing to cryptic site formation via a support vector machine (SVM), was used. We found that the highly conserved residues W134 and K135 are likely part of a cryptic site located in the *minor/small groove* (Figure 2A). Further analysis of this site with implicit solvent molecular dynamics simulations conducted with TRAPP (TRAnsient Pockets in Proteins, [34]) corroborated this finding. Specifically, below the *minor/small groove*, formed in part by

W134/R135, an accessible cavity is repeatedly detected upon different types of perturbation: either under rotamerically induced site chain perturbations with Langevin dynamics (L-RIP) [35] or via ligand-induced perturbations (RIPLig) with a probe mimicking an apolar/aromatic ligand approaching the *minor/small groove* region (Figure 2B-D and for the 1.23 Å structure Figure S4B) [35].

Our findings are in line with the fact that the *minor/small groove* of other RRMs has been previously reported to interact with different peptides [22]. However, our predicted transient site formation in the *minor/small groove* region is rather distant from the canonical RNA binding site and, it might not necessarily inhibit RNA binding. Therefore, the possibility of allosteric communication occurring between this identified site and the putative tRNA binding site (RNP1/2) was investigated.

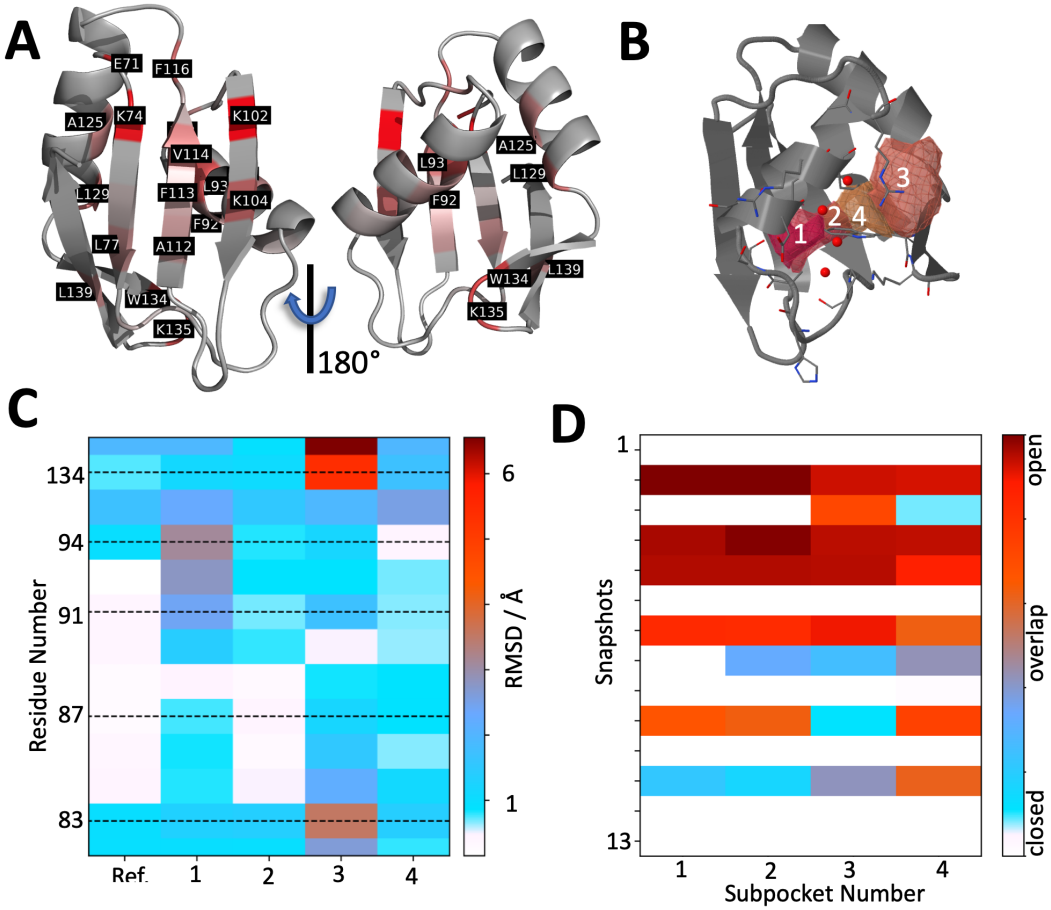


Figure 2: *Cryptic site assessment.* ***A*** *CryptoSite indices mapped onto the TRMT2A RRM (shown in cartoon representation). Values ranging from grey, unlikely part of a cryptic site to red, higher probability to contribute to a cryptic site. On the helical site of the RRM, W134 and L135 are detected as putative contributors to cryptic site formation.* ***B*** *Rear view of the RRM with the identified subpockets and consistent numbering in B, C, and D with TRAPP. All appearing transient regions are depicted as isosurfaces (at 25% occurrence over the simulation). Binding site residues are depicted as sticks.* ***C*** *Representative TRAPP binding site flexibility estimation (RMSD values in Å are color-coded from white to red) using the L-RIP/RIPlig approaches. W134, H83, and K135 show high structural variance.* ***D*** *Subpocket occurrence during the simulations shows that all subpockets are formed repeatedly.*

RNP1/RNP2 and small groove allosteric regulation. Perturbations mimicking ligand binding on the suggested tRNA-anchor-phenylalanine (F113 of RNP1) were simulated to see if it can cause a functional change in the *minor/small groove* region and thus predict the presence of

allosteric communication. To this purpose, both structures were analyzed with AlloPred [36], which uses perturbation of normal modes alongside other pocket descriptors in a machine learning approach. These results indicate that the *minor/small groove* region can allosterically modulate the β3-strand where F113 is present (see Figure S5 for the 2.0 Å structure and Figure S4C for the 1.23 Å structure, respectively). Specifically, in our analysis, W134 seems to be a major player in this allosteric communication since changes in its immediate surroundings are allosterically connected to the suggested tRNA binding site. This is in line with our water propensity distribution and conservation analysis pointing towards a key role of this residue.

Molecular dynamics and snapshot selection. To corroborate this finding, the flexibility of this putative allosteric pocket was investigated through molecular dynamics (MD) simulations. A 10 µs-long MD trajectory in explicit solvent was performed (Figure S6). While the overall structure remained stable, loop 5 appears to contribute most to the flexibility of the RRM. This is in line with the water distribution analyses in this area discussed above. We found that the C_α distance of W134 and R95 can increase from 10.62 Å in the crystal structure to more than 20 Å (Figure 3A) during the simulation, as can be seen in a representative MD snapshot (Figure S7).

During our simulation, W134 transiently passes from a buried conformations, where π-stacking with F92 is possible (Figure 1A and 3B), to solvent-exposed ones. The conformations where W134 is solvent-exposed are also characterized by loop 5 structural rearrangements (further details are provided in Figure S8 A-B of the SI). When the C_α atoms of conserved amino acids R95 and W134 are in close proximity, the consistent formation of an enclosed cavity during the simulation of the 2.0 Å structure of the RRM was observed. Moreover, an opening in the 1.23 Å structure (Figure S4 D) was also detected. Specifically, the indole ring of W134 is gradually pushed outwards and becomes solvent-exposed. This is of particular interest since W134 was indicated in our initial CryptoSite model, as well as in the follow-up TRAPP analysis, as key residue involved in putative cryptic site formation, together with R95.

Thus, frames with a C_α distance (R95-W134) cut-off of 8.5 Å were selected for further analysis (104 frames out of 100,000 total frames for the whole trajectory). Among those, six conformations had a DrugScore >0.60. The two highest-ranking receptor conformations displayed a W134 indole ring flip, which infrequently occurs during the simulation (See Figure S8 C) and were therefore not considered. The third highest-ranking conformation, however, was frame 60,942 (after 6,094.2 ns, DrugScore: 0.62), where the loop is in the process of refolding to its crystallographic conformation after a roughly 1 µs long reorganization. In this state, the volume of the *minor/small groove* is significantly increased reaching up to 589 $Å^3$ (See Figure S8 D).

By comparing this selected conformation with the original X-ray structure, we can see that the loop α2-β4 moves outward (Figure 3B). During this widening of the pocket, the side chain of tryptophan (W134) opens a small and transient cavity below the small groove. The indole ring of W134 then forms a cation-π interaction with R135. This interaction can energetically compensate the unfavorable solvent exposure of the indole ring and breaking the π-stacking with F92. The RMSD between the crystal structure and the selected structure is 1.737 Å (Figure 3B).

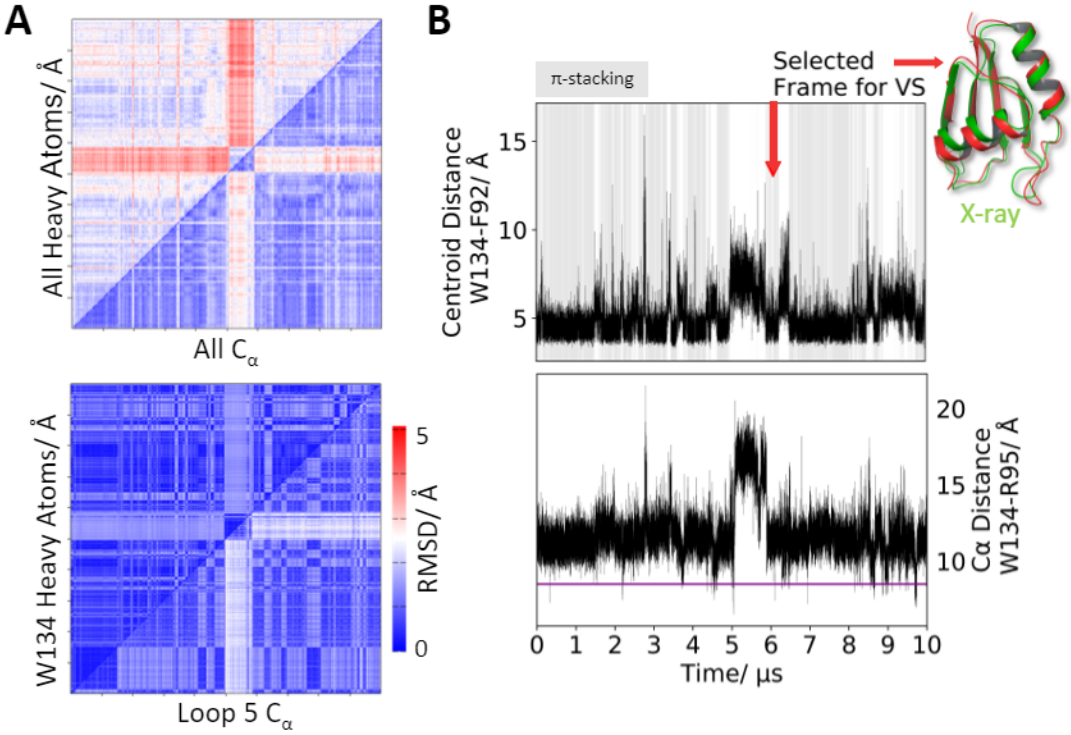


***Figure 3**: Explicit-solvent molecular dynamics results. **A** 2D-RMSD plots of C_α atoms (horizontal axes) and heavy atoms (vertical axes) over the 10 µs-long simulation. (top) Heavy atom/C_α RMSD indicate an overall stable structure, with considerable rearrangements occurring after about 5 µs. (bottom) Flexibility in the loop 5 region is mediated by the transient π-stacking of W134 and F92 (blue dashed lines in Fig. 1B) **B** (top) These structural changes are also reflected in the centroid distances of the aromatic rings involved. Frames with detected π-stacking, according to reference 37, are underlined in grey. (bottom) C_α distance between R95 (α1 helix) and W134. We consistently observed the formation of a hydrophobic pocket when these residues approach. Thus we extracted snapshots with an R95-W134 C_α distance below 8.5 Å (indicated by the magenta line) for a detailed druggability analysis. The red arrow indicates the snapshot (red) used for virtual screening (VS) and is superimposed with the low-resolution crystal structure (green).*

This partially hydrophobic pocket was next used for structure-based small inhibitor discovery. A more detailed pocket description was elucidated with Schrödinger SiteMap and is shown in Figure S8 E [38]. In addition, Replica Exchange simulations (REST2) [39] with solute tempering were also performed (see Supplement S4, Figure S9 and Figure S10). The latter technique is a powerful computational tool to enhance the exploration of proteins' conformational space [40]. Our findings suggest that the opening of the cryptic pocket is a rare event that might be facilitated by ligand binding.

Virtual Screening on TRMT2A RRM and Post Processing. To exploit this putative cryptic pocket, a structure-based virtual screening workflow was performed to prioritize small molecules that might bind into this site for *in vitro* testing. Virtual screening can be defined as a set of computational methods that analyzes large databases or collections of compounds to identify potential hit candidates [41]. As a first step, therefore, *in silico* collections of commercially available small molecules were retrieved and prepared as outlined in the methods section. Next, the prepared compound library was docked into the identified binding site. A possible binding pose for each ligand conformer was generated and then prioritized with a fast-scoring function (see Methods). Ligands with unreasonable binding mode hypotheses were discarded, while the best ones were retained and subsequently rescored using a more detailed scoring function (see Methods). Among these, 13 compounds were selected by visual inspection to undergo experimental testing (compounds 13-25, Table S2).

Further selection criteria were based on diversity of interactions formed with the TRMT2A RRM, availability, and costs. Notably, several compounds predicted to bind into the *minor/small groove* revealed a common substructure (Figure 4 A-B), despite displaying diverse binding modes. Therefore, we decided to enrich our original pool with the above-mentioned substructure. To this purpose, a second screening was performed based on a pharmacophore (i.e., an abstract description of the molecular features that are believed to be necessary for ligands to be biologically active against the selected target protein) obtained from compound 13 and 15. Three compounds were added as results of such a procedure (see Methods for details).

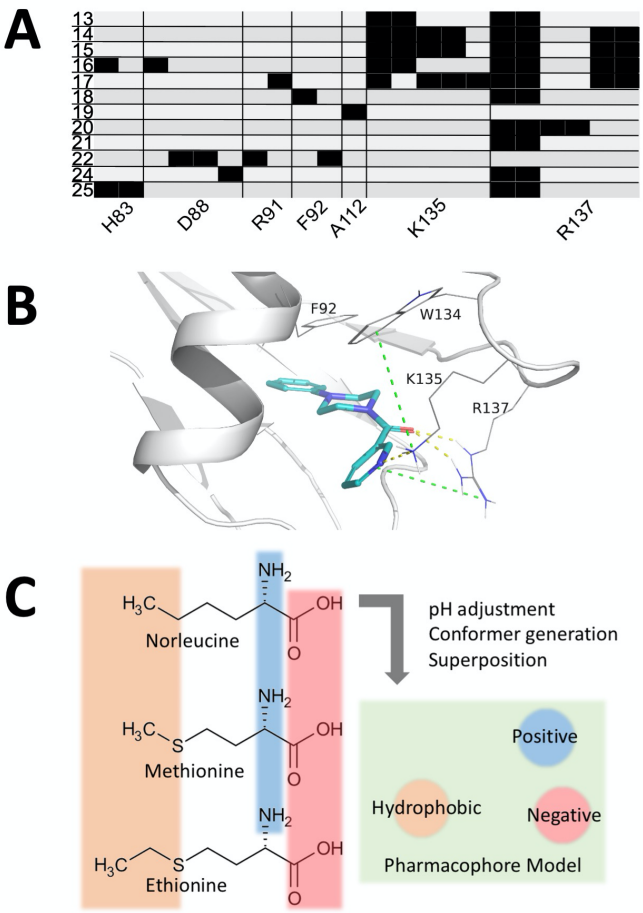


Figure 4: *Predicted protein-ligand interaction patterns.* ***A*** *Protein-Ligand Fingerprint (PLIF) of docking poses into the cryptic site depicted as barcodes. Black bars indicate one or more interactions between TRMT2A RRM residues (columns) and individual ligands (rows). First column indicates the running number of the docked compounds.* ***B*** *Interactions of docked pose of compound 13. K135 and R137 form hydrogen bonds with the ligand. Moreover, the phenyl ring of the ligand can potentially π-stack with F92. Cation-π interactions between K135 and W134, but also between R137 and the pyridinyl ring are apparent (green dotes lines).* ***C*** *Pharmacophore for three structurally similar compounds, that inhibit a E. coli homolog of TRMT2A. The angle between positive-negative-hydrophobic features is nearly a right angle (86.9°). The distance between the hydrophobic and the positive/negative (ionic) feature are 7.3 Å and 6.8 Å, respectively.*

Ligand-based virtual screening on the methyltransferase catalytic domain of TRMT2A.
To also exploit the possibility of TRMT2A inhibition with specific binders for the CD, a ligand-based virtual screening was performed. The latter uses the information present in known active ligands rather than the structure of the target protein. This is a suitable approach here

since we lack structural information on the TRMT2A CD. Specifically, a pharmacophore hypothesis was generated and exploited to screen libraries of chemical compounds to identify the ones with the required molecular and spatial features.
To perform our ligand-based virtual screening, we searched for all available inhibitors of TRMT2A homologs. Unfortunately, only data for TrmA in *E. coli* were found [42, 43, 44, 45]. Therefore, the known TrmA inhibitors L-norleucine, L-methionine, and L-ethionine, were superimposed and a pharmacophore model was built. The selection criteria of these three are explained in detail in the methods. The resulting pharmacophore model contained only three features: a hydrogen bond donor and acceptor (amino acid abstraction) and an approximately equidistant hydrophobic site feature, representing the sidechains (Figure 4C). Next, the pharmacophore was used to screen libraries of known chemical compounds to identify the ones with the same selected molecular features. Compounds were sorted according to predicted absorption, distribution, metabolism, and excretion (ADMET) properties, human oral availability, and blood-brain barrier permeability. After visual inspection, several known inhibitors from the literature and pharmacophore derived compounds were selected for experimental testing (compounds 1-12, Table S2)

***In cell* assays**

Human embryonic kidney cells (HEK293T) were used to determine the effect of *TRMT2A*-silencing on polyQ-induced toxicity and aggregation. For analysis, we first generated cells with stable expression of a short hairpin RNA (shRNA) to induce *TRMT2A* specific RNAi (sh1574). A cell line with stable expression of a scrambled shRNA served as control (shK). Analyzing abundance of TRMT2A levels in both cell lines revealed a strong reduction of TRMT2A abundance/protein levels compared to shK cells in Western blot analysis (Figure 5 A). In a next step, we asked whether reduced TRMT2A abundance protected from polyQ-induced toxicity and reduced polyQ aggregation as observed in flies [11]. We transiently transfected sh1574 and shK cells with plasmids coding *huntingtin* exon1 harboring a polyQ stretch of 103 glutamines fused to GFP (*HttQ103:GFP*). PolyQ aggregates were analyzed by filter retardation assay (FRA) two days post transfection [46]. The FRA analysis revealed a strong aggregate load in shK cells. However, in sh1574 cells, abundance of polyQ aggregates was low (Figure 5 B). Presence of PolyQ aggregates coincides with cell dysfunction/loss in several model systems [6, 47, 48]. When performing cell viability analysis, no differences were observed in control and TRMT2A-silenced cells upon expression of non-aggregating, non-toxic HttQ25 (data not shown). Expression of HttQ103:GFP strongly reduced survival in shK cells while TRMT2A-silenced cells were protected from polyQ-induced toxicity (Figure 5 C). Assuming a specific inhibition of TRMT2A with selected compounds, we would expect a strong decrease in polyQ aggregate abundance and a rescue effect on cell viability only in shK. As sh1574 almost entirely lacks TRMT2A, treatment with a TRMT2A specific inhibitor should not have any effect on polyQ aggregation and polyQ-induced cell death. Thus, we used both cell lines, shK and sh1574 to address efficacy of predicted TRMT2A inhibitors in cellular system.
Putative inhibitors were analyzed following the same experimental setup. ShK and sh1574 cells were transiently transfected with plasmids mediating HttQ103:GFP expression. Four hours post transfection, cells were either treated with inhibitors or with solvent only as control. Microscopic inspection to determine the abundance of GFP positive cells was done to ensure similar/identical transfection rates. In addition, Western blot was performed to evaluate the abundance of soluble HttQ103:GFP peptide in transfected HEK293T in each experiment. These parameters were not expected to differ dramatically between sh1574 and shK in treated and untreated condition. In FRA analysis, untreated shK cells should present high polyQ aggregate load. In case of an efficient TRMT2A inhibitor, inhibitor treated shK cells should display reduced polyQ aggregation in FRA analysis. In contrast, a specific TRMT2A inhibitor should have almost no effect on sh1574 cells.

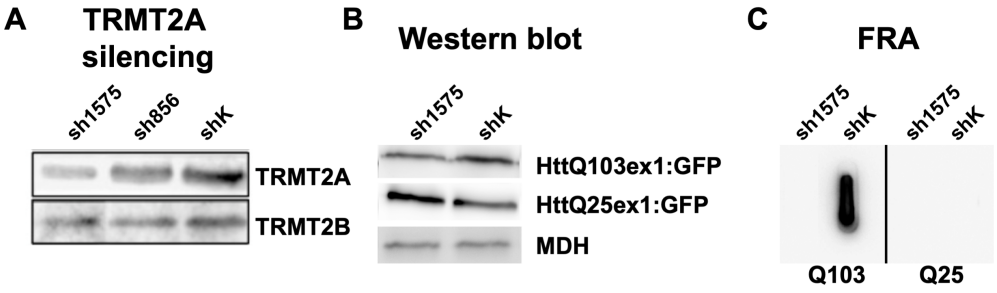


Figure 5: *PolyQ aggregation is reduced in TRMT2A-deficient cells.* ***A*** *Exemplary Western blot analysis to detect TRMT2A abundance in cell lysates derived from control (shK) and two cell lines with stable expression of short hairpin (sh) RNA to facilitate TRMT2A-silencing via RNAi. Note the strong reduction of TRMT2A (upper band) in sh1574 cell lysates. Detection of the parlog TRMT2B was used for normalization.* ***B*** *Detection of exogenous Httex1Q103:GFP and Httex1Q25:GFP peptides by Western blot analysis after transient transfection of control (shK) and TRMT2A-silenced (sh1574) cells. No difference in the abundance of GFP-tagged peptides was detected in RIPA soluble fraction. Detection of Malat Dehydrogenase (MDH) was used as loading control.* ***C*** *Detection of RIPA insoluble Httex1Q103:GFP peptides by FRA analysis. Compared to shK cells, a strong reduction of insoluble/aggregated HttQ103:GFP was observed in sh1574 cells. Non-aggregating Httex1Q25:GFP served as negative control for FRA analysis.*

In this paradigm we tested in total 25 compounds, of which four were cytotoxic and therefore excluded from further analysis (compounds 16, 19, 22 and 23). The remaining 21 compounds were used for FRA analysis. A summary of FRA analysis for the tested compounds is provided in Figure 6 A. For some compounds (grey), FRA analysis did not provide consistent data as reflected by high standard deviation (SD) (SD > 40% of control value). These compounds were 1, 2, 3, 6, 7, 8 and 9. Accordingly, these seven compounds were excluded from further analysis. Among the remaining compounds, three (5, 12 and 25) caused a strong reduction of polyQ aggregate load, comparable to the reduction observed in sh1574 cells (black bar). Seven compounds (4, 10, 13, 14, 17, 18 and 20) caused a moderate to mild reduction of polyQ aggregate load. The remaining four tested putative inhibitors of TRMT2A (11, 15, 21 and 24) did not reduce polyQ aggregation in shK cells (Figure 6 A). Accordingly, 13 compounds (4, 5, 10, 11, 12, 13, 14, 15, 17, 18, 20, 24, and 25) were further analyzed for their effect on polyQ-induced cell death in HEK293T cells (Figure 6 B). Untreated shK cells displayed roughly 60% cell death in our automated cell viability measure based on Alamar Blue staining. In untreated, but TRMT2A deficient sh1574 cells, the cell death was decreased about 20%. Thus, we would expect that shK cells treated with functional TRMT2A inhibitors should display cell death rates between 40% and 60%. Among the tested compounds, just three (11, 12 and 18) decreased cell death in a similar way as observed for sh1574 cells. Other compounds even slightly enhanced cell death (10, 24 and 25). The remaining tested compounds caused only mild or no protection towards polyQ-induced cell death.

Overexpression of polyQ peptides fused to GFP have been successfully used to model polyQ aggregation and to induce polyQ toxicity in many model organisms [48, 49, 50]. However, in human disease situation, endogenous polyQ containing proteins are expressed by their native promoter and undergo multiple post-transcriptional modifications like proteolytic cleavage and biotinylation [51]. To further evaluate our compounds as potential therapeutic agents to treat polyQ-diseased patients, we used patient-derived fibroblast (Figure 6 C). We obtained fibroblasts from patients suffering from SCA3 by the Coriell repository (https://www.coriell.org). FRA analysis revealed formation of RIPA insoluble polyQ aggregates in these cells after two days of culturing. Thus, we applied the same treatment paradigm as

for HEK293T cells to test a selection of potential TRMT2A inhibitors in fibroblasts. Compounds were chosen according to effects in previous analysis and availability. We chose inhibitors 12 and 25. Compound 12 reduced polyQ aggregation in FRA analysis and protected HEK293T cells from polyQ-induced cell death. Inhibitor 25 also strongly reduced polyQ aggregation in FRA analysis, but had no effect on polyQ-induced cell death. Like in previous analysis using HEK293T cells, treatment with these two inhibitors decreased the abundance of polyQ aggregates in SCA3-patient derived fibroblasts (Figure 6 C).

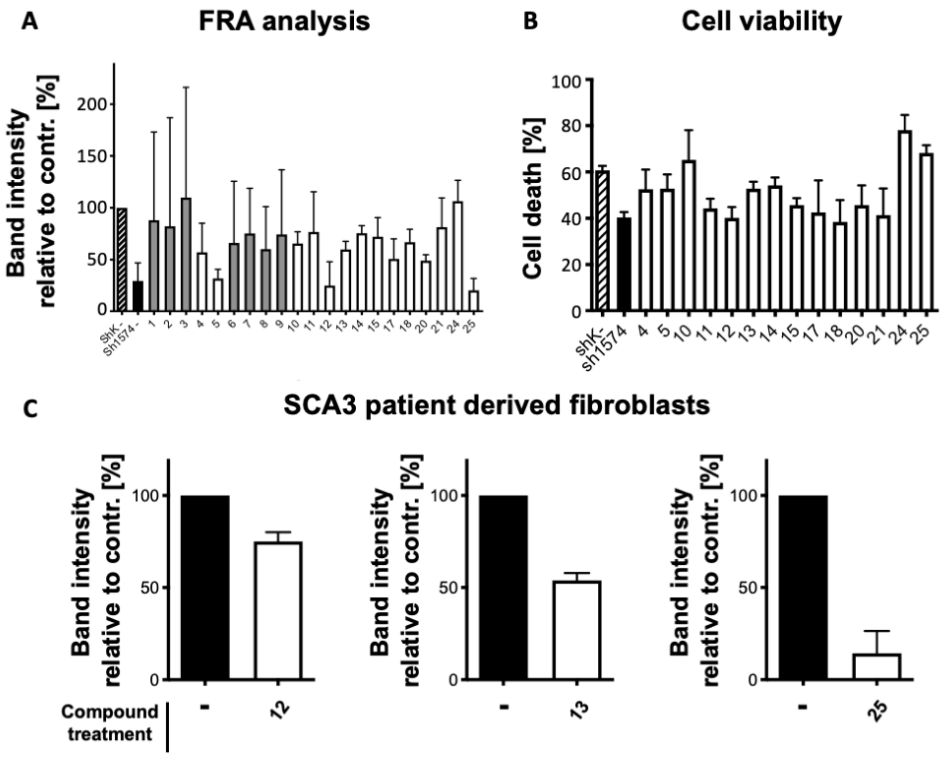


Figure 6: *Functional test of potential TRMT2A inhibitors in cells.* ***A*** *FRA analysis was performed to detect abundance of polyQ aggregates HEK293T cells after inhibitor treatment. Graph depicts summary of at least 3 replicated measures. Inconsistent results as presented by high SD are marked (grey).* ***B*** *PolyQ-induced cell death was determined in inhibitor treated HEK293T cells.* ***C*** *Detection of polyQ aggregates in fibroblasts derived from a HD and a SCA3 patient. Samples were either DMSO treated (contr., black) or treated with indicated inhibitors (white).*

From all compounds for which we performed cell death assays, compounds (4, 9, 10, 11, and 12) were previously reported in the literature as inhibitors of *E. coli* or *Saccharomyces cerevisiae* homologs of TRMT2A, respectively. Neither L-norleucine (compound 4), L-methionine (compound 9), nor L-ethionine (compound 10) was able to reduce cell death, suggesting a considerably different catalytic binding site. The two polyamines tested, putrescine (compound 11) and spermidine (compound 12), consistently decreased cell death. The organomercury agent p-chloromercuribenzoic acid (compound 5, PCMB) reacts with free thiol groups in proteins containing cysteine. In fact, the RRM of TRMT2A features a solvent-exposed cysteine (C111).

On the other hand, compounds 13-25 resulted from structure-based drug discovery and the putative cryptic allosteric pocket on the helical site of the RRM. The majority of compounds were predicted to bury at least one aromatic moiety into the cryptic pocket. Compounds 19, 22, and 23 are cytotoxic. Compound 19 was used as an internal negative control, as it contains a bulky and hydrophobic naphthalenyl group, which does not fit into the cryptic site

based on our docking study. Not only does compound 19 fail to rescue the cell lines affected by polyglutamine toxicity but also reveals additional cytotoxicity.

Interestingly, ligands featuring a phenyl group (e.g., compounds 13-15, 18, 21, and 23) predominantly decrease cell death, with the exception of the toxic compound 23 for which we detected only a subtle drop. We speculated, that if these compounds bind indeed into the cryptic site, their phenyl group would take the place of the indole ring of W134 (as seen in Fig. 1B), and then could form π-stacking interactions not only with F92 but also with the indole ring of W134, now slightly displaced with respect to the X-ray structure (see Fig. 4 B).

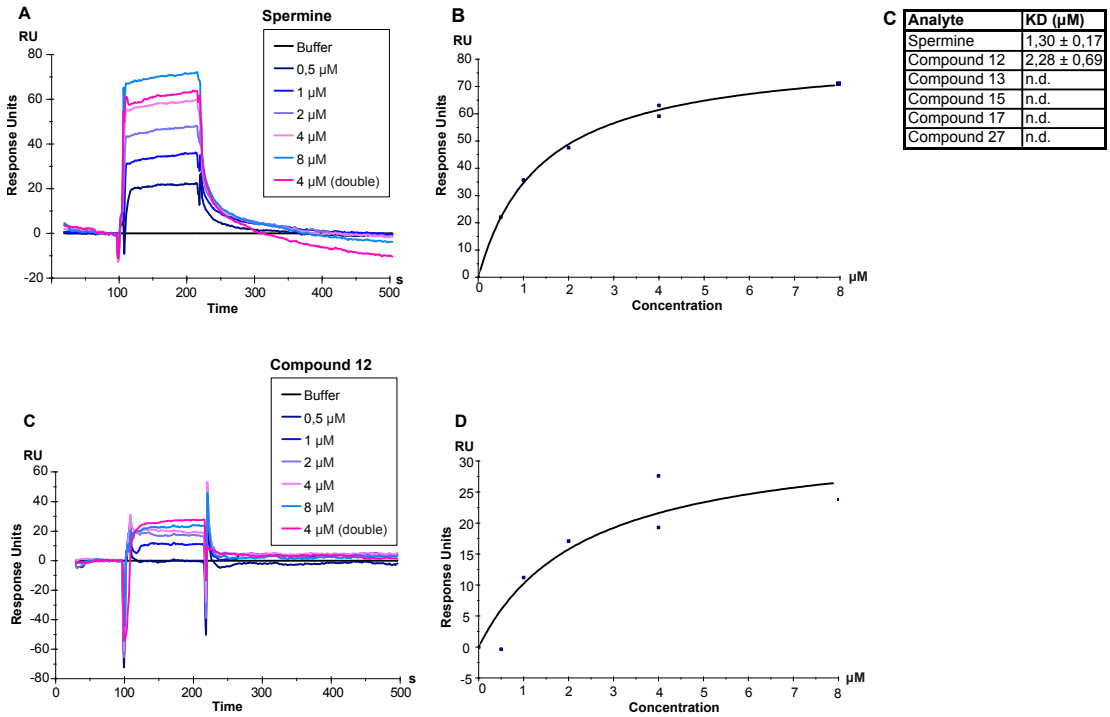
Furthermore, since we did not measure TRMT2A inhibition directly, but by means of an assay, we wanted to investigate the possibility of problematic chemotypes which frequently lead to false positives with the Pan-assay Interference (PAINS) filter [52]. The four tested compounds 18, 19, 20, and 21 did not pass the PAINS filter. As aforementioned mentioned compound 19 increased cytotoxicity. Nevertheless, for the remaining three compounds (18, 20, and 21) a false positive hit cannot be excluded.

We also evaluated the possibility of off-targets with a reversed *in silico* screening approach as implemented in SwissTargetPrediction [53] to elucidate putative human protein targets for non-polyamine bioactives. For these compounds, the most frequently predicted target class was kinases. However, this result may be biased due to the great number of known kinase inhibitors.

Notably, three novel compounds discovered by means of structure-based discovery (15, 17, and 21) along with the polyamines (spermidine and putrescine) reported in the literature for TRMT2A homologs consistently and effectively lowered cell death and aggregation in polyQ affected cells and do not contain critical substructures that often confound cell assays. Finally, compounds 15 and 21 show structural similarity in the form of a phenyl group.

Biophysical measurements of TRMT2A RRM and compounds

The compounds tested on SCA3 patient fibroblasts (12, 13 and 25) and two additional compounds that showed promising results in cell death and FRA analysis (15 and 17) were selected for further *in vitro* interaction experiments. To analyze RRM-compound interaction, surface plasmon resonance (SPR) experiments were performed. As a positive control the polyamine spermine was used. Polyamines are known to bind tRNA methyltransferases [54]. Indeed, we observed that spermine binds in the micromolar range to the TRMT2A RRM (K_D = 1.30 +/- 0.27 µM; Figure 7). In contrast, none of the selected compounds showed any binding to the protein (Figure S11) except for compound 12, the polyamine spermidine (K_D = 2.28 +/- 0.69 µM; Figure 7). Noteworthy, SPR experiments were performed with low salt concentrations in the running buffer to also allow for the occurrence and detection of lower affinity binding events. The size difference between compounds (> 0.2 kDa) and the TRMT2A RRM (9.2 kDa) was taken into consideration by coupling high amounts of protein (3000 RU) to the CM5 chip. Taken together, with the exception of polyamines spermine and compound 12/spermidine no direct binding to the RRM could be detected.



C

Analyte	KD (µM)
Spermine	1,30 ± 0,17
Compound 12	2,28 ± 0,69
Compound 13	n.d.
Compound 15	n.d.
Compound 17	n.d.
Compound 27	n.d.

Figure 7: *Results of Surface Plasmon Resonance experiments.* ***A+C*** *The TRMT2A RRM was coupled to a CM5 chip reaching 3000 RU. A concentration series from 0.5 to 8 µM spermine/compound 12 was injected, resulting in the illustrated, double-referenced binding curves. Representative curve from triplicate measurement is shown.* ***B+D*** *The response at equilibrium of the binding curve is plotted against spermine/compound 12 concentration. Steady state affinity model was fitted and allowed for the calculation of KD values* ***C****.*

Conclusion

In this work, several *in silico* strategies were performed to identify a small library of ligands able to modulate the activity of tRNA methyltransferase 2 homolog A (TRMT2A) in cells.

TRMT2A is an enzyme that converts uridine to m^5U_{54} on tRNA across species, from yeast to man [14]. Unfortunately, very little is known concerning the role of this modification in mammals [12]. It was shown that absence of TRMT2A reduces polyQ aggregation and polyQ-induced toxicity in HEK293T cells [11]. This suggests a critical role of the tRNA modification m^5U_{54} in the context of polyQ-induced toxicity. Thus, it was reasonable to test known and new (putative) inhibitors of TRMT2A with regard to their effect on polyQ aggregation and polyQ-induced toxicity. In total, we assayed 25 potential inhibitors. While four compounds turned out to be cytotoxic, the ten other compounds reduced polyQ-aggregation in FRA. Amongst the latter, three compounds (5, 12, and 25) turned out to reduce polyQ-aggregation very efficiently. The observed effect on polyQ aggregation was as strong as after *TRMT2A* silencing.

According to its similarity to yeast Trm2p, mammalian TRMT2A harbors a putative RRM in the N-terminus and the CD at the C-terminus [12]. Inhibiting its function directly interfering with the CD activity or with TRMT2A binding to the RNA might ameliorate some pathological effects linked to polyQ toxicity.

No full-length structure is available for this protein. Here we managed to crystallize and solve the structure of a sub fragment of TRMT2A corresponding to the RRM. We conducted two parallel strategies for finding inhibitors; one exploiting the structural information of the crystal structure of the TRMT2A RRM, the other exploiting the available knowledge of general methyltransferase inhibitors.

The RRM is characterized by the absence of protein pockets. However, we identified a putative cryptic site, potentially able to allosterically affect RNA binding. As first strategy, we conducted a structure-based virtual screening and the selected candidates were tested both *in vitro* and *in cell* experiments. Several compounds with similar substructures induced a reduction of polyQ aggregation *in cell*. Also, two of them, 13 and 25, were able to decrease polyQ aggregates in SCA3-patient derived fibroblasts.

For the CD, a pharmacophore model was identified and used for a ligand-based virtual screening. Also in this case, several selected candidates were able to reduce PolyQ aggregation in cell culture. Moreover, the polyamine spermidine (compound 12) caused a decrease in the number of polyQ aggregates in SCA3-patient derived fibroblasts and directly interacted with the RRM in SPR experiments.

Except for spermidine, we were not able to prove any direct binding between ligands screened against the TRMT2A RRM and this domain. There are several possible explanations for this. Ligands targeting the RRM were designed for the cryptic site in the *minor/small groove*. The lack of observed interaction of the compounds with the RRM could be due to the possibility that the cryptic site might require the presence of tRNAs for its formation. Furthermore, the rest of the protein might also play an important role in the suggested allosteric mechanism, connecting the tRNA binding site with the identified cryptic site. In addition to that, some studies show that not all of the cryptic pockets that are seen to open during MD simulations are capable of binding a ligand with substantial affinity [55]. Our model of allosteric communication between the tRNA binding site and the cryptic pocket relies on a normal mode analysis. This approach involves drastic approximations and provides reasonable prediction only if correlated thermal motions of these two areas exist [56]. Also, ligands could target the CD or other protein regions that were excluded in our assays. However, we consider this option the least likely as the *in-silico* selection of ligands was based on features derived from the experimental crystal structures of the TRMT2A RRM.

Another possible explanation is that the observed effects of inhibitors in cells might be the result of an interaction with other protein/s and/or activation of cellular pathways compensating aggregation-induced detrimental effects. In an attempt to exclude this possibility, we tested our ligands also on TRMT2A-silenced cells (sh1574) and no effect on aggregation or cell death was observed, pointing toward a TRMT2A-dependent effect of our ligands.

As mentioned above, spermidine was shown to bind the TRMT2A RRM. Like other polyamines, this compound has a well-proven affinity for nucleic acids as well as proteins. Since polyamines have been implicated in translational control and RNA binding [57], it is not entirely surprising to see also binding to the RRM. It remains to be shown how specific the interaction between the TRMT2A RRM and spermidine really is.

Several clinical trials are underway against polyQ diseases, including oligo-antisense nucleotide tominersen (formerly known as IONIS-HTT_{Rx} and RG6042) to uncover bona fide treatments [10]. Encouragingly, for tominersen the European Medicines Agency has already granted orphan drug designation for the treatment of patients with HD.
Nevertheless, no therapy has been approved to interfere with the clinical progression of the diseases in humans [9]. Moreover, oligonucleotides (ASOs) approaches still present several inherent drawbacks, like invasive administration, risky side effects and long optimization strategies [58]. Therefore, targeting a novel CAG-disease modifier such as TRMT2A is a viable therapeutic approach with several advantages over ASO-based therapies.

Methods

1. Crystal Structure

Cloning, expression and purification. Plasmid containing human TRMT2A sequence was provided by Aaron Voigt. Human TRMT2A RRM (aa 69 – 147) was cloned with a cleavable SUMO-tag at the N-terminus and expressed in Rosetta *E. coli* cells in LB medium. After isopropyl-thiogalactoside (IPTG, 0.5mM) induction over night at 18 °C, cells were harvested (4500 x g, 15 min, 4 °C) and flash frozen with liquid nitrogen. All subsequent steps were carried out at 4 °C. Typically, 6 liters of culture were resuspended in lysis buffer (50 mM HEPES/NaOH pH 8.5, 500 mM NaCl, 20 mM imidazole, 0.5% (v/v) Tween, 2% (v/v) glycerol) supplemented with one tablet of EDTA-free protease inhibitor cocktail to a total volume of 50 ml. After sonication (4 x 6 min, amplitude 40%, output: 6), the lysate was clarified by centrifugation (40000 x g, 30 min), loaded onto 5 ml His-FFTrap column (GE Healthcare), equilibrated in His-A buffer (50 mM HEPES/NaOH pH 8.5, 500 mM NaCl, 20 mM imidazole), and washed with 10 column volumes (CV) of His-A buffer, followed by a 10 CV wash with His-B buffer (50 mM HEPES/NaOH pH 8.5, 2000 mM NaCl, 20 mM imidazole). SUMO tagged protein was eluted with a 10 CV gradient of His-A buffer and His-C elution buffer (50 mM HEPES/NaOH pH 8.5, 500 mM NaCl, 500 mM imidazole). For tag cleavage, the eluate was supplemented with 100 µg of HRV 3C protease (PreScission) and dialyzed against dialysis buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 1 mM DTT) O/N. Cleaved protein was applied onto a His-Trap FF column and the flow through that contains cleaved protein was collected. This flow through was concentrated and loaded onto a size exclusion chromatography column (Superdex 75 10/300 gl) equilibrated in SEC buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl). Purified protein was concentrated to 8 mg/ml by ultrafiltration. Protein concentrations were determined by measurement of the A_{280}. Aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

Crystallization, diffraction data collection, and processing. The crystallization experiments for the RRM domain were performed at the X-ray Crystallography Platform at Helmholtz Zentrum München. Crystallization screening was done at 292 K using 8.0 mg/ml of protein with a nanodrop dispenser in sitting-drop 96-well plates and commercial screens. Crystals appeared after 1 week and were big enough for X-ray diffraction experiments. The initial data set for solving the structure was collected for a crystal grown in 0.22 M lithium sulfate, 0.1 M sodium acetate (pH 4.5) and 26% (w/v) PEG 6000 (Hampton Research PEG screen). However, the best data set used for refinement was collected for a crystal grown in 0.2 M cesium sulfate and 2.2 M ammonium sulfate (Hampton Research $AmSO_4$ screen). For the X-ray diffraction experiments, the crystals were mounted in a nylon fiber loop and flash-cooled to 100 K in liquid nitrogen. Prior freezing the crystals were protected with 25% (v/v) ethylene glycol. Diffraction data were collected at 100 K on the P11 beamline (DESY, Hamburg) and PXIII beamline (SLS, Villigen). The diffraction data were indexed and integrated using *XDS* [59] and scaled using *SCALA* [60, 61]. Intensities were converted to structure-factor amplitudes using the program *TRUNCATE* [62]. **Table S1** summarizes data collection and processing statistics.

Structure determination and refinement: The structure of the RRM with a resolution of 2.02 Å was solved by the Auto-Rickshaw pipeline [63, 64] using the sequence as input. For the molecular replacement step followed by several cycles of automated model building and refinement, the Auto-Rickshaw pipeline invoked the following X-ray crystallography software: MORDA (Vagin and Lebedev; http://www.biomexsolutions.co.uk/morda/), CCP4 [65], SHELXE [66], BUCCANEER [67, 68], RESOLVE [69], REFMAC5 [70] and PHENIX [71]. Model rebuilding was performed in COOT [72]. Initial refinement was done with the 2.02 Å dataset in REFMAC5 [70] using the maximum-likelihood target function. Further refinement was done with another dataset having the higher resolution of 1.23 Å. The stereochemical analysis of the final model was done in PROCHECK [73] and MolProbity[73]. The final model is characterized by R/R_{free} factors of

12.60 / 18.90%. Atomic coordinates and structure factors for the 2.0 Å and 1.23 Å structures have been deposited to the Protein Data Bank under accession codes 7NTN and 7NTO, respectively.

2. *In-silico* Analyses of the Crystal Structure

Agglomerative Euclidean distance clustering on the crystallographic water (oxygen atom) clusters was done with Wolfram Mathematica 11.3. The X-ray structure was prepared with Maestro Schrödinger (Schrödinger Release 2019-1: Maestro, Schrödinger, LLC, New York, NY, 2019.) using default settings unless otherwise noted. Alignments were performed with ClustalW [74], while conservation analyses were performed with AL2CO [75]. For binding site detection we used DoGSiteScorer [32], the transient site discovery the machine learning-based CryptoSite [33] approach and TRAPP [34]. TRAPP is a framework to generate and analyze large-scale secondary structure rearrangements on short molecular dynamics timescales, e.g., by disrupting sidechain atoms with short pulses (L-RIP). The system temperature is controlled by a Langevin thermostat to ensure a constant average temperature. In parallel, we investigated ligand-mediated conformational changes. We used an apolar/aromatic amino acid, as a model for hydrophobic/aromatic ligand features and brought them into contact with the protein surface (RIPlig)[35]. Finally, we investigated putative allosteric effects in TRMT2A RRM with AlloPred [36].

MD simulations. The protein was prepared using Maestro (Schrödinger Release 2016-2) and manually checked for correct protonation states. The partially occupied R137 with the side chain closer to the protein center was retained and crystallographic waters as well. To ready the protein structure for the explicit solvent molecular dynamics simulations, we used tleap (AmberTools 17) and the AMBER force field ff14SB [76]. First, the domain was embedded in a truncated octahedral TIP3P [77] water box with a buffer of 15 Å, resulting in about 6690 water molecules. Neutralizing the system was achieved by introducing nine randomly placed Cl^- counterions, placed with the addion2 application, to minimize potential perturbations of the system by adding charges too close to the protein.
After minimization with harmonic restraints on protein heavy atoms, the system was gradually heated from 100 to 300 K over 200 ps in NVT ensemble (see Wallnoefer et al. [78] for more details). A density equilibration over 1 ns was performed, followed by free simulations of the systems in NpT ensemble over 50 ns to ensure reasonable equilibration of the simulated systems. Production runs were carried out at 300 K using the Langevin thermostat [79] at 1.0 bar with 8.0 Å nonbonded cutoff. A 2.0 fs timestep was facilitated due to the usage of the SHAKE algorithm [80]. Snapshots were saved to trajectory every 500 steps or equivalent 1 ps for analysis.

Trajectories were analyzed using cpptraj from AmberTools16 [76] and PyMDLog (Version 1.0.1). Distances and interactions of protein and ligand were likewise calculated via cpptraj. The resulting trajectories were analyzed for convergence of the usual MD simulation parameters (energies, density, temperature, etc.). Depictions were created with PyMOL (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).

3. Virtual Screening

The protein conformations with the two highest DrugScores were not further investigated, as we found W134 dihedrals in these frames with low overall prevalence (See Figure SI8 C). Therefore we proceed with the third highest-ranking TRMT2A RRM conformation (according to DrugScore) and performed a virtual screening using MolPort stock compounds (7.409.777 compounds, www.molport.com) and NCI (237.771 compounds, https://www.nih.gov) retrieved as SMILES strings [81]. The libraries were prepared as follows: 1) With used built-in filters for lead-likeness [82] in the OpenEye tools (OpenEye Scientific Software, Santa Fe,

NM. http://www.eyesopen.com. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T.), 2) We discarded compounds in violation of the rule of 5 [83] and required strict atom typing. (From the MolPort dataset 2,026,118 unique entries remained, for the NCI 49,657) Additionally, the compounds were adjusted to pH 7 and desalted. 3) We performed a conformational expansion taking conformers of up to 15 kcal/mol into further consideration using OpenEye OMEGA [84]. This resulted in 123 conformers on average per molecule for MolPort compounds and an average of approximately 42 conformers for NCI compounds. 4) We prepared the selected RRM domain MD snapshot with the apopdb2receptor application with W134 as the central amino acid 5) Next, we conducted a virtual screening with the outlined pocket definition using OpenEye FRED [85] with standard precision settings. Conclusively, 1,188,185 MolPort, and 12,386 NCI compounds were successfully docked.
Finally, we considered only compounds with a CHEMGAUSS04 better -7.5, and a diverse subset of 13 compounds was selected based on a diverse set of non-covalent interactions with the RRM, visual inspection, availability, and costs for *in vitro* testing.

Enrichment. We developed a pharmacophore hypothesis based on compounds 13 and 15 by using the "Small-Molecule Drug Discovery Suite" (Schrödinger Release 2018-2: Maestro). The 2D-structures of compounds 13 and 15 were transformed into 3D-structures by using the Ligprep application in Maestro (settings described in Supplement 2). The two compounds were selected and the "Develop pharmacophore hypothesis" application was used to generate the pharmacophore model based on the selected structures (setting described in supplements 3). After ten pharmacophore models were generated, the highest-ranked model was manually edited. All features of the model were set to "required" which means that every compound that matches the pharmacophore model must exhibit all features of the model. For testing the selectivity of the pharmacophore model in a screening, a decoy set for compounds 13 and 15 was generated with the webserver DUD.E (database of useful decoys: enhanced) [86]. None of the generated decoys matched the criteria of the pharmacophore model during a test screening. Therefore, the hypothesis was regarded as sufficiently selective.

The model was used to screen the Molport database using the "Phase Ligand Screening" application in Maestro. The screening returned all molecules of the database with the respective screen scores. For eliminating structures with unfavorable properties, the scored structures were imported into Canvas (Schrödinger Release 2018-2: Canvas, Schrödinger, LLC, New York, NY, 2018). Only compounds with a screen score higher than 1.475 were retained. The remaining structures were converted into 3D structures and molecular descriptors were calculated using the QikProp application. Compounds that were unlikely to possess ADMET properties favorable for CNS drugs were removed (details in supplements 4). Additionally, all molecules that did not fulfill the rule of five (No more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, mass less than 500 daltons, logP not greater than 5) or that did not pass the PAINS2 filter were eliminated [52].

To select ten representative structures of the remaining compounds, the structures were clustered and the ten most diverse cluster centroids were selected. To allow clustering, the structures were translated into binary vectors by using the molprint2D algorithm. The Tanimoto metric was used as distance measure between the binary vectors. Based on a pairwise distance matrix a hierarchical clustering with linkage type "Mcquitty" was conducted. To determine the optimal number of clusters, the Kelley penalty score was calculated for each possible number of clusters. The set of clusters with the lowest score was considered as the optimal endpoint of the clustering.

The procedure described above was used to systematically sample compounds from the feature space constrained by compound 13 and 15. Additionally, some compounds were

sampled from a more focused feature space defined by highly similar derivatives of compound 13. Derivatives of compound 13 were returned by the MolPort search function after submitting a query with the IUPAC name of compound 13. Three very similar structures were selected based on visual inspection. The same explorative approach could not be used for compound 15 because the structure was not available in MolPort.

CD. *Selection criteria for the pharmacophore.* Some of the known TrmA inhibitors are based on the fact that methylase activity requires the cofactor S-adenosylmethionine (SAM or AdoMet) (K_m values between 0.002 - 0.018 mM in TrmA)[42]. The SAM precursor L-methionine and structurally related norleucine are weak TrmA binder (both K_m = 30 mM). The ethyl analog of L-methionine (L-ethionine) is a known non-competitive selective inhibitor. The related S-adenosyl-L-ethionine does not show such selectivity (K_m = 0.6 mM in TrmA)[43]. Ethylthioadenosine, where methionine is replaced by an ethyl group shows inhibition of K_m = 0.2 mM in TrmA. This indicated that both the adenosine and methionine substructure by themselves are sufficient for binding. Several polyamines were shown to inhibit to a different extent E. coli. TrmA, also depending on environmental conditions [44]. Among these, only L-ethionine is selective. Unfortunately, this compound was shown to be liver toxic in rats [45]. The other above-mentioned ligands are not toxic but they are not specific. Therefore, for the pharmacophore, we decided to explore the chemical space around L-ethionine to reduce toxicity, optimize pharmacological/pharmacokinetics profile in general, while retaining specificity. We specifically chose L-methionine and L-norleucine because they introduce variability in the hydrophobic moiety.
Thus we computed low energy conformers of the three selected known inhibitors with ChemAxon tools (Calculator Plugins were used for structure, property prediction, and calculation, Marvin 19.8, 2019, ChemAxon (http://www.chemaxon.com)) of the three compounds and PharmaGist [87] to align them and build the hypothesis. This pharmacophore hypothesis was used to filter the ZINC database [88]. ADMET properties were calculated with Canvas Version 2.8.014 (Schrödinger Release 2016-1: Canvas, Schrödinger, LLC, New York, NY, 2016.). Next, we calculated Qikprop properties and omitted all compounds violating the rule of 5, while retaining FDA-approved compounds. Compounds were sorted according to human oral availability and blood-brain barrier permeability. After visual inspection, several known inhibitors from the literature and pharmacophore derived compounds were selected for *in vitro* testing based on availability and associated costs.

4. Biophysical Experiments and Cell Assay

Surface Plasmon Resonance (SPR). SPR studies were performed using a BIACORE 3000 system (GE Healthcare). The TRMT2A RRM was diluted in 10 mM HEPES pH 7.5 for immobilization. Using coupling buffer (500 mM NaCl, 50 mM HEPES 7.5, 0.05% (v/v) Tween) the protein was amino coupled to a CM5 chip (GE-Healthcare) according to manufacturer's instruction reaching 3000 resonance units (RU).

Analysis of protein-compound interactions was performed at a flow rate of 30 µL/min in running buffer (150 mM NaCl, 50 mM HEPES 7.5, 0.05% (v/v) Tween). Analyte of interest was diluted in running buffer and injected for 2 minutes. To remove any residual attached compound, 2 x 2 min regeneration injections with 1 M NaCl were done in between runs. Concentration series reached from 0.5 µM to 8 µM for the positive control Spermine (Sigma Aldrich) and compound 12/Spermidine. For compounds 13, 15, 17, and 25 the concentration series reached from 0.5 µM to 32 µM.

Data were analyzed with the BIAevaluation software (GE Healthcare). Obtained binding curves were double-referenced against the signal in a protein-free reference channel and a buffer run. At equilibrium of the binding curves, the corresponding response was plotted against analyte concentration. The K_D was determined by fitting this curve to the steady-state affinity

model. All experiments were performed in quadruplets on different days. Representative results for each compound are shown.

Generation of HEKT293T with Stable RNAi-mediated Knockdown of TRMT2A

Stable RNAi-mediated knockdown of TRMT2A was achieved by infection of HEK293T cells with commercially available Lentiviral particles (MISSION® shRNA Lentviral Transduction Particles NM_182984.2-**1574**s1c1; Sigma-Aldrich). Cells with stable integration of the shRNA construct were determined by a selection of puromycin-resistant colonies (0.5 µg/ml puromycin; Invitrogen). Efficacy of *TRMT2A* silencing effect was determined by qPCR and Western blot. The cell line with the strongest reduction of TRMT2A (>90%; short sh1574) was selected for use in further experiments. The same procedure was used to generate a control cell line (short shK) expressing a scrambled shRNA (SHC002V; Sigma-Aldrich).

Transfections of HEKT293T cells. HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) FBS and 1% (v/v) penicillin/streptomycin. For transfection, cells were seeded at a density of 30 x 10^4 cells per well in a 6-well-plate. 24 h post seeding, cells were transiently transfected with a construct mediating expression of an elongated polyQ with 103 glutamines fused to GFP (polyQ103:GFP) using either Metafectene® (Biontex) or Calcium Chloride ($CaCl_2$). Metafectene® was used according to the manufacturer's instructions and $CaCl_2$-transfection was achieved as follows: The transfection mix (in relation to the total volume of cell culture medium per well) contained 1% $CaCl_2$ (1M), 4% ddH_2O, 5% 2 x HBSS; 0.075% (v/w) DNA. The mix was incubated for 5 min and added to the medium of cultured cells.

Inhibitor treatment. Inhibitors were added to fibroblasts (confluence of about 60%) by supplementation to the cell culture medium in a final concentration of 100 µM. In case of HEK293T cells, inhibitors were subjected to cell culture media in a final concentration of 100 µM four hours post transfection. Cell lysis was performed 48 hours after compound-treatment for further analysis.

Cell Death Assay Treated cells were mechanically detached from the plate by scraping and the cell/medium mixture was transferred to an Eppendorf tube. The cell/medium mixture was diluted in a 1:100 ratio with fresh DMEM medium. The diluted cell/medium mixture was supplemented with trypan blue solution (0.04%) in a 1:1 ratio. Cell viability was analyzed according to the dye exclusion test using the automated Cedex cell counter (Innovatis). For verification of observed differences in cell death, the effect of a few selected inhibitors was analyzed by a manual quantification of cell death. Here trypan blue stained cells were analyzed in the Haematocytometer (Neubauer chamber). GFP-positive cells were counted and the relative number of additionally blue cells (dead cells) was determined. 100 GFP-positive cells were analyzed per condition and three to six biological replicates were assayed.

Cell lysis and protein fractionation Cells were washed in PBS and then mechanically detached from the bottom of the plate in RIPA buffer (100 µl per six-well). Samples were centrifuged for 30 min at 13,300 rpm at 4 °C and supernatants were collected (RIPA soluble fraction). The protein concentration of the RIPA soluble fraction was determined using the Bradford protein assay (DC Protein assay, Bio-Rad) according to the manufacturer's instructions.

The pellet (RIPA insoluble fraction) was used for Dot Blot analysis. The pellet (RIPA insoluble fraction) was washed two times (resuspended and centrifuged for 30 min at 13,300 rpm at 4 °C in RIPA buffer). Finally, the pellet was resuspended in urea buffer (30 mM Tris, 7 M Urea, 2 M thiourea, 4% CHAPS pH - 8.5; 100 µl per six-well). Proteins in the pellet were solubilized by sonication (incubation in Ultrasonic Cleaner bath, VWR) for 10 min followed by an incubation for 60 min at 4 °C under rotation. Subsequently, the lysate was centrifuged at 4 °C, 9,000 x g

for 30 min. The supernatant was collected as urea soluble fraction and the urea insoluble pellet was discarded.

Western blotting 20-50 µg of the RIPA soluble fraction protein samples were supplemented with 1x Laemmli buffer and boiled 95 °C, 5 min. Samples were subjected to SDS-PAGE (100 V for 90 min). Resolved proteins were transferred onto nitrocellulose membrane (225 mA per gel for 60 min) by semi dry blotting. The membranes were then blocked with skim milk (5% in TBS-T for 60 min) followed by incubation with the primary antibody at 4°C overnight (in 5% skim milk in TBS-T). The primary antibodies used were mouse anti-Polyglutamine, Milipore MAB1574; mouse anti-Huntingtin polyQ, Developmental Hybridoma Bank, MW1 or mouse anti-GFP; Roche 11814460001. All antibodies were applied in a 1:1000 dilution. Membranes were washed three times for 10 min in TBS-T and incubated with the secondary horse raddish peroxidase (HRP) coupled antibody (GE Healthcare, NXA931V; in a 1:3000 dilution) for two hours under agitation at room temperature. Subsequently, the membranes were washed three times in TBS-T for 10 min and the chemiluminescence signal was detected using the Super Signal® West Femto Maximum Sensitivity Substrate (Thermo Scientific) detection kit according to manufacturer's instructions. Chemiluminescence signals were visualized using the Alliance UVltec system (Biometra) and processed using Image J software.

Filter retardation assay Relative to the total protein concentration in the RIPA soluble fraction (volumes kept constant), we loaded a volume that would correspond to 90 µg of protein in the RIPA soluble fraction. The total volume was adjusted to 50 µl by addition of RIPA buffer supplemented with 7 µl dot blot buffer (0.5 M Tris pH-6.8, 0.4% (w/v) SDS, 20% (v/v) glycerol, 0.2 M DTT). The samples were transferred to a nitrocellulose membrane (Protran®BA 83, Whatman, 0.2 µM pore size) by sucking the probes through the membrane using a vacuum pump attached to a dot blot chamber (Camlab, UK). The membrane was washed several times with TBS-T and blocked for 1 hour using 5% skim milk in TBS-T. Afterward, the membrane was treated as described for Western blotting using respective primary and secondary antibodies.

Acknowledgement

We like to thank Vera Roman for her technical support. This work was supported by a grant by the Federal Ministry of Education and Research (BMBF; to DN, project PolyQure).

Bibliography

1. Adegbuyiro A, Sedighi F, Pilkington AWt, Groover S, Legleiter J. Proteins Containing Expanded Polyglutamine Tracts and Neurodegenerative Disease. *Biochemistry* 2017, **56**(9)**:** 1199-1217.

2. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991, **352**(6330)**:** 77-79.

3. Burke JR, Ikeushi T, Koide R, Tsuji S, Lancet MYT. *Dentatorubral-pallidoluysian atrophy and Haw River syndrome*.

4. Paulson HL. The Spinocerebellar Ataxias. *Journal of neuro-ophthalmology : the official journal of the North American Neuro-Ophthalmology Society* 2009, **29**(3)**:** 227-237.

5. Walker FO. Huntington's disease. *The Lancet* 2007, **369**(9557)**:** 218-228.

6. Lieberman AP, Shakkottai VG, Albin RL. Polyglutamine Repeats in Neurodegenerative Diseases. *Annu Rev Pathol* 2019, **14:** 1-27.

7. Gusella JF, MacDonald ME. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat Rev Neurosci* 2000, **1**(2)**:** 109-115.

8. Bakthavachalam P, Shanmugam PST. Mitochondrial dysfunction - Silent killer in cerebral ischemia. *J Neurol Sci* 2017, **375:** 417-423.

9. Davis AK, Pratt WB, Lieberman AP, Osawa Y. Targeting Hsp70 facilitated protein quality control for treatment of polyglutamine diseases. *Cellular and Molecular Life Sciences* 2019**:** 1-20.

10. Tabrizi SJ, Leavitt BR, Landwehrmeyer GB, Wild EJ, Saft C, Barker RA*, et al.* Targeting Huntingtin Expression in Patients with Huntington's Disease. *N Engl J Med* 2019.

11. Voßfeldt H, Butzlaff M, Prüßing K, Chárthaigh R-A, Karsten P, Lankes A*, et al.* Large-Scale Screen for Modifiers of Ataxin-3-Derived Polyglutamine-Induced Toxicity in Drosophila. *PLoS ONE* 2012, **7**(11).

12. Chang YH, Nishimura S, Oishi H, Kelly VP, Kuno A, Takahashi S. TRMT2A is a novel cell cycle regulator that suppresses cell proliferation. *Biochem Biophys Res Commun* 2019, **508**(2)**:** 410-415.

13. Carter J-M, Emmett W, Mozos IR, Kotter A, Helm M, Ule J*, et al.* FICC-Seq: a method for enzyme-specified profiling of methyl-5-uridine in cellular RNA. *Nucleic Acids Research* 2019, **47**(19)**:** e113-e113.

14. Hou Y-M, Perona JJ. Stereochemical mechanisms of tRNA methyltransferases. *FEBS Letters* 2010, **584**(2)**:** 278-286.

15. Nordlund ME, Johansson MJ, von Pawel-Rammingen U, BystroeM AS. Identification of the TRM2 gene encoding the tRNA (m 5 U 54) methyltransferase of Saccharomyces cerevisiae. *Rna* 2000, **6**(6).

16. Towns WL, Begley TJ. Transfer RNA methytransferases and their corresponding modifications in budding yeast and humans: activities, predications, and potential roles in human health. *DNA Cell Biol* 2012, **31**(4)**:** 434-454.

17. Powell CA, Minczuk M. TRMT2B is responsible for both tRNA and rRNA m5U-methylation in human mitochondria. *bioRxiv* 2019**:** 797472.

18. Davanloo P, Sprinzl M, Watanabe K, Albani M, Kersten H. Role of ribothymidine in the thermal stability of transfer RNA as monitored by proton magnetic resonance. *Nucleic Acids Res* 1979, **6**(4)**:** 1571-1581.

19. Kersten H, Alni M, MÄNnlein E, Praisler R, Wurach P, Nierhaus KH. On the Role of Ribosylthymine in Prokaryotic tRNA Function. *European Journal of Biochemistry* 1981, **114**(2)**:** 451-456.

20. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 2002, **3**(3)**:** 195-205.

21. Loerch S, Kielkopf CL. Dividing and Conquering the Family of RNA Recognition Motifs: A Representative Case Based on hnRNP L. *Journal of Molecular Biology* 2015, **427**(19)**:** 2997-3000.

22. Clery A, Blatter M, Allain FH. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* 2008, **18**(3)**:** 290-298.

23. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 2005, **272**(9)**:** 2118-2131.

24. Konté Nd, Krepl M, Damberger FF, Ripin N, Duss O, Šponer J*, et al.* Aromatic side-chain conformational switch on the surface of the RNA Recognition Motif enables RNA discrimination. *Nature Communications* 2017, **8**(1).

25. Baker NA, Sept D, Joseph S, Holst MJ, McCammon AJ. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* 2001, **98**(18)**:** 10037-10041.

26. Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A. Cryptic binding sites on proteins: definition, detection, and druggability. *Current Opinion in Chemical Biology* 2018, **44:** 1-8.

27. Halperin I, Wolfson H, Nussinov R. Protein-Protein Interactions: Coupling of Structurally Conserved Residues and of Hot Spots across Interfaces. Implications for Docking. *Structure/Folding and Design* 2004, **12**(6)**:** 1027-1038.

28. Nowacka M, Boccaletto P, Jankowska E, Jarzynka T, Bujnicki JM, Dunin-Horkawicz S. RRMdb–an evolutionary-oriented database of RNA recognition motif sequences. *Database* 2019, **2019**.

29. Gavenonis J, Sheneman BA, Siegert TR, Eshelman MR, Kritzer JA. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nature Chemical Biology* 2014, **10**(9)**:** 716.

30. Li Z, Physics L-T. Water at biomolecular binding interfaces. *Physical Chemistry Chemical Physics* 2007.

31. Kovalenko A, Hirata F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chemical Physics Letters* 1998, **290**(1-3)**:** 237-244.

32. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* 2012, **28**(15)**:** 2074-2075.

33. Cimermancic P, Weinkam P, Rettenmaier TJ, Bichmann L, Keedy DA, Woldeyes RA*, et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol* 2016, **428**(4)**:** 709-719.

34. Kokh DB, Richter S, Henrich S, Czodrowski P, Rippmann F, Wade RC. TRAPP: a tool for analysis of transient binding pockets in proteins. *J Chem Inf Model* 2013, **53**(5)**:** 1235-1252.

35. Kokh DB, Czodrowski P, Rippmann F, Wade RC. Perturbation Approaches for Exploring Protein Binding Site Flexibility to Predict Transient Binding Pockets. *Journal of Chemical Theory and Computation* 2016, **12**(8).

36. Greener JG, Sternberg MJ. AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformatics* 2015, **16:** 335.

37. Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P*, et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* 2018, **7**.

38. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 2009, **49**(2)**:** 377-389.

39. Wang L, Friesner RA, Berne BJ. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J Phys Chem B* 2011, **115**(30)**:** 9431-9438.

40. Terakawa T, Kameda T, Takada S. On easy implementation of a variant of the replica exchange with solute tempering in GROMACS. *J Comput Chem* 2011, **32**(7)**:** 1228-1234.

41. Wermuth CG, Villoutreix B, Grisoni S, Olivier A, Rocher J-P. Chapter 4 - Strategies in the Search for New Lead Compounds or Original Working Hypotheses. In: Wermuth CG, Aldous D, Raboisson P, Rognan D (eds). *The Practice of Medicinal Chemistry (Fourth Edition)*. Academic Press: San Diego, 2015, pp 73-99.

42. Shugart L. Kinetic studies of Escherichia coli transfer RNA (uracil-5-)-methyltransferase. *Biochemistry* 1978, **17**(6)**:** 1068-1072.

43. Tscherne JS, Wainfan E. Selective inhibition of uracil tRNA methylases of E. coli by ethionine. *Nucleic Acids Res* 1978, **5**(2)**:** 451-461.

44. Ny T, LindstrÖM PHR, Hagervall TG, BjÖRk GR. Purification of transfer RNA (m5U54)-methyltransferase from Escherichia coli. *Eur J Biochem* 1988, **177**(3)**:** 467-475.

45. Lu LW, Chiang GH, Randerath K. Effects of DL-ethionine on mouse liver tRNA base composition. *Nucleic Acids Res* 1976, **3**(9)**:** 2243-2253.

46. Wanker EE, Scherzinger E, Heiser V, Sittler A, Eickhoff H, Lehrach H. Membrane filter assay for detection of amyloid-like polyglutamine-containing protein aggregates. *Methods Enzymol* 1999, **309:** 375-386.

47. McGurk L, Berson A, Bonini NM. Drosophila as an In Vivo Model for Human Neurodegenerative Disease. *Genetics* 2015, **201**(2)**:** 377-402.

48. Marsh JL, Lukacsovich T, Thompson LM. Animal models of polyglutamine diseases and therapeutic approaches. *J Biol Chem* 2009, **284**(12)**:** 7431-7435.

49. Pouladi MA, Morton AJ, Hayden MR. Choosing an animal model for the study of Huntington's disease. *Nat Rev Neurosci* 2013, **14**(10)**:** 708-721.

50. Riess O, Rub U, Pastore A, Bauer P, Schols L. SCA3: neurological features, pathogenesis and animal models. *Cerebellum* 2008, **7**(2)**:** 125-137.

51. Illarioshkin SN, Klyushnikov SA, Vigont VA, Seliverstov YA, Kaznacheyeva EV. Molecular Pathogenesis in Huntington's Disease. *Biochemistry (Mosc)* 2018, **83**(9)**:** 1030-1039.

52. Baell JB, Nissink WJM. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017-Utility and Limitations. *ACS chemical biology* 2017, **13**(1)**:** 36-44.

53. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Research* 2019, **47**(W1).

54. Mach M, Kersten H, Kersten W. Regulation of tRNA methyltransferase activities by spermidine and putrescine. Inhibition of polyamine synthesis and tRNA methylation by alpha-methylornithine or 1,3-diaminopropan-2-ol in Dictyostelium. *Biochem J* 1982, **202**(1)**:** 153-162.

55. Beglov D, Hall DR, Wakefield AE, Luo L, Allen KN, Kozakov D*, et al.* Exploring the structural origins of cryptic sites on proteins. *Proceedings of the National Academy of Sciences* 2018, **115**(15)**:** E3416-E3425.

56. Schueler-Furman O, Wodak SJ. Computational approaches to investigating allostery. *Current Opinion in Structural Biology* 2016, **41:** 159-171.

57. Lightfoot HL, Hall J. Endogenous polyamine function--the RNA perspective. *Nucleic Acids Research* 2014, **42**(18)**:** 11275-11290.

58. Rm L, A S, T B, M G, D N, Cf B*, et al.* Translating Antisense Technology Into a Treatment for Huntington's Disease. *Methods Mol Biol* 2018, **1780**(16 Suppl)**:** 497-523.

59. Kabsch W. XDS. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(Pt 2)**:** 125-132.

60. Evans P. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* 2006, **62**(1)**:** 72-82.

61. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR*, et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* 2011, **67**(Pt 4)**:** 235-242.

62. French S, Wilson K. On the treatment of negative intensity observations. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 1978, **34**(4)**:** 517-525.

63. Panjikar S, Parthasarathy V, Lamzin VS, Weiss, Tucker PA. Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica Section D: Biological Crystallography* 2005, **61**(4)**:** 449-457.

64. Panjikar S, Parthasarathy V, Lamzin VS, Weiss, Tucker PA. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallographica Section D: Biological Crystallography* 2009, **65**(10)**:** 1089-1097.

65. Project N. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D: Biological Crystallography* 1994, **50**(5)**:** 760-763.

66. Sheldrick GM. Macromolecular phasing with SHELXE. *Zeitschrift für Kristallographie - Crystalline Materials* 2002, **217**(12)**:** 644-650.

67. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography* 2006, **62**(9)**:** 1002-1011.

68. Cowtan K. Fitting molecular fragments into electron density. *Acta Crystallographica Section D: Biological Crystallography* 2008, **64**(1)**:** 83-89.

69. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallographica Section D: Biological Crystallography* 2000, **56**(Pt 8)**:** 965-972.

70. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D: Biological Crystallography* 1997, **53**(3)**:** 240-255.

71. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N*, et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(2)**:** 213-221.

72. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(Pt 4)**:** 486-501.

73. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 1993, **26**(2)**:** 283-291.

74. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H*, et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**(21)**:** 2947-2948.

75. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 2001, **17**(8)**:** 700-712.

76. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 2015, **11**(8)**:** 3696-3713.

77. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983, **79**(2)**:** 926-935.

78. Wallnoefer HG, Handschuh S, Liedl KR, Fox T. Stabilizing of a Globular Protein by a Highly Complex Water Network: A Molecular Dynamics Simulation Study on Factor Xa. *J Phys Chem B* 2010, **114**(21)**:** 7405-7412.

79. Adelman SA, Doll JD. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *Journal of Chemical Physics* 2008, **64**(6)**:** 2375.

80. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 1977, **23**(3)**:** 327-341.

81. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comp Sci* 1988, **28**(1)**:** 31-36.

82. Oprea TI. Property distribution of drug-related chemical databases*. *J Comput Aided Mol Des* 2000, **14**(3)**:** 251-264.

83. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* 2004, **1**(4)**:** 337-341.

84. Hawkins PC, Skillman AG, Warren GL, Ellingson BA, Stahl MT. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* 2010, **50**(4)**:** 572-584.

85. McGann M. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model* 2011, **51**(3)**:** 578-596.

86. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 2012, **55**(14)**:** 6582-6594.

87. Schneidman-Duhovny D, Dror O, Inbar Y, Nussinov R, Wolfson HJ. PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Res* 2008, **36**(Web Server issue)**:** W223-228.

88. Sterling T, Irwin JJ. ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model* 2015, **55**(11)**:** 2324-2337.