DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking

Trent E Balius*, Y. Stanley Tan, and Mayukh Chakrabarti

NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., PO Box B, Frederick, MD 21702

* Corresponding Author: trent.balius@nih.gov

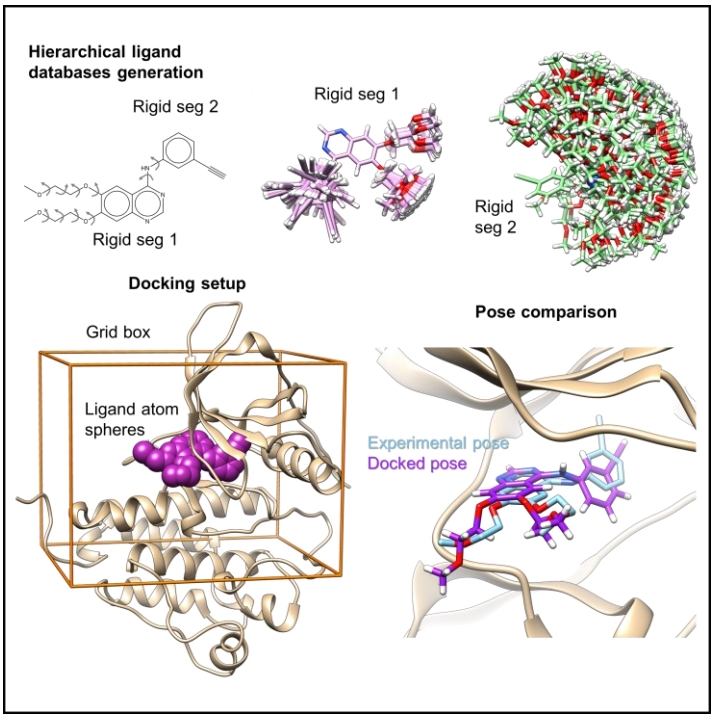
Abstract. (long)

Molecular Docking is a prevalent tool in drug discovery campaigns. DOCK 6's extensible design lends itself to implementing and testing new methods in molecular docking. Development in DOCK 3 has enabled the screening of large databases of now billions of small molecules, allowing us access to unprecedented chemical space. To allow DOCK 6 to access this scale, we have implemented into DOCK 6 a search routine that traverses precomputed ligand conformations stored in a hierarchical database format, as is done in DOCK 3.7, and we have updated the ChemGrid scoring function to mirror the one in DOCK 3.7. Features are tested on the DUDE-Z and SB2012 test sets. We also modified DOCK 6 to be used as a conformation generation tool. The hierarchical database search routine is 16 times faster than DOCK 6's default sampling routine, anchor-and-grow, docking molecules in seconds compared to minutes. Pose reproduction *docking success* rate (defined as the percentage of systems in which the best energy pose has an RMSD <= 2.0 Å from the crystal pose) over the SB2012 test set is 76% for anchor-and-grow, and 60% for the hierarchical database search routine. When we seed the hierarchy with the crystallographic conformation, we obtain a 91.5% success rate. In comparing DOCK 3.7 with DOCK 6, the enrichment performance, quantified using the area under the curve (AUC) of the receiver operating characteristic (ROC) curves, tested on the DUDE-Z benchmark, is on average similar, but DOCK 3 gives overall better enrichments (ΔAUC = 0.92, ΔlogAUC = 1.14). However, with post-docking torsion minimization, DOCK 6 surpasses DOCK 3.7. Over the DUDE-Z test set, DOCK 6 is on average 1.7 times slower than DOCK 3. A proof-of-concept large-scale virtual screen is carried out with DOCK 6, demonstrating its ability to perform docking of 23 million fragment molecules. We use current features in DOCK 6 to complement hierarchical database calculations, including *best first clustering* and *torsion minimization*, both of which are not available in DOCK 3.

Abstract. (150 words)

To allow DOCK 6 access to unprecedented chemical space for screening billions of small molecules, we have implemented features from DOCK 3.7 into DOCK 6, including a search routine that traverses precomputed ligand conformations stored in a hierarchical database. We tested them on the DUDE-Z and SB2012 test sets. The hierarchical database search routine is 16 times faster than anchor-and-grow. However, the ability of hierarchical database search to reproduce the experimental pose is 16% worse than that of anchor-and-grow. The enrichment performance is on average similar, but DOCK 3.7 has better enrichment than DOCK 6, and DOCK 6 is on average 1.7 times slower. However, with post-docking torsion minimization, DOCK 6 surpasses DOCK 3.7. A large-scale virtual screen is performed with DOCK 6 on 23 million fragment molecules. We use current features in DOCK 6 to complement hierarchical database calculations, including *torsion minimization*, which is not available in DOCK 3.7.

Graphical table of contents image.



Calculations for EGFR-erlotinib complex (PDB code: 1M17). Erlotinib chemical structure and two conformational expansions about the central quinazoline ring (rigid seg 1) and the phenylacetylene ring (rigid seg 2). Docking setup of the receptor (EGFR kinase domain) with ligand atom spheres and grid box shown. Experimental pose with the best RMSD docked pose (the best score pose does not reproduce the experimental pose for this system).

Summary (75 words)

Molecular Docking is a prevalent tool in drug discovery. DOCK 6's extensible design enables implementing and testing new methods in molecular docking. Development in DOCK 3 enabled

screening of large databases of billions of small molecules. To allow access to this unprecedented chemical space, we have implemented features from DOCK 3.7 into DOCK 6, including traversal of precomputed ligand conformations stored in a hierarchical database. We test these new features retrospectively.

Keywords. Molecular Docking, Large-scale virtual screening, Flexibase search, drug-lead discovery method.

1. Introduction.

Molecular docking methods are computational techniques for predicting the binding mode of a ligand, a small organic molecule, in a pocket on a receptor, most often a protein drug target. One of the most useful applications of docking is virtual screening, where a computational database of small molecules is screened against a pocket of interest, *i.e.*, each molecule is placed into a pocket on a protein in a compatible pose, assigned a score, and rank ordered. Molecules at the top of the list are investigated experimentally.

In docking, two principal tasks are performed: sampling and scoring. To ensure that calculations can be performed rapidly, exhaustive sampling is impractical, and approximations are made during scoring (e.g., pre-computing energies on a grid and using tri-linear interpolation). Because shortcuts are deployed, docking methods are evaluated retrospectively to determine their performance. Crucial to docking success is the ability for experimental tests to verify that the predicted ligand binds, and because of the low success rate, an adequate number of predictions must be tested. Despite its limitations, there are many successes of docking software that have been published in the literature.[1-3] Indeed, these methods have been tested and have resulted in the discovery of ligands that bind potently to a protein, as verified through experimental assays and structural predictions.[4-9]

UCSF DOCK, developed by Kuntz and colleagues, is the first molecular docking program[10], and has been used to conduct some of the earliest docking screens for inhibitor discovery[11,12], spurring the subsequent development of many other docking programs over the last four decades. Examples of such programs from the literature include DiffDock[13], DPL[14], GNINA[15], AutoDock[16,17]/AutoDock Vina[18], Glide[19], RosettaLigand[20], FlexX[21], ICM[22], Gold[23], MCDock[24], PLANTS[25], rDock[26] and many others[27]. These programs deploy an array of sampling methods, including incremental construction (as is done in DOCK 6), diffusion-based generative docking models powered by machine learning (DiffDock and DPL), stochastic methods like Genetic algorithms (AutoDOCK) and Monte Carlo (MCDock), controlled random walks, swarm-intelligence methods (PLANTS), and traversal through precomputed conformations (as is done in DOCK 3.7). Sampling methods mostly focus on ligand flexibility but may also explore local movements in the receptor pocket (like in ICM). The concept of pre-computing molecular conformations to account for ligand flexibility has been recognized and implemented in docking programs since the early 1990s.[28,29] The programs also deploy many different scoring functions. These scoring functions are generally grouped into three classes: physics-based (used in both DOCK 3 and 6), knowledge-based, and empirical.[30] Most docking software are developed with proteins as the macromolecule of interest, but a subset of software also focuses on other biomolecules, like RNA and DNA (e.g., rDOCK).

Recent advances in docking and chemistry have given us unprecedented access to chemical space, culminating in large-scale docking successes.[5,6,8,9] To keep up with the growth of chemical space, our docking methods need to be fast (and ever faster). DOCK 3.7 has been used for large-scale docking, and spends on average less than one second per molecule, allowing us to dock hundreds of millions[31]— now billions[32]—of molecules.[33] DOCK 6.9, using anchor-and-grow, spends on average one minute per molecule, making it difficult to dock at this scale. Here, we have ported features from DOCK 3.7 to DOCK 6.9. By implementing a hierarchal database search routine in DOCK 6 (see methods), which traverses through DB2 files in a similar way to that in DOCK 3, we enable DOCK 6 to perform larger screens—as in DOCK 3—and leverage all the useful features that are in DOCK 6. These new features will be released in a future version of DOCK 6.

In this paper, we present the following work. **1**. We enabled and used DOCK 6 as a conformation generation tool. **2**. We implemented a hierarchical database search routine in DOCK 6. **3**. We updated ChemGrid score (DOCK3.5 score) to match the DOCK 3.7 scoring function. **4.** We performed retrospective testing on the DUDE-Z and SB2012 test sets to evaluate the new features and compare DOCK 6 to DOCK 3. **5.** Finally, we performed a proof-of-principle virtual screen to a pocket on the important cancer drug target KRAS.

2. Methods.

For the hierarchical database (HDB) search routine, we precompute molecule conformations, store the conformations in a flexibase DB2 file[34] with a defined format (described below), read these DB2 files in DOCK, and orient and search through these conformations to find the best scoring poses. In the next two sections, we describe the HDB search implementation in detail.

DB2 file format.

The DB2 database file format is highly compressed (**Figure 1 A-D**). In the DB2 format, one molecule will have one or more hierarchies. All conformations are generated with a rigid position defined; all movement of the molecule is relative to this rigid segment (**Figure 1B**). Consider a molecule with multiple rings. Each ring will be treated as a rigid segment, and a hierarchy will be generated for each ring. If a molecule has three rings, then three hierarchies, or databases, will be generated.

The database is a tree, where the root represents the rigid segment, each node is a segment of the molecule in a specific conformation, and a branch of the tree is a completely grown conformation of the molecule (**Figure 1 B, D**). The root node, a rigid segment, is shared among all branch conformations. When visualizing the database, this rigid segment is the focus: all other atoms have more movement (**Figure 1B**).

The DB2 files are divided into sections: Header information (M), atom information (A), coordinate information (X, coord), segment information (C, conf), branch or conformer information (S, Set), and clusters (D, clust; currently the cluster information is not being used).[35]

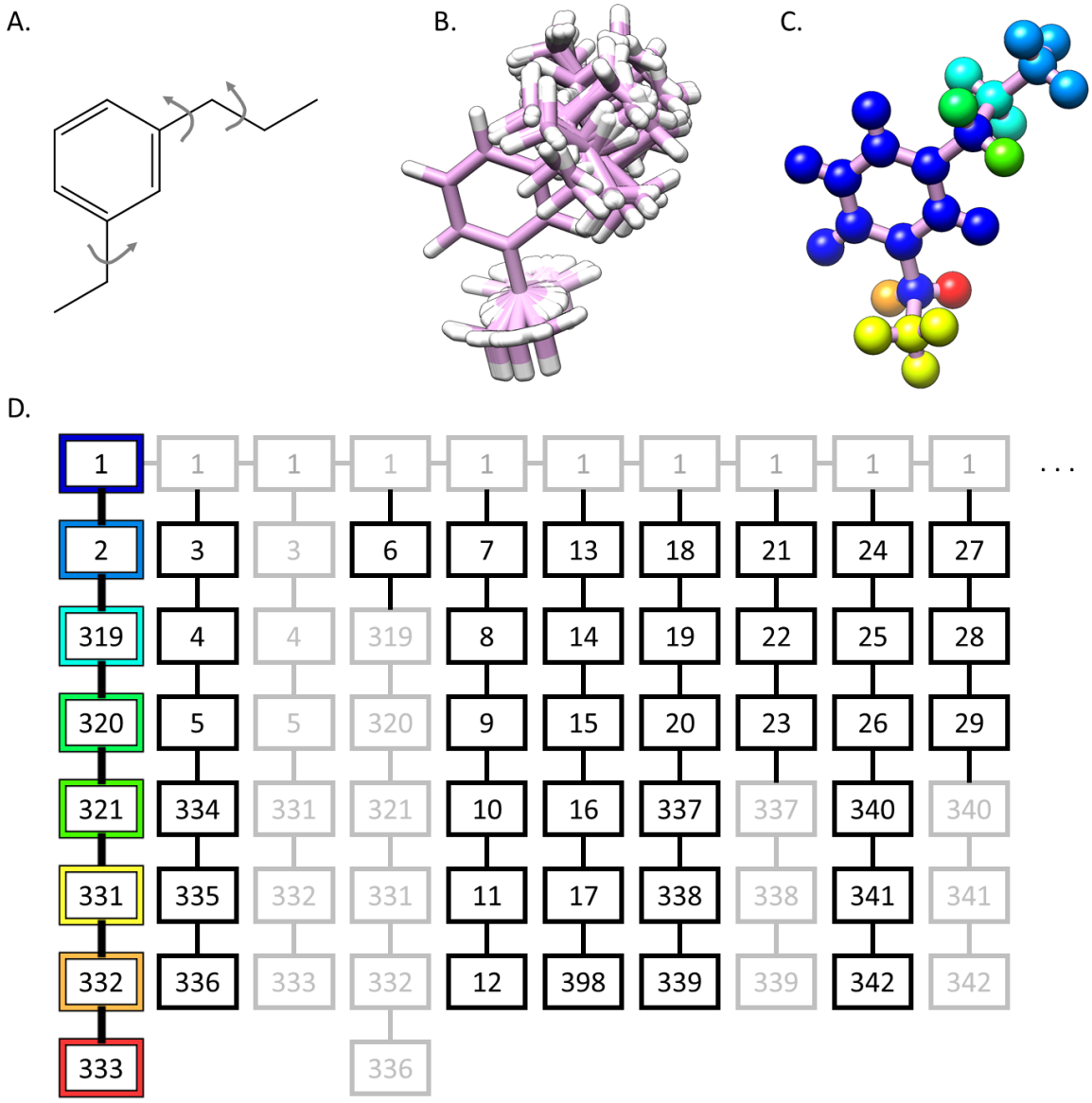


Figure 1. An image highlighting the database format for storing conformations of a molecule. A. Chemical structure. **B.** Image of all the conformations of the molecule. **C.** The first branch is shown as a representative to illustrate the correspondence between nodes and sets of atoms. (Node one is in every branch and is the rigid segment used for orienting). There are 81 branches and 411 nodes for this molecule. The conformational expansion is generated with DOCK 6. When generated with OMEGA, there are 16 branches and 46 nodes. **D.** Network depiction of the database, where the columns show a branch. The nodes are a set of atoms in a defined position (a node can be one atom or a set of atoms). The first 10 branches are shown. The black boxes indicate the first occurrence of the node, and grey boxes indicate repeats of the node.

The database file stores multiple conformations of a ligand in a compressed manner and allows for efficient search.

Unlike the earlier DB format[36] in DOCK 3.5 and DOCK 3.6, the DB2 format, introduced by Coleman *et al.*[34], does not define any part of the molecule as independent from another part of the molecule (**Figure 2**).

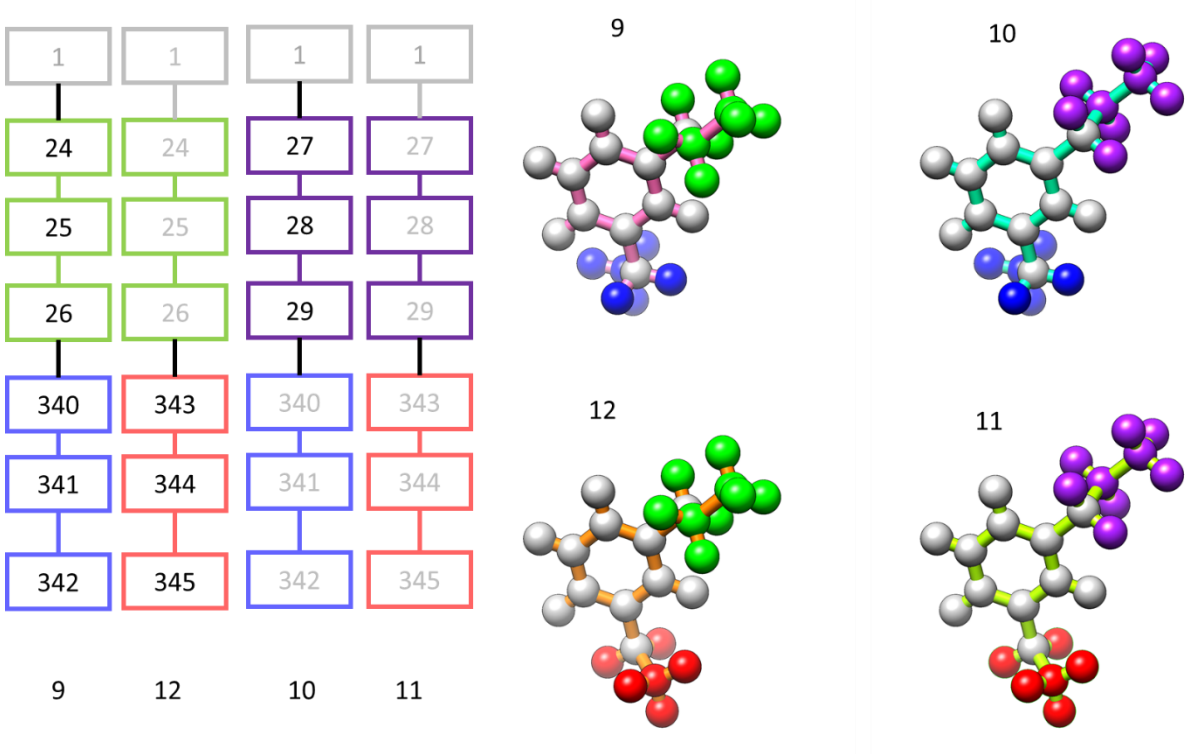


Figure 2. Illustration of 4 conformations stored in the hierarchy with overlapping conformations. Green and purple are two different conformations of one part of the molecule, and blue and red are two conformations of another part of the molecule. Although these conformations appear to be independent, they are treated as dependent on one another.

HDB implementation.

DB2 reader. We implemented a DB2 reader and DB2 data structure into DOCK 6.

DB2 search routine. We implemented a HDB search routine into DOCK 6. When possible, we leveraged existing code. The rigid segment is oriented to the spheres and scored (if the match is incompatible, it is discarded, and the next match is scored). For each set (branch), each segment conformation is evaluated one at a time. First, the segment is oriented, and then it is scored. If the segment is incompatible, it is flagged (so when the segment is encountered in another branch, it is not reevaluated), the search is stopped, and the routine proceeds to the next branch. If segments are compatible, the routine proceeds to the next segment. Once all segments have been evaluated, the routine moves to the next branch. By orienting each segment one-by-one, time is saved. As soon as one segment is determined to clash, that branch is no longer considered, and none of the following segments will need to be aligned into

the pocket. Minimization occurs at the end of the search and adjusts only the rigid 6 degrees of freedom (3 translational and 3 rotational).

There are a few parameters that are worth mentioning. In **Table 1,** the HDB parameters are shown.

Table 1. DOCK parameters that control HDB Search.		
DOCK input (dock.in) Parameters	Description	Value *
conformer_search_type	DOCK 6 has a few different search types. We have added HDB search.	HDB
num_per_search	**Added.** Number of poses kept for each orientation. The HDB hierarchy is oriented, and then the hierarchy is searched, and a user-specified number of poses are kept (here 1 or 10). (If the minimizer is turned on, these poses are all minimized.)	1 (enrichment) 10 (pose reproduction)
skip_broken	**Added.** If atoms within a conformation (a branch in the hierarchy) are too close, the conformation is flagged as broken in the DB2 file. If this parameter is “yes”, these conformations are skipped.	yes
hdb_db2_input_file	**Added.** This parameter is for passing DB2 files. It can take a split_database_index file (as in DOCK 3), which contains a list of db2.gz files, or it can take a single file db2.gz file.	sdi.txt
hdb_db2_search_score_threshold	**Added.** This is the score cutoff for each segment of the branch. If the score-cutoff is exceeded, then the segment is flagged and all branches containing the segment are halted.	10.0
* These parameter values are used unless otherwise stated.		

The orienting parameters and values are described in **Table 2**, which we used to closely match DOCK 3.7 behavior.

Table 2. Parameters that control orienting for DOCK 6.		
Parameter	Description	Value *
automated_matching	There is an automated procedure for matching. If the parameter is set to “yes”, then default parameters are used. We did not use the default, to enable better agreement in orienting behavior between DOCK 3 and DOCK 6.	no
automated_matching_iteration	**Added.** To better match DOCK 3.7 behavior. We start with distance_tolerance and we then multiply it by 1 up to 10, or until max orientations are met.	yes
distance_tolerance	This parameter behaves differently between the two codes. In DOCK 6, distance_tolerance is compared to the sum of the residuals between spheres and atoms. In DOCK 3.7, distance_tolerance is compared to the maximum residual (that is, each residual is compared to the distance_tolerance one at a	0.2

	time; as soon as one is in violation, it is discarded). The behavior of this parameter between DOCK 6 and DOCK 3 seems to differ by a factor of four (which corresponds to the number of *nodes*).	
distance_minimum	This parameter was ignored in earlier versions. It is now used to skip matches containing spheres that are too close. If set to 0.0, the parameter is turned off.	0.0
nodes_minimum	Minimum number of sphere and ligand atoms to match.	4
nodes_maximum	Maximum number of sphere and ligand atoms to match.	4
receptor_site_file	Specifies the sphere file.	dockfiles/matching_spheres.sph
max_orientations	The maximum number of orientations to pass to the "search" step.	1200
score_threshold	Score threshold: only poses with values less than this value will be written to the mol2 file.	100.0
* These parameter values are used unless otherwise stated.		

The DOCK 3.7 orienting parameters are shown in Table 3.

Table 3. Parameters that control orienting for DOCK 3.7.

Parameter	Description	Value*
match_method	There are two choices: "1" will use the number of orientations generated for a single distance tolerance. "2" will iterate until the distance maximum is met, or the match goal is exceeded.	2
distance_tolerance	For match_method 1, this is the distance_tolerance used. For match_method 2, this is the starting point.	0.05
match_goal	Match goal; this is a soft cutoff and is often exceeded.	1000
distance_step	Step size for iterations of distance tolerance.	0.05
distance_maximum	This is the cutoff for permitted matches.	0.5
nodes_maximum	Maximum number of sphere and ligand atoms to match.	4
nodes_minimum	Minimum number of sphere and ligand atoms to match.	4
bump_maximum	This is the score cutoff for each segment of the branch. If the score-cutoff is exceeded, then the segment is flagged and all branches containing the segment are halted. (Corresponds to the *hdb_db2_search_score_threshold*)	10.0
bump_rigid	This is the score cutoff for the rigid segment	10.0
mol2_score_maximum	Score threshold: only poses with values less than this value will be written to the mol2 file. (Corresponds to the *score_threshold*)	-10.0
* These parameter values are used unless otherwise stated.		

Differences between DOCK 6 and DOCK 3 implementations.

The behavior of DOCK 6 and DOCK 3 differ in a few key aspects, particularly in orienting. DOCK 6 has a hard cap on the number of orientations. DOCK 3.7 has a soft cap (the match_goal parameter), resulting in DOCK 3 generating a wide range of orientations. They also differ in minimization. Both DOCK 3.7 and DOCK 6 use a simplex minimizer. However, the implementations are different, and the minimizer itself is stochastic in nature. This is a source of variability in the results. In DOCK 3.7, only the top pose (or a select number of top-scoring poses) from each dock run are minimized, while in DOCK 6 HDB, one pose (or a select number) from each orientation is minimized. By resurrecting the *final_min* parameter, we have added the ability within DOCK 6 to only minimize the user-specified number of poses on the aggregate poses from all orientations (as is done in DOCK 3). However, this is not much faster, and the results are not as good. If the *final_min* parameter is turned on, poses are minimized immediately before they are written out. (It appears that only a small amount of time is spent in the minimizer when comparing the runtime with and without minimization.)

Scoring functions and ChemGrid update.

The default scoring function in DOCK 6 is Grid Score, which has two components: $E_{Grid} = E_{VDW} + E_{ES}$. For Grid Score, the van der Waals grids were calculated with an all-atom representation of the protein using 6-9 exponents. Electrostatic grids were calculated using Coulomb's law with a distance-dependent dielectric, $\epsilon(r) = 4r$.

DOCK 6 has several other scoring functions[37-39], including ChemGrid score, which has three components: $E_{ChemGrid} = E_{VDW} + E_{ES} + E_{lig,desolv}$. For ChemGrid Score, the van der Waals grids were calculated with a united-atom representation of the protein and 6-12 exponents. Electrostatic grids were calculated using a Poisson–Boltzmann solver[34,40-42]. The ligand desolvation component was calculated using a solvation grid (which precomputes the solvation cost of placing an atom at a specific location), and the precomputed solvation values for each atom of the ligand.[43,44]

We updated the *ChemGrid Score* (also called *DOCK 3.5* score) scoring function to match the behavior in DOCK 3.7. The program *QNIFFT*[40,41] is now used to calculate the electrostatic energies (replacing Delphi). We updated the Solvmap code to be consistent with DOCK 3.7/3.8. To get the two scoring functions to match when performing single point calculations, we added the parameter file *$DOCK6PATH/parameters/vdw_AMBER_parm94.dock3_7.defn*. This file matches the vdW parameters used by DOCK 3.7. We have updated the *grid-convert* command as well. The function will work with grids generated with *blastermaster.py* or generated with the DOCK 6 suite. The parameter *DOCK3_7_grids* controls whether one is using the DOCK 3 or DOCK 6 grids. With these changes, the DOCK3.7 score function and DOCK6 ChemGrid Score show strong agreement (See SI **Section S1** and **Figure S1**).

Conformation generation.

We use DOCK 6 as a conformation generation tool. We modified DOCK 6 in the following two ways: (**1**) We implemented a repulsive van der Waals as a standalone scoring function. (**2**) We write the anchor atom names in the dock.out file (this allows us to choose the anchors in subsequent steps). These code modifications were released in DOCK 6.10[39,45]. We run DOCK 6

specifying each anchor to perform the conformational expansion about the rigid segment (usually a ring). We have written shell scripts that generate DB2 files using DOCK 6 to codify this procedure. Details of DB2 generation are discussed in the **Database generation** section below. The shell script that wraps DOCK 6 for DB2 building will be distributed with a future version of DOCK 6.

Enabling rescoring, torsion minimization and optimization.

To enable rescoring with *ChemGrid Score*, we added solvation writing to the mol2 file—this will just reproduce the solvation atomic information that is in the input *db2* or *mol2* file. Solvation writing can be turned on or off with the parameter *write_mol_solvation*. If the user wants to rescore, minimize, or perform hydrogen optimization, they can now do so by turning on solvent writing (*write_mol_solvation yes*) while docking, and turn on solvent reading in the rescoring step. Performing a single point calculation in DOCK 3 is much more difficult, and the torsion minimization of a docked pose or a crystallographic pose is not possible, nor is hydrogen optimization. Furthermore, solvation is not written out in DOCK 3, so getting the solvation parameters is also a hurdle.

A single point calculation is performed by using the parameters listed in **Table 4**:

Table 4. Parameters for a single point calculation.	
conformer_search_type	rigid
orient_ligand	no
minimize_ligand	no

This single point score evaluation will just calculate the energy for a crystallographic or docked pose without moving the molecule.

A minimization calculation is performed by using the parameters listed in **Table 5**:

Table 5. Parameters for torsion minimization calculation.	
conformer_search_type	rigid
orient_ligand	no
minimize_ligand	yes

This will perform simplex minimization on a pose by adjusting the 3 translational, 3 rotational, and N internal degrees of freedom. It will only perform local movements.

Hydrogen optimization is performed with the parameter choices listed in **Table 6**:

Table 6. Parameters for hydrogen optimization calculation.	
conformer_search_type	flex
orient_ligand	no
minimize_ligand	yes
flex_defn_file	$DOCK6PATH/parameters/flex_just_OH.defn

This set of parameters will perform hydrogen optimization. By using conformer_search_type *flex*, we will access the growth part of the anchor-and-grow algorithm, providing a modified flex definition file, *flex_just_OH.defn*, where only hydroxyls or thiols are defined as flexible. These files will be made available in a future release of DOCK 6.

The HDB routine searches through pre-generated conformations, exploring torsions, but it does not adjust them. However, in DOCK 6, we can adjust the torsions of the molecules to the protein environment by performing a subsequent torsion minimization or hydrogen optimization experiment—this is an ability that DOCK 6 has that DOCK 3 does **not** have.

When attempting to perform minimization after docking, writing out multiple poses during docking is helpful. By giving the minimizer multiple starting poses, we may find that the most favorable minimized pose may originate from a starting pose that is not the lowest scored pose (prior to minimization). DOCK 6 has a *best first clustering* procedure, which uses an RMSD metric to ensure that the poses written out are different from one another. One issue with DOCK 3 is that poses are often very similar when writing out multiple poses, which is not an issue with DOCK 6 because of this clustering procedure. We deploy this "best first" method in docking, prior to minimization and optimization experiments. We use a threshold value, *cluster_rmsd_threshold*, of 1.0. Clustering is used in virtual screens and enrichment calculations but is not used in the pose reproduction calculations (all poses are used and are not clustered).

Enrichment calculations with DUDE-Z.

We tested the new code with the DUDE-Z database.[46] For all DUDE-Z systems, we downloaded rec.crg.pdb and xtal-lig.pdb from the web page[47]. We constructed the docking grids after missing sidechains were added in UCSF Chimera[48] using Dock Prep. We built the grids with the standard charges for residues (we do not polarize any residues). We also downloaded all molecules—the active molecules, the property-matched decoys, and the extrema decoys—from the web[47].

Enrichment. We used the scripts provided with DOCK3.7, *enrich.py* and *plot.py*, to calculate the logAUC and AUC values.

Database generation.

In our database generation for pose reproduction, we start with molecules in SMILES format. For each molecule, we perform the following five steps: **I.** We use *cxcalc* (version 19.18.0) and *molconvert* (version 19.18.0) from *ChemAxon* (https://www.chemaxon.com) to calculate protonation and tautomerization states of the molecule, to determine which titratable atoms have positive or negative charge. **II.** For each protomer/tautomer of the molecule, *Corina Classic*[49] is then used for generating a 3D conformation (including ring puckering). **III.** *AMSOL 7.1*[50] is used to calculate partial charges using *AM1-BCC* theory, and to calculate per-atom decomposition of solvation energies using a Generalized Born (GB) solvation model. *RDkit*[51] is used to calculate the formal charge of the molecule before running AMSOL. **IV.** DOCK 6 is used to perform the conformational expansion. **V.** Using the *mol2db2.py* program distributed with DOCK 3, the desolvation parameters, the partial charges, and the conformational ensemble are combined into hierarchal database DB2 files (one file for each ring of the molecule). The pipeline is scripted and will be made available in a future release of DOCK 6. The pipeline described here is similar to that described previously[5,52], with a few key replacements. OpenEye Omega[53] is replaced with DOCK 6 for the conformational expansion. OpenEye was replaced with RdKit to calculate the formal charge of the SMILES string. Enrichment[47] and virtual

screening[54] databases were downloaded from the web. These downloaded DB2 files use this previous pipeline [5,52].

SB2012 database for pose reproduction.

We use the SB2012 database[55] for pose reproduction. We downloaded the receptor and ligand mol2 files from the web[56]. We generated databases in two ways: **1.** We got the SMILES strings from the Protein Databank when possible, or by converting the mol2 to a SMILES string if the ligand consisted of multiple names using *obabel* from Open Babel 2.3.1[57]. SMILES strings are listed in SI (see **Section S2** and **Table S1**). We then built hierarchies using our pipeline from the SMILES strings. This pipeline consists of five steps, described above in the **Database generation** section. **2.** We also generated ligand databases from the mol2 file provided with SB2012. We did this by skipping the first two steps and starting with step **III**, where we calculated the formal charge by summing up the partial charges in the mol2, and then running *AMSOL 7.1*. The steps **IV** and **V** are the same as above.

Pose reproduction outcomes.

In a pose reproduction experiment, we define three outcomes: *docking success*, *sampling failure*, or *scoring failure*.[55] We use a RMSD threshold of 2.0 Å to define these outcomes. *Docking success* is when the best scoring molecule has an RMSD below the threshold. *Scoring failure* is when the best scoring molecule has a RMSD above the threshold, but there exists a pose with a RMSD below the threshold. *Sampling failure* is when none of the poses has a RMSD value below the threshold. Here, we are keeping 10 poses per molecule for each docking run.

3. Results.

Enrichment calculations.

First, we tested the code using the DUDE-Z test set to quantify the ability to enrich known binders from a background of decoys. We compare DOCK 3.7 with DOCK 6.9. In comparing DOCK 3 and DOCK 6, we quantified the enrichment in two ways: using the AUC of the ROC curve (**Figure 3**, top left panel), and by using log-adjusted AUC values to weight early enrichment more (**Figure 3**, top right). We see that a few of the DUDE-Z systems, like ACES, preferred DOCK 6, while other systems, like HSP90, had virtually no difference in enrichments. Most systems, like PUR2 or FGFR1, preferred DOCK 3 over DOCK 6. We see that the two methods are performing comparably (avg. AUC: 65.39 to 64.46 and avg. logAUC: 16.57->15.43), although DOCK 3.7 seems to perform on average better (ΔAUC = 0.90, ΔlogAUC= 1.11, **Table 7**). We compared timing **(Figure 3**, **middle left**) and the ratio (**middle right**). DOCK 3.7 is on average faster (Ratio = 1.66, **Table 8**). In addition, we looked at the distribution of scores (in kcal/mol) produced by all molecules (ligands and decoys) across the systems (**Figure 3**, bottom left). We see that the two methods result in similar behavior. We also looked at the number of orientations generated during the docking and this underscores a behavioral difference between DOCK 6 HDB and DOCK 3.7 (see **Methods** for the matching procedure differences). DOCK 3 has poses with many more orientations than DOCK 6. For example, UROK has over 10,000 orientations for the pose with the maximum number. However, on average, it produces about the same number, around 1200. There are many parameter

combinations that might increase enrichment or enable better timing; however, we endeavored to use parameters that would facilitate a direct comparison of the two DOCK versions. Specifically, we attempted to match the average number of orientations and the energy distributions.

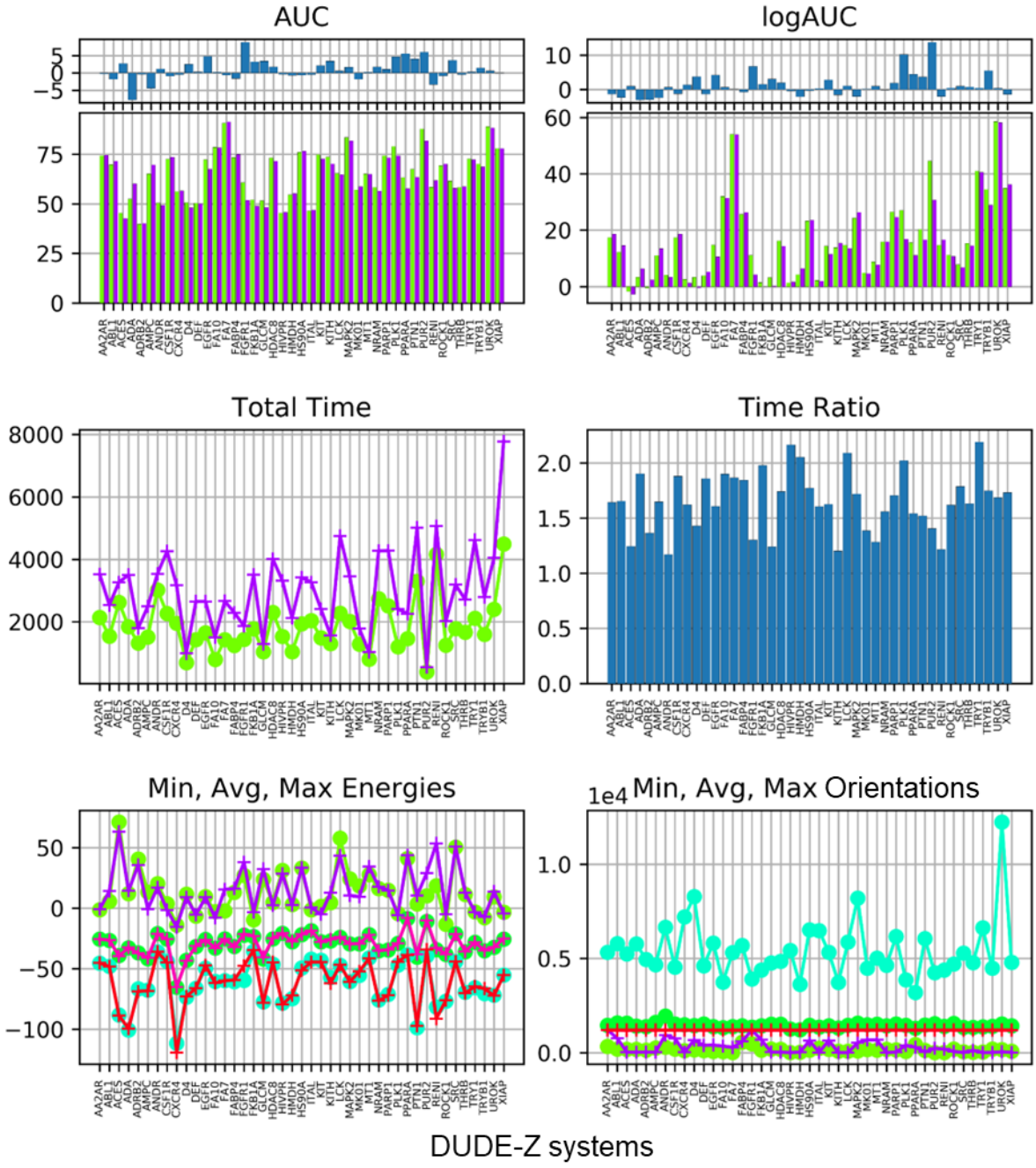


Figure 3. Comparing DOCK 3.7 with modified DOCK 6.9 using 42 DUDE-Z systems. Ligand and property-matched decoys downloaded from web are used. **Top panels**: A comparison quantifying enrichment using AUC (**left**) and logAUC (**right**), DOCK 3.7 (green), DOCK 6.9 (purple), difference (Blue). **Middle panels:** Comparison of timings of DOCK 3.7 (green), DOCK 6.9 dev (purple), ratio (Blue). **Bottom Panel**: Energy distribution (**left**) and

orientations (**right**) are shown. The plus symbols in red, pink, and purple, and circle symbols in shades of green, represent the maximum, average, and minimum values for each system.

Table 7. Calculated enrichment quantification comparing DOCK 6 with DOCK 3 and with minimization and optimization protocols.

	AUC	logAUC	ΔAUC	ΔlogAUC
DOCK 6 [a,b]	64.46	15.43	0.0	0.0
DOCK 3.7 [a]	65.39	16.57	0.93	1.14
DOCK 6_max_orient_3600	67.39	17.89	2.93	2.46
DOCK 6 kept10[c]	65.35	16.07	0.89	0.64
DOCK 6 min	64.98	16.08	0.52	0.65
DOCK 6 hopt	64.91	15.94	0.45	0.51
DOCK 6 hopt after min	65.26	16.34	0.80	0.91
DOCK 6 min after hopt	65.25	16.26	0.79	0.83
DOCK 6 min10[d]	66.05	16.78	1.59	1.35

a. rec.crg.pdb is used to generate grids. Chimera Dock Prep is used to build missing sidechains.
b. Reference for delta values.
c. 10 poses are kept during search routine (no torsion minimization).
d. 10 poses are kept during search routine, and 10 are torsion minimized.

Table 8. Timing Ratio between DOCK 3 and DOCK 6.

Min time ratio	Avg time ratio	Max time ratio
1.17	1.66	2.19
Chimera is used to build missing side chains; default charges are used.		

There are many parameters that will impact sampling (see **Tables 1-3)**. These include the max_orientations and num_per_search parameters. We increased max_orientations from 1200 to 3600 (**Table 7**), which resulted in an improvement in enrichment of ΔAUC = 2.93 and ΔlogAUC = 2.46, but also resulted in an increase in runtime (2.8 times slower). This change allowed DOCK 6 to surpass DOCK 3.7 in enrichment performance (ΔAUC = 2.0 and ΔlogAUC =1.32). We increased num_per_search from 1 to 10 (**Table 7**) and observed an increase in enrichment (ΔAUC = 0.89 and ΔlogAUC 0.64) with a subtle increase in time (1.1 times slower). See SI **Section S3** and **Figure S2** for AUC, logAUC, and energy distribution plots for the 42 DUDE-Z systems, expanding on the optimization and minimization shown in **Table 7**.

Minimization effect on enrichment.

One advantage that DOCK 6 has over DOCK 3 is the ability to rescore molecules. We can read the molecules back in, minimize, and rescore them. During minimization, DOCK will adjust the torsion angles.

Using DOCK 6, docking followed by torsion minimization and hydrogen optimization (hopt_after_min) results in an AUC and logAUC of 65.26 and 16.34, respectively (**Table 7**).

When comparing the enrichment of the optimized pose to DOCK 3.7, DOCK 6 approaches DOCK 3 performance: a ΔAUC of -0.13 and a ΔlogAUC -0.23 is observed. We can see the minimization has an impact on the scores for PTN1 and CXCR4, which receive a more favorable minimum score after minimization (on average, minimization does not have a major impact). If we keep 10 poses (fewer after clustering), then we see DOCK 6 slightly surpasses DOCK 3: a ΔAUC of 0.66 and a ΔlogAUC of 0.21 is observed.

The decoy background has a profound effect on the enrichment quantification. Surprisingly, when the property-matched decoys are used, Grid Score outperforms ChemGrid Score (ΔAUC = 4.34 and ΔlogAUC = 1.17). However, for the extrema decoy background, ChemGrid Score outperforms Grid Score (ΔAUC = -7.26, ΔlogAUC = -7.92). For extrema, there is just one system, PUR2, where Grid Score does substantially better than ChemGrid score (see below for more discussion).

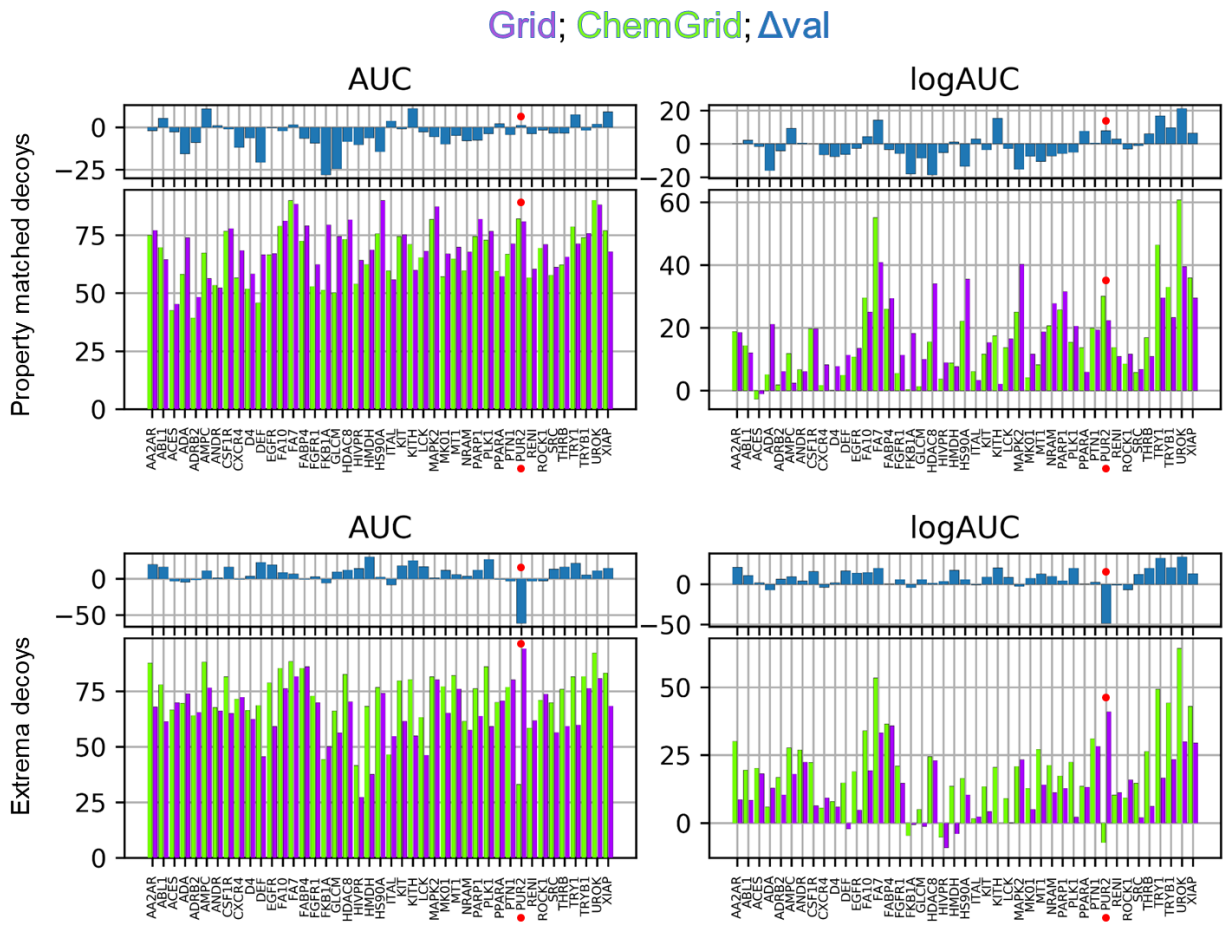


Figure 5. Comparison of enrichment for two different scoring functions. ChemGrid (green) score vs. Grid Score (purple) using Property Matched (top panel) and Extrema (bottom) decoy backgrounds. Red dots indicate PUR2, which is discussed further in the text.

Table 9. Comparison of using ChemGrid Score and Grid Score, averaged over DUDE-Z set.				
	AUC (pmd)	logAUC (pmd)	AUC (extrema)	logAUC (extrema)
Grid Score	69.91	17.30	65.52	12.48
ChemGrid Score	65.57	16.13	72.78	20.40
Δ enrichment value	4.34	1.17	-7.26	-7.92
Skip_broken no.				

PUR2 demands investigation (**Figure 5**, red dots). When we compare ChemGrid Score and Grid Score enrichments with the extrema decoy background, we see that the enrichments for the PUR2 system are clear outliers. Grid Score enrichment is significantly better than ChemGrid Score enrichment: -7.56 vs. 38.48 logAUC units (ΔlogAUC = 46.04) and 31.56 vs. 93.64 AUC units (ΔAUC = 62.08) (**see Figure 5**). When investigating the binding modes, Grid Score seems to produce better poses, where the diaminopyrimidinone ring is placed in the channel in a similar way to the crystal ligand (1NJS[58]) (see **Figures 6 A-D)**. Looking at the molecular properties of the top scoring molecules, we observe that the two scoring functions enrich for molecules with different charge distributions (**Figure 6E**). ChemGrid Score has a clear charge bias for +1 and +2 molecules, while Grid Score has little or no preference for charge (**Figure 6E**). Since all the ligands are anions (-2 and -3), we would ideally like to enrich for negatively charged molecules. For Grid Score, the vdW energy is the dominant component. Specifically, for the top 500 poses, the vdW component is ~6 times the value of the electrostatic component. We rescore Grid Score poses with the ChemGrid scoring function, and conversely, we rescore ChemGrid Score poses with Grid Score (see **Figures 6 F and G**)—which is only possible in DOCK 6. The rescoring calculation uses the single point parameters (**Table 4**) discussed in more detail in **Methods**. We observe that when the poses from Grid Score are rescored in ChemGrid Score, the desolvation energies are substantial, and prevent the poses from scoring well with ChemGrid Score. When we turn off the ligand desolvation component for ChemGrid Score, some of the poses produced resemble those from Grid Score. ChemGrid places some of the ligands in the channel (**Figure 6 D**), and enrichment improves (6.93 logAUC units and 71.41 AUC units). However, enrichment is still not as good as Grid Score, and ChemGrid Score is still enriching for +1 and +2 molecules. When the ChemGrid Score poses with ligand desolvation turned off are rescored with Grid Score, the ligand poses score well with Grid Score, and enrichment is comparable to when Grid Score is used as the pose-generating scoring function (34.11 logAUC units and 85.80 AUC units). When investigating the differences between the original ChemGrid scores and the Grid rescores, the ChemGrid vdW component appears lower in magnitude than the Grid-rescored vdW component. The electrostatic component is more variable. In many cases, the ChemGrid electrostatic component is lower when compared to the Grid-rescored electrostatic component.

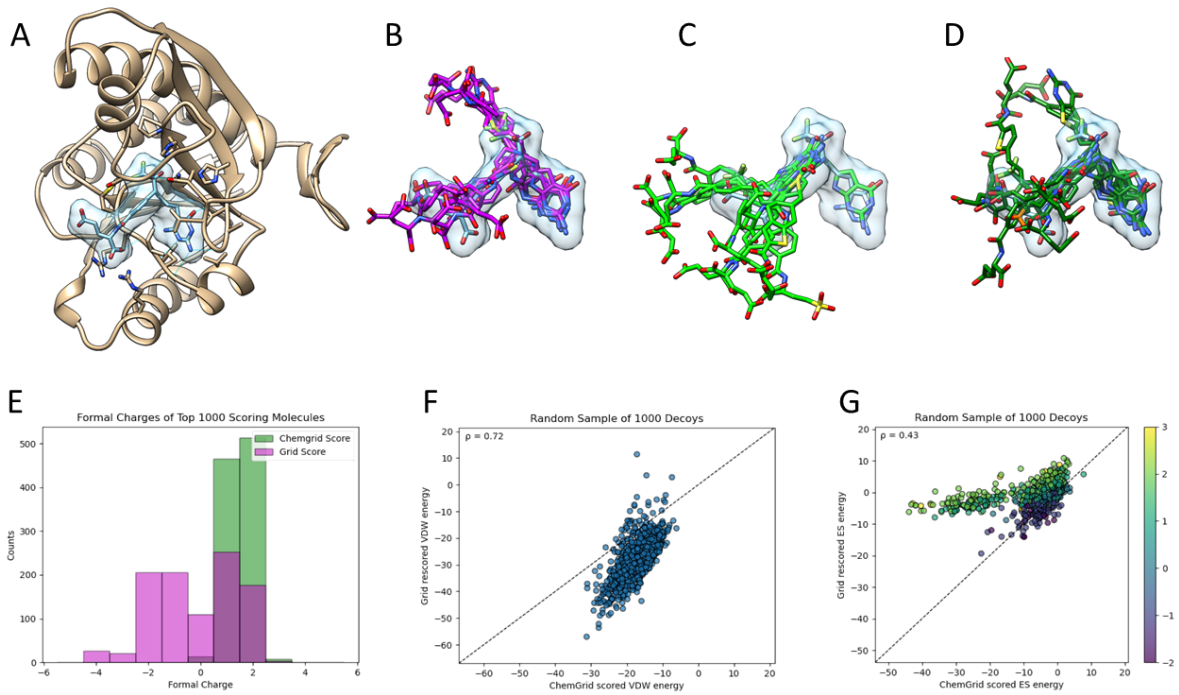


Figure 6. DOCK results for PUR2, comparing Grid Score and ChemGrid Score. A. The PUR2 crystallographic ligand and protein (PDB ID: 1NJS), with the ligand surface shown. **B.** Top ten scoring ligands with Grid Score, overlaid with the crystallographic ligand. **C.** Top ten scoring ligands with ChemGrid Score, overlaid with the crystallographic ligand. **D.** Top ten scoring ligands with ChemGrid Score, with ligand desolvation turned off, overlaid with the crystallographic ligand. **E.** Plot of the charge distributions for the top scoring 1000 molecules with Grid Score (magenta) and ChemGrid Score (green). **F.** Plot of the correlation between the ChemGrid scored vdW energies and the Grid rescored vdW energies for a random sample of 1000 decoys (Pearson correlation coefficient = 0.72). $y = x$ is the dashed line. **G.** Plot of the correlation between the ChemGrid scored electrostatic (ES) energies and the Grid rescored ES energies for the same random sample of 1000 decoys (Pearson correlation coefficient = 0.43). Color of points shows the formal charge of the ligands. $y = x$ is the dashed line.

The correlation between ChemGrid electrostatic energies and Grid-rescored electrostatic energies changes based on the charges of the molecules (**Figure 6G**). For negatively charged molecules (ligands and decoys), the magnitudes of the energies are similar (clustered around y=x). For positively charged molecules (all decoys), the ChemGrid electrostatic energies are almost always lower than the corresponding Grid rescored electrostatic energies. The lower electrostatic component results in positively charged decoys scoring better and ranking higher in the original ChemGrid score than in the Grid rescore, which explains the worse enrichment for ChemGrid. We speculate that the difference in vdW component is due to the difference between United Atom (ChemGrid) and All Atom (Grid) (**Figure 6F**). For Grid Score, we use the softened repulsive term, while in ChemGrid, we use the standard term, so the difference between 6-9 vs. 6-12 may be playing a role in the different behavior as well. We speculate the difference in electrostatic energies is due to the difference between Poisson-Boltzmann electrostatic potential grids and Coulomb's law with a distance-dependent dielectric. The distance-dependent

dielectric will squash the charge-charge interactions more quickly than the charge-charge interactions calculated in the evacuated pocket.

Pose reproduction.

We tested the HDB search method with pose reproduction computational experiments over the SB2012 test set. For all pose reproduction experiments, we use Grid Score. To do this, we first generated DB2 files. We do this in two ways: starting from SMILES strings, or from mol2 files (see **Methods** for more details). The *from_smi* experiment (**Table 10** and **Figure 7**) is the most consistent with our enrichment calculations and virtual screening procedures. The *from_smi* experiment has the poorest DOCK performance, with 52% success. When we minimize the *from_smi* poses allowing the torsions to adjust, the *docking success* rate increases to 53%. For the *from_mol2* experiment, we do not lose all the crystallographic information: we retain the bond lengths and angles, but dihedral angles are explored without consideration of the starting experimentally determined angles. For the *from_mol2* experiment, we observe that 60.02% of the systems are *docking successes*, and this number increases to 62.13% upon torsion minimization. We explored a few optimization procedures: hydrogen optimization (*hopt*), hydrogen optimization after minimization (*hopt_after_min*), and minimization after hydrogen optimization (*min_after_hopt*). Of these methods, *hopt_after_min* seems to perform the best, consistent with the enrichment results, but has only marginal improvement over minimization alone (62.13% compared to 62.32%, 0.19% improvement). The *from_mol2* experiment is more comparable to the anchor-and-grow data. Compared to *from_mol2*, the anchor-and-grow method has a better *docking success* rate of 76.03% (16.01% better or 1.27 times better); however, it has a significantly slower average time, 70.97 seconds compared to 4.29 seconds—16.5 times slower. During DB2 generation, we can increase ligand internal sampling and store more conformations of the ligand. In the *more_sampling* experiment, we increased the conformational goal from 300 to 1000. For *more_sampling*, we see an increase of success from 60.02% to 62.80% (a change of 2.78%). For the *seed* experiment, we seeded the more_sampling conformation databases with the crystallographic conformation, and 91.47% of the systems are *docking successes*. This number only changes marginally upon minimization (0.48%). These torsion minimization and optimization procedures are only possible in the new DOCK 6 implementation.

Table 10. Docking statistics for pose reproduction docking and minimization					
	Docking Success (%)	Scoring Failure (%)	Sampling Failure (%)	Did Not Dock (%)	Timing (s)
anchor-and-grow	76.03	7.48	15.82	0.67	70.97
From_smi	52.35	9.11	30.58	7.96	
From_smi_min	53.50	8.72	29.82	7.96	
From_mol2	60.02	6.71	30.11	3.16	4.29
From_mol2_min	62.13	5.27	29.43	3.16	
From_mol2_hopt	60.88	6.04	29.91	3.16	
From_mol2_hopt_after_min	62.32	5.27	29.24	3.16	
From_mol2_min_after_hopt	62.03	5.66	29.15	3.16	
more_sampling	62.80	7.67	26.94	2.59	8.99
seed	91.47	4.79	1.44	2.30	
seed_min	91.95	4.51	1.25	2.30	
Δ from_smi	1.15	-0.38	-0.77	0.00	
Δ from_mol2	2.11	-1.44	-0.67	0.00	
Δ seed	0.48	-0.29	-0.19	0.00	
N = 1043, Δ is the comparison between the corresponding min and no min values.					

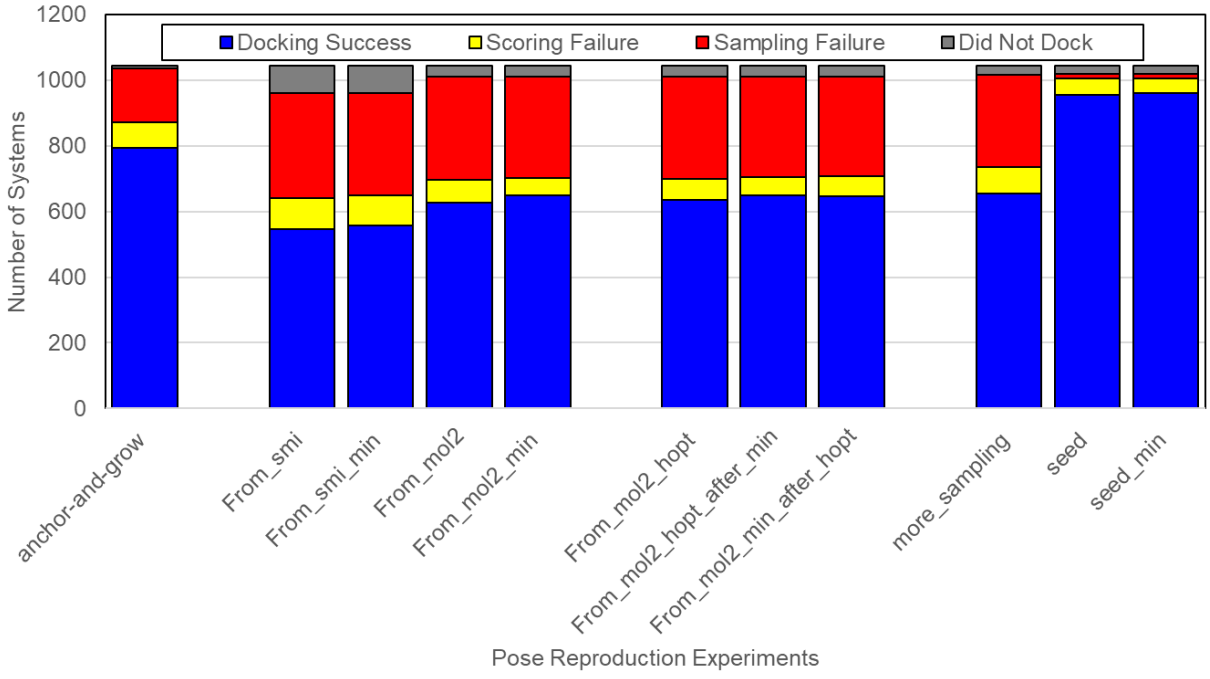


Figure 7. Stacked bar graph for pose reproduction results for anchor-and-grow and 10 HDB docking experiments. Blue, yellow, and red indicate *docking successes*, *scoring failures*,

and *sampling failures*, respectively, the three possible outcomes of a docking experiment. Grey indicates the number of systems that *failed to dock*.

The DB2 search routine is 16 times faster than anchor-and-grow. However, the time to construct the database takes on average 92 seconds. If one adds the 4 seconds for docking, that is 96 seconds total, compared to the 71 seconds for anchor-and-grow. Thus, to build and then DOCK HDB, it is 15 seconds slower than anchor-and-grow. Most of this time is spent in the Python program *mol2todb2.py* (this program has been somewhat sped up in DOCK v3.8). We are also considering all rings as rigid starting points, not just 3 rings, as is done in the ZINC (or DUDE-Z) pipeline[5,52]. It is worth this upfront cost, because for virtual screening, the cost is paid once, and databases are stored to perform many screening calculations against many targets.

Table 11. Classification for pose reproduction for number of systems that change or stay the same before and after minimization for the *from_smi* experiment.

Before \ After	Docking Success	Scoring Failure	Sampling Failure
Docking Success	533	13	0
Scoring Failure	21	73	1
Sampling Failure	4	5	310

For the *from_smi* pose reproduction, after minimization, 21 systems moved from a *scoring failure* to a *docking success*, and 13 systems move from a *docking success* to a *scoring failure* (a net change of 8 systems) (see **Table 11**). Four systems move from *sampling failures* to *docking successes*, and 0 move from docking successes to sampling failures. Five systems move from sampling failures to scoring failures, and 1 moves from *scoring failure* to *sampling failure*. The net gain of 8 + 4 = 12 systems become *docking successes*.

To understand how minimization results in improvements, we investigate the systems that swap classifications. Protein Tyrosine Phosphatase-1B (2BGE) is one example of a system that swaps upon minimization from *sampling failure* to *docking success* **(Figure 8)**. For this system, we can see the poses before minimization and the poses after minimization, and the corresponding RMSDs and scores. We see that the lowest scoring pose before minimization (black diamond, **Figures 8A, C**) and the corresponding pose after minimization (**Figure 8B**) does not resemble the crystallographic pose (**Figure 8F)**. The best RMSD before minimization (black square, **Figures 8D, E**) gets a better energy, surpassing the before-minimization-best-scoring poses (and its corresponding pose), and it gets closer to the crystallographic pose, but it is not the best pose for either RMSD or score. The same pose has the best RMSD and score after minimization (magenta triangle and the cyan upside-down triangle, **Figures 8G, H**). Its original pose is unfavorable and has a score of +28.62 kcal/mol, but after minimization, it becomes quite favorable, with a score of -64.10. The ligand has one dihedral angle which adjusts from -0.29 to -63.76 degrees. This conformational change is seen in the overlay, and the insert focuses on the dihedral (**Figure 8I**). Plots for the systems that swap classifications are shown and discussed in the SI (**Section S4** and **Figure S3**). The best RMSD and energy poses do not always initiate from the poses that hold those classifications before minimization. Although the swaps are sometimes simply poses shifting slightly over the threshold, more interesting are the systems, like 2BGE, with more dramatic movements and the transfer of label from one pose to another. This illustrates the usefulness of keeping multiple poses, particularly for rescoring and minimization.

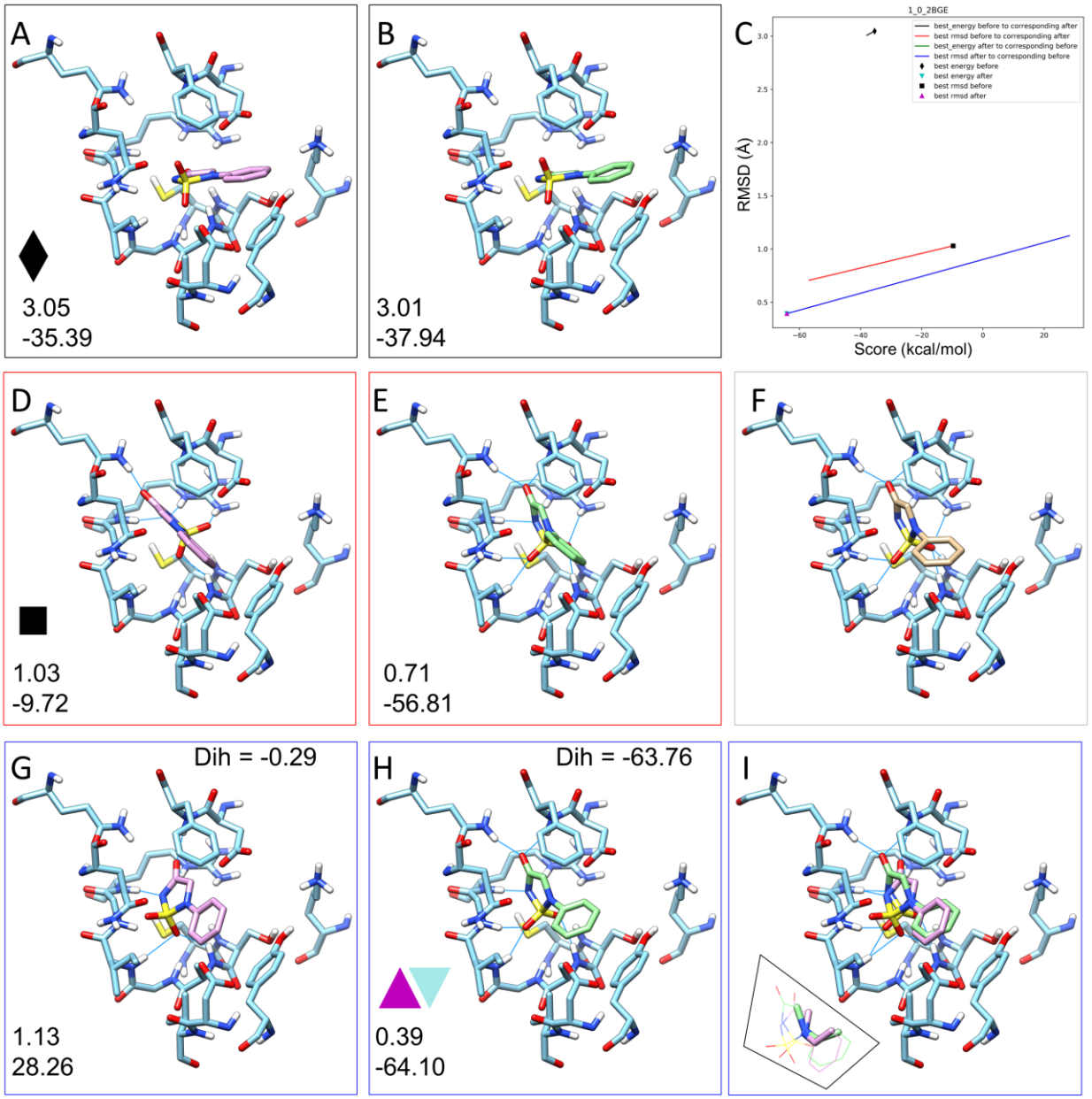


Figure 8. Protein Tyrosine Phosphatase-1B (2BGE) is an example of a system that switches from Sampling Failure to Docking Success. A. Best scoring pose from docking. **B.** torsion minimization of the pose shown in panel A. **C.** Plot of RMSD vs. DOCK Score. Four pairs of poses plotted: black diamond shows the best docked pose, and black square represents the best RMSD docked pose; the magenta triangle and the cyan upside-down triangle are the minimized best score and best RMSD poses, respectively. Lines connect to points before and after minimization: black line connects best score before minimization to the corresponding score after minimization, red connects best RMSD before minimization to the corresponding RMSD after minimization, green connects best score after minimization to the corresponding score before minimization, and blue connects best RMSD after minimization to the corresponding RMSD before minimization. **D.** Pose before minimization, minimum RMSD (red).

E. Pose after minimization (red line). **F.** Crystallographic pose. **G.** The pose before minimization. **H.** Both best scoring and best RMSD after minimization. **I.** Overlay of the poses. The trapezoidal inset showing the dihedral angle. RMSDs and scores are shown in the lower left corner in panels A,B,D,E,G,H.

Example virtual screen.

We performed an illustrative screen to a pocket on the KRAS protein. Here, we used PDB ID 6GJ8[59], which is a complex of the small molecule BI-2852 bound to KRAS G12D with the cofactor GMPPCP (which we convert to GTP for our docking). In this screen, we docked to a monomer of KRAS and to the pocket occupied by small molecule BI-2852.

Table 12. Information about the example virtual screen: Scores[a], timings and number of orientations.

System	Max Score	Avg Score	Min Score	Time[b]	Min Orient.	Avg Orient.	Max Orient.
DOCK 3.7	-46.56	-47.85	-53.52	7871186	12	1209.08	5539
DOCK 6_HDB	-46.69	-47.98	-53.56	20111511	4	970.93	1000
Re-docking[c]	-47.20	-48.57	-55.77	113550	16	1443.76	1500
Torsion Min[c]	-49.51	-51.08	-59.42	1976	NA	NA	NA

a. Score ranges are for the top 1000 poses. Scores are in kcal/mol.
b. Not run back-to-back on the same machine.
c. Number of molecules rescored is 51003, with a score cutoff of -40.00 kcal/mol. 10 poses are kept for each molecule and minimized (num_per_search = 10).

Over the screen, DOCK 6 is 2.56 times slower than DOCK 3. This is a higher ratio than those observed in the enrichment calculation, where the max was 2.36 times slower. However, this time difference seems to be system dependent. Unlike the enrichment calculations, the virtual screen results are not run back-to-back on the same machine, so we have less confidence in the time comparison. However, DOCK 6 is clearly slower. The score distributions are similar when comparing DOCK 6 with DOCK 3 (**Table 12**). DOCK 6 generates fewer orientations, because of the cap and because of differences in the *distance_tolerance* parameter calculation (see methods). With re-docking and minimization, only possible in DOCK 6, of the top 51K molecules from the original docking, we see that the scores get better. For example, we see the best score value becomes more favorable, -53.56 to -55.77 to - 59.42, as we compare DOCK 6 HDB, redocked (with more sampling), and torsion minimized runs, respectively (**Table 12**).

Table 13. Formal distribution of the formal charges.

	Top 1000 poses				Top 10 poses		
Formal_Charge:	Ori[a]	More[b]	Min[c]		Ori[a]	More[b]	Min[c]
0	1	2	3		0	0	0
1	802	780	705		8	5	3
2	196	217	291		2	5	7
3	1	1	1		0	0	0

a. DOCK 6 Screen.
b. DOCK 6 re-docking of the top ca. 51K poses, with 10 poses kept with more orienting.
c. The torsion minimization of the re-docking poses.

This pocket, with default grid generation, shows a charge preference for +1, which is the dominant charge over the top 1000 poses, and +2 charges across all sampling schemas (the

original docking, the re-docking with more sampling, and the torsion minimization) (see **Table 13**). With re-docking and minimization, there is a shift towards +2 molecules. For the top 1000, we see a shift after minimization. We see that +1 goes from 780 to 705 and +2 goes from 217 to 291, but +1 remains dominant. For the top 10 poses, the shift after minimization is clearer. Comparing before to after-minimization for +1 and +2 charges, we see that +1 goes from 5 to 3 and +2 goes from 5 to 7, and +2 becomes dominant.

Future directions and new features.

Other new features not discussed here, but under development and testing, are the following: **1.** We have implemented a GIST scoring component to account for receptor desolvation, using a Gaussian weighting method for precomputing displacement in DOCK 6. This is similar to the method described by Stein et al. in his thesis[60] which built on this paper[61]. **2.** We have implemented a new covalent docking method using an algorithm we call attach-and-grow. **3.** We are also developing a RAS test set. Other work from the Rizzo lab includes RDKit integration and new *de novo* features (personal correspondence with Robert C Rizzo).

4. Discussion.

In recent years, one focus of the docking community has been to perform virtual screens of large, accessible chemical libraries that are now billions of molecules, a number that only continues to grow.[5,6,8,31,32,52,62] There is also a push to dock bespoke libraries.[63] As these databases grow, our methods will need to adjust to access this increase in database size, while concurrently maintaining enough accuracy to take advantage of this growth. Here, we modified DOCK 6 with an eye to method development that will allow us to access larger chemical spaces.

There are four key findings that have been presented in this work. **1.** DOCK 6's extensible design enabled the implementation and testing of the HDB sampling method. **2.** DOCK 3 and DOCK 6 HDB have comparable behavior, although DOCK 3 is faster. **3.** Within DOCK 6, we compared HDB's behavior with different scoring functions, and compared it to anchor-and-grow, learning about its behavior with retrospective experiments. **4.** DOCK 6 has pre-existing functionality that can be used with HDB search, including pose clustering, rescoring, torsion minimization, and hydrogen optimization, which we deployed to complement HDB, as highlighted in the example virtual screen, as well as in the retrospective tests.

DOCK 6 has an extensible design,[38,64] which enables the implementation and testing of new features. The extensible design is most notable in the sampling and scoring portions of the code. This allows us to compare different scoring functions, such as comparing Grid Score and ChemGrid Score, as shown in **Figures 5** and **6,** and **Table 9**. By testing enrichment with both Grid Score and ChemGrid Score, we can compare their strengths and weaknesses. By testing with two different decoy backgrounds (PMD and extrema), we see the enrichment behavior is greatly impacted by background choice. Here, we see that either Grid Score or ChemGrid Score is better, depending on the decoy background. We can also compare two sampling methods (HDB and anchor-and-grow) using pose reproduction tests (discussed more below).

Through enrichment quantification, DOCK 6 HDB and DOCK 3.7 are comparable, although DOCK 3 is faster on average than DOCK 6 (**Figure 3, Table 8**). We attempted to engineer

functions and tweak existing functions within DOCK 6 to match the behavior of DOCK 3. Despite our best efforts, there remain differences. These differences include the orienting and minimization procedures. The minimization differences are due to slight differences in implementation, and the stochastic nature of the procedure. In DOCK 6, we have the ability to minimize all *num_per_search* for each orientation, or the number of poses to be written out, using the *final_min* flag. After re-enabling *final_min* within the code, preliminary results show limited speed advantage and slightly worse results, so we continued to use the *num_per_search* minimization procedure. Regarding the orienting differences, although the algorithms are almost the same, there are some implementational differences, like the *distance_threshold* parameter. Another difference in the orienting is the *max_nodes* for the matching procedure. In DOCK 3.7, this is capped at 4, while in DOCK 6, this parameter does not have a cap, and in fact the default is 10. To us, it is unclear which is better, so we kept these differences. We attempted to choose a parameter set for both programs that would match the behavior of the two code bases to enable comparison. The DUDE-Z test set was indispensable in enabling these tests for developing and evaluating these new functionalities in DOCK 6 and comparing it with DOCK 3.7.[46] As time passed, the code bases were developed separately and diverged, but, unsurprisingly, due to cross-pollination, they seem to have become similar to one another. This is illustrated through DOCK 3 history around the minimization and orienting procedures. DOCK 3.6[43] has simplex minimization, but used a bin-based method for matching. The initial implementation of DOCK 3.7[34] had no minimization, but used the same bipartite matching algorithm as DOCK versions 4, 5, and 6.[65] The simplex minimization was implemented into DOCK 3.7 to enable large-scale docking.[5]

HDB and anchor-and-grow are two different sampling methods. By having both implemented in the same program, we can better compare their strengths and weaknesses, as is seen in the pose reproduction analysis (**Table 10, Figure 7).** In pose reproduction experiments, although DOCK 6 HDB performs worse than anchor-and-grow (anchor-and-grow is 16.01% or 1.27 times better), it is much faster (16 times faster). Although the implementations differ in key aspects, the two programs are performing the same tasks with similar algorithms and good results. Anchor-and-grow (implemented in DOCK 4[66], DOCK 5[64] and DOCK 6[38]) has been tested with pose reproduction[55,67] and enrichment[68]. The HDB routine was implemented first in DOCK 3.5 using the DB file format[36], and then was modified in DOCK 3.7 to use the DB2 file format[34]. The DOCK 3 code and methods have been tested retrospectively with enrichment calculations and with pose reproduction (albeit with less focus on poses than in DOCK 6)[34,46,69], and prospectively in discovery campaigns against simplified model cavities[61,70-73] and drug targets with the aim of drug discovery[5,7]. Performing both enrichment and pose reproduction, we see how they complement one another. It is possible to have strong enrichment, but with poor poses (for example, poses dissimilar to known poses), not binding in the pocket (as is seen in PUR2) (**Figure 6**), or other artifactual issues. Likewise, it is possible to reproduce the experimentally determined poses using a scoring function, but to be unable to distinguish between actives and decoys with that same function.

By implementing HDB search (using the DB2 file format) into DOCK 6, we have enabled more integration of existing DOCK 6 functionality into large-scale docking, including re-docking, minimization, and rescoring (for example, with MM-GBSA or Amber Score). We have demonstrated re-docking with more sampling and torsion minimization in the virtual screening demonstration. Re-docking protocols have been deployed by other groups with good effect.[8] These rescoring protocols are cumbersome in DOCK 3, but are easily performed with this new

version of DOCK 6. Rescoring with MM-GBSA and AmberScore been tested extensively; they offer improvements, but also some vulnerabilities, so they need to be used with scrutiny[72].

There are three caveats that merit discussion. **1.** There are many parameters to explore, and we only explore a small fraction of possibilities. **2.** These are retrospective tests, which only evaluate the ability to reproduce known results, and not the program's ability to find new molecules. **3.** Under-sampling and approximate scoring results in errors and makes it difficult to evaluate improvements. DOCK 6 orienting and minimization does not behave the same as that of DOCK 3; although the algorithm is the same, the implementations differ, so comparing them is difficult. Although we tried many sets of parameters, they do not behave the same. Simplex minimization is stochastic in nature and can result in different minimum poses between the two programs, sometimes different in score. The noise in the calculations is in part due to under-sampling.

Conclusions. The caveats should not detract from the key findings. DOCK 6, as an extensible program, enables testing new scoring and sampling methods. The HDB routine implemented into DOCK 6 is comparable to that in DOCK 3. We tested DOCK6 HDB with different scoring functions, which showed that the decoy background matters. Moreover, we compared DOCK 6 HDB to a different sampling routine (anchor-and-grow). HDB is much faster but does not have as good of a pose reproduction ability as anchor-and-grow. Finally, HDB can be integrated with other features in DOCK 6, i.e., clustering, minimization, and other scoring functions, to enable new protocols in virtual screening. This code will be made available in a future release of the DOCK 6 codebase.

Acknowledgments.

We thank John Irwin, John Bickel, Lorenzo Finci, and Scott Brozell for reading and commenting on this manuscript, and, Brian Shoichet, Robert Rizzo, Priyanka Prakash, Vandana Kumari, Stefan Gahbauer, Lauren Prentis, Jiankun Lyu, Sudipto Mukherjee, and the DOCK community for helpful discussions. We also thank our colleagues at the NCI RAS Initiative and FNLCR, including Dwight Nissley and Frank McCormick, for a collaborative and nurturing environment, and the Frederick Research Computing Environment (FRCE) for excellent computing resources. This project was funded in whole or in part with federal funds from the National Cancer Institute, NIH Contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, and the mention of trade names, commercial products, or organizations does not imply endorsement by the US Government.

References

1. Irwin, J. J.; Shoichet, B. K. Journal of Medicinal Chemistry 2016, 59(9), 4103-4120.
2. Lionta, E.; Spyrou, G.; K. Vassilatis, D.; Cournia, Z. Current Topics in Medicinal Chemistry 2014, 14(16), 1923-1938.
3. Zhu, H.; Zhang, Y.; Li, W.; Huang, N. In International Journal of Molecular Sciences, 2022.

4. Alon, A.; Lyu, J.; Braz, J. M.; Tummino, T. A.; Craik, V.; O'Meara, M. J.; Webb, C. M.; Radchenko, D. S.; Moroz, Y. S.; Huang, X.-P.; Liu, Y.; Roth, B. L.; Irwin, J. J.; Basbaum, A. I.; Shoichet, B. K.; Kruse, A. C. Nature 2021, 600(7890), 759-764.
5. Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Nature 2019, 566(7743), 224-229.
6. Stein, R. M.; Kang, H. J.; McCorvy, J. D.; Glatfelter, G. C.; Jones, A. J.; Che, T.; Slocum, S.; Huang, X.-P.; Savych, O.; Moroz, Y. S.; Stauch, B.; Johansson, L. C.; Cherezov, V.; Kenakin, T.; Irwin, J. J.; Shoichet, B. K.; Roth, B. L.; Dubocovich, M. L. Nature 2020, 579(7800), 609-614.
7. Fink, E. A.; Xu, J.; Hübner, H.; Braz, J. M.; Seemann, P.; Avet, C.; Craik, V.; Weikert, D.; Schmidt, M. F.; Webb, C. M.; Tolmachova, N. A.; Moroz, Y. S.; Huang, X.-P.; Kalyanaraman, C.; Gahbauer, S.; Chen, G.; Liu, Z.; Jacobson, M. P.; Irwin, J. J.; Bouvier, M.; Du, Y.; Shoichet, B. K.; Basbaum, A. I.; Gmeiner, P. Science, 377(6614), eabn7065.
8. Sadybekov, A. A.; Sadybekov, A. V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.-P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N. K.; Tong, F.; Zvonok, N.; Jain, M. K.; Savych, O.; Radchenko, D. S.; Nikas, S. P.; Petasis, N. A.; Moroz, Y. S.; Roth, B. L.; Makriyannis, A.; Katritch, V. Nature 2022, 601(7893), 452-459.
9. Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.-F.; Fischer, P. D.; Coote, P. W.; Padmanabha Das, K. M.; Malets, Y. S.; Radchenko, D. S.; Moroz, Y. S.; Scott, D. A.; Fackeldey, K.; Hoffmann, M.; Iavniuk, I.; Wagner, G.; Arthanari, H. Nature 2020, 580(7805), 663-668.
10. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. Journal of Molecular Biology 1982, 161(2), 269-288.
11. Kuntz, I. D. Science 1992, 257(5073), 1078-1082.
12. Shoichet, B. K.; Stroud, R. M.; Santi, D. V.; Kuntz, I. D.; Perry, K. M. Science 1993, 259(5100), 1445-1450.
13. Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. arXiv 2023.
14. Shuya, N.; Yoshiharu, M.; Shigenori, T. bioRxiv 2022, 2022.2012.2020.521309.
15. McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D. R. Journal of Cheminformatics 2021, 13(1), 43.
16. Goodsell, D. S.; Morris, G. M.; Olson, A. J. Journal of Molecular Recognition 1996, 9(1), 1-5.
17. Goodsell, D. S.; Olson, A. J. Proteins: Structure, Function, and Bioinformatics 1990, 8(3), 195-202.
18. Trott, O.; Olson, A. J. Journal of Computational Chemistry 2010, 31(2), 455-461.
19. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Journal of Medicinal Chemistry 2004, 47(7), 1739-1749.
20. Meiler, J.; Baker, D. Proteins: Structure, Function, and Bioinformatics 2006, 65(3), 538-548.
21. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. Journal of Molecular Biology 1996, 261(3), 470-489.
22. Abagyan, R.; Totrov, M.; Kuznetsov, D. Journal of Computational Chemistry 1994, 15(5), 488-506.
23. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Journal of Molecular Biology 1997, 267(3), 727-748.
24. Liu, M.; Wang, S. Journal of Computer-Aided Molecular Design 1999, 13(5), 435-451.
25. Korb, O.; Stützle, T.; Exner, T. E. Ant Colony Optimization and Swarm Intelligence, Berlin, Heidelberg, 2006// 2006, pp 247-258.
26. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. PLOS Computational Biology 2014, 10(4), e1003571.
27. Pagadala, N. S.; Syed, K.; Tuszynski, J. Biophys Rev 2017, 9(2), 91-102.

28. Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D. Journal of Computer-Aided Molecular Design 1994, 8(5), 565-582.
29. Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. Journal of Computer-Aided Molecular Design 1994, 8(2), 153-174.
30. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Nature Reviews Drug Discovery 2004, 3(11), 935-949.
31. Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. Journal of Chemical Information and Modeling 2020.
32. Tingle, B. I.; Tang, K. G.; Castanon, M.; Gutierrez, J. J.; Khurelbaatar, M.; Dandarchuluun, C.; Moroz, Y. S.; Irwin, J. J. Journal of Chemical Information and Modeling 2023, 63(4), 1166-1176.
33. Lyu, J.; Irwin, J. J.; Shoichet, B. K. Nature Chemical Biology 2023.
34. Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K. PLOS ONE 2013, 8(10), e75992.
35. Coleman, R. G. https://sites.google.com/site/dock37wiki/home/mol2db2-format (March 02, 2023).
36. Lorber, D. M.; Shoichet, B. K. Protein Science 1998, 7(4), 938-950.
37. Jiang, L.; Rizzo, R. C. The Journal of Physical Chemistry B 2015, 119(3), 1083-1102.
38. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. RNA 2009, 15(6), 1219-1230.
39. Allen, W. J.; Balius, T.; Bickel, J.; Boysan, B.; Brozell, S.; Corbo, C.; Fochtman, B.; Jiang, L.; Lang, P. T.; McGee Jr., T. D.; Moustakas, D.; Mukherjee, S.; Pak, S.; Prentis, L.; Singleton, C.; Zhou, Y.; Rizzo, R.; Case, D.; Shoichet, B.; Kuntz, I. DOCK 6.10 Users Manual. https://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm. (May 03, 2023).
40. Gallagher, K.; Sharp, K. Biophys J 1998, 75(2), 769-776.
41. Sharp, K. A. Biopolymers 1995, 36(2), 227-243.
42. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Journal of Computational Chemistry 1992, 13(4), 505-524.
43. Mysinger, M. M.; Shoichet, B. K. Journal of Chemical Information and Modeling 2010, 50(9), 1561-1573.
44. Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Proteins: Structure, Function, and Bioinformatics 1999, 34(1), 4-16.
45. Prentis, L. E.; Singleton, C. D.; Bickel, J. D.; Allen, W. J.; Rizzo, R. C. Journal of Computational Chemistry 2022, 43(29), 1942-1963.
46. Stein, R. M.; Yang, Y.; Balius, T. E.; O'Meara, M. J.; Lyu, J.; Young, J.; Tang, K.; Shoichet, B. K.; Irwin, J. J. Journal of Chemical Information and Modeling 2021, 61(2), 699-714.
47. https://dudez.docking.org/. (July 23, 2022).
48. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Journal of Computational Chemistry 2004, 25(13), 1605-1612.
49. Gasteiger, J.; Rudolph, C.; Sadowski, J. Tetrahedron Computer Methodology 1990, 3(6, Part C), 537-547.
50. Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Zhu, T.; Thompson, J. D.; Winget, P.; Lynch, B. J. Series. AMSOL. 2004
51. Landrum, G.; Tosco, P.; Kelley, B.; Ric; sriniker; gedeck; Vianello, R.; NadineSchneider; Cosgrove, D.; Kawashima, E.; Dalke, A.; N, D.; Jones, G.; Cole, B.; Swain, M.; Turk, S.; AlexanderSavelyev; Vaucher, A.; Wójcikowski, M.; Take, I.; Probst, D.; Ujihara, K.; Scalfani, V. F.; godin, g.; Pahl, A.; Berenger, F.; JLVarjo; strets123; JP; DoliathGavid. Series. RDKit: Open-source cheminformatics. 2022
52. Sterling, T.; Irwin, J. J. Journal of chemical information and modeling 2015, 55(11), 2324-2337.

53. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Journal of Chemical Information and Modeling 2010, 50(4), 572-584.
54. https://zinc15.docking.org/. (Apr 10, 2020).
55. Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Journal of Chemical Information and Modeling 2010, 50(11), 1986-2000.
56. http://ringo.ams.stonybrook.edu/index.php/SB2012. (Sep 22, 2021).
57. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Journal of Cheminformatics 2011, 3(1), 33.
58. Zhang, Y.; Desharnais, J.; Marsilje, T. H.; Li, C.; Hedrick, M. P.; Gooljarsingh, L. T.; Tavassoli, A.; Benkovic, S. J.; Olson, A. J.; Boger, D. L.; Wilson, I. A. Biochemistry 2003, 42(20), 6043-6056.
59. Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoephel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T.; Goodwin, C.; Greb, P.; Häring, D.; Hela, W.; Hoffmann, J.; Karolyi-Oezguer, J.; Knesl, P.; Kornigg, S.; Koegl, M.; Kousek, R.; Lamarre, L.; Moser, F.; Munico-Martinez, S.; Peinsipp, C.; Phan, J.; Rinnenthal, J.; Sai, J.; Salamon, C.; Scherbantin, Y.; Schipany, K.; Schnitzer, R.; Schrenk, A.; Sharps, B.; Siszler, G.; Sun, Q.; Waterson, A.; Wolkerstorfer, B.; Zeeb, M.; Pearson, M.; Fesik, S. W.; McConnell, D. B. Proceedings of the National Academy of Sciences 2019, 116(32), 15823-15829.
60. Stein, R. M.; Singh, I.; Ghattas, M.; Kurtzman, T.; Balius, T. E.; Shoichet, B. K. In Understanding virtual solvent through large-scale ligand discovery; University of California: San Francisco, 2020, p 61-110.
61. Balius, T. E.; Fischer, M.; Stein, R. M.; Adler, T. B.; Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T.; Shoichet, B. K. Proceedings of the National Academy of Sciences 2017, 114(33), E6839-E6846.
62. Irwin, J. J.; Shoichet, B. K. Journal of chemical information and modeling 2005, 45(1), 177-182.
63. Kaplan, A. L.; Confair, D. N.; Kim, K.; Barros-Álvarez, X.; Rodriguiz, R. M.; Yang, Y.; Kweon, O. S.; Che, T.; McCorvy, J. D.; Kamber, D. N.; Phelan, J. P.; Martins, L. C.; Pogorelov, V. M.; DiBerto, J. F.; Slocum, S. T.; Huang, X.-P.; Kumar, J. M.; Robertson, M. J.; Panova, O.; Seven, A. B.; Wetsel, A. Q.; Wetsel, W. C.; Irwin, J. J.; Skiniotis, G.; Shoichet, B. K.; Roth, B. L.; Ellman, J. A. Nature 2022, 610(7932), 582-591.
64. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Journal of Computer-Aided Molecular Design 2006, 20(10), 601-619.
65. Ewing, T. J. A.; Kuntz, I. D. Journal of Computational Chemistry 1997, 18(9), 1175-1189.
66. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. Journal of Computer-Aided Molecular Design 2001, 15(5), 411-428.
67. 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. Journal of Computational Chemistry 2015, 36(15), 1132-1156.
68. Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Journal of Computer-Aided Molecular Design 2012, 26(6), 749-773.
69. Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Journal of Medicinal Chemistry 2012, 55(14), 6582-6594.
70. Barelier, S.; Boyce, S. E.; Fish, I.; Fischer, M.; Goodin, D. B.; Shoichet, B. K. PLOS ONE 2013, 8(7), e69153.
71. Fischer, M.; Coleman, R. G.; Fraser, J. S.; Shoichet, B. K. Nature Chemistry 2014, 6(7), 575-583.
72. Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Journal of Molecular Biology 2008, 377(3), 914-934.
73. Kamenik, A. S.; Singh, I.; Lak, P.; Balius, T. E.; Liedl, K. R.; Shoichet, B. K. Proceedings of the National Academy of Sciences 2021, 118(36), e2106195118.

Supplemental Material. DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking

Trent E Balius*, Y. Stanley Tan, and Mayukh Chakrabarti

NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., PO Box B, Frederick, MD 21702

* Corresponding Author: trent.balius@nih.gov

Section S1. Scoring function agreement.

We rescored the DOCK 3.7 results with DOCK 6 using ChemGrid Score, and we see that they are very correlated (**Figure S1**). Here, we have done this for AMPC and EGFR. We get the solvation parameters from the DB2 file to perform this calculation, since these values are not written out in the mol2 file.

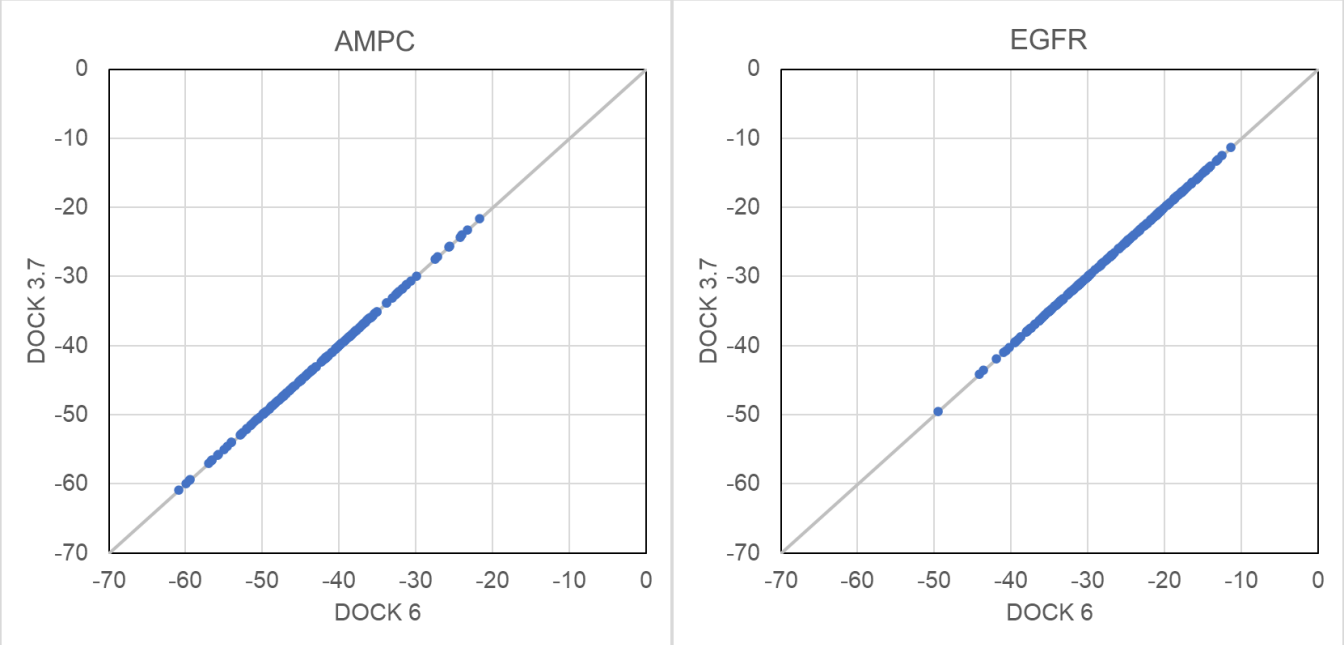


Figure S1. DOCK 3.7 rescored with developmental DOCK 6.9. We rescored 171 and 208 poses of the active molecules for AMPC and EGFR, respectively. Per-atom solvation values are taken from the DB2 files.

Section S2. SMILES stings for SB2012 database generation.

Here, in **Table S1**, we provide the SMILES strings for all the ligands from SB2012 used in pose reproduction. SMILES stings were obtained by scraping the PDB page if the ligand had one *resid*, or by converting the mol2 file to a SMILES string if there was more than one *resid* that defined the ligand.

Table S1. Chemical structures of SB2012 test set used to build DB2 files from SMILES string.

Chemical Structure (SMILES)	Lig name	PDB name
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@@](=O)(CP(=O)(O)O)O)O)O)N	GCP	121P
c1ccccc1	BNZ	181L
c1ccc2c(c1)cco2	BZF	182L
c1ccc2c(c1)CC=C2	DEN	183L
CC(C)Cc1ccccc1	I4B	184L
c1ccc2c(c1)cc[nH]2	IND	185L
CCCCc1ccccc1	N4B	186L
C(CSSCCO)O	HED	187L
Cc1ccc(cc1)C	PXY	187L
C(CSSCCO)O	HED	188L
Cc1ccccc1C	OXE	188L
CC(=O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4=CC(=O)CC[C@]34C)C	STR	1A28
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)[C@@H]([C@@H](CO)O)O)C(=O)O	ZMR	1A4G
CCCN(CCc1ccccc1)C(=O)[C@H]2[C@@H]([C@H](C=C(O2)C(=O)O)N)NC(=O)C	DPC	1A4Q
C[C@]12CC[C@@H]3c4ccc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O	EST	1A52
c1cc2c(cc1F)c(c[nH]2)CCCOP(=O)(O)O	FIP	1A5S
c1nc(c2c(n1)n(cn2)[C@@H]3C=C[C@@H]([C@H]3O)O)N	ADC	1A7A
CC(C)[C@@H](C(=O)NCc1ccc(cc1O)OC)NC(=O)[C@@H]([C@@H]([C@H](Cc2ccccc2)NC(=O)[C@H](C(C)(C)C)NC(=O)OCc3ccccc3)O)NCc4ccc(cc4)OC	2Z4	1A8G
C([C@@H]1[C@H]([C@@H]([C@H]([C@]2(O1)C(=O)NC(=O)N2)O)O)O)O	GLS	1A8I
[H]/N=C(\N)/NCCC[C@@H](C(=O)N)NC(=O)[C@H](CCC(=O)N)NC(=O)[C@@H](C[C@@H]([C@H](CC1CCCCC1)NC(=O)C)O)C(C)C	U0E	1A9M
C[S@](=O)c1ccc(cc1)c2[nH]c(c(n2)c3ccc(cc3)F)c4ccncc4	SB2	1A9U
C[C@@H](C(=O)N[C@@H](C)C(=O)N[C@@H](Cc1ccccc1)[C@H](CCC(=O)N[C@@H](C(C)C)C(=O)N[C@@H](C(C)C)C(=O)OC)O)N	PSI	1AAQ
C1[C@@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O	ARA	1ABE
C1[C@@H]([C@@H]([C@H]([C@H](O1)O)O)O)O	ARB	1ABE
C[C@@H]1[C@@H]([C@@H]([C@H]([C@H](O1)O)O)O)O	FCA	1ABF
C[C@@H]1[C@@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O	FCB	1ABF
c1ccc2c(c1)c(c3c(n2)CCCC3)N	THA	1ACJ
C([C@@H](C(=O)O)NC(=O)CP(=O)(O)O)C(=O)O	PAL	1ACM
c1cnc2c(c1N)ncn2[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O	1DA	1ADD
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@@](=O)(O)OP(=O)(O)O)OP(=O)(O)O)O)N	PAP	1AFK
c1ccc2c(c1)[C@H](C(=O)N2)Cc3ccc(cc3)N4CCN(CC4)C=O	SU2	1AGW
c1cc(cc(c1)[N+](=O)[O-])CC(=O)O	MNP	1AI5
c1cc(ccc1C2(SCCS2)CCCN3CCC(CC3)(c4ccc(cc4)Cl)O)F	THK	1AID
c1ccc(cc1)CN2[C@@H]([C@@H]([C@H]([C@H](N(S2(=O)=O)Cc3ccccc3)COc4ccccc4)O)O)COc5ccccc5	NMB	1AJV
CC(=O)Nc1nnc(s1)S(=O)(=O)N	AZM	1AZM
c1c(c2c([nH]1)c(ncn2)O)[C@H]3[C@@H]([C@@H]([C@H](N3)CO)O)O	IMH	1B8O
CCC(CC)C(=O)Nc1cc(ccc1NC(=O)C)C(=O)O	FDI	1B9S
[H]/N=C(/N)\Nc1cc(ccc1N2C(=O)CCC2(CO)CO)C(=O)O	RAI	1B9T
c1ccc(cc1)CS(=O)(=O)N[C@H]2CCCN(C2=O)CC(=O)N[C@@H](CCCNC(=[NH2+])N)C=O	0IT	1BA8
C1[C@@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O	ARA	1BAP
C1[C@@H]([C@@H]([C@H]([C@H](O1)O)O)O)O	ARB	1BAP
Cc1cc(=O)oc2c1ccc(c2)O[C@@H]3[C@@H]([C@H]([C@@H]([C@H](O3)CO)O[C@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O[C@H]5[C@@H]([C@H]([C@@H]([C@H](O5)CO)O)O)O)NC(=O)C)O)NC(=O)C)O)NC(=O)C	UMG	1BB6
Cc1cc(=O)oc2c1ccc(c2)O[C@H]3[C@@H]([C@H]([C@@H]([C@H](O3)CO)O[C@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O)O)NC(=O)C)O)NC(=O)C	GUM	1BB7
C(F)(F)(F)S(=O)(=O)N	FMS	1BCD

CC(C)[C@@H](c1[nH]ccn1)NC(=O)[C@H](Cc2ccccc2)C[C@@H]([C@H](Cc3ccccc3)NC(=O)OC(C)(C)C)O	IM1	1BDQ
CC(C)[C@@H](c1[nH]ccn1)NC(=O)[C@H](Cc2ccccc2)C[C@@H]([C@H](Cc3ccccc3)NC(=O)OC(C)(C)C)O	IM1	1BDR
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2GP	1BIR
CCCN(CCc1ccccc1)C(=O)[C@H]2[C@@H]([C@H](C=C(O2)C(=O)O)N)NC(=O)C	DPC	1BJI
[H]/N=C(\c1ccc(cc1)NC(=O)Nc2ccc(cc2)Cl)/N	GP6	1BJU
[H]/N=C(\c1ccc(cc1)NC(=O)Nc2ccc(cc2)Oc3ccccc3)/N	GP8	1BJV
c1cc(ccc1c2c(n(cn2)C3CCNCC3)c4ccnc(n4)N)F	SB4	1BL7
Cc1ccc(cc1)CNS(=O)(=O)c2ccc(s2)S(=O)(=O)N	AL5	1BN1
COc1cccc(c1)N2C=Cc3cc(sc3S2(=O)=O)S(=O)(=O)N	AL6	1BN3
COc1ccc(cc1)CNS(=O)(=O)c2ccc(s2)S(=O)(=O)N	AL9	1BN4
COc1cccc(c1)N2CCc3cc(sc3S2(=O)=O)S(=O)(=O)N	AL1	1BNN
COc1ccc(cc1)N2C[C@@H](c3cc(sc3S2(=O)=O)S(=O)(=O)N)O	AL2	1BNT
c1cc(sc1)CN2C[C@@H](c3cc(sc3S2(=O)=O)S(=O)(=O)N)O	AL3	1BNU
CN[C@@H]1CN(S(=O)(=O)c2c1cc(s2)S(=O)(=O)N)c3cccc(c3)OC	AL7	1BNV
c1cc(sc1)CNS(=O)(=O)c2ccc(s2)S(=O)(=O)N	TPD	1BNW
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	UMP	1BP0
c1c(nc2c(=O)[nH]c(nc2n1)N)[C@H]([C@@H](CO)O)O	NEO	1BR5
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4ccccc4c5c3c6n2c7ccccc7c6c8c5C(=O)NC8)NC)OC	STU	1BYG
c1cc(cc2c1cc(cc2)C(F)(F)P(=O)(O)O)C(=O)N[C@@H](CCC(=O)O)C(=O)N	TPI	1BZC
c1cc(cc2c1cc(cc2)C(F)(F)P(=O)(O)O)C(=O)O	PIC	1BZJ
CCOCn1c(c(c(=O)[nH]c1=O)C(C)C)Cc2cc(cc(c2)C)C	GCA	1C1B
CCOCn1c(c(c(=O)[nH]c1=O)C(C)C)SC2CCCCC2	612	1C1C
[H]/N=C(/c1ccc2c(c1)nc([nH]2)Cc3[nH]c4ccccc4n3)\N	BAI	1C1R
c1cc(ccc1[C@H](Cc2ccc3c(c2)c(=O)[nH]c(n3)N)C(=O)O)C(=O)N[C@@H](CCC(=O)O)C(=O)O	NHS	1C2T
[H]/N=C(\C1CCC(CC1)CNC(=O)[C@@H]2C=CCn3n2c(=O)n(c3=O)CCS(=O)(=O)c4ccc(cc4)Br)/N	IH1	1C4U
[H]/N=C(/C1CCC(CC1)CNC(=O)[C@@H]2C=C(Cn3n2c(=O)n(c3=O)CC(c4ccccc4)c5ccccc5)C)\N	IH2	1C4V
CC(=O)Nc1ccc2c(c1S(=O)(=O)O)cccc2S(=O)(=O)O	TK4	1C5C
c1cc2c(cc(s2)C(=[NH2+])N)c(c1)I	ESI	1C5N
[H]/N=C(\c1ccccc1)/N	BEN	1C5O
[H]/N=C(\c1ccccc1)/N	BEN	1C5P
c1cc2c(cc(s2)C(=[NH2+])N)c(c1)I	ESI	1C5Q
c1ccc2c(c1)cc(s2)C(=[NH2+])N	ESX	1C5S
c1cc2cc(sc2nc1)C(=[NH2+])N	ESP	1C5T
c1c[nH]c2c1cc(c(c2)NC(=O)C(=O)O)C(=O)O	OAI	1C83
c1ccc2cc(c(cc2c1)C(=O)O)NC(=O)C(=O)O	761	1C84
C1COCc2c1c(c(s2)NC(=O)C(=O)O)C(=O)O	OPA	1C86
C1COCc2c1c(c(s2)NC(=O)C(=O)O)C(=O)O	OPA	1C87
C1CNCc2c1c(c(s2)NC(=O)C(=O)O)C(=O)O	OTA	1C88
C[N@]1CC[C@@H]([C@@H](C1)O)c2c(cc(c3c2oc(cc3=O)c4ccccc4Cl)O)O	CPB	1C8K
c1ccc(c(c1)O)SCC/C=C/P(=O)(O)O	HE1	1C8V
Cc1c(c(c(cn1)COP(=O)(O)O)C=O)O	PLP	1C8V
c1cc(c(cc1F)SCC/C=C/P(=O)(O)O)O	HF1	1C9D
c1ccc(cc1)C[C@H](CC(=O)O)C(=O)O	BZS	1CBX
[H]/N=C(\c1ccccc1)/N	BEN	1CE5
CCN(CC)CCC[C@@H](C)Nc1ccnc2c1ccc(c2)Cl	CLQ	1CET
CCN[C@H]1C[C@@H](S(=O)(=O)c2c1cc(s2)S(=O)(=O)N)C	ETS	1CIL
C[C@H]1C[C@@H](c2cc(sc2S1(=O)=O)S(=O)(=O)N)N	PTS	1CIM
C[C@H]1C[C@@H](c2cc(sc2S1(=O)=O)S(=O)(=O)N)NC	MTS	1CIN
CC(C)[C@H](CO)Nc1nc(c2c(n1)n(cn2)C(C)C)Nc3ccc(c(c3)Cl)C(=O)O	PVB	1CKP
c1cc(ccc1C(=O)NCCOCCOCCNC(=O)CN)S(=O)(=O)N	EG1	1CNW
c1cc(ccc1C(=O)NCCOCCOCCN)S(=O)(=O)N	EG2	1CNX
c1ccc(cc1)C[C@@H](C(=O)NCCOCCOCCNC(=O)c2ccc(cc2)S(=O)(=O)N)N	EG3	1CNY
C1=CC(C=CC1O)(CC(=O)C(=O)O)C(=O)O	PRE	1COM
CS(=N)(=N)C[C@@H](Cc1ccccc1)C(=O)O	CPM	1CPS
C1C=CN(C(=O)N1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O	DHZ	1CTT
c1ccc(c(c1)O)[S@](=O)CCCCP(=O)(O)O	HSP	1CW2
c1cc(ccc1c2cc(nn2c3ccc(cc3)S(=O)(=O)N)C(F)(F)F)Br	S58	1CX2
c1ccc(c(c1)N)SCCCCP(=O)(O)O	NHP	1CX9
Cc1c(c(c(cn1)COP(=O)(O)O)C=O)O	PLP	1CX9
C([C@@H](C(=O)O)NC(=O)CP(=O)(O)O)C(=O)O	PAL	1D09
c1cc(ccc1c2c(c3ccc(cc3s2)O)Cc4ccc(c(c4)Br)CN5CCCC5)OCCN6CCCC6	BZT	1D3D
c1cc(ccc1Cc2c3ccc(cc3sc2c4ccc(nc4)OCCN5CCCC5)O)OCCN6CCCC6	BT3	1D3P
[H]/N=C(/c1ccc2c(c1)cc([nH]2)C(=O)N3CCC(CC3)Cc4ccccc4)\N	BPP	1D4P

CCC[C@]1(CC(=C(C(=O)O1)[C@H](CC)c2cccc(c2)NS(=O)(=O)c3ccc(cn3)C(F)(F)F)O)CCc4ccccc4	TPV	1D4Y
c1ccc(cc1)S(=O)(=O)CCn2c(=O)n3n(c2=O)[C@@H](C=CC3)C(=O)NC4CCC(CC4)c5cnc([nH]5)N	00R	1D6W
c1ccc(cc1)S(=O)(=O)CCn2c(=O)n3n(c2=O)[C@@H](C=CC3)C(=O)NCC4CCC(CC4)N	00P	1D9I
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	VDX	1DB1
CC(=O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4=CC(=O)CC[C@]34C)C	STR	1DBB
C[C@]12CC[C@H](C[C@H]1CC[C@@H]3[C@@H]2CC[C@]4([C@H]3CCC4=O)C)O	AE2	1DBJ
c1cn(c(=O)[nH]c1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	U5P	1DBT
C1[C@H]2[C@@H]([C@@H](S1)CCCCC(=O)O)NC(=O)N2	BTN	1DF8
[C@H]1([C@@H]([C@H]([C@@H]([C@H]([C@@H]1OP(=O)(O)O)O)OP(=O)(O)O)OP(=O)(O)O)O)O	I3P	1DJX
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=O)N2Cc3cccc(c3)N)Cc4cccc(c4)N)Cc5ccccc5)O)O	DMQ	1DMP
c1c(=O)[nH]c(=O)n(c1O)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	BMP	1DQX
C(C[C@@H](C(=O)O)N)CN[P@](=O)(N)NS(=O)(=O)O	PSQ	1DUV
c1c(=O)[nH]c(=O)n(n1)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	UP6	1DVJ
CC(C)NC[C@@H](COc1cccc2c1cccc2)O	SNP	1DY4
CCC1=C[C@@H]2Cc3c(c(c4ccc(cc4n3)Cl)N)[C@@H](C2)C1	HUX	1.00E+66
C([C@@H]1[C@H]([C@@H]([C@H](c2n1nnn2)O)O)O)O	NTZ	1E6Q
C([C@@H]1[C@H]([C@@H]([C@H](/C(=N/O)/N1)O)O)O)O	GOX	1E6S
C([C@@H]1[C@H]([C@@H]([C@H](/C(=N/O)/N1)O)O)O)O	GOX	1.00E+72
[H]/N=C(\c1cccc(c1)C(=O)N[C@H](c2ccccc2)C(=O)N3CCC(CC3)C(=O)c4ccccc4)/N	BPO	1EB2
C(C(=O)NO)P(=O)(O)O	PAH	1EBG
CCCC[C@@H](C(=O)N)NC(=O)[C@H](C)NC(=O)[C@H](CCC(=O)O)NC(=O)[C@H](Cc1ccccc1)NC[C@H](CC(C)C)NC(=O)[C@H](C(C)C)NC(=O)[C@H](CCCNC(=[NH2+])N)N	0Q4	1EBK
[C@H]([C@@H](C(=O)NO)O)([C@H](C(=O)[O-])O)O	XYH	1EC9
C([C@@H]([C@H](C(=O)[O-])O)O)[C@@H](C(=O)[O-])O	DXG	1ECQ
C([C@@H]1[C@@H]([C@@H]([C@H]([C@H](O1)O)O)O)O)O	GLA	1EEF
c1cc(ccc1N)O[C@@H]2[C@@H]([C@H]([C@H]([C@H](O2)CO)O)O)O	GAT	1EFI
COc1cccc(c1)c2[nH]c3cccc(c3n2)C(=O)N	BZC	1EFY
C1C(=O)NC(=O)N(C1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	BMQ	1EIX
[H]/N=C(/N)\Nc1ccc(cc1)CNC(=O)NC23CC4CC(C2)CC(C4)C3	AGB	1EJN
CC(C)C[C@@H](C(=O)Nc1ccc(cc1)C(C)C)NC(=O)[C@H](CCCC[NH3+])NC(=O)C(F)(F)F	0Z4	1ELB
CC(C)c1ccc(cc1)NC(=O)[C@H](Cc2ccccc2)NC(=O)[C@H](CCCC[NH3+])NC(=O)C(F)(F)F	0Z3	1ELC
CC(C)c1c(n(c(n1)COC(=O)N)Cc2ccncc2)Sc3cc(cc(c3)Cl)Cl	S11	1EP4
CC(=O)N[C@@H]1[C@H]([C@@H]([C@H](O[C@H]1O)CO)O)O	NAG	1EQG
C[C@@H](c1ccc(cc1)CC(C)C)C(=O)O	IBP	1EQG
C[C@@H](c1ccc(c(c1)F)c2ccccc2)C(=O)O	FLP	1EQH
C[C@]12CC[C@@H]3c4ccc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O	EST	1ERE
[H]/N=C(/c1ccc(cc1)C[C@@H](C(=O)N2CCCCC2)NS(=O)(=O)c3ccc(cc3)C)\N	4QQ	1ETT
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1EUS
COc1cc2c(cc1OC)C(=O)[C@@H](C2)CC3CCN(CC3)Cc4ccccc4	E20	1EVE
[H]/N=C(/c1cccc(c1)C[C@H]([C@@H](C)NC(=O)c2ccc(cc2)c3cccc(c3)CN)C(=O)OC)\N	RPR	1EZQ
c1cc2c(cc(s2)S(=O)(=O)N[C@H]3CCN(C3=O)Cc4ccc5ccnc(c5c4)N)nc1	815	1F0R
c1cc2c(cc(s2)S(=O)(=O)N[C@H]3CCN(C3=O)Cc4cc5ccncc5[nH]4)nc1	PR2	1F0S
[H]/N=C(\c1ccc(c(c1)CN2CC[C@@H](C2=O)NS(=O)(=O)c3cc4c(s3)cccn4)O)/N	PR1	1F0T
[H]/N=C(/c1cccc(c1)C[C@H]([C@@H](C)NC(=O)c2ccc(cc2)c3cccc(c3)CN)C(=O)OC)\N	RPR	1F0U
c1cc(ccc1CNC2=[NH+]CCCC2)N	TPM	1F3D
Cc1ccc(cc1)S(=O)(=O)N2CCC[C@@H]2C(=O)O	TPR	1F4E
c1cc(ccc1C(=O)N[C@@H](CCC(=O)O)C(=O)O)S(=O)(=O)N2CCC[C@@H]2C(=O)O	TP3	1F4F
c1cc(ccc1C(=O)N[C@@H](CCC(=O)O)C(=O)O)S(=O)(=O)N2CCC[C@@H]2C(=O)NCCC(=O)O	TP4	1F4G
C([C@H](C(=O)O)N)S	DCY	1F57
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1F8B
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)N	4AM	1F8C
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CN)O)O)C(=O)O)O	9AM	1F8D
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CN)O)O)C(=O)O)N	49A	1F8E
C1C(=Nc2c(=O)[nH]c(nc2N1)N)CO	PH2	1F9H
c1ccc(cc1)CNC(=O)c2cccc(c2)O[C@@H]3[C@@H]([C@H]([C@H]([C@H](O3)CO)O)O)O	AI1	1FD7
CCCC[C@@H](CN[C@@H](CCCC)C(=O)N[C@@H](CCC(=O)N)C(=O)N[C@@H](CCCNC(=[NH2+])N)C(=O)N)NC(=O)[C@H]([C@@H](C)CC)NC(=O)[C@H]([C@@H](C)O)NC(=O)C	2NC	1FEJ
CCCC[C@@H](C(=O)N)NC(=O)[C@H](C)NC(=O)[C@H](CCC(=O)O)NC(=O)[C@H](Cc1ccccc1)NC[C@H](CC(C)C)NC(=O)[C@H](C(C)C)NC(=O)[C@H](CCCNC(=[NH2+])N)N	0Q4	1FFF
CCCC[C@@H](C(=O)N)NC(=O)[C@H](C)NC(=O)[C@H](CCC(=O)O)NC(=O)[C@H](Cc1ccccc1)NC[C@H](CC(C)C)NC(=O)[C@H](C(C)C)NC(=O)[C@H](CCCNC(=[NH2+])N)N	0Q4	1FG8
Cc1c[nH]c(c1CCC(=O)O)/C=C\2/c3ccccc3NC2=O	SU1	1FGI
[H]/N=C(/c1ccc(c(c1)Oc2c(c(c(c(n2)Oc3cccc(c3)C4=NCCN4C)F)[N@@](C)CC(=O)O)F)O)\N	Z34	1FJS

CC1(COC(=O)CCCCCCCCCCCCOC(=O)[C@@H]2CCCCN2C(=O)C1=O)C	SB1	1FKI
Cc1c(nc(o1)c2ccccc2)CCOc3ccc(cc3)C[C@@H](C(=O)O)Nc4ccccc4C(=O)c5ccccc5	570	1FM9
C1[C@H]([C@H]([C@@H]([C@H](N1)CO)O)O)O	DMJ	1FO2
C([C@@H]1[C@H]([C@@H]([C@@H]([C@@H]2N1C(=O)C(=O)N2)O)O)O)O	KIF	1FO3
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4cccnc4	PRC	1FPU
CC(C)(C)OC(=O)N[C@@H](Cc1ccccc1)[C@H](CN[C@@H](Cc2ccccc2)C(=O)N[C@@H](CCC(=O)O)C(=O)N[C@@H](Cc3ccccc3)C(=O)N)O	0ZT	1FQX
[H]/N=C(/c1ccc(cc1)NCc2nc3cc(ccc3n2C)C4(CC4)/C(=N\OCC(=O)OCC)/c5ccccn5)\N	T87	1G2L
CO[C@@H](c1ccc(cc1)CN2[C@@H]([C@@H]([C@H]([C@H](N(S2(=O)=O)Cc3ccc(cc3)CO)COc4ccccc4)O)O)COc5ccccc5)O	AHF	1G35
[H]/N=C(/c1ccc(cc1)NCc2nc3cc(ccc3n2C)Cn4c(nc5c4cccc5)C)\N	R11	1G36
c1cc(c(c(c1)F)CNC(=O)c2ccc(cc2)S(=O)(=O)N)F	F6B	1G48
c1cc(c(c(c1)F)F)CNC(=O)c2ccc(cc2)S(=O)(=O)N	F2B	1G52
c1cc(c(c(c1)F)CNC(=O)c2ccc(cc2)S(=O)(=O)N)F	F6B	1G53
CCCCCNC(=O)[C@H](Cc1ccc(cc1)OC(C(=O)O)C(=O)O)NC(=O)N[C@@H](Cc2ccccc2)C(=O)O	INZ	1G7F
CCCCCNC(=O)[C@H](Cc1ccc(c(c1)C(=O)O)OCC(=O)O)NC(=O)[C@H](Cc2ccccc2)NC(=O)CCC(=O)O	INX	1G7G
Cc1cc(cc(c1)NC(=O)Cc2ccc(cc2)OC(C)(C)C(=O)O)C	RQ3	1G9V
c1cc(c(=O)[nH]c1)c2[nH]c3ccc(cc3n2)C(=[NH2+])N	120	1GHV
c1ccc(c(c1)c2[nH]c3ccc(cc3n2)C(=[NH2+])N)O	BMZ	1GHW
c1cc(c(=O)[nH]c1)c2[nH]c3ccc(cc3n2)C(=[NH2+])N	120	1GHZ
c1ccc(c(c1)c2[nH]c3cc(ccc3n2)C(=[NH2+])N)O	122	1GI4
c1ccc(c(c1)c2cc3cc(ccc3[nH]2)C(=[NH2+])N)O	124	1GI6
c1ccc(cc1)c2cccc(c2O)c3cc4cc(c(cc4[nH]3)Cl)C(=[NH2+])N	132	1GJ6
CC(C)C[C@H](CN(C=O)O)C(=O)N[C@H](C(=O)NC)C(C)(C)C	NFH	1GKC
CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)NC(=O)Cc3ccccc3)C(=O)O)C	PNN	1GM7
CC1([C@@H](N2[C@H]([S@H]1O)[C@@H](C2=O)NC(=O)Cc3ccccc3)C(=O)O)C	SOX	1GM8
CC1([C@@H](N2[C@H]([S@H]1O)[C@@H](C2=O)NC(=O)Cc3ccccc3)C(=O)O)C	SOX	1GM9
C/C=C/1\[C@@H]2Cc3c(ccc(=O)[nH]3)[C@]1(CC(=C2)C)N	HUP	1GPK
CC1=C[C@H]2Cc3c(ccc(=O)[nH]3)[C@@]4(C1)[C@@H]2CCCN4	HUB	1GPN
c1cc(ccc1c2cc3ccc(cc3s2)O)O	ZTW	1GWQ
C[C@]12CC[C@@H]3c4ccc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O	EST	1GWR
c1cc(cc(c1)N2CCN(CC2)CCC(=O)c3cc(c(c(c3)O)O)[N+](=O)[O-])C(F)(F)F	BIA	1H1D
c1cc(=O)[nH]c2c1[C@H](CCC2)NCCCCCCCCCCCN[C@H]3CCCc4c3ccc(=O)[nH]4	E10	1H22
c1cc(=O)[nH]c2c1[C@H](CCC2)NCCCCCCCCCCCCCN[C@H]3CCCc4c3ccc(=O)[nH]4	E12	1H23
CC(C)/N=C/[C@H](COc1cccc2c1cccc2)O	RNP	1H46
COc1ccc2c(c1)CN(CCS2)C(=O)CCN3CCC(CC3)Cc4ccccc4	K21	1HAK
c1ccc(cc1)C[C@H](C(=O)O)NC(=O)NO	INF	1HDQ
c1cc(c[n+](c1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO[P@@](=O)([O-])O[P@@](=O)(O)OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)O)O)O)C(=O)N	NAD	1HEX
CC(C)C[C@H](CC(=O)NO)C(=O)N[C@@H](Cc1ccccc1)C(=O)NC	PLH	1HFC
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)O)OP(=O)(O)O)O)N	A3P	1HI4
CC[C@H](C)[C@@H](C(=O)NCc1ccccn1)NC(=O)[C@@H]([C@H]([C@@H]([C@H](CC2CCCCC2)NC(=O)[C@H](Cc3c[nH]c[nH+]3)NC(=O)COc4cccc5c4cccc5)O)O)C(C)C	1ZK	1HIV
CCCCCCCCCCCCCCCCCOC[C@H](COCC(F)(F)F)O[P@](=O)(O)OC	MJI	1HN4
c1cc2c(cc1S(=O)(=O)N)CNCC2	SKF	1HNN
C([C@@H]([C@H]([C@@H]([C@@H](COP(=O)(O)O)O)O)O)N)O	AGP	1HOR
c1cnc(c2c1n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O)N	AD3	1HP0
CC(C)[C@@H](C(=O)N[C@@H](Cc1ccccc1)[C@H](C[C@@H](Cc2ccccc2)c3[nH]c(cn3)C(=O)C(C)C)O)NC(=O)OCc4ccccc4	RUN	1HPS
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CCOC2)O)S(=O)(=O)c3ccc(cc3)N	478	1HPV
CC(C)(C)NC(=O)[C@@H]1CSCN1C(=O)[C@H]([C@H](Cc2ccccc2)NC(=O)[C@H](CSC)NC(=O)COc3cccc4c3ccnc4)O	KNI	1HPX
C1C(=Nc2c(=O)[nH]c(nc2N1)N)CO	PH2	1HQ2
CC(C)(C)NC(=O)[C@@H]1C[N@@](CC[N@]1C[C@H](C[C@@H](Cc2ccccc2)C(=O)N[C@H]3c4ccccc4C[C@H]3O)O)Cc5cccnc5	MK1	1HSG
CC(C)(C)NC(=O)[C@@H]1C[N@@](CC[N@]1C[C@H](C[C@@H](Cc2ccccc2)C(=O)N[C@H]3c4ccccc4C[C@H]3O)O)Cc5cccnc5	MK1	1HSH
c1c([nH+]c[nH]1)C[C@@H](C(=O)O)N	HIS	1HSL
C[C@@H](c1ccc(c(c1)F)c2ccccc2)C(=O)OC	FL2	1HT5
CC(=O)N[C@@H]1[C@H]([C@@H]([C@H](O[C@H]1O)CO)O)O	NAG	1HT8
CCCOc1ccc(cc1Cl)CC(=O)O	34C	1HT8
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=NC#N)N2Cc3ccc(cc3)CO)Cc4ccc(cc4)CO)Cc5ccccc5)O)O	Q82	1HVH
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=O)N2Cc3ccc4ccccc4c3)Cc5ccc6ccccc6c5)Cc7ccccc7)O)O	XK2	1HVR
Cc1[nH]c2ccc(cc2c(=O)n1)CN(C)c3ccc(s3)C(=O)N[C@@H](CCC(=O)O)C(=O)O	D16	1HVY

CC[C@H](C)C(=O)O[C@H]1CCC=C2[C@H]1[C@H]([C@H](C=C2)C)CC[C@H](C[C@H](CC(=O)O)O)O	114	1HW8
CCC(C)(C)C(=O)O[C@H]1C[C@H](C=C2[C@H]1[C@H]([C@H](C=C2)C)CC[C@H](C[C@H](CC(=O)O)O)O)C	SIM	1HW9
CC(C)n1c2ccccc2c(c1/C=C/[C@H](C[C@H](CC(=O)O)O)O)c3ccc(cc3)F	115	1HWI
CC(C)c1c(c(c(c(n1)C(C)C)COC)c2ccc(cc2)F)CC[C@H](C[C@H](CC(=O)O)O)O	116	1HWJ
CC(C)c1c(c(c(n1CC[C@H](C[C@H](CC(=O)O)O)O)c2ccc(cc2)F)c3ccccc3)C(=O)Nc4ccccc4	117	1HWK
CC(C)c1c(c(nc(n1)N(C)S(=O)(=O)C)c2ccc(cc2)F)CC[C@H](C[C@H](CC(=O)O)O)O	FBI	1HWL
C=CCN1[C@@H]([C@@H]([C@H]([C@H](N(C1=O)CC=C)Cc2ccccc2)O)O)Cc3ccccc3	216	1HWR
C1C[C@H]([C@@H]2[C@@H]([C@@H](C[N@]2C1)O)O)O	SWA	1HWW
CC(C)(C)NC(=O)[C@@H]1C[C@@H]2CCCC[C@@H]2C[N@]1C[C@H]([C@H](Cc3ccccc3)NC(=O)[C@H](CC(=O)N)NC(=O)c4ccc5ccccc5n4)O	ROC	1HXB
CC(C)c1nc(cs1)CN(C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc2ccccc2)C[C@@H]([C@H](Cc3ccccc3)NC(=O)OCc4cncs4)O	RIT	1HXW
c1ccc(cc1)C[C@H](CC(=O)O)C(=O)O	BZS	1HYT
COc1cccc(c1)N2C(=Cc3cc(sc3S2(=O)=O)S(=O)(=O)N)CN4CCOCC4	INL	1I8Z
COCCCN1C=C(c2cc(sc2S1(=O)=O)S(=O)(=O)N)N	INM	1I90
c1cc(cc(c1)O)N2C(=Cc3cc(sc3S2(=O)=O)S(=O)(=O)N)CN4CCOCC4	INQ	1I91
c1ccc(cc1)Sc2cccc3c2c(nc(n3)N)N	TQ3	1IA1
Cc1ccc(cc1)Sc2cccc3c2c(nc(n3)N)N	TQ4	1IA2
CC(C)(C)c1ccc(cc1)Sc2cccc3c2c(nc(n3)N)N	TQ5	1IA3
c1cc2c(c(c1)Sc3ccc(cc3)N4CCOCC4)c(nc(n2)N)N	TQ6	1IA4
CC(C)[C@@H](C(=O)N[C@@H](Cc1ccccc1)[C@@H](C[N@@]2CC[C@H](C[C@H]2C(=O)NC(C)(C)C)SCc3cccnc3)O)NC(=O)c4ccc5ccccc5n4	0PO	1IDA
Cc1cccc(c1OCC(=O)N[C@@H](Cc2ccccc2)[C@@H](C[N@@]3CC[C@H](C[C@H]3C(=O)NC(C)(C)C)S(=O)(=O)c4ccncc4)O)C	0DO	1IDB
CCC(CC)(CCCO[C@H](C)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C)O	KH1	1IE8
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	VDX	1IE9
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C	STI	1IEP
C[C@H](CNC(=O)c1ccc(cc1)S(=O)(=O)N)Cn2ccc3c2cccc3	SBR	1IF7
C[C@@H](CNC(=O)c1ccc(cc1)S(=O)(=O)N)Cn2ccc3c2cccc3	SBS	1IF8
Cc1c(sc[n+]1Cc2cnc(nc2N)C)CCO	VIB	1IG3
C(C(=O)NO)OP(=O)(O)O	PGH	1IK4
c1cc2c(cc1Cl)[C@@](OC(=O)N2)(C#CC3CC3)C(F)(F)F	EFZ	1IKW
CC(=O)Nc1ccc(cc1NC(N)N)C(=O)O	ST4	1INF
CC(=O)N[C@@H]1[C@H](C[C@@H](O[C@H]1[C@@H]([C@@H](CO)O)O)P(=O)(O)O)O	AXP	1INW
[O-]S(=O)(=O)[O-]	SO4	1ITV
CC(=O)Nc1c(cc(cc1O)C(=O)O)[N+](=O)[O-]	ST1	1IVB
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1IVF
CC[C@H](C)[C@@H](C(=O)NCc1ccccn1)NC(=O)[C@@H]([C@H]([C@@H]([C@H](CC2CCCCC2)NC(=O)[C@H](Cc3c[nH]c[nH+]3)NC(=O)COc4cccc5c4cccc5)O)O)C(C)C	1ZK	1IVP
CC(C)[C@H](C[C@@H]([C@H](CC1CCCCC1)NC(=O)[C@H](CC(=O)N)NC(=O)c2ccc3ccccc3n2)O)C(=O)NCC(C)(C)C	0PX	1IVQ
C1CNC(=O)[C@H]([C@@H]1O[C@H]2[C@@H]([C@H]([C@@H](CO2)O)O)O)O	XIL	1J01
C[N+](C)(C)CCCCCCCCCC[n+]1c2cc(ccc2c3ccc(cc3c1c4ccccc4)N)N	DCU	1J07
CC1(N=C(N=C(N1OCCCOc2cc(c(cc2Cl)Cl)Cl)N)N)C	WRA	1J3I
CCc1c(c(nc(n1)N)N)c2ccc(cc2)Cl	CP6	1J3J
CC1(N=C(N=C(N1OCCCOc2cc(c(cc2Cl)Cl)Cl)N)N)C	WRA	1J3K
CC(=O)O	ACY	1JCZ
CC(=O)Nc1nnc(s1)S(=O)(=O)N	AZM	1JD0
c1ccc(cc1)C[C@@H]([C@H](Cc2ccc3c(c2)OCO3)C(=O)O)C(=O)O	BYS	1JJE
c1cc2c(cc1C[C@@H]([C@H](Cc3ccc4c(c3)OCO4)C(=O)O)C(=O)O)OCO2	BDS	1JJT
c1cc2c(cc1Cl)[C@@](OC(=O)N2)(C#CC3CC3)C(F)(F)F	EFZ	1JKH
c1cc(ccc1C(=O)N[C@@H](CCC(=O)O)C(=O)O)[C@@](Cc2ccc3c(c2)c(nc(n3)N)O)(CNCC(=O)N[C@H]4[C@@H]([C@@H]([C@H](O4)COP(=O)(O)O)O)O)O	138	1JKX
CC(C)c1c(n(c(=O)[nH]c1=O)COCc2ccccc2)Cc3ccccc3	TNK	1JLA
Cc1ccnc2c1NC(=O)c3cccnc3N2C4CC4	NVP	1JLB
CCOc1ccnc(c1F)CCNC(=S)Nc2ccc(cn2)Cl	FTC	1JLC
CCC(CC)NC(=O)C[C@@H](C(=O)NC[C@H]([C@H](Cc1ccccc1)NC(=O)[C@H]([C@@H](C)O)NC(=O)c2ccc3ccccc3n2)O)C(C)(C)C	0PP	1JLD
Cc1ccnc2c1NC(=O)c3cccnc3N2C4CC4	NVP	1JLF
Cc1c(cco1)C(=S)Nc2ccc(c(c2)OCC=C(C)C)Cl	UC1	1JLG
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	UMP	1JMF
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	UMP	1JMG
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	UMP	1JMI

c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)CO)O[P@](=O)(O)O[P@](=O)(O)OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)O	139	1JN4
c1ccc(cc1)CS(=O)(=O)N2C[C@@H]3CC[C@H](N3C(=O)C2)C(=O)NCC4CCN(CC4)/C(=N/O)/N	BLI	1JWT
c1[nH]c2c(n1)c(ncn2)N	ADE	1JYS
[H]/N=C(\c1cccc(c1)C[C@@H](C(=O)N2CCCC[C@@H]2C(=O)O)NS(=O)(=O)c3ccc4ccccc4c3)/N	FD1	1K1I
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)N2CCC(CC2)C(=O)OC)NS(=O)(=O)c3ccc4ccccc4c3)\N	FD2	1K1J
[H]/N=C(\c1cccc(c1)C[C@@H](C(=O)N2CCNCC2)NS(=O)(=O)c3ccc4ccccc4c3)/N	FD3	1K1L
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)N2CCN(CC2)C(=O)C)NS(=O)(=O)c3ccc4ccccc4c3)\N	FD4	1K1M
[H]/N=C(/N)\NCCCNC(=O)[C@@H]1CCCCN1C(=O)[C@@H](CC2CCCCC2)NCC(=O)O	IGN	1K21
[H]/N=C(\c1ccc(cc1)CNC(=O)[C@@H]2CCN2C(=O)[C@@H](C3CCCCC3)NCC(=O)O)/N	MEL	1K22
c1ccc2c(c1)c(c[nH]2)CC(=O)N[C@@H](CC(=O)O)C(=O)O	IAD	1K3U
c1ccc2c(c1)c(c[nH]2)CC(=O)NCC(=O)O	IAG	1K7E
c1cc(ccc1CCCCc2ccc(cc2)C(F)(F)P(=O)(O)O)C(F)(F)P(=O)(O)O	FEP	1KAV
CNS(=O)(=O)c1ccc(cc1)N/C=C\2/c3ccccc3NC2=O	LS1	1KE5
CNS(=O)(=O)Cc1ccc(cc1)N/N=C\2/c3c(ccc4c3scn4)NC2=O	LS2	1KE6
c1cc2c(cc1c3cnco3)/C(=C/Nc4ccc5c(c4)CS(=O)(=O)C5)/C(=O)N2	LS3	1KE7
c1ccc2c(c1)/C(=C/Nc3ccc(cc3)S(=O)(=O)Nc4nccs4)/C(=O)N2	LS4	1KE8
[H]/N=C(/N)\NS(=O)(=O)c1ccc(cc1)N/C=C\2/c3ccccc3NC2=O	LS5	1KE9
Cc1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)CO)O	THM	1KIM
Cc1c(c2c(c(c1O)C)CC[C@](O2)(C)CCC[C@@H](C)CCC[C@@H](C)CCCC(C)C)C	VIT	1KPM
c1ccc(cc1)CCCCCCC(=O)N[C@H](CCc2ccccc2)CCC(=O)O	BR4	1KQU
C[C@H]([C@@H](Cc1cccc(c1)C(=[NH2+])N)C(=O)OC)NC(=O)c2ccc(cc2)c3cc[n+](cc3)[O-]	FXV	1KSN
[H]/N=C(/c1ccc(cc1)NCc2nc3cc(ccc3n2C)C(=O)N(CCC(=O)OCC)c4ccccn4)\N	C24	1KTS
[H]/N=C(/c1ccc(cc1)Cc2nc3cc(ccc3n2C)NS(=O)(=O)c4ccccc4)\N	C02	1KTT
Cc1ccc(cc1)n2c(cc(n2)C(C)(C)C)NC(=O)Nc3ccc(c4c3cccc4)OCCN5CCOCC5	B96	1KV2
c1ccc(cc1)CCCCCCC(=O)N[C@@H](CCC(=O)O)CSc2ccc(cc2)Cc3ccccc3	OAP	1KVO
[H]/N=C(\c1cccc(c1)C[C@H](C(=O)NCCc2ccccc2)NC(=O)NC34CC5CC(C3)CC(C5)C4)/N	RUP	1KYE
Cc1ccccc1CNC(=O)[C@@H]2C(SCN2C(=O)[C@H]([C@H](Cc3ccccc3)NC(=O)c4cccc(c4C)O)O)(C)C	JE2	1KZK
CC[C@@H]1Cc2cc(ccc2C3=C1c4ccc(cc4C[C@H]3CC)O)O	ETC	1L2J
c1cc(ccc1NS(=O)(=O)c2ccsc2C(=O)O)Cl	STC	1L2S
[H]/N=C(\N)/N[C@@H]1C[C@@H]([C@H]([C@H]1[C@H](C(CC)CC)NC(=O)C)O)C(=O)O	BCZ	1L7F
[H]/N=C(\N)/N[C@@H]1C[C@@H]([C@H]([C@H]1[C@H](C(CC)CC)NC(=O)C)O)C(=O)O	BCZ	1L7G
[H]/N=C(\N)/N[C@@H]1C[C@@H]([C@H]([C@H]1[C@H](C(CC)CC)NC(=O)C)O)C(=O)O	BCZ	1L7H
C(C[C@@H](C(=O)O)N)CN	ORN	1LAH
CC(C)C[C@@H](C(=O)O)N	LEU	1LAN
Cc1cc[nH]c1	5MP	1LI6
CC(=O)N[C@@H]1[C@H]([C@@H]([C@H](O[C@H]1O[C@H]2C[C@H](N[C@@H]2CO)C(=O)NCCS(=O)(=O)O)CO)OS(=O)(=O)O)O	BUL	1LMC
c1cn(c(=O)[nH]c1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	U	1LOQ
c1c(=O)[nH]c(=O)n(c1O)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	BMP	1LOR
c1c(=O)[nH]c(=O)n(n1)[C@H]2[C@@H]([C@@H]([C@H](O2)COP(=O)(O)O)O)O	UP6	1LOS
[H]/N=C(/c1cccc(c1)Cn2c3ccc(cc3cc2C(=O)NCc4ccc(cc4)[N+](C)(C)C)OCc5ccccc5)\N	IMA	1LPG
[H]/N=C(/c1cccc(c1)Cn2c3ccccc3cc2C(=O)OCc4cccc(c4)/C(=N\[H])/N)\N	CBB	1LPK
[H]/N=C(/c1cccc(c1)Cn2c3cccc(c3cc2C(=O)NCc4cc(cc(c4)Cl)Cl)C)\N	CMB	1LPZ
c1ccc2c(c1)cccc2CC(=O)O	NLA	1LRH
C(CC[NH3+])C[C@@H](C(=O)O)N	LYS	1LST
COCCOc1cc2c(cc1OCCOC)ncnc2Nc3cccc(c3)C#C	AQ4	1M17
C[C@@H]1C[C@H]2[C@@H]3CCC4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CO)O)C)O)F)C	DEX	1M2Z
Cn1c2c(cc(c1=O)c3c(cccc3Cl)Cl)cnc(n2)Nc4cccc(c4)SC	P17	1M52
[C@H]1([C@@H]([C@H]([C@@H]([C@H]([C@@H]1OP(=O)(O)O)O)O)OP(=O)(O)O)OP(=O)(O)O)O)O	I3P	1MAI
Cc1c2c(c(c(c1OC)C/C=C(\C)/CCC(=O)O)O)C(=O)OC2	MOA	1ME7
c1[nH]c(=O)c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)O)O)O	IMP	1MEH
Cc1c2c(c(c(c1OC)C/C=C(\C)/CCC(=O)O)O)C(=O)OC2	MOA	1MEH
Cc1c2c(c(c(c1OC)C/C=C(\C)/CCC(=O)O)O)C(=O)OC2	MOA	1MEI
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=O)N2Cc3cccc(c3)N)Cc4cccc(c4)N)Cc5ccccc5)O)O	DMQ	1MER
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=O)N2Cc3ccc(cc3)CO)Cc4ccc(cc4)CO)Cc5ccccc5)O)O	DMP	1MES
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=O)N2Cc3ccc(cc3)CO)Cc4ccc(cc4)CO)Cc5ccccc5)O)O	DMP	1MET
C(C(=O)O)C(CC(=O)O)(C(=O)O)O	CIT	1MLD
CCCNC(=[NH2+])NCCC[C@@H](C(=O)O)N	3AR	1MMV
CCCC(=[NH2+])NCCC[C@@H](C(=O)O)N	VIO	1MMW
CN1CCN(CC1)Cc2csc(c2Cl)C(=O)Nc3ccc(cc3C(=O)Nc4ccc(cc4)Cl)Cl	XLC	1MQ5

CN(Cc1csc(c1Cl)C(=O)Nc2c(cc(cc2OC)Cl)C(=O)Nc3ccc(cn3)Cl)C4=NCCO4	XLD	1MQ6
c1nc2c(c(n1)N)[nH]nc2[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O	FMC	1MRK
CC[C@H](C)[C@H]1C(=O)NCCCOc2ccc(cc2)C[C@@H](C(=O)N1)NC[C@H]([C@H](Cc3ccccc3)NC(=O)OC(C)(C)C)O	PI6	1MTR
Cc1cnc(c(=O)n1CC(=O)NCc2c(cccn2)F)NCC(c3ccccn3)(F)F	CDA	1MU6
Cc1ccnc(c1F)CNC(=O)Cn2c(c[nH+]c(c2=O)NCC(c3cccc[nH+]3)(F)F)C	CDB	1MU8
c1ccc(c(c1)CNC(=O)Cn2c(cnc(c2=O)NCC(c3cccc[n+]3[O-])(F)F)Cl)F	CDD	1MUE
CC[C@@H](C)c1nccs1	TZL	1MUP
CC1(CCC(c2c1ccc(c2)C3(OCCO3)c4ccc(cc4)C(=O)[O-])(C)C)C	BM6	1MVC
CC(=O)N[C@@H]1[C@H](C[C@@](O[C@H]1[C@@H]([C@@H](CO)O)O)(C(=O)O)O)O	SIA	1MWE
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1MZ6
CC[C@]1(CCCCN(C1=O)C)c2cccc(c2)Oc3cc(ccc3C#N)[C@@](C)(c4cncn4C)N	BNE	1MZC
C([C@@H]1[C@H]([C@@H]([C@@H]([C@H](O1)O)O)O)O)O	MAN	1N1M
C([C@@H]1[C@H]([C@@H]([C@@H]([C@@H](O1)O)O)O)O)O	BMA	1N1M
CC(C)[C@@H](C(=O)N1CCCC1)N	A3M	1N1M
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1N1T
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1N1V
CC(=O)N[C@@H]1[C@H](C[C@@](O[C@H]1[C@@H]([C@@H](CO)O)O)(C(=O)O)O)O	SIA	1N1Y
CC(C)(CO)[C@H](C(=O)[O-])O	PAF	1N2J
CCCCc1[nH]c2c(n1)c(=O)[nH][nH]c2=O	BDI	1N2V
Cc1cc(cc(c1Oc2ccc(c(c2)C(C)C)O)C)n3c(=O)[nH]c(=O)cn3	PFA	1N46
CC(C)c1cc(ccc1O)Oc2c(cc(cc2Cl)CC(=O)O)Cl	IH5	1NAV
CSC[C@@H]1[C@H]([C@H]([C@@H](O1)n2ccc3c2ncnc3N)O)O	MTH	1NC1
c1nc2c(c(n1)N)[nH]nc2[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O	FMC	1NC3
c1ccc2c(c1)[nH]c(n2)NCc3ccc(s3)c4csc(n4)N=C(N)N	FR0	1NDV
c1ccc(cc1)CC[C@H](CO)n2cc(nc2)C(=O)N	FR2	1NDW
c1ccc2c(c1)cccc2CC[C@H](CO)n3cc(nc3)C(=O)N	FR3	1NDY
Cn1c2ccccc2nc1CCC(=O)Nc3ccc4ccn(c4c3)CC[C@H](CO)n5cc(nc5)C(=O)N	FR5	1NDZ
[H]/N=C(\c1cccc(c1)CN2CCN(CC2=O)S(=O)(=O)c3cc4ccc(cc4s3)Cl)/N	RRP	1NFU
c1cc(sc1/C=C/S(=O)(=O)[N@]2CCN(C(=O)C2)Cc3cc4cnccc4[nH]3)Cl	RRR	1NFW
c1cc(cc2c1cc(s2)S(=O)(=O)N3CCN(C(=O)C3)Cc4cc5cnccc5n4CCO)Cl	RDR	1NFX
[H]/N=C(\c1ccc(cc1)CN2CCN(CC2=O)S(=O)(=O)c3cc4ccc(cc4s3)Cl)/N	RTR	1NFY
c1cc(c(cc1CC(=O)O)[N+](=O)[O-])O	NPA	1NGP
c1cn(c(=O)nc1N)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	DCM	1NJA
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	UMP	1NJD
c1cn(c(=O)nc1N)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	DCM	1NJE
C[C@H]([C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O[C@@H]2[C@H](C[C@H]([C@@H]([C@H]2O)O[C@@H]3[C@@H]([C@H]([C@@](CO3)(C)O)NC)O)N)N)N)O)O)O	GET	1NJJ
c1cc(ccc1[C@@H](CCC[C@H]2[C@H](N[C@H](NC2=O)N)N)C(C(F)(F)F)(O)O)C(=O)N[C@@H](CCC(=O)O)C(=O)O	KEU	1NJS
c1[nH]c2c(n1)c(ncn2)N	ADE	1NLI
c1ccc(cc1)C[C@H]2CNCCCNCCc3ccccc3CNC(=O)Cn4c(cnc(c4=O)N2)Cl	L86	1NM6
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1NNB
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)[C@@H]([C@@H](CO)O)O)C(=O)O	ZMR	1NNC
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1NSD
c1cc2c(cc1Cl)CNC(=O)[C@@H]3CCCN3C(=O)[C@H](NC(=O)CCNC(=O)CO2)C4CCCCC4	T76	1NT1
C[S@@+](CC[C@@H](C(=O)[O-])N)C[C@@H]1[C@H]([C@H]([C@@H](O1)n2cnc3c2ncnc3N)O)O	SAM	1NW5
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CSCC[C@@H](C(=O)O)N)O)O)N	SAH	1NW7
[H]/N=C(\c1cc(cn1C)CNC(=O)c2cccn2C(=O)[C@@H](CC3CCCCC3)NCC(=O)O)/N	162	1NZQ
[H]/N=C(\c1ccc(n1C)CNC(=O)c2cccn2C(=O)[C@@H](CC3CCCCC3)NCC(=O)O)/N	163	1O0D
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)O)OP(=O)(O)O)O)N	A3P	1O0F
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)OP(=O)(O)O)O)O)N	ADP	1O0H
c1cn(c(=O)[nH]c1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)OP(=O)(O)O	U2P	1O0M
c1cn(c(=O)[nH]c1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)OP(=O)(O)O)O	U3P	1O0N
c1ccc(cc1)c2cccc(c2[O-])c3cc4cc(ccc4[nH]3)C(=[NH2+])N	696	1O2G
c1cc(c(c(c1)OC2CCCC2)[O-])c3cc4cc(ccc4[nH]3)C(=[NH2+])N	CR3	1O2H
c1cc(c(c(c1)OC2CCCC2)[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N	655	1O2I
CC(C)COc1cccc(c1[O-])c2[nH]c3ccc(cc3n2)C(=[NH2+])N	656	1O2J
CC(C)COc1cccc(c1[O-])c2[nH]c3ccc(cc3n2)C(=[NH2+])N	656	1O2K
c1ccc(cc1)c2cccc(c2[O-])c3[nH]c4cc(c(cc4n3)C(=[NH2+])N)Cl	762	1O2N
CC(C)COc1cccc(c1[O-])c2[nH]c3cc(c(cc3n2)C(=[NH2+])N)F	950	1O2O
c1cc(c(c(c1)OC2CCCC2)[O-])c3cc4cc(c(cc4[nH]3)Cl)C(=[NH2+])N	991	1O2Q
c1ccc(c(c1)c2[nH]c3ccc(cc3n2)C(=[NH2+])N)[O-]	CR4	1O2S
c1cc2c(cc1C(=[NH2+])N)nc([nH]2)c3cc(cc(c3[O-])Br)[C@@H](CC(=O)[O-])C(=O)[O-]	847	1O2U
c1cc2c(cc1C(=[NH2+])N)nc([nH]2)c3cc(cc(c3[O-])Br)[C@@H](CC(=O)[O-])C(=O)[O-]	847	1O2V

COc1ccccc1c2cc(cc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N)[C@@H](CC(=O)[O-])C(=O)[O-]	312	1O2Z
c1ccc(c(c1)c2cc(cc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N)[C@@H](CC(=O)[O-])C(=O)[O-])F	693	1O30
c1cc(c(nc1)c2[nH]c3ccc(cc3n2)C(=[NH2+])N)[O-]	801	1O31
c1ccc(c(c1)c2cc(cc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N)[C@@H](CC(=O)[O-])C(=O)[O-])N	607	1O36
c1ccc(cc1)c2cc(cc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N)CC[NH3+]	653	1O37
c1ccc(cc1)c2cccc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N	780	1O39
c1ccc(cc1)c2cccc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N	780	1O3B
c1ccc(cc1)c2cccc(c2[O-])c3[nH]c4ccc(cc4n3)C(=[NH2+])N	780	1O3D
c1ccc(cc1)c2cccc(c2[O-])c3cc4cc(ccc4[nH]3)C(=[NH2+])N	696	1O3E
c1ccc(cc1)c2cccc(c2[O-])c3cc4cc(ccc4[nH]3)C(=[NH2+])N	696	1O3F
c1ccc(cc1)c2cccc(c2[O-])c3cc4cc(ccc4[nH]3)C(=[NH2+])N	696	1O3G
Cc1cc(c(c(c1)Br)[O-])c2cc3cc(ccc3[nH]2)C(=[NH2+])N	907	1O3H
Cc1cc(c(c(c1)Br)[O-])c2cc3cc(ccc3[nH]2)C(=[NH2+])N	907	1O3I
C[C@H]1CCCC[C@@H]1Oc2cccc(c2[O-])c3[nH]c4cc(c(cc4n3)C(=[NH2+])N)F	CR9	1O5G
c1ccc(cc1)CC[C@@H](C(=O)O)N[C@@H](CCCCN)C(=O)N2CCC[C@H]2C(=O)O	LPR	1O86
c1ccc2c(c1)c(ccn2)NCCCCCCCCCNc3c4ccccc4nc5c3CCCC5	A8B	1ODC
CC(C)(C)OC(=O)N[C@@H](Cc1ccccc1)[C@@H](CNC[C@H]([C@H](Cc2ccccc2)NC(=O)OC(C)(C)C)O)O	0.00E+00	1ODW
Cc1cn(c(=O)[nH]c1=O)[C@]23C[C@H]2[C@@H]([C@H](C3)O)CO	SCT	1OF1
c1cc(ccc1C[C@H](C(=O)O)N)O	DTY	1OF6
Cc1c(cccc1O)C(=O)N[C@@H](CSc2ccccc2)[C@@H](C[N@@]3C[C@H]4CCCC[C@H]4C[C@H]3C(=O)NC(C)(C)C)O	1UN	1OHR
C1[C@@H]([C@H]([C@@H](CN1)O)O)CO	IFM	1OIF
C1[C@@H]([C@H]([C@@H]([C@H](N1)CO)O)O)O	NOJ	1OIM
CN(C)c1cccc2c1cccc2S(=O)(=O)N	MNS	1OKL
c1cc(ccc1C(=O)NCCCCNCS)S(=O)(=O)N	STB	1OKN
Cn1c2c(cc(c1=O)c3c(cccc3Cl)Cl)cnc(n2)Nc4cccc(c4)CO	P16	1OPK
Cc1ccc(cc1)c2cc(nn2c3ccc(cc3)S(=O)(=O)N)C(F)(F)F	CEL	1OQ5
c1ccc(cc1)C[C@H](C[P@@](=O)([C@H](Cc2ccccc2)N)O)C(=O)N[C@@H](Cc3ccccc3)C(=O)O	0PQ	1OS0
C(C[C@@H](C(=O)O)N)CNC(=O)CP(=O)(O)O	PAO	1OTH
[H]/N=C(\c1ccc2cc(ccc2c1)C(=O)Nc3ccccc3)/N	675	1OWE
CC(C)CCCCCO	MHN	1OYF
COc1ccc(cc1OC2CCCC2)[C@H]3CC(=O)NC3	ROL	1OYN
c1cc(ccc1CN2C(=O)[C@H]3[C@@H]4CCC[N@@]4[C@H]([C@H]3C2=O)c5ccc(cc5)C(=[NH2+])N)F	FSN	1OYT
c1cc(ccc1C(=O)N[C@@H](CCC(=O)O)C(=O)O)N(Cc2ccc3c(n2)c(=O)[nH]c(n3)N)C(=O)/C=C/c4c(ncn4[C@H]5[C@@H]([C@@H]([C@H](O5)COP(=O)(O)O)O)O)C(=O)N	MS1	1OZ0
C[N@@]1CCC[C@H]1c2cccnc2	NCT	1P2Y
c1ccc2c(c1)-c3ccccc3C2N4CCN(CC4)C(=O)c5ccc6c(c5)cc[nH]6	GEQ	1P44
c1cn(c(=O)nc1N)[C@H]2C([C@@H]([C@H](O2)CO)O)(F)F	GEO	1P62
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CSCC[C@@H](C(=O)O)N)O)O)N	SAH	1P9P
c1cc(c(cc1[N+](=O)[O-])OS(=O)(=O)O)O	CSN	1PA9
C([C@H]([C@H]([C@@H]([C@H](C(=O)O)O)O)O)O)OP(=O)(O)O	6PG	1PGP
c1cc(c[n+](c1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO[P@@](=O)([O-])O[P@@](=O)(O)OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)O)O)O)C(=O)N	NAD	1PL8
CCCn1c(c(nc1C2CCN(CC2)C)c3ccc(c(c3)Cl)Cl)c4ccnc(n4)NC5CC5	984	1PMN
[H]/N=C(/c1ccc(cc1)C[C@H](C(=O)N2CCCCC2)NC(=O)CNS(=O)(=O)c3ccc4ccccc4c3)\N	MID	1PPC
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)N2CCCCC2)NS(=O)(=O)c3ccc(cc3)C)\N	0ZG	1PPH
COc1cc(ccc1O)CN2[C@@H]([C@@H](C[N@](N(C2=O)Cc3ccc(c(c3)OC)O)Cc4ccccc4)O)Cc5ccccc5	A88	1PRO
C([C@@H]1[C@H]([C@@H]([C@@H]([C@@H]2N1C(=O)C(=O)N2)O)O)O)O	KIF	1PS3
CC(C)(C)OC(=O)N[C@@H](Cc1ccc(cc1)OC(C(=O)O)C(=O)O)C(=O)NCCCCOc2cccc(c2C(=O)OC)O	941	1PYN
CSC[C@@H]1[C@H]([C@H]([C@@H]([NH2+]1)c2c[nH]c3c2nc[nH]c3=O)O)O	MTI	1Q1G
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4ccccc4c5c3c6n2c7ccccc7c6c8c5C(=O)NC8)NC)OC	STU	1Q3D
c1ccc2c(c1)-c3c(c4cc(ccc4[nH]3)[N+](=O)[O-])CC(=O)N2	ATU	1Q3W
c1ccc2c(c1)/C(=C/3\C(=N\O)\c4ccccc4N3)/C(=O)N2	IXM	1Q41
C[C@@H](c1ccc(cc1)c2ccccc2)C(=O)O	BFL	1Q4G
c1cc(cc(c1)Cl)C2=C(C(=O)NC2=O)Nc3ccc(c(c3)C(=O)O)Cl	679	1Q4L
c1ccc2c(c1)nnn2C(Cc3ccc(cc3)C(F)(F)P(=O)(O)O)(Cc4ccc(cc4)C(F)(F)P(=O)(O)O)c5ccc(c(c5)F)F	P27	1Q6M
c1ccc(cc1)[C@@](Cc2ccc(cc2)c3cccc(c3)P(=O)(O)O)(Cc4ccc(cc4)C(F)(F)P(=O)(O)O)n5c6ccccc6nn5	213	1Q6P
Cc1ccc2cc(cc(c2n1)P(=O)(O)O)c3ccc(cc3)C[C@@](Cc4ccc(cc4)C(F)(F)P(=O)(O)O)(c5ccccc5)n6c7ccccc7nn6	214	1Q6S

CC(C)C[C@@H](c1ccc2cc(cc(c2n1)P(=O)(O)O)c3ccc(cc3)C[C@@](Cc4ccc(cc4)C(F)(F)P(=O)(O)O)(c5ccccc5)n6c7ccccc7nn6)OC	600	1Q6T
c1ccc(cc1)c2c3cc(ccc3c4ccc(cc4[n+]2CCCCCCc5cn(nn5)CCNc6c7ccccc7nc8c6CCCC8)N)N	TZ4	1Q84
C([C@@H](C(=O)O)NC(=O)CP(=O)(O)O)C(=O)O	PAL	1Q95
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CSCC[C@@H](C(=O)O)N)O)O)N	SAH	1QAN
[H]/N=C(\c1ccc(c(c1)Oc2c(c(c(c([nH+]2)Oc3cccc(c3)c4nccn4C)F)[N@]5CCC[C@H](C5)C(=O)O)F)O)/N	974	1QB1
[H]/N=C(\N)/c1cc(ccc1)Oc2[nH+]c(c(c(c2F)C)F)Oc3cc(ccc3)/C(=N/[H])/N	623	1QB6
[H]/N=C(/C)\N1CCC(CC1)Oc2ccc3c4ccccc4n(c3c2)Cc5ccc6ccc(cc6c5)/C(=N/[H])/N	806	1QB9
[H]/N=C(\c1ccc(c(c1)Oc2cc(cc(n2)Oc3cccc(c3)c4[nH]ccn4)C(=O)O)O)/N	688	1QBN
c1ccc(cc1)C[C@@H]2[C@@H]([C@H]([C@H](N(C(=O)N2Cc3ccc(cc3)CO)Cc4ccc(cc4)CO)Cc5ccccc5)O)O	DMP	1QBS
C[C@H]1[C@@H]2CC[C@]3([C@H]([C@]2(CC[C@H]1O)C)[C@@H](C[C@@H]\4[C@@]3(C[C@@H](/C4=C(/CCC=C(C)C)\C(=O)O)OC(=O)C)C)O)C	FUA	1QCA
Cc1ccc(cc1)c2c3c(ncnc3n(n2)C(C)(C)C)N	PP1	1QCF
c1ccc(cc1)C[C@@H](C(=O)N[C@@H](Cc2ccc(cc2)O)C(=O)O)NC(=O)[C@H](Cc3ccccc3)S	TI2	1QF0
CCCCC[C@@H](C(=O)N[C@@H](Cc1ccccc1)C(=O)N[C@@H](C)C(=O)O)S	TI1	1QF1
c1ccc(cc1)C[C@@H](C(=O)NCC(=O)N2[C@H](CC[C@H]2C(=O)O)c3ccccc3)S	TI3	1QF2
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O[P@@](=O)(O)O[P@@](=O)(O)OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)OP(=O)(O)O	PUA	1QHC
c1cc(ccc1N([C@H](O)SC[C@@H](C(=O)NCC(=O)O)NC(=O)CC[C@@H](C(=O)O)N)O)I	GIP	1QIN
C[C@]12CC[C@@H]3c4ccc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O	EST	1QKT
C[C@]12CC[C@@H]3c4ccc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O	EST	1QKU
c1cc(cc2c1cc(cc2)Cl)S(=O)(=O)N3CCN(CC3)C(=O)C4CCN(CC4)c5ccncc5	ZEN	1QL7
CC(C)(C)n1c2c(c([nH+]1)c3ccc(cc3)Cl)c(ncn2)N	PP2	1QPE
c1cc(c(nc1)C(=O)O)C(=O)O	NTM	1QPQ
CC(=O)N[C@@H](Cc1ccc(c(c1)O)OCC(=O)O)C(=O)NCCCCOc2cccc(c2C(=O)OC)O	429	1QXK
C[C@@H](C(=O)O)NC(=O)[C@H](Cc1ccc(cc1)c2ccccc2)C[P@@](=O)([C@H](C)N)O	BIR	1R1H
C([C@@H]1[C@H]([C@@H]([C@@H]([C@H](S1)N)O)O)O)O	LKA	1R33
CC(C)C[C@H]([C@@H](C(=O)NO)O)C(=O)N[C@H](C(=O)NC)C(C)(C)C	97	1R55
CC(C)SCC[C@H]([C@@H](C(=O)NNC(=O)c1cccc(c1)Cl)O)N	AO5	1R58
CC/C(=C(\c1ccccc1)/c2ccc(cc2)/C=C/C(=O)O)/c3ccccc3	GW5	1R5K
C[C@@H](c1ccc(c(c1)F)c2ccccc2)C(=O)O	FLP	1R9O
C([C@H]([C@H]([C@](COP(=O)(O)O)(C(=O)O)O)O)O)OP(=O)(O)O	CAP	1RBO
CC1=C(C(CCC1)(C)C)/C=C/C(=C/C=C/C(=C/CO)/C)/C	RTL	1RBP
C[C@H]1[C@H]([C@H]([C@@H]([C@@H](O1)OC)O)O)O	MFU	1RDI
C([C@@H]1[C@@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O)O	GAL	1RDK
CC(=O)N[C@@H]1[C@H]([C@@H]([C@H](O[C@@H]1O)CO)O)O	NDG	1RDN
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2AM	1RGK
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2GP	1RGL
C[C@@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C3C[C@H](C(=C)[C@@H](C3)O)O)C	VDZ	1RJK
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	VDX	1RK3
CC[C@H](C)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C3C[C@H](C(=C)[C@@H](C3)O)O)C	VD1	1RKG
CC1[C@@H](CC(=C/C=C/2\CCC[C@]3([C@H]2CC[C@@H]3[C@H](C)CCCC(C)(C)O)C)C[C@H]1O)O	VD2	1RKH
CC(C)CN(CC(=O)NO)S(=O)(=O)c1ccc(cc1)OC	NGH	1RMZ
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2GP	1RNT
c1cn(c(=O)nc1N)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)OP(=O)(O)O	C2P	1ROB
CCOc1ccc(cc1)c2ccc(cc2)C(=O)C[C@H](CCN3C(=O)c4ccccc4C3=O)C(=O)O	DEO	1ROS
CCOCn1c(c(c(=O)[nH]c1=O)C(C)C)Cc2ccccc2	MKC	1RT1
CC(C)c1c(n(c(=O)[nH]c1=O)COCc2ccccc2)Cc3ccccc3	TNK	1RT2
CCN1c2cc(ccc2N(C(=O)c3c1nccc3)C)[N+](=O)[O-]	U05	1RTH
CCC(CC)(/C=C/C=C/[C@@H](C)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C)O	EB1	1S0Z
C[C@H](/C=C/[C@H](C1CC1)O)[C@H]2CC[C@@H]\3[C@@]2(CCC/C3=C\C=C/4\C[C@H](C[C@@H](C4=C)O)O)C	MC9	1S19
Cc1c(cco1)C(=S)Nc2ccc(c(c2)OCC=C(C)C)Cl	UC1	1S1T
Cc1c(c2cc(ccc2n1C(=O)c3ccc(cc3)Cl)OC)CC(=O)O	IMN	1S2A
CN(C[C@H]1CCC2=C(C1)C(=N[C@@H](N2)N)N)c3cc(c(c(c3)OC)OC)OC	TQD	1S3V
C[N+](C)(C)C[C@@H](CC(=O)O)O	152	1S5O
c1ccc(cc1)CS(=O)(=O)N[C@H](CC2CCCCC2)C(=O)N3CCC[C@H]3C(=O)Nc4ccc5c(c4)sc(n5)N	165	1SB1
CC(C)(C)NC(=O)[C@@H]1C[N@@](CC[N@]1C[C@H](C[C@@H](Cc2ccccc2)C(=O)N[C@H]3c4ccccc4C[C@H]3O)O)Cc5cccnc5	MK1	1SDT
c1cc(ccc1/C=C/c2cc(cc(c2)O)O)O	STL	1SG0
c1cc(ccc1[C@@H]2[C@@H](Oc3ccc(cc3S2)O)c4ccc(cc4)OCCN5CCCCC5)O	E4D	1SJ0

c1cc[n+](c(c1)C(CNc2c(=O)n(c(cn2)Cl)CC(=O)NCc3cc(ccc3n4cnnn4)Cl)(F)F)[O-]	170	1SL3
Cc1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)OP(=O)(O)O	THP	1SNC
CC(C)(CO)[C@H](C(=O)NCCC(=O)O)O	PAU	1SQ5
C[C@]12CC[C@H]3[C@H]([C@@H]1CC[C@]2(C#C)O)CCC4=CC(=O)CC[C@H]34	NDR	1SQN
C[C@@H]1C[C@H]2[C@@H]3CCC4=CC(=O)C=C[C@@]4([C@]3([C@H](C[C@@]2([C@]1(C(=O)CCl)OC(=O)c5ccco5)C)O)Cl)C	MOF	1SR7
Cc1cc(ccc1O)/N=N/c2ccccc2C(=O)O	MHB	1SRG
Cc1cc(cc(c1O)C)/N=N/c2ccccc2C(=O)O	DMB	1SRI
c1ccc2c(c1)c(ccc2O)/N=N/c3ccccc3C(=O)O	NAB	1SRJ
C([C@@H](C(=O)O)N)S	CYS	1SSQ
C1[C@H]2[C@@H]([C@@H](S1)CCCCC(=O)O)NC(=O)N2	BTN	1STP
COc1ccc(cc1)C(=O)O	ANN	1SV3
c1ccc(c(c1)CC(=O)O)Nc2c(cccc2Cl)Cl	DIF	1SV9
C1[C@H]2[C@@H]([C@@H](S1)CCCCC(=O)O)NC(=O)N2	BTN	1SWP
C1[C@H]2[C@@H]([C@H](S1)CCCCC(=O)O)NC(=O)N2	BTQ	1SWP
C1[C@H]2[C@@H]([C@@H](S1)CCCCC(=O)O)NC(=O)N2	BTN	1SWR
CC[C@H](C)C(=O)O[C@H]1C[C@H](C=C2[C@H]1[C@H]([C@H](C=C2)C)CC[C@H](C[C@H](CC(=O)O)O)O)C	LVA	1T02
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	1T3R
c1cc(c(cc1F)OCC(=O)O)C(=O)NCc2nc3c(c(cc(c3s2)F)F)F	ID5	1T40
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C	STI	1T46
CCc1c(c2ccc(cc2o1)S(=O)(=O)N(C)C)C(=O)c3cc(c(c(c3)Br)O)Br	BB3	1T48
CCc1c(c2ccc(cc2o1)S(=O)(=O)Nc3ccc(cc3)S(=O)(=O)N)C(=O)c4cc(c(c(c4)Br)O)Br	892	1T49
[H]/N=C(/N)\NOCCCOc1cc(cc(c1)Cl)C(=O)N(CC=C)C2CCCC2	14A	1T4V
c1ccc(cc1)C(c2ccccc2)[C@H](C(=O)N3CCC[C@H]3C(=O)NCc4cc(ccc4Cl)Cl)N	176	1TA2
CCNC(=O)COc1ccc(cc1CNC(=O)[C@@H]2CCCN2C(=O)[C@@H](C3CCCCC3)N)Cl	177	1TA6
c1c(c(=O)[nH]c(=O)n1[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O)F	UFP	1TDB
c1ccc(cc1)CC[C@H](CC2(CCCC2)C(=O)N[C@@H](Cc3c[nH]c4c3cccc4)C(=O)O)C(=O)O	0DB	1THL
Cc1cc(ccc1n2c(c(c(n2)C)/N=N/c3cc(ccc3C(=O)O)S(=O)(=O)O)O)S(=O)(=O)O	326	1THZ
Cc1ccc(c(n1)N)C[n+]2csc(c2C)CCO[P@@](=O)(O)O[P@](=O)(O)[O-]	N1T	1TKB
C[C@H]1[C@@H]([C@H]([C@H]([C@@H](O1)O[P@](=O)(N[C@@H](CC(C)C)C(=O)N[C@@H](Cc2c[nH]c3c2cccc3)C(=O)O)O)O)O)O	RDF	1TLP
C1CCC(CC1)C[NH3+]	AMC	1TNG
c1cc(ccc1C[NH3+])F	FBA	1TNH
c1ccc(cc1)CCCCN	PBN	1TNI
c1ccc(cc1)CC[NH3+]	PEA	1TNJ
c1ccc(cc1)CCCN	PRA	1TNK
c1ccc(cc1)[C@H]2C[C@@H]2[NH3+]	TPA	1TNL
c1ccc2c(c1)c3ccccc3n2CCCC(=O)O	CRZ	1TOW
C1[C@H]([C@@H]([C@H]([S@@+]1C[C@@H]([C@@H](CO)OS([O-])([O-])[O-])O)CO)O)O	SSD	1TQT
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O	UMP	1TSY
CC(=C)[C@H]1CN[C@@H]([C@H]1CC(=O)O)C(=O)O	KAI	1TT1
C(CCCC(=O)O)CCCC(=O)O	AZ1	1TUF
C[C@H](CC#CC(C)(C)O)[C@H]1CC[C@H]\2[C@@]1(CCC/C2=C\C=C3C[C@H](C[C@@H](C3)O)O)C	TX5	1TXI
CC(C)C[C@@H](C(=O)O)NC(=O)[C@H]([C@@H](Cc1ccccc1)N)O	BES	1TXR
CC/C(=C(/CC)\c1ccc(cc1)O)/c2ccc(cc2)O	DES	1TZ8
Cc1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)COP(=O)(O)O)O[P@@](=O)([O-])O[P@@](=O)([O-])OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)OP(=O)(O)O	PAX	1U1B
c1ccc(cc1)Cc2cn(c(=O)[nH]c2=O)COCCO	BAU	1U1C
C([C@@H]1[C@H]([C@@H]([C@H](/C(=N/O)/N1)O)O)O)O	GOX	1U2Y
CO[C@@H]1[C@H](O[C@@H]([C@@H]([C@H]1O)O)O[C@@H]2[C@H](O[C@@H]([C@@H]([C@H]2O)O)O[C@@H]3[C@H](N/C(=N\O)/[C@@H]([C@H]3O)O)CO)CO)CO	LM2	1U33
c1cc2c(cc1O)onc2c3ccc(cc3O)O	272	1U3Q
c1cc(cc2c1cc(cc2)O)c3c4ccc(cc4on3)O	797	1U3S
c1c[nH]c2c1/C(=C\3/C(=O)NC(=N3)N)/CCNC2=O	DBQ	1U4D
c1cc(ccc1c2cc3cc(ccc3o2)O)O	397	1U9E
c1[nH]c2c(n1)c(=O)[nH]c(n2)N	GUN	1ULB
C([C@H]([C@@H]([C@H](C(=O)COP(=O)(O)O)O)O)O)OP(=O)(O)O	2FP	1UMG
c1ccc(cc1)CCC(=O)Nc2ccc3ccn(c3c2)CC[C@H](CO)n4cc(nc4)C(=O)N	FR4	1UML
CCCCc1c2c([nH]c1c3ccc(cc3)O)nccn2	ALH	1UNG
c1ccc2c(c1)/C(=C/3\C(=N\O)\c4ccccc4N3)/C(=O)N2	IXM	1UNH
CC[C@H](CO)Nc1nc(c2c(n1)n(cn2)C(C)C)NCc3ccccc3	RRC	1UNL
c1ccc(cc1)[N@@]2CCc3cc(ccc3[C@H]2c4ccc(cc4)OCCN5CCCCC5)O	PTI	1UOM
[H]/N=C\1/CCCN1Cc2c(c(=O)[nH]c(=O)[nH]2)Cl	CMU	1UOU

c1cc(c[n+](c1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO[P@@](=O)([O-])O[P@@](=O)(O)OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)O)O)O)C(=O)N	NAD	1UR5
CNC(=O)[C@H](Cc1c(c(c(c(c1F)F)F)F)F)NC(=O)Nc2nnc(s2)S	IN9	1USN
c1ccc2c(c1)c(c3c([nH+]2)CCCC3)NCCCCCCCC[NH3+]	A8N	1UT6
c1ccc(cc1)CN	ABN	1UTN
C(C(CO)(CO)[NH3+])O	TRS	1UTN
c1ccc(cc1)CC[NH3+]	PEA	1UTO
c1ccc(cc1)CCCCN	PBN	1UTP
Cc1cc(cc(c1)OCCNc2cc[nH+]cc2)NS(=O)(=O)c3ccccc3	I48	1UVT
CCCCn1c(nc2c1ncnc2N)Cc3cc(c(c(c3)OC)OC)OC	PU3	1UY6
CCCCn1c(nc2c1ncnc2N)Cc3ccc(cc3)OC	PU4	1UY7
CCCCn1c(nc2c1ncnc2N)Cc3cccc(c3)OC	PU5	1UY8
CCCCn1c(nc2c1ncnc2N)Cc3ccc4c(c3)OCO4	PU6	1UY9
CCCCn1c(nc2c1ncnc2N)Cc3cc(ccc3OC)OC	PU7	1UYC
CCCCn1c(nc2c1ncnc2N)Cc3cc(c(c(c3Cl)OC)OC)OC	PU8	1UYD
COc1cc(c(c(c1OC)OC)Cl)Cc2nc3c(ncnc3n2CCCC#C)N	PU9	1UYE
COc1cc(c(c(c1OC)OC)Cl)Cc2nc3c(nc(nc3n2CCCC#C)F)N	PU1	1UYF
COc1ccc(c(c1)Cc2[nH]c3c(nc(nc3n2)F)N)OC	PU2	1UYG
CCCCn1c(nc2c1nc(nc2N)F)Cc3cc(ccc3OC)OC	PU0	1UYH
COc1ccc(c(c1)Cc2nc3c(nc(nc3n2CCCC#C)F)N)OC	PUZ	1UYI
CCCCn1c(nc2c1nc(nc2N)F)Cc3ccc4c(c3)OCO4	PUX	1UYK
C1[C@@H]([C@H]([C@@H](C(=O)N1)O)O)CO	IFL	1UZ1
CC(C)[C@H](CO)Nc1nc(c2c(n1)n(cn2)C(C)C)Nc3ccc(c(c3)Cl)C(=O)O	PVB	1V0P
C1[C@H]([C@@H](C(=C[C@]1(C(=O)O)O)F)O)O	FA3	1V1J
[H]/N=C(\c1ccccc1)/N	BEN	1V2J
c1cc(cc2c1cc(cc2)Cl)S(=O)(=O)N3CCN(CC3)C(=O)C4CCN(CC4)c5ccncc5	ZEN	1V2K
[H]/N=C(\c1ccccc1)/N	BEN	1V2L
[H]/N=C(\c1ccccc1)/N	BEN	1V2M
[H]/N=C(\N)/c1ccc(cc1)C[C@H]2C(=O)[C@@H](CCCC2)Cc3ccc(cc3)/C(=N\[H])/N	BBA	1V2N
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)OC)NC(=O)CNS(=O)(=O)c2ccc(cc2)C)\N	ANH	1V2O
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)OC)NC(=O)CNS(=O)(=O)c2ccc(cc2)C)\N	ANH	1V2Q
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)OC)NC(=O)CNS(=O)(=O)c2ccc(cc2)C)\N	ANH	1V2R
[H]/N=C(\c1ccccc1)/N	BEN	1V2S
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)OC)NC(=O)CNS(=O)(=O)c2ccc(cc2)C)\N	ANH	1V2T
[H]/N=C(\c1ccccc1)/N	BEN	1V2U
[H]/N=C(\c1ccccc1)/N	BEN	1V2V
[H]/N=C(/c1cccc(c1)C[C@@H](C(=O)OC)NC(=O)CNS(=O)(=O)c2ccc(cc2)C)\N	ANH	1V2W
C[S@@+](CC[C@@H](C(=O)[O-])N)C[C@@H]1[C@H]([C@H]([C@@H](O1)n2cnc3c2ncnc3N)O)O	SAM	1V2X
CC(=O)N[C@@H]1[C@H](C[C@](O[C@H]1[C@@H]([C@@H](CO)O)O)(C(=O)O)O)O	SLB	1V3C
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	1V3D
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)[C@@H]([C@@H](CO)O)O)C(=O)O	ZMR	1V3E
C[N@@]1CCc2c(sc(n2)C(=O)N3CCN(C[C@@H]3C(=O)N)S(=O)(=O)c4ccc5cc(ccc5c4)Cl)C1	D76	1V3X
c1nc2c(=O)[nH]c(nc2n1CCCCC(F)(F)P(=O)(O)O)N	HA1	1V48
Cn1ccnc1Sc2cc(c(cc2F)N)C(=O)Nc3nccs3	MRK	1V4S
CCC(CC)Nc1cc(ccc1N2C(=O)CC[C@@]2(CN)CO)C(=O)O	IBA	1VCJ
c1cc(ccc1NC(=O)NC2CCCCC2)I	CIU	1VJ5
C/C=C/1\[C@@H]2Cc3c(ccc(=O)[nH]3)[C@]1(CC(=C2)C)N	HUP	1VOT
Cc1ccnc2c1NC(=O)c3cccnc3N2C4CC4	NVP	1VRT
Cc1ccc(c(c1)N[C@@H](c2c(cccc2Cl)Cl)C(=O)N)C(=O)C	AAP	1VRU
[H]/N=C(\c1ccc(cc1)[C@H]2[C@@H]3[C@H]([C@H]4[N@@]2C[C@@H](C4)O)C(=O)N(C3=O)Cc5ccc6c(c5)OCO6)/N	SHY	1VZQ
C1C[C@H]2C(=O)NCC(=O)N2C1	GIO	1W1P
Cc1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)CO)O	THM	1W2G
C1[C@H]([C@@H]([C@H](ON1)CO)O)O	OXZ	1W3J
COc1ccc2c3c1O[C@@H]4[C@@]3(CC[N+](=C2)CCCCCCCCN5C(=O)c6ccccc6C5=O)C=C[C@@H](C4)O	GL8	1W4L
c1cn(c(=O)[nH]c1=O)[C@H]2[C@H]([C@@H]([C@H](O2)CO)OP(=O)(O)O)O	UA3	1W4O
c1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)CO)OP(=O)(O)O	UM3	1W4P
c1cn(c(=O)[nH]c1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)OP(=O)(O)O)F	UMF	1W4Q
[H]/N=N/C(=O)c1ccncc1	ISZ	1W6F
C[N@@]1CC[C@@]23C=C[C@@H](C[C@@H]2Oc4c3c(ccc4OC)C1)O	GNT	1W6R
C[N@@]1CC[C@@]23C=C[C@@H](C[C@@H]2Oc4c3c(ccc4OC)C1)O	GNT	1W76
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@@](=O)(C[P@](=O)(O)OP(=O)(O)O)O)O)O)N	APC	1WC0
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@@](=O)(C[P@](=O)(O)OP(=O)(O)O)O)O)O)N	APC	1WC5

CC(C)(C)c1nnc(o1)C(=O)[C@@H]2CCCN2	PTB	1WM1
c1cc(ccc1c2cc3cc(cc(c3o2)C#N)O)O	697	1X76
c1cc(ccc1c2cc3cc(cc(c3o2)CC#N)O)O	244	1X78
C=Cc1cc(cc2c1oc(n2)c3ccc(c(c3)F)O)O	41	1X7B
c1cc(ccc1c2cc3cc(cc(c3o2)CC#N)O)O	244	1X7E
c1cc(ccc1c2coc3cc(cc(c3c2=O)O)O)O	GEN	1X7J
c1cc(ccc1c2coc3cc(cc(c3c2=O)O)O)O	GEN	1X7R
c1cc(ccc1C[C@@H](C(=O)O)N)O	TYR	1X8X
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C	STI	1XBB
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4ccccc4c5c3c6n2c7ccccc7c6c8c5C(=O)NC8)NC)OC	STU	1XBC
C[C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O[C@@H]2[C@H](O[C@@H]([C@@H]([C@H]2O)O)O)CO)O)O)N[C@H]3C=C([C@H]([C@@H]([C@H]3O)O)O[C@@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O[C@@H]5[C@@H]([C@H]([C@@H]([C@H](O5)CO)O)O)O)O)O)CO	ARE	1XD0
C[C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O[C@@H]2[C@H](O[C@@H]([C@@H]([C@H]2O)O)OC[C@@H]3[C@H]([C@@H]([C@H]([C@@H](O3)O)O)O)O)CO)O)O)N[C@H]4C=C([C@H]([C@@H]([C@H]4O)O)O[C@@H]5[C@@H]([C@H]([C@@H]([C@H](O5)CO)O[C@@H]6[C@@H]([C@H]([C@@H]([C@H](O6)CO)O)O)O)O)O)CO	6SA	1XD1
c1ccc(c(c1)O)O	CAQ	1XEP
c1cc(cc(c1)[N+](=O)[O-])NS(=O)(=O)c2ccsc2C(=O)O	NST	1XGI
c1cc(c(cc1C(=O)O)O)NS(=O)(=O)c2ccsc2C(=O)O	HTC	1XGJ
[H]/N=C(\c1cccc(c1)c2cccc(c2)[C@H](CCCNc3ccncc3)C(=O)O)/N	4PP	1XKA
[H]/N=C(\c1cccc(c1)c2cccc(c2)[C@H](CCCNc3ccncc3)C(=O)O)/N	4PP	1XKB
CS(=O)(=O)CCNCc1ccc(o1)c2ccc3c(c2)c(ncn3)Nc4ccc(c(c4)Cl)OCc5cccc(c5)F	FMM	1XKK
COc1ccc(cc1OC2CCCC2)C3(CCC(CC3)C(=O)O)C#N	CIO	1XLX
C/C(=N\OC(=O)N)/c1ccc(c(c1)OC2CCCC2)OC	FIL	1XLZ
COc1ccc(cc1OC2CCCC2)C(=O)Nc3c(cncc3Cl)Cl	PIL	1XM4
CCCOc1cc(ccc1OC)[C@@]2(CNC(=O)O2)C	5RM	1XM6
c1cc(c(cc1C(=O)Nc2c(cncc2Cl)Cl)OCC3CC3)OC(F)F	ROF	1XMU
COc1ccc(cc1OC2CCCC2)[C@H]3CC(=O)NC3	ROL	1XMY
COc1ccc(cc1OC2CCCC2)[C@H]3CC(=O)NC3	ROL	1XN0
CC(C)C[C@@H]([C@H]1[C@@H](C[C@@H](N1)C(=O)O)C(=O)OC)NC(=O)C	ABX	1XOE
CC(C)C[C@@H]([C@H]1[C@@H](C[C@@H](O1)C(=O)O)C(=O)OC)NC(=O)C	ABW	1XOG
COc1ccc(cc1OC2CCCC2)C3(CCC(CC3)C(=O)O)C#N	CIO	1XOM
COc1ccc(cc1OC2CCCC2)C(=O)Nc3c(cncc3Cl)Cl	PIL	1XON
c1cc(c(cc1C(=O)Nc2c(cncc2Cl)Cl)OCC3CC3)OC(F)F	ROF	1XOQ
COc1cc(ccc1OC(F)F)c2ccc(=O)[nH]n2	ZAR	1XOR
CCCc1c2c(c(=O)[nH]c(n2)c3cc(ccc3OCC)S(=O)(=O)N4CCN(CC4)C)n(n1)C	VIA	1XOS
CCCc1nc(c2n1[nH]c(nc2=O)c3cc(ccc3OCC)S(=O)(=O)N4CCN(CC4)CC)C	VDN	1XOT
C[C@@]1(CC[C@@H]2[C@@]1(C=CC3=C4CCC(=O)C=C4CC[C@@H]23)C)O	R18	1XOW
CN1CC(=O)N2[C@@H](C1=O)Cc3c4ccccc4[nH]c3[C@H]2c5ccc6c(c5)OCO6	CIA	1XOZ
CCCc1nc(c2n1[nH]c(nc2=O)c3cc(ccc3OCC)S(=O)(=O)N4CCN(CC4)CC)C	VDN	1XP0
C[C@@]1(CC[C@@H]2[C@@]1(C=CC3=C4CCC(=O)C=C4CC[C@@H]23)C)O	R18	1XQ3
c1ccc(cc1)[N@@]2CCc3cc(ccc3[C@H]2c4ccc(cc4)[N@]5CC[N@@]6CCCC[C@H]6C5)O	AEJ	1XQC
COc1ccc(cc1Nc2ncc(o2)c3cccc(c3)c4cccnc4)S(=O)(=O)NCC5CC5	AAX	1Y6B
CC(=O)NO	HAE	1Y93
CC(=O)Nc1nnc(s1)S(=O)(=O)N	AZM	1YDB
C[C@H]1CNCCN1S(=O)(=O)c2cccc3c2ccnc3	IQP	1YDR
[H]/N=C(/c1ccc(cc1O)N[C@H](c2cc(cc(c2F)OCC)OCC)C(=O)NS(=O)(=O)c3cccc(c3)N)\N	905	1YGC
C[C@@H]1c2c(ccc(c2F)O)O[C@H]([C@H]1c3ccc(cc3)O)c4ccc(cc4)OCCN5CCCCC5	CM3	1YIN
Cc1cc(ccc1F)S(=O)(=O)N[C@H](C2CCOCC2)C(=O)NO	915	1YQY
Cc1ccc2c(c1)C(=O)[C@@]3(CCN(C3=N2)c4ccccc4)O	BIT	1YV3
c1ccc(cc1)C(=O)N/C(=C\c2ccc(cc2)Oc3ccccc3Br)/C(=O)O	PH7	1YVF
C[C@@H](c1ccccc1)Nc2nccc(n2)c3c(c(=O)n(n3C)C4CCNCC4)c5ccc(cc5)F	LI9	1YWR
c1cc(ccc1c2ccc3c(c2)ccc(c3Cl)O)O	4NA	1YY4
c1cc(cc2c1cc(cc2C#N)c3ccc(c(c3)F)O)O	196	1YYE
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)O)O)O)N	AMP	1YZ0
C[N@@](CC[C@@H](C(=O)O)N)C[C@@H]1[C@H]([C@H]([C@@H](O1)n2cnc3c2ncnc3N)O)O	SA8	1Z3C
CN(C)Cc1nccn1c2ccc(c(c2)F)NC(=O)c3cc(nn3c4ccc5c(c4)c(no5)N)C(F)(F)F	IK8	1Z6E
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)O)O)O)N	AMP	1Z6S
C[C@H](c1cccc(c1)Cl)NC(=O)Cc2c(ccc([n+]2[O-])NCC(c3cccc[n+]3[O-])(F)F)Cl	L17	1Z71
C[C@](CS(=O)(=O)c1ccc(cc1)F)(C(=O)Nc2ccc(c(c2)C(F)(F)F)C#N)O	198	1Z95
c1ccc(cc1)C[C@H](CS)NC(=O)CC(=O)O	RRT	1Z9G
c1cc(ccc1C2=C(c3ccc(cc3C2=O)O)Br)O	789	1ZAF
c1ccc(cc1)C[C@H](CS)C(=O)NCC(=O)O	TIO	1ZDP
c1ccc2c(c1)c(c3c(n2)CCCC3)NCCCCCCCCCCN[C@H]4CCCc5c4ccc(=O)[nH]5	A2E	1ZGC
c1ccc2c(c1)-c3c[n+](ccc3[C@]2(C(=O)N4CCC[C@H]4C(=O)NCc5cc(ccc5CN)Cl)O)[O-]	62	1ZRB

C[C@]12CCC(=O)C=C1CC[C@@H]3[C@@H]2[C@H](C[C@]4([C@H]3CC[C@@H]4C(=O)CO)C=O)O	AS4	2AA2
CC(=O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4=CC(=O)CC[C@]34C)C	STR	2AA5
CC(=O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4=CC(=O)CC[C@]34C)C	STR	2AA6
C[C@]12CC[C@H]3[C@H]([C@@H]1CC[C@@H]2C(=O)CO)CCC4=CC(=O)CC[C@]34C	1CA	2AA7
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2GP	2AAD
C[C@]12CC(=O)[C@H]3[C@H]([C@@H]1CC[C@@]2(/C(=C/O)/O)O)CCC4=CC(=O)C=C[C@]34C	PDN	2AAX
CC(=O)S[C@@H]1CC2=CC(=O)CC[C@@]2([C@@H]3[C@@H]1[C@@H]4CC[C@]5([C@]4(CC3)C)CCC(=O)O5)C	SNL	2AB2
CC[N+](C)(C)c1cccc(c1)O	EDR	2ACK
c1nc2c(c(n1)O)NC[N@]2[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O	HPR	2ADA
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)O)O)O)N	AMP	2AK3
C[C@@]1(CC[C@@H]2[C@@]1(C=CC3=C4CCC(=O)C=C4CC[C@@H]23)C)O	R18	2AO6
C[C@H](CCC(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)O)C	CHD	2AZY
C[C@H](CCC(=O)NCCS(=O)(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)O)C	TCH	2AZZ
C[C@H](CCC(=O)NCC(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)O)C	GCH	2B00
C[C@H](CCC(=O)NCCS(=O)(=O)O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)C	TUD	2B01
CC(=O)N[C@@H]1[C@H](C[C@@](O[C@H]1[C@@H]([C@@H](CO)O)O)(C(=O)O)O)O	SIA	2BAT
Cc1cc(cc(c1)Oc2c(c([nH]c(=O)c2I)C)CSCc3ccco3)C	R22	2BE2
c1ccc(cc1)[N@@]2CC(=O)NS2(=O)=O	T2D	2BGE
c1ccc(cc1)CCc2cc(cnc2)C(=O)N3CCC(CC3)c4cccc(c4)CN	PM2	2BM2
[H]/N=C(/c1ccc(cc1)[C@H]2[C@@H]3[C@H]([C@H]4[N@@]2CCC4)C(=O)N(C3=O)CCC[N+](C)(C)C)\N	784	2BOK
CC(C)N1CCC(CC1)NC(=O)c2cc3ccccc3n2Cc4cccc(c4)OC	IID	2BQ7
COc1ccc(cc1)c2c3c(ncnc3oc2c4ccc(cc4)OC)NCCO	PFP	2BR1
CCNC(=O)c1c(c([nH]n1)c2cc(c(cc2O)O)Cl)c3ccc(cc3)OC	BSM	2BSM
C1[C@@H]([C@H]([C@@H](CN1)O[C@H]2[C@@H]([C@H]([C@@H]([C@H](O2)CO)O)O)O)O)CO	ISX	2BVD
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@@](=O)(C[P@](=O)(O)OP(=O)(O)O)O)O)O)N	APC	2BW7
c1ccc(cc1)CN	ABN	2BZA
c1ccc(cc1)C(c2ccccc2)N/C(=N/CC(=O)O)/Nc3ccc(cc3)C#N	GAS	2CGR
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)c3ccc4cc(ccc4c3)Cl	GSK	2CJI
c1ccc2c(c1)c(c3c(n2)CCCC3)NCCCCCCCNc4c5ccccc5nc6c4CCCC6	AA7	2CKM
c1ccc2c(c1)c(c3c(n2)CCCC3)NCCCCCNc4c5ccccc5nc6c4CCCC6	F11	2CMF
Cc1ccc(cc1Nc2nccc(n2)c3cncnc3)NC(=O)c4ccc(c(c4)C(F)(F)F)C[N@@]5CC[C@@H](C5)N(C)C	406	2E2B
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@@](=O)(NP(=O)(O)O)O)O)O)N	ANP	2EB3
CN(C)CCOc1ccc(cc1)/C(=C(\CCCO)/c2ccccc2)/c3ccc(cc3)O	TXF	2EWP
[H]/N=C(\N)/N[C@@H]1C[C@@H]([C@H]([C@H]1[C@H](C(CC)CC)NC(=O)C)O)C(=O)O	BCZ	2F10
Cc1cc([nH]n1)Nc2cc(nc(n2)Sc3ccc(cc3)NC(=O)C4CC4)N5CCN(CC5)C	VX6	2F4J
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	2F80
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	2F81
Cc1nc(cs1)COc2ccc(cc2)C[C@@H]([C@@H](CN(CC(C)C)S(=O)(=O)c3ccc4c(c3)OCO4)O)NC(=O)O[C@H]5CO[C@@H]6[C@H]5CCO6	385	2FDD
Cc1nc(cs1)COc2ccc(cc2)C[C@@H]([C@@H](CN(CC(C)C)S(=O)(=O)c3ccc4c(c3)OCO4)O)NC(=O)O[C@H]5CO[C@@H]6[C@H]5CCO6	385	2FDE
COc1ccccc1C2CCN(CC2)Cc3cnc(c4c3cc(c(c4)OC)OC)Cc5ccc(c(c5)OC)OC	RO0	2FVJ
c1ccc(c(c1)C[N@]2CC[C@H](C2)O)c3ccc(cc3)N4CCc5c(n(nc5C(F)(F)F)c6ccc7c(c6)c(no7)N)C4=O	5QC	2FZZ
CN(C)Cc1ccccc1c2ccc(cc2)N3CCc4c(n(nc4C(F)(F)F)c5cccc(c5)C(=O)N)C3=O	4QC	2G00
C([C@@H]1[C@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O)O	BGC	2GBP
CCCC[C@@]12CCC(=O)C(=C1c3ccc(cc3C2)O)Br	FBR	2GIU
Cc1cccc(c1NC(=O)c2cnc(s2)Nc3cc(nc(n3)C)N4CCN(CC4)CCO)Cl	1N1	2GQG
CCO[C@@H](Cc1ccc2c(c1)ccn2Cc3c(oc(n3)c4ccccc4Cl)C)C(=O)O	208	2GTK
CC(=O)Nc1nnc(s1)S(=O)(=O)N	AZM	2H4N
C[N+](C)(C)CCOC(=O)CCC(=O)OCC[N+](C)(C)C	SCK	2HA2
CCC[C@H]1[C@@H](C/C(=C/C=C/2\CCC[C@]3([C@H]2CC[C@@H]3[C@H](C)CCCC(C)(C)O)C)/C(=C)[C@H]1O)O	C33	2HAM

C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H]([C@@H]([C@@H](C3=C)O)OCCCO)O)C	OCC	2HAR
CCCO[C@H]1[C@@H](C/C(=C/C=C/2\CCC[C@]3([C@H]2CC[C@@H]3[C@H](C)CCCC(C)(C)O)C)/C(=C)[C@H]1O)O	C3O	2HAS
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H]([C@@H]([C@@H](C3=C)O)CCCO)O)C	O1C	2HB7
C[C@H]1[C@@H](C/C(=C/C=C/2\CCC[C@]3([C@H]2CC[C@@H]3[C@H](C)CCCC(C)(C)O)C)/C(=C)[C@H]1O)O	MVD	2HB8
C[C@]12CCC/C(=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)/[C@@H]1CC[C@@H]2C#CC#CC(C)(C)O	XE4	2HBH
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	VDX	2HC4
C[C@]12CCC/C(=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)/[C@@H]1CC[C@@H]2C(CCCC(C)(C)O)CCCC(C)(C)O	BIV	2HCD
Cc1ccnc2c1NC(=O)c3cccnc3N2C4CC4	NVP	2HND
Cc1ccnc2c1NC(=O)c3cccnc3N2C4CC4	NVP	2HNY
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	2HS2
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	2HT7
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	2HT8
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)[C@@H]([C@@H](CO)O)O)C(=O)O	ZMR	2HTQ
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	2HTR
[H]/N=C(\N)/N[C@@H]1C[C@@H]([C@H]([C@H]1[C@H](C(CC)CC)NC(=O)C)O)C(=O)O	BCZ	2HTU
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	2HTW
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	2HU0
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	2HU4
CCCc1c2ccc(cc2ccc1OCCCn3ccc4c3ccc(c4)OCC(=O)O)C(=O)c5ccccc5	DRY	2HWQ
CCCc1c2ccc(cc2ccc1OCCCn3ccc4c3cccc4OC(C)(C)C(=O)O)C(=O)c5ccccc5	DRD	2HWR
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C	STI	2HYY
CCCc1ccc(cc1C(F)(F)F)NC(=O)c2ccccc2NCc3ccc(=O)[nH]c3	GIN	2HZ0
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4c(c5c3c6n2c7ccccc7c6c8c5C(=O)NC8)CCCC4)NC)OC	4ST	2HZ4
Cc1cc(ccc1F)Nc2ncc3cc(c(=O)n(c3n2)C)c4c(cccc4Cl)Cl	JIN	2HZI
c1cc(cc(c1)NC(=O)Nc2ccc(cc2)Oc3ccncc3)C(F)(F)F	KIN	2HZN
c1cc(ccc1[C@H]2[C@H]3CCC[C@H]3c4cc(ccc4O2)O)O	I0G	2I0G
c1cc(ccc1[C@H]2[C@H]3CCC[C@H]3c4cc(ccc4O2)O)O	I0G	2I0J
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	2IDW
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	2IEN
CC(C)CN(C[C@H]([C@H](Cc1ccccc1)NC(=O)O[C@H]2CO[C@@H]3[C@H]2CCO3)O)S(=O)(=O)c4ccc(cc4)N	17	2IEO
COc1cc2c(cc1OCCCN3CCOCC3)/c(=N/c4ccc(c(c4)Cl)F)/nc[nH]2	IRE	2ITO
CCN1CCN(CC1)Cc2ccc(cc2)c3cc4c([nH]3)ncnc4N[C@H](C)c5ccccc5	AEE	2ITP
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4c(c5c3c6n2c7ccccc7c6c8c5C(=O)N=C8)CCCC4)NC)OC	ITQ	2ITQ
CCN1CCN(CC1)Cc2ccc(cc2)c3cc4c([nH]3)ncnc4N[C@H](C)c5ccccc5	AEE	2ITT
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4c(c5c3c6n2c7ccccc7c6c8c5C(=O)N=C8)CCCC4)NC)OC	ITQ	2ITU
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@@](=O)(NP(=O)(O)O)O)O)O)N	ANP	2ITV
C[C@@]12[C@@H]([C@@H](C[C@@H](O1)n3c4c(c5c3c6n2c7ccccc7c6c8c5C(=O)N=C8)CCCC4)NC)OC	ITQ	2ITW
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@@](=O)(NP(=O)(O)O)O)O)O)N	ANP	2ITX
COc1cc2c(cc1OCCCN3CCOCC3)/c(=N/c4ccc(c(c4)Cl)F)/nc[nH]2	IRE	2ITY
[H]/N=C/1\N[C@H]2CS[C@H]([C@H]2N1)CCCCC(=O)O	IMI	2IZL
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)c3cc4cc(ccc4[nH]3)Cl	GSQ	2J2U
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)c3cc4ccc(cc4s3)Cl	GS6	2J34
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)c3cc4cc(ccc4s3)Cl	GS5	2J38
CCN1CCN(CC1)Cc2ccc(cc2)c3cc4c([nH]3)ncnc4N[C@H](C)c5ccccc5	AEE	2J6M
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)c3nc([nH]n3)c4ccc(s4)Cl	G15	2J94
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)c3ccc(s3)c4ccc(s4)Cl	GSX	2J95
COCc1cc(cc2c1O[C@H]([C@@H]3[C@H]2CCC3)c4ccc(cc4)O)O	JJ3	2JJ3
c1ccc2c(c1)c(ccc2c3ccc(cc3)O)/C=N/O	555	2NV7
C/C(=C\1/CC[C@@H]\2[C@@]1(CCC/C2=C\C=C3C[C@H](C(=C)[C@@H](C3)O)O)C)/CCCC(C)(C)O	VD5	2O4R

CC[C@H]1C(=O)Nc2ccc(cc2N1C(=O)OC(C)C)F	HBQ	2OPS
c1ccc(c(c1)Cl)Cl	YAN	2OTY
CNc1ccccc1	1MR	2OTZ
Cn1cccc1	MR3	2OU0
c1ccc2c(c1)c(c[nH]2)CC(=O)O	IAC	2OYF
COc1ccc(cc1)n2c3c(c(n2)C(=O)N)CCN(C3=O)c4ccc(cc4)N5CCCCC5=O	GG2	2P16
CNc1nccn1Cc2csc(c2Cl)C(=O)Nc3c(cc(cc3OC)Cl)C(=O)Nc4ccc(cn4)Cl	993	2P3T
Cn1ccnc1Cc2csc(c2Cl)C(=O)Nc3c(cc(cc3OC)Cl)C(=O)Nc4ccc(cn4)Cl	663	2P3U
Cc1cc(ccc1Cl)O	43M	2P7A
CC(C)(c1ccc(cc1)O)c2ccc(cc2)O	2OH	2P7G
CC/C(=C(\c1ccc(cc1)O)/c2ccc(cc2)OCCN(C)C)/c3ccccc3	OHT	2P7Z
c1ccn(c(=O)c1)c2ccc(cc2)C(=O)NCCNC(=O)c3ccc(s3)Cl	ME1	2P93
c1ccn(c(=O)c1)c2ccc(cc2)C(=O)N[C@H]3CCCC[C@H]3NC(=O)c4ccc5c(c4)[nH]cc5Cl	ME4	2P94
c1ccn(c(=O)c1)c2ccc(cc2)C(=O)N[C@H]3CCC[C@H]3NC(=O)c4ccc(s4)Cl	ME5	2P95
c1ccc(cc1)C2(CCCCC2)N3CCCCC3	1PC	2PCP
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@](=O)(O)OC[C@@H]4[C@H]([C@H]([C@@H](O4)O)O)O)O)O)N	APR	2PHH
Cc1ccc(cc1)n2cc(c3c2cccc3)CNCC4CCCCC4	47	2PJL
C(CCC(=O)O)CCN	ACA	2PK4
CO[C@@H]1C[C@@H](N(C1)C(=O)Nc2ccc(cc2)Cl)C(=O)Nc3ccc(cc3F)c4ccccc4S(=O)(=O)C	237	2PR3
CCCCN(C(=O)Nc1ccc(cc1)Cl)C2(CC2)C(=O)Nc3ccc(cc3F)c4ccccc4S(=O)(=O)C	FXI	2Q1J
CC(C)n1c(c(c(c1C(=O)Nc2ccccc2)c3ccc(cc3)F)c4ccc(cc4)F)CC[C@H](C[C@H](CC(=O)O)O)O	882	2Q1L
Cc1c(c2cc(ccc2n1C(=O)c3ccc(cc3)OC)OC(F)(F)F)Cc4cccc(c4)O[C@@H](C)C(=O)O	241	2Q5P
c1ccc(cc1)Sc2c3cc(ccc3n(c2C(=O)O)Cc4ccc(cc4)Cl)Cl	NZA	2Q5S
c1ccc(cc1)Cn2c3ccc(cc3c(c2C(=O)O)Sc4ccccc4)Cl	SF1	2Q61
CC(C)n1c(c(c2c1C(=O)N(CCC2)c3ccccc3)c4ccc(cc4)F)CC[C@H](C[C@H](CC(=O)O)O)O	HR2	2Q6B
CC(C)n1c(c(c2c1c(=O)n(c3c2cccc3)c4ccccc4)c5ccc(cc5)F)CC[C@H](C[C@H](CC(=O)O)O)O	HR1	2Q6C
COc1cccc(c1)Cn2c3ccc(cc3c(c2C(=O)O)Sc4ccccc4)Cl	SF2	2Q6R
c1cnc(nc1)Oc2ccc(c(c2)C(=O)O)NC(=O)c3ccc(cc3Cl)Cl	PLB	2Q6S
c1cc2c(cc1O)[nH]nc2c3ccc(cc3O)O	KN2	2QA6
COCc1cc(cc2c1O[C@H]([C@@H]3[C@H]2CCC3)c4ccc(cc4)O)O	JJ3	2QE4
C[C@]12C[C@@H]([C@@H]3c4ccc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O)COC	EED	2QGT
c1cc(ccc1n2c(c3cc(ccc3n2)O)Cl)O	EES	2QGW
CCOC(=O)[C@@H]1[C@@H]([C@@H]2C(=C([C@H]1O2)c3ccc(cc3)O)c4ccc(cc4)O)C(=O)OCC	ODE	2QH6
CN(C)C(=O)c1cc(cnc1)c2cc3c(c[nH]c3nc2)c4ccccc4OC	P3Y	2QOH
COc1ccc2cc-3[n+](cc2c1OC)CCc4c3cc5c(c4)OCO5	BER	2QVD
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	2QWC
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)N	4AM	2QWD
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)[C@@H]([C@@H](CO)O)O)C(=O)O	ZMR	2QWE
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)C(=O)N(C)CCC)C(=O)O	G20	2QWF
CCN(CC)C(=O)[C@H]1[C@@H]([C@H](C=C(O1)C(=O)O)N)NC(=O)C	G28	2QWG
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	2QWH
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)C(=O)N(C)CCC)C(=O)O	G20	2QWI
CCN(CC)C(=O)[C@H]1[C@@H]([C@H](C=C(O1)C(=O)O)N)NC(=O)C	G28	2QWJ
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	2QWK
Cn1c2cc(cnc2nc1N)c3ccccc3	PIQ	2QXM
C[C@H](c1ccccc1)N(C)C(=O)c2c(c(n(n2)c3ccc(cc3)F)CC[C@H](C[C@H](CC(=O)O)O)O)C(C)C	RIE	2R4F
CC1CCN(CC1)CCOc2ccc(cc2)C(=O)c3c4ccc(cc4sc3c5ccc(cc5)O)O	LLB	2R6W
c1cc(ccc1c2c(c3ccc(cc3s2)O)C(=O)c4ccc(cc4)OCCN5CCCC5)O	LLC	2R6Y
CN(C)Cc1nccn1c2ccc(c(c2)F)c3ccc4c(c3F)n(nc4C(=O)N)c5ccc6c(c5)c(no6)N	JNJ	2RA0
c1cc2c(cc1Br)c(c([nH]2)C(=O)N)S(=O)(=O)N3CCCC3	MRX	2RF2
c1ccc(cc1)Cn2c(nnc2SCc3ccccc3Cl)c4cccs4	TT1	2RKI
CC(=O)N[C@@H]1[C@H](C=C(O[C@H]1[C@@H]([C@@H](CO)O)O)C(=O)O)O	DAN	2SIM
c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@@](=O)(NP(=O)(O)O)O)O)O)N	ANP	2SRC
CC[C@]1([C@H](C1(Cl)Cl)C)C(=O)N[C@H](C)c2ccc(cc2)Cl	CRP	2STD
CC(C)C[C@@H](C(=O)N)NP(=O)(O)O	0FA	2TMN
c1ccc(cc1)COC(=O)N[C@H](CC(=O)Nc2nnc(s2)S)c3ccccc3	IN8	2USN
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)/C=C/c3ccc(s3)Cl	895	2UWL
C[C@@H](C(=O)N1CCOCC1)N2CC[C@@H](C2=O)NS(=O)(=O)/C=C(\C)/c3ccc(s3)Cl	701	2UWO
CC(C)(C)NC[C@@H](COc1cccc2c1CC(=N2)C#N)O	P32	2VT4
C[C@H](CCc1ccc(cc1)O)NCCc2ccc(c(c2)O)O	Y00	2Y00
C[C@H](CCc1ccc(cc1)O)NCCc2ccc(c(c2)O)O	Y00	2Y01
C[C@H](Cc1ccc(cc1)OC)NC[C@@H](c2ccc(c3c2ccc(=O)[nH]3)O)O	WHJ	2Y02
CC(C)NC[C@@H](c1ccc(c(c1)O)O)O	5FW	2Y03
CC(C)(C)NC[C@@H](c1ccc(c(c1)CO)O)O	68H	2Y04

C(C(=O)O)OP(=O)(O)O	PGA	2YPI
c1cc(ccc1[C@H]2[C@H]3CC(C[C@H]3c4cc(ccc4O2)O)(F)F)O	DC8	2Z4B
CN(C)C(=O)c1cc(cnc1)c2cc3c(c[nH]c3nc2)c4ccccc4OC	P3Y	2Z60
CC(C)(c1ccccc1)c2ccc(cc2)O	1OH	2ZAS
Cc1cc(cc(c1Nc2ccnc(n2)Nc3ccc(cc3)C#N)C)/C=C/C#N	T27	2ZD1
CCC(CC)(c1ccc(c(c1)C)OC[C@H](CO)O)c2ccc(c(c2)C)OC[C@@H](C(C)(C)C)O	YR3	2ZFX
CCC(CC)(CS[C@@H](C)C1=CC[C@@H]\2[C@@]1(CCC/C2=C\C=C3C[C@H](C([C@@H](C3)O)OCCO)O)C)O	VDA	2ZL9
CCC(CC)(CS[C@H](C)C1=CC[C@@H]\2[C@@]1(CCC/C2=C\C=C3C[C@H](C([C@@H](C3)O)OCCO)O)C)O	VDB	2ZLA
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	VDX	2ZLC
c1cc(ccc1CN2CCCC2)N/C=C\3/c4cc(ccc4C(=O)NC3=O)Br	575	2ZM3
C[C@H](/C=C/[C@H](C12CC3CC(C1)CC(C3)C2)O)[C@H]4CC[C@@H]\5[C@@]4(CCC/C5=C\C=C6C[C@H](C(=C)[C@@H](C6)O)O)C	NYA	2ZMH
C[C@@H](/C=C/C[C@H](C12CC3CC(C1)CC(C3)C2)O)[C@H]4CC[C@@H]\5[C@@]4(CCC/C5=C\C=C6C[C@H](C(=C)[C@@H](C6)O)O)C	TT2	2ZMI
C[C@@H](/C=C/C[C@H](CC12CC3CC(C1)CC(C3)C2)O)[C@H]4CC[C@@H]\5[C@@]4(CCC/C5=C\C=C6C[C@H](C(=C)[C@@H](C6)O)O)C	MI4	2ZMJ
CCCC[C@@H](CCO)[C@@H](C)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	JB1	2ZXM
CCCC[C@@H](CCO)[C@H](C)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	JC1	2ZXN
C[C@@H](C[C@H]1CC(=C)C(=O)O1)[C@H]2CC[C@@H]\3[C@@]2(CCC/C3=C\C=C/4\C[C@H](C[C@@H](C4=C)O)O)C	TEJ	3A2H
C[C@@H](C[C@H]1CC(=C)C(=O)O1)[C@H]2CC[C@@H]\3[C@@]2(CCC/C3=C\C=C/4\C[C@H](C[C@@H](C4=C)O)O)C	TEJ	3A2I
C[C@@H](C[C@H]1CC(=C)C(=O)O1)[C@H]2CC[C@@H]\3[C@@]2(CCC/C3=C\C=C/4\C[C@H](C[C@@H](C4=C)O)O)C	TEJ	3A2J
C[C@H]1[C@@H](C/C(=C/C=C/2\CCC[C@]3([C@H]2CC[C@@H]3[C@@]4(C[C@@H](CO4)CC(C)(C)O)C)C)/C(=C)[C@H]1O)O	2MV	3A3Z
C[C@H]1[C@@H](C/C(=C/C=C/2\CCC[C@]3([C@H]2CC[C@@H]3[C@@]4(C[C@H](CO4)CC(C)(C)O)C)C)/C(=C)[C@H]1O)O	23R	3A40
C[C@H](CCCC(C)(C)O)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@@H](C[C@@H](C3=C)O)O)C	3EV	3A78
CCCC[C@@H](C[C@H](C(C)C)O)[C@@H](C)[C@H]1CC[C@@H]\2[C@@]1(CCC/C2=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)C	ICJ	3AFR
CC(=O)N[C@@H]1[C@H](C[C@](O[C@H]1[C@@H]([C@@H](CO)O)O)(C(=O)O)O)O	SLB	3B50
CC(C)c1c(c(c(n1CC[C@H](C[C@H](CC(=O)O)O)O)c2ccc(cc2)F)c3ccccc3)S(=O)(=O)N4CCOCC4	RID	3BGL
Cc1cc(cc(c1Nc2ccnc(n2)Nc3ccc(cc3)C#N)C)/C=C/C#N	T27	3BGR
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2GP	3BIR
CC(C)c1c(c(c(n1CC[C@H](C[C@H](CC(=O)O)O)O)c2ccc(cc2)F)c3ccccc3)C(=O)Nc4ccccc4O	3HI	3CCT
CC(C)c1c(nc(n1CC[C@H](C[C@H](CC(=O)O)O)O)c2ccc(cc2)F)C(=O)NCc3ccccc3	4HI	3CCW
CC(C)c1c(nc(n1CC[C@H](C[C@H](CC(=O)O)O)O)c2ccc(cc2)F)C(=O)N[C@H](CO)c3ccccc3	5HI	3CCZ
CC(C)n1c(c(nc1C(=O)NCc2ccc(cc2)F)c3ccc(cc3)F)CC[C@H](C[C@H](CC(=O)O)O)O	6HI	3CD0
CCc1c(c2c(n1CC[C@H](C[C@H](CC(=O)O)O)O)CCCCC2)C(=O)Nc3ccc(cc3)c4ccccc4	7HI	3CD5
CC(C)n1c(c(c(c1C(=O)Nc2ccccc2)c3ccc(cc3)F)c4ccc(cc4)F)CC[C@H](C[C@H](CC(=O)O)O)O	882	3CD7
CC(C)n1c(c(c(c1C(=O)Nc2ccc(cc2)S(=O)(=O)N)c3ccccc3)c4ccc(cc4)F)CC[C@H](C[C@H](CC(=O)O)O)O	8HI	3CDA
CC(C)c1c(c(c(n1CC[C@H](C[C@H](CC(=O)O)O)O)c2ccc(cc2)F)c3ccc(cc3)F)S(=O)(=O)Nc4ccc(cc4)C(=O)N	9HI	3CDB
[H]/N=C(\N)/N[C@H]1C=C(O[C@H]([C@@H]1NC(=O)C)[C@@H]([C@@H](CO)O)O)C(=O)O	ZMR	3CKZ
CCC(CC)O[C@@H]1C=C(C[C@@H]([C@H]1NC(=O)C)N)C(=O)O	G39	3CL0
c1cc(ccc1[C@H]([C@@H](CO)NC(=O)C(Cl)Cl)O)[N+](=O)[O-]	CLM	3CLA
C[C@]12CCC/C(=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)/[C@@H]1CC[C@@H]2[C@@]4(C[C@@H](CO4)CC(C)(C)O)C	COV	3CS4
C[C@]12CCC/C(=C\C=C/3\C[C@H](C[C@@H](C3=C)O)O)/[C@@H]1CC[C@@H]2[C@@]4(C[C@H](CO4)CC(C)(C)O)C	0CO	3CS6
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)C(=O)Nc4cc(cc(c4)n5cc(nc5)C)C(F)(F)F	NIL	3CS9
CN1CCN(CC1)C2CC(C2)c3nc(c4n3ccnc4N)c5ccc6ccc(nc6c5)c7ccccc7	D94	3D94
CN(Cc1cnc2c(n1)c(nc(n2)N)N)c3ccc(cc3)C(=O)N[C@@H](CCC(=O)O)C(=O)O	MTX	3DFR
c1ccc(cc1)C(=O)c2cc(ccc2OCC(=O)Nc3ccccc3)Cl	GFA	3DLE
Cc1cc(ccc1NC(=O)COc2ccc(cc2C(=O)c3cc(cc(c3)F)C(F)(F)F)Cl)S(N)(O)O	GWE	3DLG
CCC(=O)NS(=O)(=O)c1ccc(c(c1)C)NC(=O)COc2ccc(cc2C(=O)c3cc(cc(c3)Cl)C#N)Cl	GWI	3DOL
C[C@@]1(CC[C@H](C1(C)C)/C=C/C=C2C[C@H](C[C@@H](C2)O)O)CCC#CC(C(F)(F)F)(C(F)(F)F)O	C5D	3DR1
Cc1cc2cc(ccc2c(c1c3cccc(c3)O)Oc4ccc(cc4)O)O	369	3DT3

c1cc(c(c(c1Cc2c3ccnnc3[nH]n2)F)Oc4cc(cc(c4)Cl)C#N)Br	PZL	3DYA
CC(C)C[C@H](C(=O)O)NS(=O)(=O)c1ccc(cc1)c2ccccc2	BDL	3EHX
C[C@H](C(=O)O)NS(=O)(=O)c1ccc(cc1)OC	TBL	3EHY
CC/C(=C(/CC)\c1ccc(cc1)O)/c2ccc(cc2)O	DES	3ERD
c1cc(ccc1c2c(n(cn2)C3CCNCC3)c4ccnc(n4)N)F	SB4	3ERK
CC/C(=C(\c1ccc(cc1)O)/c2ccc(cc2)OCCN(C)C)/c3ccccc3	OHT	3ERT
COc1ccc2c(c1)c(c[nH]2)CCC(=O)O	ET0	3ET0
COc1ccc(cc1)S(=O)(=O)n2cc(c3c2ccc(c3)OC)CCC(=O)O	ET1	3ET1
COc1ccc(cc1)S(=O)(=O)n2cc(c3c2ccc(c3)OC)CCC(=O)O	ET1	3ET2
COc1ccc(cc1)S(=O)(=O)n2cc(c3c2ccc(c3)OC)CCC(=O)O	ET1	3ET3
COc1ccc(cc1)S(=O)(=O)N(C[C@@H](CO)O)CC(=O)N=O	HS1	3F15
COc1ccc(cc1)S(=O)(=O)N[C@H](CO)C(=O)N=O	HS3	3F16
c1ccc(cc1)c2ccc(cc2)S(=O)(=O)NCC(=O)NO	HS4	3F17
c1cc(ccc1F)S(=O)(=O)N(CCO)CC(=O)N=O	HS5	3F18
c1cc(ccc1F)S(=O)(=O)NCC(=O)N=O	HS6	3F19
c1ccc(cc1)S(=O)(=O)NCC(=O)N=O	HS7	3F1A
Cn1ccnc1Sc2ccc(cc2Cl)Nc3c4cc(c(cc4ncc3C#N)OCCC[N@@](C)CCO)OC	741	3F5P
CCO[C@@H](Cc1ccc(cc1C)OCc2csc(n2)c3ccc(cc3)Cl)C(=O)O	CTM	3FEI
CCO[C@@H](Cc1ccc(cc1C)OCc2csc(n2)c3ccc(cc3)Cl)C(=O)O	CTM	3FEJ
c1ccc2c(c1)cc(cn2)Oc3c(cc(cc3Cl)NS(=O)(=O)c4ccc(cc4Cl)Cl)Cl	Z12	3FUR
Cc1c(c2cc(ccc2n1C(=O)c3ccc(cc3)Cl)OC)CC(=O)O	IMN	3H1X
c1cc(c[n+](c1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO[P@@](=O)([O-])O[P@@](=O)(O)OC[C@@H]3[C@H]([C@H]([C@@H](O3)n4cnc5c4ncnc5N)O)O)O)O)C(=O)N	NAD	3HAD
COC1=CC(=O)C=CC1=O	MCW	3HSW
Cc1ccc(cc1C#Cc2cnc3n2nccc3)C(=O)Nc4ccc(c(c4)C(F)(F)F)CN5CCN(CC5)C	0LI	3IK3
C(C[C@@H](C(=O)O)N)CN	ORN	3JDW
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C	STI	3K5V
Cc1ccc2c(c1/C=C/n3cnc4c3ncnc4Nc5ccc(cc5)P(=O)(C)C)cn[nH]2	B90	3KF4
Cc1ccc(cc1/C=C/n2cnc3c2ncnc3NC4CC4)C(=O)Nc5cc(cc(c5)n6cc(nc6)C)C(F)(F)F	B91	3KFA
CC(=O)C1=C(C=C2[C@@](C1=O)(c3c(cc(c(c3O2)C(=O)NCc4cccc5c4cccc5)OC)O)C)O	CEK	3LMP
c1ccc(cc1)Oc2ccc(cc2O)C=O	FT0	3LSY
Cn1cc(c2c1cc(c(n2)OC)OC)c3cc4c(ccnc4[nH]3)Cl	PDR	3LVP
c1cc2c(c[nH]c2c(c1)C(=O)NC3CCN(CC3)CCCCc4c[nH]c5c4cc(cc5)C#N)C#N	CCX	3LW0
c1cc2c(nc(nn2c1)[C@H]3CCCN3C(=O)c4ccc(nc4)F)Nc5cc([nH]n5)C6CC6	LGX	3NW5
c1cc2c(nc(nn2c1)[C@@H]3CCCN3C(=O)c4ccc(nc4)F)Nc5cc([nH]n5)C6CC6	LGW	3NW6
c1cc2c(nc(nn2c1)CNC(=O)c3ccc(nc3)F)Nc4cc(n[nH]4)C5CC5	LGV	3NW7
Cc1ccc(cc1C#Cc2cnc3n2nccc3)C(=O)Nc4ccc(c(c4)C(F)(F)F)CN5CCN(CC5)C	0LI	3OXZ
Cc1ccc(cc1C#Cc2cnc(n2C)C(=O)N)C(=O)Nc3ccc(c(c3)C(F)(F)F)CN4CCN(CC4)CCO	XY3	3OY3
CC(C)(c1cc(c(c(c1)Br)O)Br)c2cc(c(c(c2)Br)OS(=O)(=O)O)Br	ZXG	3PBA
CC(C)(C)c1cc(n(n1)c2ccc3c(c2)cccn3)NC(=O)Nc4ccc(cc4F)Oc5ccnc(c5)C(=O)NC	919	3QRI
CC(C)(C)c1cc(n(n1)c2ccc3c(c2)cccn3)NC(=O)Nc4ccc(cc4F)Oc5ccnc(c5)C(=O)NC	919	3QRJ
CC(C)(C)c1cc(n(n1)c2ccc3c(c2)C[C@H](NC3)C(=O)O)NC(=O)Nc4cccc(c4Cl)Cl	9DP	3QRK
c1ccc(cc1)C(CCNc2c3ccccc3nnc2C#N)c4ccccc4	MQ0	3STD
CC[C@@H](c1ccccc1)c2c(c3ccc(cc3oc2=O)OC)O	U03	3UPJ
c1c2c([nH]c1C(=O)O)-c3c(cc(nc3C(=O)C2=O)C(=O)O)C(=O)O	PQQ	4AAH
Cc1c(c2cc(ccc2n1C(=O)c3ccc(cc3)Cl)OC)CC(=O)O	IMN	4COX
C(C(=O)C(=O)O)C(=O)[O-]	OAA	4CTS
CN(Cc1cnc2c(n1)c(nc(n2)N)N)c3ccc(cc3)C(=O)N[C@@H](CCC(=O)O)C(=O)O	MTX	4DFR
CC1(CCC(c2c1ccc(c2)[C@@H](C(=O)Nc3ccc(cc3F)C(=O)O)O)(C)C)C	961	4LBD
c1ccc(cc1)C[C@H](CC(C[C@@H](Cc2ccccc2)C(=O)N[C@H]3c4ccccc4C[C@H]3O)O)C(=O)N[C@H]5c6ccccc6C[C@H]5O	VAC	4PHV
c1cn(c(=O)[nH]c1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)OP(=O)(O)O)O	U3P	4RSK
CC(=O)N[C@@H]1[C@H](C[C@](O[C@H]1[C@@H]([C@@H](CO)O)O)(CC=C)C(=O)O)O	CNP	4SLI
C[C@H](c1ccc(cc1)Br)NC(=O)c2cc(ccc2O)F	BFS	4STD
C([C@H](C(=O)O)OP(=O)(O)O)O	2PG	4TIM
CC(C)C[C@@H](C(=O)NO)N	LNO	4TLN
C[C@@H](C(=O)O)NC(=O)[C@H](CC(C)C)N[P@](=O)([C@H](Cc1ccccc1)NC(=O)OCc2ccccc2)O	0PK	4TMN
c1cc(ccc1C[C@@H](C(=O)O)N)O	TYR	4TS1
CC[C@@H](c1cccc(c1)NC(=O)CNC(=O)OC(C)(C)C)c2c(c3ccccc3oc2=O)O	U04	4UPJ
C([C@H]([C@H]([C@@H]([C@H](CO)O)O)O)O)O	SOR	4XIA
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)OP(=O)(O)O)N	2GP	5BIR
C([C@H](C(=O)O)OP(=O)(O)O)O	2PG	5ENL
C([C@@H]([C@H](C(=O)O)O)C(=O)O)C(=O)O	ICT	5ICD
c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO[P@](=O)(O)O[P@@](=O)(NP(=O)(O)O)O)O)O)N	GNP	5P21

C[C@@H](C(c1ccccc1)c2ccccc2)Nc3c4cc(c(cc4ncn3)F)F	UNN	5STD
CC(C)C[C@@H](C(=O)N[C@@H](CC(C)C)C(=O)O)N[P@](=O)(CNC(=O)OCc1ccccc1)O	0PJ	5TMN
C([C@H](C([C@H](CO)O)O)O)O	XYL	5XIA
C[C@@H](C(=O)N[C@@H](C)[P@@](=O)(O)O[C@@H](Cc1ccccc1)C(=O)O)NC(=O)OCc2ccccc2	ZAF	6CPA
C(C(=O)O)OP(=O)(O)O	PGA	6ENL
C[C@@H]1[C@](C1(Cl)Cl)(C(=O)N[C@H](C)c2ccc(cc2)Br)[S@@](=O)C	MS2	6STD
C([C@H](COP(=O)(O)O)O)O	G3P	6TIM
CC(C)C[C@@H](C(=O)O)NC(=O)[C@H](CC(C)C)O[P@](=O)(CNC(=O)OCc1ccccc1)O	0PI	6TMN
CC(C)[C@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)OCc2ccccc2)[P@@](=O)(O)O[C@@H](Cc3ccccc3)C(=O)O	FVF	7CPA
C[C@@H](C(=O)Nc1ccc(cc1)C(F)(F)F)NC(=O)[C@H](CC(C)C)NC(=O)C(F)(F)F	0Z2	7EST
CC[C@]1([C@H](C1(Cl)Cl)C)C(=O)N[C@H](C)c2ccc(cc2)Cl	CRP	7STD
C(C(=O)NO)OP(=O)(O)O	PGH	7TIM
c1ccc(cc1)S(=O)(=O)Nc2cccc(c2)[C@@H](c3c(c4c(oc3=O)CCCCCC4)O)C5CC5	INU	7UPJ
c1cn2c(n1)[C@@H]([C@H]([C@@H]([C@H]2CO)O[C@H]3[C@@H]([C@H]([C@@H]([C@H](O3)CO)O)O)O)O)O	IDC	8A3H
C([C@@H]1[C@@H]([C@@H]([C@H]([C@H](O1)O)O)O)O)O	GLA	8ABP
C([C@@H]1[C@@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O)O	GAL	8ABP
C([C@@H](C(=O)O)NC(=O)CP(=O)(O)O)C(=O)O	PAL	8ATC
C[C@@H](C(=O)NC[P@@](=O)(O)O[C@@H](Cc1ccccc1)C(=O)O)NC(=O)OCc2ccccc2	AGF	8CPA
C([C@@H]([C@H](C(=O)O)O)C(=O)O)C(=O)O	ICT	8ICD
Cc1c(c(c(cn1)COP(=O)(O)O)CN)O	PMP	9AAT
CC(C)[C@@H](C(=O)N[C@@H](Cc1ccccc1)C([C@H](Cc2ccccc2)NC(=O)[C@H](C(C)C)NC(=O)OCc3ccccc3)O)NC(=O)OCc4ccccc4	0E9	9HVP
c1cc(cc(c1)F)Cn2c3ccc(cc3cn2)Nc4c(c(ncn4)N)CNN5CCCCC5	HYZ	2RGP
COCCO/N=C/c1c(ncnc1Nc2ccc3c(c2)cnn3Cc4cccc(c4)F)N	POX	3BEL
N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)O)CC(C)C)CC(=O)O)CCC(=O)O	1A30_lig	1A30
N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N)CCCC)C)CCC(O)O)Cc1ccccc1)CC(C)C)C(C)CCCNC(N)N	1A8K_lig	1A8K
[C@H]1([C@@H]([C@H]([C@@H](C(=C1)CO)O)O)O)N[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1C)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)O)O)O)O)O)O	1AGM_lig	1AGM
[C@@H]1([C@H](NC(=O)C)[C@@H](O)[C@H](O)[C@H](O1)CO)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O)NC(=O)C)O	1BB5_lig	1BB5
C1(=O)CCCOc2ccc(C[C@H](NC(=O)[C@@H](N1)CC(=O)N)[C@@H](O)CN1[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N)C(C)C)[C@H](CC)C)CCC1)cc2	1CPI_lig	1CPI
N(=C\c1c(c(ncc1COP(O)(O)O)C)O)/[C@H](C(=O)O)CCSC	1D6S_lig	1D6S
N[C@@H](C(=O)N1[C@H](C(=O)N([C@H](C=O)CCCNC(N)N)C)CCC1)CC1CCCCC1	1EB1_lig	1EB1
OC[C@H]1O[C@@H]2C[C@@H]1OP(O)(O)OC[C@H]1O[C@H](C[C@@H]1O)n1c(=O)nc([C@@H]3N2C(=O)NC(=O)[C@]3(C)O)c(C)c1	1EHL_lig	1EHL
C(=O)(C)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)O)CCCNC(N)N)CCC(=O)N)CCCC)CCCC)[C@H](CC)C)[C@H](O)C	1FFI_lig	1FFI
C(=O)(C)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)O)CCCNC(N)N)CCC(=O)N)CCCC)CCCC)[C@H](CC)C)[C@H](O)C	1FG6_lig	1FG6
C(=O)(C)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)O)CCCNC(N)N)CCC(=O)N)CCCC)CCCC)[C@H](CC)C)[C@H](O)C	1FGC_lig	1FGC
[C@@H]1([C@@H]([C@H]([C@@H](CO1)O)O)O)O[C@H]1[C@@H]([C@H](CNC1)O)O	1FH7_lig	1FH7
[C@@H]1([C@@H]([C@H]([C@@H](CO1)O)O)O)O[C@H]1[C@@H](CNCC1)O	1FH8_lig	1FH8
[C@@H]1([C@@H]([C@H]([C@@H](CO1)O)O)O)O[C@H]1[C@@H]([C@H]([C@H](NC1)NO)O)O	1FH9_lig	1FH9
[C@@H]1([C@@H]([C@H]([C@@H](CO1)O)O)O)O[C@H]1[C@@H]([C@H]([C@H]2N(C1)CCN2)O)O	1FHD_lig	1FHD
[C@H]1([C@@H]([C@H]([C@@H](C(=C1)CO)O)O)O)N[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1C)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O)O[C@@H]1CO)O)O)O)O)O)O	1GAH_lig	1GAH
[C@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)C1)O)O)O)N[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1C)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O)O[C@@H]1CO)O)O)O)O)O)O	1GAI_lig	1GAI
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O[C@H]1[C@@H]([C@H]([C@@H]([C@@H](CO)O1)O[C@H]1[C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O)NC(=O)C)O	1HEW_lig	1HEW
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O)NC(=O)C)O	1HVQ_lig	1HVQ
[C@@H]1([C@H](O)[C@@H](O)[C@H](O)[C@H](O1)CO)S[C@@H]1[C@H]([C@H](O)O[C@@H]([C@H]1O)CO)O	1J8V_lig	1J8V

P(O)(O)(OC[C@H]1O[C@H]([C@@H]([C@@H]1O)O)n1cnc2c(N)ncnc12)OP(O)(O)O	1JU9_lig	1JU9
C(=O)(OCc1ccccc1)N[C@H](C(=O)O)[C@H](O)C	1KJO_lig	1KJO
C(=O)(OCc1ccccc1)N[C@H](C(=O)O)CCC(=O)O	1KJP_lig	1KJP
C(=O)(OCc1ccccc1)N[C@H](C(=O)O)CC(=O)O	1KKK_lig	1KKK
C(=O)(OCc1ccccc1)N[C@H](C(=O)O)C	1KL6_lig	1KL6
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O)NC(=O)C)O)NC(=O)C)O)NC(=O)C)O	1KQY_lig	1KQY
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O)NC(=O)C)O)NC(=O)C)O	1KQZ_lig	1KQZ
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O)NC(=O)C)O)NC(=O)C)O	1KR0_lig	1KR0
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O)NC(=O)C)O)NC(=O)C)O	1KR1_lig	1KR1
C(=O)(OCc1ccccc1)N[C@@H](C(=O)O)CCC(=O)O	1KR6_lig	1KR6
C(=O)(OCc1ccccc1)N[C@H]([C@H](C)O)C(=O)O	1KRO_lig	1KRO
C(=O)(OCc1ccccc1)N[C@@H](C(=O)O)CC(=O)O	1KS7_lig	1KS7
C(=O)(OCc1ccccc1)N[C@H](C)C(=O)O	1KTO_lig	1KTO
[C@H]1([C@@H]([C@H]([C@@H](C(=C1)CO)O)O)O)N[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1C)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@@H](O)O[C@@H]1CO)O)O)O)O)O)O	1LF9_lig	1LF9
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O)O[C@@H]1CO)NC(=O)C)O	1LJN_lig	1LJN
[C@@H]1([C@@H]([C@@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@@H]2[C@H]([C@@H]1CO)O[C@H](N2)N(C)C)O)NC(=O)C)O	1LLO_lig	1LLO
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O	1LMO_lig	1LMO
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O)NC(=O)C)O)NC(=O)C)O	1LMP_lig	1LMP
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O[C@@H]1CO)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O)NC(=O)C)O)NC(=O)C)O	1LMQ_lig	1LMQ
c1ccc2c(c1)cc(n2Cc1cc(ccc1)C(N)N)C(=O)NCc1ccc(cc1)[N](C)(C)C	1LQE_lig	1LQE
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@H](O[C@H]2[C@@H]([C@H]([C@H](O)O[C@@H]2CO)NC(=O)C)O)O[C@@H]1CO)NC(=O)C)O	1LZB_lig	1LZB
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@H](O[C@H]2[C@@H]([C@H]([C@H](O[C@H]3[C@@H]([C@H]([C@H](O)O[C@@H]3CO)NC(=O)C)O)O[C@@H]2CO)NC(=O)C)O)O[C@@H]1CO)NC(=O)C)O	1LZC_lig	1LZC
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@H](O[C@H]2[C@@H]([C@H]([C@H](O[C@H]3[C@@H]([C@H]([C@H](O)O[C@@H]3CO)NC(=O)C)O)O[C@@H]2CO)NC(=O)C)O)O[C@@H]1CO)NC(=O)C)O	1LZE_lig	1LZE
[C@@H]1([C@@H]([C@H](C[C@@H](CO)O1)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)NC(=O)C)O	1LZG_lig	1LZG
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@H](O[C@@H]1CO)O)NC(=O)C)O	1LZY_lig	1LZY
n1(cnc2c(N)ncnc12)C	1MRG_lig	1MRG
c12c3O[C@H](CP(O)(O)N[C@H](C(=O)N[C@H]2CC(C)C)CC(C)C)CCc3cc(c1)C	1PE5_lig	1PE5
c12cc(ccc1OCCP(O)(O)N[C@H](C(=O)N[C@H]2CC(C)C)CC(C)C)C	1PE7_lig	1PE7
C1COCCP(O)(O)N[C@H](C(=O)N[C@H]1CC(C)C)CC(C)C	1PE8_lig	1PE8
NC(=O)CC[C@@H](C(=O)O)NC(=O)c1ccc(cc1)[C@@H](C(C(F)(F)F)(O)O)CCCc1c(N)nc([nH]c1=O)N	1RBM_lig	1RBM
[C@H]1([C@H]([C@@H]([C@@H]([C@H](C)O1)O)O)O)O	1RDJ_lig	1RDJ
[C@H]1([C@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)O)O	1RDL_lig	1RDL
C1(=O)[C@@H](NC(=O)[C@@H](N(C(=O)[C@@H](N2C(=O)[C@@H](NC(=O)[C@@H](NC(=O)[C@H]([C@H](O1)C)NC(=O)[C@H](NC(=O)[C@@H](COS(O)(O)O)O)CC(C)C)CCCNC(N)N)CC[C@H]2O)CC(C)C)C)Cc1ccc(O)cc1)C(C)C	1TPS_lig	1TPS
N1C(=O)/C(=C\c2[nH]cc(c2CCC(=O)O)C)/c2ccccc12	1VR2_lig	1VR2
C(=O)(CCCCCCCCCCCC)OC[C@@H]([C@H](/C(=C(\C(O)O)/O)/O)O)O	1W3Y_lig	1W3Y
CC(=O)N[C@@H](N)NCCC[C@@H]1NC(=O)[C@@H](NC(=O)[C@@H](N2C(=O)[C@@H](NC(=O)[C@@H]3N(C1=O)CCC3)C[C@H]2O)Cc1cnc[nH]1)CCCC(=O)O	1W9U_lig	1W9U

CNC(=O)N[C@@H](N)NCCC[C@H]1C(=O)N(C)[C@H](C(=O)N[C@@H](CC(=O)N[C@@H](CC(=O)N[C@@H](C(=O)N1)C)C(=O)O)C(=O)O)Cc1ccccc1	1W9V_lig	1W9V
CNC(=O)N/C(=N/CCC[C@H]1C(=O)N(C)[C@H](C(=O)N[C@@H](CC(=O)N[C@@H](CC(=O)N[C@@H](C(=O)N1)C)C(O)O)C(O)O)Cc1ccccc1)/N	1WB0_lig	1WB0
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)O)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)O)O	1Z3T_lig	1Z3T
[C@@H]1([C@@H]([C@H]([C@H]([C@@H](CO)O1)O)O)O)O[C@H]1[C@@H]([C@H]([C@H](O)O[C@@H]1CO)O)O	1Z3V_lig	1Z3V
N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(O)O)C)C)C)C)C)C)C	2HPE_lig	2HPE
N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(O)O)C)C)C)C	2HPF_lig	2HPF
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O)O[C@@H]1CO)NC(=O)C)O[C@H](C)C(=O)N[C@@H](C)C(=O)N[C@H](CCC(=O)O)C(=O)N	2IXU_lig	2IXU
[C@@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O)O)NC(=O)C)O[C@H]1[C@@H]([C@H]([C@@H](O)O[C@@H]1CO)NC(=O)C)O[C@H](C)C(=O)N[C@@H](C)C(=O)N[C@H](CCC(=O)N)C(=O)O	2IXV_lig	2IXV
N[C@H](C(=O)N[C@H](C(=O)NCC=O)C)[C@H](CC)C	2OXW_lig	2OXW
NCC(=O)N[C@H](C(=O)O)Cc1ccc(cc1)O	3CPA_lig	3CPA
C(=O)(C)N1[C@H](C(=O)N[C@H](C(=O)N2[C@H](C(=O)N[C@H](C(O)O)Cc3ccccc3)CCC2)C)CCC1	4SGA_lig	4SGA
[C@H]1([C@@H]([C@H]([C@@H]([C@@H](CO)O1)O[C@H]1[C@@H]([C@H]([C@H]([C@@H](CO)O1)O)O)O)O)NC(=O)C)O	5GAL_lig	5GAL
C(=O)(C)N[C@H](C(=O)N[C@H](C(=O)N[C@@H](CC(C)C)[C@@H](O)CC(=O)N[C@H](C(=O)N[C@@H](CC(C)C)[C@@H](O)CC(O)O)C)C(C)C)C(C)C	5HVP_lig	5HVP
C(=O)(C)N1[C@H](C(=O)N[C@H](C(=O)N2[C@H](C(=O)N[C@H](C(O)O)Cc3ccc(cc3)O)CCC2)C)CCC1	5SGA_lig	5SGA
O=NC(=O)[C@@H](Cc1ccccc1)C(=O)N[C@@H](C)C(=O)NCCO	5TLN_lig	5TLN
N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@@H](CC(C)C)[C@@H](O)C[C@@H](C(C)C)C(=O)N[C@H](C(=O)N[C@H](C(O)O)C(C)C)[C@H](CC)C)CC(=O)N)CCC(=O)N)CO)C(C)C	8HVP_lig	8HVP
N([C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N1[C@H](C(=O)N)CCC1)CC(C)C)CCCCN(C)C)C1CCCCC1)Cc1ccc(cc1)C(N)N)C(=O)C	8KME_lig	8KME
P(O)(O)(OC[C@H]1O[C@H]([C@@H]([C@@H]1O)OP(O)(O)O)n1cnc2c(N)ncnc12)O	9ICD_lig	9ICD

Section S3. Enrichment comparisons for optimization procedures.

We can deploy different optimization procedures, including torsion minimization, hydrogen optimization, and combinations of them (hopt before minimization and minimization before hopt). Values are shown in main text **Table 7**. **Figure S2** shows the 42 DUDE-Z systems for these different optimization procedures.

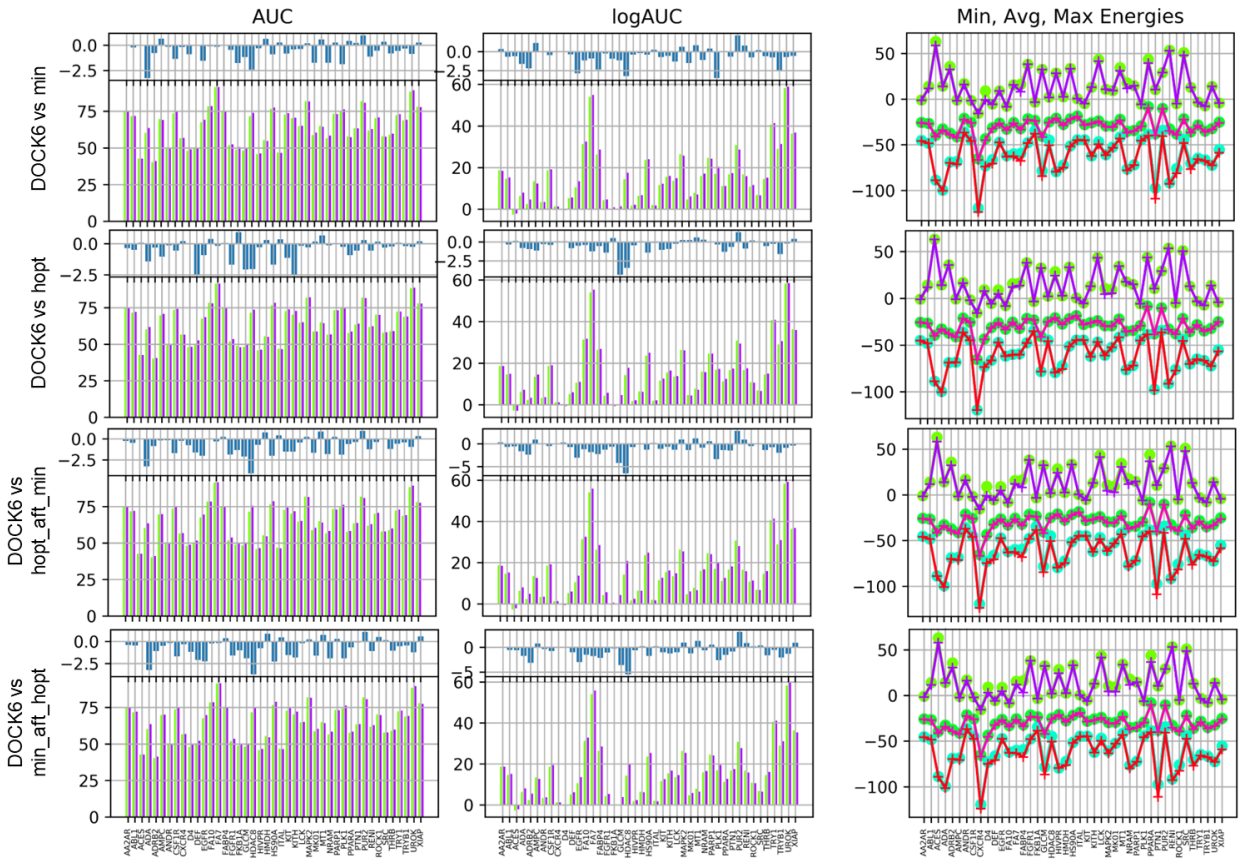


Figure S2. Comparing DOCK 6 with rescoring minimization and optimization runs over the 42 DUDE-Z systems. Ligand and property matched decoys downloaded from the web are used. A comparison quantifying enrichment using AUC (**left**) and logAUC (**right**) with DOCK 6 (green), DOCK minimized or optimized (purple), and the difference (Blue). Energy distributions are shown in the left panels.

Section S4. Pose reproduction swaps.

We show the plots for all the pose reproduction systems that swap from one category to another (**Figure S3**). All systems that change classifications from the *from_smi* pose reproduction experiments, which are the off-diagonal entries from **Table 11**, are shown in **Figure S3**. The final 10 graphs show movements or pose RMSD or poses close to the cutoff. These are mostly subtle movements. See main text and **Figure 8** for a detailed look at the system with PDB ID 2BGE, a *scoring failure* to *docking success*. 2BGE is one of the most interesting: before minimization, the best score and best RMSD are different poses. After minimization, the same pose is both the best score and RMSD. The origin of the after-minimization best scoring and RMSD pose is not either of the other poses. The docking-success-to-scoring-failure systems 1O36 and 1X75 have all poses switching, i.e., there are 4 non-overlapping lines, each containing one of the symbols. This indicates that the best energy and RMSD for the minimized pose originates from a different pose. This highlights the importance of considering multiple poses when rescoring or minimizing as post-docking steps. 1JMG (scoring-failure-to-docking-success) also has 4 non overlapping lines.

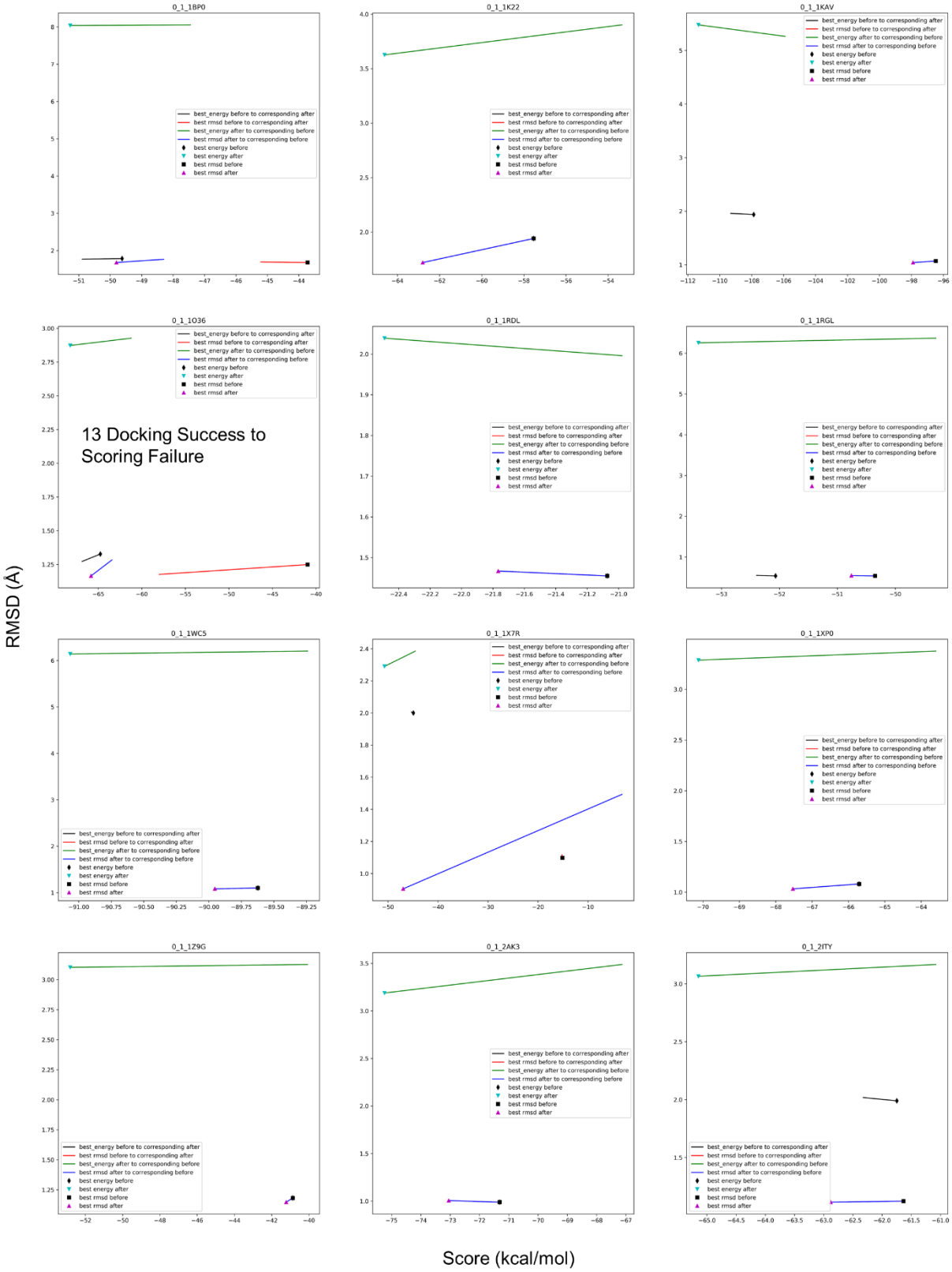

0_1_1BP0
0_1_1K22
0_1_1KAV
0_1_1O36
0_1_1RDL
0_1_1RGL
0_1_1WC5
0_1_1X7R
0_1_1XP0
0_1_1Z9G
0_1_2AK3
0_1_2ITY
best_energy before to corresponding after
best rmsd before to corresponding after
best_energy after to corresponding before
best rmsd after to corresponding before
best energy before
best energy after
best rmsd before
best rmsd after
13 Docking Success to Scoring Failure
RMSD (Å)
Score (kcal/mol)

Figure S3. (start)

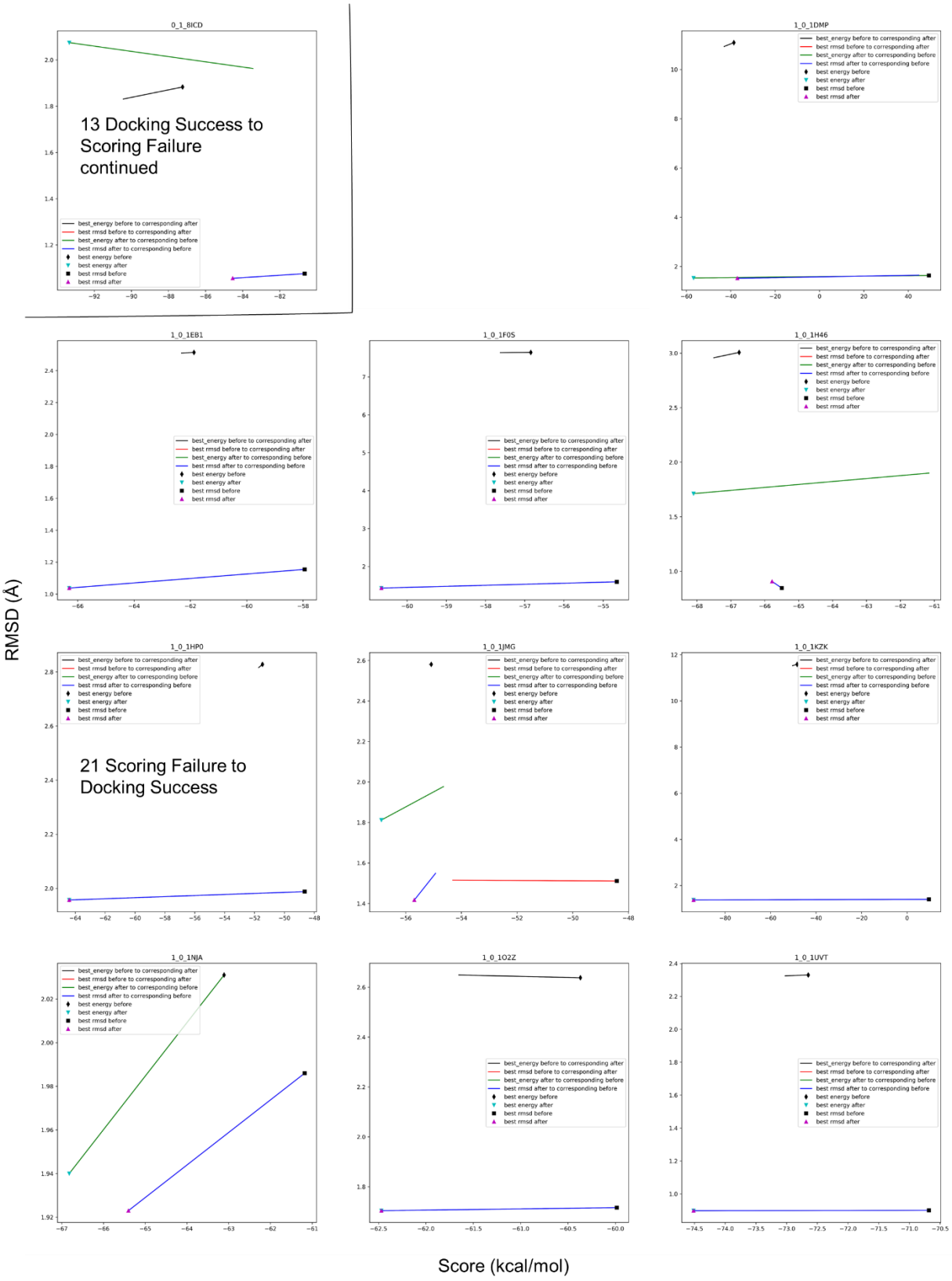


Figure S3. (continued)

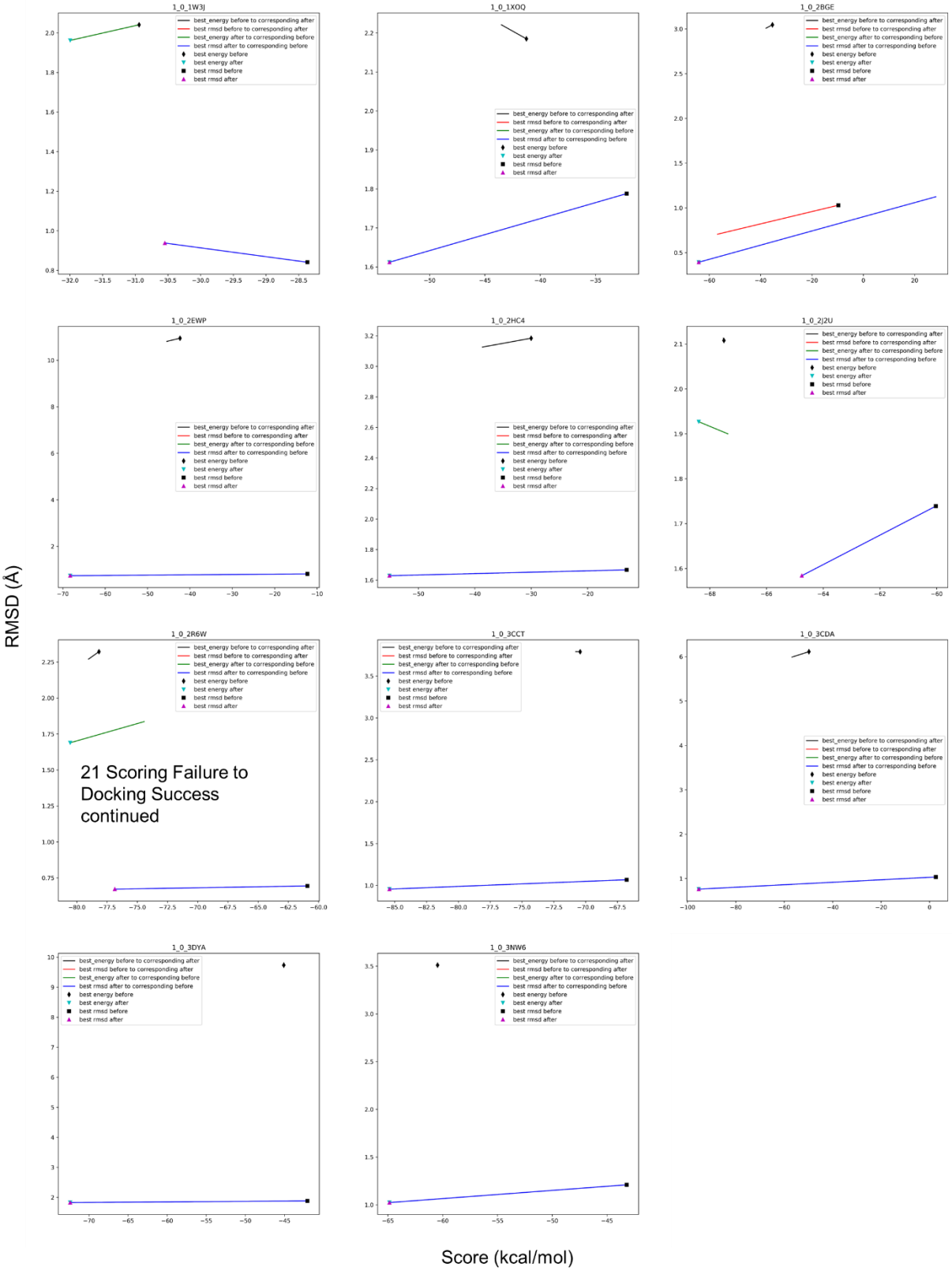


Figure S3. (continued)

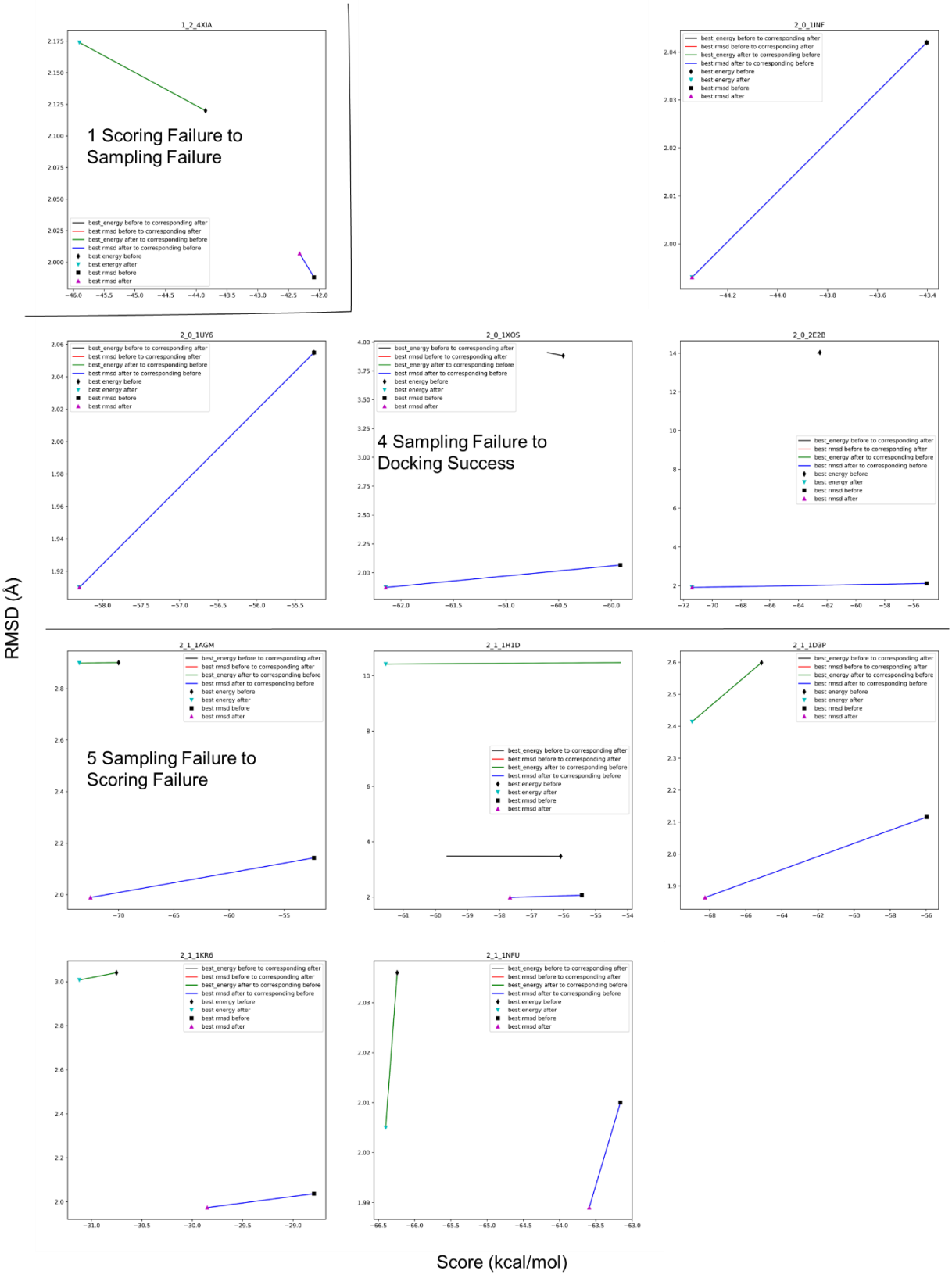


Figure S3. All the systems that switch classifications, comparing before and after torsion minimization for the *from_smi* experiment (Table 11). For each graph of RMSD vs. Grid Score, four pairs of poses are plotted. The black diamonds show the best docked pose. The black squares represent the best RMSD docked pose. The magenta triangles and the cyan upside-down triangles are the minimized best score and best RMSD poses, respectively. Lines connect to points before and after minimization: the black line connects best score before minimization to the corresponding score after minimization, red connects best RMSD before minimization to the corresponding RMSD after minimization, green connects best score after minimization to the corresponding score before minimization, and blue connects the best RMSD after minimization to the corresponding RMSD before minimization. The first 13 graphs are the *docking successes* to *scoring failures*. The next 21 are *scoring failures* to *docking successes*. The next plot is the only *scoring failure* to *sampling failure*. The next four plots are *sampling failures* to *docking successes*. The final 5 graphs are *sampling failures* to *scoring failures*. For each panel, the PDB name is shown in the title.