Ligands and receptors with broad binding capabilities have common structural characteristics

Journal of Medicinal Chemistry

Perspective

Cite This: J. Med. Chem. XXXX, XXX, XXX−XXX

pubs.acs.org/jmc

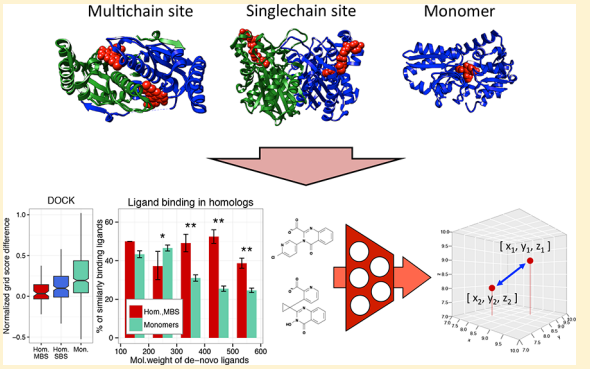
Ligands and Receptors with Broad Binding Capabilities Have Common Structural Characteristics: An Antibiotic Design Perspective

György Abrusán* and Joseph A. Marsh

MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, U.K.

Supporting Information

ABSTRACT: The spread of antibiotic resistance is one of the most serious global public-health problems. Here we show that a particular class of homomers with binding sites spanning multiple protein chains is particularly suitable for targeting by broad-spectrum antibacterial agents because due to the slow evolutionary change of such binding pockets, ligands of such homomers are much more likely to bind their homologs than ligands of monomers, or homomers with a single-chain binding site. Additionally, using de novo ligand design and deep learning, we show that the chemical compounds that can bind several different receptors have common structural characteristics and that halogens and fragments similar to the building blocks existing antimicrobials are overrepresented in them. Finally, we show that binding multiple receptors selects for flexible compounds, which are less likely to accumulate in Gram-negative bacteria; thus there is trade-off between reducing the emergence of resistance by multitargeting and broad-spectrum antibacterial activity.



INTRODUCTION

Currently the world faces an emerging antibiotic resistance crisis because the rate of developing new antibiotics has not caught up with the pace of the spread of antibiotic resistance.[1,2] This is caused by several factors: antibiotic resistance is an ancient evolutionary phenomenon and is unavoidable,[1,2] while the small number of novel antibiotics entering the market might be partly caused by the limitations of existing compound libraries used in the pharmaceutical industry and the possible lack of unexplored, "low-hanging-fruit" drug classes[3,4] (but see also ref 5). Socioeconomic factors also contribute, including the irresponsible use of antibiotics promoting resistance and the relatively low profitability of novel antimicrobials, which, exactly to prevent the emergence of resistance, are likely to be used as last-resort drugs rather than first-line medications. Since antibiotic resistance can emerge quickly, even in laboratory settings,[6] developing drugs that reduce the likelihood of resistance is a central goal of the field.[7−9] Resistance can emerge due to several factors, like changes in the proteins targeted by the antibiotic, changes in the rate of removal or uptake of the antibiotic, or changes in the degradation rate of the antibiotic. However, the analysis of currently available antibiotics indicates that most successful antibiotics or antibiotic classes bind several protein targets, e.g., β-lactam antibiotics, fluoroquinolones (or target substrates rather than enzymes, e.g., vancomycin[10]), while resistance emerges much more quickly for antibiotics that target only a single protein (e.g., sulfonamides, trimetophrim), and such drugs are used mostly in combination with other drugs.[7,11] The most likely cause of this phenomenon is that in the case of single target drugs, a few mutations at a single binding site can be sufficient to make the drug ineffective, whereas for multitarget drugs, several binding sites have to be mutated to achieve resistance.

As a consequence, the strategies that have been employed to slow down the emergence of resistance typically rely on targeting several proteins simultaneously, by either a single drug or "cocktails" of drugs. The central goal is obviously to find novel drug classes, but an alternative and very promising strategy is to create hybrid molecules that contain the core pharmacophores of several existing drugs, connected by a linker.[7,12−14] For several difficult to treat infections like *Helicobacter pylori* or *Mycobacterium tuberculosis* (but also for pathogens like HIV or *Plasmodium*), combination therapy is already the only effective treatment, and multitarget drugs are currently being developed.[15] Additionally, it has been shown that the emergence of resistance for one antibiotic frequently influences (and sometimes increases) the sensitivity to other antibiotics,[16−20] suggesting that, aside from combination therapy, developing drugs that target the "right" protein combinations may significantly reduce the speed at which resistance develops.

Received: February 1, 2019
Published: June 12, 2019

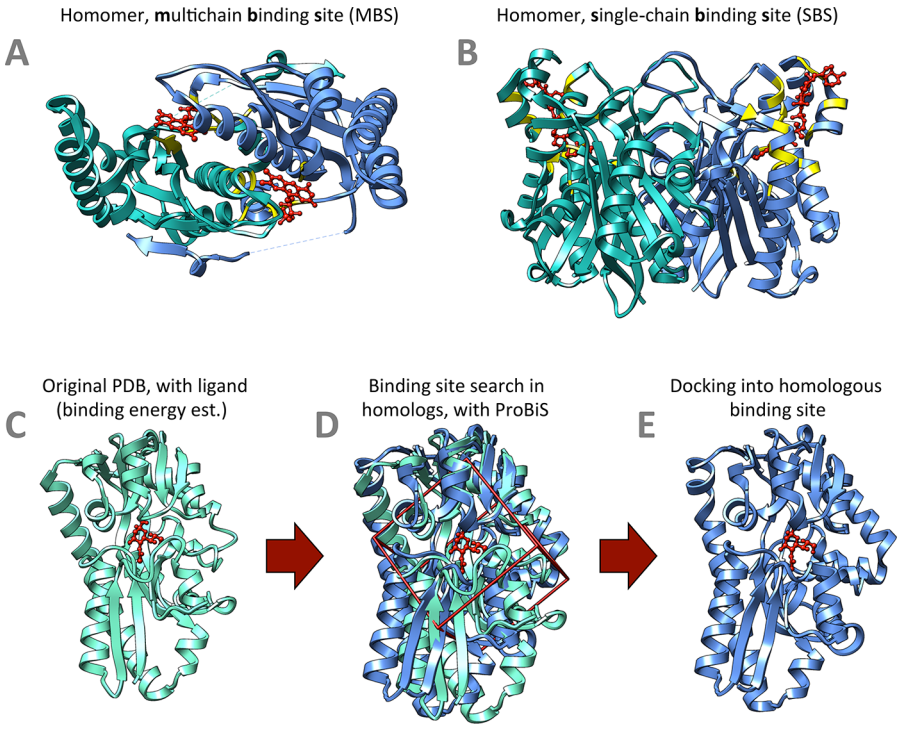


Figure 1. Examples of homomers with multichain binding sites (MBS), single-chain binding sites (SBS), and an overview of binding site identification in homologs. (A) Structure of nitroreductase ydjA from *E. coli* (PDB code 3bm1). The dimer structure has two multichain binding sites, both "sandwiched" between the two chains of the complex. The ligand (flavin-mononucleotide) is displayed in red, and ligand binding residues are in yellow. (B) Structure FabH protein from *E. coli* (PDB code 1ebl). The dimer has two binding sites, both restricted to a single chain. The ligand (coenzyme A) is displayed in red, and ligand binding residues are in yellow. (C) Structure of SiaP protein from *Haemophilus influenzae* (PDB code 2wyk). The protein is a monomer, and has a single binding site, with its ligand (*N*-glyconeuraminic acid) displayed in red. (D) We searched for similar binding sites in homologous proteins with ProBiS and defined the region of binding sites for grid building and docking through local structural alignments. The structural alignment shows the *H. influenzae* SiaP protein binding site superposed with the binding site (identified by ProBis) of the homologous c4-dicarboxylate-binding protein of *Pseudomonas aeruginosa* (PDB code 4nf0); the red box indicates the region of the binding sites. Note that the alignment optimized the superposition of the binding sites and not the global protein structures. (E) Once the binding site has been identified in c4-dicarboxylate-binding protein, the ligand of *H. influenzae* SiaP protein was docked into it, and the binding energies (i.e., grid score) in the two structures were compared.

A theoretical problem in designing multitarget antibiotics is that the proteins targeted by them may not be similarly druggable in different species. First, species that are distantly related (i.e., Gram positive and Gram negative bacteria) may not share all the proteins targeted by the drug, even if they are essential. Second, and more importantly, the binding sites of homologous proteins in distantly related species may not be structurally identical, and thus the drug might not be able to bind efficiently all of them, even if they are present. These problems are expected to scale exponentially with the number of targeted proteins, and as a consequence, drugs that are multitarget in one species might effectively be single target in another one and therefore not be "resistance resistant". Therefore, selecting targets that have similar binding sites and functions across the broadest possible spectrum of bacterial species is of great importance to the selection of antibiotic targets.

Recently, we have found that the structure of ligand binding sites has profound consequences for the evolution of protein function and structural divergence of binding pockets, particularly in proteins that form homomers—protein complexes made up of multiple copies of the same polypeptide subunit.[21] In prokaryotes, homomers are by far the most common type of protein complexes, with >50% of bacterial proteins of known structure falling into this category.[21,22] Binding sites that are formed by multiple protein chains show much higher structural conservation than sites that are formed by the residues of individual protein chains, and have more similar ligands than binding sites of monomers, or homomers with a single chain binding site.[21] Since similar binding pockets generally bind similar ligands,[23] proteins where distant homologs have similar binding pockets are potentially good candidates for broad-spectrum antibiotic targets. In this paper, using in silico experiments, we show that ligands and de novo designed ligands of homomers with multichain binding sites (MBS, see Figure 1A) are more likely to bind their homologs than ligands of monomers or homomers with a single-chain binding site (SBS, see Figure 1B). This shows that considering the quaternary structure of proteins and the structures of their ligand-binding sites can aid the selection of protein targets for new broad-spectrum antibiotics. In addition, using de novo ligand design and methods based on deep learning, we also test whether the chemical compounds that can bind several different receptors have common structural characteristics. We show that fragments similar to the building blocks of existing antibiotics are overrepresented in them and that there is likely to be a trade-off between preventing the emergence of drug resistance and broad spectrum activity. We expect these findings to be useful in selecting targets of novel leads and in generating targeted fragment libraries for antibiotic design.

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

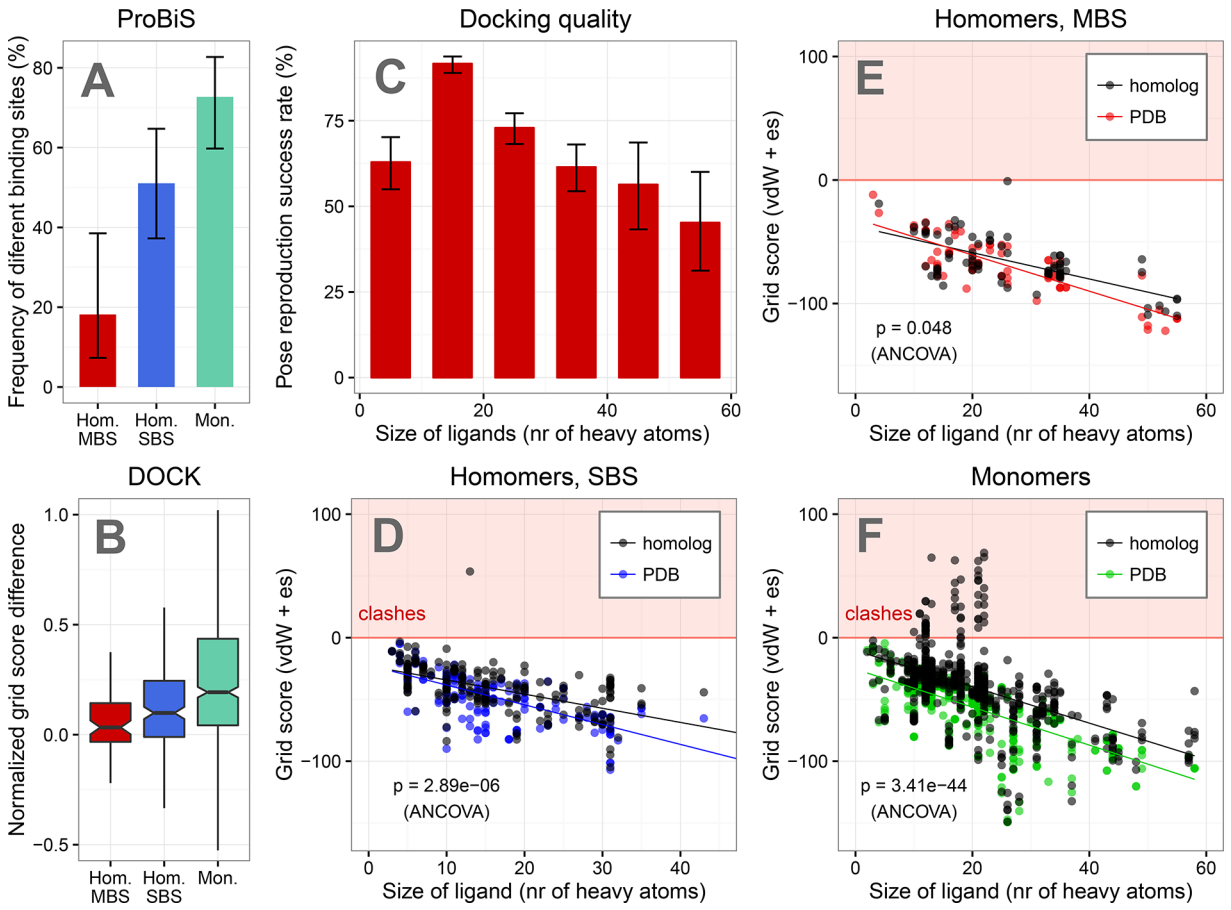


Figure 2. Ligands of homomers are more likely to bind their homologs than the ligands of monomers. (A) Frequency of homologous protein pairs without similar ligand binding sites, (ProBiS *Z* score of >2). See also ref 21 for general trends in ligand binding site evolution. MBS homomers are much more likely to have an identical binding site in their homologs than SBS homomers ($p = 0.029$, test of proportions) or monomers ($p = 0.00011$, test of proportions). (B) The difference between the estimated ligand binding energies (grid score) in the original binding sites and homologous binding sites shows that homomers, and particularly MBS homomers, are much more likely to bind the ligands of their homologs than monomers. Normalized grid score difference was calculated as $(\text{score}_{\text{original}} - \text{score}_{\text{homolog}})/\text{score}_{\text{original}}$. All three possible comparisons are significant ($p = 0.013$ for MBS vs SBS homomer; $p \ll 0.005$ for MBS/SBS homomer vs monomers; Kruskal−Wallis rank sum tests). (C) Pose reproduction success rate of DOCK for ligands of different sizes. The high flexibility of ligands above 40 heavy atoms in this particular data set results in a relatively low (~50%) success rate for large ligands. (D−F) The relationships between ligand size and binding energy (grid score) follow a linear trend in the original ligand binding sites and the binding sites of homologs. The difference between the original binding sites and the binding sites of homologs is lowest in MBS homomers and largest in monomers (ANCOVA).

RESULTS

Ligands of MBS Homomers Bind Their Homologs Significantly Better than Ligands of SBS Homomers or Monomers. In the first step of the analysis we identified bacterial proteins that are likely to be suitable targets for antibiotics. Using BLAST and the prokaryotic proteins present in the Protein Data Bank (PDB), we compiled a data set of 687 pairs of homologous bacterial proteins. We only used proteins that form homomers or monomers. The proteins of the homologous pairs were selected according to the following criteria: (1) none of the proteins have a homolog in the human genome with BLAST *e*-value $<10^{-3}$; (2) the pairs have homologous regions, with BLAST *e*-value $<10^{-5}$; (3) at least one protein of the pair is essential; (4) the sequence similarity between them is less than 40% because above 40% protein structures are very similar (see ref 24 for examples); (5) both proteins have a biologically relevant ligand, i.e., present in the BioLiP database;[25] and (6) their quaternary structure and binding site type (MBS vs SBS) are similar. Next, in each PDB structure of the protein pairs, the ligand-binding site of every small molecule ligand was identified (Figure 1C), and we identified the best binding site candidate in the structures of their homologs with ProBiS[26,27] (Figure 1D). Once the location of the putative binding site was determined in the homolog, we docked the ligand with DOCK[28] to the original binding site and the binding site in the homolog and compared their estimated binding energies (i.e., grid score, Figure 1E). Only those pockets (and proteins) were used where the performance of DOCK was acceptable; i.e., it could reproduce the position of the original ligand in the top 10 scoring poses (see Methods for details). In addition, the binding sites of metals and cofactors were excluded from the analysis because metals typically have very small binding sites, while the binding pockets of cofactors are structurally so conserved that they are likely to be similar in bacterial and mammalian proteomes.[21] This procedure resulted in a set of 1007 binding pocket pairs (78 in MBS homomers, 212 in SBS homomers, and 717 in monomers) between 1464 different binding pockets of 262 different proteins (see Table S1 for a list of pocket pairs). Note that the pocket pairs contain redundancies; the number of nonredundant pockets (excluding

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

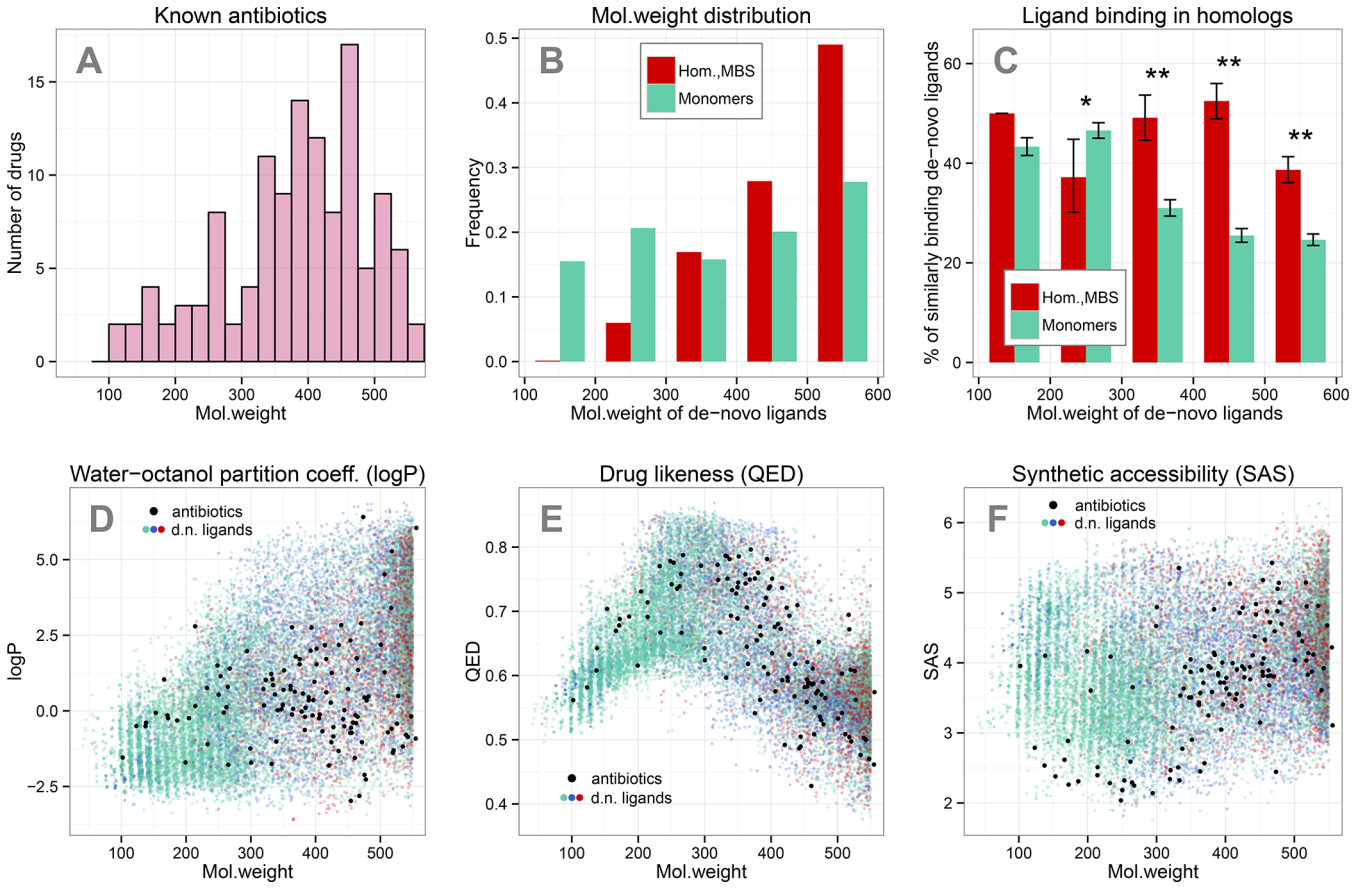


Figure 3. General properties of antibiotics and de novo ligands. For each receptor, the 50 best scoring ligands were included. (A) Molecular weight distribution of Drugbank antibiotics, below 600 Da. (B) Molecular weight distributions of de novo ligands in MBS homomers and monomers. The de novo ligands of MBS homomers are significantly larger than of monomers, while the ligands of SBS homomers (Figure S1) do not show the same trend. (C) The frequency of ligands that bind similarly in homologs and their original binding site is significantly higher among MBS homomers than monomers above 300 Da (**, $p \ll 0.005$, tests of proportions). (D) The log *P* values of antibiotics and de novo ligands. For most antibiotics log *P* is below 3, but for a significant number of de novo ligands it is above 3, particularly above 300 Da. (E) The druglikeness of antibiotics and de novo ligands follows the same trend and declines above molecular weight of 300. (F) Below 300 Da the synthetic accessibility of de novo ligands is considerably worse than of real antibiotics.

pockets with metals and cofactors) is 524 in the original PDB receptors and 940 in their homologs.

The binding pocket searches with ProBiS indicate that this data set, despite its much smaller size, shows a qualitatively similar pattern of binding pocket similarity as the entire PDB in our previous study:[21] the frequency of MBS homomer pairs that have no highly similar (ProBiS *Z* score >2) binding pockets is much smaller than in SBS homomers or monomers (Figure 2A, $p = 0.029$ and $p = 0.0001$, respectively; tests of proportions). Docking of the ligands into their original binding site and the best binding site of their homologs shows a similar trend (Figure 2B). The normalized grid score $[(\text{score}_{\text{original}} - \text{score}_{\text{homolog}})/\text{score}_{\text{original}}]$ indicates that for MBS homomers, the binding of the ligand in their homolog is nearly as strong as in the original binding site, while in SBS homomers and especially monomers, ligands bind much more weakly at the homologous biding sites (Figure 2B, $p = 0.013$ between MBs and SBS homomers, and $p \ll 0.001$ for other comparisons, Dunn's test [Kruskal−Wallis rank sum test]). The performance of DOCK was highly dependent on the size of the ligand (and number of rotatable bonds), as reported previously[28,29] (Figure 2C), and performed best on ligands with 10−30 heavy atoms. The estimated binding energy of ligands (i.e., DOCK grid score, which estimates interaction strength based on a simplified force field) generally follows a linear trend in all three quaternary structure types, with the ligands of monomers having many more clashes in their homologs than homomers (Figure 2D−F). Statistical analyses with ANCOVA indicate a similar trend as the normalized grid scores: there is a marginally significant difference between MBS homomers and their homologs (Figure 2E), and the difference is largest between monomers and their homologs (Figure 2F). Taken together, these findings indicate that ligands of MBS homomers bind much more strongly in their homologs than ligands of SBS homomers or monomers.

Characteristics of Antibiotics and de Novo Ligands. We next determined whether de novo designed ligands show similar trends as the previous analysis of the original ligands of the PDB structures. We used de novo DOCK[30] to design novel ligands in each binding site where the pose reproduction of the original ligand was successful, excluding cofactor and metal binding sites and also sites where the original ligand had only one or no rotatable bonds (see Methods for details), with a maximum molecular weight cutoff of 550 Da. This resulted in 453 binding sites where de novo ligands were built (see Table S3 for the full list). From all de novo ligands that were generated in each receptor we selected the 50 with the lowest grid scores (i.e., the 50 best binders) and used only these in the further analyses, altogether 22 650. Note that de novo ligands were not designed

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

in receptors identified through homology. We also downloaded the SMILES strings of FDA approved antibiotics present in DrugBank[31] from the ZINC15 database[32] to compare the properties of de novo ligands with real antimicrobial agents. Antibiotics with molecular weight of >600 Da were not included for two reasons. First, only a relatively small number of compounds fall into this size category (glycopeptides [e.g., vancomycin], lipopeptides [e.g., daptomycin], or macrolides), and these are typically active only against Gram positive bacteria.[33] Additionally, these compounds generally have complex, cyclic structures, and designing functional ligands of this size and complexity is beyond the capabilities of the current de novo design tools. The list of antibiotics included is available in Table S2.

The comparison of molecular weight distributions of antibiotics and de novo ligands shows that the de novo ligands of MBS homomers are somewhat biased toward larger molecular weights than antibiotics, which have a peak between 400 and 500 Da (Figure 3A,B), while monomers (Figure 3B) and SBS homomers (Figure S1A) are characterized by more uniform distributions. Next, the 50 best scoring de novo ligands of each receptor were re-docked into their original binding site and to the corresponding binding site in their homologous proteins, and we determined the fraction of ligands that have an approximately similar molecular weight to the original ligand (min. 85%), and bind in the homologous binding sites similarly as in the original site (min 85% of grid score). The comparison of de novo ligands in the different quaternary structures indicates that in homomers, and particularly MBS homomers, the fraction of ligands that bind similarly in homologs and in the original binding site is much higher than in monomers, particularly above 300 Da (Figure 3C, Figure S1B; **, $p \ll 0.005$; tests of proportions). The difference is primarily caused by the higher similarity of binding sites and not by the qualitatively different performance of de novo DOCK in proteins with different binding sites (although it may somewhat contribute to the pattern in SBS homomers) because the fraction of strongly binding ligands (grid score of de novo ligand is min 85% of the original ligand) is comparable in all quaternary structure types (Figure S2). Surprisingly, SBS homomers perform somewhat better than MBS homomers or monomers (Figure S2), although the difference is less than 20% for most molecular weight bins.

Next we examined whether there are consistent qualitative differences between antibiotics and de novo ligands for the three different quaternary structure types. We used a chemical variational autoencoder (VAE), a recently developed method based on deep learning,[34] to estimate the synthetic accessibility[35] (SAS), quantitative estimate of druglikeness[36] (QED), and octanol−water partition coefficient (log P) values for antibiotics and each de novo ligand using their SMILES (Figure 3D−F). The solubility in organic solvents (log P values) of de novo ligands gradually increase with molecular weight and, for ligands with molecular weight above 300 Da, can substantially exceed the values observed in antibiotics, which are generally below 3 (Figure 3D). The quantitative estimate of druglikeness (QED, higher is better) is a more modern estimate of druglikeness than Lipinski's rule of five and is a combination of eight different molecular descriptors.[36] Antibiotics and de novo ligands show a qualitatively similar, nonlinear pattern, with the most druglike compounds having molecular weight of 200−400 Da (Figure 3E). However, even the compounds with molecular weight above 400 Da have reasonably good (>0.5) QED estimates, and since antibiotics and natural compounds are known to not follow the same trend in druglikeness as synthetic drugs,[33] the druglikeness of de novo ligands should be generally seen as satisfactory. The largest difference between antibiotics and de novo ligands is in their synthetic accessibility (SAS, lower is better), which estimates the ease of synthesis of chemical compounds (Figure 3F). Antibiotics show a clearly increasing trend with molecular weight, with SAS being below 3 for small compounds (easy synthesis) and gradually increasing to ∼6 (more difficult synthesis) above 300 Da. In contrast, de novo ligands show almost no change with molecular weight, with high variability in every weight class (Figure 3F). Moreover, their SAS below 300 Da is much higher than of real antibiotics. This is in agreement with previous reports that the synthetic accessibility of de novo designed ligands can be low. However, it should be noted that most natural products also fall in the range of SAS = 3−6, and thus the values we observe are not prohibitively high.

Finally, using the VAE, we tested whether there are consistent differences in the chemical composition of de novo ligands of MBS homomers, SBS homomers, and monomers. The VAE is fundamentally a pair of neural networks that "encode" (convert to) and "decode" (convert back) discrete chemical compounds into a continuous, high dimensional space called "latent space", where they are represented as a numerical vector. The vector representation offers several powerful operations on chemical compounds, like generating entirely novel molecules, interpolating between molecules, or optimizing existing molecules[34] (see also further analyses below). Additionally, since the latent space is structured, where similar compounds are located close to each other, it can also be used to visualize whether there are any structural clusters in the de novo ligands. We transformed each de novo ligand into a vector in the latent space using the encoder module of the VAE. The latent space has 196 dimensions, and to reduce its dimensionality, we used Barnes−Hut t-SNE[37] for clustering. This transforms the position of chemical compounds in a high dimensional space into 2D, which is suitable for visualization. The 2D maps of de novo ligands show that monomers and MBS homomers have different distributions in the chemical space, and SBS homomers show an intermediate pattern (Figure S3). However, much of the variation is simply due to differences in size (see Figure 3), and aside from these coarse grained differences, no distinct, quaternary structure specific clusters could be identified: in the areas of the plot where all quaternary structure types are present, their distribution is homogeneous (e.g., red circle, Figure S3).

The Best Ranking de Novo Ligands Show Similar Binding Patterns as the Original Ligands of the Receptors. Next we tested whether the de novo ligands show similar patterns in binding the homologous binding sites as the original ligands of the receptors (Figure 2). We found that this is the case for the best ranking ligands, although the pattern is less pronounced than for the original ligands (Figure 4) most likely due to the higher flexibility of de novo ligands. For the 10 best scoring de novo ligands of every receptor, the normalized grid scores are significantly different in all three possible comparisons ($p \ll 0.001$ for all three comparisons, Dunn's test, [Kruskal−Wallis rank sums test], Figure 4A). The frequency of clashes (Figure 4B) is significantly higher in monomers than in MBS and SBS homomers ($p < 0.001$ in both cases, tests of proportions), but there is no difference between MBS and SBS homomers ($p = 0.369$, test of proportions). The grid scores of de novo ligands show a similar linear scaling with their size, in their original and homologous binding sites, with somewhat

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

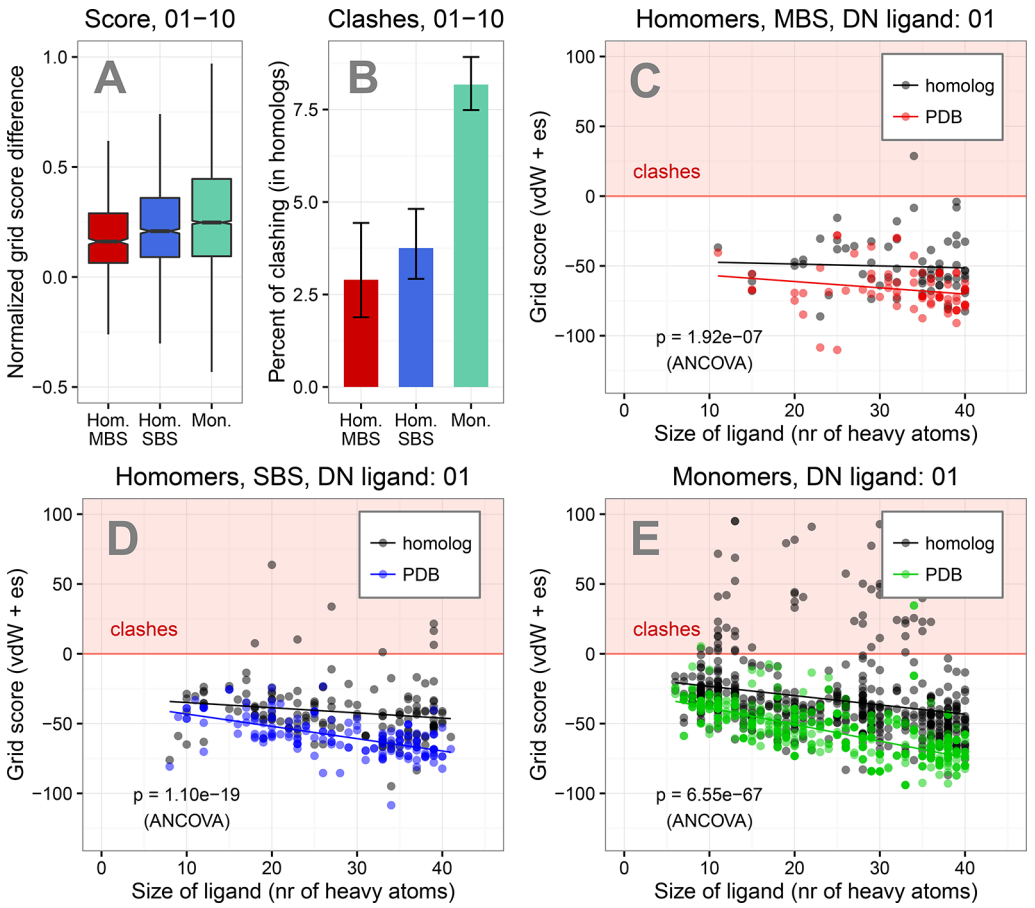


Figure 4. The best scoring de novo ligands show qualitatively similar, although weaker trends than the original ligands of the receptors. (A) Difference between the normalized grid score of the 10 best ligands of each receptor. All three possible comparisons are significant ($p < 0.0001$ Kruskal−Wallis rank sum tests). (B) The frequency of clashes is significantly higher in monomers than in homomers ($p < 0.001$ for both MBS and SBS homomers, tests of proportions), but there is no difference between MBS and SBS homomers homomers ($p = 0.43$). (C−E) Relationships between ligand size and estimated binding energy (grid score) of the best scoring ligand of each receptor in the original ligand binding site and the binding sites of homologs. Similar to the original ligands of the receptor, the relationship between size and grid score is linear, and the difference between the original binding sites and the binding sites of homologs is smallest in MBS homomers and largest in monomers.

weaker binding in homologs (Figure 4C−E). For the lower ranking ligands, we found that the difference between ligands of MBS and SBS homomers is not consistent; however, de novo ligands of monomers have consistently worse (higher) scores in the binding sites of homologs than MBS or SBS homomers (not shown).

Characteristics of de Novo Ligands That Can Target Several Proteins. Next, we examined whether de novo ligands that are likely to bind several proteins have common structural characteristics. We performed the analysis in two steps. First, since de novo DOCK uses a fixed fragment library to build the ligands, we performed a fragment enrichment analysis of the de novo ligands that have structurally similar ligands in multiple receptors. Second, we optimized/evolved these ligands using the VAE to test whether structural motifs that were absent in the fragment library of de novo DOCK emerge during the optimization.

From the de novo ligands (using the best 50 in each receptor), we selected ligands predicted to target several proteins with the following procedure. (1) First we converted each de novo ligand into its vector representation in the latent space of the VAE and calculated all possible (256 million) distances between the 22 650 de novo ligands in the latent space, using their vector representation (see Figure 5A). (2) Next we selected the pairs of de novo ligands where their distance in the latent space is less than 13, and both de novo ligands have lower (better) grid score than the original ligand of the binding site. The distance in the latent space correlates with the structural similarity of compounds; however, it also depends on the size of the ligand: in the case of ligands of 300−400 Da, the distance 13 roughly corresponds to Tanimoto similarity 0.7 (Figure 5B,C), while for larger ligands, larger distances have to be used for the same structural similarity (see Figure S4). We chose the distance cutoff 13, because the druglikeness of de novo ligands in our data set is highest below 400 Da; see Figure 3E. (3) From the ligand pairs, we selected the ligands that have a low (<13) scoring pair in at least two more bacterial species than the query ligand. (4) We discarded ligands with properties that are substantially different from real antibiotics: molecular weight of <300 Da, log $P > 3$, and where SAS > (molecular weight/100) + 1. These steps reduced the total pool of ligands to only 56 (1 duplicate ligand), which are expected to be druglike, have structurally similar ligands in the receptors of minimum three different species, and are likely to be easy to synthesize (see Supplmentary Data for the list of ligands and their structures).

The statistical variability of such a small set of ligands is expected to be much larger than for the whole data set of 22 650 molecules. Therefore, to estimate the uncertainty associated

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

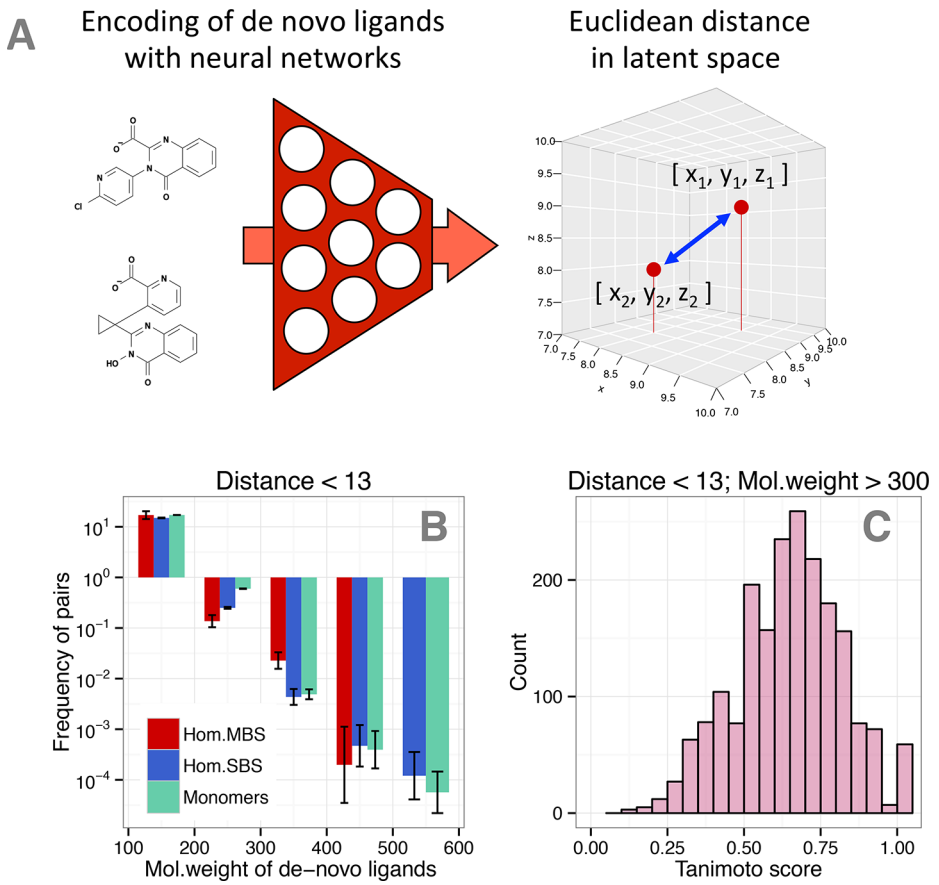


Figure 5. Outline of the structural comparisons using the VAE vector representation. (A) De novo ligands were converted to vector, which represents the parameters of a statistical distribution and defines the location of the compound in the latent space. The Euclidean distance between the vectors was used to measure the structural similarity between all ~256 million possible pairs of compounds. (B) The frequency of pairs with distance below 13 is negatively correlated with their molecular weight. (C) For molecules above 300 Da the distance cutoff 13 results in an approximate Tanimoto similarity of 0.7 (most of them are in the range of 300−400 Da).

with our analysis, we repeated the entire de novo design process and generated a second, fully independent set of de novo ligands using the same methods. In this second set, de novo ligands were built in 452 binding sites (see Table S3). This resulted in an additional, independent set (set 2) of 73 putatively antibiotic-like ligands (with three duplicate ligands; see Supplementary Data).

Several of the selected de novo ligands are reminiscent of compounds and drugs with known antimicrobial activity, particularly of quinolones, quinazolinones, oxadiazoles, and morpholine antifungals (Figure 6A; Figure S5A; note the top left compound of Figure 6A, which resembles a hybrid between a quinolone and an oxazolidinone antibiotic, and the bottom left compound, which is similar to morpholine antifungal). Quinazolinones and oxadiazoles represent two new groups of antibacterials that were shown to be effective against methicillin and vancomycin resistant *Staphylococcus aureus*.[38−44] From the 380 fragments of the fragment library, 11 (set 1) and 12 (set 2) are significantly enriched in the compounds ($p < 0.05$, tests of proportions), and there are overlaps between the two sets of fragments (Figure 6B−D; Figure S5B−D). The most common fragment is the carboxyl group (side chain 23), and in both sets halogen-containing fragments are enriched. Additionally, in set 1 the oxadiazole linker (lnk.36), and the 4-morpholinyl side chain (sid.68) that is the core pharmacophore of several ergosterol synthesis inhibiting antifungals (amorolfine, fenpropimorph, tridemorph),[45−47] is enriched, while in set 2, a quinolone-like linker (quinazolinone, lnk.308, Figure S5C) is enriched. However, although being significant in only one of the sets, quinolone-like (quinazolinone) and morpholine groups are common in both sets (see Supplementary Data; note that only quinazolinone and not quinolone fragments was present in the fragment library).

Next we tested whether the selected ligands originate from binding sites of proteins with similar functions to the proteins that are targeted by the antibiotics they are similar to. Quinolones target the DNA binding gyrase and topoisomerase IV,[48] quinazolinones and oxadiazoles target penicillin binding protein 2a,[38,39,44] and morpholine antifungals inhibit ergosterol synthesis.[46] The de novo ligands of the two sets originate from the binding sites of 23 and 25 proteins (set 1 and set 2, respectively, see Table S4 and S5). Neither their quaternary structure (Figure 7A, Figure S6A) nor their gene ontology (GO) terms show significant enrichment compared to the full set of proteins we used. The frequency of cellular component GO terms indicates that they are present in several different cellular components, including the cytoplasm, cell wall, or periplasmic space (Figure 7B, Figure S6B). The frequency of molecular function GO terms show that most of them have enzymatic functions that are not related to the functions of quinolone, quinazolinone, oxadiazole, or morpholine antimicrobials, although many of them utilize NADP cofactors and ATP in catalysis (Figure 7C,D; Figure S6C, Tables S4 and S5). While we excluded cofactor-binding sites from our analysis, the binding

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

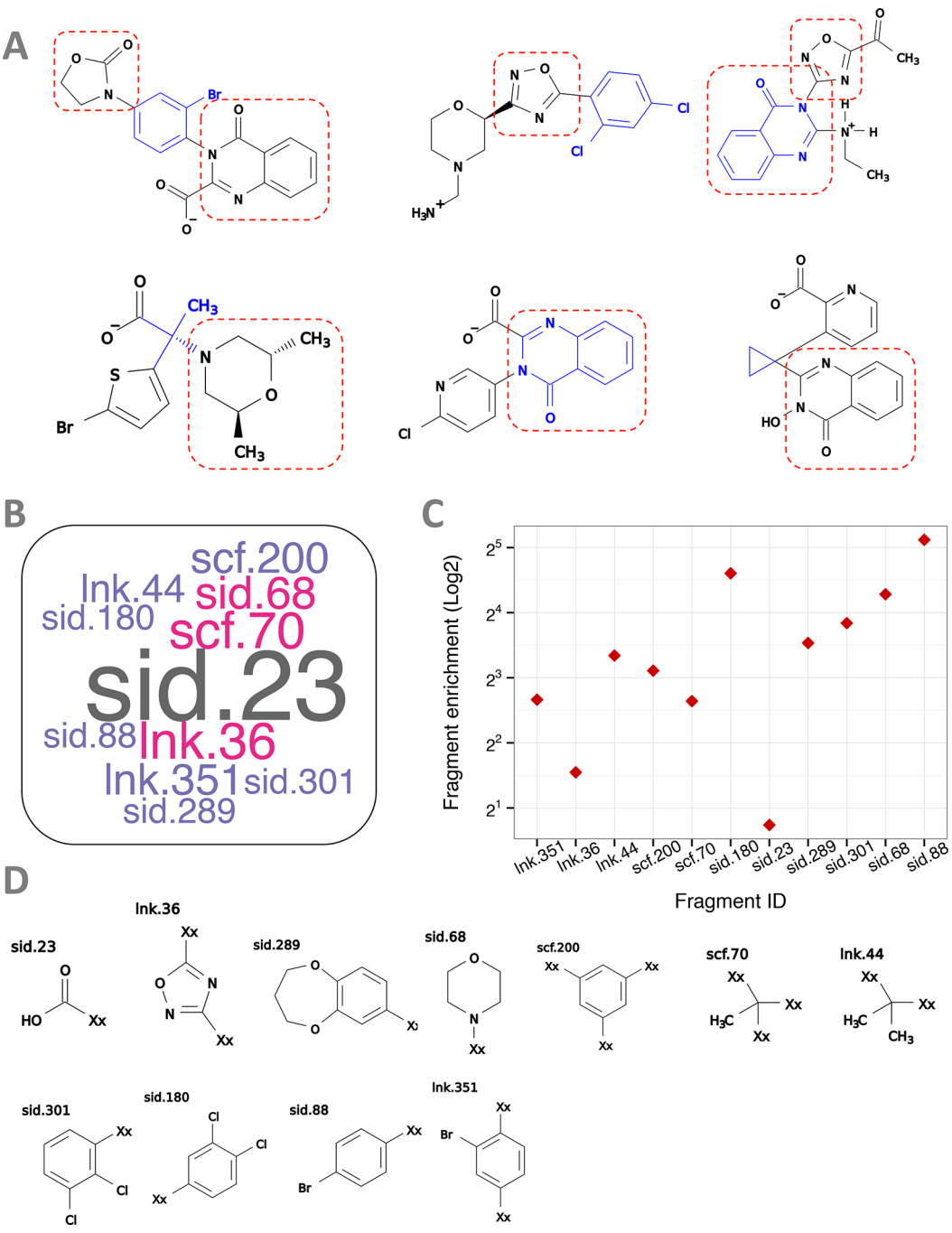


Figure 6. Properties of the selected de novo ligands in set 1. (A) Examples of interesting de novo ligands. See Supplementary Data for the full list. Several of them are reminiscent of quinolones, quinazolinones, oxadiazoles, and morpholine antifungals. Oxazolidinone, quinazolinone, oxadiazole, and morholine groups are highlighted with red, and the DOCK anchor fragments are highlighted with blue. A particularly interesting case (top left) contains a quinazolinone linker and an oxazolidinone side chain (oxazolidinones are among the newest antibiotics, e.g., linezolid, tedizolid,[9] used to treat vancomycin resistant Gram-positive bacteria). The compound at the top center is reminiscent of an oxadiazole antibiotic, while the bottom left compound resembles morpholine antifungals. (B) Word cloud of the significantly enriched fragments ($p < 0.05$, tests of proportions; sid. = side chain, a fragment with only one connection; lnk. = linker, a fragment with two connections; scf. = scaffold, a fragment with three connections). The size of the symbols corresponds to the abundance of the fragments. (C) Enrichment of the fragments. Enrichment was calculated as frequency in set 1/frequency in all de novo ligands with similar size and properties to set 1. (D) 2D structures of the significantly enriched fragments. Several fragments with halogen groups (Cl/Br) are present among them and also a morpholine side chain (sid.68).

sites of substrates and cofactors frequently form a continuum, and the structural constraints imposed by nucleic acid containing cofactors or ATP could contribute to the similarity of several de novo ligands to quinolones.

Optimization and Evolution of de Novo Ligands Using AI. The main limitation of using fragment libraries by de novo design tools is that the extent of chemical space that can be explored by the de novo ligands is limited by the fragment library. To overcome this, we further optimized the selected sets of de novo ligands (56 and 73 compounds) using the VAE (Figure 8A). We applied the following, so-called "hill climbing" protocol for ligand optimization. (1) First we docked the

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

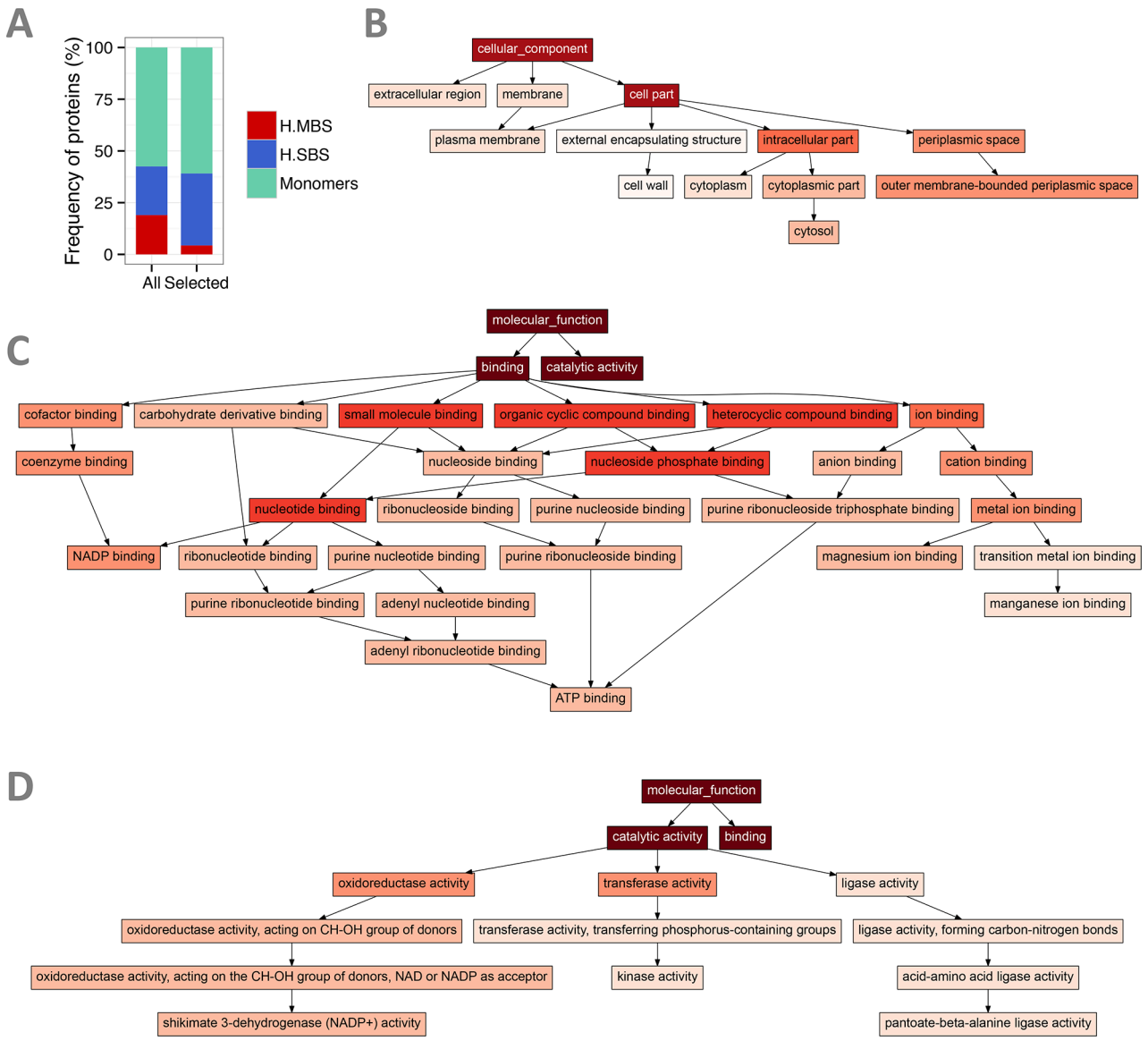


Figure 7. Gene ontology analysis of the receptors of selected ligands. (A) The quaternary structure composition is not significantly different from the total set of proteins ($p = 0.42$), although the frequency of MBS homomers is lower, as a result of the requirement to have similar structures in at least three species. (B) Graph of cellular component terms associated with the proteins. Note that the color-coding of terms (intensity of red) corresponds to the frequency of terms and not statistical significance. The proteins are present in most cellular components, including cell wall, periplasmic space, plasma membrane, and cytoplasm. (C, D) Graphs of molecular function terms associated with the proteins. The two highest-level terms are binding (C) and catalytic activity (D), with nucleoside/nucleotide binding and oxidoreductase activity being the most common terms.

selected ligands to all the receptors where a de novo ligand with distance in the latent space less than 13 was found and selected the three receptors, each from a different species, where the grid score is the lowest. (2) Next we encoded the ligand and sampled its neighborhood in the latent space for structurally related molecules (see Figure 8A and Figure S7). We used several different *Z*-distance cutoffs (see ref 34 and Methods), from 0.5 to 50, which correspond to increasingly larger added noise levels to the encoded molecule. (3) The neighbors were converted to 3D structures and were docked to the three previously selected receptors. If at least two neighbors had lower (better) grid scores than the encoded molecule, the best performing neighbor was selected, and the cycle was repeated until no improvement was detected (Figure 8A). This process generally resulted in lower grid scores (see Figure S8 for an example), and it also takes into account the similarity of properties like log *P*, QED, or SAS during the optimization of the ligand. (4) In the final step, using the ligands of the entire optimization process (including the neighbors that were not selected for further optimization), we selected the ligand with the best average grid score that also satisfied the same criteria as the original de novo ligand: log $P <$ 3, SAS < (molecular weight/100) + 1, and molecular weight of >300.

The majority of unoptimized de novo ligands could bind all three receptors without clashes (78% and 75%, in set 1 and set 2, respectively), and 98 and 95% of them could bind at least two. Nevertheless, the optimization resulted in a clear improvement of grid scores in both sets of molecules (Figure 8B, Figure S10A). While the final optimized molecules frequently differed, there was little difference in the magnitude of the grid score improvement when different noise levels (*Z*-distances) were used. This is due to several processes: first the characteristics of

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

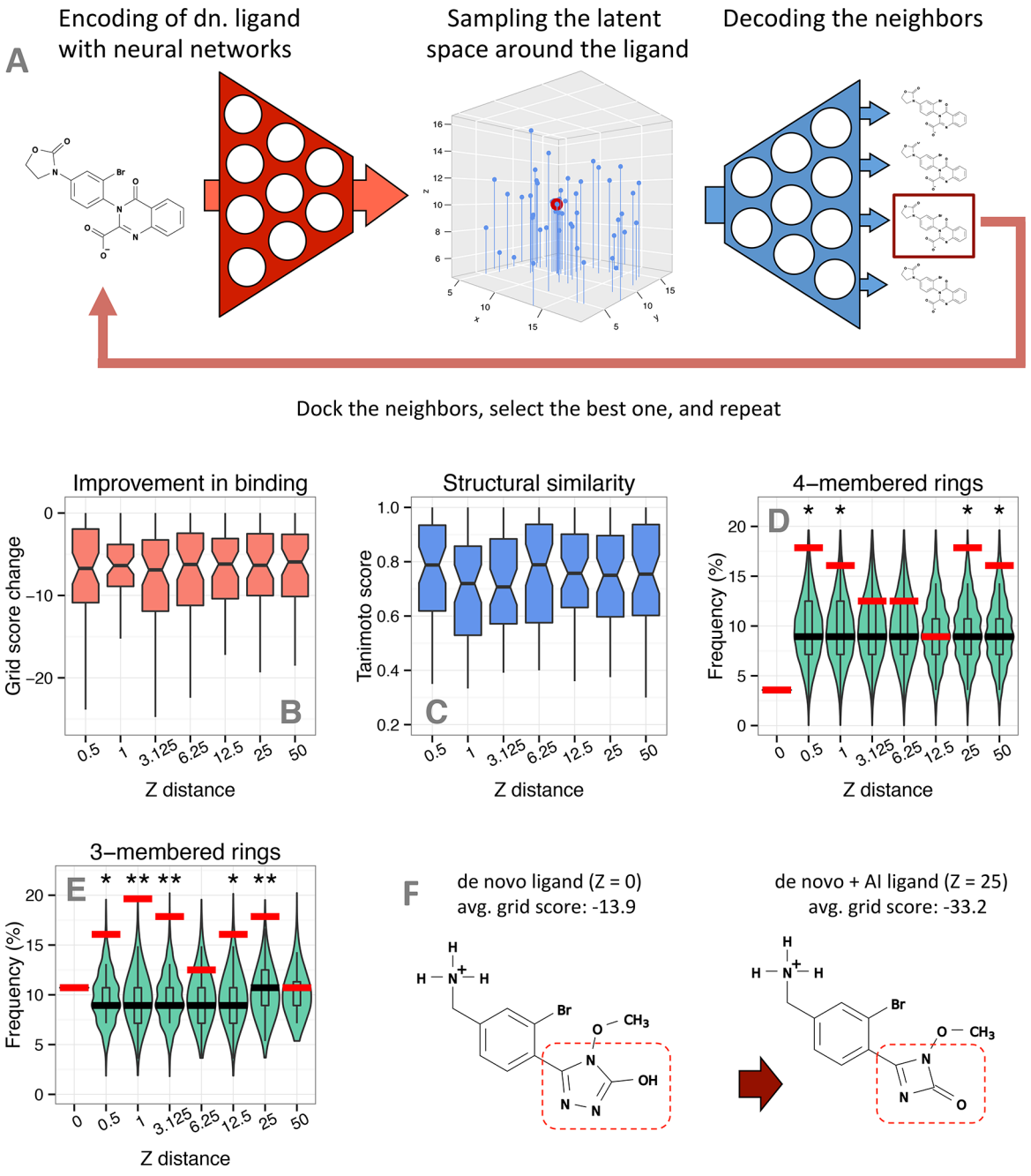


Figure 8. Optimization and evolution of ligands with AI. (A) Outline of the method. First, the selected de novo ligands were encoded with the VAE, and the latent space was sampled in their vicinity to search for related structures. Next the, samples were "decoded" into chemical structures and were docked into three different receptors, and the best performing (lowest scoring) molecule was used to repeat the cycle until there was no more improvement in the estimated binding energies (grid score). See also Figure S8. (B) The optimization resulted in a comparable improvement in binding energies (grid score), irrespective of the *Z* distance cutoff used. (C) The median Tanimoto similarity of optimized ligands and de novo ligands is 0.7. (D, E) The frequency of four- and three-membered rings in the optimized ligands is significantly higher than in the original de novo ligands and the random expectation (∗, $p < 0.05$; ∗∗, $p < 0.005$, randomization tests). Distance 0 indicates the original de novo ligands, red horizontal bars indicate the observed frequency of four- or three-membered rings, while black horizontal bars indicate their expected frequency. Violin plots show the frequency distribution of 10 000 random replicates. Outliers above 20% were not plotted for clarity. Note that in the case of four-member rings, the expected frequency is higher than their frequency among the de novo ligands ($Z = 0$) because they are more common in the VAE samples. (F) Example of the emergence of a β-lactam-like ring in a de novo ligand.

the VAE sampling play a role, as the structural diversity of the sampled SMILES changes little with the magnitude of the selected noise (*Z*-distance) level (Figure S9). With larger noise levels, the probability of a obtaining a valid SMILES string decreases, and in consequence most decoded ligands still originate from the neighborhood of the input even when the added noise (*Z*-distance) is large. (Note that the VAE is inherently stochastic, and decoding the same vector several times, without any added noise, results in variability in the returned SMILES.) Second, the ligands can probably be improved only to a limited degree, and the different runs reached different (and comparable) local optima. Similarly, the

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

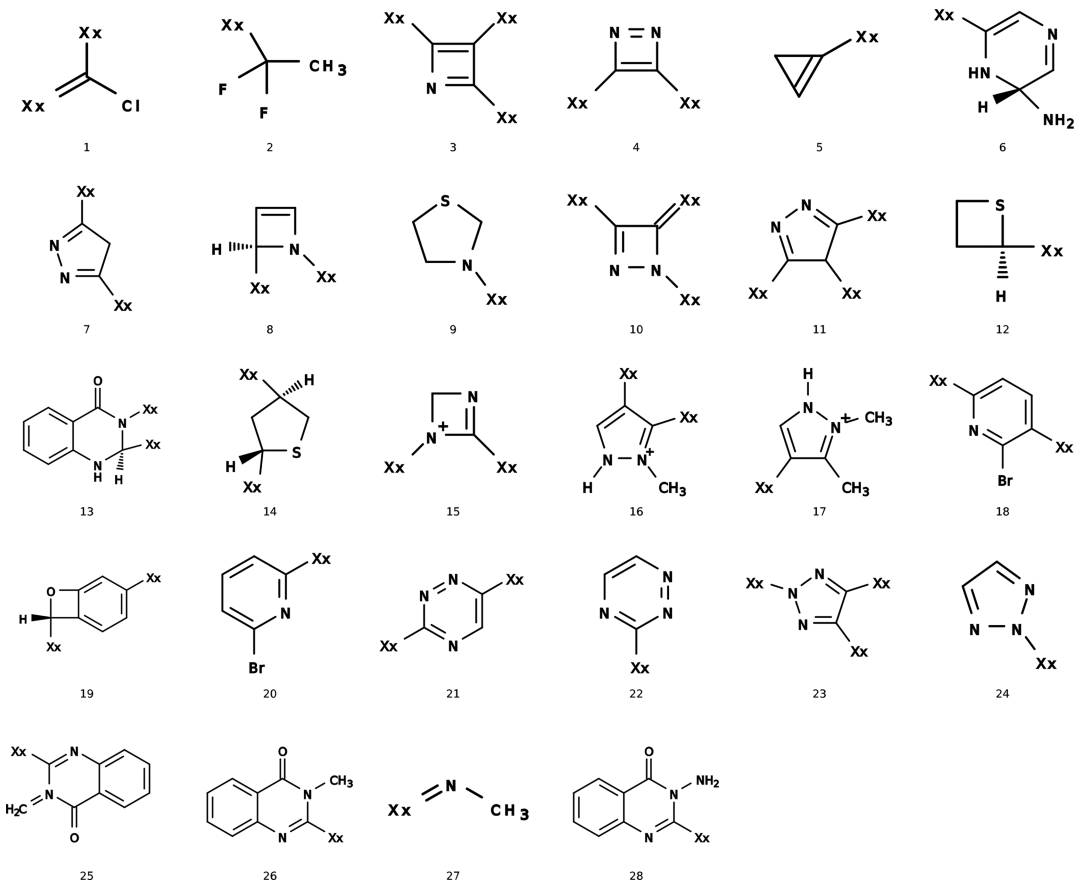


Figure 9. Fragments of the optimized ligands, which are present in at least two compounds in any of the VAE batches (excluding fragments that are likely to be toxic). Several halogenated compounds and four-membered rings are present among them. Since the four-membered rings frequently contain reactive double bonds, in practice their saturated versions should be used.

Tanimoto similarity score of the final optimized molecules with the original de novo ligand depends little on the *Z*-distance used and is approximately ∼0.75 for all *Z* cutoffs (Figure 8C, Figure S10B).

The optimized ligands show structural differences compared to the original de novo ligands. Quinolone-like and aromatic rings were generally not modified substantially by the optimization, while morpholine groups (and nonaromatic rings) were more variable and were frequently modified or substituted (see Supplementary Data Sets). Currently it is unclear to what degree these differences are caused by the nonhomogeneous nature or other characteristics of the latent space, as chemical generative methods are still under rapid development[49,50] (see also Segler et al.[51] for a different, recurrent neural network based approach). We tested the optimized (and de novo) ligands for toxicity and the presence of pan-assay interference compounds (PAINS)[52,53] with FAF-Drugs4[54] and the rd_filters tool, using the Glaxo filter.[55] We found that none of the de novo ligands and only a small fraction of optimized ligands (∼2%; 4 in set 1 and 7 in set 2, using all *Z*-distances) contain PAINS fragments. The fraction of putatively toxic ("rejected") compounds is, 6% and 9%, while the fraction of compounds with low-risk structural alerts is 50−60% both in de novo and in optimized ligands in the two sets using FAF-Drugs4 (note that low-risk structural alerts are frequent in existing drugs). The Glaxo filters flag ∼16% of the de novo and ∼20−21% of the optimized compounds as problematic in both sets (see Table S6 and Supplementary Data). However, most likely both tools underestimate the frequency of toxic or unstable fragments.

Generally the optimized ligands are characterized by a significantly higher frequency of three- and four-membered rings than the original de novo ligands (red bars, Figure 8D,E, Figure S10C,D). This pattern can be the result of two separate processes. First, in the molecules returned by the VAE sampling, the frequency of small rings can be higher than in the original de novo ligands due to the characteristics of the training set of the VAE. Second, such fragments can be enriched due to the structural constraints imposed by binding in several receptors. We separated these two processes by Monte Carlo simulations. We determined the presence of three- and four- membered rings in the molecules returned by the VAE sampling for each de novo ligand with RDKit. Next, we resampled the returned molecules 10 000 times to determine the random expectation for the frequency of three- and four-membered rings in the optimized ligands. The results show that the frequency of three- and four-membered rings in the final, optimized ligands is significantly higher (randomization test, see Methods) than the random expectation for most *Z*-distances, despite the fact that in the case of four-membered rings the expected frequency is considerably higher than their frequency in de novo ligands (Figure 8D,E, Figure S10C,D). This indicates that the necessity to bind multiple receptors selects for small rings, and their enrichment is not a simple byproduct of the sampling characteristics of the VAE.

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

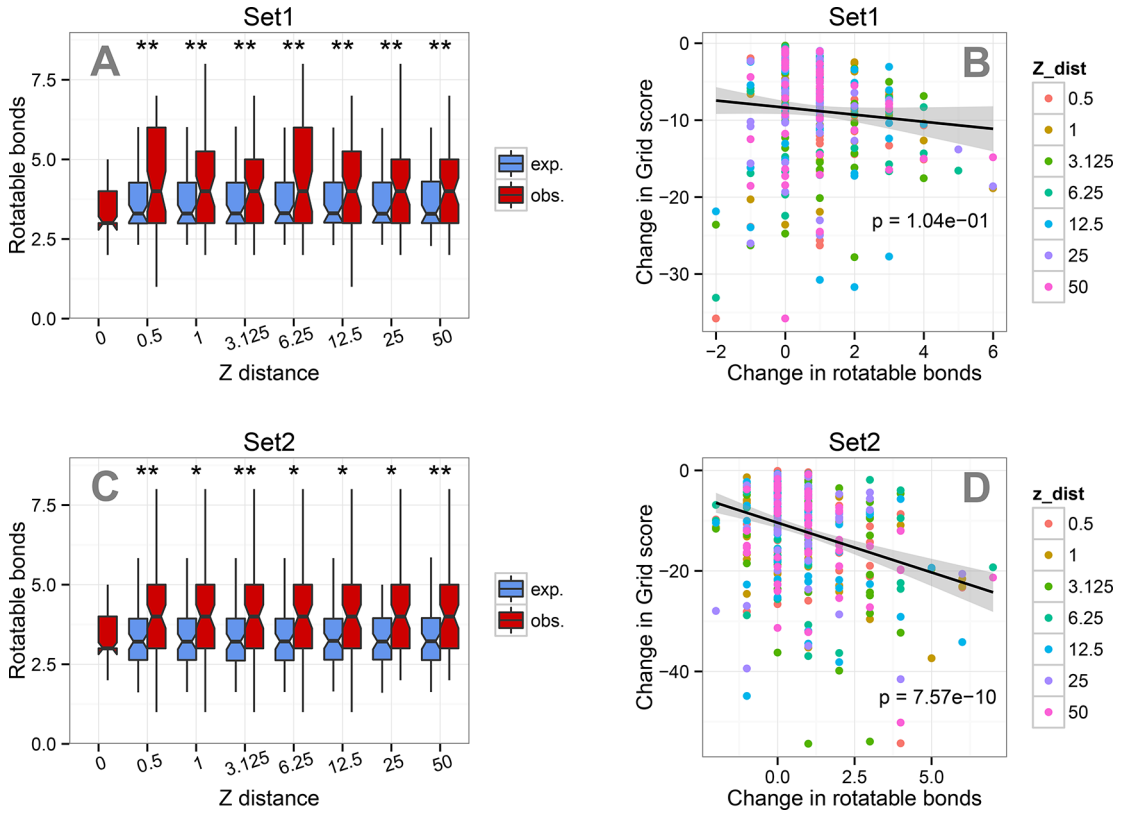


Figure 10. Ligands optimized for binding multiple receptors have more rotatable bonds than the original de novo ligands. (A) Number of rotatable bonds in de novo and VAE optimized ligands of set 1. Dark red represents the original de novo ($Z = 0$) or final optimized ligands, and blue indicates the expected number of rotatable bonds in the compounds returned by the VAE, without selection for multitarget binding. This was calculated as the averages of all ligands returned by the VAE when the de novo ligands were used as the input. In all VAE batches the number of rotatable bonds is significantly higher than in the expectation or de novo ligands (**, $p < 0.005$, Wilcoxon tests on the differences from de novo ligands). (B) Relationship between the change in rotatable bonds and grid score in set 1, using the pooled data of all VAE batches. The correlation is not significant due to a few outliers with rotatable bond change of −2; excluding them results in significant correlation ($p = 7.88 \times 10^{-4}$). (C) Number of rotatable bonds in de novo and VAE optimized ligands of set 2. The color coding is the same as on panel A. Similar to set 1, in all VAE batches the number of rotatable bonds is significantly higher than in the expectation or de novo ligands (**, $p < 0.005$, Wilcoxon tests on the differences from de novo ligands). (D) In set 2, the correlation between the change in rotatable bonds and grid score is highly significant ($p = 7.57 \times 10^{-10}$), indicating that the more flexible is the molecule, the better it can bind multiple receptors.

Neither the fragment library of DOCK nor the compounds of the VAE were specifically tailored for antibiotic design. The fragment library of DOCK was extracted from 13 million ZINC compounds by selecting the most common ones,[30] while the VAE was built from 250K randomly selected ZINC molecules.[34] Using the optimized compounds to update the de novo fragment library and repeating the entire de novo design and VAE optimization process several times may be useful in evolving fragment libraries tailored for specific tasks from a generic one (both for computational or experimental screening[56]). Using DOCK, we extracted the fragments from the optimized ligands and filtered out those already present in the input de novo ligands with Open Babel and also those lacking a carbon atom. This resulted in a very high diversity of fragments: altogether we identified 352 new fragments in the two sets (Figure S11), which do, however, contain several fragments that are toxic, reactive, or unstable, like epoxide, aziridine, or cyclobutadiene-like rings (and others). Note that different fragments can contain the same substructure if the number and location of their attachment points (−Xx) are different. However, the majority of them are present in only in a single structure, and selecting the ones that are present in at least two structures in the same VAE batches (and excluding putatively toxic ones) results in a much smaller set of 28 fragments (Figure 9), of which four contain halogens and seven contain four-membered rings, that are almost completely absent in the fragment library of DOCK. Several of them contain reactive double bonds; thus in practice their saturated versions should be used. The effect of such (e.g., azetidine-like) four-membered rings on toxicity is less clear; azetidine-based inhibitors of herpesvirus proteases are known and being developed,[57] and azetidines have been suggested for use as peptidomimetics in medicinal chemistry.[58] However, exactly due to its incorporation into proteins (as a substitute of proline), azetidine-2-carboxylic acid is known to be toxic.

Trade-Off between Preventing Resistance and Accumulation in Gram-Negative Bacteria. One of the key difficulties in developing broad spectrum antibiotics that are active against both Gram-positive and Gram-negative bacteria is due to the outer membrane and efficient efflux pumps in the latter, which prevent the accumulation of antibiotics. In general, accumulating compounds are expected to be small (<600 Da), but currently there are no established rules that would enable the design of compounds with a high likelihood off accumulation in Gram-negative bacteria. Recently it has been suggested that a few key parameters, like the number of rotatable bonds, the globularity of the molecule, and the presence of ionizable nitrogen (particularly primary amines) might be important for

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

accumulation and passing through the porins of the outer membrane.[59−61] It is unclear how general these rules are, as they were derived from a relatively small set of accumulating molecules; broad-spectrum β-lactam antibiotics (but also polymyxins, macrolides) typically have many more rotatable bonds than recommended, and neither fluoroquinolones nor broad-spectrum β-lactams are particularly enriched in ionizable primary amines (in fact, frequently lack them). However, some compounds active only against Gram-positive bacteria could be successfully converted to broad-spectrum antibiotics using these rules,[59] and recent measurements of accumulation in Gram-negative bacteria[62] also provide some support for them, which indicates that they are nevertheless a step in the right direction.

To test whether optimization to bind multiple receptors might influence the properties necessary for accumulation in Gram-negative bacteria, we examined whether it influences the number of rotatable bonds and globularity of the designed compounds (rotatable bonds were calculated ignoring hydrogens; thus methyl or amino groups were not assigned as rotatable; see Methods). We separated the possible biases of the VAE from the effect of optimization by calculating the number of rotatable bonds (and globularity) in all the compounds returned by the VAE when de novo ligands were used as the input (expectation) and compared it with the number of rotatable bonds (and globularity) in the final optimized ligands. We find that in both sets 1 and 2, optimization resulted in a highly significant increase in the number of rotatable bonds in all VAE optimized batches (Figure 10) and that there is a weak but significant correlation between the change in rotatable bonds and improvement in grid score (note that in the case of set 1 it is significant [$p = 7.88 \times 10^{-4}$] only after removing the outliers with a rotatable bond change of −2). The globularity of compounds (measured with the plane of best fit [PBF] algorithm;[63] see Methods) does not show a similar consistent change (Figure S12), although in some VAE batches of set 2, we do observe a slightly increased globularity compared to de novo ligands, mostly due to biases of the VAE (Figure S12C). The frequency of compounds with primary amines is relatively high in the de novo ligands (23% and 13% in set 1 and set 2, respectively), and somewhat lower in the optimized ligands (not shown). However, this is most likely due to the low frequency of primary amines in the compounds used to train the VAE (2.7%) rather than the result of binding multiple targets.

Taken together these results indicate that binding multiple receptors selects for flexible compounds with more rotatable bonds. Thus, preventing the evolution of resistance by multitargeting is likely to have the side effect that compounds capable of binding several receptors efficiently are at the same time less likely to accumulate in Gram-negative bacteria. Since high flexibility is the result of the necessity to adapt to binding sites with somewhat different topologies, one possible way to overcome this trade-off is to target binding sites with high structural similarity; thus MBS homomers may be preferable targets over other SBS homomers or monomers also for this reason.

DISCUSSION

Drug discovery is an extremely lengthy and costly process: on average, developing a new drug takes ∼14 years and costs an estimated 1−2 billion USD,[64] more than the cost of sending a spacecraft to an asteroid.[65] One of the earliest, and most critical steps in the drug discovery process is the identification of a druggable target protein, for which a ligand that modifies its function and has a therapeutic effect can be designed. Selecting the right drug target is critical, as poor target protein selection is one of the main causes of the low (∼10%) success rate of drug candidates entering the clinical phase.[64] Our results indicate that considering the quaternary structure of proteins can help in the selection of drug targets where the goal is targeting several different pathogen species. Ligands and de novo ligands of homomers, particularly MBS homomers, are much more likely to bind the binding sites of their equally diverged homologs than monomers (Figures 2, 3, 4). Since high conservation is always the consequence of strong constraints on function, the higher structural conservation of multichain binding sites in homomers[21] also indicates that such sites are less likely to accumulate mutations than the binding sites of SBS homomers or monomers. Finally, their more conserved binding sites reduce the need for flexibility and might improve the chances of accumulation in Gram-negative bacteria.

Successful antiretroviral drugs offer insights into this hypothesis, as the very high evolutionary rates of drug targets is a particularly severe problem in the treatment of retroviral infections like HIV. Interestingly, several antiretroviral drugs bind residues from more than one protein chain in their targets. HIV protease is an MBS homomer, and drugs targeting it bind both chains of its binding site.[66] Besides binding the active site, some HIV protease inhibitors (darunavir and tipranavir) also inhibit the dimerization of the protease, which has been suggested as a factor in the slow evolution resistance for these compounds.[66] Most HIV reverse transcriptase (RT) inhibitors are nucleoside analogs, although some of the non-nucleoside analog inhibitors like rilpivirine and etravirine bind residues from both chains of the RT heterodimer.[67,68] Drugs targeting HIV integrase typically bind the catalytic site, which interacts with DNA; however compounds not targeting the catalytic site bind allosteric pockets with residues from multiple chains.[69] Interestingly, one HIV integrase inhibitor, elvitegravir, is also characterized by a fluoroquinolone-like structure.[70]

Taken together, our results indicate that the high binding site similarity of MBS homomers makes them promising targets for broad spectrum antibacterial agents. Moreover, their slow structural change during evolution[21] suggests that targeting MBS homomers might also slow down the evolution of resistance, due to the high functional constraints on such sites. Targeting multichain binding sites might also result in the inhibition of protein complex formation itself, which is likely to have a significant effect on the evolution of resistance. Finally, our recent results show that allostery is much more frequent among MBS complexes than among SBS complexes, particularly in the case of homomers.[71] This indicates that MBS homomers also offer more targetable pockets than SBS homomers for drug development, and more diverse mechanisms can be exploited for developing novel inhibitors. Although our work focused primarily on homomers, we also note that a "trivial" way of achieving multitargeting is to target multichain binding sites of heteromeric protein complexes that are formed by multiple distinct polypeptide subunits. In humans, such sites are characterized by the highest frequency of pathogenic mutations in all quaternary structure types, indicating strong purifying selection.[21] However, MBS heteromers are generally not characterized by more conserved binding sites[21] or much higher frequency of allostery than monomers.[71]

The analysis of the structural properties of de novo ligands with similar de novo structures in several receptors indicates that such ligands do have common structural characteristics (Figure

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

6, Figure S5): they are enriched in halogen-containing, quinolone-like (quinazolinone), oxadiazole, and morpholine-like groups. Such fragments are characteristic in current broad-spectrum fluoroquinolone antibiotics (also the HIV integrase inhibitor elvitegravir), quinazolinone and oxadiazole antibacterials, and morpholine antifungals, suggesting that morpholine-based inhibitors could also be developed for bacteria. Additionally, the high frequency of quinazolinone fragments in our data sets suggests that quinazolinones are promising compounds for further development.

Halogens, particularly fluorine, are routinely used in medicinal chemistry for optimization and to improve ligand−target interactions.[72,73] However, the fact that halogen groups primarily interact with the carbonyl oxygens of the protein backbone and not with the side chains[72−74] makes them particularly suitable for use in "resistance-resistant" drugs,[75] as point mutations affect primarily the side chains of amino acids and change the protein backbone much less. Ligand−backbone interactions are already used in HIV reverse trascriptase inhibitors[75] and the protease inhibitor darunavir[66] to slow down the emergence of resistance (although in the latter it is not a halogen that interacts with the backbone). Additionally, our results suggest that the high overall frequency of halogens in drugs (~40% [72]) might be partly the result of interacting with several targets, even in the case of drugs that were originally designed to target only a single protein.

Most of the selected de novo compounds could bind several receptors without modifications, and they could be further improved by optimization with the VAE (Figure 8, Figure S10). In the majority of the cases the optimization resulted in compounds that are not dramatically different from the original de novo ligands, with Tanimoto similarity of ~0.7−0.75. However, in a fraction of the cases, it resulted in substantially different compounds (see Supplementary Data Sets). A characteristic structural difference in the optimized ligands is the appearance of small, three- and four-membered (sometimes β-lactam-like; see Figure 8F) rings, of which the latter are effectively absent in the original de novo ligands.

While the selected compounds contain many novel topologies, most of the interesting enriched fragments (quinolones/quinazolinones, oxadiazoles, morpholines, four-member rings [β-lactams/monobactams], organohalogens) are already used in the core pharmacophores of existing antimicrobials. This indicates that our approach has the potential to identify fragments relevant for antibiotic design but also that the chemical space of broad-spectrum antibiotics is not unlimited and, unfortunately, that the perception of the industry that many of the "low hanging fruit" antibiotic classes might have already been discovered[3,4] is not unfounded. From the strategies that help to preserve our current antibiotics, combination therapy is likely to be a successful strategy[17] as it effectively implements multitargeting, while recent work suggests that antibiotic cycling and mixing may not be very effective in practice.[76,77] We did not customize the fragment library used by de novo DOCK toward antibiotics, and also the VAE was based on a random selection of compounds[34] without any specific tailoring for antibiotic design. Thus, the emergence of compounds enriched in antibiotic-like fragments is not the result of such biases and indicates that the combination of fragment library based tools like DOCK, and deep-learning based tools like the VAE, that can explore the chemical space in ways fragment library based tools cannot is a powerful combination and can also be used to customize fragment libraries for a specific task. Its main current limitation is that some of the resulting compounds can be toxic, difficult to synthesize, or unstable (see Table S6 and Supplementary Data), and that occasionally the combination of certain fragments can result in compounds with protonation states that are not relevant at physiologically relevant pH.

Another limitation of our analysis is that the protein targets that resulted in the ligand sets used for the enrichment analysis and evolution using AI represent a relatively small set of molecular functions (Figure 7, Figure S6, Tables S4 and S5). Currently only a small fraction of the essential proteins of clinically relevant bacteria have a 3D structure deposited in the PDB, and even the available structures frequently cover only fragments or individual domains of them. Thus, our analysis was by necessity limited by the available structures, and a large, possibly global effort to determine the structures of the "essentialome" in the clinically most problematic microbes (e.g., ESKAPE) could have a major impact on designing new antibiotics with fundamentally novel structures.

METHODS

Selection of Bacterial Proteins. We selected the bacterial proteins for the analysis as follows. First, using the OGEE,[78] CEG,[79] and DEG[80] essential gene databases, we identified the known essential prokaryotic genes in the PDB. Next, using BLAST with an *e*-value cutoff of 10^{-3}, we removed those proteins from the data set that have a homolog in the human genome, the ones that form heteromeric protein complexes, and ones that have no structure with a small molecule ligand in the BioLiP database.[25] Finally, using BLAST with an *e*-value of 10^{-5} cutoff, we identified pairs of homologous bacterial proteins in the PDB where (1) the sequence similarity is below 40%, (2) at least one member of the pair is essential, (3) none of them have a homolog in the human proteome (*e*-value of 10^{-3}), (4) their structures overlap in the alignment, thus their structures also contain homologous regions, (5) they have similar quaternary structure, and (6) the type of ligand binding (MBS or SBS) is similar. The list of protein pairs and the PDB codes used in the analysis is available in Supporting Information Table S1.

Docking, Preparation of Receptors and Ligands, and de Novo Design of Ligands. We used DOCK 6.8[28] and its utility tools for binding site and ligand preparation, and docking. The first biological assembly was used for all PDB entries. Since some DOCK utility tools are unable to process large proteins and complexes, before docking we identified the residues of the receptor within 10 Å of the ligand with ProBiS, discarded all other residues of the structure, and built the grids using this substructure of the receptor. This did not have any measurable effect on the performance or accuracy of DOCK (the grids were built within 5 Å of the ligand) and enabled us to process large complexes. Receptor proteins and ligands were prepared with a standard procedure: in the receptor, incomplete side chains were completed, hydrogens were added, residues with low occupancy were removed, and Amber charges were added with the dock-prep tool of Chimera.[81] Ligands were converted to mol2 format, and hydrogens and Amber charges were added with Chimera. In a small number of cases where Chimera was unable to process the ligand, we added hydrogens and (Gasteiger) charges with OpenBabel.[82] To estimate the pose reproduction success rates, we first minimized the ligand through rigid docking, next docked it to its receptor with the FLX (flexible) algorithm.[28,83] The parameters were similar as in ref 83, the main parameters being max_orients = 1000, pruning_max_orients = 1000, pruning_clustering_cutoff = 100 (see http://ringo.ams.sunysb.edu/downloads/SB2010/FLX.in for the original parameters). We used only those receptors in further analyses where at least one of the first 10 clusterheads had RMSD < 2 Å with the original position of the ligand. We used this relaxed strategy because in the case of large ligands with several rotatable bonds, it is common that the core of the ligand is in the correct position, but certain side chains are not, pushing their RMSD above 2 Å. Furthermore, ligands in the PDB are not necessarily in a location that is the energetically most favorable, and in such cases a

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

"successful" pose reproduction is actually an error. Further docking was performed with the FLX protocol, both for the original binding site of the ligand and for the homologous binding site.

De novo ligands were built with the de novo algorithm of DOCK (a prerelease version of 6.9),[30] and we used the fragment library distributed with it, which was built using 13 million compounds of the ZINC database.[32] Only those binding sites were used where pose reproduction was successful, and binding sites of cofactors or metals were not used. We followed the following procedure for ligand generation: (1) First, we identified the number of rotatable bonds of the original ligand of the receptor with DOCK. (2) Next, we decomposed the original ligand of the receptor to fragments and identified the largest one. (3) From the fragment library we randomly selected a fragment with ±1 heavy atoms as the largest fragment of the original ligand for anchor. (For ligands where the largest fragment had 9 or more heavy atoms, fragments with 9−12 heavy atoms were used). (4) Next, the number of rotatable bonds of the ligand of the original binding site was determined with DOCK, and we set the number of growth layers as the number of rotatable bonds/2. Only binding sites of ligands with two or more rotatable bonds were processed. We used the graph sampling method, and the maximum allowed molecular weight of the de novo ligands was set to 550. (5) We ran 30 independent replicates for each receptor (i.e., we used 30 different anchors), and from the output of the 30 independent runs we selected a total of 50 de novo ligands with the best grid score. Overall this procedure resulted in de novo ligands of comparable size and complexity as the original ligand.

Ligand Characterization. The characteristics of the 50 best de novo ligands of each receptor were calculated with the chemical variational autoencoder (VAE) developed by the Aspuru-Guzik lab,[34] using the autoencoder distributed with the tool itself ("zinc_properties"). We used the VAE to calculate log *P*, synthetic accessibility score (SAS),[35] and quantitative druglikeness (QED)[36] for each de novo ligand, and also a vector corresponding to their representation in the latent space (with 196 dimensions). Molecular weight was calculated with the obprop tool of the Open Babel suite. Clustering of the latent space representation of de novo ligands (for visualization) was performed with Barnes−Hut t-SNE,[37] with two dimensions and perplexity 50.

Optimization and Evolution of de Novo Ligands with Docking and AI. The selected de novo ligands were further optimized and evolved with a strategy that used DOCK 6.8 and the VAE (Figure 8). First, from the list of binding sites that had a de novo ligand with a distance less than 13, we selected the binding site from the three different species where the selected de novo ligand could be docked with the best grid score. Next, each selected de novo ligand was converted to a SMILES string with RDKit, and we used the VAE to sample the latent space for structurally related chemicals, using *Z* cutoffs of 0.5, 1, 3.125, 6.25, 12.5, 25 and 50, taking 30 000 samples. This usually returned 10−100 unique chemical structures. If the number of returned chemicals was higher than 50, we randomly selected 50 of them. The structures returned by the VAE were docked to the three receptors, and the one with the best average score (if its score was an improvement in at least two of the receptors) was chosen and used in the next round of sampling/docking. This cycle was repeated as long as there was an improvement in the grid score of the new ligands.

Monte Carlo Simulations (Randomization Test). We performed Monte Carlo simulations to test whether small rings are enriched in the optimized molecules due to the constraints of binding in multiple receptors. First, in the SMILES returned in the first step of the optimization (see Figure S8 and FigureS8_FullResult.txt for an example) we determined for every atom whether it is part of a three-membered and four-membered ring using RDKit, and the number of such rings. Next we took 10 000 samples using all de novo ligands by selecting one SMILES randomly from the returned molecules of every de novo ligand and calculated the frequency of three- and four-membered ring in every random sample. Finally, we estimated whether the observed frequency of three- and four-membered rings is significantly higher than the expectation using the formula $p = (n + 1)/(N + 1)$, where N is the total number of samples (10 000) and n is the number of samples with equal or higher frequency of small rings than their observed frequency. The expected frequency of small rings was determined as the median of random samples.

Globularity and Rotatable Bond Estimation. Globularity was measured with the plane of best fit (PBF) method.[63] SMILES of de novo and VAE ligands were first converted to 3D structures with Open Babel using the --gen3d flag, which uses the MMFF94 force field. PBF scores were calculated on hydrogen-free structures,[63] with the pbf function of RDKit. A fraction of compounds (~3% in the optimized ligands, 5−10% in the raw ligands returned by the VAE) were excluded, due to the inability of RDKit to process their 3D structure. The number of rotatable bonds in each ligand was calculated with the obprop tool of the Open Babel suite, ignoring hydrogens; thus methyl or amino groups are not counted as rotatable bonds.

Visualization and Statistics. All statistical tests were performed with in-house Perl scripts and R and were corrected for multiple testing with the Benjamini−Hochberg method.[84] Protein structures were visualized with Chimera (version 1.11.2), chemical structures with Open Babel 2.3.1.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.9b00220.

Six tables of data on receptors, antibiotics, de novo sites, proteins, and PAINS (XLSX)

Two data sets (ZIP)

Twelve figures showing comparison of ligands, performance, chemical characteristics, weight−frequency relationships, gene ontology analysis results, sampled molecules, optimization steps, and globularity results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +44 131 651 8500. E-mail: gyorgy.abrusan@igmm.ed.ac.uk.

ORCID

György Abrusán: 0000-0003-4375-1552

Author Contributions

G.A. conceived the project, performed the analyses, and wrote the first version of the manuscript. J.A.M. contributed to the analysis and corrected the manuscript.

Notes

The authors declare no competing financial interest.

Biographies

György Abrusán obtained his M.Sc. in Biology in 1997 from the University of Debrecen (Hungary) and his Ph.D. in Hydrobiology from Warsaw University (Poland) in 2003. Currently he is a Research Fellow at the University of Edinburgh. His research interests include comparative genomics and computational structural biology, and focus on the factors governing the evolution of quaternary structure of proteins, and its dependence on interactions with ligands.

Joseph A. Marsh obtained his B.Sc. in Microbiology and Immunology from the University of British Columbia in 2003 and his Ph.D. in Biochemistry from the University of Toronto in 2010. This was followed by postdoctoral research at the MRC Laboratory of Molecular Biology and the EMBL European Bioinformatics Institute in Cambridge. He moved to the MRC Human Genetics Unit, University of Edinburgh, in 2014, where he is now a Reader. His research uses computational methods to study the structure, assembly, and evolution

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

of protein complexes and to investigate the implications of this for understanding human disease.

■ ACKNOWLEDGMENTS

We thank Lauren Prentis, Stephen Telehany, Scott Brozell, and Robert Rizzo for help with de novo DOCK and for providing the prerelease version of the source code of DOCK 6.9. We thank Neil Clark and Asier Unciti-Broceta for critical reading of the manuscript. We acknowledge the use of the Eddie3 computing cluster of the University of Edinburgh and thank the support of John Ireland and Steve Thorn from Research Computing. This work was supported by the Medical Research Council (Career Development Award MR/M02122X/1 for J.A.M.).

■ ABBREVIATIONS USED

GO, gene ontology; MBS, multichain binding site; PBF, plane of best fit; QED, quantitative estimate of druglikeness; SAS, synthetic accessibility; SBS, single-chain binding site; VAE, variational autoencoder

■ REFERENCES

(1) Bowater, L. *The Microbes Fight Back: Antibiotic Resistance*, Gld ed.; Royal Society of Chemistry: Cambridge, U.K., 2016.
(2) Davies, J.; Davies, D. Origins and Evolution of Antibiotic Resistance. *Microbiol Mol. Biol. Rev.* **2010**, *74* (3), 417−433.
(3) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for Bad Bugs: Confronting the Challenges of Antibacterial Discovery. *Nat. Rev. Drug Discovery* **2007**, *6* (1), 29−40.
(4) Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. Retrospective Analysis of Natural Products Provides Insights for Future Discovery Trends. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (22), 5601−5606.
(5) Skinnider, M. A.; Magarvey, N. A. Statistical Reanalysis of Natural Products Reveals Increasing Chemical Diversity. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (31), E6271−E6272.
(6) Toprak, E.; Veres, A.; Michel, J.-B.; Chait, R.; Hartl, D. L.; Kishony, R. Evolutionary Paths to Antibiotic Resistance under Dynamically Sustained Drug Selection. *Nat. Genet.* **2012**, *44* (1), 101−105.
(7) Silver, L. L. Multi-Targeting by Monotherapeutic Antibacterials. *Nat. Rev. Drug Discovery* **2007**, *6* (1), 41−55.
(8) Oldfield, E.; Feng, X. Resistance-Resistant Antibiotics. *Trends Pharmacol. Sci.* **2014**, *35* (12), 664−674.
(9) Singh, S. B.; Young, K.; Silver, L. L. What Is an "ideal" Antibiotic? Discovery Challenges and Path Forward. *Biochem. Pharmacol.* **2017**, *133*, 63−73.
(10) Dhanda, G.; Sarkar, P.; Samaddar, S.; Haldar, J. Battle against Vancomycin-Resistant Bacteria: Recent Developments in Chemical Strategies. *J. Med. Chem.* **2019**, *62* (7), 3184−3205.
(11) Brötz-Oesterhelt, H.; Brunner, N. A. How Many Modes of Action Should an Antibiotic Have? *Curr. Opin. Pharmacol.* **2008**, *8* (5), 564−573.
(12) Hu, Y.-Q.; Zhang, S.; Xu, Z.; Lv, Z.-S.; Liu, M.-L.; Feng, L.-S. 4-Quinolone Hybrids and Their Antibacterial Activities. *Eur. J. Med. Chem.* **2017**, *141*, 335−345.
(13) Hubschwerlen, C.; Specklin, J.-L.; Sigwalt, C.; Schroeder, S.; Locher, H. H. Design, Synthesis and Biological Evaluation of Oxazolidinone-Quinolone Hybrids. *Bioorg. Med. Chem.* **2003**, *11* (10), 2313−2319.
(14) Domalaon, R.; Idowu, T.; Zhanel, G. G.; Schweizer, F. Antibiotic Hybrids: The Next Generation of Agents and Adjuvants against Gram-Negative Pathogens? *Clin. Microbiol. Rev.* **2018**, *31* (2), No. e00077-17.
(15) Li, K.; Schurig-Briccio, L. A.; Feng, X.; Upadhyay, A.; Pujari, V.; Lechartier, B.; Fontes, F. L.; Yang, H.; Rao, G.; Zhu, W.; Gulati, A.; No, J. H.; Cintra, G.; Bogue, S.; Liu, Y.-L.; Molohon, K.; Orlean, P.; Mitchell, D. A.; Freitas-Junior, L.; Ren, F.; Sun, H.; Jiang, T.; Li, Y.; Guo, R.-T.; Cole, S. T.; Gennis, R. B.; Crick, D. C.; Oldfield, E. Multitarget Drug Discovery for Tuberculosis and Other Infectious Diseases. *J. Med. Chem.* **2014**, *57* (7), 3126−3139.
(16) Lázár, V.; Singh, G. P.; Spohn, R.; Nagy, I.; Horváth, B.; Hrtyan, M.; Busa Fekete, R.; Bogos, B.; Méhi, O.; Csörgő, B.; Pósfai, G.; Fekete, G.; Szappanos, B.; Kégl, B.; Papp, B.; Pál, C. Bacterial Evolution of Antibiotic Hypersensitivity. *Mol. Syst. Biol.* **2013**, *9* (1), 700.
(17) Pál, C.; Papp, B.; Lázár, V. Collateral Sensitivity of Antibiotic-Resistant Microbes. *Trends Microbiol.* **2015**, *23* (7), 401−407.
(18) Suzuki, S.; Horinouchi, T.; Furusawa, C. Prediction of Antibiotic Resistance by Gene Expression Profiles. *Nat. Commun.* **2014**, *5*, 5792.
(19) Chait, R.; Craney, A.; Kishony, R. Antibiotic Interactions That Select against Resistance. *Nature* **2007**, *446* (7136), 668−671.
(20) Michel, J.-B.; Yeh, P. J.; Chait, R.; Moellering, R. C.; Kishony, R. Drug Interactions Modulate the Potential for Evolution of Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (39), 14918−14923.
(21) Abrusán, G.; Marsh, J. A. Ligand Binding Site Structure Influences the Evolution of Protein Complex Function and Topology. *Cell Rep.* **2018**, *22* (12), 3265−3276.
(22) Lynch, M. The Evolution of Multimeric Protein Assemblages. *Mol. Biol. Evol.* **2012**, *29* (5), 1353−1366.
(23) Klabunde, T. Chemogenomic Approaches to Drug Discovery: Similar Receptors Bind Similar Ligands. *Br. J. Pharmacol.* **2007**, *152* (1), 5−7.
(24) Abrusán, G.; Marsh, J. A. Alpha Helices Are More Robust to Mutations than Beta Strands. *PLoS Comput. Biol.* **2016**, *12* (12), No. e1005242.
(25) Yang, J.; Roy, A.; Zhang, Y. BioLiP: A Semi-Manually Curated Database for Biologically Relevant Ligand-protein Interactions. *Nucleic Acids Res.* **2012**, *41* (D1), D1096−D1103.
(26) Konc, J.; Janežič, D. ProBiS Algorithm for Detection of Structurally Similar Protein Binding Sites by Local Structural Alignment. *Bioinformatics* **2010**, *26* (9), 1160−1168.
(27) Konc, J.; Janežič, D. ProBiS Tools (algorithm, Database, and Web Servers) for Predicting and Modeling of Biologically Interesting Proteins. *Prog. Biophys. Mol. Biol.* **2017**, *128*, 24−32.
(28) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36* (15), 1132−1156.
(29) Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a Pose Generation and Database Enrichment Tool. *J. Comput.-Aided Mol. Des.* **2012**, *26* (6), 749−773.
(30) Allen, W. J.; Fochtman, B. C.; Balius, T. E.; Rizzo, R. C. Customizable de Novo Design Strategies for DOCK: Application to HIVgp41 and Other Therapeutic Targets. *J. Comput. Chem.* **2017**, *38* (30), 2641−2663.
(31) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074−D1082.
(32) Sterling, T.; Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324−2337.
(33) O'Shea, R.; Moser, H. E. Physicochemical Properties of Antibacterial Compounds: Implications for Drug Discovery. *J. Med. Chem.* **2008**, *51* (10), 2871−2878.
(34) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268−276.
(35) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminf.* **2009**, *1*, 8.
(36) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* **2012**, *4* (2), 90.

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

(37) Van der Maaten, L. Accelerating T-SNE Using Tree-Based Algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 3221−3245.
(38) Bouley, R.; Kumarasiri, M.; Peng, Z.; Otero, L. H.; Song, W.; Suckow, M. A.; Schroeder, V. A.; Wolter, W. R.; Lastochkin, E.; Antunes, N. T.; Pi, H.; Vakulenko, S.; Hermoso, J. A.; Chang, M.; Mobashery, S. Discovery of Antibiotic (E)-3-(3-Carboxyphenyl)-2-(4-Cyanostyryl)quinazolin-4(3H)-One. *J. Am. Chem. Soc.* **2015**, *137* (5), 1738−1741.
(39) Bouley, R.; Ding, D.; Peng, Z.; Bastian, M.; Lastochkin, E.; Song, W.; Suckow, M. A.; Schroeder, V. A.; Wolter, W. R.; Mobashery, S.; Chang, M. Structure-Activity Relationship for the 4(3H)-Quinazolinone Antibacterials. *J. Med. Chem.* **2016**, *59* (10), 5011−5021.
(40) Gatadi, S.; Gour, J.; Shukla, M.; Kaul, G.; Das, S.; Dasgupta, A.; Malasala, S.; Borra, R. S.; Madhavi, Y. V.; Chopra, S.; Nanduri, S. Synthesis of 1,2,3-Triazole Linked 4(3H)-Quinazolinones as Potent Antibacterial Agents against Multidrug-Resistant Staphylococcus Aureus. *Eur. J. Med. Chem.* **2018**, *157*, 1056−1067.
(41) Gatadi, S.; Gour, J.; Kaul, G.; Shukla, M.; Dasgupta, A.; Akunuri, R.; Tripathi, R.; Madhavi, Y. V.; Chopra, S.; Nanduri, S. Synthesis of New 3-Phenylquinazolin-4(3H)-One Derivatives as Potent Antibacterial Agents Effective against Methicillin- and Vancomycin-Resistant Staphylococcus Aureus (MRSA and VRSA). *Bioorg. Chem.* **2018**, *81*, 175−183.
(42) O'Daniel, P. I.; Peng, Z.; Pi, H.; Testero, S. A.; Ding, D.; Spink, E.; Leemans, E.; Boudreau, M. A.; Yamaguchi, T.; Schroeder, V. A.; Wolter, W. R.; Llarrull, L. I.; Song, W.; Lastochkin, E.; Kumarasiri, M.; Antunes, N. T.; Espahbodi, M.; Lichtenwalter, K.; Suckow, M. A.; Vakulenko, S.; Mobashery, S.; Chang, M. Discovery of a New Class of Non-B-Lactam Inhibitors of Penicillin-Binding Proteins with Gram-Positive Antibacterial Activity. *J. Am. Chem. Soc.* **2014**, *136* (9), 3664−3672.
(43) Spink, E.; Ding, D.; Peng, Z.; Boudreau, M. A.; Leemans, E.; Lastochkin, E.; Song, W.; Lichtenwalter, K.; O'Daniel, P. I.; Testero, S. A.; Pi, H.; Schroeder, V. A.; Wolter, W. R.; Antunes, N. T.; Suckow, M. A.; Vakulenko, S.; Chang, M.; Mobashery, S. Structure-Activity Relationship for the Oxadiazole Class of Antibiotics. *J. Med. Chem.* **2015**, *58* (3), 1380−1389.
(44) Janardhanan, J.; Chang, M.; Mobashery, S. The Oxadiazole Antibacterials. *Curr. Opin. Microbiol.* **2016**, *33*, 13−17.
(45) Campoy, S.; Adrio, J. L. Antifungals. *Biochem. Pharmacol.* **2017**, *133*, 86−96.
(46) Mercer, E. I. Morpholine Antifungals and Their Mode of Action. *Biochem. Soc. Trans.* **1991**, *19* (3), 788−793.
(47) Prasad, R.; Shah, A. H.; Rawal, M. K. Antifungals: Mechanism of Action and Drug Resistance. In *Yeast Membrane Transport*; Ramos, J., Sychrová, H., Kschischo, M., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, Switzerland, 2016; pp 327−349.
(48) Aldred, K. J.; Kerns, R. J.; Osheroff, N. Mechanism of Quinolone Action and Resistance. *Biochemistry* **2014**, *53* (10), 1565−1574.
(49) Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; Kadurin, A.; Nikolenko, S.; Aspuru-Guzik, A.; Zhavoronkov, A. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *ArXiv* **2018**, 1811.12823.
(50) Brown, N.; Fiscato, M.; Segler, M. H. S.; Vaucher, A. C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59* (3), 1096−1108.
(51) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4* (1), 120−131.
(52) 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.* **2010**, *53* (7), 2719−2740.
(53) Baell, J. B.; Nissink, J. W. M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. *ACS Chem. Biol.* **2018**, *13* (1), 36−44.
(54) Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O. FAF-Drugs3: A Web Server for Compound Property Calculation and Chemical Library Design. *Nucleic Acids Res.* **2015**, *43* (W1), W200−W207.
(55) Walters, P. A Script to Run Structural Alerts Using the RDKit and ChEMBL. 2018. https://github.com/PatWalters/rd_filters (accessed Apr 14, 2019).
(56) Lamoree, B.; Hubbard, R. E. Using Fragment-Based Approaches to Discover New Antibiotics. *Slas Discovery* **2018**, *23* (6), 495−510.
(57) Gable, J. E.; Acker, T. M.; Craik, C. S. Current and Potential Treatments for Ubiquitous but Neglected Herpesvirus Infections. *Chem. Rev.* **2014**, *114* (22), 11382−11412.
(58) Glawar, A. F. G.; Jenkinson, S. F.; Thompson, A. L.; Nakagawa, S.; Kato, A.; Butters, T. D.; Fleet, G. W. J. 3-Hydroxyazetidine Carboxylic Acids: Non-Proteinogenic Amino Acids for Medicinal Chemists. *ChemMedChem* **2013**, *8* (4), 658−666.
(59) Richter, M. F.; Drown, B. S.; Riley, A. P.; Garcia, A.; Shirai, T.; Svec, R. L.; Hergenrother, P. J. Predictive Compound Accumulation Rules Yield a Broad-Spectrum Antibiotic. *Nature* **2017**, *545* (7654), 299−304.
(60) Spaulding, A.; Takrouri, K.; Mahalingam, P.; Cleary, D. C.; Cooper, H. D.; Zucchi, P.; Tear, W.; Koleva, B.; Beuning, P. J.; Hirsch, E. B.; Aggen, J. B. Compound Design Guidelines for Evading the Efflux and Permeation Barriers of Escherichia Coli with the Oxazolidinone Class of Antibacterials: Test Case for a General Approach to Improving Whole Cell Gram-Negative Activity. *Bioorg. Med. Chem. Lett.* **2017**, *27* (23), 5310−5321.
(61) Richter, M. F.; Hergenrother, P. J. The Challenge of Converting Gram-Positive-Only Compounds into Broad-Spectrum Antibiotics. *Ann. N. Y. Acad. Sci.* **2019**, *1435* (1), 18−38.
(62) Acosta-Gutiérrez, S.; Ferrara, L.; Pathania, M.; Masi, M.; Wang, J.; Bodrenko, I.; Zahn, M.; Winterhalter, M.; Stavenger, R. A.; Pagès, J.-M.; Naismith, J. H.; van den Berg, B.; Page, M. G. P.; Ceccarelli, M. Getting Drugs into Gram-Negative Bacteria: Rational Rules for Permeation through General Porins. *ACS Infect. Dis.* **2018**, *4* (10), 1487−1498.
(63) Firth, N. C.; Brown, N.; Blagg, J. Plane of Best Fit: A Novel Method to Characterize the Three-Dimensionality of Molecules. *J. Chem. Inf. Model.* **2012**, *52* (10), 2516−2525.
(64) Nicolaou, K. C. Advancing the Drug Discovery and Development Process. *Angew. Chem., Int. Ed.* **2014**, *53* (35), 9128−9140.
(65) OSIRIS-REx. Wikipedia, 2018.
(66) Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* **2016**, *59* (11), 5172−5208.
(67) Das, K.; Bauman, J. D.; Clark, A. D.; Frenkel, Y. V.; Lewi, P. J.; Shatkin, A. J.; Hughes, S. H.; Arnold, E. High-Resolution Structures of HIV-1 Reverse transcriptase/TMC278 Complexes: Strategic Flexibility Explains Potency against Resistance Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (5), 1466−1471.
(68) Lansdon, E. B.; Brendza, K. M.; Hung, M.; Wang, R.; Mukund, S.; Jin, D.; Birkus, G.; Kutty, N.; Liu, X. Crystal Structures of HIV-1 Reverse Transcriptase with Etravirine (TMC125) and Rilpivirine (TMC278): Implications for Drug Design. *J. Med. Chem.* **2010**, *53* (10), 4295−4299.
(69) Fader, L. D.; Malenfant, E.; Parisien, M.; Carson, R.; Bilodeau, F.; Landry, S.; Pesant, M.; Brochu, C.; Morin, S.; Chabot, C.; Halmos, T.; Bousquet, Y.; Bailey, M. D.; Kawai, S. H.; Coulombe, R.; LaPlante, S.; Jakalian, A.; Bhardwaj, P. K.; Wernic, D.; Schroeder, P.; Amad, M.; Edwards, P.; Garneau, M.; Duan, J.; Cordingley, M.; Bethell, R.; Mason, S. W.; Bös, M.; Bonneau, P.; Poupart, M.-A.; Faucher, A.-M.; Simoneau, B.; Fenwick, C.; Yoakim, C.; Tsantrizos, Y. Discovery of BI 224436, a Noncatalytic Site Integrase Inhibitor (NCINI) of HIV-1. *ACS Med. Chem. Lett.* **2014**, *5* (4), 422−427.
(70) Hare, S.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. Retroviral Intasome Assembly and Inhibition of DNA Strand Transfer. *Nature* **2010**, *464* (7286), 232−236.

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX

(71) Abrusán, G.; Marsh, J. A. Ligand Binding Site Structure Shapes Allosteric Signal Transduction and the Evolution of Allostery in Protein Complexes. *Mol. Biol. Evol.* **2019**, DOI: 10.1093/molbev/msz093.
(72) Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G. The Halogen Bond. *Chem. Rev.* **2016**, *116* (4), 2478−2601.
(73) Lu, Y.; Liu, Y.; Xu, Z.; Li, H.; Liu, H.; Zhu, W. Halogen Bonding for Rational Drug Design and New Drug Discovery. *Expert Opin. Drug Discovery* **2012**, *7* (5), 375−383.
(74) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen Bonds in Biological Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (48), 16789−16794.
(75) Himmel, D. M.; Das, K.; Clark, A. D.; Hughes, S. H.; Benjahad, A.; Oumouch, S.; Guillemont, J.; Coupa, S.; Poncelet, A.; Csoka, I.; Meyer, C.; Andries, K.; Nguyen, C. H.; Grierson, D. S.; Arnold, E. Crystal Structures for HIV-1 Reverse Transcriptase in Complexes with Three Pyridinone Derivatives: A New Class of Non-Nucleoside Inhibitors Effective against a Broad Range of Drug-Resistant Strains. *J. Med. Chem.* **2005**, *48* (24), 7582−7591.
(76) Van Duijn, P. J.; Verbrugghe, W.; Jorens, P. G.; Spöhr, F.; Schedler, D.; Deja, M.; Rothbart, A.; Annane, D.; Lawrence, C.; Nguyen Van, J.-C.; Misset, B.; Jereb, M.; Seme, K.; Šifrer, F.; Tomiç, V.; Estevez, F.; Carneiro, J.; Harbarth, S.; Eijkemans, M. J. C.; Bonten, M.; Goossens, H.; Malhotra-Kumar, S.; Lammens, C.; Vila, J.; Roca, I. The Effects of Antibiotic Cycling and Mixing on Antibiotic Resistance in Intensive Care Units: A Cluster-Randomised Crossover Trial. *Lancet Infect. Dis.* **2018**, *18* (4), 401−409.
(77) Beardmore, R. E.; Peña-Miller, R.; Gori, F.; Iredell, J. Antibiotic Cycling and Antibiotic Mixing: Which One Best Mitigates Antibiotic Resistance? *Mol. Biol. Evol.* **2017**, *34* (4), 802−817.
(78) Chen, W.-H.; Lu, G.; Chen, X.; Zhao, X.-M.; Bork, P. OGEE v2: An Update of the Online Gene Essentiality Database with Special Focus on Differentially Essential Genes in Human Cancer Cell Lines. *Nucleic Acids Res.* **2017**, *45* (D1), D940−D944.
(79) Ye, Y.-N.; Hua, Z.-G.; Huang, J.; Rao, N.; Guo, F.-B. CEG: A Database of Essential Gene Clusters. *BMC Genomics* **2013**, *14*, 769.
(80) Luo, H.; Lin, Y.; Gao, F.; Zhang, C.-T.; Zhang, R. DEG 10, an Update of the Database of Essential Genes That Includes Both Protein-Coding Genes and Noncoding Genomic Elements. *Nucleic Acids Res.* **2014**, *42* (D1), D574−D580.
(81) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605−1612.
(82) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 33.
(83) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *J. Chem. Inf. Model.* **2010**, *50* (11), 1986−2000.
(84) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57* (1), 289−300.

DOI: 10.1021/acs.jmedchem.9b00220
J. Med. Chem. XXXX, XXX, XXX−XXX