PROTEIN DESIGN

Expansive discovery of chemically diverse structured macrocyclic oligoamides

Patrick J. Salveson[1,2], Adam P. Moyer[1,2], Meerit Y. Said[1,2], Gizem Gökçe[1,3], Xinting Li[1,2], Alex Kang[1,2], Hannah Nguyen[1,2], Asim K. Bera[1,2], Paul M. Levine[1,2], Gaurav Bhardwaj[1,3], David Baker[1,2]*

Small macrocycles with four or fewer amino acids are among the most potent natural products known, but there is currently no way to systematically generate such compounds. We describe a computational method for identifying ordered macrocycles composed of alpha, beta, gamma, and 17 other amino acid backbone chemistries, which we used to predict 14.9 million closed cycles composed of >42,000 monomer combinations. We chemically synthesized 18 macrocycles predicted to adopt single low-energy states and determined their x-ray or nuclear magnetic resonance structures; 15 of these were very close to the design models. We illustrate the therapeutic potential of these macrocycle designs by developing selective inhibitors of three protein targets of current interest. By opening up a vast space of readily synthesizable drug-like macrocycles, our results should considerably enhance structure-based drug design.

Macrocycles composed of combinations of alpha, beta, gamma, delta, epsilon, aminobenzoic, aminophenylacetic, aminomethylbenzoic, and aminomethylpicolinic acids along with oxazoles, thiazoles, oxazolidines, thiazolidines, triazoles, and thioethers have bioactivities ranging from antifungal or antibiotic properties to cancer cytotoxicity to pain relief (*1*–*4*). Despite the diversity of bioactivities observed in natural products of this class, exploration of this space has been largely limited to variants of those present in nature—indeed, most macrocyclic drugs currently approved for use in humans are derived from natural products (*5*–*7*). The exploration of macrocycle chemical space in large-scale display techniques is largely constrained to those composed of primarily α–amino acids (*8*, *9*). Although current diversity-oriented synthesis techniques in principle enable the combinatorial synthesis of 10s to 1000s of macrocycles from more-exotic building blocks, these approaches are challenged by limitations in available chemistries (*10*–*13*).

Computational methods that enable the rapid and comprehensive exploration of the space of possible macrocycles could greatly facilitate the discovery of natural product–like compounds with distinctive bioactivities, but existing methods are similarly constrained by the diversity of chemistries accessible to them. Generative models derived from AlphaFold and RoseTTAFold enable designing small peptides composed of only the 20 canonical α–amino acids (*14*, *15*). Enumerative methods that rely on random sampling of backbone torsion angles to identify closed conformations of macrocycles become intractable when attempting to enumerate macrocycles built from more than a single backbone building block (*16*, *17*). There are nearly 60,000 two-, three-, and four-residue combinations of the 22 backbone building blocks depicted in Fig. 1A, and each of the 60,000 can be further diversified with millions of combinations of different side chains. This chemical space is far too large to employ explicit conformational sampling to identify the small fraction of linear sequences that can be closed to form a cycle.

We set out to develop a computationally tractable method for sampling the very large chemical space of possible small macrocycles (Fig. 1; see methods in the supplementary materials). We devised a three-step approach for carrying out this search given an input set of monomers with widely varying chemical structures. In the first step, we use an adaptive grid-based search over the AIMNet (*18*, *19*) potential energy landscape for each monomer to identify low-energy conformers. In the second step, the rigid body transformations associated with these monomer conformations, and all dipeptide conformations constructed from pairwise combinations of these monomers, are computed and stored in hash tables. In the third step, macrocycles are rapidly and systematically generated by identifying pairs of entries in the hash tables for which the combination of the rigid body transforms is close to zero (so that, together, the fragments form a cycle). This transform-based approach allows rapid identification of two-, three-, or four-residue macrocycles from every combination of the monomers depicted in fig. S1 in far less compute time than would be required to explicitly build backbone coordinates and evaluate closure for each of the different monomer combinations. The approach is agnostic to the number, identity, and connectivity of atoms between the terminal amides, as only their relative orientations in space are considered for determining closure.

Enumeration of closed macrocycles

We used our approach to explore the very large space of three- and four-residue macrocycles constructed from 130 monomers and, where appropriate, their enantiomers. These monomers fall into 22 classes, hereafter referred to as chemotypes, which are distinguished by the atomic number and hybridization of the atoms in the backbone of each monomer (Fig. 1A); we refer to each chemotype by a letter (“a” for α–amino acids, “b” for β–amino acids, “d” for δ–amino acids, etc.; see Fig. 1A caption). We constructed hash tables for all low-energy conformers identified from the potential energy surfaces for each of the 130 monomers (fig. S1) and from dimers constructed from these low-energy conformers, and then we systematically searched through all combinations of monomer-dimer, dimer-monomer, and dimer-dimer hash tables, generating ensembles for each chemical for which matching hash values could be found. These hash tables contain nearly 9 billion conformers, and generating all combinations explicitly to check for closure would require building ~10^{19} macrocycles; our hashing approach reduces the complexity required to evaluate all of these combinations from $O(n^2)$ to $O(n)$.

Systematically searching through the hash tables yielded 14.9 million closed macrocycles containing 9- to 32-membered rings belonging to 3494 three-residue and 38,544 four-residue chemotypes (accounting for all circular permutations of the cyclic chemotypes). The chemical space spanned by the three- and four-residue chemotypes greatly expands upon what has been previously explored (Fig. 2A and fig. S3): 206 chemicals belonging to 23 chemotypes have high-resolution structures deposited in the Cambridge Structural Database (CSD), and 13,932 chemicals belonging to 397 chemotypes are described in the PubChem database. Previous computationally designed macrocycles have been built from just one chemotype (α–amino acids).

Each chemotype samples distinct 3D shapes (Fig. 2, B and C, and fig. S5). To characterize the structural diversity for each chemotype, we binned backbone torsion angles into 60° bins and represented each macrocycle conformer by a string of these bins. The many thousands of conformers generated for the macrocycles belonging to each chemotype for the most part have only a handful of bin strings, which likely reflect the torsional preferences of the constituent amino acids. Principal moment of inertia (PMI) analysis of conformers spanning all the bin strings sampled for each sequence revealed that different

---

[1] Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [2] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [3] Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195, USA.
*Corresponding author. Email: dabaker@uw.edu

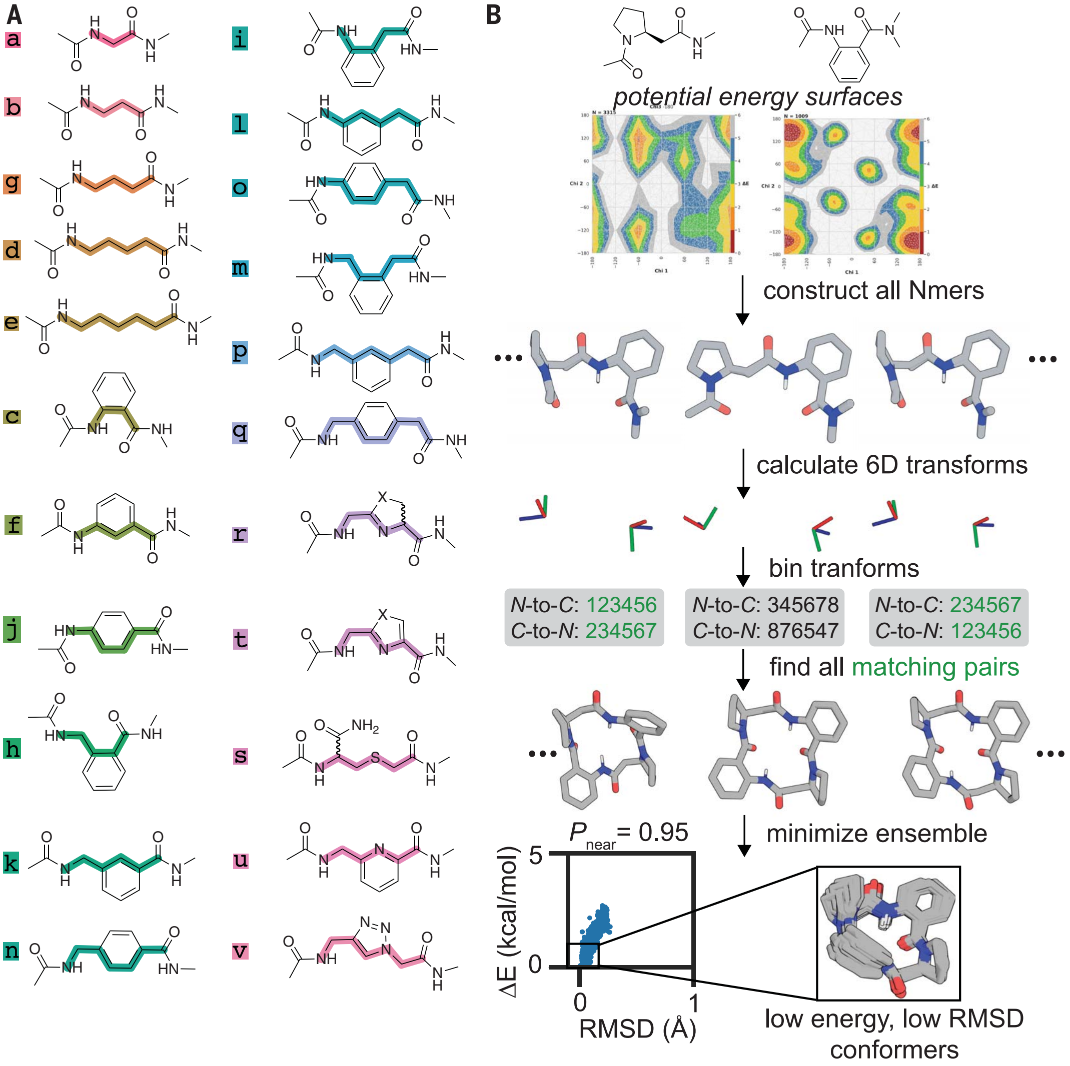

Fig. 1. Overview of macrocycle discovery approach. (**A**) Chemical structures of monomeric chemotypes used to construct small drug-like macrocycles: amino acids (a, b, g, d, and e), aminobenzoic acids (c, f, and j), aminomethylbenzoic acids (h, k, and n), aminophenylacetic acids (i, l, and o), aminomethylphenylacetic acids (m, p, and q), oxazolidines or thiazolidines (r), oxazoles or thiazoles (t), thioethers (s), aminomethylpicolinic acids (u), and triazoles (v). The backbone atoms are colored to highlight the differences between chemotypes. The elaborated monomers used throughout this work that belong to these chemotypes are presented in fig. S1. (**B**) Assembly and identification of low-energy conformers of cyclo-(βpro-ben2-βpro-ben2) beginning from calculated potential energy surfaces of Ac-βpro-NHMe and Ac-ben2-NMe2. Combining these low-energy conformers yields thousands of Ac-βpro-ben2-NMe2 dimers. The 6D transform between the planes defined by the terminal amides is calculated for each dimer and then binned. All pairs of dimers that close one another (indicated by green in this example) are identified in a quick lookup, and the resulting ensemble of macrocycle conformers is minimized.

macrocycle chemotypes sample distinct regions of shape space (*20*). In some cases—for example, aaaa and aaac—the specific 3D shapes sampled are quite restricted by the closure constraint, regardless of sequence or torsional diversity.

For each monomer sequence, the hash-based closure generates an ensemble of macrocycle structure models built from low-energy monomer conformers. Given sufficient compute power, for each of these ensembles, we could carry out minimization in AIMNet (to incorporate monomer–monomer interactions, closure strain, etc.) and evaluate the degree to which the sequence of monomers encodes a single low-energy conformer by considering the energy landscape mapped by the set of all low-energy conformers for the sequence [using, for example, the P_{near} approximation to the Boltzmann weight (*21*)]. Indeed, we found that the hash sampling procedure coupled with AIMNet calculations can predict the structures of small macrocycles from their sequence of monomers with backbone root mean square deviations (RMSDs) between the predicted and experimental structures of <0.3Å (fig. S6). Carrying out the full energy calculations on all 14 million generated ensembles is, however, not computationally tractable: ~10^8 CPU-hours would be needed to minimize the ~23 billion conformers sampled across the 14 million ensembles. Instead, we focused on two subclasses of closed macrocycles that we hypothesized would be particularly likely to have single ground states: (i) those with local interactions strongly favoring one or a small number of closed states and (ii) those containing nonlocal hydrogen bonds between backbone amides.

Locally encoded macrocycles

To identify the first locally encoded subclass, we generated conformational ensembles from hash tables containing only monomers within

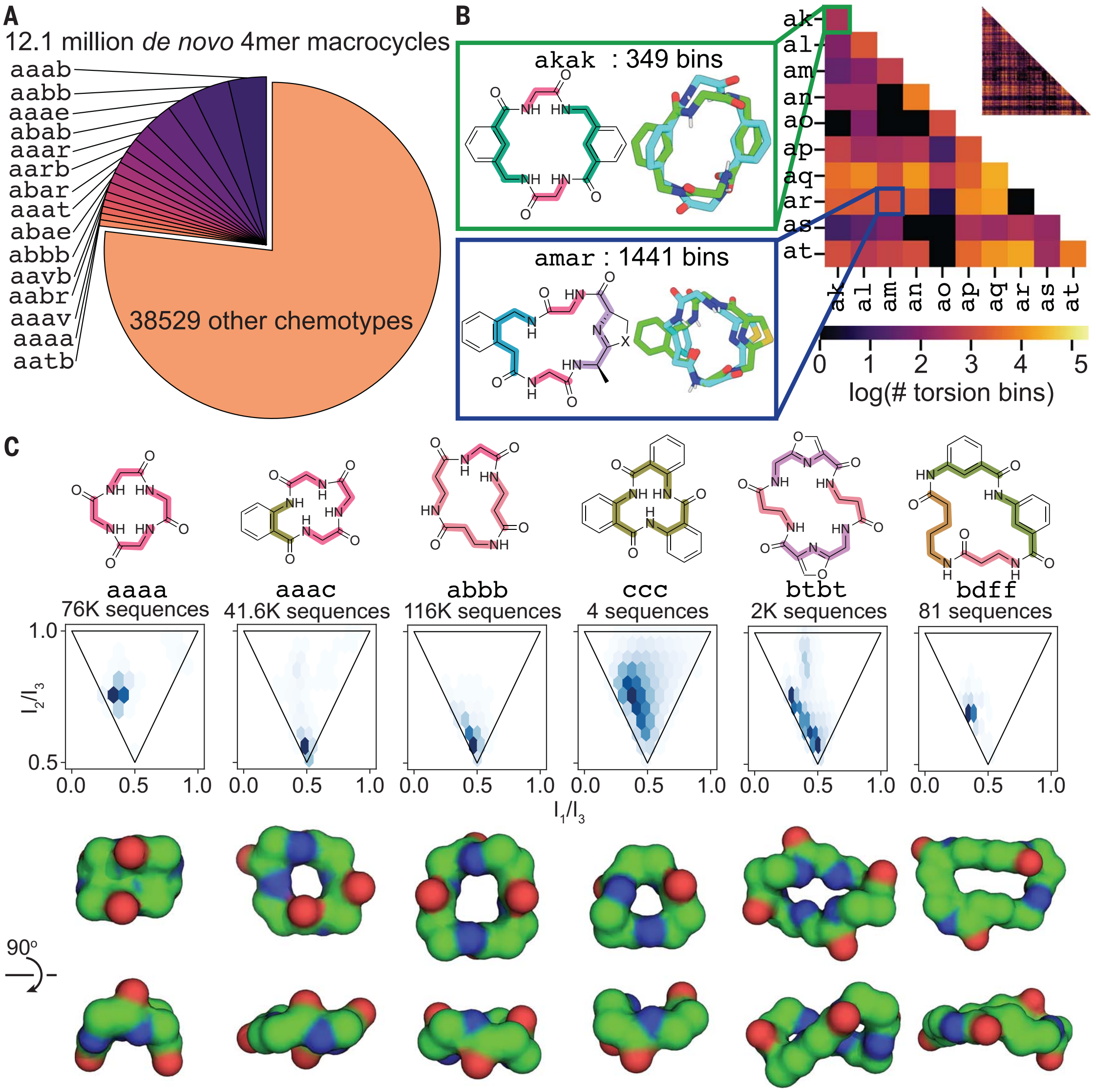

Fig. 2. Chemical and structural diversity of small de novo macrocycles. (**A**) Distribution of sequences sampled across all four-residue macrocyclic chemotypes. The 15 most populated chemotypes are labeled. (**B**) Torsional diversity of sampled four-residue macrocycle chemotypes. Heatmap pixels represent four-residue macrocycle chemotypes generated from the two residue fragments on the *x* and *y* axes and are colored by the number of backbone torsion bins sampled across all sequences belonging to that chemotype. The inset at top right shows the full map over all 22^4 monomer chemotype combinations; a 10-by-10 subset of this is blown up in the main panel for clarity. Examples of two representative torsion bins are shown in the boxes at left for two chemotypes. (**C**) Principal moment of inertia distributions (second row) and representative conformers (bottom) sampled by macrocycles constructed from different chemotypes (top row).

1 kcal/mol of their lowest energy state; such macrocycles are torsionally optimized from the perspective of the composite building blocks. As the resulting macrocycle space is still very large, for computational tractability, we explored the use of the number of different torsion bin strings sampled in these ensembles as an indicator of the extent to which the torsional biases together with the closure constraint are sufficient to specify a single low-energy minimum. We found, after AIMNet minimizations of a subset of the ensembles, that in cases where five or fewer torsion bin strings were sampled there were generally deep energy minima in the AIMNet landscapes around a single state with $P_{\text{near}} > 0.9$ (fig. S7). We identified ~380,000 three- and four-residue sequences for which the hash-generated ensembles contained >50 closed conformers that spanned five or fewer torsion bins and carried out AIMNet minimization on ~4800; roughly 85% of these had single low-energy structures with $P_{\text{near}} > 0.9$.

We synthesized 13 such macrocycles and were able to determine structures of 10 using x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (Fig. 3 and fig. S8). Six macrocycles grew crystals in vapor diffusion experiments of sufficient quality for structure determination using direct methods; seven macrocycles had NMR spectra in deuterated dimethyl sulfoxide (DMSO-d6) with sufficient dispersion to enable the unambiguous assignment of key rotating frame Overhauser effects (ROEs) in ROESY spectra—for these, we generated 3D models by relaxing conformers generated from integrated ROE-based distance constraints in unconstrained molecular dynamics (MD) simulations and selecting the lowest-energy conformers (see methods). Eight of the 11 experimentally determined structures superimpose on their respective design models with RMSDs of 0.8 Å or less, with small differences in ring puckering of five- and six-membered rings, in single

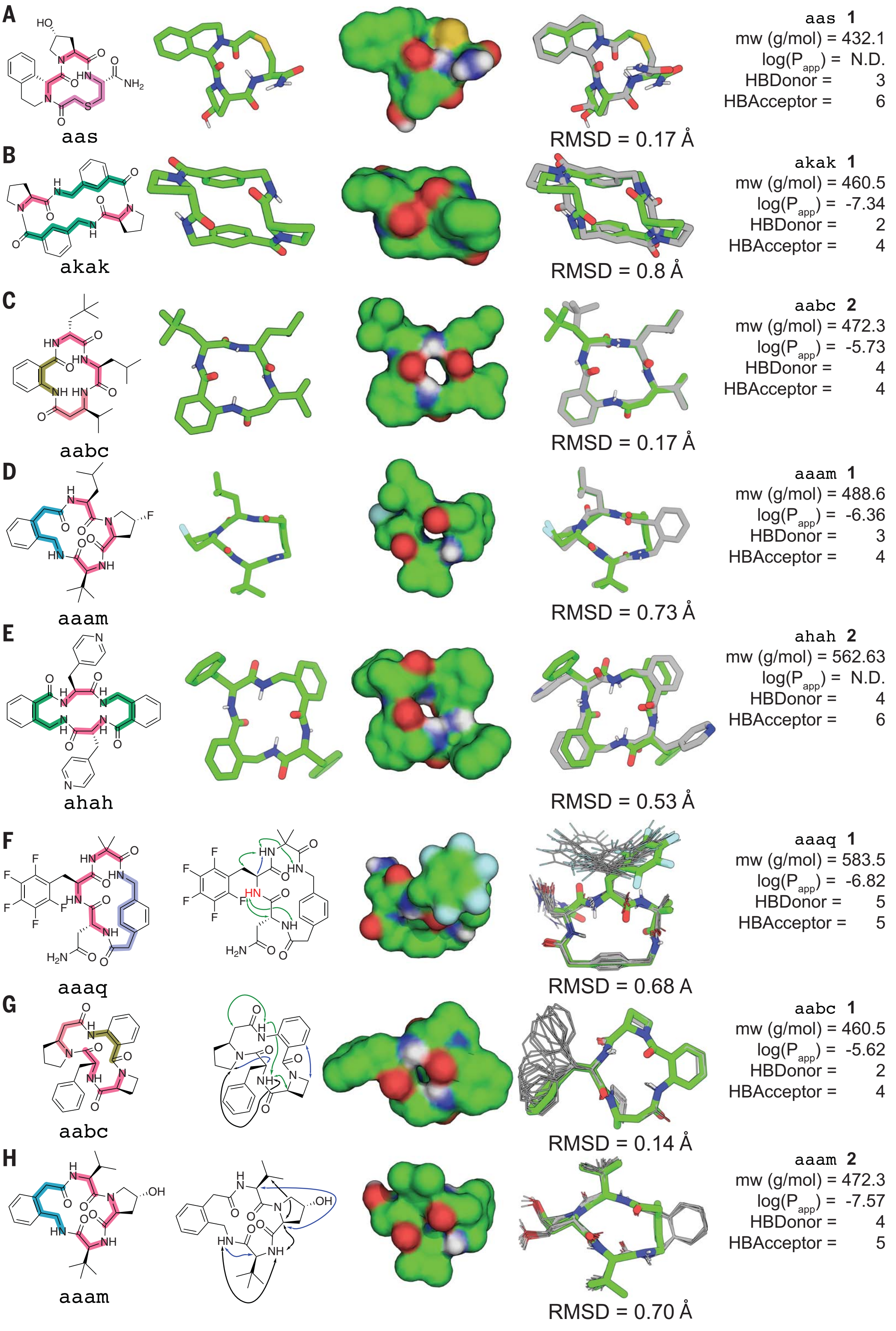

Fig. 3. X-ray and NMR structures of locally encoded macrocycles are very close to the design models. (**A** to **E**) Crystal structures. Column 1, chemical structure colored by chemotype; column 2, design model; column 3, surface representation of design models; column 4, superposition of the design model (green) and experimentally determined coordinates (gray); column 5, chemical properties and apparent permeabilities; the number of hydrogen bond donors (HBDonor) and hydrogen bond acceptors (HBAcceptor) was determined using RDKit. ahah **2** was modeled with phenylalanines and chemically synthesized with pyridylalanines (E). mw, molecular weight. (**F** to **H**) NMR structures. In column 2, in place of the design model, we show ROEs used for distance-constrained ensemble generation. Blue arrows, distances < 2.5 Å; green arrows, distances between 2.5 and 3.5 Å; black arrows, distances > 3.5 Å. Experimental structure to design model RMSDs are reported over all backbone atoms for x-ray structures and over ensemble average backbone coordinates for NMR structures.

side-chain torsion angles, and in pairs of backbone torsion angles [larger differences were observed for the other three macrocycles (fig. S8)]. We were able to determine both x-ray crystallographic structures and NMR structures for two of the 11 macrocycles. In these two cases, both the NMR structure and the x-ray crystallographic structures are very similar to their respective design models and to each other (fig. S9).

The 10 structures (Fig. 3) are built from a total of eight different monomer chemotypes arranged into six macrocycle chemotypes. Design aas **1** contains two hydrogen bond interactions: one between backbone amides and the second involving the primary amide of the s monomer. Both aaam macrocycles and macrocycle akak **1** contain proline-like monomers sampling *cis*-amides. In the NMR structures of both aaam macrocycles, these motifs lead to strong ROEs between the Cα proton of the proline-like monomer and the Cα proton of the monomer immediately N-terminal to it; these macrocycles contain α–amino acids in very extended β sheet–like conformations, and the *cis*-amides rapidly turn the backbone to enable closure (fig. S10). Macrocycle aaaq **1** contains an extended pentafluorophenylalanine residue that buries its amide NH against the 4-aminomethylphenylacetic acid, thereby shielding it from solvent and causing the amide NH to shift downfield with a temperature shift coefficient (T_{coeff}) of 0.9 parts per billion (ppb)/K (the backbone amides in the remaining monomers of this design shift upfield with T_{coeff} values all less than −5.0 ppb/K; see supplementary data). The aaam **1**, aaam **2**, and aaaq **1** designs provide three different ways to mimic the extended conformations of poly-α–amino acid peptides, such as those adopted by many protease and kinase substrates (*22*, *23*), and could be useful for targeting these enzymes.

Hydrogen bond–stabilized macrocycles

To identify the second class of macrocycles containing transannular hydrogen bonds, we constructed dimer hash tables with conformers for which the terminal amides make a hydrogen bond, increasing the monomer energy cutoff for hash table inclusion to increase sampling of these rare interactions. To reduce local strain, after identification of macrocycle chemotypes that contained long-range hydrogen bonding interactions, we chose monomer identities and side-chain rotamers with low energy given the backbone torsion angles. After hash table–based ensemble enumeration and AIMNet minimization, we selected for experimental characterization designs with much lower energy than predicted by a simple monomer composition–based model (see methods).

We prepared 17 macrocycles designed to contain long-range hydrogen bonding interactions and were able to determine the high-resolution structures of seven (Fig. 4). Four macrocycles grew crystals in vapor diffusion experiments of sufficient quality for structure determination using direct methods; three macrocycles displayed NMR spectra with sufficient dispersion to facilitate the unambiguous assignment of key ROEs in ROESY spectra. Eight of the 17 macrocycles displayed 1D-NMR in DMSO-d6, indicative of a mixture of states; six of these collapse into a single major species in deuterated chloroform ($CDCl_3$). Although only three macrocycles showed sufficient dispersion for unambiguous assignment, 10 of the 17 had sufficient dispersion to enable measuring amide T_{coeff} values that are consistent with the presence of the designed transannular hydrogen bonds in all 10 cases (supplementary data). All seven of the experimentally determined structures superimpose on their respective design models with RMSDs of <0.8 Å. Like the macrocycles described in Fig. 3, differences between the designed model and the experimental structures of these hydrogen bond–containing macrocycles involve differences in ring-puckering and 180° rotation of backbone amides.

The macrocycles depicted in Fig. 4 are constructed from eight different monomer chemotypes arranged into six macrocyclic chemotypes and are stabilized by a transannular hydrogen bond between backbone amides involving non-α-amino acids. Macrocycle aalm **1** contains a hydrogen-bonding fragment built from monomers with predominantly sp^2 hybridized atoms in the backbone (lm), macrocycle aagb **1** contains a hydrogen-bonding fragment whose backbone contains many more sp^3 hybridized atoms than present in α–amino acid backbones (gb), and macrocycle aabi **2** contains a hydrogen-bonding fragment that blends these two features (bi). Six of the macrocycles contain contiguous fragments of α–amino acids that mimic β-turns common in protein–protein interfaces (*24*–*27*): The designed macrocycles belonging to the aagb, aabi, and aaam chemotypes contain turns akin to type I β-turns, and the aalm-based macrocycle contains a turn akin to the type II β-turn. The *N*-methylated amino acid residues present in both aaap-based macrocycles adopt *cis*-amides, resulting in type VI–like turns. Despite the presence of these obvious β-turn–like features in the aaam **6** macrocycle, the spacing and orientation of the two phenylalanine side chains mimic the spacing and orientation of side chains positioned at *i* and *i*+4 of an α helix (fig. S11; this spacing is not present in either aaap **3** or aaap **4**, which are isomers of aaam **6** at the level of the backbone atoms). This mimicry of β-turn and helical arrangements should be useful in targeting proteins recognizing these structural elements.

We measured the passive membrane diffusion of all 29 macrocycles designed with the two approaches using the parallel artificial membrane permeability assay (PAMPA) [Figs. 3 and 4 and fig. S12; roughly half of the 14.9 million closed macrocycles satisfy the rule-of-five (Ro5) criteria for drug-like compounds; fig. S4 (*28*)]. Almost all of the macrocycles were quite membrane permeable: 16 of the 29 had $\log(P_{\mathrm{app}})$ values (P_{app} is the apparent permeability) greater than −6, and only three were not detected in the acceptor well. Under the same conditions, the small-molecule drug propranolol achieved $\log(P_{\mathrm{app}})$ of −5.36 (fig. S12). As expected, designs with exposed polar groups, particularly side-chain hydroxyls and primary amides, had lower permeabilities. Overall, the hydrogen bond–containing designs were more permeable than the torsionally optimized designs, likely owing to fewer exposed backbone polar groups (*29*–*31*). The vast majority of the macrocycles are stable to serum proteases; the half-lives in reconstituted serum of all but a single macrocycle were >24 hours (fig. S19).

Designing macrocyclic inhibitors of protein targets

We sought to use the above methodology to identify chemically diverse macrocycles capable of inhibiting protein targets of interest starting from known inhibitors with limited potency and/or selectivity. We used a motif scaffolding approach to identity inhibitors of histone deacetylase 6 (HDAC6) and Mpro, the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We started from dipeptide fragments derived from the previously described crystal structure of cyclic peptide des4.3.1 bound to HDAC6 and the crystal structure of the linear C-terminal prosequence bound to Mpro (see methods) (*32*, *33*). We identified candidate macrocyclic inhibitors by rapidly searching for closures across all hydrogen bond–containing hash tables with these dipeptides, building the resulting four-residue macrocycles in their respective pockets, and discarding candidates that clashed with the surrounding protein residues. This search embeds the starting dipeptide motif in nearly 38,000 possible macrocycles whose 3D coordinates are constructed in the context of the target protein binding site. We prioritized synthesizing candidate macrocycles on the basis of three criteria: (i) the degree to which the designed sequence preorganizes the dipeptide fragment into the binding-competent conformation assessed by P_{near} analysis of the generated ensembles, (ii) recovery of the designed complex in ligand docking simulations performed in GALigandDock (*34*), and (iii) predicted binding free energy of the designed sequence compared with that of the initial dipeptide motif. After filtering the candidate macrocycles with the above criteria, we synthesized 11 macrocycles belonging to six chemotypes designed to inhibit HDAC6 and six macrocycles belonging to five chemotypes designed to inhibit Mpro.

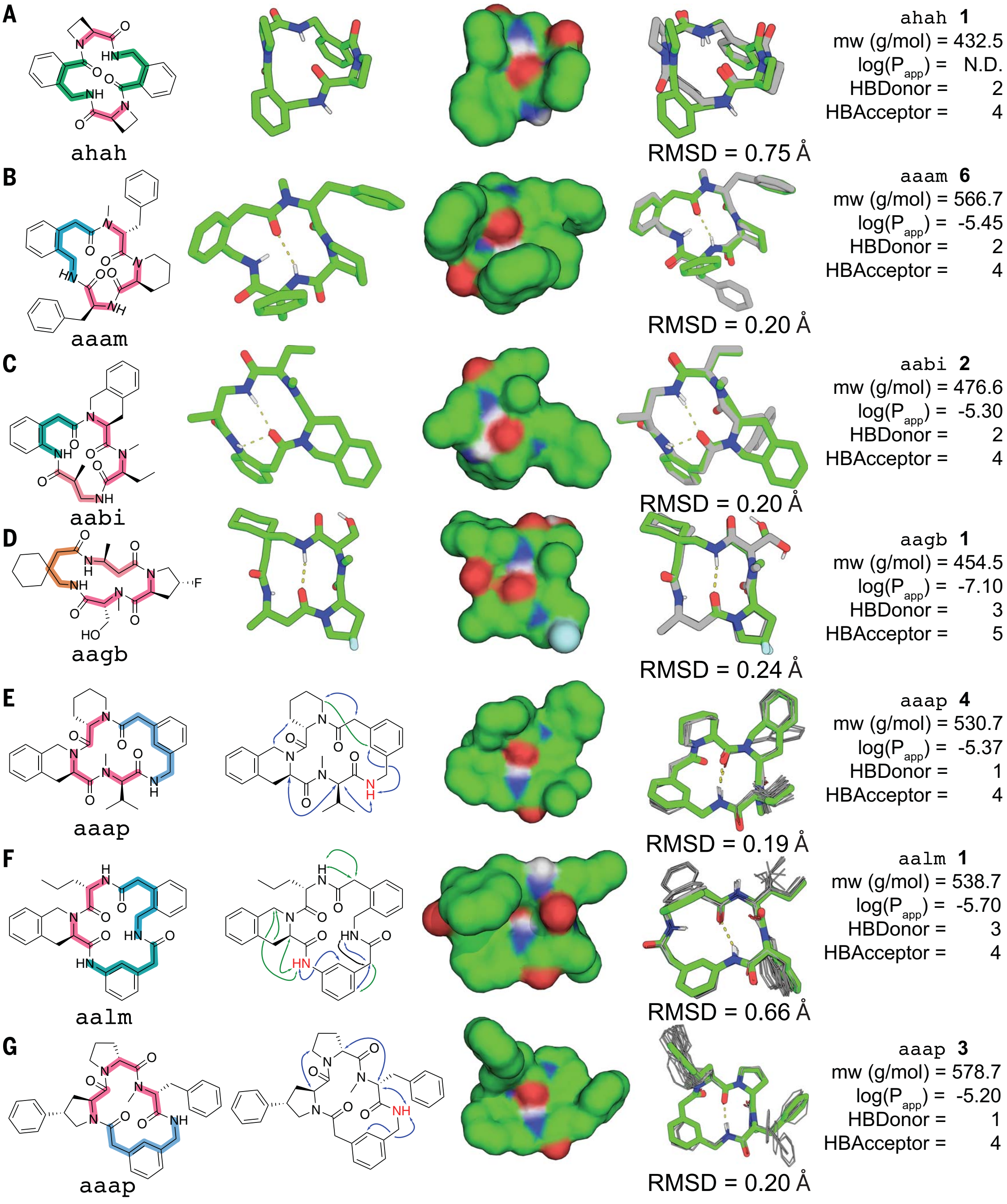

Fig. 4. X-ray crystallographic and NMR ensemble structures of hydrogen bond–containing macrocycles are very close to their design models. (**A** to **G**) Columns 1 to 5 are as in Fig. 3. Column 2 rows in (E) to (G): ROEs used for distance-constrained ensemble generation. Arrows are as in Fig. 3. Red amide NH indicates protons with T_{coeff} > −4.5 pbb/K.

Ten of the 11 chemically synthesized macrocycles displayed median inhibitory concentration (IC_{50}) values against HDAC6 ranging from 1.5 to 70 nM, a 10-fold to 400-fold improvement in potency compared with the minimal alkylthiol warhead present in each design (Fig. 5, A to C; fig. S20; and table S11). The most selective of these macrocycles belong to three distinct chemotypes and are 100-fold to >1000-fold more selective for HDAC6 over the other HDACs tested. The observed selectivity of the designed macrocycle is comparable to several advanced HDAC6-selective inhibitors (table S11). Of the six designed Mpro inhibitors, only one, belonging to the aabi chemotype, displayed a modest IC_{50} of ~16 μM (Fig. 5D; fig. S21; and table S12). We subsequently constructed and redocked a focused in silico library of homologs of this modest inhibitor using GALigandDock keeping the backbone chemistry and its predicted structure unchanged and varying the side chains. From this in silico library, we selected and chemically synthesized 27 homologs, 17 of which displayed IC_{50} values against Mpro ranging from 4 to 0.88 μM (table S12). Linear analogs of three active designs showed IC_{50} values of >200 μM, supporting the importance of macrocyclization of the sequence.

We next sought to identify macrocycles that inhibit the interaction of Bak (BCL2 homologous antagonist/killer) and MCL1 (myeloid cell leukemia-1) using the large number of hash-generated macrocycle ensembles as an in silico

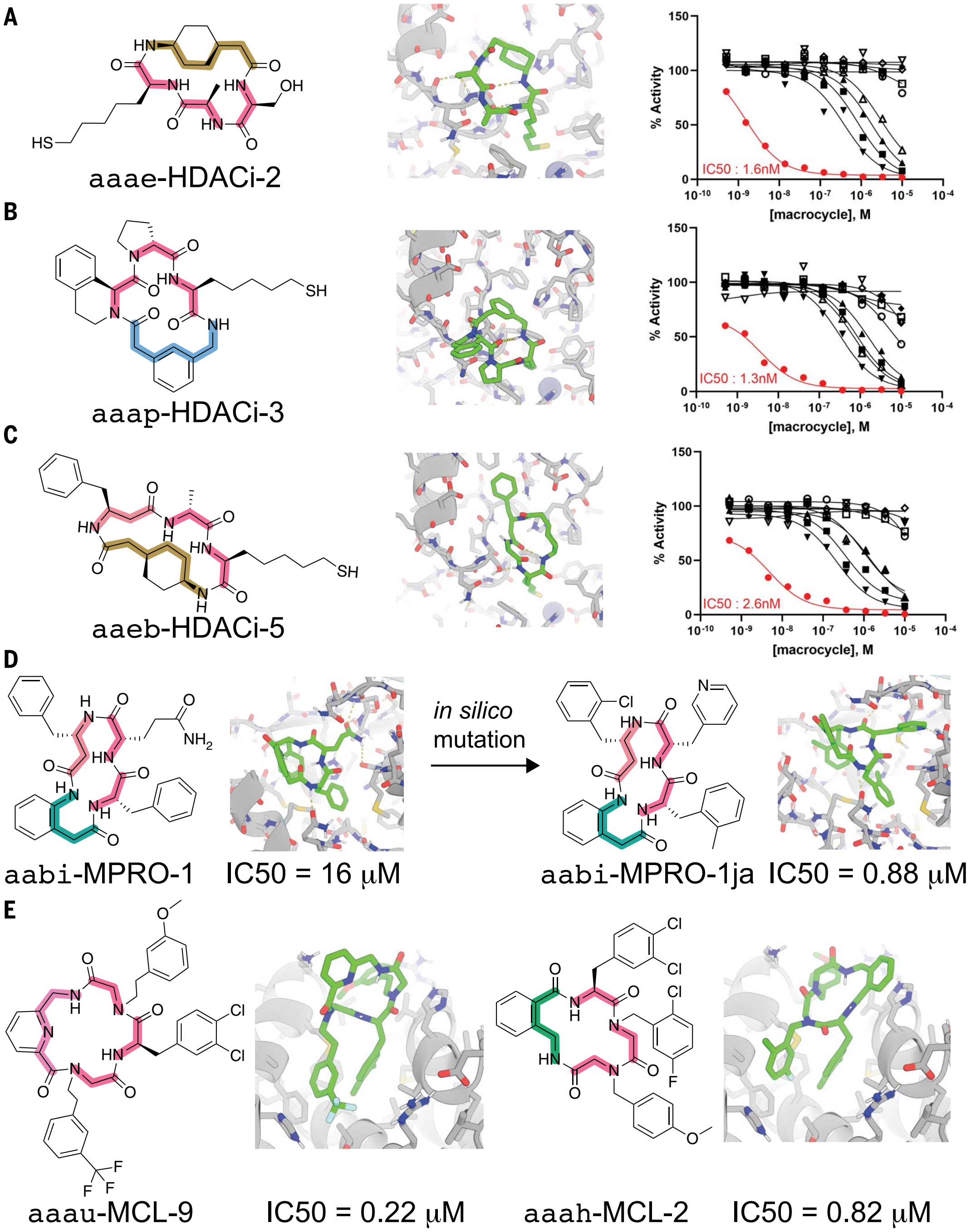

Fig. 5. De novo design of macrocycles that inhibit enzymes and protein–protein interactions. (**A** to **C**) Chemical structures (left), design models (middle), and concentration dependence of inhibition (right) of HDAC6 (red) and HDACs 1 (closed square), 2 (closed triangle), 3 (closed inverted triangle), 4 (closed diamond), 5 (open circle), 7 (open square), 8 (open triangle), 9 (open inverted triangle), and 11 (open diamond) by designed macrocyclic HDAC inhibitors (HDACi). The macrocycles were designed to contain a nonselective alkylthiol moiety. The designed macrocycles show 100-fold to >1000-fold greater selectivity for HDAC6 than for other HDACs. (**D**) Chemical structures and design models of the initial hit versus Mpro (left) and most potent (right) designed macrocycle identified from a focused in silico library of mutants of the initial hit. (**E**) Chemical structures and design models of two of the most potent designed MCL1: Bak PPI inhibitors.

virtual library. We generated MCL1-directed virtual libraries by first modifying each macrocycle to contain a side-chain dichlorobenzene moiety [a functional group previously described to interact with MCL1 at the MCL:Bak interface (*35*)] and then mutated the remaining positions of the macrocycles to a host of other noncanonical α-amino acids, peptoids, and β3-amino acids beyond those for which we explicitly generated energy landscapes (see methods). From a virtual library of ~2.7 million compounds, we prioritized macrocycles for synthesis on the basis of the calculated binding free energy resulting from GALigandDock and visual inspection. Three of the initial 18 synthesized macrocycles showed inhibitory activity near or below 50 μM in the Förster resonance energy transfer assay (table S13). We subsequently generated and docked a larger virtual library of ~27 million macrocycles and identified 12 active designs from the 30 synthesized macrocycles (Fig. 5E; figs. S22 and S23; and table S13). We suspect that the higher in silico and experimental success of the larger library is due to a greater diversity of shapes represented (*36*); the small library contained macrocycles belonging to only 22 chemotypes, whereas the larger library contained macrocycles derived from 3991 different chemotypes. The identified hits spanned IC_{50} values from tens of micromolar to hundreds of nanomolar, with the most potent being 220 nM.

Discussion

Small macrocycles found in nature generally populate a mixture of conformers in solution. Our approach enables the rapid exploration of the large and chemically diverse macrocycle space to identify those that populate primarily a single conformer. The two strategies used here—using conformationally restricted building blocks to favor specific geometries or incorporating longer-range hydrogen bonding interactions—both resulted in macrocycles with well-defined structures [another strategy is to introduce large barriers to backbone bond rotation (*37*)]. Our observation that the chemotype of a macrocycle, regardless of its side-chain identity, largely defines its globular shape suggests that diversity-oriented synthesis methodology seeking to span a wide range of shapes should focus on expanding the number of chemotypes that can be synthesized combinatorially as opposed to the diversity of side chains arrayed on a single or small number of chemotypes [the importance of macrocycle chemotype diversity has been noted in DNA-encoded libraries (*38*)]. Encouragingly for future therapeutic applications, nearly half of the macrocycles we characterized are membrane permeable, with $\log(P_{\mathrm{app}}) > -6.0$ in PAMPA, and this fraction would likely have been considerably higher had we explicitly designed for permeability, for example, by disfavoring compounds with exposed NH groups. The chemical space can readily be expanded by including more-diverse side chains or additional noncanonical backbones in the hashing step, and almost all compounds in this class can be readily synthesized using standard manual Fmoc-based solid-phase peptide synthesis followed by solution-phase cyclization (see methods). Our binder design results show that diverse biochemical functions can be introduced into de novo designed small macrocycles with binding selectivities (as in the HDAC6 case) rivaling those of advanced small-molecule therapeutic candidates.

The very large and diverse set of drug-like compounds described here opens up exciting avenues in drug discovery. The rigidity of the molecules could translate into higher on-target affinity, owing to the lower entropy loss upon target binding, and lower off-target binding, as there are fewer alternative states (*39*–*43*); Ro5 compliance during generation should lead to membrane permeability and other desirable pharmacological properties. Libraries of Ro5-compliant macrocycles populating single states can be screened in silico and/or experimentally to identify new lead compounds binding targets of interest. For targets for which small-molecule fragments are already known to bind, custom rigid macrocycle libraries incorporating this specific functionality can be readily generated by searching the hash tables for all closed macrocycles that incorporate the fragments. These approaches should allow systematic discovery of new macrocyclic therapeutic candidates.

REFERENCES AND NOTES

1. V. Sarojini, A. J. Cameron, K. G. Varnava, W. A. Denny, G. Sanjayan, *Chem. Rev.* **119**, 10318–10359 (2019).
2. J. T. Mhlongo, E. Brasil, B. G. de la Torre, F. Albericio, *Mar. Drugs* **18**, 203 (2020).
3. T. Degenkolb, W. Gams, H. Brückner, *Chem. Biodivers.* **5**, 693–706 (2008).
4. M. J. Ferracane *et al.*, *ACS Chem. Neurosci.* **11**, 1324–1336 (2020).
5. E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discov.* **7**, 608–624 (2008).
6. M. Muttenthaler, G. F. King, D. J. Adams, P. F. Alewood, *Nat. Rev. Drug Discov.* **20**, 309–325 (2021).
7. D. Garcia Jimenez, V. Poongavanam, J. Kihlberg, *J. Med. Chem.* **66**, 5377–5396 (2023).
8. Y. Goto, H. Suga, *Acc. Chem. Res.* **54**, 3604–3617 (2021).
9. K. Josephson, A. Ricardo, J. W. Szostak, *Drug Discov. Today* **19**, 388–399 (2014).
10. K. T. Mortensen, T. J. Osberger, T. A. King, H. F. Sore, D. R. Spring, *Chem. Rev.* **119**, 10288–10317 (2019).
11. G. Sangouard *et al.*, *Angew. Chem. Int. Ed.* **60**, 21702–21707 (2021).
12. S. Habeshian *et al.*, *Nat. Commun.* **13**, 3823 (2022).
13. A. Roy, E. Koesema, T. Kodadek, *Angew. Chem. Int. Ed.* **60**, 11983–11990 (2021).
14. S. A. Rettie *et al.*, bioRxiv 2023.02.25.529956 [Preprint] (2023); https://doi.org/10.1101/2023.02.25.529956.
15. J. L. Watson *et al.*, *Nature* **620**, 1089–1100 (2023).
16. P. Hosseinzadeh *et al.*, *Science* **358**, 1461–1466 (2017).
17. E. A. Coutsias, K. W. Lexa, M. J. Wester, S. N. Pollock, M. P. Jacobson, *J. Chem. Theory Comput.* **12**, 4674–4687 (2016).
18. R. Zubatyuk, J. S. Smith, J. Leszczynski, O. Isayev, *Sci. Adv.* **5**, eaav6490 (2019).
19. E. Caldeweyher *et al.*, *J. Chem. Phys.* **150**, 154122 (2019).
20. W. H. B. Sauer, M. K. Schwarz, *J. Chem. Inf. Comput. Sci.* **43**, 987–1003 (2003).
21. G. Bhardwaj *et al.*, *Nature* **538**, 329–335 (2016).
22. P. K. Madala, J. D. A. Tyndall, T. Nall, D. P. Fairlie, *Chem. Rev.* **110**, PR1–PR31 (2010).
23. C. J. Miller, B. E. Turk, *Trends Biochem. Sci.* **43**, 380–394 (2018).
24. J. Gavenonis, B. A. Sheneman, T. R. Siegert, M. R. Eshelman, J. A. Kritzer, *Nat. Chem. Biol.* **10**, 716–722 (2014).
25. M. Guharoy, P. Chakrabarti, *Bioinformatics* **23**, 1909–1918 (2007).
26. B. L. Sibanda, T. L. Blundell, J. M. Thornton, *J. Mol. Biol.* **206**, 759–777 (1989).
27. P. N. Lewis, F. A. Momany, H. A. Scheraga, *Biochim. Biophys. Acta* **303**, 211–229 (1973).
28. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **64**, 4–17 (2012).
29. G. Bhardwaj *et al.*, *Cell* **185**, 3520–3532.e26 (2022).
30. G. J. Saunders, A. K. Yudin, *Angew. Chem. Int. Ed.* **61**, e202206866 (2022).
31. A. A. Rzepiela, L. A. Viarengo-Baker, V. Tatarskii, R. Kombarov, A. Whitty, *J. Med. Chem.* **65**, 10300–10317 (2022).
32. P. Hosseinzadeh *et al.*, *Nat. Commun.* **12**, 3384 (2021).
33. T. Muramatsu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12997–13002 (2016).
34. H. Park, G. Zhou, M. Baek, D. Baker, F. DiMaio, *J. Chem. Theory Comput.* **17**, 2000–2010 (2021).
35. J. W. Johannes *et al.*, *ACS Med. Chem. Lett.* **8**, 239–244 (2016).
36. J. Lyu, J. J. Irwin, B. K. Shoichet, *Nat. Chem. Biol.* **19**, 712–718 (2023).
37. D. B. Diaz *et al.*, *Nat. Chem.* **13**, 218–225 (2021).
38. E. Koesema *et al.*, *Angew. Chem. Int. Ed.* **61**, e202116999 (2022).
39. K. T. Butler, F. J. Luque, X. Barril, *J. Comput. Chem.* **30**, 601–610 (2009).
40. A. C. Brueckner *et al.*, *J. Med. Chem.* **64**, 3282–3298 (2021).
41. J. W. Liebeschuetz, *J. Med. Chem.* **64**, 7533–7543 (2021).
42. A. N. Jain, A. C. Brueckner, A. E. Cleves, M. Reibarkh, E. C. Sherer, *J. Med. Chem.* **66**, 1955–1971 (2023).
43. V. K. Mulligan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2012800118 (2021).
44. P. Salveson *et al.*, Data for Expansive discovery of chemically diverse structured macrocyclic oligoamides, Zenodo (2023); https://doi.org/10.5281/zenodo.10109705.

ACKNOWLEDGMENTS

We thank L. Goldschmidt and K. VanWormer for maintaining the computational and wet lab infrastructure at the Institute for Protein Design. We thank L. Stewart for thoughtful discussion during the development of the methodologies. We thank J. Wang and L. Stewart for compound management support, PAMPA assay support, and research management support. Crystallographic work was based on research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences through the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. This work made use of the Integrated Molecular Structure Education and Research Center (IMSERC) at Northwestern University, which has received support from the NIH (1S10OD012016-01 and 1S10RR019071-01A1), Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the State of Illinois, and the International Institute for Nanotechnology (IIN). We thank the University of Washington's Hyak supercomputing team and the volunteers of the Rosetta@home project for providing compute resources. **Funding:** This work was supported with funds provided by the Audacious Project at the Institute for Protein Design (D.B., A.K.B., A.P.M., and A.K.); Juvenile Diabetes Research Foundation International grant 2-SRA-2018-605-Q-R (D.B. and P.J.S.); Helmsley Charitable Trust Type 1 Diabetes Program grant 2019PG-T1D026 (D.B. and X.L.); the Nordstrom Barrier Institute for Protein Design Directors Fund (P.J.S.); the Wu Tsai Translational Fund (P.J.S.); the Open Philanthropy Project Improving Protein Design Fund (D.B. and A.K.B.); Defense Threat Reduction Agency grant HDTRA1-19-1-0003 (D.B., M.Y.S., and P.M.L.); and the Higgins family (M.Y.S.). G.B. is supported by funds from the DARPA Harnessing Enzymatic Activity for Lifesaving Remedies (HEALR) program (HR001120S0052 contract HR0011-21-2-0012), the Defense Threat Reduction Agency (DTRA) (HDTRA1-19-1-0003), HHMI Emerging Pathogens Initiative, and start-up support from the University of Washington's Department of Medicinal Chemistry and Institute for Protein Design. **Author contributions:** P.J.S. and A.P.M. developed the hash-based closure methodology. A.P.M. developed the adaptive sampling methodology P.J.S. and A.P.M. developed binder-design and virtual screening methodology. P.J.S. generated designed macrocycles for structure determination experiments. P.J.S. and A.P.M. performed motif scaffolding and virtual screening calculations. A.K.B., A.K., H.N., and P.J.S. collected and processed data for x-ray crystallography. P.J.S. and M.Y.S. processed data for NMR spectroscopy. P.J.S. performed PAMPA experiments. G.G. performed Mpro inhibition experiments. P.M.L. performed serum stability assays. X.L. and P.M.L. orchestrated compound synthesis, handling, and testing at myriad CROs. P.J.S. and D.B. wrote and edited the manuscript. G.B. and D.B. offered supervision throughout the project. All authors read and contributed to the manuscript. **Competing interests:** Provisional patents (patent application numbers 63/369,441 and 63/333,077) covering the design methodology and designed macrocycles described in this paper have been filed by the University of Washington, Seattle. P.J.S., A.P.M., and D.B. are inventors on these patents. P.J.S., A.P.M., G.B., and D.B. are cofounders and shareholders of Vilya, an early-stage biotechnology company that has licensed the provisional patent. **Data and materials availability:** All data are available in the main text or supplementary materials. X-ray crystallographic structures were deposited into the Cambridge Crystallographic Data Centre under deposition numbers CCDC 2130362 (aas **1**), 2130361 (akak **1**), 2130363 (ahah **1**), 2130360 (aabc **2**), 2130359 (aaam **1**), 2130358 (aaam **6**), 2130356 (aabi **2**), 2130354 (aagb **1**), 2130352 (bbbb **3**), and 2203552 (ahah **2**). All design scripts are available through Zenodo (*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk1687
Materials and Methods
Figs. S1 to S24
Tables S1 to S15
Analytical Characterization Data
NMR Spectral Data
References (*45*–*52*)
MDAR Reproducibility Checklist

Submitted 7 August 2023; accepted 22 March 2024
10.1126/science.adk1687