science.sciencemag.org/content/369/6508/1227/suppl/DC1

# Supplementary Materials for

## A defined structural unit enables de novo design of small-molecule–binding proteins

Nicholas F. Polizzi* and William F. DeGrado*

*Corresponding author. Email: nicholas.polizzi@ucsf.edu (N.F.P.); william.degrado@ucsf.edu (W.F.D.)

Published 4 September 2020, *Science* **369**, 1227 (2020)
DOI: 10.1126/science.abb8330

**This PDF file includes:**



**Other Supplementary Materials for this manuscript include the following:**
(available at science.sciencemag.org/content/369/6508/1227/suppl/DC1)

Data S1 (.txt)
MDAR Reproducibility Checklist

## Materials and Methods

Curation of van der Mers

PDB database for vdM generation

We downloaded protein structures from the RCSB with 30% sequence homology, X-ray diffraction resolution ≤ 2.0 Å, and $R_{obs} \leq 0.3$. We used the program Reduce (*38*) to add hydrogens to the structures and to perform any necessary rotamer-flips of Asn, Gln, and His residues. We then used the program Molprobity (*39*) to obtain the Molprobity score for each structure. We subsequently constructed biological assemblies of the PDBs with Molprobity score ≤ 2, using the program Prody (*40*). The final list of accession codes/chain IDs for van der Mer (vdM) searching can be found in the supporting file, Data S1. The non-redundant structural database contains a total of 8743 PDBs with 9189 unique chains. Note that while we used biological assemblies to search for vdMs, we only searched through the non-redundant chains in the structure, such that contacts could be found across subunits of the assembly, without artificial duplication of vdMs.

Defining protein/chemical group contacts for vdM generation

We approximated chemical groups (CGs) as fragments of amino-acid sidechain or mainchain, in order to increase sampling statistics. For example, our database contains 348,067 residues contacting a carboxamide derived from Asn or Gln sidechains. Of these, 189,849 residues have interactions with carboxamide that are distant in sequence (> 7 amino acids away in the linear polypeptide chain), which avoids bias from nonspecific proximity effects. In this work, we further winnowed the number of interacting residues by considering only H-bonded interactions (85,750 residues). To define a vdM, we next categorize the interactions by residue type (e.g., 5,785 Tyr residues H-bond with a carboxamide).

We used the program Probe (*41*) to determine which amino-acid residues are in van der Waals (vdW) contact with a given chemical group (CG), as well as the nature of the contact (H-bond, close vdW contact, wide vdW contact). For example, to search for vdMs of carboxamide, we iterated through every Asn and Gln residue in each unique protein chain in the database. For each Asn and Gln residue in the chain, we used Probe to detect other residues in the biological assembly that are within vdW contact of the sidechain's carboxamide (e.g., CB, CG, OD1, ND2, HD21, HD22 atoms of Asn). To find vdMs of carbonyl (C=O), we used the backbone carbonyl of Gly and Ala residues. We then used only the subset of vdMs that formed H-bonded interactions. These vdMs were grouped in two ways: by superposition on mainchain for sampling, and by superposition on mainchain and chemical group coordinates for scoring (see text below and **Fig. S2**).

vdM cluster score

We scored vdMs based on their prevalence in the non-redundant protein structural database. Instead of aligning vdMs exactly by amino-acid backbone atoms, we performed a pair-wise all-against-all superposition of backbone (N, Ca, C) and CG atoms for every vdM of a particular amino-acid type. Using both backbone and CG in the superposition helps to alleviate the lever-arm effect, where small changes in backbone

coordinates lead to large changes in the location of a CG. The all-against-all pairwise RMSD matrix was used to cluster vdMs by RMSD < 0.5 Å, using a greedy clustering algorithm. Much of the interaction space sampled by proteins in our database is captured in a small number of these clusters. For example, only 31 clusters of Asp / carboxamide vdMs are needed to capture half of the observed interactions (**Fig. S3**). The corresponding curves for each amino acid are provided in **Fig. S4**.

A single cluster may use a variety of sidechain rotamers to position the chemical group in the same location relative to the residue's backbone atoms, and the sidechain dihedral angles of vdMs appear to follow the same distribution as canonical rotamers (**Fig. S1**), which may prove beneficial for generation of synthetic vdMs that employ non-canonical chemical groups. (Many non-canonical chemical groups, such as halogens, can be found in protein-cocrystal structures in the PDB with bound drugs. A limited set of vdMs could be generated based on these structures as well.)

We defined a cluster score ($C$) of a vdM as a quantitative measure for how representative that cluster's interaction geometry is for that residue type in the PDB. The score is based on placement of a CG relative to the protein backbone, since backbone and CG coordinates are the only coordinates involved in clustering. Sidechain conformation (rotamer) is not explicitly considered in the clustering and therefore not in $C$. We compare the size of the cluster $k$ to the average cluster size of that vdM type by $C^{(k)} = \ln N^{(k)} / \langle N \rangle$, where $N^{(k)}$ is the number of members in cluster $k$ and $\langle N \rangle$ is the average cluster size (**Fig. 1E**). Positive $C$ indicates the location of the CG relative to the backbone, represented by the cluster, is enriched relative to other locations of the CG. We used only interactions with positive $C$ in the design of ABLE.

vdM representatives for sampling

For sampling we used more fine-grained clusters, which would allow sampling over finer elements of conformational space (**Fig. S2**). To create these sub-clusters, we aligned the vdMs exactly by backbone atoms (N, Ca, C), and tightly clustered them (using a greedy clustering algorithm) by sidechain and CG coordinates (all-heavy-atom RMSD of 0.1 Å). The centroids of each cluster were used directly in sampling of vdMs on protein backbones. We refer to this fine-grained set as "vdM representatives"; by this definition, each vdM can be divided to a smaller number of vdM representatives. In this way, we can sample through representative members of a given vdM cluster without over-sampling very closely related members. In summary, we refer to a vdM as the cluster defined using a 0.5 Å RMSD cutoff, vdM cluster members as the individual members of the set, and vdM representatives as sub-clusters used for sampling.

Design protocol

Generation of parametric helical bundles

We aimed to create a highly stable protein that not only folds to the desired structure but also binds a ligand, which further restrains the sequence space in addition to the requirements for folding. We therefore sought to use a highly designable scaffold that can accommodate many sequences but is still tractable to computationally design from

scratch. Consequently, we parametrically generated a small set [32] of antiparallel 4-helix bundles using Crick parameters that are similar to those describing natural heme-binding proteins, such as helical bundles in cytochrome BC1, and to those describing non-natural porphyrin-binding proteins, such as the de novo bundle PS1 (*13*). Using the CCCP server (*10*), we sampled parameters on a grid that varied the bundle radius from 7.9 Å to 8.2 Å, and covaried the superhelical phases of two helices by 14°, resulting in bundles that had wide interfaces that varied between 108 and 120 ° (interhelical Cα distances of ~ 8.2 - 9.8 Å). These parameters were chosen because they result in highly designable backbones that can accommodate a variety of sequences (see structural bioinformatics below), as well as provide a variable-sized binding cavity for the ligand. Bundle parameters can be found in **Table S1**.

Structural bioinformatics of ABLE parametric backbone

We used the program Master (*42*) to query a structural database of approximately 20,000 protein crystal structures filtered at 50% sequence homology and with resolution < 2.5 Å ($R_{obs}$ < 0.3). A four-helix query of the database (10 residues each helix) returned 319 unique proteins with structural matches with Cα RMSD < 2 Å (**Table S2**). A query of the tightly interfaced helix-helix pair (10 residues each helix) of the parametric backbone returned 1466 unique proteins with structural matches with Cα RMSD < 0.7 Å.

The backbone of ABLE was defined by parametric design (*28, 43*), using a simple algebraic expression with a handful of adjustable parameters to define a highly symmetrical backbone with reasonable bond lengths and angles. The resulting backbone nevertheless served as a scaffold for design of proteins that bind a highly complex and asymmetric ligand. Curious about other proteins that might use this scaffold functionally, we probed the structural similarity of this backbone to natural four-helix bundle proteins in the PDB. We found hundreds of structural matches to a wide variety of proteins both natural and designed, with natural proteins ranging from the meiotic synaptonemal protein complex (*44*) to a superoxide oxidase (*45*); and with de novo proteins designed to form internal hydrogen bonds (*46*) or to bind porphyrins (*13*) (**Table S2**). One very recent structure (pdb 5xub) of a domain from a chemotaxis protein (*47*), deposited subsequently to the design of ABLE, binds citrate in approximately the same location of a four-helix bundle as the location of apixaban in ABLE. This collection of bundles illustrates the emergence of diverse complex functions from relatively minor (< 2 Å Cα RMSD) tweaks to an otherwise fully symmetrical scaffold.

Ligand conformation

We used the conformation of apixaban from the co-crystal structure with factor Xa (pdb 2p16, **Fig. S6**). We added hydrogens with the program Avogadro and created a Rosetta params file (see supplementary text) for use in flexible-backbone sequence design. This conformation is similar to its relaxed *in vacuo* conformation but is slightly higher in energy. The carboxamide of apixaban is internally H-bonded to the pyrazole nitrogen, creating a stable energy well for this conformation, which is observed in all small-molecule crystal structures of apixaban in the Cambridge Structural Database (**Fig. S7**). Indeed, the conformation of apixaban in complex with ABLE differs slightly from the factor Xa geometry (0.6 Å RMSD, **Fig. 4C**), but is almost identical to that observed in small-molecule crystal structures, as well as the quantum chemically optimized

geometry via DFT (*48*) using the B3LYP functional and 6-31G* basis set (**Fig. 4**, **Fig. S7**, and **Fig. S17**). We also computationally explored three higher-energy alternate conformations of apixaban, related by torsion about the methoxy-phenyl and the terminal 2-oxopiperidine moieties (**Fig. S7A**). For these conformations, we generated ligand-appended vdMs (see below) and searched for binding sites in the same way as in the design of ABLE and LABLE. These searches did not discover any better-scoring binding sites than those found using the apixaban conformer from factor Xa, so we did not experimentally investigate designs for these alternate-ligand conformations.

We ordered apixaban as a solid from Combi-Blocks and made DMSO stock solutions varying from 1mg/mL to 18 mg/mL.

COMBS strategy

The collective process of generating vdMs, loading vdMs on a backbone, sampling ligand poses, and selecting protein–ligand interactions is called COMBS (convergent motifs for binding sites). Below, we describe the process by which COMBS finds binding sites that achieve H-bonded interactions with the ligand apixaban (**Fig. S5**).

Interior vs exterior defined by convex hull algorithm

The design process starts with the coordinates of a poly-glycine backbone only. We used a restricted set of residues (H, S, T, Y, W) for sampling buried vdMs of carboxamide and carbonyl in the interior of the protein bundle and used a more expanded set for intermediate and exterior positions (H, S, T, Y, W, Q, N, D, E, R, K). We defined interior, intermediate, and exterior positions with a convex hull algorithm (*49*). We first make an all-Ala version of the protein, which defines the positions of Cβ atoms. The convex hull algorithm uses Cα and Cβ coordinates of the protein to define two surfaces. If the Cβ atom lies on the surface of the Cβ hull, that residue is exposed. If a Cβ atom lies in the interior of the Cα surface, then that residue is either buried or intermediate. Intermediate residues are those that are also part of the Cβ hull. The algorithm can limit the size of the radius of the sphere (alpha sphere) that is used to define the exterior surface, which limits the surface coarseness. We used an alpha-sphere size of 9 Å.

Sampling of vdMs on a backbone

We sample vdMs by aligning a set of vdM representatives (see above) to a backbone position. This has the effect of placing a chemical group (CG) in space relative to the backbone (sidechain is also placed). Similar to the program Probe, we use van der Waals radii of the atoms to define clashes of vdM sidechain and CG with the surrounding mainchain atoms, taking into account close approaches due to H-bonding. We do not sample vdMs one at a time in a conventional rotamer-sampling algorithm, but instead load them simultaneously onto a backbone scaffold to concurrently enumerate all possible CG locations (see Nearest neighbors graph of CGs). Multiple vdMs can occupy the same residue position on the backbone.

For sampling, we divided vdMs into 4 interaction types: 1) those making only backbone Cα and/or N-H contacts with the CG (called bbNH vdMs); 2) those making only backbone C=O contacts with the CG (called bbCO vdMs), 3) those making only sidechain contacts with the CG (called SC vdMs); and 4) those making both mainchain and sidechain contacts with the CG (called φψ vdMs). For each parametrically generated

helical bundle, we aligned vdMs of each category to the backbone by superposing, respectively, by 1) Cα, N, H atoms, 2) Cα, C, O atoms, 3) N, Cα, C atoms, and 4) N, Cα, C atoms. This allows for a finer sampling of vdMs that have interactions that are dependent on only φ, only ψ, or both φ and ψ. bbNH vdMs are φ-dependent, and bbCO vdMs are ψ-dependent. For sampling, we treated SC vdMs as φ/ψ independent, although φ/ψ dependence of the rotamer is implicitly considered when we remove any vdMs that clash with the mainchain. Because φψ vdMs are inherently φ/ψ dependent, we only sampled them from vdMs with φ/ψ in a bin of ± 30 ° of φ/ψ of the scaffold residue onto which they were aligned.

We sampled vdMs over a 14-residue span of each ~ 40 residue helix. We loaded vdMs onto 14 x 4 residue positions and created an array of CG coordinates for construction of a nearest neighbors graph, which we used to discover vdMs that are consistent with the position of a ligand.

Nearest neighbors graph of CGs

We construct a nearest-neighbors graph from the CG coordinates of the vdMs once they have been superimposed onto the backbone scaffold. For carboxamide, we used an RMSD of 1.0 Å for the CG (Cb, Cg, Od1, Nd2 atoms of Asn, and Cg, Cd, Oe1, Ne2 atoms of Gln). For carbonyl (backbone C and O atoms of Gly and Ala), we used an RMSD of 0.7 Å. We used the nearest-neighbors implementation in the Python package sci-kit learn. This allows for very fast lookups of neighbors given query coordinates, which we take from placed ligands (see below). The neighbors tell us precisely which vdMs place a chemical group within the RMSD threshold of the query coordinates, as well as the RMSD distance of each from the query. The next step in the design process is to determine which of these neighboring vdMs possess sidechains that do not clash with the placed ligand, and then to score the clash-free remainder by *C* (see above).

Ligand-placement algorithms

Previous computational approaches to sample ligand positions have focused on either geometric overlap of entire ligands (*6, 50, 51*) or on ligand placement with one user-defined contact (*23*). For example, after sampling ligand-appended rotamers on protein backbones, candidate binding sites were defined as those that placed the full ligand in the same region of space (*6*). These approaches suffer from the lever-arm effect, where small deviations in protein–ligand contact geometry amplify to large changes of the ligand position remote from the contacting region. Massive amounts of sampling are required to overcome the lever-arm effect (*4, 6, 8, 23*), yet only a fraction of the total possible conformational space is available for sampling on a reasonable timeframe, even on large computing clusters. COMBS instead uses a set of ligand-superimposed vdMs to initially place a ligand in the binding site (see below) but then looks for nearest-neighbors vdMs of the ligand's chemical groups, instead of matches to full ligand locations. COMBS currently searches through static conformers only, such that searching through multiple conformers of a ligand requires the generation of a different set of ligand-superimposed vdMs for each conformer. Searches through multiple conformers can then be run in parallel.

Generation of ligand poses

To generate ligand placements relative to the protein backbone, we first curate a set of vdMs with the ligand superimposed by the CG. We remove all vdM / ligand combinations that are clashing after superposition. We then load this set of ligand-superimposed vdMs onto the backbone scaffold in the same way we load vdMs. This has the advantage of placing the ligand with a least one vdM-derived CG contact, that of its superimposed vdM. We remove any ligand-superimposed vdMs with ligand or sidechain that is clashing with the backbone. We further remove any ligand-superimposed vdMs based on ligand burial. For design of ABLE and LABLE, we required at least 60 percent of apixaban's apolar heavy atoms to be buried in the interior of the protein, as defined by the convex hull (see above).

With the coordinates of the other CGs within the ligand now defined relative to the backbone, we use these coordinates as queries to the nearest-neighbors graph of carboxamide and carbonyl. We look for overlap of the ligand's CGs in their respective nearest-neighbors graphs instead of overlap of an entire ligand in order to reduce the lever arm effect, which amplifies small deviations in local geometry to affect large swings in distant parts of a ligand. The use of CG graphs allows us to find binding interactions for a particular ligand location consistent with small local deviations in the interactions that would otherwise be missed by a search for full ligand overlap. By sampling the ligand position with superposed ligands onto vdMs, we experience the lever arm effect only once (during the superposition), instead of multiple times (one time per CG) in the ligand.

Selection of ligand poses for further design

We selected poses of apixaban based on ligand burial and satisfaction of H-bonding constraints to its buried CGs. We required that the two carbonyls and the carboxamide of apixaban be engaged in a vdM-derived H-bond if buried in the interior of the protein. We selected individual vdMs (among all the nearest neighbors) for a ligand pose based on maximizing $C$ while avoiding vdW clashes between vdM sidechains. We chose 6 poses based on apixaban burial and $\sum C$ that explored three distinct placements of apixaban (**Fig. S8**). We further checked for the robustness of the pose by clustering ligand poses by ligand position across the 32 bundles. Poses from large clusters with the same vdM-derived interactions suggested these interactions could be consistent with small-scale structural fluctuations on the order of 1 Å Cα RMSD.

Flexible backbone sequence design

After vdM-derived ligand placement and H-bonded interactions were found to apixaban, we performed a custom protocol for flexible-backbone sequence design in the program Rosetta (*26*) (linux version 2018.33.60351). We froze the identities and rotamers of the H-bonded residues, and constrained the H-bond distances using a harmonic potential. We generated a parameter file for apixaban for use in Rosetta, which defines its partial charges (see supplemental text). We did not allow the ligand conformer to be flexible during design.

We automatically generated Rosetta residue files based on burial and secondary structure of each position in the backbone. To do so, we applied the convex hull algorithm described above, as well as the secondary structure assignment program DSSP, to the entire PDB dataset (~9,000 proteins) to create burial and secondary structure

propensities for each residue type, based on backbone coordinates only. The propensity is defined as $p = f_{aa}(\text{burial, ss}) / f_{aa}$ where $f_{aa}$(burial, ss) is the frequency that amino acid *aa* occurs in that burial assignment (exposed, intermediate, or buried) with secondary structure ss, and $f_{aa}$ is the frequency of the amino acid *aa* in the database. We used residues at each position that had a burial and secondary structure propensity $p \geq 0.9$. For 3 of the 6 designs, including that of ABLE, we allowed Ala, Ser, Thr, and Val residues at solvent exposed positions during design to lower the surface polarity in order to promote crystallization. Scripts for flexible-backbone sequence design can be found in supplementary text below. The outputted backbones (500 total) varied on average from their starting structure by ~ 1 Å Cα RMSD. We selected designs for advancement to the next stage of computation by considering the packing of the core residues (pstat score in Rosetta) and the overall energy (ref2015 weights).

Loop construction

Loops connecting helices are selected from a database of natural α-helical protein structures and spliced onto the backbone to minimize Cα distance with the helices (*52*). The loop sequences were allowed to vary in the flexible backbone design process, with the set of possible residues selected in the automated fashion describe above.

Negative design of surface residues

We used a simple Monte Carlo protocol to bias the desired folded topology, by searching for charged surface residues that stabilize the desired topology and destabilize the reverse topology (*53*). The protocol results in a surface pattern of negatively and positively charged residues. We modified the Rosetta residue file to account for this surface patterning by disallowing the opposite charge at positions specified by the surface pattern (The residues were still allowed to be neutral and polar.). We find that this protocol results in bundles that exhibit well-defined *ab initio* folding funnels with single minima (e.g. **Fig. S20A**). Without this negative design element, folding funnels often show multiple minima representing different folded topologies. Scripts for surface patterning can be found in topology.py within the COMBS software package.

Model selection

We selected final, single-chain designs (among 500 total outputted models for each of the 6 designs) by considering the packing of the core residues (pstat score in Rosetta) and the overall energy (ref2015 weights). We used the convex hull algorithm mentioned above and a custom python script based on the program Probe to detect any buried residues with polar atoms not engaged in an H-bond, such as Tyr or Trp residues. We selected designs that did not feature any "unsatisfied" H-bonding residues. Computational models of the designs are freely available at the online repository zenodo (*37*).

*Ab initio* folding

Rosetta *ab initio* folding (*54*) was performed on the final designed sequences. The command line input for folding simulations can be found in the supplementary text. RMSD was calculated to all Cα atoms of the input model. Of the 6 designed sequences, only 2 were predicted to fold to a structure that maintained an open, solvent-accessible

binding site, ABLE and LABLE (**Fig. S20**). Three of the other designs showed a collapsed hydrophobic binding site with no space for binding, and one design was predicted not to fold. We expressed and characterized all 6 of them. Interestingly, the two designs with a predicted open binding site tightly bound apixaban. The other designs did not bind apixaban, suggesting that *ab initio* folding predictions of binding-site collapse are a good indicator for design success. All designs were helical as measured by circular dichroism spectroscopy (**Fig. S9**).

Code availability

The code for COMBS is available at github (https://github.com/npolizzi/combs_pub). The scripts for flexible-backbone sequence design in Rosetta can be found in supplementary text.

Protein expression

The genes coding for the 6 protein sequences were ordered from GenScript, and were cloned into the IPTG-inducible pet-11a plasmid (cloning site NdeI-BamHI). The sequence of each design also coded for an N-terminal 6xHis-tag followed by a TEV protease cleavage sequence.

Cloned gene sequence of ABLE

ATGCACCACCACCACCACCACGAAAACCTGTACTTCCAGAGCGTGAAGAGCG
AGTATGCGGAAGCTGCGGCGGTTGGTCAAGAAGCGGTGGCGGTTTTCAACAC
CATGAAGGCGGCGTTTCAGAACGGCGATAAAGAGGCGGTTGCGCAATACCTG
GCGCGTCTGGCGAGCCTGTATACCCGTCACGAGGAACTGCTGAACCGTATCC
TGGAAAAGGCGCGTCGTGAGGGTAACAAAGAAGCGGTGACCCTGATGAACG
AGTTCACCGCGACCTTTCAGACCGGCAAGAGCATTTTCAACGCGATGGTTGC
GGCGTTTAAAAACGGCGACGATGACAGCTTTGAGAGCTACCTGCAGGCGCTG
GAAAAGGTGACCGCGAAAGGCGAGACCCTGGCGGACCAAATCGCGAAAGCG
CTGTAA

Expressed protein sequence of ABLE

MHHHHHHENLYFQ/SVKSEYAEAAAVGQEAVAVFNTMKAAFQNGDKEAVAQY
LARLASLYTRHEELLNRILEKARREGNKEAVTLMNEFTATFQTGKSIFNAMVAAF
KNGDDDSFESYLQALEKVTAKGETLADQIAKAL

where the "/" defines the cleavage site of TEV protease. TEV-cleaved ABLE is 126 residues. The plasmids were transfected into *E. coli* BL21(DE3) cells (Invitrogen), which were grown in LB/ampicillin media until OD @ 600 nm = 0.6. The cells were then induced with IPTG and allowed to grow for 4 more hours. Cells were then centrifuged and frozen. The frozen cell pellets were thawed and lysed by sonication, purified by Ni NTA affinity column (Invitrogen), and purified protein was confirmed by gel electrophoresis. The buffer was exchanged to a TEV protease buffer (5 mM DTT, 50 mM Tris, 0.5 mM EDTA, pH 8.0), and proteins were incubated with His-tagged TEV protease for 1 day at room temperature. The cleaved protein was collected from the flow-through of a Ni NTA column and concentrated in a stock of 50 mM NaPi, 100 mM NaCl, pH 7.4 buffer. Both TEV-cleaved and His-tagged proteins were used in experiments, as

they showed no significant differences in binding. ABLE had an approximate yield of 200 mg/L.

Expressed protein sequence of LABLE

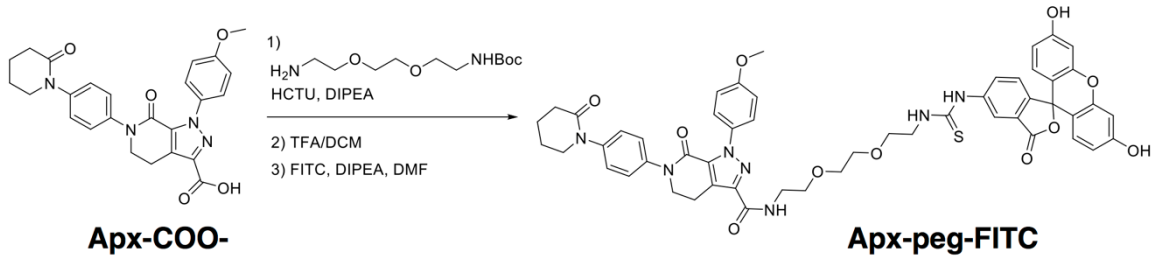
MHHHHHHENLYFQ/SSEEDQLDKLLKEFKAVFNHGKKVFEQMKQAWERMASA
FKNNQNASELLDELAKYISELNEVTKHGQELAKKIRDAAERANASDEWRKTFDE
AAKVGQAFIKTWEAFVRTWEAFEKAYKNGDDEKNLKAYLEQLKKYLEQLESYL
RQHDELLQKLEELWKKIKS

Construction of ABLE mutants

Single amino-acid mutations to the ABLE sequence were made via the Agilent QuikChange II protocol, with primers designed using the QuikChange primer design program. Sequences of mutants were confirmed by Sanger sequencing (GeneWiz) of transformed bacterial colonies.

Synthesis of Apx-peg-FITC

Apixaban-peg-FITC was synthesized from apixaban acid ($ApxCOO^-$) by coupling with Boc-$(PEG)_2$-amine followed by deprotection and reaction with FITC (**Scheme S1**). To a solution of apixaban acid (200 mg, 0.43 mmol) in DMF (2 mL) were added Boc-$(PEG)_2$-amine (108 mg, 0.43 mmol), DIPEA (174 uL, 1 mmol) and HCTU (169 mg, 041 mmol). The mixture was stirred at room temperature for 3 h and diluted with ethyl acetate. The organic layer was successively washed with 1 M HCl, sat. $NaHCO_3$, and brine. After drying over $Na_2SO_4$, the mixture was concentrated under reduced pressure and a solution of TFA in DCM (50%, 5 mL) was added. The mixture was stirred for 1h and the volatiles were removed under reduced pressure. To the solution of crude amine in DMF (4 mL) were added DIPEA (134 uL) and FITC (136 mg). After stirring 2 h, the mixture was diluted with ethyl actetate and washed with sat. $NaHCO_3$. The organic layer was concentrated and purified by RP-HPLC. $^1H$ NMR (DMSO-d6, 300 MHz) 8.30 (1H, br s), 7.73 (1H, d, J=8.0 Hz), 7.49 (2H, d, J=8.9 Hz, 1H), 7.33 (2H, d, J=8.8 Hz), 7.26 (2H, d, J=8.8 Hz), 7.16 (1H, d, J=8.3 Hz), 6.98 (2H, d, J=9 Hz), 6.67 (2H, s), 6.56 (m, 4H), 4.04 (m, 2H), 3.79 (s, 3H), 3.5-3.53 (m, 12H), 3.42-3.35 (m, 2H), 3.25-3.15 (m, 2H), 2.38 (2H, br s), 1.85-1.83 (4H, m) ESI-MS ($MH^+$) 980.5




**Scheme S1.** Synthetic scheme of Apx-peg-FITC used for fluorescence polarization experiments.

Determination of binding dissociation constant

We used spectral titration and fluorescence polarization experiments to determine the binding dissociation constants for ABLE. ABLE was purified via HPLC (C4 reverse

phase column), lyophilized, and reconstituted in buffer (50 mM NaPi, 100 mM NaCl, pH 7.4). Aliquots of apixaban from 2 mM, 1 mM, or 0.5 mM stocks in DMSO were serially added to 2 mL solutions of ABLE at 20 μM, 10 μM, and 5 μM concentration, respectively. (Final DMSO concentration was kept below 2%.) Absorbance changes at 305 nm, due to the restricted torsional conformation of apixaban in the bound state, were fit to Equation 1 using a single-site, protein-ligand binding model for the [Apx·ABLE] complex (Equation 2) (**Fig. 3G, H** and **Fig. S11**). Global parameters of the fit were $\Delta\varepsilon_{305}$, $K_D$, and $N$, where $\Delta\varepsilon_{305}$ is the change in extinction coefficient at 305 nm of the bound complex relative to free apixaban and protein. We used $\varepsilon_{305}$, the extinction coefficient of apixaban at 305 nm, as a local fitting parameter for each concentration, and the results of these locally fit extinction coefficients were within experimental error of each other (also agreeing with that measured for apixaban alone). Results of the fit are listed in **Fig. S11** (legend). The best-fit value of $N$ was 1.4 (stoichiometry of 0.7 ligand to 1 protein); deviation from unity was either due to experimental errors in concentrations or a small population of the protein that has less affinity toward apixaban. Multiple different starting parameters converged onto those listed in **Fig. S11**. Individual fitting at different concentrations gave good fits, but a limited degree of covariation between $N$ and $K_D$. This was eliminated by global fitting at multiple concentrations. Randomness of the residuals confirmed goodness of fit. While ABLE contains no Trp residues, the presence of several Trp residues in LABLE that spectrally overlap with apixaban precluded the spectral titration experiment with LABLE.

We performed fluorescence anisotropy experiments (*55, 56*) of ABLE and LABLE using a FITC fluorophore conjugated to apixaban as the fluorescent probe (**Fig. S12** and **Scheme 1**). We serially diluted a concentrated protein solution containing 25 nM Apx-peg-FITC, holding constant the concentration of Apx-peg-FITC. Parallel and perpendicularly polarized emission at 528 nm (10 nm slit width, 510 nm long pass filter) were integrated for 10 seconds after 485 nm (5 nm slit width) excitation of the FITC fluorophore of Apx-peg-FITC. The data were fit to a single-site binding model (Equation 2, with $N = 1$) (**Fig. S12B**). We performed ligand competition experiments (*55, 56*) by adding aliquots of a concentrated ligand stock in DMSO into an approximately 50% bound complex of protein and Apx-peg-FITC (25 nM). The decrease in anisotropy as a function of competing ligand was fit to a competitive binding model (*55, 56*) (**Fig. 4D and Fig. S12**, **C** and **D**). Because effects of DMSO on the anisotropy of bound ABLE / Apx-peg-FITC complex become more pronounced at high concentrations of DMSO (> 4 %), we fit only the initial data points at low DMSO concentration (% DMSO < 2 %, **Fig. 4D**). Comparison of a fit to the full titration with apixaban as competitor showed similar results (**Fig. S12C**), indicating that DMSO does not significantly affect the $K_D$ of apixaban to ABLE. DMSO may have a minor effect on competitive binding with weaker competitors (i.e., apixaban$COO^-$ and rivaroxaban), where high concentrations of competitor are needed to decrease anisotropy. As such, the reported dissociation constants of these weaker competitors can be viewed as a lower limit to the $K_D$. Fluorescence polarization competition with apixaban as competitor gives a $K_D$ of 7 (± 1) μM for the apixaban / ABLE complex (**Fig. 4D**), which agrees with that of the more precise spectral titration [$K_D$ = 5 (± 1) μM]. Binding experiments were performed at least twice to confirm reproducibility, and the reported errors correspond to the uncertainty in the fit parameters.

**Equation 1**

$$OD_{305}([\mathbf{Apx}]_T) =$$

$$OD_{305}([\mathbf{Apx}]_T = 0) + \varepsilon_{305}[\mathbf{Apx}]_T + \Delta\varepsilon_{305}[\mathbf{Apx} \cdot \mathbf{ABLE}]([\mathbf{ABLE}]_T, [\mathbf{Apx}]_T, K_D, N)$$

**Equation 2**

$$[\mathbf{Apx} \cdot \mathbf{ABLE}]([\mathbf{ABLE}]_T, [\mathbf{Apx}]_T, K_D, N) =$$

$$\frac{1}{2}\left[K_D + [\mathbf{Apx}]_T + \frac{[\mathbf{ABLE}]_T}{N} - \sqrt{\left(K_D + [\mathbf{Apx}]_T + \frac{[\mathbf{ABLE}]_T}{N}\right)^2 - 4[\mathbf{Apx}]_T\frac{[\mathbf{ABLE}]_T}{N}}\right]$$

Steady-state electronic absorption spectroscopy

Electronic absorption spectra were collected using a HP 8453 spectrophotometer in 1 cm quartz optical cells. The noise level of the instrument was maintained at 0.1 mOD.

Thermal stability

CD spectra were collected on a Jasco J-810 CD spectrometer in a 0.1 cm path length quartz cuvette (**Fig. S9**). Full spectra were collected from 200 nm to 250 nm in continuous scanning mode, with a band width of 2 nm, scanning speed of 50 nm/min, data pitch of 2 nm, response of 8 sec (standard sensitivity), and an average of 3 accumulations. Designs were prepared in 6 or 12 μM concentrations in 50 mM NaPi pH 7.4, 100 mM NaCl buffer. Temperature-dependent data of liganded and unliganded ABLE (**Fig. S15)** were collected at 222 nm from 20 to 95 °C with an interval of 5 °C and an increase rate of 3 °C/minute, and an average of 5 accumulations. ABLE was prepared at 10 μM in 50 mM NaPi pH 7.4, 100 mM NaCl buffer. Apixaban-bound ABLE solution contained 30 μM apixaban (0.27% final concentration of DMSO). To aid in direct comparison to the bound complex, the unliganded protein solution also contained 0.27% DMSO.

Oligomerization state

We determined oligomerization state by size exclusion chromatography on an Akta Pure FPLC using a Superdex 75 5/150 analytical column. Both drug-free- and drug-bound ABLE eluted at elution volumes equivalent to its molecular weight (**Fig. S14**).

X-ray crystallography

We screened crystallization conditions for unliganded- and liganded ABLE in 96-well hanging drop trays from Hampton Research. His-tag-cleaved ABLE was concentrated in water at 30 mg/mL. For preparation of drug-bound ABLE, we added 1.1 equivalents of apixaban from a concentrated DMSO stock, resulting in a DMSO

concentration of 12%. Both drug-bound- and drug-free ABLE readily crystallized in multiple conditions from Hampton Peg Ion 2 screen and the ammonium sulfate ($AmSO_4$) screen. We looped the crystals and submerged them in paratone cryoprotectant before freezing them in liquid nitrogen. Diffraction data was collected remotely using an Eiger 16M detector at the 24-IDE (NE-CAT) beamline of the Advanced Photon Source at Fermi Lab. Multiple conditions gave high-quality diffraction with resolution below 2 Å. The well condition that gave the best diffraction for both drug-bound and drug-free ABLE was 2.6 M AmSO4, 0.1 M Na acetate. Crystals of both proteins diffracted to 1.3 Å resolution in this condition. Reflections were processed and merged using RAPD (https://rapd.nec.aps.anl.gov/). The structures were solved by molecular replacement with Phaser in Phenix, using the design model with apixaban removed. The structures were iteratively refined in Phenix and Coot. Diffraction data and refinement statistics of apixaban-bound- and drug-free ABLE are shown in **Table S3**. Crystals of the H49A mutant of ABLE were grown in a 24-well hanging drop plate with well solution 0.03 M Citric acid, 0.07 M BIS-TRIS propane / pH 7.6 with 20% w/v Polyethylene glycol 3,350 (Hampton PEG/Ion 2 screen condition 40). Crystals were looped in paratone and frozen in liquid nitrogen, and diffraction data to 1.6 Å resolution was collected on a PILATUS3 6M detector at the 8.3.1 beamline at the Advanced Light Source and Lawrence Berkeley National Labs. Reflections were processed and merged via XDS program and the structure was solved by molecular replacement with Phaser in Phenix, using the drug-free ABLE protein structure as the search model. The structure was iteratively refined in Phenix and Coot. Diffraction data and refinement statistics of unliganded H49A ABLE are shown in **Table S4.**

**Supplementary Text**

Command lines and flags for flexible backbone design algorithm

```
~/rosetta_bin_linux_2018.33.60351_bundle/main/source/bin/rosetta_script
s.static.linuxgccrelease -database
~/rosetta_bin_linux_2018.33.60351_bundle/main/database/ -s input.pdb -
nstruct 500 -extra_res_fa APX.params -parser:protocol
flexbb_design_protocol.xml -packing:multi_cool_annealer 10 -
packing:linmem_ig 10
```

RosettaScript for flexible backbone design (flexbb_design_protocol.xml)

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="ref15" weights="ref2015">
      <Reweight scoretype="atom_pair_constraint" weight="1"/>
    </ScoreFunction>
    <ScoreFunction name="ref15_1" weights="ref2015">
      <Reweight scoretype="aa_composition" weight="1" />
      <Reweight scoretype="netcharge" weight="1.0" />
      <Reweight scoretype="atom_pair_constraint" weight="1"/>
      <Set aa_composition_setup_file="no_met_thr_ser_asn.comp" />
      <Set netcharge_setup_file="netcharge.charge" />
    </ScoreFunction>
```

```
  </SCOREFXNS>
  <RESIDUE_SELECTORS>
  </RESIDUE_SELECTORS>
  <TASKOPERATIONS>
    <InitializeFromCommandline name="ifcl"/>
    <ReadResfile name="resfile" filename="resfile.txt"/>
    <ExtraRotamersGeneric name="extrachi" ex1="1" ex2="1"
                ex1_sample_level="1" ex2_sample_level="1"
                extrachi_cutoff="14"/>
    <IncludeCurrent name="include_curr" />
  </TASKOPERATIONS>
  <FILTERS>
    <PackStat name="pstat" confidence="0" threshold="0" repeats="10"/>
    <PackStat name="pstat_mc" threshold="0" repeats="10"/>
    <ScoreType name="total_score_1" scorefxn="ref15_1" score_type="total_score"
          threshold="0"/>
   </FILTERS>
  <MOVERS>
    <ConstraintSetMover name="atomic" cst_file="vdM_Hbonds.cst"/>
    <PackRotamersMover name="pack" scorefxn="ref15_1"
              task_operations="ifcl,resfile,include_curr,extrachi"/>
    <PackRotamersMover name="pack_fast" scorefxn="ref15_1"
              task_operations="ifcl,resfile,include_curr"/>
    <MinMover name="min_bb" scorefxn="ref15" tolerance="0.0000001"
max_iter="1000" chi="false" bb="true">
      <MoveMap name="map_bb">
        <Span begin="1" end="125" bb="true" chi="false" />
        <Span begin="126" end="999" bb="false" chi="false"/>
      </MoveMap>
    </MinMover>
    <Idealize name="idealize"/>
    <MinMover name="min_sc" scorefxn="ref15" tolerance="0.0000001"
max_iter="1000" chi="true" bb="false">
      <MoveMap name="map_sc">
        <Span begin="1" end="125" bb="false" chi="true" />
        <Span begin="126" end="999" bb="false" chi="false"/>
      </MoveMap>
    </MinMover>
    <MinMover name="min_sc_bb" scorefxn="ref15" tolerance="0.0000001"
max_iter="1000" chi="true" bb="true">
      <MoveMap name="map_sc_bb">
        <Span begin="1" end="125" bb="true" chi="true" />
        <Span begin="126" end="999" bb="false" chi="false"/>
      </MoveMap>
    </MinMover>
    <ParsedProtocol name="parsed_pack_fast" >
```

```
        <Add mover_name="pack_fast"/>
        <Add mover_name="min_bb"/>
      </ParsedProtocol>
      <ParsedProtocol name="parsed_pack" >
        <Add mover_name="pack"/>
        <Add mover_name="min_bb"/>
        <Add mover_name="min_sc"/>
      </ParsedProtocol>
      <GenericMonteCarlo name="pack_mc" preapply="0" trials="3" temperature="0.03"
                  filter_name="pstat_mc" sample_type="high"
mover_name="parsed_pack">
          <Filters>
            <AND filter_name="total_score_1" temperature="15" sample_type="low"/>
          </Filters>
      </GenericMonteCarlo>
      <GenericMonteCarlo name="pack_fast_mc" preapply="0" trials="2"
temperature="0.03"
                  filter_name="pstat_mc" sample_type="high"
mover_name="parsed_pack_fast">
          <Filters>
            <AND filter_name="total_score_1" temperature="15" sample_type="low"/>
          </Filters>
      </GenericMonteCarlo>
    </MOVERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
      <Add mover="atomic"/>
      <Add mover_name="parsed_pack_fast"/>
      <Add mover_name="pack_fast_mc"/>
      <Add mover_name="pack_mc"/>
      <Add mover_name="min_sc_bb"/>
      <Add filter_name="pstat"/>
    </PROTOCOLS>
    <OUTPUT scorefxn="ref15_1"/>
</ROSETTASCRIPTS>
```

Contents of constraint file (vdM_Hbonds.cst) for ABLE

```
AtomPair NE2 48A O3 1X HARMONIC 2.7 0.3
AtomPair OG1 111A O2 1X HARMONIC 2.8 0.3
AtomPair HG1 111A O2 1X HARMONIC 2.0 0.3
AtomPair HE2 48A O3 1X HARMONIC 1.7 0.3
```

Contents of Apixaban parameters file (APX.params) for use in Rosetta

```
NAME APX
IO_STRING APX Z
```

```
TYPE LIGAND
AA UNK
ATOM  C8  CNH2  X   0.27
ATOM  O3  ONH2  X   -0.27
ATOM  N5  Npro  X   -0.27
ATOM  C7  aroC  X   0.04
ATOM  C22 aroC  X   -0.04
ATOM  C18 aroC  X   -0.04
ATOM  C16 aroC  X   0.04
ATOM  N2  Npro  X   -0.27
ATOM  C19 CNH2  X   0.22
ATOM  O2  ONH2  X   -0.28
ATOM  C23 CH2   X   0.02
ATOM  C25 CH2   X   -0.04
ATOM  C21 CH2   X   -0.04
ATOM  C20 CH2   X   0.02
ATOM  H14 Hapo  X   0.05
ATOM  H15 Hapo  X   0.05
ATOM  H16 Hapo  X   0.03
ATOM  H17 Hapo  X   0.03
ATOM  H23 Hapo  X   0.03
ATOM  H24 Hapo  X   0.03
ATOM  H19 Hapo  X   0.04
ATOM  H20 Hapo  X   0.04
ATOM  C14 aroC  X   -0.04
ATOM  C44 aroC  X   -0.04
ATOM  H25 Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  H13 Haro  X   0.06
ATOM  H18 Haro  X   0.06
ATOM  C24 CH2   X   0.02
ATOM  C17 CH2   X   -0.01
ATOM  C12 aroC  X   0.02
ATOM  C10 aroC  X   0.15
ATOM  N6  Nhis  X   -0.16
ATOM  N1  Npro  X   -0.23
ATOM  C4  aroC  X   0.07
ATOM  C3  aroC  X   -0.03
ATOM  C2  aroC  X   -0.02
ATOM  C1  aroC  X   0.12
ATOM  O4  OH    X   -0.50
ATOM  C15 CH3   X   0.08
ATOM  H8  Hapo  X   0.07
ATOM  H9  Hapo  X   0.07
ATOM  H10 Hapo  X   0.07
ATOM  C6  aroC  X   -0.02
```

```
ATOM  C5  aroC  X   -0.03
ATOM  H5  Haro  X   0.06
ATOM  H6  Haro  X   0.07
ATOM  H1  Haro  X   0.07
ATOM  H2  Haro  X   0.06
ATOM  C13 aroC  X   0.13
ATOM  C11 CNH2  X   0.26
ATOM  O1  ONH2  X   -0.27
ATOM  N3  NH2O  X   -0.32
ATOM  H3  Hpol  X   0.15
ATOM  H4  Hpol  X   0.15
ATOM  H11 Hapo  X   0.03
ATOM  H12 Hapo  X   0.03
ATOM  H21 Hapo  X   0.05
ATOM  H22 Hapo  X   0.05
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   O4  1
BOND_TYPE  C1   C6  4
BOND_TYPE  N1   C4  1
BOND_TYPE  N1   N6  4
BOND_TYPE  N1   C13 4
BOND_TYPE  O1   C11 2
BOND_TYPE  C2   C3  4
BOND_TYPE  N2   C16 1
BOND_TYPE  N2   C19 4
BOND_TYPE  N2   C20 1
BOND_TYPE  O2   C19 2
BOND_TYPE  C3   C4  4
BOND_TYPE  N3   C11 4
BOND_TYPE  O3   C8  2
BOND_TYPE  C4   C5  4
BOND_TYPE  O4   C15 1
BOND_TYPE  C5   C6  4
BOND_TYPE  N5   C7  1
BOND_TYPE  N5   C8  4
BOND_TYPE  N5   C24 1
BOND_TYPE  N6   C10 4
BOND_TYPE  C7   C22 4
BOND_TYPE  C7   C44 4
BOND_TYPE  C8   C13 1
BOND_TYPE  C10  C11 1
BOND_TYPE  C10  C12 4
BOND_TYPE  C12  C13 4
BOND_TYPE  C12  C17 1
BOND_TYPE  C14  C16 4
BOND_TYPE  C14  C44 4
```

```
BOND_TYPE  C16  C18 4
BOND_TYPE  C17  C24 1
BOND_TYPE  C18  C22 4
BOND_TYPE  C19  C23 1
BOND_TYPE  C20  C21 1
BOND_TYPE  C21  C25 1
BOND_TYPE  C23  C25 1
BOND_TYPE  C2   H1  1
BOND_TYPE  C3   H2  1
BOND_TYPE  N3   H3  1
BOND_TYPE  N3   H4  1
BOND_TYPE  C5   H5  1
BOND_TYPE  C6   H6  1
BOND_TYPE  C14  H7  1
BOND_TYPE  C15  H8  1
BOND_TYPE  C15  H9  1
BOND_TYPE  C15  H10 1
BOND_TYPE  C17  H11 1
BOND_TYPE  C17  H12 1
BOND_TYPE  C18  H13 1
BOND_TYPE  C20  H14 1
BOND_TYPE  C20  H15 1
BOND_TYPE  C21  H16 1
BOND_TYPE  C21  H17 1
BOND_TYPE  C22  H18 1
BOND_TYPE  C23  H19 1
BOND_TYPE  C23  H20 1
BOND_TYPE  C24  H21 1
BOND_TYPE  C24  H22 1
BOND_TYPE  C25  H23 1
BOND_TYPE  C25  H24 1
BOND_TYPE  C44  H25 1
CHI 1  C2   C1   O4   C15
CHI 2  N6   N1   C4   C3
CHI 3  C18  C16  N2   C19
CHI 4  C8   N5   C7   C22
CHI 5  C12  C10  C11  O1
NBR_ATOM  C8
NBR_RADIUS 10.865233
ICOOR_INTERNAL    C8     0.000000    0.000000    0.000000   C8    O3    N5
ICOOR_INTERNAL    O3     0.000000  180.000000    1.231236   C8    O3    N5
ICOOR_INTERNAL    N5     0.000000   56.912601    1.396232   C8    O3    N5
ICOOR_INTERNAL    C7     1.861545   58.604695    1.397270   N5    C8    O3
ICOOR_INTERNAL    C22  116.605296   62.137596    1.469823   C7    N5    C8
ICOOR_INTERNAL    C18 -175.989024   61.379327    1.445469   C22   C7    N5
ICOOR_INTERNAL    C16    0.813034   61.902299    1.464716   C18   C22   C7
```

```
ICOOR_INTERNAL   N2  179.326815   62.708149   1.369606  C16  C18  C22
ICOOR_INTERNAL   C19  84.211713   61.381606   1.362129  N2   C16  C18
ICOOR_INTERNAL   O2   -0.750882   57.093251   1.230451  C19  N2   C16
ICOOR_INTERNAL   C23 -178.336703   61.046447   1.518330  C19  N2   O2
ICOOR_INTERNAL   C25  -22.826129   66.916174   1.522564  C23  C19  N2
ICOOR_INTERNAL   C21   54.381415   72.125433   1.513394  C25  C23  C19
ICOOR_INTERNAL   C20  -65.929287   72.585866   1.518530  C21  C25  C23
ICOOR_INTERNAL   H14  168.321198   71.890144   1.070033  C20  C21  C25
ICOOR_INTERNAL   H15  121.939979   73.687057   1.069996  C20  C21  H14
ICOOR_INTERNAL   H16  119.682070   70.015091   1.070011  C21  C25  C20
ICOOR_INTERNAL   H17  119.410071   69.333109   1.070013  C21  C25  H16
ICOOR_INTERNAL   H23  119.750803   70.132011   1.070003  C25  C23  C21
ICOOR_INTERNAL   H24  119.538025   69.598481   1.069963  C25  C23  H23
ICOOR_INTERNAL   H19 -120.537286   71.438077   1.070005  C23  C19  C25
ICOOR_INTERNAL   H20 -121.238076   72.646354   1.069999  C23  C19  H19
ICOOR_INTERNAL   C14  179.366426   56.231249   1.466344  C16  C18  N2
ICOOR_INTERNAL   C44    0.087443   61.996128   1.445609  C14  C16  C18
ICOOR_INTERNAL   H25 -178.446203   59.324002   1.032000  C44  C14  C16
ICOOR_INTERNAL   H7   179.999779   59.004934   1.032021  C14  C16  C44
ICOOR_INTERNAL   H13  179.998713   59.051198   1.032029  C18  C22  C16
ICOOR_INTERNAL   H18  179.997900   59.310494   1.032007  C22  C7   C18
ICOOR_INTERNAL   C24  173.707419   60.978627   1.486888  N5   C8   C7
ICOOR_INTERNAL   C17   37.416194   65.870980   1.527253  C24  N5   C8
ICOOR_INTERNAL   C12  -47.932429   73.211144   1.502706  C17  C24  N5
ICOOR_INTERNAL   C10 -148.407600   51.860214   1.395453  C12  C17  C24
ICOOR_INTERNAL   N6  -179.979754   74.754024   1.343428  C10  C12  C17
ICOOR_INTERNAL   N1    -0.278628   69.048228   1.377285  N6   C10  C12
ICOOR_INTERNAL   C4  -175.554672   57.266493   1.413742  N1   N6   C10
ICOOR_INTERNAL   C3   -90.725691   59.397750   1.475217  C4   N1   N6
ICOOR_INTERNAL   C2  -179.340753   60.676707   1.470124  C3   C4   N1
ICOOR_INTERNAL   C1     0.236073   60.608076   1.467605  C2   C3   C4
ICOOR_INTERNAL   O4   178.772123   64.434523   1.399458  C1   C2   C3
ICOOR_INTERNAL   C15 -176.870679   55.844358   1.427541  O4   C1   C2
ICOOR_INTERNAL   H8   179.998785   70.530029   1.070023  C15  O4   C1
ICOOR_INTERNAL   H9  -119.996609   70.529409   1.070005  C15  O4   H8
ICOOR_INTERNAL   H10 -120.001345   70.528187   1.070054  C15  O4   H9
ICOOR_INTERNAL   C6  -179.437977   58.893416   1.473075  C1   C2   O4
ICOOR_INTERNAL   C5     0.527365   60.480609   1.459726  C6   C1   C2
ICOOR_INTERNAL   H5  -179.954255   59.848491   1.031988  C5   C6   C1
ICOOR_INTERNAL   H6   179.997958   59.760967   1.032020  C6   C1   C5
ICOOR_INTERNAL   H1   179.999202   59.695644   1.032059  C2   C3   C1
ICOOR_INTERNAL   H2  -179.999632   59.662272   1.032050  C3   C4   C2
ICOOR_INTERNAL   C13  176.060856   72.216496   1.373705  N1   N6   C4
ICOOR_INTERNAL   C11 -176.364992   51.621280   1.403966  C10  C12  N6
ICOOR_INTERNAL   O1   -23.545652   57.511727   1.233689  C11  C10  C12
ICOOR_INTERNAL   N3   178.564758   65.559612   1.349119  C11  C10  O1
```

```
ICOOR_INTERNAL    H3   179.996754   59.997725    0.984485   N3    C11   C10
ICOOR_INTERNAL    H4  -179.993947   60.001315    0.984475   N3    C11   H3
ICOOR_INTERNAL    H11  119.578039   69.861239    1.069967   C17   C24   C12
ICOOR_INTERNAL    H12  119.249436   68.969508    1.070017   C17   C24   H11
ICOOR_INTERNAL    H21 -120.690135   71.708371    1.069980   C24   N5    C17
ICOOR_INTERNAL    H22 -121.644804   73.257780    1.069992   C24   N5    H21
```

Contents of netcharge.charge

```
DESIRED_CHARGE -5
PENALTIES_CHARGE_RANGE -10 -1
PENALTIES 10 0 0 0 0 0 0 0 0 10
BEFORE_FUNCTION QUADRATIC
AFTER_FUNCTION QUADRATIC
```

Contents of no_met_thr_ser_asn.comp

```
PENALTY_DEFINITION
TYPE THR
DELTA_START 0
DELTA_END 1
PENALTIES 0 100
ABSOLUTE 8
BEFORE_FUNCTION CONSTANT
AFTER_FUNCTION QUADRATIC
END_PENALTY_DEFINITION

PENALTY_DEFINITION
TYPE SER
DELTA_START 0
DELTA_END 1
PENALTIES 0 100
ABSOLUTE 8
BEFORE_FUNCTION CONSTANT
AFTER_FUNCTION QUADRATIC
END_PENALTY_DEFINITION

PENALTY_DEFINITION
TYPE MET
DELTA_START 0
DELTA_END 1
PENALTIES 0 100
ABSOLUTE 3
BEFORE_FUNCTION CONSTANT
AFTER_FUNCTION QUADRATIC
END_PENALTY_DEFINITION

PENALTY_DEFINITION
```

TYPE ASN
DELTA_START 0
DELTA_END 1
PENALTIES 0 100
ABSOLUTE 12
BEFORE_FUNCTION CONSTANT
AFTER_FUNCTION QUADRATIC
END_PENALTY_DEFINITION

Command line for ab initio folding of designed sequences in Rosetta

```
~/rosetta_bin_linux_2018.33.60351_bundle/main/source/bin/AbinitioRelax.
static.linuxgccrelease -database
~/rosetta_bin_linux_2018.33.60351_bundle/main/database/ -in:file:frag3
aat000_03_05.200_v1_3 -in:file:frag9 aat000_09_05.200_v1_3 -
abinitio:relax -relax:fast -abinitio::increase_cycles 10 -
abinitio::rg_reweight 0.5 -abinitio::rsd_wt_helix 0.5 -
abinitio::rsd_wt_loop 0.5 -use_filters true -psipred_ss2
t000_.psipred_ss2 -kill_hairpins t000_.psipred_ss2 -out:file:silent
silent.out -nstruct 20000 -in:file:native able_design.pdb
```

Contents of residue file for ABLE flexible-backbone sequence design

start
1 X NATRO
13 A NATRO
48 A NATRO
111 A NATRO
1 A PIKAA ISQYARMDFEKWGPAVNLVH
2 A PIKAA KSQPAVNRHT
3 A PIKAA SQAVPNHT
4 A PIKAA ESQAVPNDHT
5 A PIKAA IGMWSAVYLAVHFT
6 A PIKAA SQAVPNHT
7 A PIKAA ESQAVPNDHT
8 A PIKAA IGMWSAVYLAVHFT
9 A PIKAA KSQPAVNRHT
10 A PIKAA SQAVPNHT
11 A PIKAA ISQYRMDFEKWAVLNVHT
12 A PIKAA IGMWSAVYLAVHFT
14 A PIKAA ESQAVPNDHT
15 A PIKAA IGMWSAVYLAVHFT
16 A PIKAA KSQPAVNRHT
17 A PIKAA SQAVPNHT
18 A PIKAA ISQYRMDFEKWAVLNVHT
19 A PIKAA IGMWSAVYLAVHFT
20 A PIKAA SQAVPNHT
21 A PIKAA ESQAVPNDHT

22 A PIKAA IGMWSAVYLAVHFT
23 A PIKAA KSQPAVNRHT
24 A PIKAA SQAVPNHT
25 A PIKAA SQAVPNHT
26 A PIKAA IGMWSAVYLAVHFT
27 A PIKAA SQAVPNHT
28 A PIKAA ESQAVPNDHT
29 A PIKAA KEGSRQPAVNDAT
30 A PIKAA EKGSRQPAVNDHT
31 A PIKAA EKRSQPAVNDHT
32 A PIKAA EKRSQPAVNDHT
33 A PIKAA EKRSQPAVNDHT
34 A PIKAA IGMWSAVYLAVHFT
35 A PIKAA ESQAVPNDHT
36 A PIKAA SQAVPNHT
37 A PIKAA IGWSVAVYLAMHFT
38 A PIKAA IGWSVAVYLAMHFT
39 A PIKAA SQAVPNHT
40 A PIKAA KSQPAVNRHT
41 A PIKAA IGWSVAVYLAMHFT
42 A PIKAA ESQAVPNDHT
43 A PIKAA SQAVPNHT
44 A PIKAA IGWSVAVYLAMHFT
45 A PIKAA IGWSVAVYLAMHFT
46 A PIKAA SQAVPNHT
47 A PIKAA KSQPAVNRHT
49 A PIKAA ESQAVPNDHT
50 A PIKAA ESQAVPNDHT
51 A PIKAA ISQYRMFKWAVLNVHT
52 A PIKAA IGWSVAVYLAMHFT
53 A PIKAA SQAVPNHT
54 A PIKAA KSQPAVNRHT
55 A PIKAA IGWSVAVYLAMHFT
56 A PIKAA IWSVQAVYLNMHFT
57 A PIKAA ESQAVPNDHT
58 A PIKAA EKRSQPAVNDHT
59 A PIKAA IGMWSAVYLAVHFT
60 A PIKAA EKRSQPAVNDHT
61 A PIKAA EKRSQPAVNDHT
62 A PIKAA EKRSQPAVNDHT
63 A PIKAA KEGSRQPAVNDAT
64 A PIKAA ISQYARMDFEKWGPAVNLVH
65 A PIKAA KSQPAVNRHT
66 A PIKAA ESQAVPNDHT
67 A PIKAA IGMWSAVYLAVHFT
68 A PIKAA SQAVPNHT

69 A PIKAA ESQAVPNDHT
70 A PIKAA IGMWSAVYLAVHFT
71 A PIKAA IGMWSAVYLAVHFT
72 A PIKAA KSQPAVNRHT
73 A PIKAA ESQAVPNDHT
74 A PIKAA IGMWSAVYLAVHFT
75 A PIKAA KSQPAVNRHT
76 A PIKAA SQAVPNHT
77 A PIKAA IGMWSAVYLAVHFT
78 A PIKAA IGMWSAVYLAVHFT
79 A PIKAA SQAVPNHT
80 A PIKAA ESQAVPNDHT
81 A PIKAA IGMWSAVYLAVHFT
82 A PIKAA KSQPAVNRHT
83 A PIKAA SQAVPNHT
84 A PIKAA ISQYRMDFEKWAVLNVHT
85 A PIKAA IGMWSAVYLAVHFT
86 A PIKAA SQAVPNHT
87 A PIKAA ESQAVPNDHT
88 A PIKAA IGMWSAVYLAVHFT
89 A PIKAA KSQPAVNRHT
90 A PIKAA SQAVPNHT
91 A PIKAA SQAVPNHT
92 A PIKAA IGMWSAVYLAVHFT
93 A PIKAA KSQPAVNRHT
94 A PIKAA ESQAVPNDHT
95 A PIKAA KEGSRQPAVNDAT
96 A PIKAA EKGSRQPAVNDHT
97 A PIKAA EKRSQPAVNDHT
98 A PIKAA EKRSQPAVNDHT
99 A PIKAA KSQPAVNRHT
100 A PIKAA IGWSVAVYLAMHFT
101 A PIKAA ESQAVPNDHT
102 A PIKAA SQAVPNHT
103 A PIKAA IGWSVAVYLAMHFT
104 A PIKAA IGWSVAVYLAMHFT
105 A PIKAA SQAVPNHT
106 A PIKAA KSQPAVNRHT
107 A PIKAA IGWSVAVYLAMHFT
108 A PIKAA ESQAVPNDHT
109 A PIKAA KSQPAVNRHT
110 A PIKAA IGWSVAVYLAMHFT
112 A PIKAA SQAVPNHT
113 A PIKAA KSQPAVNRHT
114 A PIKAA IGWSVAVYLAMHFT
115 A PIKAA ESQAVPNDHT

116 A PIKAA KSQPAVNRHT
117 A PIKAA IGWSVAVYLAMHFT
118 A PIKAA IGWSVAVYLAMHFT
119 A PIKAA ESQAVPNDHT
120 A PIKAA SQAVPNHT
121 A PIKAA IGWSVAVYLAMHFT
122 A PIKAA EKRSQPAVNDHT
123 A PIKAA EKRSQPAVNDHT
124 A PIKAA EKRSQPAVNDHT
125 A PIKAA ISQYARMDFEKWGPAVNLVH

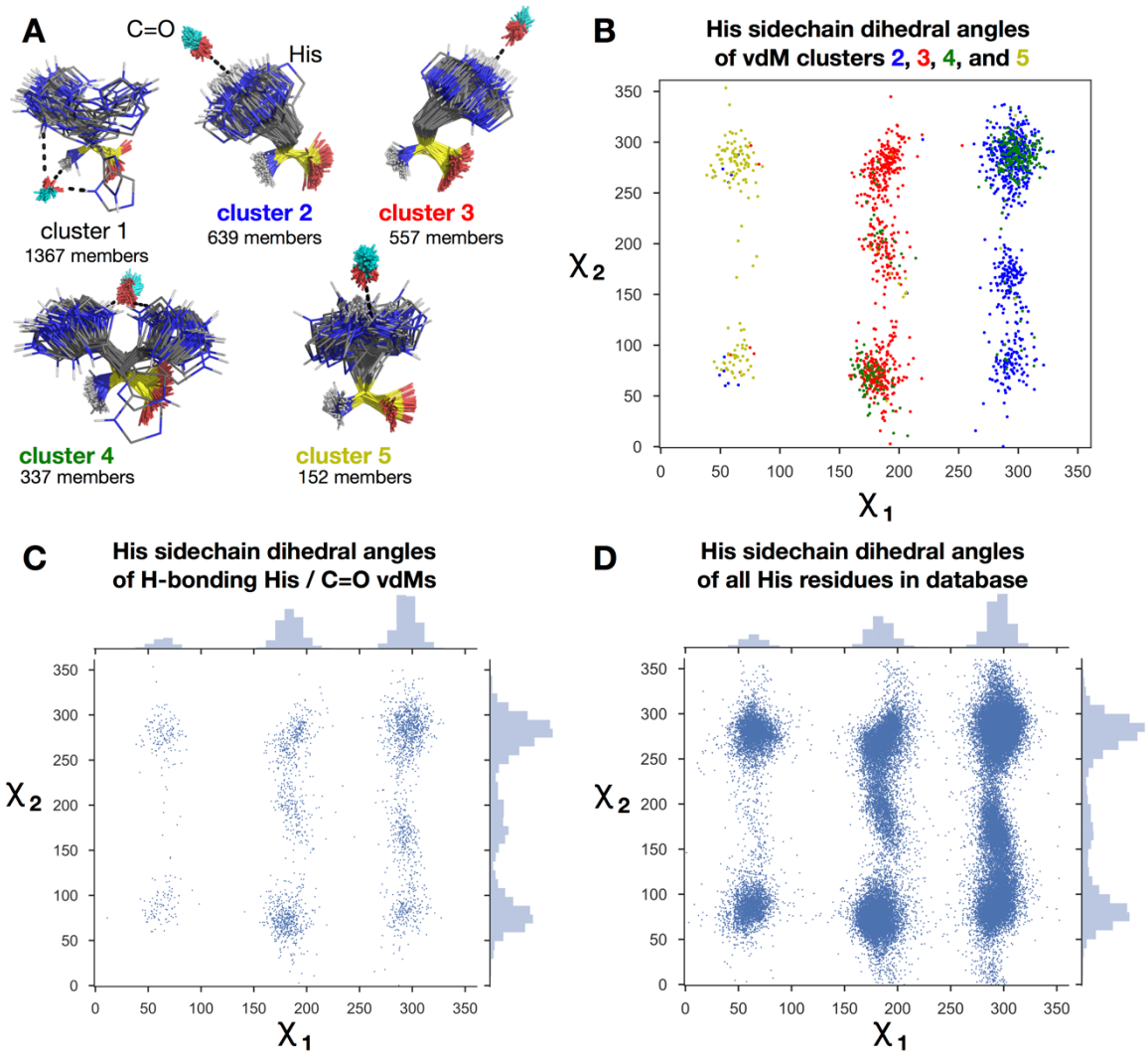



**Fig. S1. van der Mers (vdMs) have similar sidechain dihedral-angle distributions as traditional rotamers.**

**A**) The five most prevalent His/C=O vdMs, after superimposing and clustering by N, Ca, C atoms of His mainchain and C=O atoms of the chemical group. His mainchain is yellow, sidechain is gray, and C=O chemical group is cyan. The top vdM (cluster 1) makes predominantly mainchain H-bonds to C=O, with some multivalent His sidechain H-bonds. Clusters 2, 3, 4, and 5 H-bond to C=O predominantly via His sidechain. The dihedral angles of the His sidechain for clusters 2 through 5 are plotted in **B**, and are color-coded according to the cluster label in **A**. **C**) Sidechain dihedral angles of each His/C=O vdM with His sidechain H-bonding to C=O. A total of 2,692 His/C=O vdMs are plotted. **D**) Sidechain dihedral angles of each His in the COMBS protein database. A total of 51,033 His rotamers are plotted. The $X_1$ and $X_2$ distributions are similar between **C** and **D**.

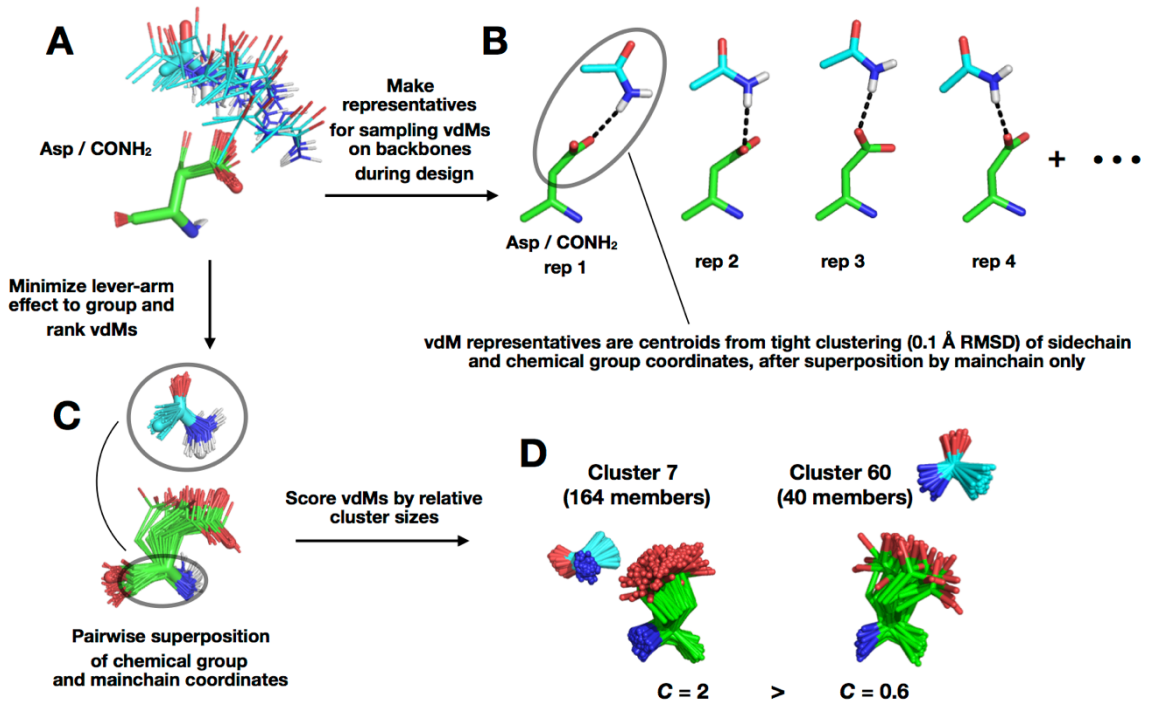



**Fig. S2. Using vdMs for sampling and scoring.**

**A, B**) Representative vdMs are made for sampling onto backbone coordinates by aligning all vdMs by mainchain atoms and then tightly clustering (0.1 Å heavy atom RMSD) by coordinates of the sidechain and chemical group. **C, D**) vdMs are scored by grouping them based on a different criterion than what is used for sampling. We score vdMs by pairwise superimposing them via mainchain and chemical group coordinates (sidechain coordinates are not considered) and clustering by heavy atom RMSD (0.5 Å) of these coordinates. This reduces lever-arm effects that occur when small movements of backbone sweep out a large solid angle causing large movement of a distant chemical group. For example, (**A**) and (**C**) show the same vdMs but superimposed only by mainchain in (**A**) while superimposed by both mainchain and chemical group in (**C**). **D**) The cluster score *C* quantifies the prevalence of a vdM in the PDB. For example, cluster 7 with 164 members has a higher *C* score than cluster 60 with 40 members. We use representatives in (**A**, **B**) for sampling vdMs on a protein backbone but keep track of their *C* scores from the alternative clustering performed in (**C, D**) for ranking binding sites later in the design process.

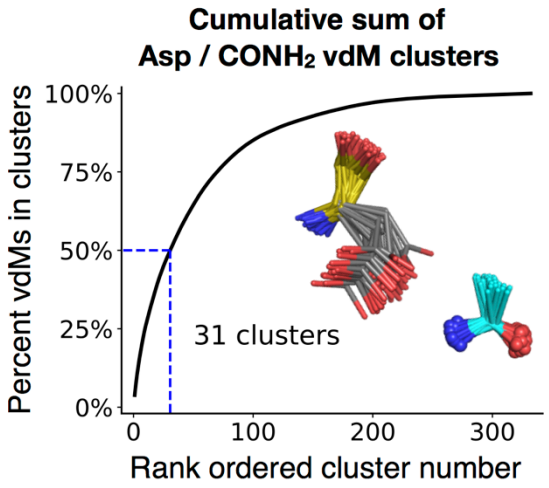



**Fig. S3. The conformational space of observed interaction geometries is largely captured by a small number of vdM clusters.**

H-bonding vdMs of Asp with carboxamide ($CONH_2$) were clustered by root mean squared deviation (R.M.S.D. less than 0.5 Å) of coordinates of backbone (N, Ca, C) and chemical-group heavy atoms, after all-by-all pairwise superposition of vdMs by these same coordinates. The plot shows the number of clusters needed to account for the total percent of observed H-bonded Asp / $CONH_2$ vdMs. Blue line indicates the point at which half of the vdMs are found in clusters. The inset shows the 7$^{th}$ largest vdM cluster, which is mainly comprised of alpha-helical residues.

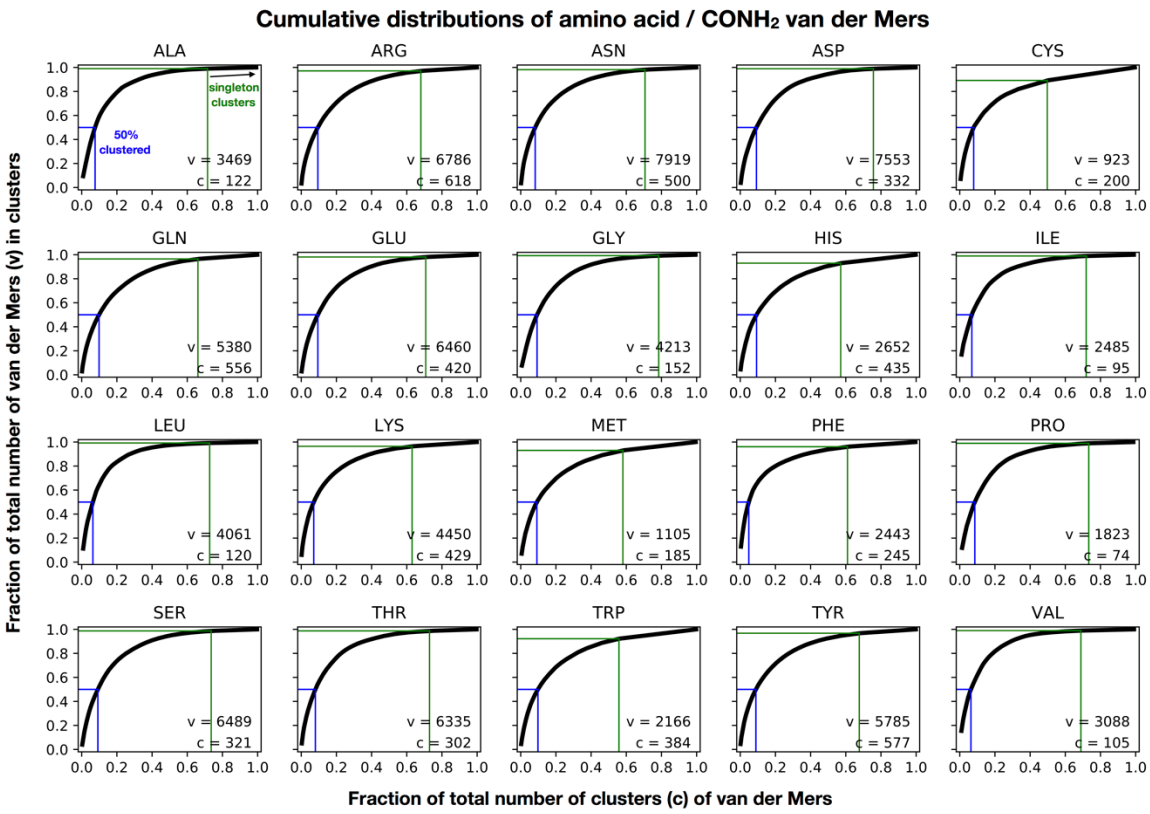



**Fig. S4. The conformational space of observed interaction geometries is largely captured by a small number of vdM clusters.**

H-bonding vdMs of $CONH_2$ for each amino acid type were clustered by root mean squared deviation (R.M.S.D. less than 0.5 Å) of coordinates of backbone (N, Ca, C) and chemical-group heavy atoms, after all-by-all pairwise superposition of vdMs by these same coordinates. The plots show the number of clusters needed to account for the total percent of observed H-bonded amino acid / carboxamide ($CONH_2$) vdMs. Blue lines indicate the point at which half of the vdMs are found in clusters. Green lines indicate the point at which the clusters to the right are singletons. Insets list the total number of H-bonded vdM members (v) and the total number of clusters (c) for that amino acid / $CONH_2$ vdM type. Hydrophobic amino acids H-bond with $CONH_2$ groups primarily through main-chain interactions, but aromatic H-bonds (e.g. N-H of $CONH_2$ with phenyl of Phe) are also included.

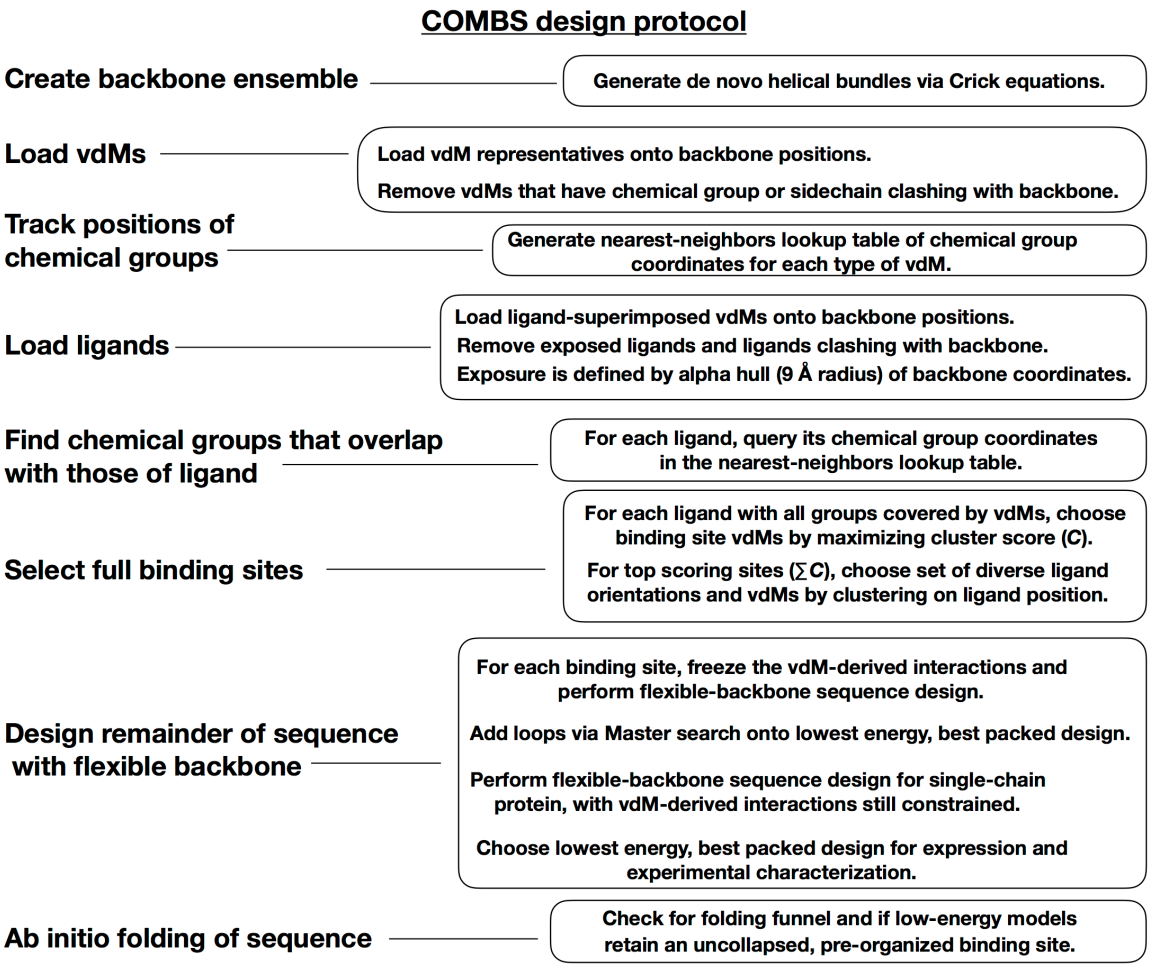



**Fig. S5. Convergent Motifs for Binding Sites (COMBS) design protocol.**

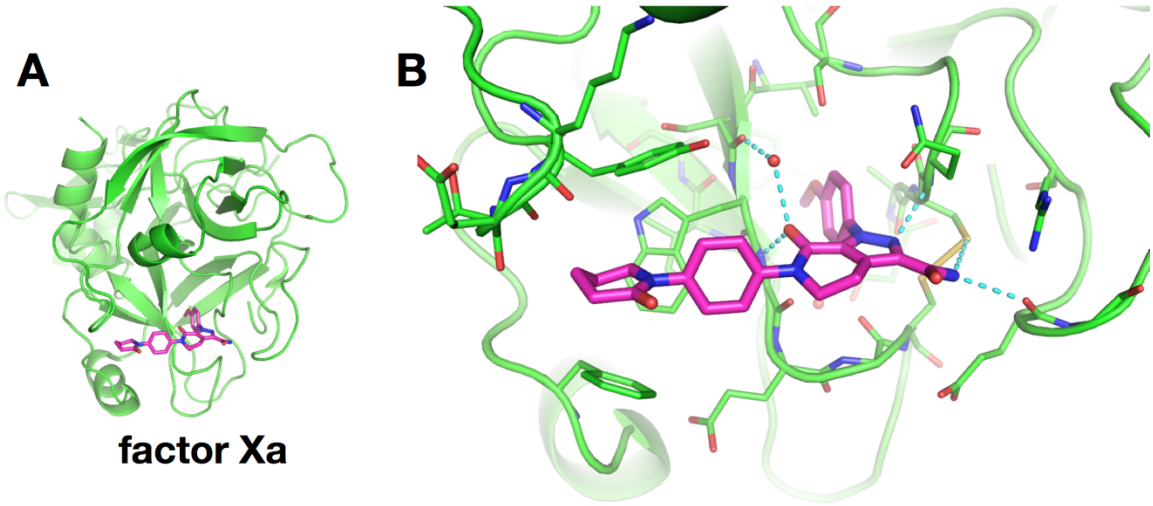



**Fig. S6. Apixaban-bound factor Xa structure.**

**A)** Overview of the factor Xa structure (green), with apixaban (purple) bound in the active site. **B**) View of the inhibitor binding site. Apixaban makes H-bonds (cyan dashes) with backbone amides in loops, as well as a water-mediated H-bond to a backbone amide.

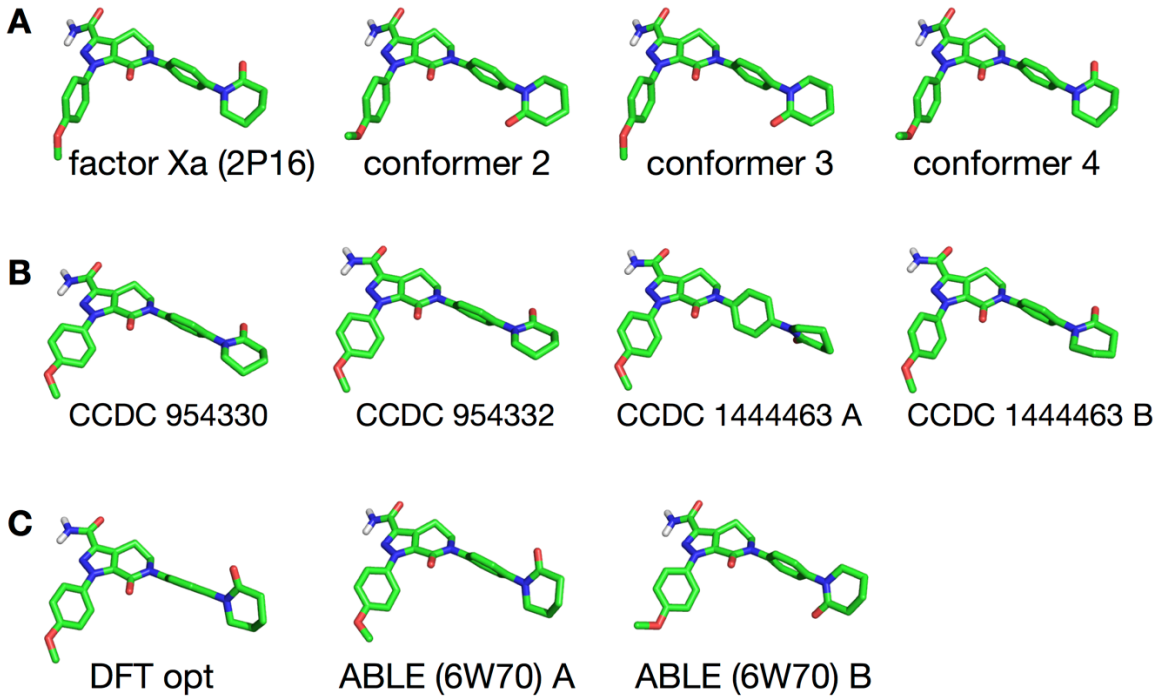



**Fig. S7. Apixaban conformers.**

**A**) Conformers searched with COMBS for design of Apixaban-binding proteins. Binding sites of conformers 2, 3, and 4 scored lower (less favorably) than binding sites using the apixaban conformer from the factor Xa complex, so designs of these conformers were not tested experimentally. **B**) Small-molecule crystal structures of apixaban from the Cambridge Structural Database show conformers similar to that in factor Xa and ABLE. **C**) DFT (B3LYP/6-31G*)-optimized geometry of apixaban and two apixaban conformers found in ABLE crystal structure.

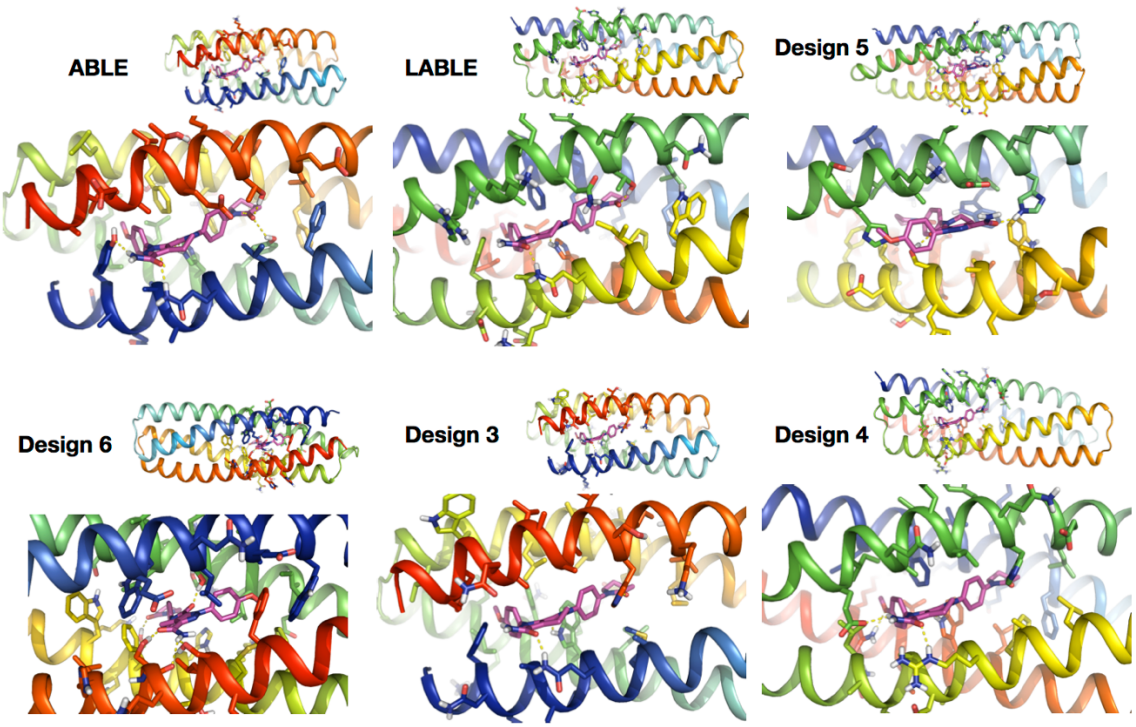



**Fig. S8. Computational models of designed proteins.**

Overview and close-up of binding sites of the 6 proteins designed via the COMBS strategy. The N terminus is colored blue, and the C-terminus is red. Apixaban is shown in purple. All sidechains within an 8 Å radius of apixaban are shown. Note the different topologies, bundle lengths, binding modes and binding residues. ABLE and LABLE share most of the same vdM-derived binding residues and the same binding position of apixaban, but differ in topology, size, and overall sequence (22% sequence homology).

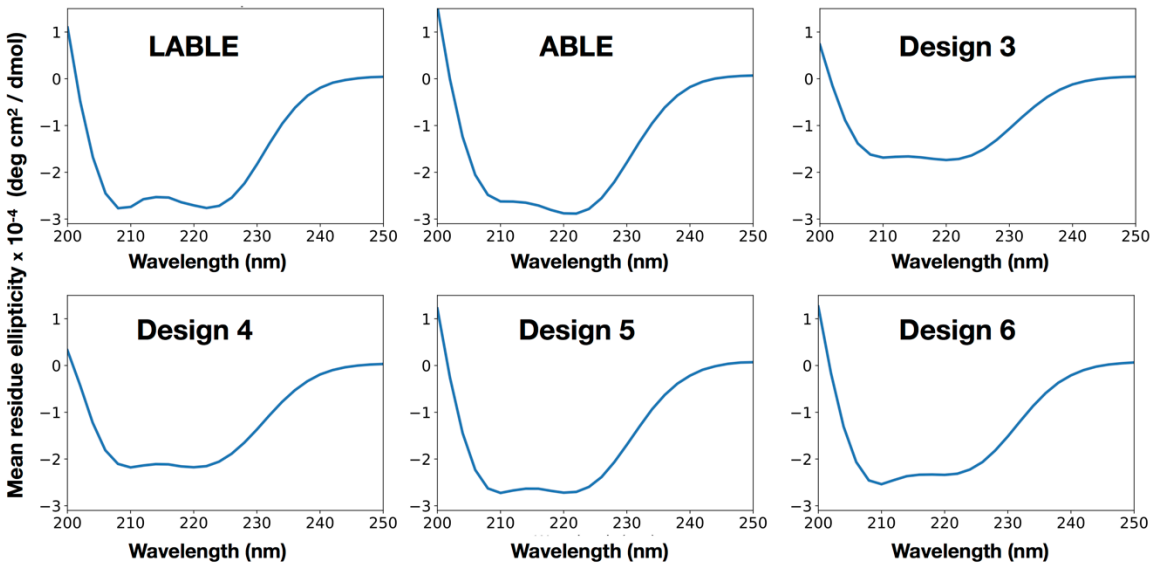



**Fig. S9. Circular dichroism spectra of the designed proteins show that all are helical.**

Spectra were collected at room temperature in 50 mM NaPi, 100 mM NaCl, pH 7.4 buffer.

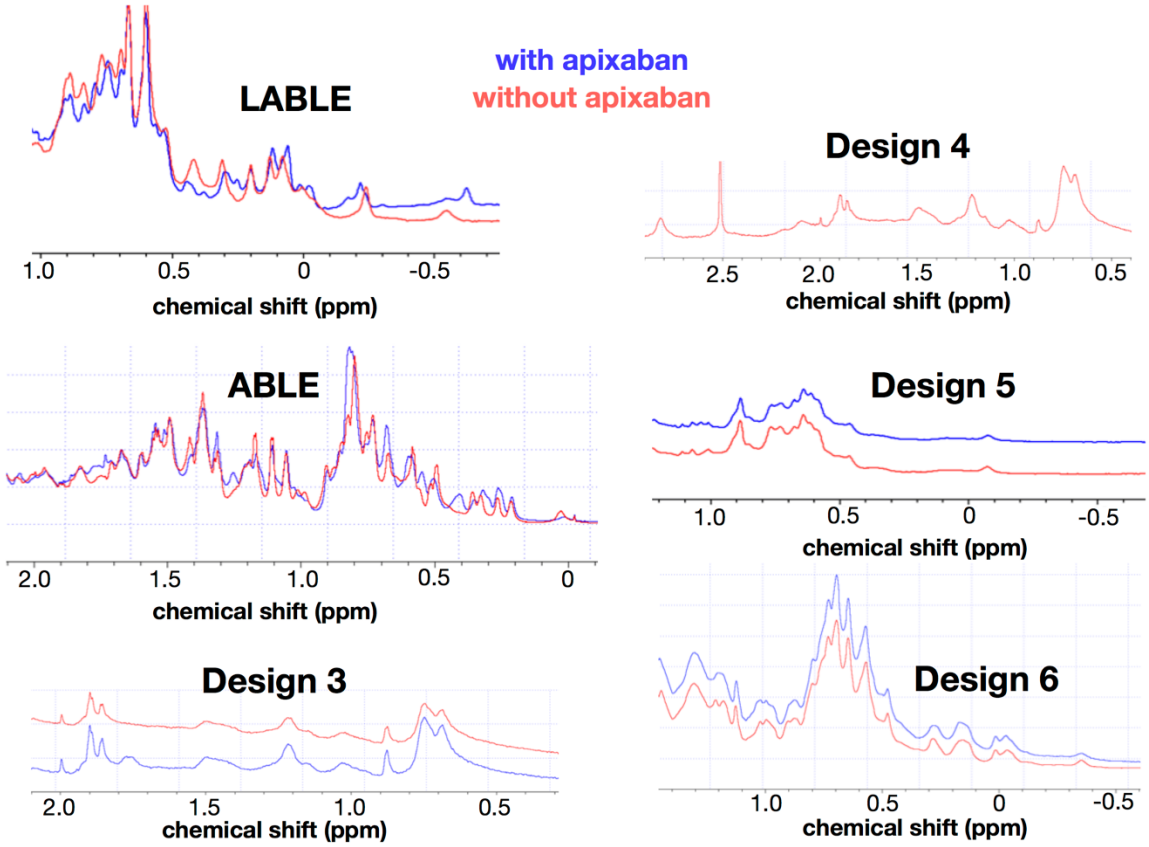



**Fig. S10. Nuclear magnetic resonance spectroscopy shows ABLE and LABLE bind apixaban.**

1-dimensional $^{1}$H spectra of LABLE, ABLE, and other designs (150 μM) with (blue) and without (red) 1 equivalent of apixaban. The 1D-spectra of LABLE and ABLE are well-dispersed and show clear differences in chemical shift upon addition of apixaban. The remainder of the designs showed no change of chemical shift upon addition of apixaban or did not display well-dispersed chemical shifts. Design 4 showed broad peaks in the methyl region, indicative of a molten globule, and was not tested for binding. 1-dimensional $^{1}$H spectra were recorded on Bruker 800 MHz spectrometer. Buffer for all experiments was 50 mM NaPi, 100 mM NaCl, pH 7.4 with 5% $d_6$-dimethylsulfoxide.

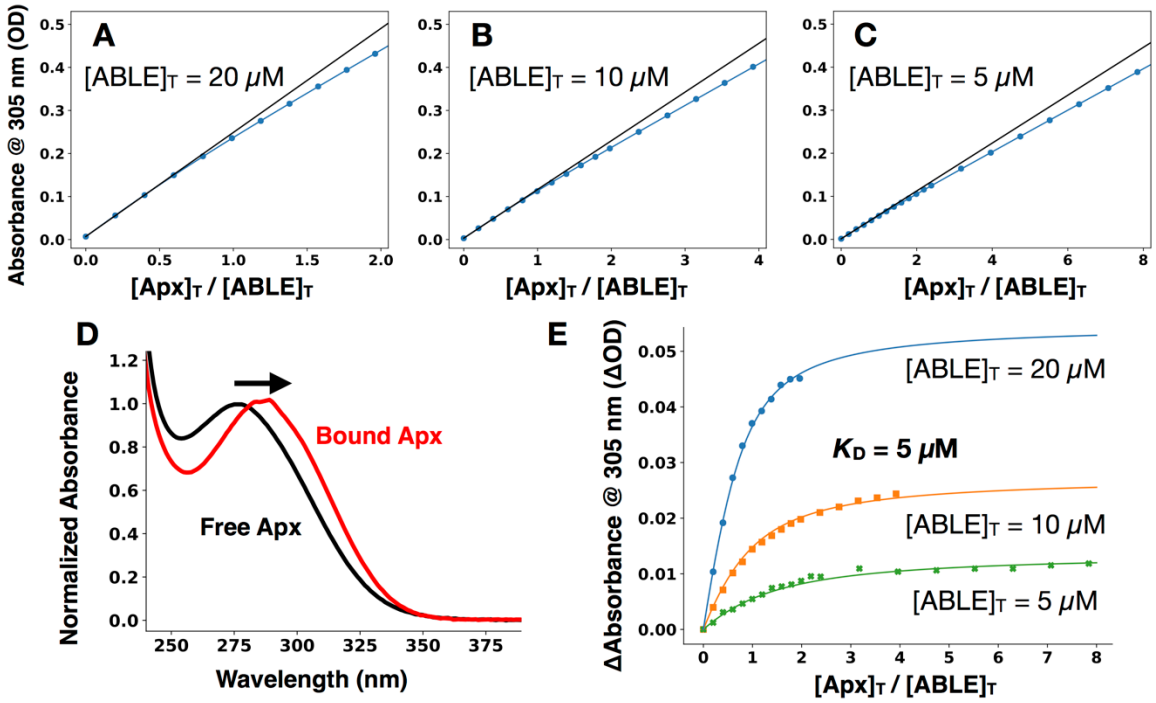



**Fig. S11. Spectral titration of apixaban into a solution of ABLE shows low-μM binding.**

**A-C)** Absorbance at 305 nm was monitored for several concentrations of ABLE ($[ABLE]_T$ = 20, 10, and 5 uM) as a function of total apixaban concentration ($[Apx]_T$). The black solid line is an extrapolation of a linear fit to the first 3 datapoints for low $[Apx]_T$. Deviation from the line is evidence of binding (approaching saturation). The data in A-C (blue circles) were globally fit to a single-site binding model (see supplemental text), and the results of the fit are shown with blue lines (A-C). **D**) The electronic absorbance spectrum of apixaban is red-shifted upon binding to ABLE. The black spectrum shows apixaban (4 μM) in 50mM NaPi, 100mM NaCl, pH 7.4 buffer. The red spectrum is the difference of the absorbance spectrum of ABLE alone (20 μM) and the spectrum of ABLE (20 μM) with apixaban (4 μM). The spectra were normalized to the peak maximum for comparison. **E**) Global fit of a single-site binding model to the absorbance changes at 305 nm upon titration of apixaban into 5, 10, and 20 μM solutions of ABLE. The data points shown are transformations of the raw data in A-C. The first two terms in Equation 1 were subtracted from the raw data to isolate the contribution from the bound complex. Solid lines show the result of the global fit. The parameters from the fit (see Equation 1) were Δε (305 nm) = 3900 (± 420) $M^{-1}$ $cm^{-1}$, $K_D$ = 5 (± 1) μM, N = 1.4 (± 0.07), and ε (305 nm) = 9570 (± 28) $M^{-1}$ $cm^{-1}$ [20 μM ABLE], 9530 (± 53) $M^{-1}$ $cm^{-1}$ [10 μM ABLE], 9680 (± 93) $M^{-1}$ $cm^{-1}$ [5 μM ABLE]. Errors are the standard deviation of the fitted parameters. The experiment was repeated three times with new samples.

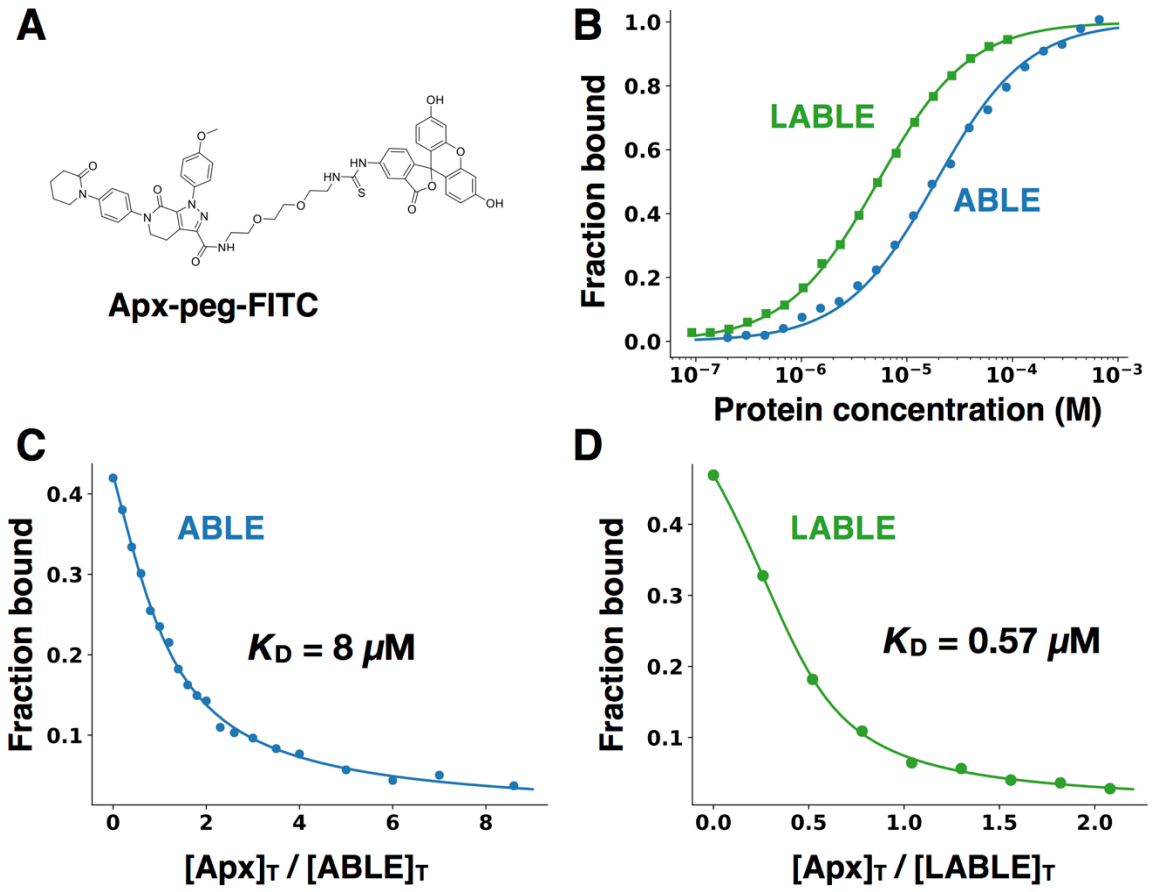



**Fig. S12. Fluorescence anisotropy binding experiments of ABLE and LABLE with apixaban.**

**A**) Apx-peg-FITC was used for fluorescence anisotropy experiments to assess binding of apixaban to ABLE and LABLE. **B**) Anisotropy data was converted to fraction bound (circles) of Apx-peg-FITC to ABLE or LABLE and fit to a single-site binding model (solid lines). $K_D$ of LABLE to Apx-peg-FITC was found to be 5.3 (± 0.2) μM and that of ABLE to Apx-peg-FITC was 18.8 (± 1.3) μM. **C** and **D**) Competition experiments with apixaban show that ABLE and LABLE bind apixaban with $K_D$ of 8.1 (± 1.2) μM and 0.57 (± 0.07) μM, respectively. Errors are the standard deviation of the fitted parameters. The experiments were repeated at least twice.

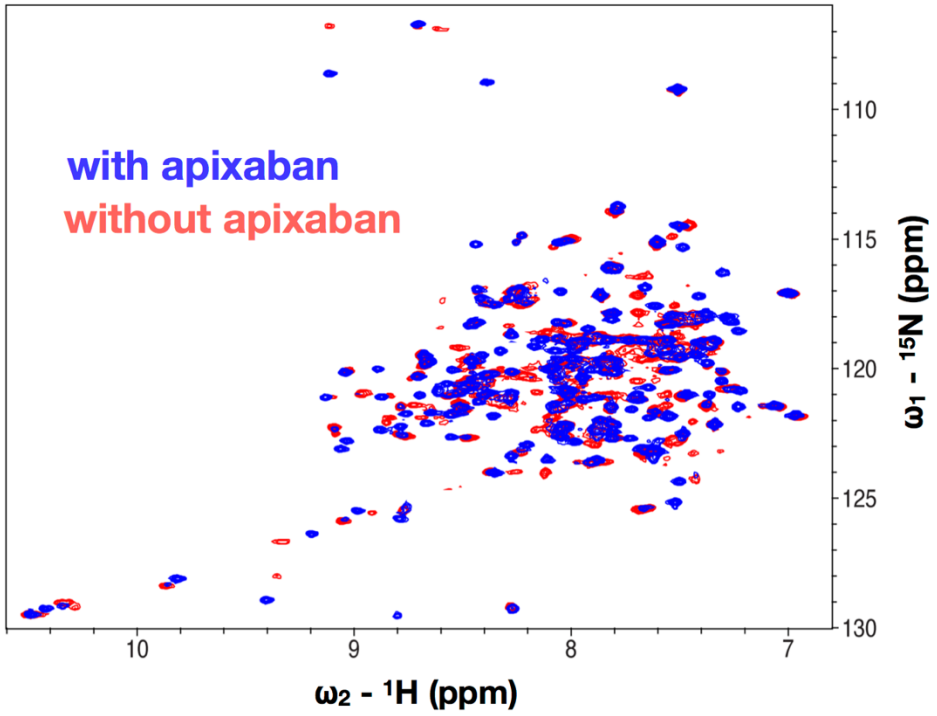



**Fig. S13. $^{1}$H-$^{15}$N HSQC spectrum of 400 µM LABLE with (blue) and without (red) 1 equivalent of apixaban.**

The 2D-spectrum of LABLE is well-dispersed and show clear differences in chemical shift upon additional of apixaban. 2-dimensional $^{1}$H-$^{15}$N HSQC spectra were recorded on Bruker 800 MHz spectrometer. Buffer was 50 mM NaPi, 100 mM NaCl, pH 7.4 with 5% $d_6$-dimethylsulfoxide.

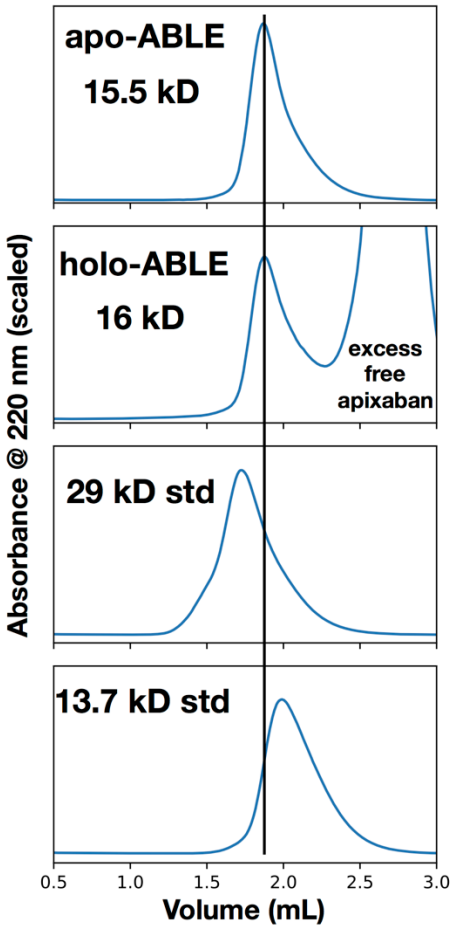



**Fig. S14. Analytical gel filtration analysis of apo- and holo-ABLE shows a monomeric protein.**

Samples were run on a Superdex 75 5/150 column concentrations of 140 µM and 75 µM for apo and holo, respectively, in 50 mM NaPi, 100 mM NaCl, pH 7.4 buffer. The peak near 2.7 ml elution volume in holo-ABLE is attributed to both DMSO and unbound (excess) apixaban, which was added in excess (200 uM). The 6xHis-tag and TEV cleavage linker on the N-terminus add to the MW of the protein.

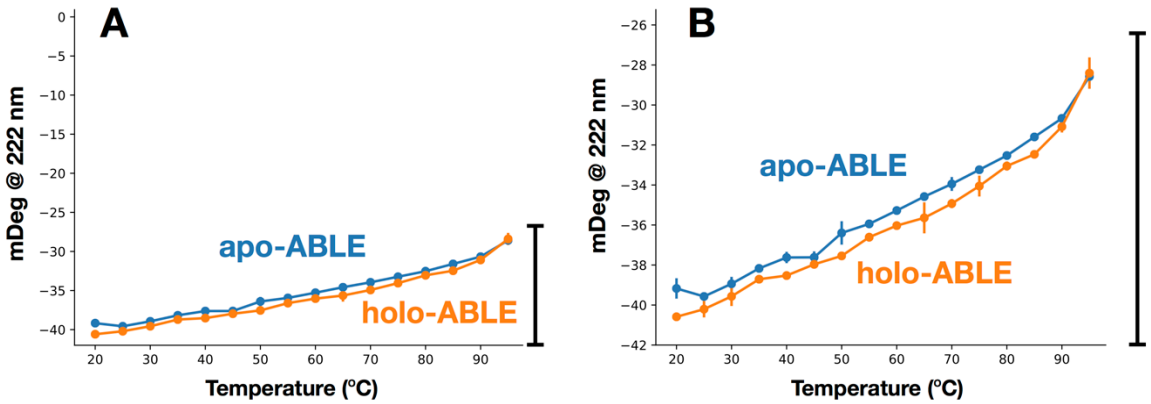



**Fig. S15. Apo- and holo-ABLE are thermostable.**

Circular dichroism signal at 222 nm for various temperatures shows that apo- and holo-ABLE have melting temperatures > 95 ºC. Markers show the average CD signal with error bars denoting the standard deviation of two sequential experiments. **B**) is the same as (**A**) but scaled to a smaller region of the plot.

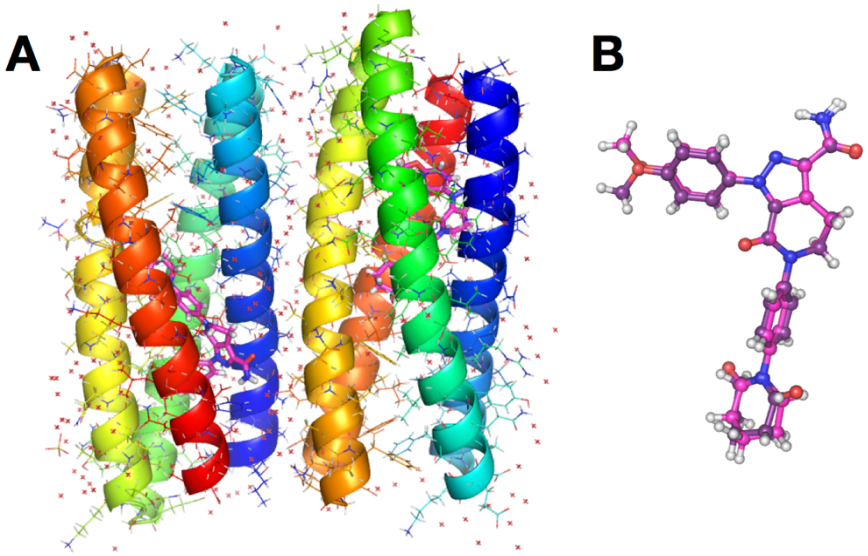



**Fig. S16. Crystallographic asymmetric unit of apixaban-bound ABLE.**
**A)** Two highly similar monomers were found in the asymmetric unit (0.64 Å Cα RMSD). Apixaban is colored magenta in both monomers, and protein chains are colored from blue (N-terminus) to red (C-terminus). **B)** Superimposed conformations of apixaban (magenta and dark purple) in the two subunits largely agree, differing by rotation around the methoxy-phenyl bond and the terminal oxopiperidine.

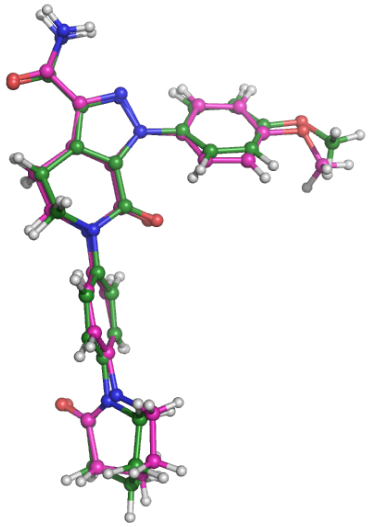


**Fig. S17. DFT-optimized structure of apixaban (green) and structure from drug-bound ABLE (purple).**

The conformation of apixaban relaxes slightly in drug-bound ABLE relative to its initial starting geometry, which was taken from the co-crystal structure of factor Xa (PDB 2p16).

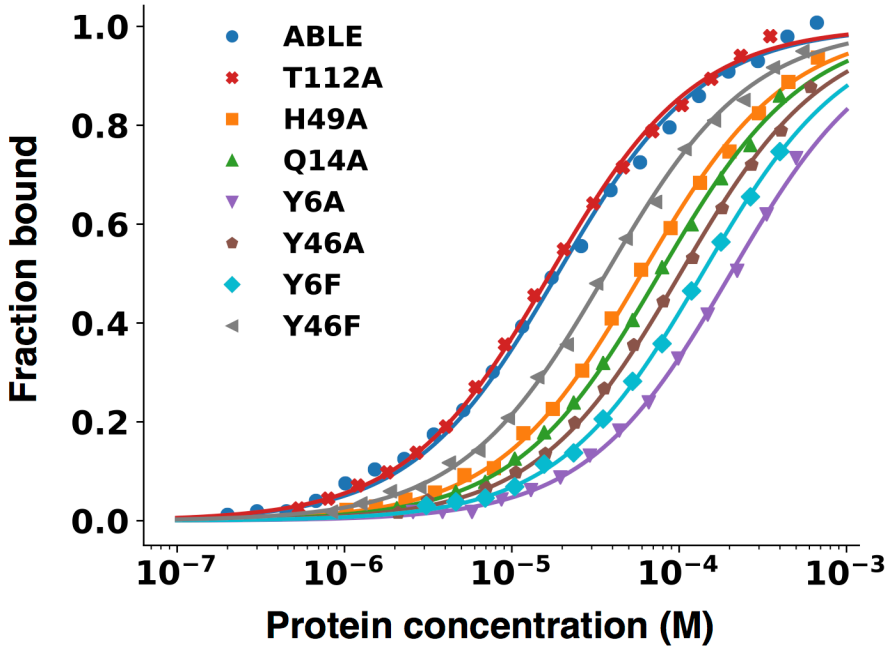



**Fig. S18. Anisotropy data of single-site mutants of ABLE.**

Anisotropy data was converted to fraction bound (markers) of Apx-peg-FITC to ABLE or LABLE and fit to a single-site binding model (solid lines). Dissociation constants ($K_D$) to Apx-peg-FITC are: 17.0 (± 0.6) µM for T112A, 36.5 (± 1.3) µM for Y46F, 60 (± 3) µM for H49A, 76 (± 3) µM for Q14A, 102 (± 4) µM for Y46A, 137 (± 6) µM for Y6F, 230 (± 10) µM for Y6A. Errors are the standard deviation of the fitted parameters.

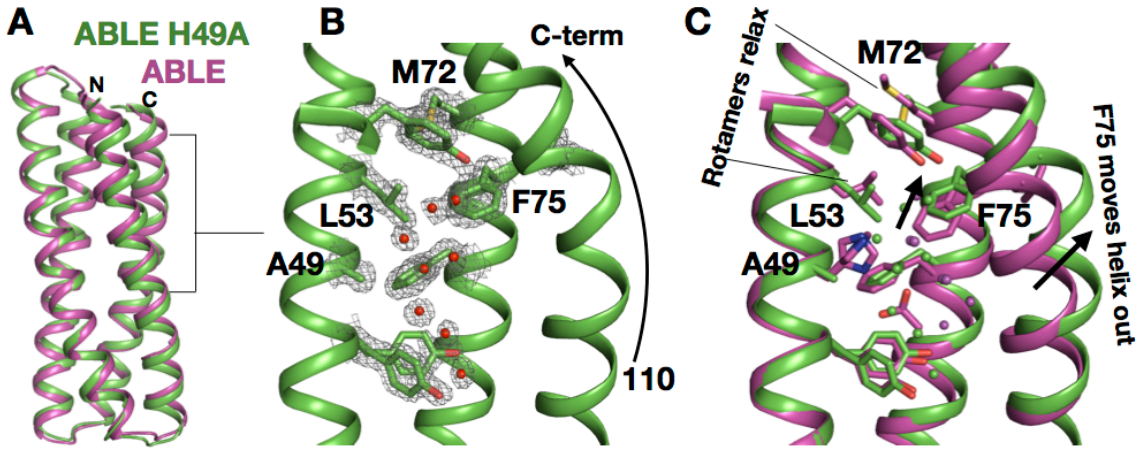



**Fig. S19. Structure of H49A mutant of ABLE.**

**A**) The 1.6 Å resolution structure of the H49A mutant of ABLE (green) is superimposed on that of apo-ABLE (purple) via Cα coordinates of residues 1-110 (0.9 Å Cα RMSD). **B**) The drug-free binding site of the H49A mutant shows a water-filled pocket with several relaxed rotamers, e.g. L53 and M72. **C**) Comparison of H49A mutant with apo-ABLE (purple) shows that the absence of H49A allows L53 to adopt its preferred rotamer, which in turn allows methionine M72 to adopts its preferred rotamer. Phenylalanine F75 allows for space of the L53 rotamer by abutting the C-terminal helix, which kinks near residue 113 to avoid sterically clashing with F75. Waters from unliganded ABLE are shown as purple spheres, and waters from H49A mutant are shown as green spheres.

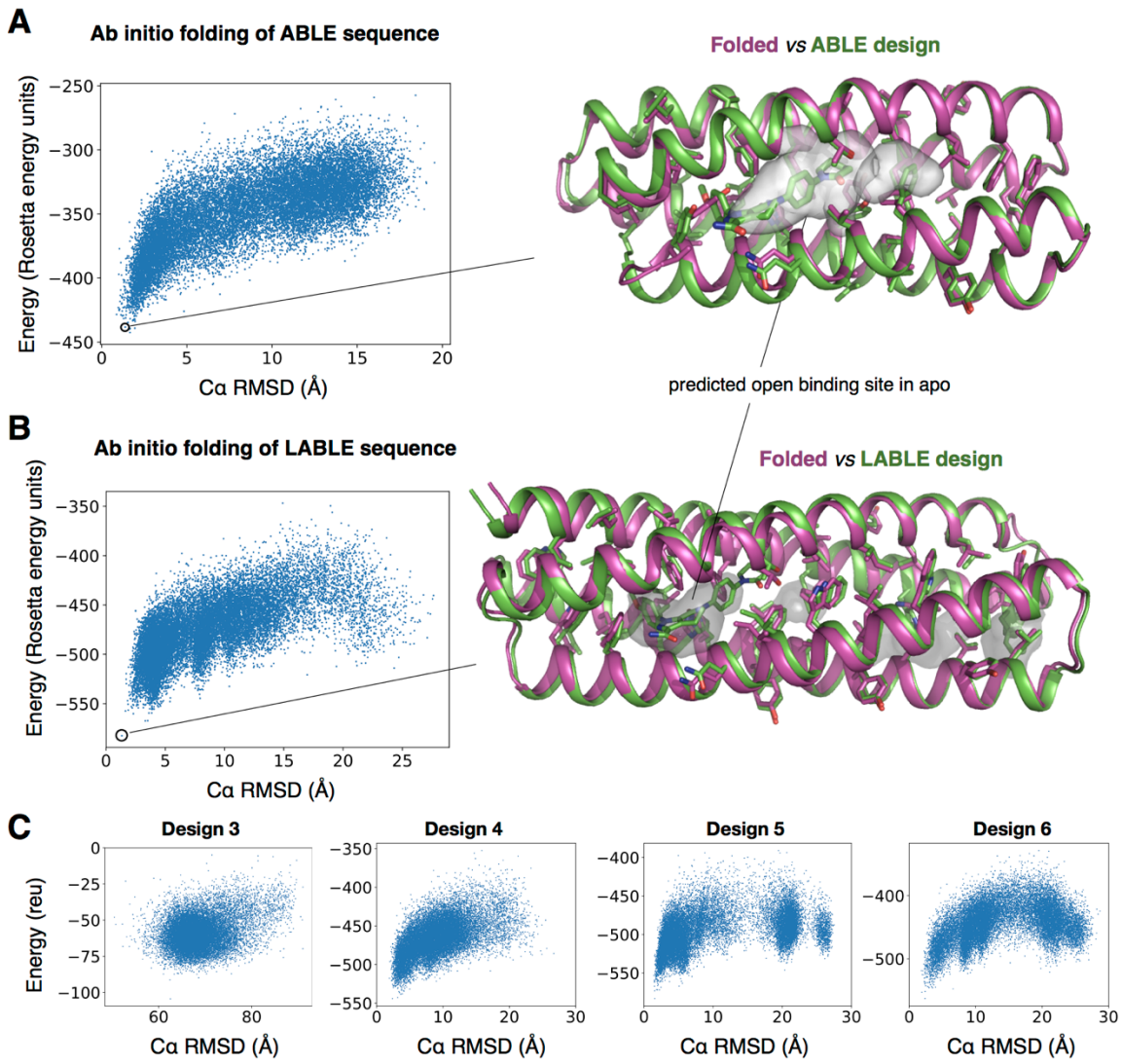



**Fig. S20.** ***Ab initio*** **folding of designed sequences.**

**A** and **B**) The ABLE and LABLE design models (green) are accurately predicted by the lowest energy, lowest RMSD (< 2 Å) *ab initio* models (purple). **C**) Other designs were predicted either not to fold or showed collapsed binding sites with higher RMSD (> 2 Å).

**Table S1. Crick parameters* for generation of parametric 4-helix bundle ensemble**

| Crick parameters | Model 1 | Model 32 |
|---|---|---|
| $R_0$ (Å) | 7.902 | 8.202 |
| $R_1$ (Å) | 2.250 | 2.250 |
| $\omega_0$ (°/res) | -3.114 | -3.114 |
| $\omega_1$ (°/res) | 102.325 | 102.325 |
| α (°) | -16.626 | -17.277 |
| $\varphi_1$ for chain A (°) | -60.432 | -60.432 |
| $\varphi_1$ for chain B (°) | -50.045 | -50.045 |
| $\varphi_1$ for chain C (°) | -70.284 | -70.284 |
| $\varphi_1$ for chain D (°) | -52.207 | -52.207 |
| $\Delta\varphi_0$ for chain B (°) | -76.093 | -76.093 |
| $\Delta\varphi_0$ for chain C (°) | 164.337 | 178.717 |
| $\Delta\varphi_0$ for chain D (°) | 88 | 102.380 |
| $\Delta Z_{aa'}$ for chain B (Å) | 4.045 | 4.045 |
| $\Delta Z_{aa'}$ for chain C (Å) | 0.673 | 0.673 |
| $\Delta Z_{aa'}$ for chain D (Å) | 4.027 | 4.027 |

*Bundles were generated using the CCCP server (*10*): http://arteni.cs.dartmouth.edu/cccp/index.fit.php

Shaded parameters ($R_0$ and $\Delta\varphi_0$) were interpolated to create the ensemble. The $\Delta\varphi_0$ values were covaried for chains C and D to preserve the designable packing geometry between these chains.

**Table S2. PDB accession codes and Cα RMSD values of matches to a 4-helix query (10-residues each helix) of the initial parameterized backbone of ABLE.**

1ezv (0.90397), 3cx5 (0.92492), 2qjy (0.93155), 5o0t (0.93779), 2a06 (1.0068), 2fyu (1.041), 1ntm (1.0703), 4fm3 (1.1008), 4ogq (1.1697), 4o6y (1.1811), 5cx2 (1.2003), 5vjt (1.2629), 2j8w (1.2804), 1kqf (1.2915), 2p0n (1.3116), 2ykz (1.3794), 2yfa (1.3933), 4cpc (1.3949), 6f6p (1.4037), 4w7z (1.4043), 4jea (1.4056), 5xup (1.4095), 5fnp (1.4136), 4dri (1.4181), 2a3q (1.4277), 1cpq (1.4454), 5j0k (1.4582), 2qzc (1.4589), 2y43 (1.4665), 5j10 (1.4715), 5ly9 (1.4739), 3d19 (1.4764), 5iy5 (1.4911), 2v66 (1.4934), 1vyk (1.4947), 2y69 (1.4991), 5b1a (1.5008), 6g94 (1.5019), 1m56 (1.5096), 5xub (1.5099), 4g7v (1.5158), 6f63 (1.5159), 1hul (1.5253), 3dby (1.5313), 4ecg (1.5331), 5tw9 (1.538), 3l34 (1.5385), 2yev (1.5414), 3hr0 (1.548), 5we0 (1.5555), 4akk (1.5644), 1gqa (1.5685), 3t9o (1.5709), 1r4a (1.5726), 5nl1 (1.5748), 2q5z (1.5753), 1zke (1.5788), 5gpg (1.5797), 3fx7 (1.5826), 2q73 (1.5873), 4nsw (1.5947), 3ls0 (1.5989), 2mhr (1.6018), 3rfy (1.6018), 3cax (1.6022), 2hmz (1.6065), 4xwj (1.6067), 2nrj (1.6077), 5xa5 (1.6138), 4fmt (1.6229), 4psm (1.6235), 4qpk (1.6253), 5ic0 (1.626), 2j9w (1.6262), 3e7k (1.6274), 4e17 (1.6292), 1f1m (1.6301), 5hyl (1.6314), 5oc0 (1.6357), 4w8p (1.6374), 1w9c (1.6409), 5ayn (1.6422), 5ipx (1.6478), 4geh (1.648), 3crk (1.6514), 3w8i (1.6521), 3cve (1.6533), 5a61 (1.6558), 5olm (1.6583), 4ulv (1.6609), 5lcy (1.6616), 4kyo (1.6701), 3k6c (1.6728), 1upt (1.6742), 4k7b (1.6746), 4euk (1.6747), 3zsu (1.6757), 4q20 (1.6809), 5n5f (1.6815), 2pnr (1.6841), 4f7g (1.6842), 4bix (1.6851), 4cni (1.6875), 3t6g (1.6882), 5w93 (1.6887), 2vpz (1.6902), 4nn5 (1.6909), 3w8h (1.6916), 1ydi (1.693), 3e0s (1.6945), 4e18 (1.6964), 5j13 (1.6996), 2p06 (1.7005), 2ch7 (1.7013), 4dkc (1.7026), 3nce (1.7034), 2gs4 (1.7039), 1dow (1.7045), 1w2y (1.7077), 3dkq (1.7079), 3dyj (1.7107), 3cra (1.7116), 3s90 (1.7121), 4qgp (1.7164), 2lig (1.7176), 4jnu (1.7196), 5da5 (1.722), 3p8c (1.7244), 4k0d (1.7257), 1uz3 (1.7275), 4rbr (1.7316), 1yjg (1.7337), 3efz (1.7367), 4e01 (1.7399), 6db1 (1.7409), 1cnt (1.741), 5j71 (1.7423), 3fyq (1.7425), 1nfn (1.7517), 3waq (1.7527), 3oyv (1.7536), 4afl (1.7553), 4mhl (1.7568), 4z9h (1.7578), 4p9t (1.7588), 4djg (1.7589), 2yxh (1.7589), 1bz4 (1.7608), 4n6h (1.761), 2j0n (1.7614), 4tx5 (1.7635), 2gsc (1.7685), 4jkv (1.7706), 1xzp (1.7784), 1jmw (1.7789), 1gs9 (1.7797), 3oq3 (1.7801), 3nzz (1.7805), 3h6p (1.7812), 5n5e (1.7824), 4ryo (1.7829), 2d4u (1.7836), 1lki (1.7859), 2bs2 (1.7864), 5fmn (1.7883), 3piv (1.7894), 5xfl (1.7895), 6exs (1.7896), 4e40 (1.7919), 4fa8 (1.7935), 2wdq (1.794), 3r84 (1.7965), 3zyl (1.7971), 2y6x (1.7979), 5cs0 (1.8002), 4cih (1.8017), 5uwc (1.8018), 3myf (1.803), 4nb5 (1.8031), 5lxf (1.8033), 1fch (1.8052), 4u7o (1.8078), 3aai (1.8089), 4xek (1.8145), 4bwe (1.8192), 3uit (1.8192), 4gs7 (1.8199), 2vs0 (1.8214), 5gw9 (1.8238), 3d36 (1.8265), 4xpx (1.8269), 4qnc (1.8271), 3lrq (1.8298), 2e8g (1.8305), 6f1e (1.8306), 3cjl (1.8326), 5iu4 (1.8327), 4k1p (1.833), 2z3q (1.8338), 5adu (1.8348), 4er9 (1.8351), 5xyf (1.8352), 2c0l (1.8366), 3c90 (1.8383), 4phq (1.8418), 5owo (1.8425), 4i0x (1.844), 3r9v (1.845), 5tsa (1.8471), 3mq7 (1.8488), 6c1r (1.8513), 3s84 (1.8519), 2xl4 (1.8519), 2itb (1.8524), 4uc1 (1.8534), 1orj (1.8539), 3ejj (1.8549), 3va9 (1.8555), 4pgr (1.8567), 3pf0 (1.8572), 2zrr (1.8585), 3cx8 (1.859), 4zp0 (1.8601), 1hwg (1.862), 1gvn (1.8674), 1yuz (1.8695), 2rld (1.8698), 6ffi (1.8699), 3dza (1.8699), 5x56 (1.8713), 5an6 (1.8724), 4hfq (1.8733), 4gcz (1.8734), 1ax8 (1.8745), 3rkl (1.8765), 2vrw (1.877), 1nig (1.8772), 4xev (1.8782), 5ue0 (1.8811), 4xnv (1.8823), 3gwk (1.8824), 1wu3 (1.8839), 3c3w (1.885), 2j9u (1.8866), 4zf7 (1.8874), 1qkr (1.8905), 5j1n (1.8917), 3ogh (1.892), 2bl8 (1.8951), 1s0p (1.8968), 4gof (1.9019), 1mxr (1.9031),

3m6j (1.9044), 5xsj (1.9065), 4zmu (1.9068), 3pwx (1.9081), 5h5m (1.9094), 3gf9 (1.91), 3x0u (1.9105), 2gyq (1.9111), 3teq (1.9112), 2qsb (1.9125), 1r0d (1.9152), 3ld9 (1.9232), 5d50 (1.9275), 5ee7 (1.9283), 4bgo (1.9322), 2ap3 (1.9324), 6dfp (1.9326), 4wzx (1.9363), 4abm (1.9386), 4iwb (1.9395), 1yar (1.9399), 5hjf (1.9412), 4bmo (1.9441), 3vrc (1.9448), 3mtu (1.9478), 4m7c (1.948), 5tia (1.9488), 3frr (1.9496), 2vkp (1.9499), 5jhh (1.9522), 4m3l (1.9528), 6f6e (1.957), 5iji (1.9576), 3u9j (1.9587), 1t3u (1.9623), 1huw (1.9623), 3hl1 (1.9643), 2fu2 (1.9646), 4x1h (1.9654), 3nrh (1.9658), 5n7e (1.9663), 4bjs (1.9678), 6az6 (1.9691), 4pxz (1.9703), 5c39 (1.9721), 4wd8 (1.9733), 3wsc (1.9734), 1ory (1.975), 2nox (1.9773), 5cj1 (1.9795), 1xu9 (1.9821), 4iwk (1.9833), 5ghe (1.9875), 4fx5 (1.9878), 4ysx (1.9882), 2vxx (1.9911), 5kay (1.9916), 4yon (1.9916), 3ee4 (1.9956), 6ek7 (1.9956), 5ukv (1.9962), 2ip6 (1.9964), 1i5n (1.9971), 1he1 (1.998), 1h6g (1.9984), 5v4t (1.999), 1tjo (1.9994)

**Table S3. Data collection and refinement statistics of drug-free- and drug-bound ABLE.**

| | drug-free ABLE (6W6X) | apixaban-bound ABLE (6W70) |
|---|---|---|
| **Wavelength** | 0.9792 | 0.9792 |
| **Resolution range** | 45.9 - 1.297 (1.344 - 1.297) | 39.22 - 1.296 (1.342 - 1.296) |
| **Space group** | C 1 2 1 | P 1 21 1 |
| **Unit cell** | 51.426 43.648 91.8355 90 91.356 90 | 35.1935 78.445 43.1335 90 106.953 90 |
| **Total reflections** | 328760 (29902) | 656950 (16786) |
| **Unique reflections** | 50139 (4876) | 54943 (4971) |
| **Multiplicity** | 6.6 (6.1) | 12.0 (3.3) |
| **Completeness (%)** | 98.85 (98.41) | 98.86 (88.97) |
| **Mean I/sigma(I)** | 13.34 (2.01) | 27.92 (1.40) |
| **Wilson B-factor** | 16 | 13.57 |
| **R-merge** | 0.173 (0.6034) | 0.2924 (0.7827) |
| **R-meas** | 0.1872 (0.6566) | 0.3037 (0.9055) |
| **R-pim** | 0.07086 (0.2552) | 0.08143 (0.4402) |
| **CC1/2** | 0.987 (0.85) | 0.987 (0.542) |
| **CC*** | 0.997 (0.959) | 0.997 (0.839) |
| **Reflections used in refinement** | 49894 (4876) | 54871 (4971) |
| **Reflections used for R-free** | 2472 (215) | 2702 (271) |
| **R-work** | 0.1664 (0.2666) | 0.1607 (0.4573) |
| **R-free** | 0.2135 (0.3205) | 0.1835 (0.4404) |
| **CC(work)** | 0.975 (0.883) | 0.964 (0.819) |
| **CC(free)** | 0.959 (0.740) | 0.952 (0.832) |
| **Number of non-hydrogen atoms** | 2303 | 2259 |

| | | |
|---|---|---|
| **macromolecules** | 1964 | 1906 |
| **ligands** | 23 | 81 |
| **solvent** | 316 | 272 |
| **Protein residues** | 252 | 252 |
| **RMS(bonds)** | 0.009 | 0.015 |
| **RMS(angles)** | 0.93 | 1.2 |
| **Ramachandran favored (%)** | 98.79 | 100 |
| **Ramachandran allowed (%)** | 1.21 | 0 |
| **Ramachandran outliers (%)** | 0 | 0 |
| **Rotamer outliers (%)** | 0 | 1.1 |
| **Clashscore** | 2.81 | 0.26 |
| **Average B-factor** | 22.15 | 18.9 |
| **macromolecules** | 20.15 | 17.83 |
| **ligands** | 64.17 | 16.55 |
| **solvent** | 31.58 | 27.14 |

Statistics for the highest-resolution shell are shown in parentheses.

**Table S4. Data collection and refinement statistics of H49A mutant of unliganded ABLE.**

| H49A mutant (6X8N) | |
|---|---|
| **Wavelength** | 1.045 |
| **Resolution range** | 41.08 - 1.6 (1.657 - 1.6) |
| **Space group** | P 1 |
| **Unit cell** | 36.75 46.44 47 90.131 117.661 106.889 |
| **Total reflections** | 112953 (11087) |
| **Unique reflections** | 32545 (3202) |
| **Multiplicity** | 3.5 (3.5) |
| **Completeness (%)** | 94.73 (93.49) |
| **Mean I/sigma(I)** | 7.91 (1.09) |
| **Wilson B-factor** | 21.22 |
| **R-merge** | 0.0784 (1.038) |
| **R-meas** | 0.09316 (1.23) |
| **R-pim** | 0.04959 (0.6512) |
| **CC1/2** | 0.998 (0.612) |
| **CC*** | 0.999 (0.872) |
| **Reflections used in refinement** | 32472 (3201) |
| **Reflections used for R-free** | 1381 (131) |
| **R-work** | 0.2058 (0.3782) |
| **R-free** | 0.2404 (0.4233) |
| **CC(work)** | 0.957 (0.828) |
| **CC(free)** | 0.935 (0.751) |
| **Number of non-hydrogen atoms** | 2220 |
| **macromolecules** | 1957 |
| **solvent** | 263 |
| **Protein residues** | 252 |

| | |
|---|---|
| **RMS(bonds)** | 0.011 |
| **RMS(angles)** | 1.07 |
| **Ramachandran favored (%)** | 99.6 |
| **Ramachandran allowed (%)** | 0.4 |
| **Ramachandran outliers (%)** | 0 |
| **Rotamer outliers (%)** | 1.57 |
| **Clashscore** | 6.4 |
| **Average B-factor** | 32.83 |
| **macromolecules** | 31.98 |
| **solvent** | 39.09 |

Statistics for the highest-resolution shell are shown in parentheses.

**Additional Data S1 (.txt)**

This file contains the PDB accession codes and chain IDs of the proteins used to compile vdM databases.

**References and Notes**


1. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973). doi:10.1126/science.181.4096.223 Medline

2. I. V. Korendovych, W. F. DeGrado, *De novo* protein design, a retrospective. *Q. Rev. Biophys.* **53**, e3 (2020). doi:10.1017/S0033583519000131 Medline

3. B. Kuhlman, P. Bradley, Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 (2019). doi:10.1038/s41580-019-0163-x Medline

4. J. Dou, L. Doyle, P. Greisen Jr., A. Schena, H. Park, K. Johnsson, B. L. Stoddard, D. Baker, Sampling and energy evaluation challenges in ligand binding protein design. *Protein Sci.* **26**, 2426–2437 (2017). doi:10.1002/pro.3317 Medline

5. E. Marcos, B. Basanta, T. M. Chidyausiku, Y. Tang, G. Oberdorfer, G. Liu, G. V. T. Swapna, R. Guan, D.-A. Silva, J. Dou, J. H. Pereira, R. Xiao, B. Sankaran, P. H. Zwart, G. T. Montelione, D. Baker, Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201–206 (2017). doi:10.1126/science.aah7389 Medline

6. C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, D. Baker, Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013). doi:10.1038/nature12443 Medline

7. A. L. Day, P. Greisen, L. Doyle, A. Schena, N. Stella, K. Johnsson, D. Baker, B. Stoddard, Unintended specificity of an engineered ligand-binding protein facilitated by unpredicted plasticity of the protein fold. *Protein Eng. Des. Sel.* **31**, 375–387 (2018). doi:10.1093/protein/gzy031 Medline

8. J. Dou, A. A. Vorobieva, W. Sheffler, L. A. Doyle, H. Park, M. J. Bick, B. Mao, G. W. Foight, M. Y. Lee, L. A. Gagnon, L. Carter, B. Sankaran, S. Ovchinnikov, E. Marcos, P.-S. Huang, J. C. Vaughan, B. L. Stoddard, D. Baker, De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018). doi:10.1038/s41586-018-0509-0 Medline

9. E. P. Barros, J. M. Schiffer, A. Vorobieva, J. Dou, D. Baker, R. E. Amaro, Improving the efficiency of ligand-binding protein design with molecular dynamics simulations. *J. Chem. Theory Comput.* **15**, 5703–5715 (2019). doi:10.1021/acs.jctc.9b00483 Medline

10. G. Grigoryan, W. F. Degrado, Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011). doi:10.1016/j.jmb.2010.08.058 Medline

11. P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014). doi:10.1126/science.1257481 Medline

12. K. Szczepaniak, G. Lach, J. M. Bujnicki, S. Dunin-Horkawicz, Designability landscape reveals sequence features that define axial helix rotation in four-helical homo-oligomeric antiparallel coiled-coil structures. *J. Struct. Biol.* **188**, 123–133 (2014). doi:10.1016/j.jsb.2014.09.007 Medline

13. N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S.-Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien, W. F. DeGrado, De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017). doi:10.1038/nchem.2846 Medline

14. G. G. Rhys, C. W. Wood, J. L. Beesley, N. R. Zaccai, A. J. Burton, R. L. Brady, A. R. Thomson, D. N. Woolfson, Navigating the structural landscape of de novo α-helical bundles. *J. Am. Chem. Soc.* **141**, 8787–8797 (2019). doi:10.1021/jacs.8b13354 Medline

15. A. J. Reig, M. M. Pires, R. A. Snyder, Y. Wu, H. Jo, D. W. Kulp, S. E. Butch, J. R. Calhoun, T. Szyperski, E. I. Solomon, W. F. DeGrado, Alteration of the oxygen-dependent reactivity of de novo Due Ferri proteins. *Nat. Chem.* **4**, 900–906 (2012). doi:10.1038/nchem.1454 Medline

16. A. N. Lupas, J. Bassler, S. Dunin-Horkawicz, in *Fibrous Proteins: Structures and Mechanisms*, D. A. D. Parry, J. M. Squire, Eds. (Springer, Cham, 2017), pp. 95–129.

17. A. Lombardi, F. Pirro, O. Maglio, M. Chino, W. F. DeGrado, De novo design of four-helix bundle metalloproteins: One scaffold, diverse reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019). doi:10.1021/acs.accounts.8b00674 Medline

18. J. R. Desjarlais, T. M. Handel, De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 (1995). doi:10.1002/pro.5560041006 Medline

19. J. Janin, S. Wodak, M. Levitt, B. Maigret, Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**, 357–386 (1978). doi:10.1016/0022-2836(78)90408-4 Medline

20. M. J. McGregor, S. A. Islam, M. J. E. Sternberg, Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295–310 (1987). doi:10.1016/0022-2836(87)90314-7 Medline

21. J. W. Ponder, F. M. Richards, Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791 (1987). doi:10.1016/0022-2836(87)90358-5 Medline

22. B. I. Dahiyat, S. L. Mayo, Protein design automation. *Protein Sci.* **5**, 895–903 (1996). doi:10.1002/pro.5560050511 Medline

23. J. K. Lassila, H. K. Privett, B. D. Allen, S. L. Mayo, Combinatorial methods for small-molecule placement in computational enzyme design. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16710–16715 (2006). doi:10.1073/pnas.0607691103 Medline

24. J. Singh, J. M. Thornton, *Atlas of Protein Side-Chain Interactions* (Oxford Univ. Press, 1992).

25. A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker, New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006). doi:10.1110/ps.062353106 Medline

26. K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan, J. Meiler, Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987–2998 (2010). doi:10.1021/bi902153g Medline

27. R. Ferreira de Freitas, M. Schapira, A systematic analysis of atomic protein-ligand interactions in the PDB. *MedChemComm* **8**, 1970–1981 (2017). doi:10.1039/C7MD00381A Medline

28. B. North, C. M. Summa, G. Ghirlanda, W. F. DeGrado, $D_n$-symmetrical tertiary templates for the design of tubular proteins. *J. Mol. Biol.* **311**, 1081–1090 (2001). doi:10.1006/jmbi.2001.4900 Medline

29. D. H. Williams, E. Stephens, D. P. O'Brien, M. Zhou, Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew. Chem. Int. Ed.* **43**, 6596–6616 (2004). doi:10.1002/anie.200300644 Medline

30. S. K. Tan, K. P. Fong, N. F. Polizzi, A. Sternisha, J. S. G. Slusky, K. Yoon, W. F. DeGrado, J. S. Bennett, Modulating integrin αIIbβ3 activity through mutagenesis of allosterically regulated intersubunit contacts. *Biochemistry* **58**, 3251–3259 (2019). doi:10.1021/acs.biochem.9b00430 Medline

31. F. Thomas, W. M. Dawson, E. J. M. Lang, A. J. Burton, G. J. Bartlett, G. G. Rhys, A. J. Mulholland, D. N. Woolfson, De novo-designed α-helical barrels as receptors for small molecules. *ACS Synth. Biol.* **7**, 1808–1816 (2018). doi:10.1021/acssynbio.8b00225 Medline

32. J. Park, B. Selvaraj, A. C. McShan, S. E. Boyken, K. Y. Wei, G. Oberdorfer, W. DeGrado, N. G. Sgourakis, M. J. Cuneo, D. A. A. Myles, D. Baker, De novo design of a homo-trimeric amantadine-binding protein. *eLife* **8**, e47839 (2019). doi:10.7554/eLife.47839 Medline

33. A. A. Glasgow, Y.-M. Huang, D. J. Mandell, M. Thompson, R. Ritterson, A. L. Loshbaugh, J. Pellegrino, C. Krivacic, R. A. Pache, K. A. Barlow, N. Ollikainen, D. Jeon, M. J. S. Kelly, J. S. Fraser, T. Kortemme, Computational design of a modular protein sense-response system. *Science* **366**, 1024–1028 (2019). doi:10.1126/science.aax8780 Medline

34. N. Tokuriki, D. S. Tawfik, Protein dynamism and evolvability. *Science* **324**, 203–207 (2009). doi:10.1126/science.1169375 Medline

35. T. J. Stout, C. R. Sage, R. M. Stroud, The additivity of substrate fragments in enzyme-ligand binding. *Structure* **6**, 839–848 (1998). doi:10.1016/S0969-2126(98)00086-0 Medline

36. D. A. Keedy, Z. B. Hill, J. T. Biel, E. Kang, T. J. Rettenmaier, J. Brandão-Neto, N. M. Pearce, F. von Delft, J. A. Wells, J. S. Fraser, An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *eLife* **7**, e36307 (2018). doi:10.7554/eLife.36307 Medline

37. N. Polizzi, npolizzi/combs_pub: Combs, Version v0.0.1, Zenodo; https://doi.org/10.5281/zenodo.3910780.

38. J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999). doi:10.1006/jmbi.1998.2401 Medline

39. V. B. Chen, W. B. Arendall 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure

validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010). doi:10.1107/S0907444909042073 Medline

40. A. Bakan, L. M. Meireles, I. Bahar, ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011). doi:10.1093/bioinformatics/btr168 Medline

41. J. M. Word, S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, J. S. Richardson, D. C. Richardson, Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285**, 1711–1733 (1999). doi:10.1006/jmbi.1998.2400 Medline

42. J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci.* **24**, 508–524 (2015). doi:10.1002/pro.2610 Medline

43. A. Lombardi, C. M. Summa, S. Geremia, L. Randaccio, V. Pavone, W. F. DeGrado, Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6298–6305 (2000). doi:10.1073/pnas.97.12.6298 Medline

44. J. M. Dunce, O. M. Dunne, M. Ratcliff, C. Millán, S. Madgwick, I. Usón, O. R. Davies, Structural basis of meiotic chromosome synapsis through SYCP1 self-assembly. *Nat. Struct. Mol. Biol.* **25**, 557–569 (2018). doi:10.1038/s41594-018-0078-9 Medline

45. C. A. K. Lundgren, D. Sjöstrand, O. Biner, M. Bennett, A. Rudling, A.-L. Johansson, P. Brzezinski, J. Carlsson, C. von Ballmoos, M. Högbom, Scavenging of superoxide by a membrane-bound superoxide oxidase. *Nat. Chem. Biol.* **14**, 788–793 (2018). doi:10.1038/s41589-018-0072-x Medline

46. S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. DiMaio, J. H. Pereira, B. Sankaran, G. Seelig, P. H. Zwart, D. Baker, De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016). doi:10.1126/science.aad8865 Medline

47. Y. Hong, Z. Huang, L. Guo, B. Ni, C.-Y. Jiang, X.-J. Li, Y.-J. Hou, W.-S. Yang, D.-C. Wang, I. B. Zhulin, S.-J. Liu, D.-F. Li, The ligand-binding domain of a chemoreceptor from Comamonas testosteroni has a previously unknown homotrimeric structure. *Mol. Microbiol.* **112**, 906–917 (2019). doi:10.1111/mmi.14326 Medline

48. M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong, NWchem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**, 1477–1489 (2010). doi:10.1016/j.cpc.2010.04.018

49. J. Liang, K. A. Dill, Are proteins well-packed? *Biophys. J.* **81**, 751–766 (2001). doi:10.1016/S0006-3495(01)75739-6 Medline

50. N. D. Clarke, S.-M. Yuan, Metal search: A computer program that helps design tetrahedral metal-binding sites. *Proteins* **23**, 256–263 (1995). doi:10.1002/prot.340230214 Medline

51. M. Lee, T. Wang, O. V. Makhlynets, Y. Wu, N. F. Polizzi, H. Wu, P. M. Gosavi, J. Stöhr, I. V. Korendovych, W. F. DeGrado, M. Hong, Zinc-binding structure of a catalytic amyloid

from solid-state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6191–6196 (2017). doi:10.1073/pnas.1706179114 Medline

52. S. J. Lahr, D. E. Engel, S. E. Stayrook, O. Maglio, B. North, S. Geremia, A. Lombardi, W. F. DeGrado, Analysis and design of turns in α-helical hairpins. *J. Mol. Biol.* **346**, 1441–1454 (2005). doi:10.1016/j.jmb.2004.12.016 Medline

53. C. M. Summa, M. M. Rosenblatt, J.-K. Hong, J. D. Lear, W. F. DeGrado, Computational de novo design, and characterization of an $A_2B_2$ diiron protein. *J. Mol. Biol.* **321**, 923–938 (2002). doi:10.1016/S0022-2836(02)00589-2 Medline

54. P. Bradley, K. M. S. Misura, D. Baker, Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871 (2005). doi:10.1126/science.1113801 Medline

55. S. L. Reid, D. Parry, H.-H. Liu, B. A. Connolly, Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: A study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry* **40**, 2484–2494 (2001). doi:10.1021/bi001956p Medline

56. A. M. Rossi, C. W. Taylor, Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365–387 (2011). doi:10.1038/nprot.2011.305 Medline