nature machine intelligence

Supplementary information

https://doi.org/10.1038/s42256-025-01082-y

Target-specific de novo design of drug candidate molecules with graph-transformer-based generative adversarial networks

In the format provided by the authors and unedited

Table of Contents

List of Figures and Tables

Supplementary Text

S1. A short review of the fundamental literature on AI-based molecule design

Most of the initial AI-based de novo design methods focused on designing valid molecules that are typically small in terms of molecular size (via training the models on datasets such as QM9 (~9 heavy atoms on average) [1,2] and ZINC250K (~23 heavy atoms on average) [3], which is an easier task compared to designing molecules that have the same size distribution as the existing drugs and bioactive molecules (~29 heavy atoms in ChEMBL molecules, on average).

One of the first generative modelling architectures used in de novo molecule design was variational autoencoders (VAE) [4]. Gomez-Bombarelli et al. devised a molecule generation method based on VAE, where the encoding network converts discrete SMILES expressions of molecules into continuous vectors, with the decoder network subsequently reconstructing SMILES from this continuous space. Furthermore, an additional network was implemented into the system that guides the decoder by predicting properties such as drug-likeness and synthetic accessibility of the representations in the latent space [3]. Another generative modelling architecture, Generative Adversarial Networks (GAN) [5], originally developed for image generation, has been employed to design de novo molecules. GANs are trained via a battle between generator and discriminator networks in a zero-sum game,

where each agent tries to beat the other one by performing better at each move. The model called MolGAN uses a multilayer perceptron-based generator and graph convolutional discriminator to handle the molecule generation process [6]. This method was one of the first studies to implement GANs for de novo drug design. A few GAN-based models, similar to MolGAN, were implemented in the following years [7]. Another study set the training objective as predicting the masked node and edge labels on molecular graphs to enhance the efficiency of the molecular design process [8]. Denoising diffusion probabilistic models have surfaced as a promising approach as a part of the recent advancements in generative modelling [9,10]. Diffusion models learn the underlying data distribution by constructing a forward diffusion process and utilising a denoising model. One notable study by Hoogeboom et al. [11] proposes an equivariant diffusion model (EDM) for molecule generation in 3-D. The model uses forward diffusion to add noise to input atom coordinates and types and reverse diffusion to denoise the positions and types. This approach enables the generation of diverse molecular structures in the learned Euclidean space. Although diffusion models have the potential to provide high-quality samples, training and inference with them can be challenging due to difficulties in convergence, lengthy iterations, and the resource-intensive nature of computations. While other 3-D graph- or coordinate-based models exist [12,13], their higher computational demands and inconsistent performance on complex problems underscore the reliability advantage of 2-D graph-based models. This causes an overall preference for 2-D graph-based models over their 3-D counterparts due to their ability to offer cost-effective modelling with high generalisation capabilities, facilitating rapid analysis of extensive datasets.

S2. About protein kinase B (PKB) / AKT

One of the intriguing proteins with biomedical relevance to be targeted could be the AKT protein, also known as protein kinase B (PKB). AKT is an evolutionarily conserved serine-threonine kinase that plays a critical role in the development and progression of cancer [14]. AKT is a key component of the PI3K/AKT signalling pathway, which regulates various cellular processes such as proliferation, cell growth, metabolism, and survival [15]. The three isoforms—AKT1, AKT2, and AKT3—have distinct functions and tissue expressions, sharing a structurally conserved composition (85-90% sequence similarity) [16]. The shared compositions comprise an N-terminus PH domain, a central kinase domain, and a C-terminus regulatory domain encompassing a hydrophobic motif [17]. In numerous cancer types, the activation of AKT signalling is frequently observed, leading to increased cell survival and uncontrolled proliferation [18–23]. Therefore, inhibition of AKT signalling has emerged as a promising therapeutic approach for various cancers, as targeting this pathway could inhibit tumour growth and improve the efficacy of other treatments. AKT1 inhibitors may compete with adenosine triphosphate (ATP) for its binding site in the kinase domain (Supplementary Fig. 1). Alternatively, they may go for allosteric binding sites (i.e., the pleckstrin homology - PH domain) [24] (Supplementary Fig. 1). Some important ATP-competent AKT inhibitors that reached clinical trials include uprosertib (GSK2141795), ipatasertib (GDC-0068), and capivasertib (AZD5363) [17]. Recently, the latter was approved by the FDA [25]. Capivasertib and ipatasertib were also crystallised within the ATP binding site of the AKT1 protein (PDB IDs: 4GV1 [26] and 4EKL [27], respectively). Numerous studies in industry and academia highlight the current need for novel inhibitors that effectively target the AKT pathway.

S3. Overall performance evaluation and comparison (details)

Considering individual metrics, DrugGEN displayed high performance on most fronts and did not suffer from issues such as low validity, novelty, or uniqueness—problems that afflict many competing models. Among the competing approaches, some are based on GANs. ORGAN relies on a GAN structure composed of an RNN (serving as the generator) and a CNN (serving as the discriminator) to produce conditioned molecules, while STAGAN was introduced to mitigate the mode collapse observed in conventional GANs; however, it exhibits diminished validity when processing considerably larger molecular structures within the ZINC250k dataset (with approximately 23 heavy atoms on average). In comparison, DrugGEN achieves a superior overall ranking (1 versus 12 for ORGAN and 3 for STAGAN) by maintaining robust validity (0.931, compared to 0.379 for ORGAN and 0.482 for STAGAN) alongside nearly perfect novelty (0.991, compared to 0.687 for ORGAN). On the other hand, graph-based methods offer an alternative route: MARS, which combines a graph neural network with MCMC sampling for multi-objective optimisation, shows extremely high validity (0.997), yet its novelty (0.333) remains low, yielding an overall ranking of 8, while MGM—based on an MPNN and operating as a masked graph model—achieves moderate novelty (0.722) and perfect uniqueness, although its validity (0.849) is lower than that of DrugGEN, resulting in an overall ranking of 4.

Furthermore, moving to transformer-based models, MolGPT employs the conventional sequence-based transformer decoder architecture—originally developed for natural

language processing and trained on SMILES representations of molecules—although its novelty score (0.782) remains subpar. Additionally, CMGN, a transformer encoder–decoder model designed to incorporate target-specific information for generating inhibitor candidates, faces limitations in terms of uniqueness (0.409) and novelty (0.585), resulting in an overall ranking considerably lower than that of DrugGEN. Lastly, TRIOMPHE-BOA, a transformer-based variational autoencoder enhanced with Bayesian optimisation, demonstrates a relatively high QED score (0.819). However, its low uniqueness (0.445) and limited internal diversity (0.586) suggest that it captures only a narrow segment of the molecular space, resulting in an overall performance that does not match the balanced metrics achieved by DrugGEN.

Similarly, RL-based approaches such as QADD, MolDQN, and REINVENT typically aim to optimise specific properties (for example, QED) but often exhibit low novelty because they tend to generate molecules already present in the training set. DrugGEN, on the other hand, produces highly original compounds (novelty 0.991, compared to 0.307 for REINVENT, 0.341 for QADD, and 0.360 for MolDQN) while also balancing multiple metrics to secure the top overall ranking. Related to this, we also calculated the average maximum and mean pairwise Tanimoto-based similarity values between DrugGEN's de novo molecules and all real molecules in the ChEMBL dataset and found 55% and 11% similarity, respectively—underscoring the high originality of our de novo molecules. DrugGEN utilises graph transformers with modified attention modules, likely contributing to its robust performance.

In addition, a separate group of architectures directly integrates 3D structural information of both drugs and/or proteins. For instance, RELATION employs an encoder–decoder framework that uses 3D grid representations of ligand–receptor complexes along with bidirectional transfer learning to generate target-specific molecules, achieving competitive performance (overall ranking 2) but with slightly lower validity (0.854) than DrugGEN; Pocket2Mol leverages an E(3)-equivariant graph neural network combined with an efficient conditional sampling algorithm to generate molecules from detailed 3D protein pocket features—although it achieves excellent validity (0.962), its low uniqueness (0.365) and modest QED (0.400) imply that it tends to generate similar, non-drug-like compounds; ResGen, which employs a hierarchical autoregression approach via a residue–atom parallel multiscale modelling scheme with SE(3)-equivariant modules, shows markedly lower uniqueness (0.289) and QED (0.218) and thus falls short relative to other models; and TargetDiff adopts a non-autoregressive 3D diffusion framework with an SE(3)-equivariant network to jointly generate atom coordinates and types, ensuring spatial consistency with the protein binding site, yet its overall performance (ranking 3) remains below that of DrugGEN, particularly in terms of validity. Finally, EDM (equivariant diffusion model), which failed to generate valid molecules (0.106) when trained on ChEMBL molecules in our analysis—likely due to the elevated complexity of the drug-like molecules in our ChEMBL training dataset (with an average of ~28.5 heavy atoms compared to ~20 in the GEOM dataset originally used for training EDM)—demonstrates that even models specifically designed for 3D data can struggle when faced with complex molecular structures. Moreover, the remaining architectures include BIMODAL, which generate molecules from SMILES strings using recurrent neural network architecture incorporating leveraging bidirectional sequence generation)—attain high validity (0.997) but suffer from low novelty (0.314), resulting in overall ranking of 9 that are considerably inferior to DrugGEN's balanced performance.

The QED metric highlights a model's ability to generate compounds with desirable drug-like properties; while some models (e.g., QADD, molDQN, and MARS) in Table 1 specifically aim to maximise QED values of their de novo molecules—hence those methods are omitted from the QED-based comparison—DrugGEN does not explicitly optimise for QED. However, it still generates molecules with relatively high QED values. Finally, the internal diversity (IntDiv) metric indicates the structural diversity among generated samples. DrugGEN exhibited notably high performance in IntDiv, demonstrating its ability to learn varied molecular structures from the training dataset.

S4. Molecular docking and Drug-likeness analysis (details)

To examine target binding characteristics of molecules on the AKT1 and CDK2 proteins, we conducted a molecular docking analysis (details are provided in the Methods section) on the molecules generated by DrugGEN and competing targeted generation methods, i.e., RELATION, TRIOMPHE-BOA, ResGen, Pocket2Mol and TargetDiff. Table 2 and Fig. 3A display median docking scores (i.e., binding free energies - ΔG in kcal/mol) of the top 10% of molecules with the best binding properties from each group/model, where a lower binding free energy value indicates a stronger affinity. Binding free energies for all molecules are given as histograms in Supplementary Fig. 4.

In Fig. 3A, the percentages shown next to each box represent the docking performance of generative models in comparison to the real AKT1/CDK2 inhibitors. For this, we applied min-max normalisation to the median docking score of each model (parameters: maximum=-8.39 and minimum=-3.56 kcal/mol, which are the median of the top and bottom 10% docking scores of real AKT1 inhibitors, respectively). DrugGEN achieved the highest performance with 99.98% and -8.39 kcal/mol (Fig. 3A), exhibiting a noteworthy contrast to other target-based models, especially TRIOMPHE-BOA and ResGen. Furthermore, the difference between the DrugGEN and the next best model, RELATION (i.e., -8.39 and -8.10 kcal/mol, respectively), was found to be statistically significant according to the Mann-Whitney U rank test (p-value < 0.0001). Here, DrugGEN struck a good balance between drug-like properties and target binding, making it a highly competitive model. Following DrugGEN and RELATION, TargetDiff and Pocket2Mol exhibit satisfactory docking scores (91.59% and -7.98 kcal/mol for TargetDiff, and 91.10% and -7.96 kcal/mol for Pocket2Mol). They also rank below DrugGEN and RELATION in most drug-likeness metrics. TRIOMPHE-BOA displayed the weakest docking performance (-6.92 kcal/mol) but achieved the highest scores in drug-likeness, including QED, SA, and Lipinski's and Veber's rules. This indicates that while the generated compounds possess favourable pharmacokinetic properties, they are not well-suited for targeting a specific protein. ResGen performed poorly in both docking and drug-likeness evaluations. Remarkably, DrugGEN surpassed the majority of the compared models and exhibited comparable performance to RELATION in terms of fragment (0.721) and scaffold (0.246) similarities, as well as the FCD score (20.416). It generated de novo molecules with structural features resembling known inhibitors (Table 2), while also being novel and diverse (Table 1).

For CDK2, Pocket2Mol exhibited the strongest binding affinity with a normalised docking performance of 88.21% and a median docking score of -8.80 kcal/mol, closely followed by DrugGEN at 86.99% (-8.75 kcal/mol). The difference in docking scores was statistically

significant (Mann-Whitney U test, $p < 0.001$). However, Pocket2Mol showed suboptimal drug-likeness characteristics, with lower QED and higher synthetic accessibility (SA) scores (Table 2), indicating challenging synthetic complexities. Also, DrugGEN outperformed Pocket2Mol in Lipinski's and Veber's rules, as well as PAINS filters, suggesting a higher potential for yielding pharmacologically promising compounds. When comparing DrugGEN to TargetDiff, DrugGEN exhibited significantly better docking performance (Mann-Whitney U test, $p < 0.0001$). On the other hand, TargetDiff had high QED and PAINS scores, indicating that it generated chemically attractive molecules. Furthermore, fragment and scaffold similarities and the FCD value of DrugGEN (0.931, 0.311 and 10.611, respectively) highlight its ability to perform meaningful structural transformations while preserving a high degree of novelty compared to the benchmark models. This further confirms that DrugGEN can generate molecules whose physicochemical properties closely align with known CDK2 inhibitors.

S5. The effect of input molecules on the generative performance

DrugGEN operates by taking real molecules at the input level. To investigate how target-specific properties change based on molecular structures and characteristics, we employed our trained model and generated two sets of de novo molecules: one using known/real AKT1 inhibitors and the other using randomly selected ChEMBL molecules as inputs. We then compared the generated molecules in terms of their structural similarity to real AKT1 inhibitors and their docking properties. Molecules inferred from real AKT1 inhibitors demonstrated high fragment similarity (0.712) and moderate scaffold similarity (0.395), with an FCD of 12.723, indicating considerable structural and physicochemical overlap with existing AKT1 inhibitors. In contrast, molecules generated using random ChEMBL molecules showed comparable fragment similarity (0.714), a lower scaffold similarity (0.042), a higher FCD of 21.248 (lower is better). While both groups maintain similar fragment-level characteristics, the marked difference in scaffold similarity suggests that molecules inferred from random ChEMBL possess distinct structural cores. This architectural divergence positions these molecules as promising candidates for novel scaffold discovery, potentially expanding the structural diversity of therapeutic approaches. Whereas, utilising known AKT1 inhibitors as input during de novo generation can yield high bioactivity values, especially suitable for molecule optimisation. The advantages of these approaches differ, enabling preference based on specific objectives.

We also checked whether DrugGEN improves the binding characteristics of real AKT1 inhibitors when used as input molecules. For this, we compared the docking scores of de novo molecules with their corresponding input AKT1 inhibitor inputs. We found that top 10% de novo molecules have improved the docking scores of their inputs (i.e., -8.232 and -7.680 kcal/mol, respectively), although the real AKT1 inhibitors has a slightly better score when we directly consider the top 10% molecules on both sides, without following up with the starting molecules (i.e., -8.232 and -8.387 kcal/mol, for de novo and real inhibitors, respectively). This indicates that, in general, DrugGEN can generate novel molecules that dock as well as the known, real inhibitors of the target. Supplementary Fig. 3 shows three example real AKT1 inhibitors used as input, the corresponding de novo molecules generated from them, and their respective docking, QED and SA scores. In these examples, the de novo molecules achieved better docking scores than their references; however, reference

molecules tend to be more synthesisable, as reflected in their lower synthetic accessibility (SA) scores. Also, drug-likeness (QED) varies.

S6. Ablation study (details)

In this analysis, we evaluated the impact of (i) the graph transformer architecture and (ii) its specialised attention mechanism on molecular generation performance. We compared the outputs of various DrugGEN and DrugGEN-NoTarget model variants with each other, as well as with the molecules in the training datasets. For consistency, we employed the same set of fundamental benchmarking metrics as in previous sections (Validity, Novelty, Uniqueness, IntDiv, QED, and SA). All models were trained using identical hyperparameter value ranges and datasets, with modifications applied only to the core module responsible for learning molecular representations. Notably, for each model variant, the best-performing model was selected based on training metrics. Furthermore, no additional post-processing steps—i.e., SMILES correction and filtering—were performed during the molecule generation process. Consequently, the results reported here may differ from those presented in other sections of the text. In addition, all metrics, with the exception of validity, were calculated solely on the subset of valid molecules. We generated 10,000 de novo molecules from each model and ran them through the metrics mentioned above. Table S1 summarises the performance of each model variant in terms of these fundamental benchmarking metrics.

To evaluate the impact of the graph transformer architecture, we utilised six distinct models: (i) DrugGEN: the default model; (ii) DrugGEN-NoTarget: a DrugGEN model that generates only random molecules featuring the same architecture as DrugGEN, but with a discriminator that accepts random real molecules as input instead of target-specific inhibitors (with the sole objective of generating drug-like molecules); (iii) MLP baseline: a targeted baseline model employing multi-layer perceptron (MLP) modules for both the generator and discriminator; (iv) MLP-NoTarget baseline: a non-targeted baseline model based on MLPs; (v) GCN baseline: a targeted baseline model employing graph convolutional network (GCN) modules, for both the generator and discriminator; and (vi) GCN-NoTarget baseline: a non-targeted baseline model based on GCNs. These baseline models are differentiated from DrugGEN by replacing the graph transformer encoder blocks with either simple MLP layers or GCN layers (please refer to the “Methods” section for further details). The purpose of comparing the primary DrugGEN model and its non-targeted version with the MLP- and GCN-based baselines is to highlight the advantages of employing graph transformers for molecule generation.

DrugGEN’s generator module takes random molecules as the initial input (i.e., starting molecules). Therefore, measuring a second novelty score against the starting molecules used during the generation process is also essential. We named this metric novelty at inference (NI) and provided the scores in Table S1 in the same column as the regular novelty measure (i.e., novelty against training data). We observed that NI scores are slightly lower than the standard novelty values, which was expected since the input of the inference run has a more direct effect on generated samples. It is also important to note that the NI scores of DrugGEN are still higher than the regular novelty scores of most of the models listed in Table 1.

In the target-specific setting, although the MLP-based targeted baseline exhibits relatively favourable quantitative estimation of drug-likeness (QED) scores, its validity score (0.512) is notably lower than that of DrugGEN (0.713) (Table S1). In contrast, the GCN-based targeted baseline achieves a higher validity score (0.754) but demonstrates poor drug-likeness metrics, as indicated by a markedly low QED (0.271) and a high synthetic accessibility (SA) score (5.215), suggesting that the GCN layers are less effective at capturing the drug-likeness related features of molecular structures. In the non-targeted setting, the DrugGEN-NoTarget model exhibits moderate performance. In contrast, the corresponding MLP-NoTarget baseline displays an even lower NI value (0.201), indicating a tendency to reproduce its input molecules, while the GCN-NoTarget variant suffers from relatively low validity (0.639) along with suboptimal QED (0.135) and SA (6.429) scores. These comparisons underscore that the graph transformer architecture within DrugGEN provides a more effective trade-off between structural validity and desirable molecular properties. Clearly, the use of graph transformer blocks contributes substantially to generating valid molecules with balanced drug-like properties.

Secondly, we performed an attention mechanism-based analysis to assess how various approaches for incorporating molecular graphs' edges (i.e., the atomic bonds) into attention affect the generative performance. We tested five attention variants: (i) $Att \odot A_m \odot (A_m + 1)$ (the official DrugGEN model), (ii) $Att \odot A_m$ (the original graph transformer), (iii) $Att \odot (A_m + 1)$, (iv) $Att \odot A_{mul} + A_{add}$ (FiLM-inspired modulation [28]), and (v) Att (without edge-based modification). Where Att denotes the standard attention matrix computed from the transformer's scaled dot product of query and key matrices, before applying a softmax, A_m denotes the adjacency matrix, and A_{mul} , A_{add} are multiplicative and additive terms used in FiLM-based edge modulation. The purpose of comparing different attention mechanism variants—including the official DrugGEN model, FiLM-inspired modulation, edge modulation, edge+1 modulation, and the version without edge-based modification—is to systematically evaluate how incorporating molecular bond information into the transformer's attention matrix influences the generative performance of the models, thereby demonstrating that an effective edge incorporation scheme is essential for capturing long-range structural dependencies and generating high-quality, target-specific molecules.

As shown in Table S1, the official DrugGEN model exhibits a substantially higher validity (0.713) compared to other attention variants (0.110, 0.464, and 0.117 for the film, edge, and edge+1 variants, respectively). This finding suggests that the edge incorporation scheme is effective in learning molecular structures. Additionally, the variant without edge-based modifications shows a dramatic drop in validity (0.081), underscoring the critical role of bond information in generating valid molecular structures. Although some alternative variants may yield higher QED scores, these results should be interpreted with caution because the scores were computed solely on valid molecules, which was considerably low for models with reduced validity. Furthermore, the lower NI scores observed for these variants (0.503, 0.527, and 0.677 for the film, edge, and edge+1 variants, respectively) suggest that the apparent improvements in QED and SA do not necessarily reflect a genuine increase in the diversity or quality of the generated molecules. In contrast, the official DrugGEN model, with its notably high NI value of 0.924, more reliably produces molecules that are distinct from their starting points. Overall, these results demonstrate that the specific design of the attention mechanism, which incorporates the $Att \odot A_m \odot (A_m + 1)$ formulation, is pivotal in

enabling the model to capture long-range structural dependencies and generate high-quality, target-specific molecules.

Collectively, the ablation results confirm that the graph transformer architecture, along with the optimised attention mechanism design, is essential for achieving superior targeted molecular generation performance relative to both MLP- and GCN-based alternatives.

S7. Analysis of physicochemical distributions

We drew histograms (density plots) displaying the physicochemical property value distributions of real molecules in our training dataset and de novo generated molecules from DrugGEN models and other target-based generation models (Supplementary Fig. 5). Details of this analysis are given in Supplementary Material (S7). In summary, de novo molecules belonging to the DrugGEN model and real AKT1 inhibitors share a similar region of the physicochemical property space in general (i.e., DrugGEN's distribution shape and mean more closely resemble those of real AKT1 inhibitors compared to other models). This finding indicates that DrugGEN has effectively learned the physicochemical properties of the real inhibitors of the selected target protein. The details are provided below.

In Supplementary Fig. 5, each panel displays a different physicochemical property. In the left side plot of each panel, non-targeted molecules are compared (i.e., all ChEMBL molecules in our training dataset vs non-targeted de novo molecules generated by DrugGEN-NoTarget). In contrast, on the right side plot of each panel, targeted molecules are compared (i.e., the real AKT1 inhibitors vs targeted de novo molecules generated by RELATION, TargetDiff, Pocket2Mol, TRIOMPHE-BOA, ResGen and DrugGEN models). The results given for the DrugGEN model are based on real AKT1 inhibitor starting molecules. It is observed from Supplementary Fig. 5 that, in general, property distributions of non-targeted de novo molecules (i.e., DrugGEN-NoTarget) are similar to ChEMBL molecules, which represent this model’s training dataset, and distributions of targeted de novo molecules (i.e., DrugGEN) resemble real AKT1 inhibitors. The real AKT1 inhibitors exhibit higher molecular weights and a greater number of heavy atoms and aromatic rings than ChEMBL molecules (e.g., Supplementary Fig. 5A, B and H). The same pattern is also evident in DrugGEN molecules when compared with DrugGEN-Notarget. Among the target-specific models, DrugGEN has the closest distributions to AKT1 inhibitors, followed by Pocket2Mol, which demonstrated mean values closely aligned with AKT1 inhibitors for heavy atoms, aromatic rings and molecular weight. RELATION and TargetDiff exhibited slight deviations from the AKT1 inhibitor distributions, while TRIOMPHE-BOA’s molecules are concentrated within a narrow range due to how molecule generation is handled by this model (i.e., producing derivatives of a single molecule). Finally, the ResGen model exhibits the distributions most distant from the real AKT1 inhibitors. TargetDiff also exhibited a substantial divergence, particularly considering the aromatic rings.

Considering individual properties, DrugGEN and Pocket2Mol successfully learned the hydrogen bond acceptor (HBA). At the same time, DrugGEN and TargetDiff most accurately captured the hydrogen bond donor (HBD) value distributions of AKT1 inhibitors, as demonstrated in Supplementary Fig. 5E and F. Apart from that, Supplementary Fig. 5C displays that ChEMBL and AKT1 molecules exhibit almost identical AlogP values (i.e., a

measure of the relative solubility of a molecule in two liquids: octanol and water); however, DrugGEN and DrugGEN-Notarget models both yielded a shifted mean compared to the distributions of their training datasets. RELATION achieved the closest AlogP values to AKT1 inhibitors, whereas TargetDiff and TRIOMPHE-BOA stand out as outliers. Furthermore, polar surface area (PSA) (Supplementary Fig. 5D) and rotatable bond (RB) (Supplementary Fig. 5G) value distributions of DrugGEN and DrugGEN-Notarget do not reflect the ratio between AKT1 and ChEMBL datasets; therefore, we conducted a statistical analysis to observe whether there is a notable difference between DrugGEN and DrugGEN-Notarget, although the distributions of their training datasets could not be captured. According to the Mann-Whitney U rank test, PSA distributions (Supplementary Fig. 5D) of DrugGEN and DrugGEN-Notarget are significantly different from each other (p-value < 0.0001); however, the same outcome was not obtained for RB distributions (Supplementary Fig. 5G) (p-value > 0.05). For the polar surface area, Pocket2Mol showed the greatest alignment with AKT1 inhibitors, whereas TargetDiff, TRIOMPHE-BOA and ResGen models showed higher mean values than AKT1 inhibitors. Conversely, for rotatable bonds, these models exhibited lower means compared to AKT1 inhibitors, except for TargetDiff, which achieved the best alignment considering this property. In contrast, RELATION displayed a density distribution similar to DrugGEN's.

The relatively modest performance of DrugGEN-NoTarget in this analysis could be explained by the fact that its training dataset (i.e., random ChEMBL molecules) is highly diverse regarding physicochemical properties since it is not limited to a particular target.

S8. Evaluation of DrugGEN molecules using Lipinski, Veber and PAINS filters

To assess their drug-related properties, de novo molecules generated by DrugGEN are subjected to the following filters: (i) Lipinski's Rule of 5 [29], (ii) Veber's rules [30], and (iii) the pan-assay INterference compoundS (PAINS) [31] filter, which identifies and eliminates false positives in biological screening assays such as redox cyclers, toxoflavins, polyhydroxylated natural phytochemicals, etc. In this analysis, we utilised the same 10,000 de novo generated molecules subjected to the benchmarking analysis. In the end, 41.15% (4,115 out of 10,000) of the de novo molecules successfully passed through the filters. Subsequently, we subjected ChEMBL and AKT1 molecules to the same filters for comparison. It is observed that 57.18% (1,342 out of 2,347) of AKT1 inhibitors and 50.80% (50,802 out of 100,000) of randomly sampled ChEMBL dataset molecules have passed, providing roughly similar percentages to DrugGEN molecules.

S9. Docking analysis of the synthesised molecules (details)

Molecule 1 (MOL_02_045762) exhibited the highest binding affinity (-8.948 kcal/mol), forming hydrogen bonds with Ala230 and Asp292, stabilised by a salt bridge with Asp292. Molecule 2 (MOL_02_045795) (-7.686 kcal/mol) demonstrated a halogen bond with Glu228, hydrogen bonds with Thr291 and Asp292, and a salt bridge with Asp292. Molecule 5 (MOL_02_045627) (-5.341 kcal/mol) displayed moderate binding, primarily through hydrogen bonds with Lys163, Lys179, Glu278, and Thr291, along with a salt bridge with Asp292. Molecule 4 (MOL_02_045609) (-8.383 kcal/mol) exhibited hydrogen bonds with

Lys158, Lys179, and Asn279, salt bridges with Glu234 and Asp292, and a π-cation interaction with Lys179. Molecule 3 (MOL_02_045758) (-8.655 kcal/mol) formed hydrogen bonds with Thr211, Glu228, and Glu234, along with salt bridges involving Asp292. Molecule 1, 2, 4, and 5 show hydrophobic interactions as well, while Molecule 3 does not exhibit any (Fig. 4B).

S10. Investigation of the binding properties of synthesised molecules using molecular dynamics (details)

DFG-in motive of the structure was chosen to search for binding poses via docking. In this scenario, the polar interactions with the backbone atoms of the hinge domain residues, which help maintain the DFG motive as the -in conformation, as well as residence within the binding cavity, are important for AKT1 inhibitory effects.

As seen in the Root Mean Square Deviation (RMSD) values (Extended Data Fig. 2C and Supplementary Fig. 8), all compounds remained stable, with values consistently below 1.6 Å. Similarly, Root Means Square Fluctuation (RMSF, Extended Data Fig. 2D and Supplementary Fig. 8) values were low and no exceptional movements were observed in most simulations, except for one copy of Molecule 2. Throughout conducted simulations with 4 copies and 500 ns, capivasertib's pyrrolopyrimidine moiety maintained hydrogen bonds for nearly the entire duration of the simulation (99%) with both Tyr229 and Ala230. Additional hydrogen bonds were formed between the aliphatic hydroxypropyl group and Glu278 (74%) residue in the catalytic loop. Water mediated hydrogen bonds were also observed with Lys158, Asn279, Asp292 residues, occurring on more than 30% of the simulations (Figure 4D).

The most potent de novo generated compound, Molecule 1, also formed hydrogen bonds between isoquinoline moiety and hinge domain residue Ala230 (96%). Its quaternary amine interacted with the surrounding acidic residues; Glu234 (29%), Glu278 (11%), Asp292 (46%). An additional hydrogen bond was observed between carbonyl group of the linker amide and the side chain of Lys179 (15%) (Extended Data Fig. 2B). Since the compound exhibited the most potent AKT1 inhibitory effect in the low μM range among the de novo generated compounds, we hypothesize that this is due to its shared interaction network with capivasertib. However, as its inhibitory value hasn't reached nanomolar range, we further hypothesize that this is related to the absence of additional contacts with hinge domain and a lack of strong hydrogen bond networks within the binding cavity. Molecule 2, on the other hand, forms halogen bonds instead of hydrogen bonds with the hinge domain residues Tyr229 (p-chlorine 30%) and Ala230 (o-chlorine 17%, p-chlorine 19%). Additionally, it forms hydrogen bonds with surrounding residues, including Lys179(23%), Glu234 (17%), Glu278 (23%), Thr291 (13%), Asp292 (21%) (Extended Data Fig. 2B). Molecule 2 exhibits the second most potent inhibitory effect within the μM range against the AKT1 enzyme.

Molecule 3 include an aminopurine moiety and oxolane ring, similar to ATP. While its structure shares the hydrogen bond network with the residues surrounding the cavity, it hasn't shown efficacy under 100 μM (Supplementary Fig. 8). We hypothesize that this is due to the compound's short length and the presence of agonist-like fragmentation. Molecule 4 lacks heterocycles capable of forming polar interactions with the hinge domain, with all its

polar interactions concentrated within the core of the binding cavity. This observation, combined with the absence of hinge domain interactions, may explain its lack of efficacy below 100 μM (Supplementary Fig. 8). Similarly, Molecule 5 was found incapable of forming hydrogen bonds with the hinge domain residues. Its structure features aminopurine ring at the core, which forms hydrogen bonds with the surrounding polar residues, and oxolanediol moiety forms hydrogen bonds with the glycine-rich loop. However, the weak hydrogen bond network within the cavity and absence of interactions with the hinge domain might be the reason to show its lack of efficacy below 100 μM (Supplementary Fig. 8).

We also conducted the same MD analysis with an additional de novo showcase molecule from the list of 30 promising compounds apart from the in vitro tested ones, to show that our positive results were not isolated cases. For this, we selected MOL_01_027820 (Extended Data Fig. 1) and evaluated its binding characteristics with AKT1. We compared it with the simulation of the reference co-crystal complex structure “AKT1 - capivasertib”. According to results, both compounds maintained stable binding conformations within the AKT1 binding site. Further details about this experiment are provided in Supplementary Material (S11) and Supplementary Fig. 9.

S11. Investigation of binding properties of the additional showcase molecule using molecular dynamics (details)

We showcase another molecule among the 30 (shown with the identifier: “MOL_01_027820” in Extended Data Fig. 1), which can be denoted as a “Pyrrolo[1,2-a]pyrimidin-4(1H)-one” derivative. Supplementary Fig. 9A displays the reference crystal complex structure of AKT1 with the ligand capivasertib (model PDB ID: 4GV1) [26], for which the binding free energy was measured as -9.517 kcal/mol in the molecular docking analysis (on the same ligand conformer). Supplementary Fig. 9B shows the best docking pose of MOL_01_027820 with the binding free energy -9.686 kcal/mol obtained via the same docking protocol. This molecule is investigated further below to comprehend its binding-related dynamic properties.

Capivasertib (Supplementary Fig. 9A) and the showcased de novo molecule “AKT1 - MOL_01_027820” (Supplementary Fig. 9B) share the binding pattern with Ala177, Lys179 and Thr291. Hydrophobic interactions are more prominent in the AKT1 - MOL_01_027820 complex, whereas hydrogen bonds are evident in the crystal structure. This may contribute to the difference in binding affinities. The replacement of the adenine ring of the ATP with various heterocyclic moieties resulted in selective and potent inhibition of kinases, and this mode of binding constitutes more than 90% of kinase inhibitors [32]. These inhibitors bind the same binding pocket as ATP and mimic the adenine group’s interactions with the hinge residues. Here, we see a similar orientation with capivasertib and MOL_01_027820 in the AKT1 kinase binding site. However, the hydrogen bond interactions with the backbones of the hinge residues were not directly observed in the docking-derived binding pose of MOL_01_027820 (Supplementary Fig. 9B). To further evaluate the relaxation and adaptation of the binding site and to reveal binding characteristics of capivasertib and MOL_01_027820, 500 ns long molecular dynamics (MD) simulations were conducted with four copies (please see “Methods” for details).

The output of MD simulations is shown in Supplementary Fig. 9, panels C to I. The AKT1 structure mostly consists of α helices, and visual inspection of the simulations displayed the conservation of secondary structures during the simulations of both complexes. As seen in RMSD plots (Supplementary Fig. 9 panels F to I), both simulations (with their simulation copies) reached stable binding patterns. Hydrogen bonding networks between the hinge domain residues Glu228 and Ala230 occurred with high ratios in all simulations. Van der Waals interactions were also observed with the hydrophobic moieties of the surrounding residues. The gatekeeper residue Met227 remained closed and supported the residence of the ligand within the binding cavity for both complexes. Additionally, Lys179 in β-sheet continued to form a salt bridge with Glu234 in the αC-helix located at the N terminal domain of AKT1. This was a supporting factor to keep the ligand within the binding site. With the help of this ionic interaction, the protein stayed in DFG-in and αC-helix in conformation in the simulations.

Capivasertib remained stable during all simulations (RMSD <1, Supplementary Fig. 9 panels F and H). The interaction analysis of capivasertib showed that hydrogen bonds with the hinge domain (via Glu228 -99%- and Ala230 -99%- residues) were conserved during the simulation. An additional hydrogen bond was seen with Glu278 (74%). MOL_01_027820 showed two different binding patterns (pattern-1 and -2, shown in Supplementary Fig. 9 panels D and E, respectively). Those patterns were seen in two individual copies of the simulations. Pattern-1 was observed in the second and the third replica simulations, and pattern-2 was observed in the first and the fourth replicas (Supplementary Fig. 9G). The first pattern was similar to the docking-derived starting conformation ($RMSD_{average}$ = 1.2, Supplementary Fig. 9 panels D, G, and I) and relatively close to the binding pattern with the compound capivasertib. However, the second pattern showed 110.2 degrees of dihedral tilting ($RMSD_{average}$ 3.1, Supplementary Fig. 9 panels E, and G) following atom number 15 and above (Supplementary Fig. 9I) but still conserved the essential interactions with the hinge domain and remained within the binding cleft and conserved the protein's general conformation. Further comments about MD results of the showcase molecule can be found below.

<u>Capivasertib–AKT1:</u>

Capivasertib remained stable during all simulations (RMSD <1, Supplementary Fig. 9 panels F and H). The interaction analysis of capivasertib showed that hydrogen bonds with the hinge domain (via Glu228 -99%- and Ala230 -99%- residues) were conserved during the simulation. An additional hydrogen bond was seen with Glu278 (74%). Also, several water bridges were observed between the amine group located on the piperidine ring and Glu278 (74%), Asp292 (90%) residues, amide group and Lys158 (33%), Glu234 (11%), Asn279 (28%), Asp292 (36%) residues, and hydroxyl group and the Lys276 (19%) residue. The quaternary amine group formed a salt bridge with Glu294 (94%), and the p-Chlorobenzyl group showed cation-π interactions with Lys179 (11%) (Supplementary Fig. 9C).

<u>MOL_01_027820–AKT1:</u>

MOL_01_027820 showed two different binding patterns (pattern-1 and -2, shown in Supplementary Fig. 9 panels D and E, respectively). Those patterns were seen in two individual copies of the simulations. Pattern-1 was observed in the second and the third

replica simulations and pattern-2 was observed in the first and the fourth replicas (Supplementary Fig. 9G). The first pattern was similar to the docking-derived starting conformation ($RMSD_{average}$ = 1.2, Supplementary Fig. 8 panels D, G, and I) and relatively close to the binding pattern with the compound capivasertib. However, the second pattern showed 110.2 degrees of dihedral tilting ($RMSD_{average}$ 3.1, Supplementary Fig. 9 panels E, and G) following atom number 15 and above (Supplementary Fig. 9I) but still conserved the essential interactions with the hinge domain and remained within the binding cleft and conserved the protein's general conformation. The hydrogen interactions with Glu228 (54% and 65%, respectively, given for pattern-1 and -2) and Ala230 (83% and 66%, respectively, given for pattern-1 and -2) backbone atoms were formed (which did not exist in the docking-derived pose) with high occupancy values in the hinge domain. However, the remaining part of the molecule showed torsional rotations observed after the 12th and, more importantly, after the 16th atoms (Supplementary Fig. 9I). Therefore, slight differences in the bound clefts and the interaction networks were seen during simulations. Pattern-1 showed that the hydroxyl group connected to the fluorobenzyl moiety forms hydrogen bonds with Glu234 (44%) and water bridges with Lys158 (48%), Glu234 (20% and 63%), and Tyr437 (52%) with the same residue. The second hydroxyl group located at the centre of the ligand showed hydrogen bonds with Leu156 (22%) and Glu234 (16%) and water bridges with Glu234 (16%) and Leu156 (19%) residues. π-π interactions were seen with fluorophenyl and Phe237 (53%) and Phe442 (18%) residues (Supplementary Fig. 9D). In pattern-2, the water network around the acid hole (Glu234 and Asp292) was mostly broken, and only two water bridges remained with Leu158 (37%) and Glu234 (10%). Hydrogen bonds were observed between the central hydroxyl group and Leu156 (11%), Glu234 (67%) residues and amine group in the heterocyclic ring and Asp292 (10%) residue. The rotated conformation was found to form Van der Waals interactions with the hydrophobic portions of the surrounding lysine residues and orthogonal multipolar interactions with fluorine atoms of the ligand and carbonyl group of Lys163 (Supplementary Fig. 9E).

S12. Experimental procedures for enzymatic assays (details)

Based on Reaction Biology's radiometric ^{33}PanQinase^{TM} activity assay, the IC_{50} profiles of the synthesised de novo compounds were retrieved against AKT1 and the IC_{50} values were determined by testing ten different concentrations of each compound, ranging from 100 µM to 3 nM with semi-logarithmic dilution, in singlicate. The compounds were later dissolved to 0.01 M with DMSO. 100 µl of those stock solutions were transferred into the master plate. 100% DMSO were also used as a control group.

As a parameter for assay quality, the Z´-factor [33] for the low and high controls of each assay plate (n = 8) was used. While the low control includes substrate but no enzyme and no test compound, the high control includes the enzyme, substrate, but not the test compound. RBE´s criterion for repetition of an assay plate is a Z´-factor below 0.4 [34]. The Z´-factor for this study was calculated as 0.79. Additionally, staurosporine, a protein kinase inhibitor, was tested in parallel as a quality control. Staurosporine's IC_{50} value was in the expected range for the protein kinase. AKT1 enzyme was produced from human cDNAs and affinity tags were removed during the purification. The purity of the enzyme was examined by SDS-PAGE/Coomassie staining, and the structure was controlled by mass spectroscopy.

The assay was performed on 96-well ScientiPlates™ from PerkinElmer (Boston, MA, USA) in 50 µL reaction volume containing 11.9 nM AKT1. The assay environment was also composed of 25 µL of assay buffer (standard buffer/[γ-^{33}P]-ATP), 10 µL of ATP solution in H_2O (1 µM), 5 µL of test compound (in 10 % DMSO), 10 µL of AKT1/GSK3(14-27) mixture where the amount of the substrate is 2 µg. The reaction cocktails were incubated at 30°C for 1 hour. The reaction was stopped with 50µL of 2 % (v/v) H_3PO_4. The plates were aspirated and washed twice with 200 µL 0.9 % (w/v) NaCl. Incorporation of $^{33}P_i$ was determined with microplate scintillation counter (Microbeta, Wallac). The assay buffer contains 70 mM HEPES-NaOH pH 7.5, 3 mM $MgCl_2$, 3 mM $MnCl_2$, 3 µM Na-orthovanadate, 1.2 mM DTT, 50 µg/ml PEG_{20000} and [γ-^{33}P]-ATP contains approximately $8x10^5$ cpm per well. Below is Supplementary Equation (SE1), which outlines the calculation of response for each concentration.

$$\text{Remaining Activity}(\%) = 100 \times [(\text{cpm of compound} - \text{low control}) / (\text{high control} - \text{low control})] \quad \text{(SE1)}$$

The remaining activities of each concentration of the compound and the IC_{50} values were calculated using *Quattro Workflow V3.1.1* (Quattro Research GmbH, Munich, Germany; www.quattro-research.com). The IC_{50} values were determined using a Sigmoidal response (variable slope) with parameters ranging from 100% (for the top value) to 0% (for the bottom value), and fitted with least-squares using 10 data points (n = 10). Hill slopes were reported, and those with a value higher than -0.4 are indicative that the curve is not sigmoidal, is very flat, or is not descending. The results are provided in Supplementary Fig. 17.

S13. Attention map visualisation (additional information)

Transformer-based models weigh intrinsic pairwise relationships between data tokens through the attention mechanism, which can be utilised to visualise the important relationships in the data and, therefore, interpret what the model learns. Adopting this strategy, we visualised the average attention scores (produced by the transformer encoder module of the generator network) for each atom in the selected de novo generated molecules to assess critical atoms and bonds according to the model's perspective. We aimed to observe the correspondence between the atoms that received high attention scores and those interacting with AKT1 protein's binding pocket residues according to the molecular docking analysis. A high correspondence would indicate that the model learned which atoms are critical for enabling physical interaction with the target.

In Fig. 5, interactions between the atoms of de novo molecules and residues in the AKT1 binding pocket are visualised via 2-D protein-ligand interaction diagrams for three different de novo molecules (Fig. 5A, C, and E), using a 4 Angstrom cutoff. Also, in Fig. 5, the same de novo molecules' atoms with the highest average attention scores are visualised with green, yellow and red, indicating the importance appointed by the model (Fig. 5B, D, and F). The de novo molecules employed in this analysis were randomly selected from the 30 promising molecules given in Extended Data Fig. 1.

Considering all three cases in Fig. 5, the model identifies most of the atoms interacting with the target residues by assigning high attention scores. For example, in Fig. 5A and B, N^+H atoms forming both salt bridges and hydrogen bonds with the residue Glu234, and HN and N

atoms forming hydrogen bonds with the residues Asp274 and Ala230, respectively, are assigned the highest attention values (i.e., the green coloured ligand atoms). The same pattern is observed for other interactions in Fig. 5. Overall, 20 out of 22 (~91%) ligand atoms in pairwise interactions were detected by the generator network of DrugGEN. Looking at the same examples from the other angle, 20 out of 24 (~%83) ligand atoms with top attention scores were in contact with the target residues, thus playing a critical role in the interaction. Even though the model did not use any data regarding pairwise molecular interactions during training and validation, it still managed to learn substructures crucial for binding to AKT1 and designed molecules accordingly. We provided two additional examples in Supplementary Fig. 10 with a similar outcome.

References

1. Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 1–7 (2014).
2. Ruddigkeit, L., Van Deursen, R., Blum, L. C. & Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
3. Gómez-Bombarelli, R. *et al.* Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
4. Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *Found. Trends® Mach. Learn.* **12**, 307–392 (2019).
5. Goodfellow, I. *et al.* Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).
6. De Cao, N. & Kipf, T. MolGAN: An implicit generative model for small molecular graphs. (2018) doi:10.48550/ARXIV.1805.11973.
7. Zou, J., Yu, J., Hu, P., Zhao, L. & Shi, S. STAGAN: An approach for improve the stability of molecular graph generation based on generative adversarial networks. *Comput. Biol. Med.* **167**, 107691 (2023).
8. Mahmood, O., Mansimov, E., Bonneau, R. & Cho, K. Masked graph modeling for molecule generation. *Nat. Commun.* **12**, 3156 (2021).
9. Ho, J., Jain, A. & Abbeel, P. Denoising Diffusion Probabilistic Models. in *Advances in Neural Information Processing Systems* (eds. Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H.) vol. 33 6840–6851 (Curran Associates, Inc., 2020).
10. Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N. & Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. (2015) doi:10.48550/ARXIV.1503.03585.
11. Hoogeboom, E., Satorras, V. G., Vignac, C. & Welling, M. Equivariant Diffusion for Molecule Generation in 3D. (2022) doi:10.48550/ARXIV.2203.17003.
12. Peng, X. *et al.* Pocket2mol: Efficient molecular sampling based on 3d protein pockets. in

International Conference on Machine Learning 17644–17655 (PMLR, 2022).

13. Schneuing, A. *et al.* Structure-based drug design with equivariant diffusion models. *ArXiv Prepr. ArXiv221013695* (2022).
14. Mroweh, M. *et al.* Targeting Akt in Hepatocellular Carcinoma and Its Tumor Microenvironment. *Int. J. Mol. Sci.* **22**, 1794 (2021).
15. Manning, B. D. & Cantley, L. C. AKT/PKB Signaling: Navigating Downstream. *Cell* **129**, 1261–1274 (2007).
16. Matheny, R. W. & Adamo, M. L. Current Perspectives on Akt Akt-ivation and Akt-ions. *Exp. Biol. Med.* **234**, 1264–1270 (2009).
17. Martorana, F. *et al.* AKT Inhibitors: New Weapons in the Fight Against Breast Cancer? *Front. Pharmacol.* **12**, 662232 (2021).
18. George, B. *et al.* AKT1 Transcriptomic Landscape in Breast Cancer Cells. *Cells* **11**, 2290 (2022).
19. Rychahou, P. G. *et al.* Akt2 overexpression plays a critical role in the establishment of colorectal cancer metastasis. *Proc. Natl. Acad. Sci.* **105**, 20315–20320 (2008).
20. Banno, E. *et al.* Clinical significance of Akt2 in advanced pancreatic cancer treated with erlotinib. *Int. J. Oncol.* **50**, 2049–2058 (2017).
21. Bellacosa, A. *et al.* Molecular alterations of theAKT2 oncogene in ovarian and breast carcinomas. *Int. J. Cancer* **64**, 280–285 (1995).
22. Stahl, M., Guba, W. & Kansy, M. Integrating molecular design resources within modern drug discovery research: the Roche experience. *Drug Discov. Today* **11**, 326–333 (2006).
23. Chen, J. *et al.* Differential regulation of AKT1 contributes to survival and proliferation in hepatocellular carcinoma cells by mediating Notch1 expression. *Oncol. Lett.* (2018) doi:10.3892/ol.2018.8193.
24. Thomas, C. C., Deak, M., Alessi, D. R. & van Aalten, D. M. High-resolution structure of the pleckstrin homology domain of protein kinase b/akt bound to phosphatidylinositol (3, 4, 5)-trisphosphate. *Curr. Biol.* **12**, 1256–1262 (2002).
25. Turner, N. C. *et al.* Capivasertib in Hormone Receptor–Positive Advanced Breast Cancer. *N. Engl. J. Med.* **388**, 2058–2070 (2023).
26. Addie, M. *et al.* Discovery of 4-Amino- *N* -[(1 *S*)-1-(4-chlorophenyl)-3-hydroxypropyl]-1-(7 *H* -pyrrolo[2,3- *d*]pyrimidin-4-yl)piperidine-4-carboxamide (AZD5363), an Orally Bioavailable, Potent Inhibitor of Akt Kinases. *J. Med. Chem.* **56**, 2059–2073 (2013).
27. Lin, K. *et al.* An ATP-Site On-Off Switch That Restricts Phosphatase Accessibility of Akt. *Sci. Signal.* **5**, (2012).
28. Perez, E., Strub, F., de Vries, H., Dumoulin, V. & Courville, A. FiLM: visual reasoning with a general conditioning layer. in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial*

Intelligence 3942–3951 (AAAI Press, New Orleans, Louisiana, USA, 2018).

29. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).

30. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).

31. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).

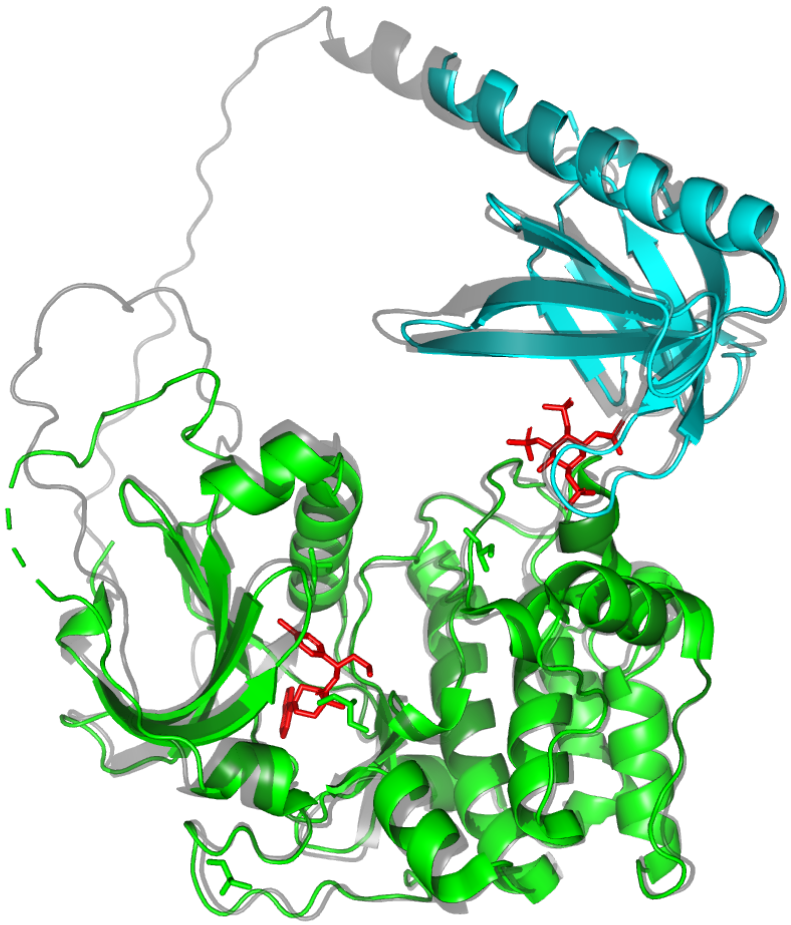
32. Sharma, V. & Gupta, M. Designing of kinase hinge binders: A medicinal chemistry perspective. *Chem. Biol. Drug Des.* **100**, 968–980 (2022).

33. Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *SLAS Discov.* **4**, 67–73 (1999).

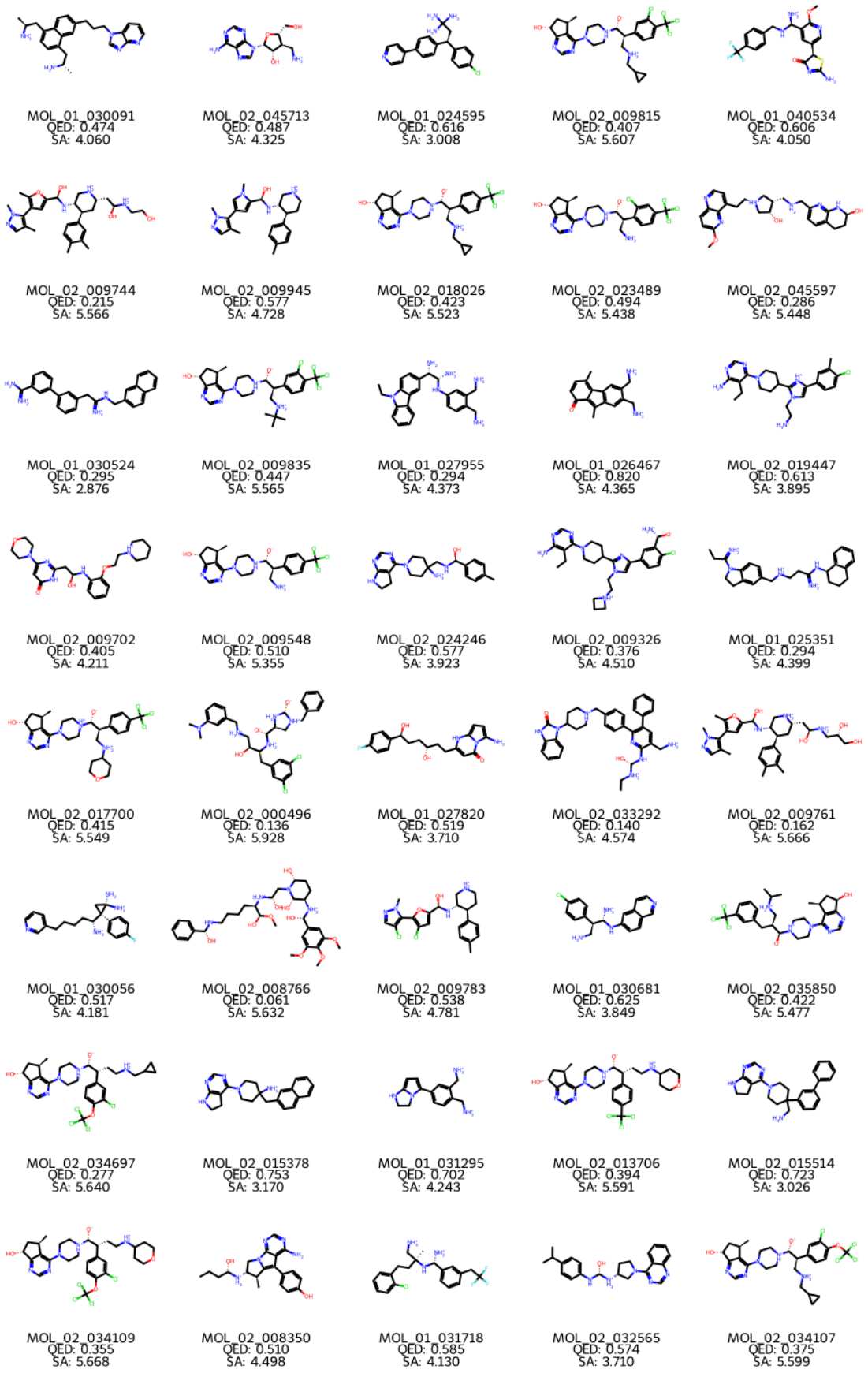
34. Iversen, P. W., Eastwood, B. J., Sittampalam, G. S. & Cox, K. L. A Comparison of Assay Performance Measures in Screening Assays: Signal Window, Z′ Factor, and Assay Variability Ratio. *SLAS Discov.* **11**, 247–252 (2006).

35. Van Linden, O. P. J., Kooistra, A. J., Leurs, R., De Esch, I. J. P. & De Graaf, C. KLIFS: A Knowledge-Based Structural Database To Navigate Kinase–Ligand Interaction Space. *J. Med. Chem.* **57**, 249–277 (2014).

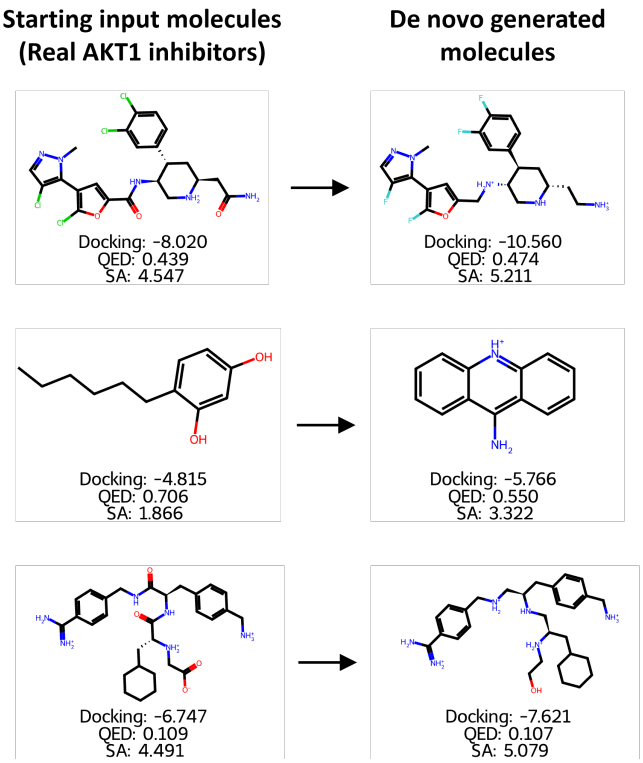
Supplementary Figures



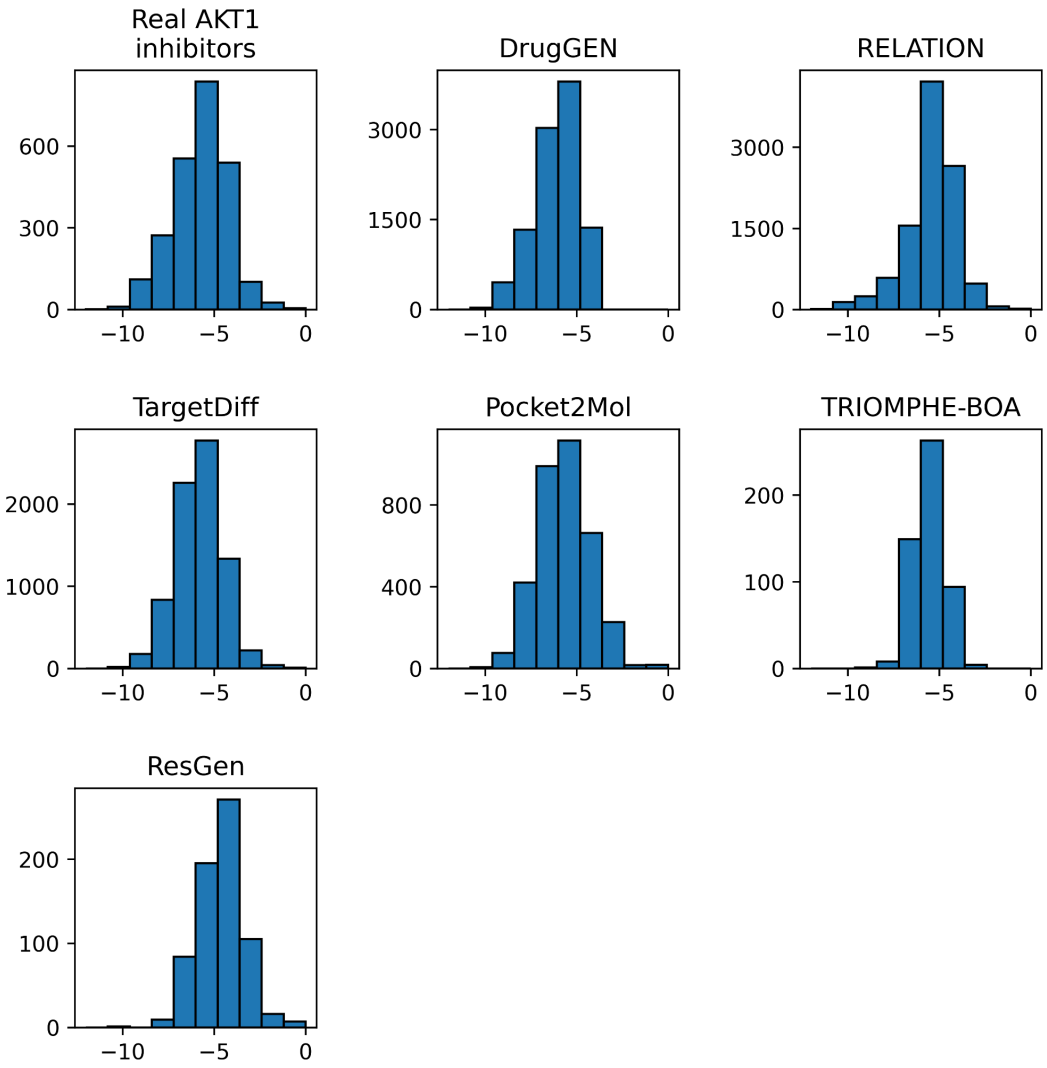
Supplementary Fig. 1: General overview of full AKT1 protein structure (gray: AlphaFold structure based on Uniprot ID: P31749, green: kinase (orthosteric) domain (PDB ID: 4GV1), cyan: PH (allosteric) domain (PDB ID: 1H10), red: ligands).



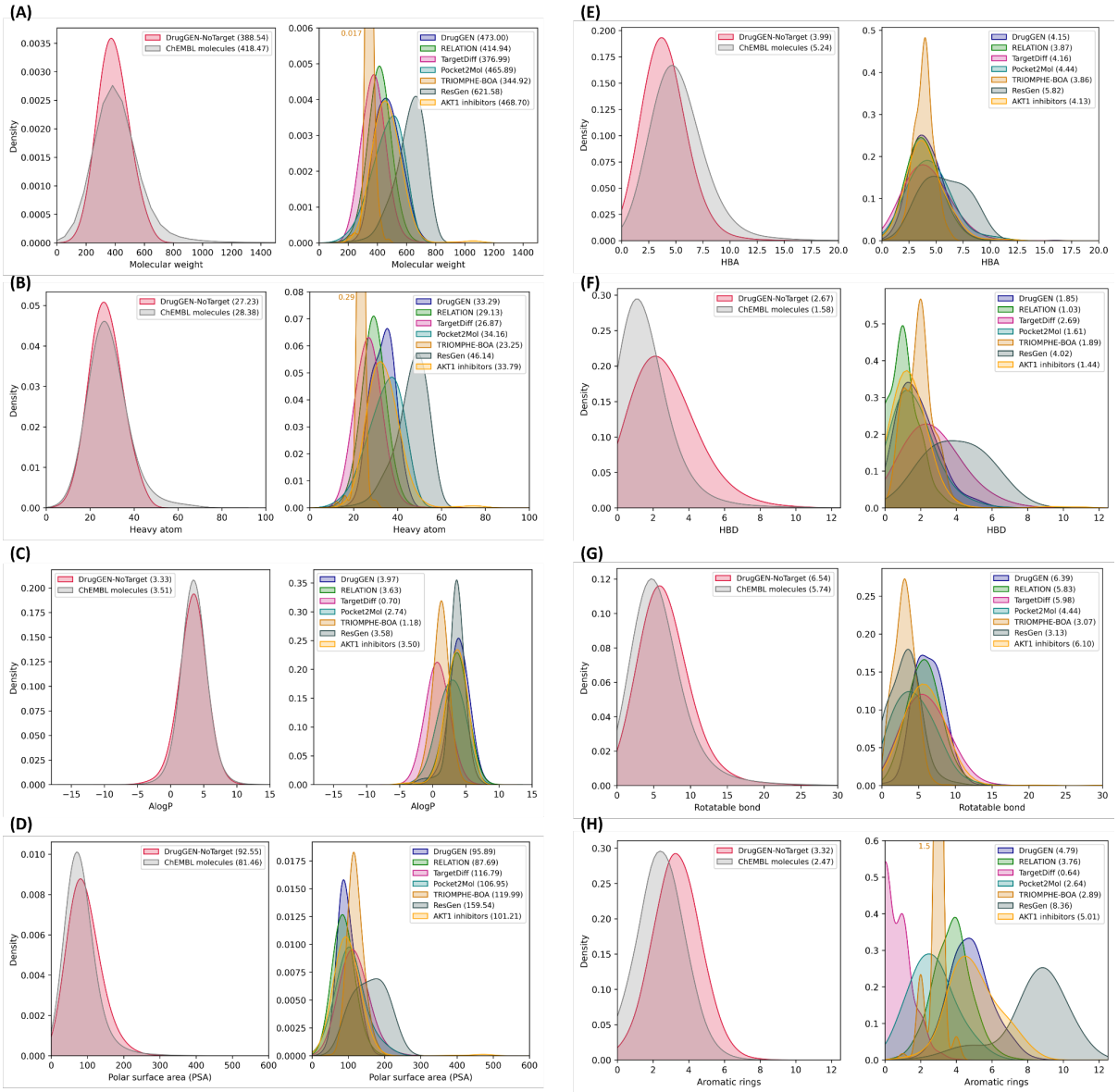
Supplementary Fig. 2: Forty molecules with predicted docking scores surpassing that of the native ligand capivasertib in the crystal complex structure of AKT1 “4GV1” (i.e., -9.517 kcal/mol).



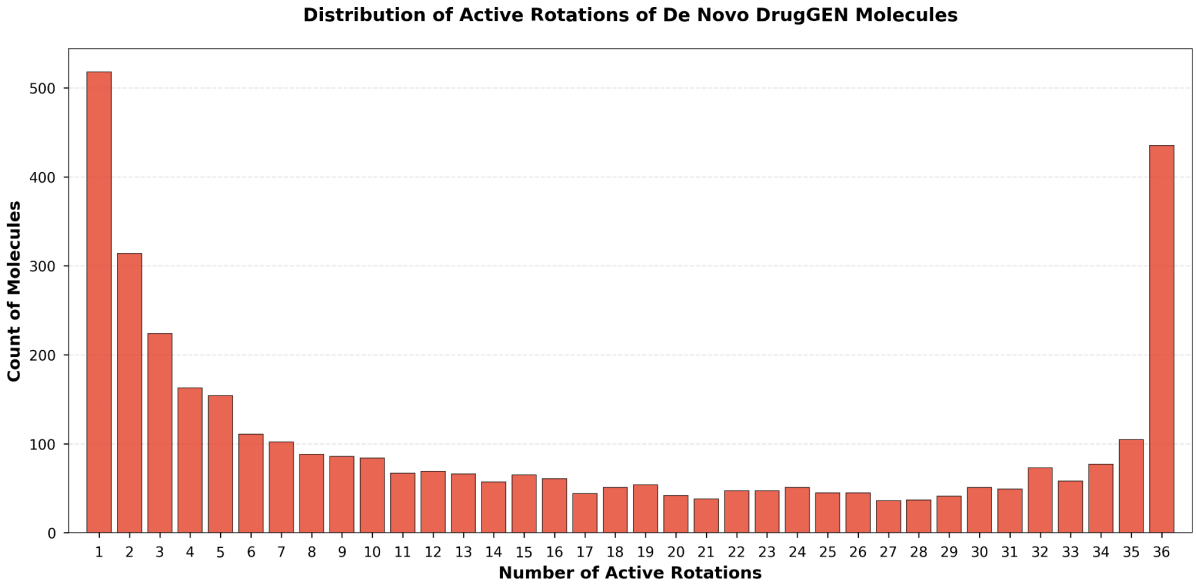
Supplementary Fig. 3: Three selected starting input molecules and their corresponding de novo-generated molecules. The left column shows the reference real AKT1 inhibitors, while the right column displays the de novo-generated molecules along with their docking scores, QED, and SA values. For the first pair, the Tanimoto similarity is 0.238 for Morgan fingerprints and 0.794 for MACCS, with Cl-to-F substitutions having a considerable impact on similarity. For the second pair, the lowest similarity (0.073 for Morgan and 0.109 for MACCS) indicates the highest structural novelty. For the last pair, scores of 0.435 for Morgan and 0.755 for MACCS suggest a moderate transformation, with the removal of oxygen atoms with double bonds contributing to the difference.



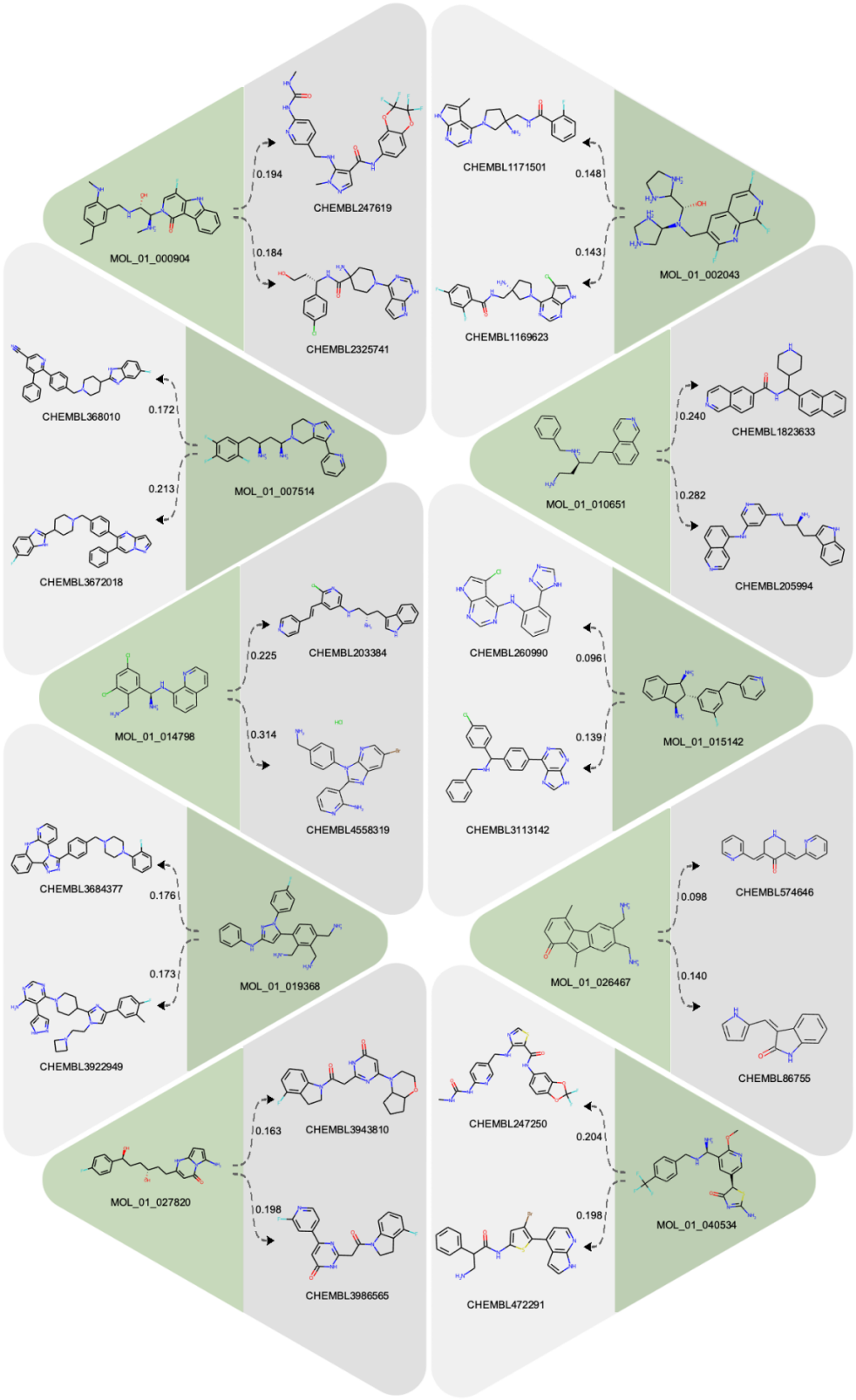
Supplementary Fig. 4: Docking score distributions of real AKT1 inhibitors, DrugGEN and other target-centric molecule generation models, including RELATION, TargetDiff, Pocket2Mol, TRIOMPHE-BOA and ResGen.



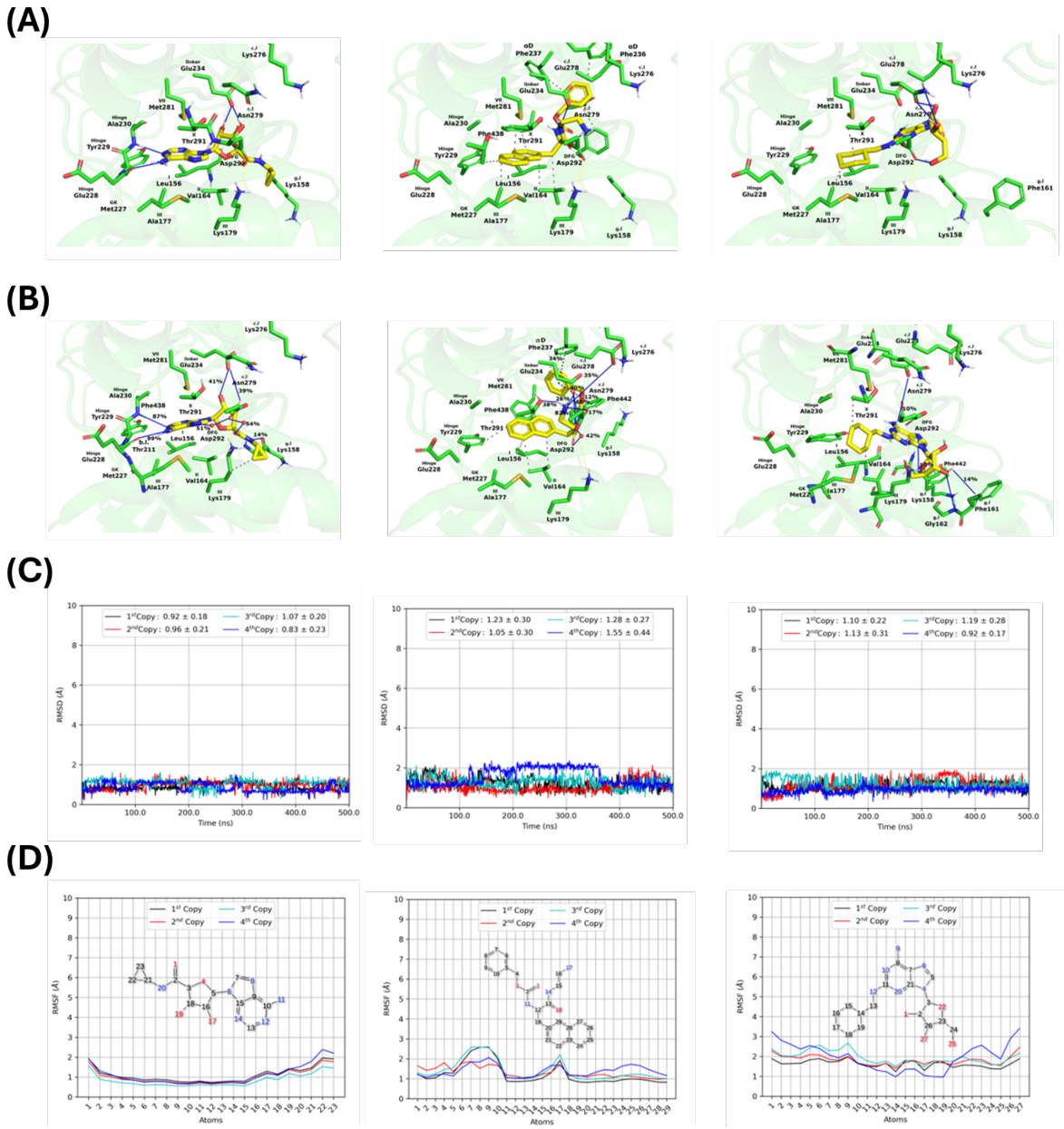
Supplementary Fig. 5: Physicochemical property value distributions of de novo molecules generated by target-centric and non-targeted models, randomly selected ChEMBL molecules and real AKT1 inhibitors for **(A)** molecular weight, **(B)** number of heavy atoms, **(C)** AlogP, **(D)** polar surface area (PSA), **(E)** hydrogen bond acceptor (HBA), **(F)** hydrogen bond donor (HBD), **(G)** rotatable bonds, and **(H)** aromatic rings. In each panel, DrugGEN-NoTarget and random ChEMBL molecules are given in the plot on the left side, whereas the targeted models' distributions are on the right side (i.e., DrugGEN, RELATION, TargetDiff, Pocket2Mol, TRIOMPHE-BOA, ResGen and real AKT1 inhibitors). The numbers in parentheses in figure legends indicate mean values.



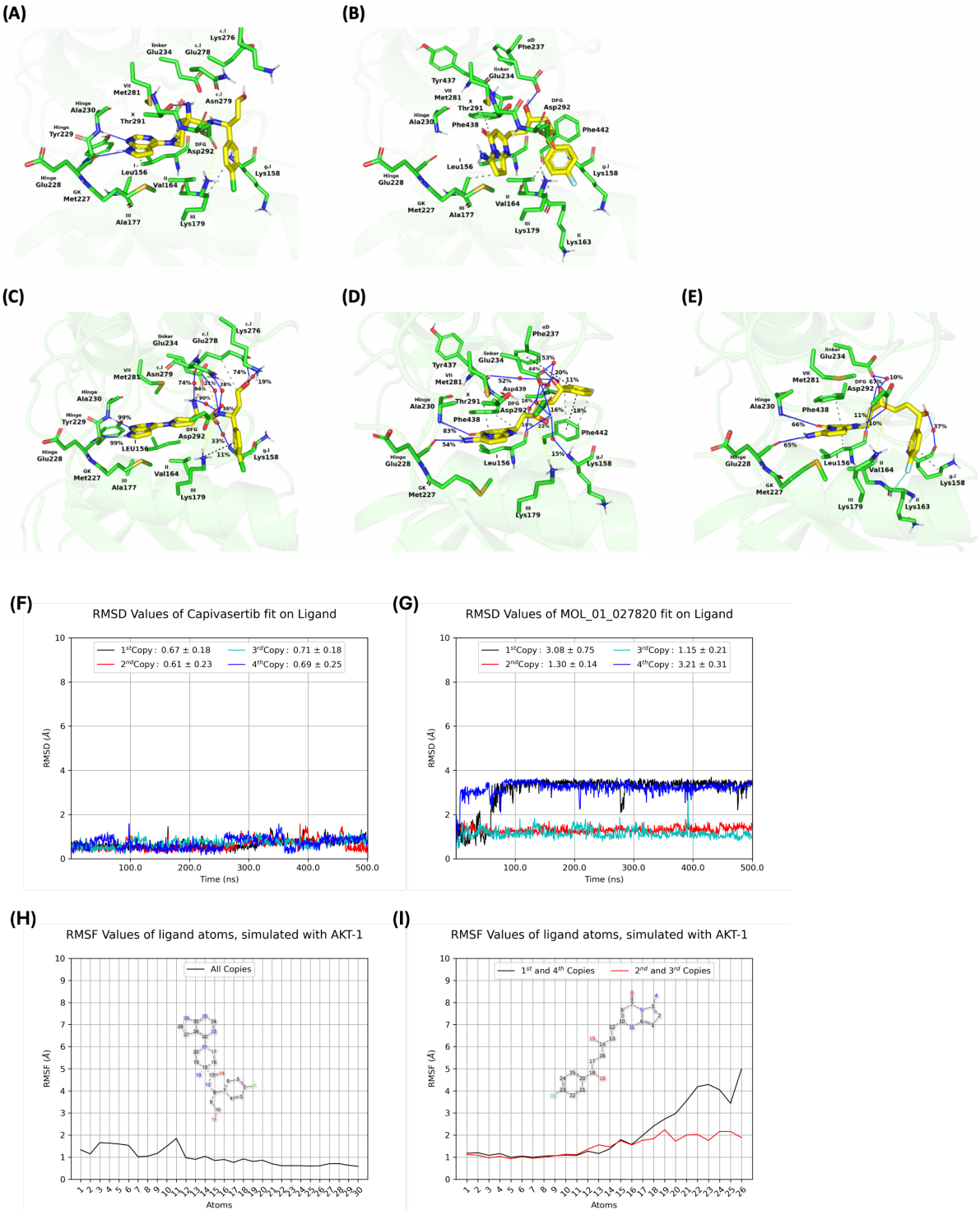
Supplementary Fig. 6: DEEPScreen drug-target interaction prediction analysis results (see “Methods” in the main text). The plot shows the confidence level histogram of 3,655 de novo compounds generated by the DrugGEN model with at least 1 rotated molecular image (out of the total of 36 rotated images generated for a single molecule) predicted as active against the AKT1 protein. 1,382 of these molecules are predicted with a sufficiently high confidence score to be accepted as bioactive molecules against AKT1 (i.e., a 50% score cut-off, corresponding to 18 actively predicted rotated images).



Supplementary Fig. 7: Ten randomly selected samples from 30 promising de novo molecules (displayed in Extended Data Fig. 1 of the main text) are highlighted in green. Each DrugGEN molecule is shown with the most structurally similar two training dataset (ChEMBL) molecules (highlighted with grey). Real (ChEMBL) molecules are selected based on their substructure similarity to the de novo designs using the Tanimoto coefficient on MACCS fingerprints.

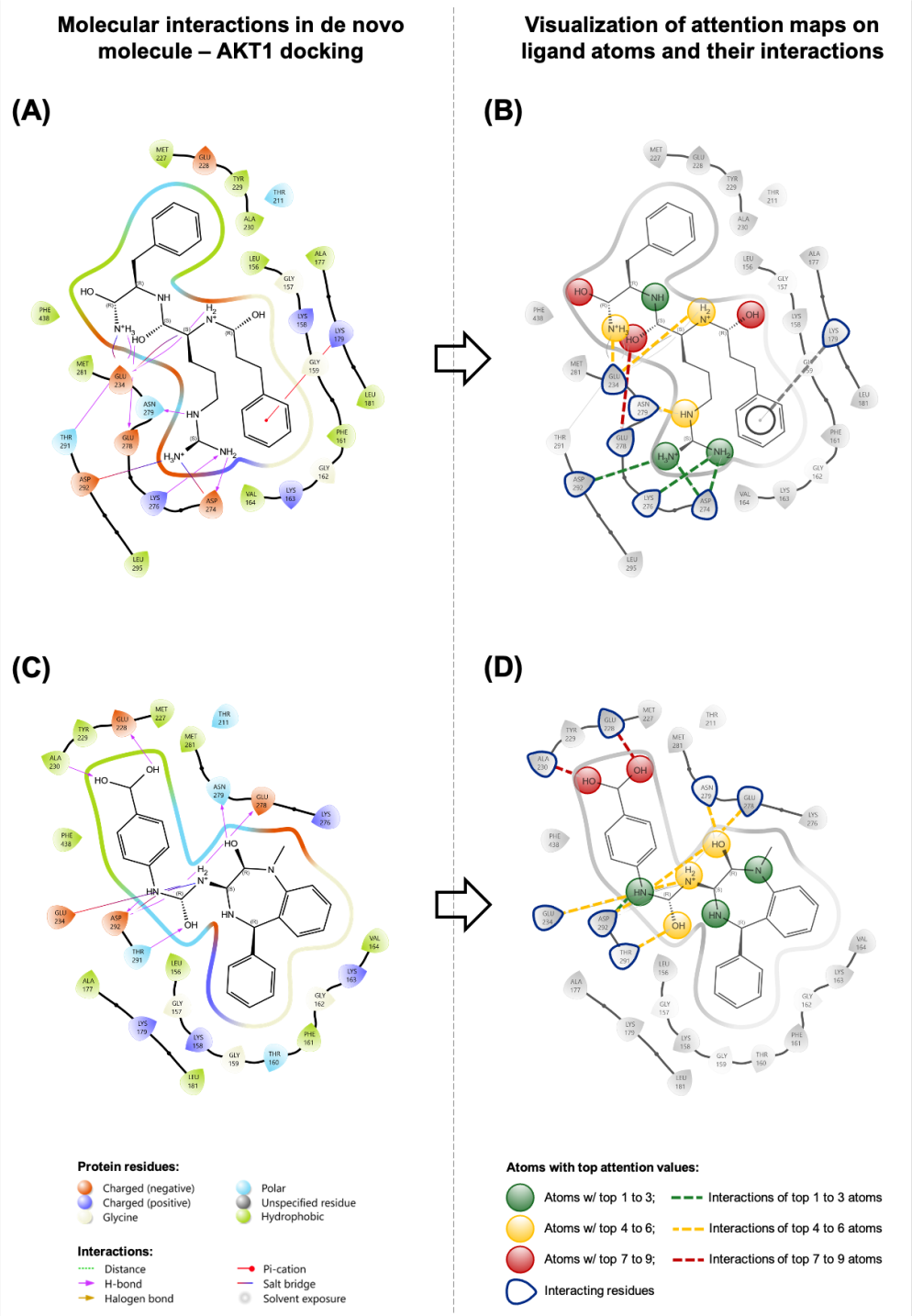


Supplementary Fig. 8: Structural analysis of three de novo generated molecules (Mol. ID 3–5) selected for experimental validation. **(A)** Initial binding orientations of Molecules 3, 4, and 5 at the starting point of molecular dynamics (MD) simulations. **(B)** Key protein-ligand interactions observed during MD simulations, visualized with interacting residues and interaction types. The depicted poses represent the most populated conformations from each simulation. **(C)** Root-mean-square deviation (RMSD) values of Molecules 3, 4, and 5 in complex with AKT1. **(D)** Root-mean-square fluctuation (RMSF) values of ligand atoms in the same complexes. Abbreviations: Exp. Mol. ID: Experimental molecule identifier. I-VII represents β-sheet numbers, g.l represents glycine-rich loop, c.l represents catalytic loop, GK represents gatekeeper residue, and xDFG represents highly conserved kinase residues, linker represents the loop that connects the hinge domain to αC-helix. Gray dashed lines represent Van der Waals interactions. Blue lines represent hydrogen bonds and water bridges. Green lines indicate halogen bonds. Yellow dashed lines represent salt bridges. Directional interactions were noted only when the occupancy value exceeded 10%; however, for visual clarity, occupancy values of the water bridges were stated only when they exceeded 30%.



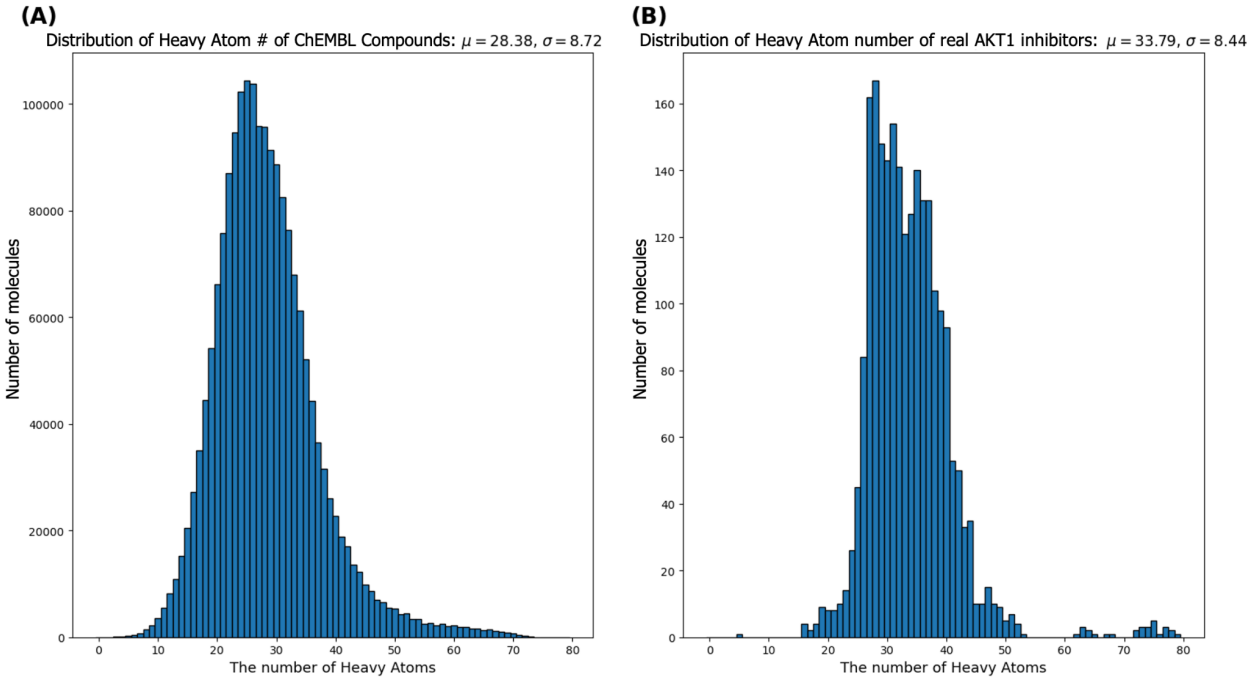
Supplementary Fig. 9: The output of the structural analysis of selected molecules: Capivasertib (the ligand from the crystal structure PDB ID: 4GV1) [26] and the showcase de novo generated molecule: "Pyrrolo[1,2-a]pyrimidin-4(1H)-one" derivative (MOL_01_027820 in Extended Data Fig. 1). Block I: Molecular docking results: **(A)** AKT1 crystal complex structure with the co-crystallised ligand: Capivasertib (binding free energy in docking: -9.517 kcal/mol), and **(B)** The best pose in the molecular docking of the showcase de novo generated molecule MOL_01_027820, to the structurally resolved binding site of AKT1 (binding free energy in docking: -9.686 kcal/mol). Gray dashed lines represent Van der Waals interactions. Blue lines represent hydrogen bonds. Yellow dashed lines represent salt

bridges. Block II: The molecular dynamics simulation output. The occupancy values of: **(C)** the single binding pattern of capivasertib, **(D)** the first binding pattern of MOL_01_027820, and **(E)** the second binding pattern of MOL_01_027820 in complex with AKT1 obtained from the MD simulations. Noninteracting backbone atoms are hidden for visual clarity. Only the interactions above 10% occupancy values are reported. Domain names are noted on the figure as given in the KLIFS database [35]. Abbreviations: I-VII represents β-sheet numbers, g.l represents glycine-rich loop, c.l represents catalytic loop, GK represents gatekeeper residue, and xDFG represents highly conserved kinase residues, linker represents the loop that connects the hinge domain to αC-helix. Gray dashed lines represent Van der Waals interactions. Blue lines represent hydrogen bonds and water bridges. Yellow dashed lines represent salt bridges. The cyan line represents orthogonal multipolar interaction. MD-derived occupancy values are shown on each interaction. Block III: RMSD plots of the compounds in MD simulations: **(F)** capivasertib and **(G)** MOL_01_027820 in each MD simulation. The illustration of rotation by RMSF plots of compounds: **(H)** capivasertib and **(I)** MOL_01_027820 for different binding patterns observed in MD simulations. For visual clarity, similar RMSF values on ligand atoms are calculated together.

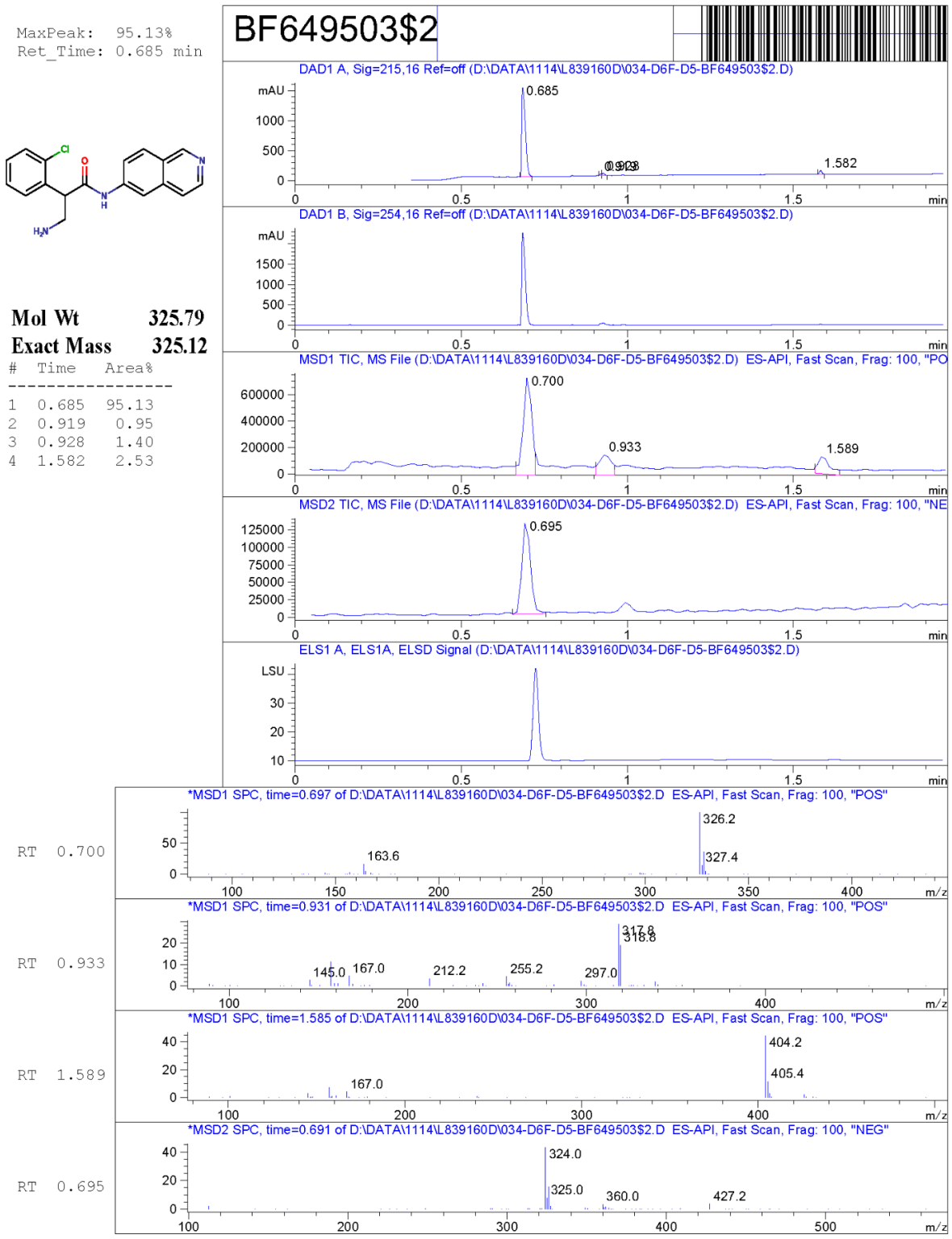


Supplementary Fig. 10: Visualisation of DrugGEN attention maps on two additional de novo molecules. The **left**-side column depicts protein-ligand interaction diagrams obtained from the molecular docking of DrugGEN's three de novo molecules (in A, C and E) with the AKT1 protein structure (PDB ID: 4GV1). The docked ligands are located in the binding pocket, with interactions between residues and the ligand shown as lines, coloured by the interaction type. Protein residues are coloured according to their physicochemical properties. The **right**-side column depicts the attention maps of the same three de novo molecules (in B, D and F) retrieved from the graph transformer module of the DrugGEN generator network. Atoms that receive the highest attention scores are highlighted with colours (green: 1st, 2nd and 3rd atoms; yellow: 4th, 5th and 6th atoms; and red: 7th, 8th and 9th atoms with the

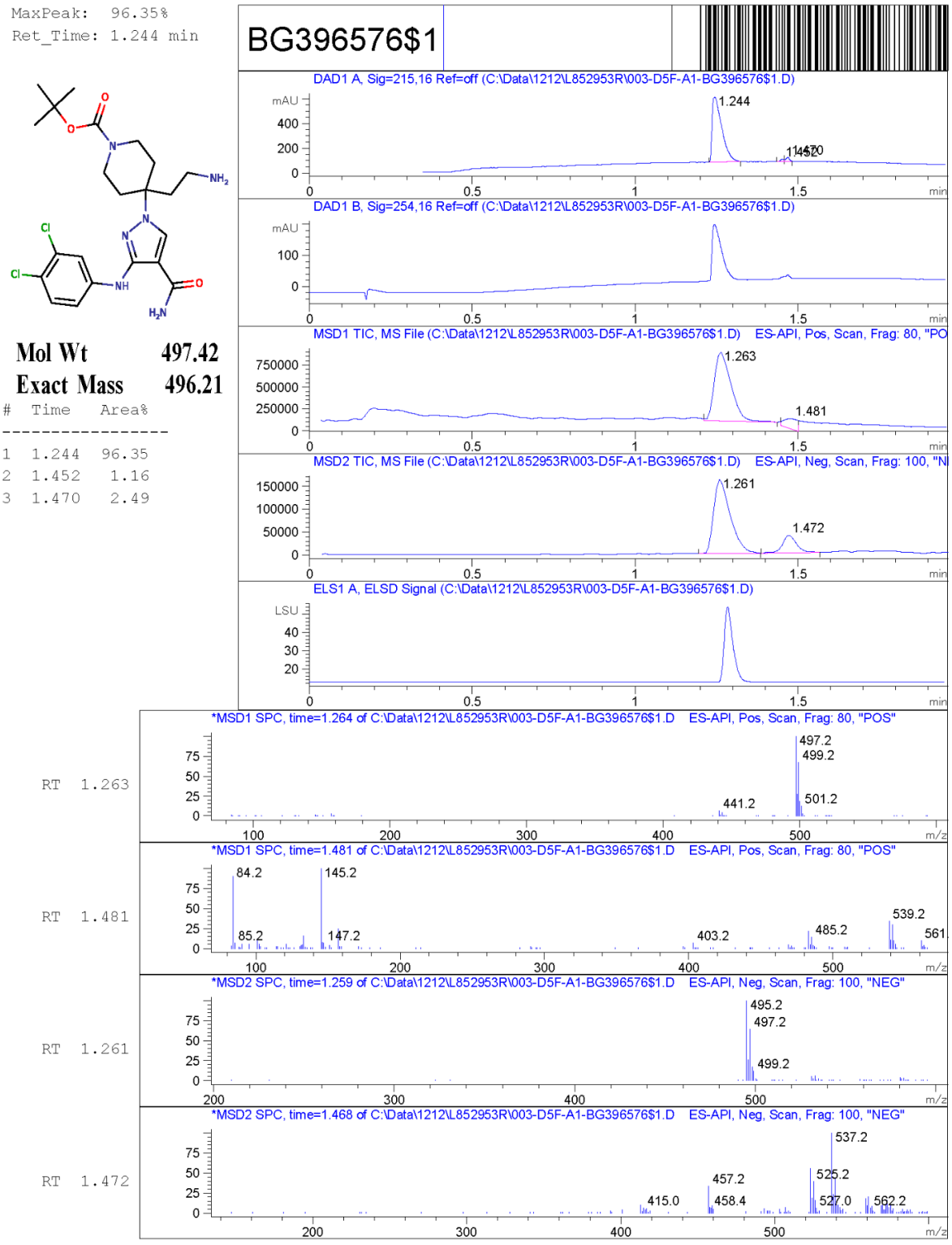
highest attention scores). Atoms with lower attention scores are not coloured. In the right-side plots, receptor-ligand interactions (i.e., those obtained from the docking analysis) are represented by dashed lines in the attention-score-based colour of the molecule atom involved in the respective interaction. If an interacting atom could not be retrieved (i.e., the atom received a low attention score), its interaction is given in grey colour. **(A, B)** molecule id: MOL_02_045598, docking score: -9.055 kcal/mol, **(C, D)** molecule id: MOL_02_000546, docking score: -9.051 kcal/mol.



Supplementary Fig. 11: The heavy atom distribution histogram of **(A)** all of the small molecules in the ChEMBL database, and **(B)** recorded AKT1 binding small molecules in the ChEMBL database (v30).

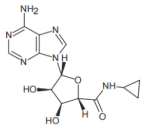


Supplementary Fig. 12: LC-MS analysis result of Molecule 1.

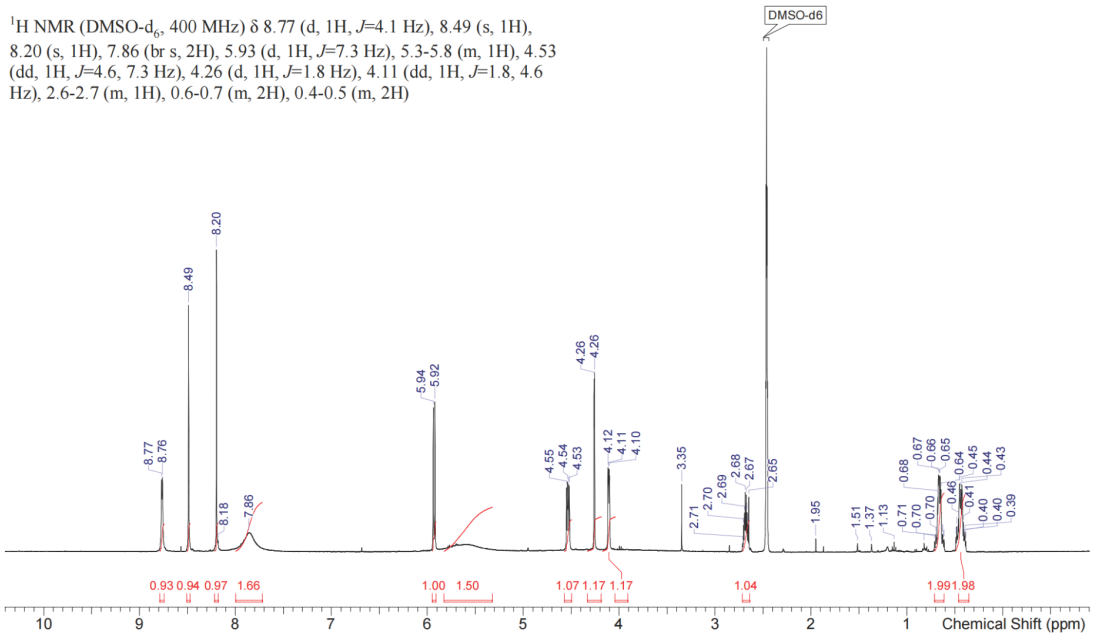


Supplementary Fig. 13: LC-MS analysis result of Molecule 2.

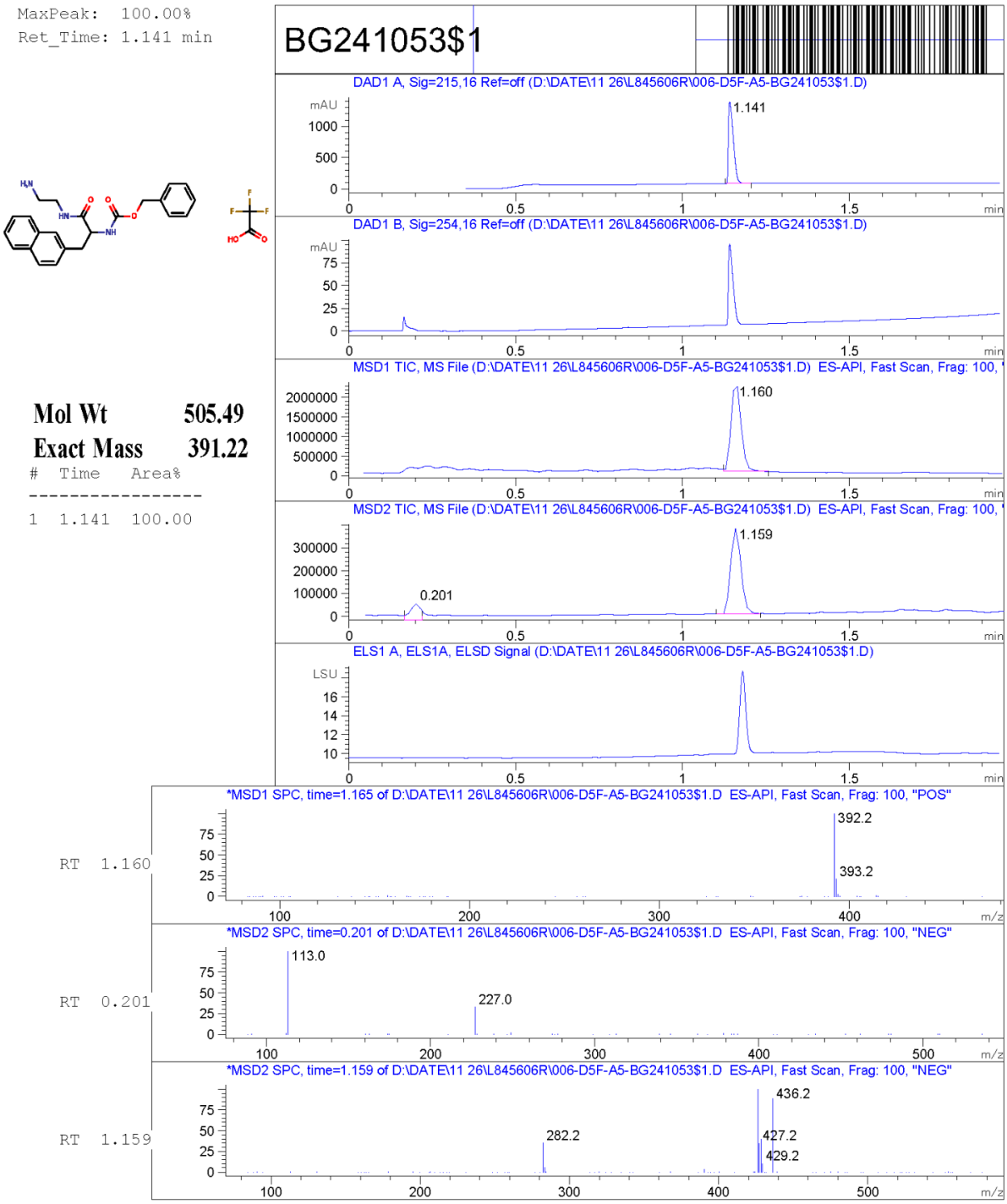
5'-(N-Cyclopropyl)carboxamidoadenosine



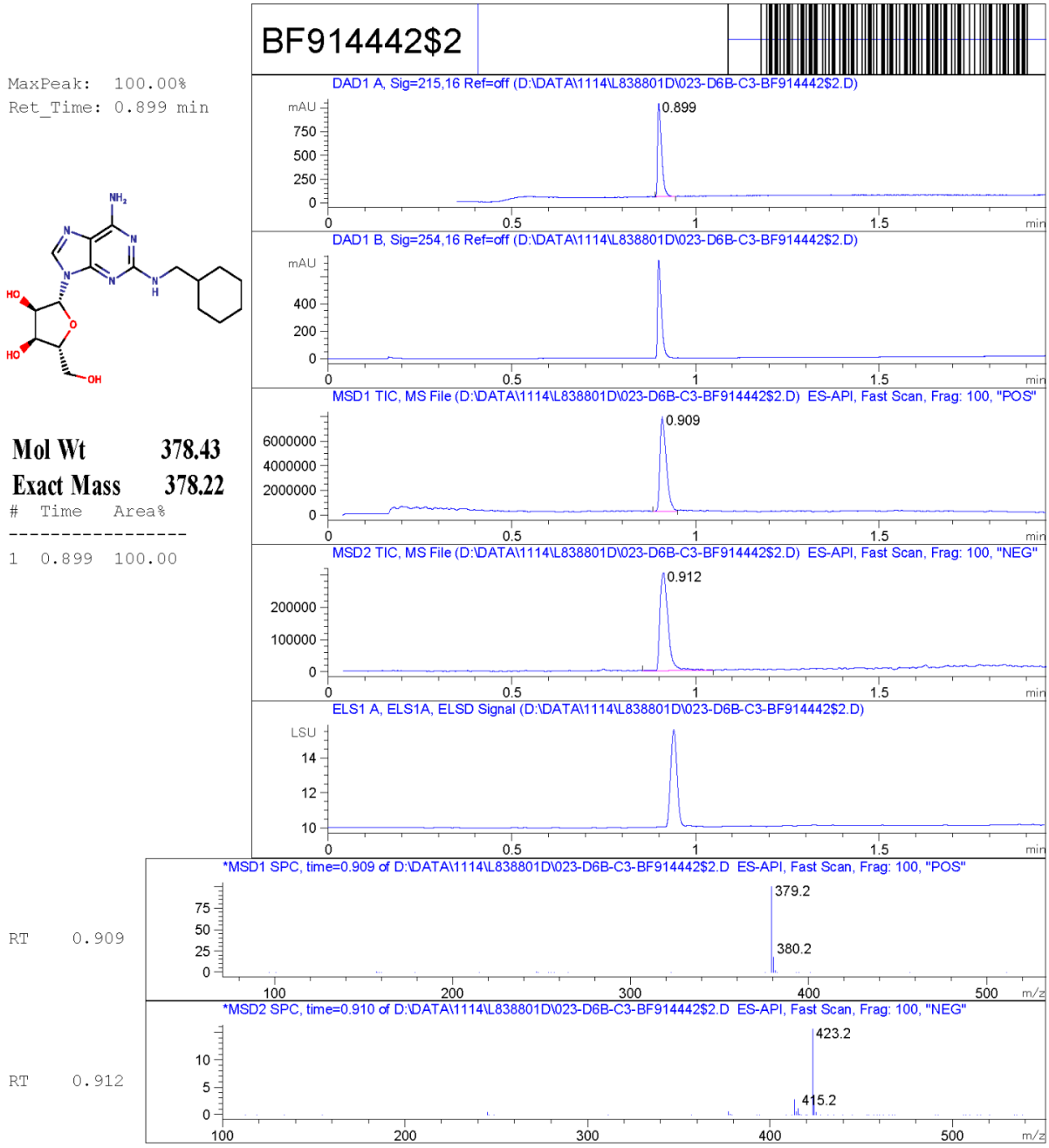
^{1}H NMR (DMSO-d_6, 400 MHz) δ 8.77 (d, 1H, *J*=4.1 Hz), 8.49 (s, 1H), 8.20 (s, 1H), 7.86 (br s, 2H), 5.93 (d, 1H, *J*=7.3 Hz), 5.3-5.8 (m, 1H), 4.53 (dd, 1H, *J*=4.6, 7.3 Hz), 4.26 (d, 1H, *J*=1.8 Hz), 4.11 (dd, 1H, *J*=1.8, 4.6 Hz), 2.6-2.7 (m, 1H), 0.6-0.7 (m, 2H), 0.4-0.5 (m, 2H)



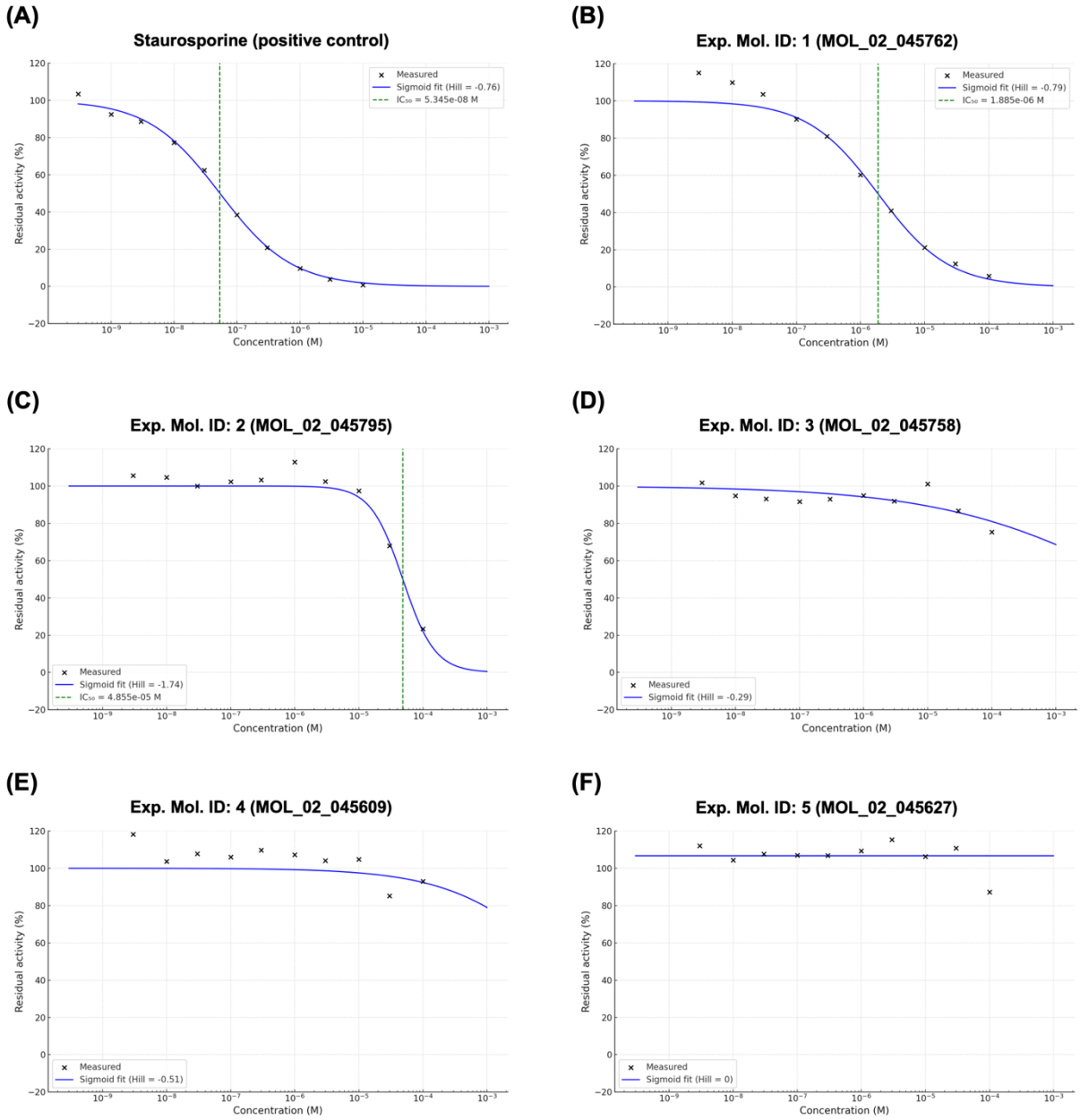
Supplementary Fig. 14: ^{1}H-NMR analysis result of Molecule 3.



Supplementary Fig. 15: LC-MS analysis result of Molecule 4.



Supplementary Fig. 16: LC-MS analysis result of Molecule 5.



Supplementary Fig. 17: IC50 curves of the compounds **(A)** Staurosporine -positive control- (IC50: 0.05 µM), **(B)** Molecule 1 (IC_{50}: 1.89 µM), **(C)** Molecule 2 (IC_{50}: 48.60 µM), **(D)** Molecule 3 (IC_{50}: -), **(E)** Molecule 4 (IC_{50}: -), **(F)** Molecule 5 (IC_{50}: -), achieved via radiometric AKT1 binding assay. Ten different concentrations were tested for each compound, ranging from 100 µM to 3 nM, using a semi-logarithmic dilution in singlicate, except for the positive control, which ranged from 10 µM to 0.3 nM. The IC_{50} values were determined using a Sigmoidal response (variable slope) with parameters ranging from 100% (for the top value) to 0% (for the bottom value) and fitted with least-squares (i.e., the blue curves) using 10 data points, i.e., dose concentrations, for each molecule (n = 10). IC_{50} values are shown with green vertical dashed lines. The plots without an IC_{50} value indicate inactive compounds up to the concentration of 100 µM; therefore, no IC_{50} was measured. The x-axis shows the molar concentration of the compounds, and the y-axis shows the per cent residual activity (i.e., remaining activity) of the compounds against the enzyme.

Supplementary Tables

Supplementary Table 1: Ablation study performance results in terms of (i) fundamental benchmarking metrics (i.e., validity, novelty, uniqueness, and IntDiv), (ii) metrics measuring physicochemical and drug-likeness-related properties (i.e., QED, and SA), and novelty at inference (NI), which is a novelty metric measured against the starting molecules during inference, whereas the regular novelty is measured against the training dataset. NI is given together with the regular novelty (in parentheses). The arrows indicate the direction in which the performance increases. The best performances are displayed in bold font, separately for non-targeted and targeted design model groups, where all scores within 1% of the highest value are included, as the differences among these closely related values are negligible.

Model objective / data properties	Model/dataset name	Highlight of the Model/Dataset	Validity (↑)	Novelty (NI*) (↑)	Uniq. (↑)	IntDiv (↑)	QED (↑)	SA (↓)
Training Dataset (real molecules)	ChEMBL molecules	Training data (random mol.*)	-	-	-	0.879 ± 0.018	0.537± 0.223	3.008 ± 0.965
	Real AKT1 inhibitors	Training data (targeted mol.)	-	-	-	0.862 ± 0.022	0.552 ± 0.189	3.225 ± 0.833
Targeted molecule design (AKT1 targeting)	DrugGEN	Targeted gen. via graph transformer with modified attention (att * edge * (edge + 1))	0.713	**0.993 (0.924)**	**0.995**	**0.878 ± 0.019**	0.512 ± 0.175	2.968 ± 0.862
	DrugGEN-att-FiLM	Attention changed to FiLM-inspired modulation (att * edge_mul + edge_add)	0.110	0.988 (0.503)	**0.996**	**0.881 ± 0.020**	**0.576 ± 0.158**	**2.847 ± 0.852**
	DrugGEN-att-edge	Attention changed to att * edge (official graph transformer; Dwivedi and Bresson, 2021)	0.464	**0.991 (0.527)**	**0.994**	**0.879 ± 0.018**	0.537 ± 0.180	3.025 ± 0.897
	DrugGEN-att-edge+1	Attention changed to att * (edge + 1)	0.117	0.961 (0.677)	0.988	**0.885 ± 0.019**	**0.580 ± 0.174**	**2.818 ± 0.929**
	DrugGEN-att-noedge	Attention without any edge-based modification	0.081	0.984 (0.959)	**0.995**	0.850 ± 0.032	0.496 ± 0.143	**2.841 ± 0.932**
	MLP* (baseline)	Targeted gen. via MLPs	0.512	**0.992 (0.952)**	**1.000**	0.857 ± 0.023	0.545 ± 0.187	3.117 ± 0.923
	GCN* (baseline)	Targeted gen. via GCNs	**0.754**	**0.999 (1.000)**	**0.997**	0.849 ± 0.022	0.271 ± 0.173	5.215 ± 0.912
Non-targeted molecule design	DrugGEN-NoTarget	Non-targeted gen. via graph transformer with modified attention (att * edge * (edge + 1))	0.811	**0.990 (0.552)**	**1.000**	**0.874 ± 0.020**	0.502 ± 0.214	3.584 ± 0.875
	MLP-NoTarget (baseline)	Non-targeted gen. via MLPs	**0.839**	**0.999 (0.201)**	**0.998**	**0.879 ± 0.018**	**0.576 ± 0.199**	**3.215 ± 0.951**
	GCN-NoTarget (baseline)	Non-targeted gen. via GCNs	0.639	**1.000 (1.000)**	**0.991**	0.800 ± 0.025	0.135 ± 0.043	6.429 ± 0.429

*mol: molecules, NI: Novelty at Inference (novelty measured against the real molecules used as input at inference), MLP: multi-layered perceptron, GCN: graph convolutional network. The best performances are displayed in bold font.

Supplementary Table 2: The number of transformer encoder blocks (depth) optimisation results for the DrugGEN model. Scores are provided in terms of per cent-based performance loss (-) or gain (+) over the default model.

	Depth=1 (default)	Depth=2	Depth=4	Depth=6
Validity	-	-35%	-34%	-89%
Uniqueness	-	0%	-4%	-8%
Novelty	-	-2%	-3%	+9%

Supplementary Table 3: IC_{50} values of the de novo generated compounds experimented with AKT1. Mol. ID represents the unique identifier assigned during molecule generation, while Exp. Mol. ID refers to the experimental molecule ID given to the synthesised molecules. Hill slopes higher than -0.4 indicate that the curve is not sigmoidal, or very flat, or not descending.

Exp. Mol. ID	Mol. ID	IC_{50} (μM)	Hill slope
1	MOL_02_045762	1.89	-0.79
2	MOL_02_045795	48.55	-1.74
3	MOL_02_045627	>100	0.00
4	MOL_02_045609	>100	-0.51
5	MOL_02_045758	>100	-0.29