## Supplementary Methods

**General considerations about the β-barrel fold.** In the past 30 years, descriptive work has been carried out to understand the geometric constraints in closed β-barrel structures. These findings laid the important foundation for the design work presented in this paper and are summarized in this subsection. This supplementary text is an addition to Figure 1. As indicated in the main text, β-barrels are characterized by the total number of strands (n) and the shear number (S), which is the total number of shifts of β-strand register between the first and the last strands. Because of the alternation of the hydrogen bonded and non-hydrogen bonded pairs of residues in anti-parallel β-sheets, the parameter S must be an even number to maintain continuous strand pairing. Together, n and S define the type of the β-barrel fold and its basic geometric properties. The ideal radius (r) of β-barrels has been described as a function of the number of strands (n), the shear number (S), the average distance between two β-strands (D) and the average distance between two residues on a β-strand (d):

$$r = [\,(Sd)^2 + (nD)^2\,]^{1/2} / [\,2n\ sin\ (\pi/n)\,] \qquad \text{eq. 1}$$

The β-strand staggering angle to the main axis of the barrel ( θ ) was defined as:

$$\tan\,\theta = \frac{S \cdot d}{n \cdot D} \qquad \text{eq. 2}$$

Because of the relationships defined above, the ideal barrel radius and strands staggering angle are not continuous -- but discrete values. They are constrained by the hydrogen bonding pattern of the β-sheet.

The shear number also defines the pattern of the residues packing in the β-barrel (Extended Data Figure 1b-d). Because of the alternation of residues facing in and out on the β-strands, the side-chains of the residues lining up in the direction of the hydrogen bonds (perpendicular to the β-strands) face the same side of the β-sheet. We named such arrangements “Cβ-strips” (Fig. 1c). The total number of Cβ-strips in the barrel is equal to the shear number, and half of them point into the barrel. Therefore, a barrel of type (n=8; S=8) is packed with 4 Cβ-strips, resulting in a 4-fold symmetric arrangement of side-chains in the hydrophobic core. A barrel of type (n=8; S=10) is packed with the intertwinned side-chains provided by 5 Cβ-strips. Soluble β-barrels large enough to accommodate a ligand binding cavity characterized so far have 8 or 10 strands with a shear number ranging from 8 to 12. Almost all β-barrels with 8 strands and a shear number of 8 belong to the superfamily of TIM barrels with parallel β-strands. A small family of anti-parallel (n=8, S=8) transferrin-binding β-barrels has been characterized, but they are stable only in the presence of a lipid membrane. The theoretical radius of β-barrels with 10 strands, on the other hand, is too large to be packed with natural hydrophobic amino acids and their cavity is usually squished to allow hydrophobic packing of the core. The same is true for β-barrels of type (n=8; S=12) (lipocalins family), which require disulfides for folding and/or stability. Therefore, we started this work focusing on beta-barrels of type (n=8; S=10), which can be packed with hydrophobic side-chains and are large enough to accommodate a small molecule binding site.

**Parametric models for β-barrel folds.** Hyperboloids defined by the quadratic equation(eq.3),

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} - \frac{z^2}{C^2} = 1 \qquad \text{eq. 3}$$

have previously been used to fit the structures of β-barrel proteins, where the β-strands pass through the generating lines of hyperboloid surface and the equatorial planes of the hyperboloids cut through the

β-sheets. Although this model assumes divergence of the β-strands away from the equatorial plane of the hyperboloid, it has the advantage of representing the β-strands as straight lines that are easily populated with residues. The parameters A and B in eq. 3 are the radii of the elliptical equatorial plane of the hyperboloid ($r_A$ and $r_B$ in Fig. 1a). The parameter C can be expressed as a function of the ideal tilt angle of the β-strands to the main axis of the barrel (Z axis in Fig. 1a), $\theta$, the parameters A and B, and the angular coordinates of the strand in the equatorial plane $\alpha$ [14]:

$$C = cot(\theta) \cdot [A^2 \cdot sin^2\alpha + B^2 \cdot cos^2\alpha]^{1/2} \quad \text{eq. 4}$$

The angle $\alpha$ is a function of the strand number n, since the n strands are regularly arranged around the elliptic equatorial plane:

$$\alpha = 2 \cdot \pi / n \quad \text{eq. 5}$$

The β-strands are modeled as equidistant vectors v along the hyperboloid surface:

$$v = [-A \cdot sin(\alpha),\ B \cdot cos(\alpha),\ C] \quad \text{eq.6}$$

These vectors are then populated with Cα, assuming ideal distance between residues on a β-strand and ideal backbone hydrogen bonds, perpendicular to direction of the β-strands. The other backbone heavy atoms were built over these Cα traces using the BBQ software (see Supplementary Data for the parameters and scripts used in this study).

**Backbone construction and sequence design.** The hyperboloid model presented above was used to generate an ensemble of β-strands arrangements that were further minimized with geometric constraints to enforce hydrogen bonds between strands. To select the backbones that would allow optimal hydrophobic packing of the core, low energy sequences were designed using a RosettaScripts flexible-backbone design protocol with filters reporting the hydrophobic packing density inside the barrel (see Supplementary Data). Based on good packing metrics, a limited number of starting backbones (P1 to P12 in Supplementary Table 2) were selected for loop closure using the loop hash protocol implemented in RosettaRomodel (see Supplementary Data). Two rounds of loop closure were applied to each backbone to connect the strands with ideal, short hairpins. Combinations of different lengths of β-turns were sampled by constructing several blueprint files. The β-barrels with the best hydrogen bond energy were selected for the next stage of sequence design using a RosettaScripts flexible-backbone design protocol (see Supplementary Data). In this design stage, the sequence of the β-turns was constrained to the consensus sequence obtained from native proteins. Each turn was identified by its ABEGO type sequence and a sequence profile was generated (in the format of Rosetta resfile, see Supplementary Data).

**Test of the proposed principles for constructing β-barrel backbones.** The hypothetical roles of glycine kinks and β-bulges in β-barrels were tested computationally using a similar approach. First, a set of 50 to 100 poly-valine backbones was generated and minimized with constraints to ensure hydrogen bond connectivity across the β-barrel. We applied the Rosetta Relax protocol (iterative backbone minimization and repacking of side-chains) to each set of backbones and the number of retained backbone hydrogen bonds was evaluate (Extended Data Fig. 2a). We built three sets of disconnected β-strands based on different parametric models for β-barrels: 1) The parametric hyperboloid model was described above; 2) A parametric cylindrical model was adapted from[13]; and 3) A coiled coil parametric model was adapted from the helical bundle generation model, where the parameters of the small helices were modified to generate β-strands[57]. For comparison, we built another set of disconnected β-barrels using Rosetta fragment assembly [58,21]. To test our final 2D map design with strategically chosen glycine kink positions,

β-bulges and β-turns, we built sets of β-strands connected with β-turns and attributed an ABEGO type to each residue in the 2D map to enforce local torsional deviations from the canonical β-sheet B space. Kinks and bulges were modeled as residues with ABEGO types E and A, respectively. β-turns were modeled with the ABEGO types GG (canonical type I' β-turn), AA (type I β-turn) or AAG (β-turn with intrinsic G1 bulge).

**Fragments-based approach to β-barrel design.** As discussed in the main text, Rosetta fragment-based backbone assembly approach allows the introduction of local torsional deviation from ideal β-strands -- such as β-bulges and glycine kinks -- at specific positions. Several previous studies pointed out the presence and relative abundance of these structural features in native β-barrels. However, the precise role of β-bulges and glycine kinks and their localization in the β-sheet remained unclear. Here, we described in detail the rationale for building the 2D map describing the (n=8, S=10) β-barrel presented in the main text. We use β-bulges and glycine kinks to release the tension in the β-sheet and shape the cavity of the barrel. We also present the computation methods used to implement the sequence design. We defined the side of the β-barrel with the N- and C-termini as the "bottom" of the barrel, and the opposite side flanked with four β-turns and accommodating the ligand-binding cavity as the "top". This bottom/top definition was used throughout the text and the figures.

**Design of the 2D map of the residues connectivity in a β-barrel of type (n=8, S=10).** By collecting and analysing the native β-barrel structures, we found that the longest β-strands in soluble β-barrels of type (n=8; S=10), and even larger (such as n=10; S=12), span 12 residues (this number excludes any β-bulges). Therefore, the maximum length of β-strands in the 2D map was set to 12. We also found that an up-and-down β-barrel with antiparallel strands connected by short β-turns could not be composed of strands of the same length and that the length of each strand is constrained by the shear number S. We sought to distribute the total register shift of 10 residues across the 8 strands of the β-barrel. We chose to shift 3 out of 4 hairpins on each side of the barrel by 2 residues (this is the minimal register shift to conserve continuous hydrogen bonds), and shift the last hairpin by 4 residues (this is a "double" register shift, resulting in a hairpin with a longer unpaired edge). Because of the alternation of β-turns between the top and bottom hairpins and the staggering pattern between strands in each hairpin, the strands with odd numbers need to be shorter than the even strands if they all to be connected with short β-turns. As a result of this constraint, the even strands in our 2D map span 12 or 10 residues and the odd strands span 8 or 10 residues (Fig. 1d, left). The specific length of a strand depends on the position to the "double" register shift.

We further defined the backbone torsion angle bins (Fig. 1d, right) for each residue in the map. Regular β-strand positions were attributed the space B, while the glycine kinks positions were modeled in the E ABEGO space. One glycine kink was placed in each Cβ-strips to relieve the strain inherent to the closure of the β-sheet on itself -- for a total of 5 glycine kinks in our barrel of type (n=8;S=10), as described in the main text. We sculpted a square shaped cavity by placing each glycine kink at the fifth position of each Cβ-strips from the bottom of the barrel. The vector between the Cα of the first residue on the Cβ-strip to the Cα of the glycine (which has an estimated length of 4 times the average distance between two β-strands (D) -- 17.6 Å) can be projected on the equatorial plane of the β-barrel by being multiplied by $\sin(\theta)$, where $\theta$ is the ideal staggering angle of the β-strands to the central axis of the barrel (estimated to 43° for a barrel of type (n-8;S=10) using the eq. 2 above). The resulting vector (Fig. 1d) has a length of

$4D \times \sin(\theta)$, which corresponds to a shift of 12 Å around the circumference of a β-barrel -- or approximately a quarter of the average circumference of a barrel of type (n=8;S=10) (estimated to 50 Å using eq. 1). Three of such vectors correspond to three sides of the square design. The remain two vectors (both starting on the β-strand bearing the "double" register shift) partially overlap and correspond to the fourth side of the square.

We noted that the top hairpin bearing the double register shift had a tendency to curve away from the center of the β-barrel. This is likely because the longer register shift allows more twist, which results in convex curvature of the edge β-strand residues (Extended Data Fig. 3a). To correct the twist of the strand, we introduced an additional glycine kink. It forces concave curvature in the area of the register shift, giving the longer hairpin a conformation similar to the other top hairpins.

In the main text, we show that β-bulges are necessary to relieve the strain associated with high β-sheet curvature at the extremities of the β-hairpins (in proximity to the β-turns). All four β-bulges in the top barrel are located in the same position in the corresponding β-hairpin, so are the bottom β-bulges. According to the ββ rule[58], if two antiparallel β-strands are connected with a short β-turn, the last residue of the first strand (position -1 relative to the β-turn) and the first residue of the following strand (position +1 relative to the β-turn) form a hydrogen bonded pair. Based on this ββ rule and the preferred hydrogen bond connectivity of β-bulges[22,23,59], β-bulge locations were constrained to:

(i) the un-paired edge of each hairpin, which corresponds to the first strand of the bottom hairpins and the second strand on the top hairpins; and (ii) preceding the closest hydrogen bonded pair of residues to the β-turns. Therefore, the ideal placement of β-bulges in our 2D map is at position -2 from the "bottom" β-turns (preceding the paired β-strand residue at position -1) and position +1 from the "top" β-turns (preceding and replacing the β-strand residue at position +1, which now shifts to position +2).

When considering the type of β-turns to use for connecting the strands whose twist has been altered with β-bulges, we found that, in native proteins, the non-canonical type I β-turn (with the ABEGO type sequence AA) is prefered when a β-bulge is located in position -2 ( the location decided for the bottom β-bulge, Extended Data Fig. 3c). The use of a type I β-turn in the bottom hairpin was further supported by its higher frequency in native β-barrels, compared to the canonical type I' β-turn (with ABEGO type sequence GG) (Extended Data Fig. 3d). When all native β-sheet protein structures are considered, the canonical GG type turn is much more common than the AA type β-turn, presumably because its twist matches most β-strands without bulges. The AA β-turn is the mirror image of the GG turn and twists in the opposite direction. The most common β-turn in the native β-barrels is the 3-residues AAG turn (Extended Data Fig.3d). This type of turn has an intrinsic G1[60-62] β-bulge at the third (G) position preceding the first residue of the β-strand and modifies the hydrogen bond pattern of the pair of β-strands residues flanking the turn. The G1 bulge is ideally positioned to satisfy the localization constraints of the top β-bulges and was used to connect all four top β-hairpins.

We next sought to specify structural features that can control the precise registry between the first and the last strands, as a strategy for negative design against competing alternatives such as amyloid-like aggregates. We found that a conserved "tryptophan corner" in lipocalins was proposed to stabilize the native barrel structure relative to the misfolded structures[25,63]. Therefore, we decided to introduce a tryptophan on the first strand (immediately after a sharp glycine turn) and a buried arginine on the last strand to guide the pairing between both strands with specific side-chain interactions.

**Constraints for maintaining the hydrogen bond connectivity defined in the 2D map.** Each backbone hydrogen bond in the 2D map was described with three pairs of constraints (Extended Data Fig. 5b). Additionally, we added a set of distance and dihedral constraints to specify the interactions specific to the "tryptophan corner" motif, which is otherwise constrained by interactions involving side-chains, which are absent from the backbones generated with Rosetta's centroid model. The specific torsion tolerances for the residues of the motif were determined by analyzing the torsion angles of a set of native tryptophan corner motifs extracted from the PDB (Extended Data Fig. 3g-j).

**Constraints on amino acid identity with a resfile.** A Rosetta resfile was used to constraint the sequence of the β-turn positions to prefered profiles observed in native proteins. An analysis of the sequence preference for the AAG and AA β-turns and flanking position was carried out to and specified in a Rosetta resfile. Our design models and crystal structure revealed that the prefered sequence for the β-turn resulted good intra- and inter-hairpin networks of hydrogen bonds (Fig. 2h&j). Additionally, the resfile defines the amino acid identity of the key residues of the tryptophan corner (G9, W11 and R107), the positions of glycine kinks and the positions constrained to proline. The dedicated positions for prolines were introduced as an attempt to improve the stability of the protein and were chosen based on several observation:

- Pro8 belongs to the tryptophan corner and is located between the residue participating to hydrophobic core packing and the glycine. Although the sequence profile of native tryptophan corner motifs does not show a preference for proline, backbone torsion distributions at that position (Cα-3 in Extended Data Fig. 3i) and the preceding position(Cα-4 in Extended Data Fig. 3h) are compatible with torsion preferences for prolines and pre-prolines.
- Pro31 and Pro50 were placed to protect the edges of the β-strand with double register shifts from forming intermolecular strand-to-strand interactions, as a negative design consideration.

**Scaffolds preparation for RIF docking.** Several sets of β-barrel scaffolds were constructed as input for RIF docking. A first backbone set was generated using the same centroid-level fragment assembly and minimization protocol developed for nonfunctional β-barrel design, during which hydrogen bond connectivity was maintained by three types of geometric constraints illustrated in Extended Data Fig. 5b. A second set of backbones, featuring broader phi, psi torsion angles distributions and overall higher structural diversity (estimated by all-to-all RMSD), was generated following the same protocol with only two pairs of constraints per hydrogen bond (the N-H-O angle constraint was left out). Output backbones were filtered for backbone torsion geometry (phi, psi, omega), backbone clashing and hydrogen bond energy. 104 and 96 backbones from each set were selected for RIF docking. These two sets of backbones yielded only a small number of RIF docking solutions (162 for 104 scaffolds and 225 for 96 scaffolds respectively, blue bars in Extended Data Fig. 5d).

We sought to further improve the backbone geometry and diversity of these backbones by performing two rounds of flexible-backbone design calculation using Rosetta full-atom energy without hydrogen bond constraints. Because we did not want to optimize the backbones for a specific non-functional sequence, this "pre-design" calculation was carried out using a generalized Ramachandra statistical potential (Rosetta/main/database/scoring/socore_functions/rama/flat/Rama_XPG_3level.txt) that only distinguishes three amino acid side chains (glycines, prolines, and all other amino acids). Such design protocol did not produce reasonable sequences to fold beta-barrels because of the low sampling (the cores were not packed

well and core residues were mostly methionines and valines) but improved the backbone torsion angles (phi, psi, omega) and diversified the backbones. The all-to-all backbone Cα RMSD of these backbones was 1.4 ± 0.3 Å. These 200 "pre-designed" scaffolds yielded 2,102 non-redundant RIF docking solutions in total (Extended Data Fig. 5d).
Sequence design calculations were continued using RIF docking solutions from "pre-designed" scaffolds. A uniformly-defined "binding site" was applied to all the BB1-type β-barrel scaffolds. Sixteen positions in the upper half of the barrel except glycine and β-turn positions were used to define the searching grids for RIF docking (Supplementary Data). The sequences selected for ordering were originated from 20 of the 200 input scaffolds. That "productive" subset of input scaffolds had an average omega score of 13.8 Rosetta Energy Unit (REU); an average rama score of -9.8 REU and an average hbond_lr_bb score of -69.8 REU.

**DFHBI ligand preparation.** 3D atom coordinates of DFHBI was downloaded from PubChem database (PubChem CID: 70808995) and converted to .mol2 file by Avogadro software (Avogadro Chemistry). Partial charges assigned by Avogadro was corrected according to Amber force field (Supplementary Data). Parameter files that define the atom and bond types were generated according to the internal definition of Rosetta (Supplementary Data). The planar *Z*- conformation of DFHBI was used as the only conformer during RIF docking and Rosetta design calculations in order to obtain selective binding interactions compatible with fluorescence activation. The torsional degree of freedom in DFHBI was eliminated for the same purpose, which was added back during the post-design ligand docking simulation. DFHBI was placed into the β-barrel scaffolds by RIF docking with its two fluorine atoms replaced by protons. At the time of this work was done, the RIF docking database did not recognize fluorine atoms and the internal RIF docking ranking scores does not include electrostatic potentials. Since the fluorine atoms are of similar size as protons for most hydrophobic packing interactions, we think this replacement did not affect the search for interactions. Fluorines were added back during Rosetta design calculations with proper partial charges.

**Rosetta sequence design following RIF docking.** Rosetta energy-based sequence design was carried out for all 2,102 RIF docking solutions. The design protocol first defined four regions of the β-barrel: the ligand binding site (Cα within 10Å distance from atom C4 (the bridging carbon between the two rings) in DFHBI), the protein core, surface and boundary regions (based on the number of neighboring side chains [30]). Each position in the protein was assigned to one of these four regions. In order to further reduce the sequence design space, each position was also categorized by its secondary structure. The combination of "regional" and secondary structure definitions was taken to refine the sequence search space. The amino acid propensities for helix, coil, sheet in protein core, boundary and surface (9 categories in total) in natural proteins were analyzed according to the definition of protein depth (see below, and Supplementary Table 17) and implemented in the design protocol (Supplementary Data). Special positions including glycine kinks, β-turns, β-bulges and tryptophan corner were specially treated using a resfile.

Design calculation started with a fixed-backbone calculation to optimize the residues in ligand binding site for better interface energy while the coordinating residues from RIF docking were kept fixed. Then the rest of scaffold was designed with backbone flexibility from energy-based gradient minimization[21],

during which, the binding-site residues were kept fixed with flexible sidechain torsions ("packable"). The sequence-guided full-protein backbone refinement introduced in the second step would change the geometry of the ligand binding site and the original binding configuration would no longer be the optimal solutions. To continue, the same fixed-backbone binding site design was repeated with the updated backbone (In the ideal case, we would expect covergencies on both backbone conformation and binding-site sequences with these dual optimization goals). This two-step design calculation was repeated three times to search for sequences that not only accommodate the DFHBI-binding interactions from RIF docking but also form a coherently-packed protein core. Output models were evaluated by total energy of the complex, backbone omega geometry, backbone beta-pairing hydrogen bonding energy, interface energy, interface shape complementarity, and the number of buried unsatisfied polar atoms on the interface (Supplementary Data). About 13% of output designs fell into the top half of all these evaluated metrics, and were continued for next round of optimization. A major problem found after the first round sequence design was that the protein core were mainly packed by methionines with long hydrophobic side chain touching each "side" of the rectangular barrel. An amino-acid composition control was introduced in the second round of design trying to bias the searching towards aromatic residues. Similar criteria based on the distribution were used to select out "improved" designs. The third round of design was done to release the fixation imposed on the ligand-coordinating interactions from RIF docking so that the binding configuration can be adjusted by energy-based optimization. We also noticed that the geometric constraints used to construct β-barrel backbones distorted the peptide omega angles and design calculations done in torsional space were not able to recover the right backbone geometry completely. Instead, an energy-based gradient minimization in Cartesian space was performed to correct the peptide bond geometry. An additional round of design (using the same Round 3 design protocol) were carried out to refine the sequences. 460 output designs from on 44 different RIF docking solutions (based on 32 input scaffolds) were selected for profile-based sequence refinement[64].

**Statistical analysis of amino acid preferences in natural proteins.** A non-redundant PDB list from PISCES[65] was used with the following cutoffs: sequence identity <30%, resolution <3.0 Å, length >30 (date 03/14/2015). The chain list was further filtered to remove any transmembrane proteins using the annotation from PDBTM[66]. To allow for more accurate computation of secondary structure (SS) and solvent accessibility (SA), each chain along with the interacting neighbors in the biologically-relevant assembly were extracted from RCSB[67]. STRIDE[68] was used to extract the SS classes. The 8 SS classes were reduced to 3. H, G, and I (from STRIDE) were reduced to (H)elix. E, B, and b were reduced to (E)xtended or sheet. T and C were reduced to (C)oil. DEPTH[69] was used to compute the SA. Residues with depth 0 to 5 Å were classified as exposed and 5 to 10 Å as buried. Using this classification we computed propensities of each amino acid (AA) given SS and SA. More specifically, we calculated: $\log_2(P(AA|SS,exposed)/P(AA))$, $\log_2(P(AA|SS,buried)/P(AA))$ and $\log_2(P(AA|SS)/P(AA))$ listed in Supplementary Table 17.

**Profile-based sequence refinement.** After four rounds of Monte Carlo-based design calculations described above, additional two rounds profile-based sequence refinement were carried out. 460 top designs were naturally clustered based on their starting RIF docking solutions. A sequence profile was generated for each cluster. The sequence design calculation was then restricted to residue identities that

already appeared in those best solutions from the preceding design calculations. This was implemented by presenting the cluster-specific sequence profiles in the format of Rosetta resfile (Supplementary Data). Within each cluster, the "functional site", which was the coordinating residues from RIF docking in this case, were totally conserved while the hydrophobic packing core showed high divergence. We inspected each cluster and manually introduced residues for potential improvement. For example, whenever a buried unsatisfied Tryptophan was seen in the packing core, cross-strand Serine and Threonine (which were not allowed for designing a simple 100% hydrophobic core) were added to the resfile in the hope that Rosetta would detect a favorable hydrogen bond between those residues. The same procedure was performed twice to re-enforce the appearing structural features. 42 final designs spanning 22 clusters were selected for experimental characterization (Supplementary Table 4).

**Post-design Analysis.** Rosetta *ab initio* folding simulations were performed for the final designs using a modified protocol. Since the secondary structure prediction methods have a low success rate for predicting beta structures with irregular features (Gly-kinks and β-bulges), we modified the *ab initio* folding protocol by providing the secondary structure profile that matches the design model for picking short peptide fragments. Thus, the folding simulation would use only beta fragments to assess if there were alternative low-energy conformations. With this assumption, all the 42 designs had a typical funnel-like folding landscape. The energy gap was scaled to 0 to 1 with 1 representing a perfect funnel-like landscape[35] (Extended Data Fig. 5e). With all the biochemical characterization results, there was no clear correlation between scaled folding energy gap and β-barrel formation. Most of the failed designs found alternative conformation by associating with multiple peptide chains (7/42 designs were insoluble and 14/42 designs form soluble aggregates, Supplementary Table 3). Far-UV CD spectra indicated that designs forming soluble aggregates formed structured beta conformations instead of random coils (Supplementary Table 3).
To validate the ligand binding interactions, we built the model for the *apo* protein and performed ligand docking simulations. The unbound protein conformation was sampled by running independent short-time MD simulations starting from the protein conformation in the designed complex model. The average structure from MD simulations was used to perform 2,000 independent docking simulations with flexible side chains and backbones in the binding site[70]. As a comparison, the protein conformation in the designed complex before MD refinement was used to perform the same ligand docking simulations (Extended Data Fig. 5f).

**Modeling mutations.** Mutations from the deep mutational scanning maps (Extended Data Fig. 8a-c) and variants arising from yeast library selection (Extended Data Fig. 8f, 10a&b) were modeled using RosettaScripts[71]. Mutations were first introduced into the parent design model, then a full-scale flexible-backbone relaxation calculation was done using FastRelax protocol with three cycles (Supplementary Data).

**Experimental Materials.** DFHBI was purchased from Lucerna (Brooklyn, NY), and were dissolved in DMSO as instructed by manufacturer. Acridine yellow was purchased from Sigma Aldrich (St. Louis, MO) and dissolved in EtOH. FITC-conjugated anti-cMyc antibodies were purchased Immunology Consultants Labs (catalog number: CMYC-45F) and used as instructed with 50-fold dilution. Thrombin

was purchased from MilliporeSigma (Burlington, MA). Trypsin-EDTA (0.25%) solution was purchased from Life Technologies (Danvers, MA). **α**-Chymotrypsin from bovine pancreas was purchased from Sigma-Aldrich (St. Louis, MI).

**DNA synthesis.** Designs BB1-4 and 41 designs from parametric design were ordered from Genscript (Piscataway, NJ) as cloned in pET29 with a C-terminal His6 tag. Codon usage was optimized for *E.coli* (Supplementary Table 1 and 2) by the internal algorithms used by Genscript. 56 designs for DFHBI binding and fluorescence activation were ordered from Gen9 (acquired by Ginkgo, Boston, MA) as cloned in pET28b with a C-terminal His6 tag (Supplementary Table 4). Codon usage was optimized for *E.coli* using DNAworks[72]. 36 designs for b11 loop insertion were ordered from Genscript as cloned between the NdeI and XhoI restriction sites of pETCON2[73] for yeast display (Supplementary Table 5). Genes encoding b11L5F and five designs based b11L5F.1 (nC1-5) were ordered from Integrated DNA Technologies(Coralville, IA) as gblock fragments and cloned into pET15 with N-terminal His6 tag followed by thrombin cleavage sequence, with codon usage optimized for *E.coli* using DNAworks. All the DNA oligos and primers used in this work were ordered from Integrated DNA Technologies(Skokie, IL).

**Yeast surface display assays.** 36 designs for b11 loop insertion were tested for binding using yeast surface display using the protocol presented in[46]. EBY100 cells transformed with designs were inoculated into 1mL SDCAA media and grown at 30℃ overnight. 1e7 yeast cells from overnight SDCAA culture were collected by centrifugation at 8,000rpm for 2min and resuspended in 1mL SGCAA media to induce surface protein display. After 24 hrs induction at 22℃ shaker, 5e7 cells were collected by centrifugation and washed twice by PBSF. Yeast surface protein display level was monitored by incubating the cells with FITC-conjugated anti-cMyc antibody for 10min at room temperature. DFHBI-binding and fluorescence signal was assessed by directly labeling the cells with DFHBI for 20-30min at room temperature. Cells labeled with FITC-conjugated anti-cMyc antibody were washed once by 100uL PBSF before reading the signal on a flow cytometer (Accuri C6, DB). DFHBI-labeled cells were analyzed by the flow cytometer without washing. For both labels, a 488nm laser was used for excitation and a 520nm band pass filter for emission (Extended Data Fig. 8f).

**b11L5F deep mutational scanning.** b11L5F deep mutational scanning library was constructed by site-directed PCR mutagenesis using the protocol described in[44]. DNA oligos were ordered with degenerate NNK at the each targeted position for introducing 20 codon variations, together a reverse complementary DNA oligo cover the flanking region (Supplementary Table 8). For each position in b11L5F, three PCR were carried out sequentially to construct the mutagenized full-length gene. Final PCR products of 111 positions were verified on 1% agarose gel to confirm the right length, then pooled together for yeast transformation by electroporation[46]. pETCON2 plasmids were digested by *XhoI* and *NdeI* restriction enzymes (NEB) and gel purified using a commercial kit (Qiagen). Pooled library DNA mixed with cut pETCON2 vector were concentrated to a volume of < 10μL using a centrifugal vacuum concentrator (Savant SpeedVac). Two libraries were transformed independently with different amount of DNA: 1μg vector and 4μg insert genes for the first library; 2μg and 6μg for the second. Number of transformants were estimated by plating a small fraction of transformed libraries after serial dilutions.

Both libraries had more than 2e7 survived clones, which was far above the theoretical size of the library (110*19+1= 2,091). Transformed libraries were passaged into 250mL of C-Ura-Trp medium and induced by SGCAA at 22℃ for > 24hrs according to[46].
Transformed naive libraries underwent a series of selections by fluorescence activated cell sorting (FACS). Protein stability selection was done by sorting out variants with protease resistance as described in[30]. Functional selection was conveniently carried out by selecting fluorescence-activating cells after DFHBI incubation. Labeling conditions and sorting parameters were given in Supplementary Table 16. 2 naive libraries and 20 selected libraries (see Supplementary Table 8 for a complete list) were sequenced by a MiSeq sequencer using a 300-cycle reagent kit (Illumina, CA). Sample preparation was done using the same protocol described in[47].

**Sequencing data analysis.** Pair-end reads from MiSeq sequencer were first combined using the PEAR program[48]. The counts analysis was done using scripts adapted from Enrich[49]. 20 sets of sequencing data (2 naive libraries, 16 libraries treated with trypsin or chymotrypsin at 4 different concentrations and 2 control selections without protease treatment) for stability selection were analyzed using the methods and scripts developed in[30], where the unfolded states were modeled without disulfide bonds (cysteine were replaced by serine). After subtracting wild type sequence stability score, relative scores in trypsin dataset were in the range of -2 to +2 while the range in chymotrypsin dataset were from -3 to +0.5. Since trypsin dataset has a better dynamic range after backbone subtraction, all the individual mutational effects on protein stability were using trypsin proteolysis data. The average effect (used to color the models in Fig. 4b and Extended Data Fig. 7b) of all the amino acid substitutions on each position were using the average stability scores from trypsin and chymotrypsin treatments. 4 sets of sequencing data for functional selections were analyzed using the statistical methods described in[50]. Python scripts written to implement the Enrich2 methods and data visualization scripts were provided in the Supplementary Data as a Python interactive notebook (b11L5F_DMS_analysis.ipynb).

**Combinatorial library construction and selection.** Combinatorial libraries for b11L5F.1 and b11L5F.2 were designed to include all the beneficial mutations and their similar amino acid substitutions from the deep mutational scanning results. 20 positions with 3 to 5 mutations were selected. To control the library size (which is limited by the maximum yeast transformants, usually 1e8), 8 positions (A3, V13, M15, M27, F29, L37, Q42, and L107) were doped with a low percentage of alternative variants (1-2%) by requesting non-catalog synthesis from Integrated DNA Technologies (IDT). By lowering the mutagenesis rate of those positions, the theoretical sizes of the libraries are around 2e6, 99% of which contain additional 0 to 3 mutations at doped positions. Full-length DNA libraries were assembled using a recursive PCR protocol[45]. Oligos received from IDT were dissolved in diH2O to 100μM stock concentrations and further diluted to 8.3μM and 12.5μM for assembling b11L5F.1 and b11L5F.2 libraries respectively, followed by another PCR using the flanking primers to amplify the full-length genes. Assembled genes were purified by gel extraction (Qiagen) and further amplified to obtain enough amount of DNA for yeast transformation. Two electroporations were performed for each library with varied amount of DNA: 3μg vector and 10μg insert genes for the first library; 4μg and 12μg for the second. Transformed yeast cells for the same library were pooled together for growth and display induction as described above. Number of transformants were estimated the same way: 8.8e7 for b11L5F.1 library and

5.5e7 for b11L5F.2 library. Given the complexity of the doped amino acid substitutions, the theoretical size the library was determined by the number of transformants and the FACS experiments were carried out to analyze at least 2 fold of that number in the initial selections. In total, 5 and 6 rounds of sorting and cell growth were carried out for b11L5F.1 and b11L5F.2 libraries, respectively. Detailed sorting parameters and statistics were provided in Supplementary Table 16.

Cells after sort 5 and sort 6 were plated on selective agar plates (C-Ura-Trp) for yeast colony PCR and Sanger sequencing. b11L5F.1 library converged to two sequences named as mFAP0 and mFAP1; b11L5F.2 library converged to one sequence after sort 5, named as mFAP2. Genes encoding mFAP0-2 were subcloned into E.coli vector pET15b for protein purification and biochemical and structural characterizations (Extended Data Fig. 8g&h).

**Error-prone library construction and selection.** Besides the combinatorial libraries described above, an error-prone library was constructed by amplifying the gene with Mutazyme II (Agilent). To achieve the desired average mutation rate of 2-3 mutations per gene, the reaction was carried out with 25ng, 10ng, 1ng, 0.1ng and 0.01ng of DNA template b11L5F.1 in pETCON2. The PCR products obtained with 1ng and 0.1ng of template were mixed together in equal amounts to build the library. The first reaction (with 1ng of template DNA) produced genes with 1 to 4 mutations, with an average mutation rate of 2 mutations per genes and 25% of WT. The second reaction (0.1ng of template DNA) produced genes with 1 to 6 mutations, with an average of 3 mutations per gene and 7% WT. The mutational bias in the library was similar to the mutational spectrum described by the manufacturer of the polymerase.

Additional PCR amplification was carried out to obtain enough amount of DNA for yeast transformation. Two electroporations were performed with varied amount of DNA: 3µg vector plus 10µg insert genes, and 4µg plus 12µg. Cells were pooled together for growth and display induction as described above. Number of transformants were estimated to be 6.1e7. Five rounds of cell sorting and growth were carried out and the final selected cells were Sanger sequenced. Three of individual clones were tested for binding and confirmed to improve the fluorescence activation (Supplementary Data).

**Protein sample preparation for crystallography.** BB1 was expressed in *E.coli* BL21(NEB) with a C-terminal His6 tag, purified by gravity flow over Ni-NTA resin (Qiagen, Germany), followed by size-exclusion chromatography in a Akta Pure FPLC machine (GE Healthcare) using a Superdex 75 increase 10/300 GL column (GE Healthcare).

b10 was subcloned into pCDB24 with a N-terminal His8 tag followed by SUMO domain that could be recognized and cleaved by sumo protease 2[74]. After overnight 18℃ *E.coli* expression with IPTG induction, b10 was first purified by gravity flow over Ni-NTA resin, then incubated with homemade sumo protease overnight at 4℃ to cleave the tag. Cleavage reaction was confirmed by SDS-Page gel and purified by Ni-NTA resin where the unbound flow-through was collected and concentrated for size-exclusion purification.

mFAP0 and mFAP1 were subcloned into pET15 to have a N-terminal His6 tag followed by thrombin cleavable sequence. After overnight 18℃ *E.coli* expression with IPTG induction, mFAP0 and mFAP1 were first purified by Ni-NTA gravity flow, then incubated with thrombin overnight at 4℃ to cleave the His6 tag. Cleavage reaction was treated the same way as described above.

**Confocal image acquisition.** Mammalian cell imaging of mFAP1 and mFAP2 was performed in NIH3T3 cells (Flp-In-3T3, Thermo Fisher Scientific, Inc). NIH3T3 cells were cultured in high-glucose DMEM, 4 mM L-glutamine, 10% fetal bovine serum (FBS, Life Technologies, Inc) at 37 ℃, 5% $CO_2$. Cells were plated at $4x10^4$ cells/mL in 35 mm glass-bottomed dishes (Matek, Inc) that were coated with poly-D-lysine. Cells were transfected 24 hours after plating with Lipofectamine 3000 (Thermo Fisher Scientific, Inc) at a ratio of 3 μL reagent: 1 μL DNA, according to manufacturer's instructions, with 1.25 μg pCDNA5 plasmids of mFAPs or mFAP fusions (1.25 μg mCherry plasmid was added to the cytosolic constructs as a transfection control). Right before imaging, cell media was replaced with FluorBrite DMEM (Thermo Fisher Scientific, Inc) media supplemented with GlutaMax (Thermo Fisher Scientific, Inc) and 10% v/v FBS, and 20 μM DFHBI. Cells were imaged on a heated stage (37 ℃). A Leica SP8X system was used for confocal microscopy. A white light laser of 488 was used to excite the DFHBI and detected by a HyD detector, over a range of 495-550 nm. All images were taken using a 63x objective with oil, at 1024x1024 resolution.

**E.coli and yeast sample preparation for confocal imaging.** 500uL *E.coli* Lemo21 cells (NEB) expressing mFAP1 and mFAP2 were washed three times by 1mL M9 medium. After incubated with 20μM DFHBI for 30 min at room temperature, E.coli cells were transferred to a 1mm-thinn 1% agarose gel on a glass slide and waited for 10min for immobilization. Cells were imaged between the glass cover slide and the agarose gel. 5e6 yeast EBY100 cells[46] displaying mFAP1 and mFAP2 were washed once by 1mL PBSF buffer and incubated with 20μM DFHBI for 30 min at room temperature. Yeast cells were transferred to a 35mm glass-bottomed dish (Matek, Inc) and waited for 5min at room temperature for cells to settle down to the bottom of the dish before imaging. Imaging of *E. coli* and yeast expressing mFAPs, or Aga2p-mFAPs (respectively) was performed on the same microscopy described above without the heating stage.

**Extinction coefficients.** Absorbance spectra were measured using a Thermo Scientific BioMate 3S UV-vis spectrophotometer (1 nm interval, 800 nm/min).

*DFHBI-mFAPs complexes.* A solution was prepared containing 1 μM DFHBI and 10 μM of mFAP1 or mFAP2 in PBS (pH 7.4) and allowed to equilibrate for at least 30 min, producing ~1 μM of mFAP1-DFHBI or mFAP2-DFHBI. The absorbance was measured for each solution, and the extinction coefficients were calculated using Beer's Law (eq. 7)

$$A = \varepsilon bc \qquad \text{eq. 7}$$

where $A$ is peak absorbance, $\varepsilon$ is extinction coefficient, $b$ is path length (1 cm), and $c$ is concentration (1 μM).

*Uncomplexed DFHBI.* The absorbance spectrum of DFHBI in PBS (pH 7.4) was measured for several stock solutions over the range 1-14 μM. The peak absorbance of each was plotted vs concentration and fitted to a line whose slope, when divided by path length 1 cm, is the reported extinction coefficient (Extended Data Fig. 10d).

**Fluorescence quantum yield.** A Perkin-Elmer LS-B Luminescence Spectrophotometer (10 nm bandwidth, 1 nm interval, 100 nm/min) was used for relative quantum yield measurements. A Hamamatsu C9920-12 integrating sphere instrument was used for absolute quantum yield measurements (6 nm excitation bandwidth, 1 nm interval).

*Relative quantum yield.* The fluorescence emission spectra of complexes mFAP1-DFHBI and mFAP2-DFHBI (in PBS, pH 7.4) and reference dye Acridine Yellow G (in methanol) were measured. The quantum yield was calculated using the equation[75](eq. 8)

$$\Phi_c = \Phi_r \times \frac{1-10^{A_r(\lambda_{ex})}}{1-10^{A_c(\lambda_{ex})}} \times \frac{\int F_c(\lambda)d\lambda}{\int F_r(\lambda)d\lambda} \times \frac{n_c^2}{n_r^2} \quad \text{eq. 8}$$

where  is quantum yield, $A(\lambda_{ex})$ is absorbance at the excitation wavelength $\lambda_{ex}$ ($\lambda_{ex}$=440nm), $F$ is fluorescence emission, $n$ is refractive index of the solution (1.335 for PBS and 1.3284 for methanol), and the subscripts 'c' and 'r' refer to the mFAP1-DFHBI or mFAP2-DFHBI complexes and the reference dye, respectively. A reference quantum yield value =0.57 was used for Acridine Yellow G (in methanol)[76].

*Absolute quantum yield.* Absolute quantum yields were measured for solutions of 1 μM DFHBI and 10 μM of mFAP1 or mFAP2 in PBS (pH 7.4) that were allowed to equilibrate for at least 30 min, producing ~1 μM of mFAP1-DFHBI or mFAP2-DFHBI. Samples were excited at 440 nm and absolute quantum yields were calculated according to the equation (eq. 9)

$$\Phi_c = \frac{f_{em}}{f_{abs}} \quad \text{eq. 9}$$

where  is the emitted photon flux and is the absorbed photon flux. A similar procedure was used to measure absolute quantum yields of two control samples (Acridine Yellow G and fluorescein) which agreed well with literature values[76,77].

**Dissociation constant.** Dissociation constants ($K_D$) for b32, b11, b11L5F, mFAP1 and mFAP2 binding DFHBI were determined by measuring the fluorescence intensity of the complex in a 96-well plate (Corning 3650) on a Synergy neo2 plate reader (BioTek, Inc). Binding reactions were performed at 200μL total volume in PBS pH7.4 buffer. For low-affinity b32, b11 and b11L5F, protein concentration was kept at 10μM with varied concentrations of DFHBI in the range of 0.5μM to 200μM. Background fluorescence from DFHBI alone at the same concentration was subtracted from each observed binding signal. For high-affinity DFHBI binders, mFAP1 and mFAP2, DFHBI concentration was kept at 0.25μM with varied concentrations of mFAP1 or mFAP2 in the range of 0.025μM to 5μM. Data were fitted to an equilibrium binding model (eq. 10) by nonlinear regression analysis in R[78]. Fitting curves were available on https://dx.doi.org/10.5281/zenodo.1216229.

$$F_{observed} = F_{free} + \frac{F_{bound} - F_{free}}{L_{total}} \times \frac{(K_d + P_{total} + L_{total}) - \sqrt{(K_d + P_{total} + L_{total})^2 - 4 \times P_{total} \times L_{total}}}{2} \quad \text{eq.10}$$

In eq. 10, $F_{observed}$ is the fluorescence signal measured by plate reader; $F_{free}$ is the fluorescence intensity from free ligand in the absence of protein; $F_{bound}$ is the fluorescence intensity from the bound ligand when binding reaction reaches saturation; $P_{total}$ and $L_{total}$ are the total concentrations of protein and ligand in the reaction; $K_d$ is the dissociation constant. By fitting the titration data iteratively, $F_{observed}$, $F_{bound}$ and $K_d$ were derived from the model.

# Supplementary Tables

**Supplementary Table 1: List of sequences for nonfunctional β-barrel designs from fragment-based approach.** Protein sequences and DNA encoding sequences (optimized for *E.coli* codon usage) of four designs (BB1-4) are provided in this table.

| Design ID | Protein sequence | DNA Sequence |
|---|---|---|
| BB1 | MVDAAQYFPGTWEFRFRSSDGK<br>EYRGTVEMQPRTPTEIEIRFKGQ<br>SSDGRPVEGRGSIEVRSPYEYRF<br>EMQSSDGARWEGTLQVRSPDSV<br>EVRFKSSDGREYSGEFRRQEGS | ATGGTTGACGCTGCTCAGTACTTCCCGGGTACCTGGGAGTTCAGGTTCCGTTCTTCT<br>GACGGTAAAGAATACCGTGGTACCGTTGAAATGCAGCCGCGTACCCCGACCGAAA<br>TCGAAATCCGTTTCAAAGGTCAGTCTTCTGACGGTCGTCCGGTTGAAGGTCGTGGT<br>TCTATCGAAGTTCGTTCTCCGTACGAATACCGTTTCGAAATGCAGTCTTCTGACGGT<br>GCTCGTTGGGAAGGTACCCTGCAGGTTCGTTCTCCGGACTCTGTTGAGGTTAGGTTC<br>AAATCTTCTGACGGTCGTGAATACTCTGGTGAATTCCGTCGTCAGGAAGGTTCT |
| BB2 | MTTAAWKYPGEWDFEFKSSDG<br>KEYRGKVRIQPETPTKIEFELEGQ<br>SSDGKPFKGHGYFEVKSPTRMR<br>LEFTSKDGRRFEGEVEEKSPHEV<br>EIRFKESDGREYRGKMRRREGS | ATGACCACCGCTGCGTGGAAATACCCAGGCGAGTGGGACTTCGAGTTCAAGTCTTC<br>TATGGTAAAGAGTACCGTGGCAAAGTTCGTATCCAGCCGGAAACCCCGACCAAAA<br>TCGAGTTGAACTGGAAGGTCAGTCCTCTGACGGTAAACCGTTCAAAGGTCACGGTT<br>ACTTTGAAGGAAGTCTCCGACCCGTATGCGTCTGGAATTCACCTCTAAAGATGGCC<br>GTCGTTTCGAAGTGAAGTTGAAGAAAAATCTCCGCACGAGGTTGAGATCCGTTTCA<br>AAGAATCCGACGGTGCGAATATCGTGGTAAGATGCGTCGTCGTGAGGGTTCT |
| BB3 | MKTISQAFPGTWRFDVTSSDGSR<br>WEGRVEVRPRTPTEFEVRFEGKS<br>SDGRPFHGRGEVHVETPDKVEV<br>RFRSSDGREYRGYMELKSPTELE<br>FRFRSSDGKEFRGRLERRGGS | ATGAAAACCATCTCTCAGGCGTTCCCGGGTACCTGGCGTTTTGACGTTACCTCTTCT<br>GATGGTTCTCGTTGGGAAGGTCGTGTTGAAGTTCGTCCGCGTACCCCAACCGAGTT<br>CGAGTGCGCTTCGAGGGTAAATCTTCCGATGGTCGTCCGTTCCACGGTCGCGGCGA<br>AGTTCACGTTGAAACCCCGGACAAAGTGGAAGTGCGTTTTCGTTCCTCTGACGGTC<br>GTGAATACCGTGGTTACATGGAACTGAAATCTCCGACCGAACTGGAGTTCCGTTTC<br>CGCTCCTCCGACGGCAAAGAATTTCGTGGTCGCCTGGAACGTCGTGGTGGCTCT |
| BB4 | METASKALPGEWRFDVKSSDGR<br>RWEGRIEVRPKTPTRFEVRFEGK<br>ESDGRPFHGHGEMRVRSPTKVE<br>VRFKSEDGREFRGSLTLRSPYEM<br>EIRFKSSDGKEYRGRLERIGGS | ATGGAAACCGCGTCTAAAGCGCTGCCGGGTGAATGGCGCTTCGACGTTAAGTCCTC<br>CATGGTCGTCGTTGGGAAGGTCGTATCGAGGTTCGTCCGAAAACCCCGACCCGTTT<br>TGAAGCGCTTTGAAGGTAAAGAATCTGATGGCCGTCCGTTCCACGGTCACGGTGAA<br>ATGCGGTTCGTTCTCCAACCAAAGTTGAAGTTCGTTTTAAGTCTGAAGACGGTCGT<br>GAATTCCGTGTTCTCTGACCCTGCGTTCCCCGTACGAAATGGAAATCCGTTTCAAAT<br>CTTCTGACGGCAAAGAATACCGTGGTCGCCTGGAACGTATCGGTGGTTCC |

**Supplementary Table 2: List of tested β-barrel designs from parametric design approach.** Geometric parameters, protein sequences and stage of failure of forty-one designs are summarized in this table.

| Batch | Design | Design description | Sequence | Stage of failure |
|---|---|---|---|---|
| Batch 1 | 1_5_0062 | b=4.35; $r_{tw}$=1.0; $r_a$=1.1; $r_b$=0.95 (P1, S=10) | MKYKVREEIYGNYYYFFVYIGNVVIVLQVYTFGDVTLVNLYVGNYFTT<br>LKVEVYGNVIVVVWGNYYQYFVYVGNYYLYFLEGGS | |
| | 1_10_0184 | b=4.35; $r_{tw}$=1.0; $r_a$=1.1; $r_b$=0.95 (P1, S=10) | MTLKVEEKKHGNVYIFIVHKGNERYYFIVVINGNDVYVEVISGNESRTF<br>RVETHGNLFIVEVGNVYYYFEKKGNFYLLKTEGGS | |
| | 6_6_0064 | b=4.35; $r_{tw}$=1.0; $r_a$=1.1; $r_b$=0.95 (P1, S=10) | MKYKVREEKHGNTYFFYVYDGNKTVLYVVKVHGNEIYVEVYSGNQS<br>RTFEVREHGNVFIVRSGNRYFYFVKKGNFYLLYEEGGS | |

| | | | | |
|---|---|---|---|---|
| | 4_4_0020 | b=4.20; $r_{tw}$=1.0; $r_a$=1.0; $r_b$=1.15 (P6, S=10) | MKILVIKQTFGNVTYLRIYIGNYVWTVRLEVYGNVIRLELVVGNYVFEL ELRIFGNVITVYVYFGNVILEIRVEVYGNYTLWETRGGS | |
| | 7_1_0012 | b=4.20; $r_{tw}$=1.0; $r_a$=1.0; $r_b$=1.15 (P6, S=10) | MKVKLVEQQQGNTIYIEVRRGDQTFTIRVESRGDTYRLELRQGNRTIEL RLERRGDTYHIEVRSGNTRLSLRVERRGNYFLVEERGGS | |
| | 3_2_0065 | b=4.60; $r_{tw}$=1.0; $r_a$=1.05; $r_b$=0.9 (P8, S=8) | MRTELTLERSGNRFRLTVGDVTLEFETRGNTIEVRLQWGNTTITLRVEI RGNYTELELRRGNTTYRLKFESRGDYWRVEVEGGS | |
| | 5_10_0080 | b=4.60; $r_{tw}$=1.0; $r_a$=1.05; $r_b$=0.9 (P8, S=8) | MRQKLTLEIYGDVTILIWGDFYFYFDIFGNVVRVELYYGNFFLTLELRV FGNVTILRVVTGNYVYYFYFEVVGNVTTVYLEGGS | |
| | 10_6_0048 | b=4.60; $r_{tw}$=1.0; $r_a$=1.05; $r_b$=0.9 (P8, S=8) | MRTELTLERSGNRFRLTVGDVTLEFETRGNTIEVRLQWGNTTITLRVEI RGNYTELELRRGNTTYRLKFESRGDYWRVEVEGGS | |
| Batch 2 | P1_charge | (P1, S=10), +15 net charge | MELRVRSRRHGDNRQFYVISGDKRIRLSVRRHGDKVFVRLQSGDKQRR LQVRRHGDKIFLRSGDRISILIRSGDFYIYRRREGS | Insoluble |
| | P1_thr3 | (P1, S=10), +7 net charge | MSLSVQSRQHGNKYIFIVRSGNRTYYFIVSIHGSRQFVQVISGNTSTTLS VRTSGDKFVVQSGSRYYIFQRSGSTYLLFTQEGS | Insoluble |
| | P6_charge | (P6, S=10), +15 net charge | MSLRVIRRRHGDRVYLILVSGDKRLRIQVSRHGDKLSLQLISGDKRFQL RLFIHGDKYQIQVRSGDKRYRISVSRHGDVYSLRRRGGS | Insoluble |
| | P6_thr3 | (P6, S=10), +10 polar charge | MQIILIQSRSGDRFRVSLISGNKQLQLTVSQHGNKYRIQLQSGNKQYQL TLRSHGSKFTITVSSGNRRFTLSVSQHGSTYSVQQSHGS | No expression |
| | P6_diverse | (P6, S=10), loop diversification, +3 net charge | MEITVTVQTHGNKLRWTFTYNGKTFTVTVTQHNGEYIITWQYGDTTW TLRLQRSGDTYTVTVVSGNKRFTVTLQQSGNTYVLRITMGS | No Expression |
| | P8_hydro | (P8, S=8), 2 hydrophobics on surface | MTKTLTLTTRGTTYTLTNGTLTLTITKRGRTLTVSLTAGTTTFTLTITRK GKTLTLTLTSGTTTITFTFTKSGTTYTVLVTSGS | No expression |
| | P8_polar | (P8, S=8), +8 net charge | MTKTLTLTRRGTTFTLTDGTNTFTLTKRGRTYTIKVTSGTTTYTVTLTQ KGKTLTLTWTNGTTTYTYIFTKRGKTYTVLVTGGS | Toxique - could not be transformed |
| | P8_thr3 | (P8, S=8), +8 net charge, down-weight threonine | MKQQLTLRRHGSTYQLRSGDLSFSISQHGNKLSVRLQYGNTTYQLTLS RHGDRLRLTVSHGNKTYIYYFQRSGDTYQVLVQGGS | Insoluble |
| | P12_div | b=4.30; $r_{tw}$=1.0; $r_a$=1.10; $r_b$=1.0 (P12, S=8) | MRTTLTVSTHNNELVLTYNNNTVRISRHGDTWTVLYGNTTLTVTRSGD TYVLTSGDLTIIISQHGDTYRLTVTGGS | Toxique - could not be transformed |
| | P12_charge | b=4.30; $r_{tw}$=1.0; $r_a$=1.10; $r_b$=1.0 (P12, S=8) | MEQTLTLRSHGDKWSVRSGSNTLVVIQHNNRFLVTYNNQTLVLQTHG NRLRLTYGNKTVSIQRHGDTFRLTVTGGS | Insoluble |
| Batch 3 | P6_negative | (P6, S=10), -15 net charge | MKREVEREEHGDKEYLRLRSGDKDLELEVDSHGDKFRLRLESGDKEFE IEWEESGDKFELRVEHGDREDRLEVEESGDKYLLDEQSGS | Poor solubility, Random coil |
| | P6_normal | (P6, S=10), +4 net charge | MKLKVIRTRHGDKEYLILVSGNKKLEIEVQRHGDTLKLRLKSGNTEITL RWREHGSRFELELESGDTRKTLTVEKSGSTYLLDTQCGS | Insoluble |
| | P6_positive | (P6, S=10), +20 net charge | MKLKVIEKRHGDKVFLRLKSGNKDYRIEVRKHGSKLKLRLESGNKRFK LRLERSGNKYRVEVRSGNKRYKIKVERSGNTYLWEKRGS | Poor expression, insoluble |
| | P8_alt_layers | (P8, S=8), 0 net charge | MRQTYELKEHGNTYKLTSGNREFRLERHGNTVRIELKYGNTTYTLTLS VHGNTWELETRSGNTTERFEFEKSGNTFTVREECGS | Insoluble |

| | | | | |
|---|---|---|---|---|
| | P8_negative | (P8, S=8), -22 net charge | MREEYELEEHGDTEELESGDREFRFEEHGDKFRVEVESGDTDYELELER SGDTQELDFESGDTRERFEFEKSGDTYRVEVECGS | Toxic - cell lysis after induction |
| | P8_neg_ ssbond | (P8, S=8), -20 net charge | MCGDEEEYELESHGDTYELESGDKELRFEEHGDKFRVELDYGDTRFEL ELQRHGDTWELDLRSGDREERFEFEESGDRFRVEVYCIEGS | No expression |
| | P8_normal_1 | (P8, S=8), +1 net charge | MTETYTLDRHGNKYRLTSGNKTFTFDQHGNTVRVELKSGNTTYTLTLS THGNTQRLDVQSGNTQETYRFEKHGDTFRVEVERGS | Unstable expression, Insoluble |
| | P8_normal_2 | (P8, S=8), +2 net charge | MPETYTLKRHGSTYKLTSGDFRLRFREHGNTFKVELEYGNTTYTLTLSS HGDTWKLDLKSGNTQLVYLFERHGDKFRVEVYGS | Unstable expression, Insoluble |
| | P8_positive | (P8, S=8), +17 net charge | MRKKYEVKRHGDRYELKSGNLKLEIRRHGNKFKVKLKSGNRTYTLKL QRHGDKWKLELRSGNRRLEFEFVRSGDKFKVYVKGS | No expression |
| | P8_pos_ ssbond | (P8, S=8), +21 net charge | MKGCKRKYTLDRHGDRWRLKSGNFQFEFKKHGDKFKVKVKYGNRK YKFKLKRHGDRFKLKVRSGNREYEYYFYRHGDKFKVVVKCRDGS | No expression |
| | P8_ssbond_ normal | (P8, S=8), -1 net charge | MGCRTETYQLRRHGSKYELTSGDRELRFEEHGDTIRVEVESGNTTYTV TLEKHGNKLELRLTHGNTEFKFEFEKSGDTFKVTVYKCGS | No expression |
| Batch 4 | Cb_short_1 | | MHMPESTLELRVDGKTLTVLVSGDTIRIESGDTEITVTKDPSNNLFRLK VNGQTYRLRQEDKNRRLEVDSNDRKTWEVQEKGSLE | Very low expression |
| | Cb_short_2 | | MHMTERRVELRVDGETWTVKVEDRTGTITITSRNKKTFEIRKSGNTYSL EYNGQQLKVEQEDDNRKFRVKSGNKTWELQEKGSLE | Insoluble |
| | Cb_short_3 | | MHMPEYTLRLEYDGKELTVYWSGDTIEIVSGNKRLEVRKDPSNNVFRL EHNGQELTVKEEKDNKVFRVTYNNRKTLRLQSEEGSLE | Insoluble |
| | Short_idl | | MHMPEYKLELRSGNKTLTVYSSGDKYEIESGNRRYEVRRSGNTFLVKS EGRTLRLEERKGKYQIESEGKTLELQQTSGSLE | No expression |
| | P6_helix | P6, with helical capping of one side of the barrel | MHMKVELRQERHGDTEKIELRSGNNRLEIEVRRHGNKLTLRVKYGNK EIKVEWRDHGSNFSLRIEYGNKRFTVKVEQSGNKYLVEGNSPAQAAKE SGSLE | No expression |
| Batch 5 | Cb1_ff | Cb_short_1 with remodeled turn 7 | MHMPQSTFELRKDGRTLEVREDGDTITIRDGNTSLTVQKDPSSGTFRID KNGKDLELRKDPDSGELRVRDEDGKTWELRKHGSLE | Oligomer |
| | Cb1_ssm | Cb_short_1 with a point mutation in a turn 7 | MHMPESTLELRVDGKTLTVLVSGDTIRIESGDTEITVTKDPSNNLFRLK VNGQTYRLRQEDKNRRLEVDSNDGKTWEVQEKGSLE | Insoluble |
| | Cb1_ff_neg | Cb_short_1 with remodeled turn 7 | MHMPEHEFEVDKDGRELKLEEDGDEIRVEDGDTEIEVRRDPDSGTFRW DVDGRDLEEEEDPDSGERRVRDEDGKTWEVRDRGSLE | No expression |
| | Cb1_ff_pos | | MHMPKRRYELKKDGRRYEVRVKGDEIEIRDGNRRWEVRRDPKTGRYR VRKDGKDWELEKDPKSGRFRIRDSDGKTLRIERRGSLE | No expression |
| | Cb1_ff_ ssbond | | MHMPQCKFELEKDGRTYEVRDDGDEIEIRDGNTNIQVQKDPSSGTYRL DVNGKDLTLEEDPSSGTRRVRDSDGKTLRVETKSCGSLE | No expression |
| | Id_ff_1 | | MHMPRRRFRLDKNGKDITVEYDPSTGVFRIHDGNTELKIERDGNTYYLI KNGKRFEVRQDGDTFYIYEGNETLRLTHDGSLE | No expression |
| | Id_ff_2 | | MHMPTYTLRVHKDGREFTLLKDPSTGTFEIRDGNDQYEIRKDPSTGLY RVHKDGRTYELYEDGNKYVIYKGNEKITVRQEGSLE | No expression |

**Supplementary Table 3:** Experimental characterization of DFHBI-binding fluorescence activating β-barrel designs. Fifty-six β-barrel designs were expressed in *E.coli* and purified by affinity chromatography, followed by SEC and CD measurements. Soluble monomeric designs were tested for DFHBI fluorescence activation. Results are summarized in this table. For protein expression and purification experiments: n=5 biological replicates for b11 and b32; n=3 for other designs; for CD spectra recording: n=1.

*Designs with a disulfide bond have the parent design ID in the parentheses.

†E-value calculated by BLAST against the non-redundant protein database.

| Scaffold ID | RIF docking solution ID | Design ID* | E-value† | *E.coli* Expression | Solubility | SEC | β CD spectrum | Fluorescence Activation |
|---|---|---|---|---|---|---|---|---|
| 13input0059 | 1 | b01 | 0.9100 | yes | yes | oligomers | yes | |
| | | b07 | 0.1600 | yes | yes | oligomers | | no |
| | 18 | b14 | 0.0650 | yes | yes | oligomers | yes | no |
| 3input0012 | 5 | b40 | 1.2000 | yes | yes | oligomers | yes | |
| | | b46 | 4.8000 | yes | no | | | |
| | | b04 | 0.7100 | yes | yes | oligomers | | |
| | | b55 | 0.1400 | yes | yes | **monomer** | | no |
| | | b48 (b55) | | yes | yes | **monomer** | | no |
| 13input0010 | 0 | b31 | 0.4800 | yes | yes | oligomers | yes | no |
| | | b05 (b31) | | yes | no | | | |
| 8input0012 | 7 | b17 | 2.4000 | no | | | | |
| | | b23 | 7.5000 | no | | | | |
| | | b06 (b23) | | yes | yes | **monomer** | | no |
| | | b52 (b23) | | yes | yes | **monomer** | | no |
| | | b29 | 1.8000 | yes | no | | | |
| | | b12 (b29) | | yes | no | | | |
| 37input0094 | 7 | b09 | 0.0090 | yes | no | | | |
| | | b16 | 0.0130 | no | | | | |
| | | b22 | 0.0020 | yes | yes | **monomer** | | no |
| | | b28 | 0.0004 | no | | | | |
| 8input0010 | 13 | b10 | 1.5000 | yes | yes | **monomer** | yes | no |
| | | b56 | 2.8000 | yes | yes | **monomer** | | no |
| 15input0040 | 0 | b08 | 0.3400 | yes | yes | tetramer | yes | |
| | | b18 (b08) | | yes | yes | oligomers | | no |
| | | b24 (b08) | | yes | yes | **monomer** | yes | no |
| | | b51 (b08) | | yes | no | | | |
| | | b15 | 0.7800 | yes | yes | **monomer** | yes | no |
| 14input0065 | 12 | b26 | 0.4100 | yes | yes | **monomer** | yes | no |
| | | b32 | 6.1000 | yes | yes | **monomer** | yes | **YES** |
| | | b38 | 4.5000 | yes | yes | oligomers | | |
| | | b11 (b38) | | yes | yes | **monomer** | yes | **YES** |
| 10input0034 | 3 | b53 | 0.0100 | yes | yes | oligomers | | no |
| | | b47 (b53) | | yes | no | | | |
| | | b13 | 0.1500 | yes | yes | oligomers | yes | |
| 11input0067 | 0 | b19 | 4.6000 | yes | yes | **monomer** | yes | **maybe** |
| | | b25 | 1.3000 | yes | yes | oligomers | yes | |
| 24input0071 | 0 | b21 | 2.6000 | yes | yes | **monomer** | | no |
| | | b27 | 2.0000 | yes | yes | **monomer** | | no |
| | | b33 | 0.3100 | yes | yes | **monomer** | yes | no |
| | | b30 (b33) | | yes | yes | oligomers | | |
| 39input0072 | 9 | b34 | 0.6800 | yes | yes | **monomer** | | no |
| | | b54 (b34) | | yes | yes | **monomer** | | no |
| | | b42 (b34) | | yes | yes | **monomer** | | no |
| 10input0001 | 0 | b35 | 6.7000 | yes | yes | oligomers | | no |
| | 1 | b41 | 0.2400 | yes | no | | | no |
| 28input0026 | 2 | b39 | 0.0040 | yes | yes | oligomers | | |
| | | b36 (b39) | | yes | yes | **monomer** | | no |
| | | b49 | 0.0210 | yes | yes | **monomer** | yes | no |
| | | b45 | 0.0150 | yes | yes | oligomers | | |
| 13input0015 | 7 | b37 | 0.0010 | yes | no | | | |
| | | b43 | 0.2200 | yes | no | | | |
| 9input0068 | 17 | b50 | 0.2900 | yes | yes | **monomer** | | no |
| 15input0019 | 9 | b02 | 0.0400 | no | | | | |
| 14input0074 | 8 | b44 | 0.1400 | yes | no | | | |
| 14input0016 | 4 | b20 | 0.0650 | no | | | | |
| 36input0056 | 4 | b03 | 0.0000 | yes | yes | oligomers | | no |

**Supplementary Table 4: List of sequences for DFHBI-binding designs.** Protein sequences and DNA encoding sequences (optimized for *E.coli* codon usage) of fifty-six designs (HBI_b_01 to _56) are provided in this table.

| Design ID | Protein Sequence | DNA Sequence |
|---|---|---|
| HBI_b_01 | MGKNVAQALPGTWKVDLTQS<br>DGSKYTGQITVKPTTPYTFDIK<br>TRGTVSDGRPLTGKGKVTVKT<br>PTTVDVTMTLSDGSTSTGKMT<br>VDSPTQFKLDVTASDGTKATG<br>TVQRQS | ATGGGCAAAAACGTTGCGCAGGCGCTGCCGGGTACCTGGAAAGTTGACCTGACTCAGTCT<br>GATGGCTCTAAATACACCGGTCAGATCACTGTTAAGCCGACTACCCCGTACACCTTCGACA<br>TCAAAACCCGTGGTACCGTGTCCGACGGTCGTCCGCTGACTGGCAAAGGTAAAGTTACCGT<br>TAAAACCCCGACCACTGTTGACGTGACCATGACCCTCTCTGACGGCTCTACCTCCACCGGT<br>AAAATGACCGTTGACTCTCCGACCCAGTTCAAACTGGACGTTACTGCGTCCGATGGCACCA<br>AAGCGACCGGTACTGTGCAGCGTCAGTCT |
| HBI_b_02 | MGAPVVEFLPGTWQINVTVSD<br>GLQFTGQMHITPRTPETLTVTS<br>RGQVEDGTPYKGQGHLTLTSP<br>TTVKFTAKAEDGADTQGHLTI<br>RTPTQFDVNMTVADGQTATG<br>KLTRHE | ATGGGCGCTCCGGTTGTTGAATTCCTGCCGGGTACCTGGCAGATCAACGTTACCGTTTCTG<br>ACGGTCTGCAGTTCACCGGTCAGATGCACATCACCCCGCGTACCCCGGAAACCCTGACCGT<br>TACCTCTCGTGGTCAGGTTGAAGACGGTACCCCGTACAAAGGTCAGGGTCACCTGACCCTG<br>ACCTCTCCGACCACCGTTAAATTCACCGCTAAAGCTGAAGACGGTGCTGACACCCAGGGTC<br>ACCTGACCATCCGTACCCCGACCCAGTTCGACGTTAACATGACCGTTGCTGACGGTCAGAC<br>CGCTACCGGTAAACTGACCCGTCACGAA |
| HBI_b_03 | MGANMKDLAPGTWTWTLTQ<br>EDGLTVQGQTDVQPRTPTTFD<br>LRSHGQTADGTPYHGNGQLH<br>VRSPDQVDVTARVKDGRDAT<br>GTTTMKTPTTLDVTMTVGDG<br>VTSQGKVTRTE | ATGGGCGCGAACATGAAAGACCTGGCTCCGGGTACTTGGACCTGGACTCTGACTCAGGAA<br>GACGGCCTCACTGTTCAGGGTCAAACCGACGTTCAGCCGCGTACCCCAACCACCTTCGACC<br>TGCGTTCTCACGGCCAGACCGCGGACGGTACCCCGTACCACGGTAACGGTCAGCTGCACGT<br>TCGTTCTCCGGACCAGGTTGACGTGACCGCGCGTGTTAAAGACGGTCGTGACGCGACCGGT<br>ACCACCACCATGAAAACCCCGACTACCCTGGACGTTACCATGACTGTTGGTGACGGTGTTA<br>CCTCTCAGGGTAAAGTTACCCGTACTGAA |
| HBI_b_04 | MGKDAASVLPGKWKFNTTAE<br>DGVTITGTITMQPRTPTTFDVT<br>LKGHQSDGRPTKGNGQITVKT<br>PDTVDSRFTLSDGRTFQGKLQ<br>LDSPDTLTINWTMQDGTTQTG<br>HVTRQE | ATGGGCAAAGACGCGGCGTCTGTTCTGCCGGGTAAATGGAAATTCAACACCACCGCGGAA<br>GACGGTGTTACCATCACTGGTACTATCACTATGCAGCCGCGTACCCCGACCACCTTCGACG<br>TGACCCTGAAAGGTCACCAGTCCGACGGTCGTCCGACCAAAGGTAACGGTCAGATCACCG<br>TTAAAACCCCGGACACTGTTGACTCTCGTTTCACCCTGTCTGATGGCCGTACCTTCCAGGGT<br>AAACTGCAGCTGGACTCTCCGGATACCCTGACCATCAACTGGACCATGCAGGACGGTACC<br>ACCCAGACCGGCCACGTTACCCGCCAGGAA |
| HBI_b_05 | MGPCAKTVLPGKWDLNFTSS<br>DGTTFTGKMTVQPKTPDTVDV<br>TIKGKQSDGNPTNGQGQLHVE<br>SCTTFTWDVTYADGKTFKGKT<br>QLKTPTTLQVDVRAADGSKST<br>GYLTRKD | ATGGGCCCGTGCGCGAAAACCGTTCTGCCGGGTAAATGGGACCTGAACTTTACCTCCTCTG<br>ACGGTACCACCTTCACCGGTAAAATGACTGTTCAGCCGAAGACCCCGGACACCGTTGACGT<br>GACCATTAAAGGCAAACAGTCCGACGGCAACCCGACTAACGGTCAGGGTCAGCTGCACGT<br>TGAATCTTGCACCACTTTCACCTGGGACGTTACCTACGCGGATGGTAAGACTTTCAAAGGT<br>AAAACCCAGCTGAAAACCCCGACCACCCTGCAAGTGGACGTGCGTGCGGCGGACGGCTCT<br>AAATCTACCGGTTACCTGACCCGTAAAGAC |
| HBI_b_06 | MGCLGWTVLPGTWKFTVTWS<br>DGQTSTGQVHFQPRTPTTLQV<br>HFRGRSSDGRPFNGKGHVTCK<br>TPTTFDVNVTQSDGATSTGKIT<br>MKSPTTIDVTFTIEDGQTATGQ<br>MHRQS | ATGGGCTGCCTGGGTTGGACCGTTCTGCCGGGTACCTGGAAATTCACCGTTACCTGGTCTG<br>ATGGTCAGACCTCTACTGGTCAGGTGCACTTTCAGCCGCGTACTCCGACTACTCTGCAAGT<br>TCACTTTCGTGGTCGTTCTTCCGACGGTCGTCCGTTCAACGGTAAAGGCCACGTGACCTGC<br>AAAACCCCAACTACCTTTGACGTTAATGTTACCCAGTCCGATGGCGCGACCTCCACCGGCA<br>AAATCACCATGAAATCTCCGACCACTATCGACGTTACTTTCACCATCGAAGACGGTCAAAC<br>CGCGACCGGCCAGATGCACCGCCAATCT |
| HBI_b_07 | MGQQVAQALPGTWKFDLTQS<br>DGSKYTGQITIKPETPTTLTVK<br>TKGTVSDGRPLTGKGKVTMK<br>TPETMDVTMTLSDGSTSTGKM<br>RLRSPDTFDLDVTASDGTKAK<br>GQVHRQS | ATGGGCCAGCAGGTTGCGCAGGCGCTGCCGGGTACCTGGAAATTTGACCTGACTCAGTCC<br>GACGGTTCTAAATACACCGGTCAGATCACCATCAAACCGGAGACCCCGACCACCCTGACC<br>GTTAAAACCAAAGGTACCGTTTCTGACGGTCGTCCGCTGACTGGTAAGGGTAAAGTTACCA<br>TGAAAACCCCGGAAACTATGGACGTGACCATGACCCTGTCTGATGGCTCTACCTCCACCGG<br>TAAGATGCGTCTGCGTTCTCCGGACACCTTCGACCTGGACGTTACCGCTTCCGATGGTACT<br>AAAGCGAAAGGTCAGGTTCACCGTCAATCT |
| HBI_b_08 | MGTAAVQFLPGTWKFDVTAE<br>DGSQFTGKVTVQPDSPDTVKI<br>TFNGTQSDGKPATGQGTLTMT<br>SPETVKITVTYSDGKTFTGYVT<br>LRTPTQFQVDATANDGTKSTG<br>YMRRTE | ATGGGCACTGCGGCGGTTCAGTTCCTGCCGGGTACTTGGAAATTCGACGTTACCGCGGAAG<br>ACGGTTCTCAGTTCACTGGTAAAGTTACTGTTCAGCCGGACTCTCCAGACACCGTGAAAAT<br>CACCTTCAACGGTACCCAGTCTGATGGCAAACCGGCGACCGGCCAGGGTACTCTGACCAT<br>GACCTCTCCGGAAACCGTTAAGATTACCGTTACCTACTCCGACGGCAAAACCTTCACCGGC<br>TATGTTACCCTGCGTACCCCGACCCAGTTCCAGGTTGACGCGACCGCGAACGACGGTACTA<br>AATCTACCGGTTATATGCGTCGTACCGAA |

| | | |
|---|---|---|
| HBI_b_09 | MGRAAVQFLPGTWKMTSHYE<br>DGTQMQGHVHVQPRSPDTVD<br>VTVTGKASDGKPMQGQGKIT<br>VDSPDQVQWHLTSSDGTQAK<br>GSSQIDSPTQLKLDLTASDGTR<br>LTGTFQRTS | ATGGGCCGTGCGGCGGTGCAGTTCCTGCCGGGTACCTGGAAAATGACCTCCCACTATGAA<br>GACGGTACCCAGATGCAAGGTCATGTTCACGTTCAGCCGCGTTCCCCGGACACCGTTGACG<br>TTACCGTTACTGGTAAAGCGTCCGACGGTAAACCGATGCAGGGCCAGGGTAAAATCACCG<br>TGGATTCTCCGGACCAGGTTCAGTGGCACCTGACTTCTTCTGATGGTACTCAGGCGAAAGG<br>TTCTTCTCAGATCGACTCTCCGACCCAGCTGAAACTGGACCTGACCGCGTCTGACGGCACC<br>CGCCTGACTGGTACCTTCCAGCGTACCTCT |
| HBI_b_10 | MGSALAQQLPGTWKMDVTSE<br>DGVRTTGQMHIQPKTPTTMDV<br>TLTGTHADGKPFTGQGKITVK<br>TPTTVDITVTYEDGSTATGQLT<br>VDSPTQFKFDMTASDGTRFTG<br>TVQRQS | ATGGGCTCTGCGCTGGCGCAGCAACTGCCGGGTACCTGGAAAATGGACGTGACCTCTGAA<br>GACGGTGTTCGTACCACCGGCCAGATGCACATCCAGCCGAAAACCCCGACCACTATGGAT<br>GTTACCCTGACCGGTACCCATGCGGACGGTAAACCGTTTACTGGTCAAGGTAAAATCACCG<br>TTAAGACTCCGACTACCGTTGACATTACCGTGACTTACGAGGACGGTTCTACCGCGACTGG<br>CCAACTGACTGTTGACTCTCCGACCCAGTTCAAATTCGACATGACCGCGTCTGACGGTACT<br>CGTTTCACCGGCACCGTGCAACGTCAGTCT |
| HBI_b_11 | MGCRAASLLPGTWQVTMTNE<br>DGQTSQGQMHFQPRSPYTLDV<br>KAQGTMSDGRPIQGKGKVTC<br>KTPDTMDVDITYSDGKQVQG<br>QVTLDSPTQFKFDVTTSDGSK<br>VTGTLQRQE | ATGGGCTGCCGTGCGGCGTCTCTGCTGCCGGGTACCTGGCAAGTTACTATGACCAACGAAG<br>ACGGTCAGACCTCTCAGGGTCAGATGCACTTCCAGCCGCGTTCTCCGTACACCCTCGACGT<br>GAAAGCGCAGGGTACTATGTCCGACGGTCGTCCGATTCAAGGCAAAGGTAAAGTGACCTG<br>CAAAACCCCGGACACTATGGATGTTGACATCACCTACTCTGACGGCAAACAGGTTCAGGG<br>CCAGGTTACCCTGGACTCTCCGACCCAGTTTAAGTTTGACGTTACCACCTCTGATGGTTCTA<br>AAGTTACCGGCACCCTGCAACGTCAGGAA |
| HBI_b_12 | MGPCNVQVLPGTWQFQVTFS<br>DGQTSRGHVHVQPRTPTTVQV<br>HFTGRSSDGRPFNGKGHMTCK<br>TPTTVDVNVTQSDGATSTGKF<br>TMKSPTTLDVRFTIEDGQTAQ<br>GQLHRQS | ATGGGCCCGTGCAACGTGCAGGTTCTGCCGGGTACCTGGCAGTTCCAGGTTACCTTCTCTG<br>ATGGTCAGACTTCTCGTGGTCACGTTCATGTTCAACCGCGTACCCCGACCACTGTTCAGGT<br>GCATTTCACCGGTCGTTCTTCTGACGGTCGCCCGTTCAACGGCAAAGGTCACATGACCTGC<br>AAAACCCCAACCACCGTTGACGTTAATGTGACTCAGTCCGACGGTGCGACCTCCACCGGTA<br>AATTCACCATGAAATCTCCGACTACCCTGGACGTTCGTTTTACCATCGAAGACGGTCAAAC<br>CGCGCAGGGTCAGCTGCACCGCCAATCT |
| HBI_b_13 | MGGDLSKVVPGTWDLDATNE<br>DGATIKGQIDIQPRTPDKFQLT<br>SRGQYSDGKPMKGTGSFKLDT<br>PTTVTVTFHLSDGRTIQGKLTV<br>KTPTTLDINVTASDGSTSTGQV<br>TRRE | ATGGGCGGTGACCTGTCTAAAGTTGTTCCGGGTACCTGGGACCTGGACGCGACCAACGAA<br>GACGGTGCGACCATCAAAGGTCAGATTGACATCCAGCCGCGTACCCCGGACAAATTCCAG<br>CTGACCTCTCGTGGTCAATATTCTGACGGCAAACCGATGAAAGGTACCGGTTCTTTCAAAC<br>TGGACACCCCAACCACCGTTACTGTTACCTTCCACCTGTCCGACGGTCGTACCATCCAGGG<br>TAAACTGACCGTGAAAACCCCGACCACCCTCGACATCAACGTTACCGCGTCTGATGGCTCT<br>ACCTCTACCGGTCAGGTTACCCGTCGTGAA |
| HBI_b_14 | MGAEVAQALPGTWQMDLTQS<br>DGSQAKGRFTVKPTTPTTFKL<br>TYKGTISDGRPTNGQGTMTVR<br>SPDTVDLKMTLSDGATIQGKL<br>TIDSPTQLKVDLTMSDGTRAT<br>GTVQRQS | ATGGGCGCGGAAGTTGCGCAGGCGCTGCCGGGCACCTGGCAGATGGACCTGACTCAGTCT<br>GATGGTTCTCAGGCGAAAGGTCGTTTCACCGTTAAACCGACCACCCCGACTACCTTCAAAC<br>TGACTTACAAAGGTACCATCTCCGATGGCCGTCCGACCAACGGCCAGGGCACCATGACCG<br>TTCGTTCTCCGGACACCGTTGATCTGAAAATGACCCTCTCTGACGGTGCGACTATCCAGGG<br>TAAGCTCACCATCGATTCTCCGACCCAGCTGAAAGTTGACCTCACTATGTCCGACGGCACC<br>CGTGCGACCGGCACCGTGCAGCGTCAGTCT |
| HBI_b_15 | MGTAAVQFLPGTWKFDVTAE<br>DGSQFKGKVHIQPDSPDTVKV<br>TFNGTQSDGKPATGQGTLTMT<br>SPETVKLQVTYEDGKTFTGYM<br>TLRTPTQFQLDAKANDGTKST<br>GYMRRTE | ATGGGCACCGCGGCTGTTCAGTTCCTGCCGGGTACCTGGAAATTCGATGTTACCGCGGAGG<br>ACGGTTCTCAGTTCAAAGGTAAAGTTCACATCCAGCCGGACTCCCCGGATACTGTTAAAGT<br>TACCTTCAACGGTACCCAGTCTGATGGTAAACCGGCGACCGGCCAGGGTACTCTGACCATG<br>ACCTCTCCGGAAACCGTTAAACTGCAGGTTACCTACGAAGACGGTAAAACCTTTACCGGTT<br>ACATGACCCTGCGTACCCCGACCCAGTTTCAGCTGGACGCGAAAGCGAACGATGGCACCA<br>AATCCACTGGTTATATGCGTCGTACCGAA |
| HBI_b_16 | MGQAAVQLMPGTWDITSHYE<br>DGQTMQGKVHVKPRSPDTVDI<br>TVTGQASDGKPMQGQGQLTM<br>KSPHQVQVRLTSSDGTQAQGT<br>VTMESPTRFRWDLTASDGVRL<br>TGTTQRTS | ATGGGCCAGGCTGCTGTTCAGCTGATGCCGGGTACCTGGGACATCACCTCTCACTACGAAG<br>ACGGTCAGACCATGCAGGGTAAAGTTCACGTTAAACCGCGTTCTCCGGACACCGTTGACAT<br>CACCGTTACCGGTCAGGCTTCTGACGGTAAACCGATGCAGGGTCAGGGTCAGCTGACCAT<br>GAAATCTCCGCACCAGGTTCAGGTTCGTCTGACCTCTTCTGACGGTACCCAGGCTCAGGGT<br>ACCGTTACTATGGAATCTCCGACCCGTTTCCGTTGGGACCTGACCGCTTCTGACGGTGTTCG<br>TCTGACCGGTACCACCCAGCGTACCTCT |
| HBI_b_17 | MGPAAVQVLPGTWKFTVTWS<br>DGQTSTGQVHVQPRTPTTVQV<br>HFRGRSSDGRPFNGKGHLTMK<br>TPTTLDVNVTQSDGATSTGKF<br>TMKSPTTIDVTFTIEDGQTATG<br>QMHRQS | ATGGGCCCGGCTGCTGTTCAGGTTCTGCCGGGTACCTGGAAATTCACCGTTACCTGGTCTG<br>ACGGTCAGACCTCTACCGGTCAGGTTCACGTTCAGCCGCGTACCCCGACCACCGTTCAGGT<br>TCACTTCCGTGGTCGTTCTTCTGACGGTCGTCCGTTCAACGGTAAAGGTCACCTGACCATG<br>AAAACCCCGACCACCCTGGACGTTAACGTTACCCAGTCTGACGGTGCTACCTCTACCGGTA<br>AATTCACCATGAAATCTCCGACCACCATCGACGTTACCTTCACCATCGAAGACGGTCAGAC<br>CGCTACCGGTCAGATGCACCGTCAGTCT |

| | | |
|---|---|---|
| HBI_b_18 | MGRCEAVRLPGTWKFDVTAE<br>DGSQFTGKVTVQPDSCDTVKI<br>TFNGTQSDGKPATGQGTLTMT<br>SPETVKITVTYSDGKTFTGYVT<br>LRTPTQFQVDATANDGTKSTG<br>YMRRTE | ATGGGCCGTTGCGAAGCTGTTCGTCTGCCGGGTACCTGGAAATTCGACGTTACCGCTGAAG<br>ACGGTTCTCAGTTCACCGGTAAAGTTACCGTTCAGCCGGACTCTTGCGACACCGTTAAAAT<br>CACCTTCAACGGTACCCAGTCTGACGGTAAACCGGCTACCGGTCAGGGTACCCTGACCATG<br>ACCTCTCCGGAAACCGTTAAAATCACCGTTACCTACTCTGACGGTAAAACCTTCACCGGTT<br>ACGTTACCCTGCGTACCCCGACCCAGTTCCAGGTTGACGCTACCGCTAACGACGGTACCAA<br>ATCTACCGGTTACATGCGTCGTACCGAA |
| HBI_b_19 | MGPAAAQVFPGTWDFQFTAE<br>DGSTFRGKVTFQPRTPTTLDVT<br>MKGTQSDGRPLTGKGKVHVE<br>SPTTVQINVTYSDGRTIQGKLT<br>VKTPTTVDVDARFSDGTKSTG<br>KVRRTS | ATGGGCCCGGCTGCGGCGCAGGTTTTCCCGGGTACCTGGGACTTCCAGTTCACCGCGGAAG<br>ACGGTTCTACTTTCCGTGGTAAAGTTACCTTCCAGCCGCGCACTCCGACCACCCTCGACGTT<br>ACTATGAAAGGTACTCAGTCTGACGGTCGTCCGCTGACCGGTAAAGGCAAAGTGCACGTT<br>GAATCTCCAACCACCGTGCAGATCAACGTTACCTACTCCGATGGTCGTACCATCCAGGGTA<br>AACTGACCGTTAAAACCCCGACTACCGTTGATGTTGACGCTCGCTTCTCCGACGGTACCAA<br>ATCTACTGGTAAGGTTCGTCGTACTTCT |
| HBI_b_20 | MGALVQNALPGKWQMHLQM<br>SDGSSFTGYVTFQPRSPTTFDV<br>HTQGQASDGQPSQGTGKVTV<br>KTPETVDVTTTMKDGRQVTG<br>KFTVKSPTHLQVDLQQADGST<br>VSGTMKRSE | ATGGGCGCTCTGGTTCAGAACGCTCTGCCGGGTAAATGGCAGATGCACCTGCAGATGTCTG<br>ACGGTTCTTCTTTCACCGGTTACGTTACCTTCCAGCCGCGTTCTCCGACCACCTTCGACGTT<br>CACACCCAGGGTCAGGCTTCTGACGGTCAGCCGTCTCAGGGTACCGGTAAAGTTACCGTTA<br>AAACCCCGGAAACCGTTGACGTTACCACCACCATGAAAGACGGTCGTCAGGTTACCGGTA<br>AATTCACCGTTAAATCTCCGACCCACCTGCAGGTTGACCTGCAGCAGGCTGATGGCTCTAC<br>CGTAAGCGGTACCATGAAACGTTCTGAA |
| HBI_b_21 | MGPEAVNILPGDWDVQLHSE<br>DGSTFRGTLRVQPKTPTTLDV<br>TMQGTVSDGRPSDGQGQVHV<br>DSPHDVKITMTMSDGSTATGT<br>LKLHSPTTFQVTLTYADGFTA<br>QGRFTRDG | ATGGGCCCGGAAGCGGTTAACATCCTGCCGGGTGACTGGGACGTTCAGCTCCACTCTGAA<br>GACGGCTCTACCTTCCGTGGTACCCTGCGTGTTCAGCCGAAAACCCCAACCACCCTGGACG<br>TTACCATGCAAGGTACTGTTTCTGATGGTCGTCCGTCCGACGGTCAAGGTCAAGTTCACGT<br>TGACTCTCCGCACGACGTTAAAATCACCATGACCATGTCTGACGGTTCTACCGCGACCGGC<br>ACCCTGAAACTGCACTCCCCGACCACCTTCCAGGTTACCCTCACTTACGCGGACGGTTTCA<br>CCGCTCAGGGTCGTTTCACTCGTGACGGT |
| HBI_b_22 | MGRAAVQFLPGTWNITSTYED<br>GTTMQGTVHVTPRSPETFDITV<br>QGQASDGKPMRGQGKVTVQS<br>PHQVQVNLTSEDGTQAQGVF<br>QVDSPTRVKVDLTASDGVRLT<br>GTLQRTS | ATGGGCCGTGCGGCGGTTCAGTTCCTGCCGGGTACTTGGAACATCACTTCTACCTACGAAG<br>ATGGCACCACCATGCAGGGTACTGTTCATGTTACCCCGCGTTCTCCGGAAACCTTCGACAT<br>CACCGTGCAAGGCCAAGCGTCTGACGGTAAACCGATGCGTGGTCAGGGTAAAGTTACCGT<br>TCAGTCTCCGCACCAGGTTCAAGTTAATCTGACCTCTGAAGACGGTACCCAGGCGCAGGGT<br>GTTTTCCAGGTTGACTCTCCGACCCGTGTTAAAGTTGACCTGACCGCGTCCGACGGTGTTC<br>GTCTCACCGGTACTCTGCAACGTACCTCT |
| HBI_b_23 | MGPEAYQVLPGTWKFTVTWS<br>DGQTSTGQVHFQPRTPTTLQV<br>HFRGRSSDGRPFNGKGHVTM<br>KTPTTFDVNVTQSDGATSTGKI<br>TMKSPTTIDVTFTIEDGQTATG<br>QMHRQS | ATGGGCCCGGAAGCTTACCAGGTTCTGCCGGGTACCTGGAAATTCACCGTTACCTGGTCTG<br>ACGGTCAGACCTCTACCGGTCAGGTTCACTTCCAGCCGCGTACCCCGACCACCCTGCAGGT<br>TCACTTCCGTGGTCGTTCTTCTGACGGTCGTCCGTTCAACGGTAAAGGTCACGTTACCATGA<br>AAACCCCGACCACCTTCGACGTTAACGTTACCCAGTCTGACGGTGCTACCTCTACCGGTAA<br>AATCACCATGAAATCTCCGACCACCATCGACGTTACCTTCACCATCGAAGACGGTCAGACC<br>GCTACCGGTCAGATGCACCGTCAGTCT |
| HBI_b_24 | MGPCHALELPGTWKFDVTAE<br>DGSQFTGKVTVQPDSPDTVKI<br>TFNGTQSDGKPATGQGTLTCT<br>SPETVKITVTYSDGKTFTGYVT<br>LRTPTQFQVDATANDGTKSTG<br>YMRRTE | ATGGGCCCGTGCCACGCTCTGGAACTGCCGGGTACCTGGAAATTCGACGTTACCGCTGAAG<br>ACGGTTCTCAGTTCACCGGTAAAGTTACCGTTCAGCCGGACTCTCCGGACACCGTTAAAAT<br>CACCTTCAACGGTACCCAGTCTGACGGTAAACCGGCTACCGGTCAGGGTACCCTGACCTGC<br>ACCTCTCCGGAAACCGTTAAAATCACCGTTACCTACTCTGACGGTAAAACCTTCACCGGTT<br>ACGTTACCCTGCGTACCCCGACCCAGTTCCAGGTTGACGCTACCGCTAACGACGGTACCAA<br>ATCTACCGGTTACATGCGTCGTACCGAA |
| HBI_b_25 | MGPAAAQYFPGTWKVQFTVE<br>DGSTFTGRVDFQPRTPTTLDV<br>RFQGTQSDGKPVQGKGKVHV<br>DSPTTLTVNVTYSDGRTIQGKL<br>TLKTPTKFDVDATFSDGTKST<br>GTVHRTS | ATGGGCCCGGCTGCTGCTCAGTACTTCCCGGGCACCTGGAAAGTTCAGTTCACTGTTGAAG<br>ACGGTTCTACCTTCACCGGTCGTGTGGACTTCCAGCCGCGTACCCCAACTACCCTGGACGT<br>TCGTTTCCAAGGTACCCAGTCTGACGGCAAACCAGTTCAGGGTAAAGGTAAAGTGCACGTT<br>GACTCTCCGACCACCCTGACCGTTAATGTTACCTACTCCGACGGTCGTACCATTCAGGGCA<br>AGCTGACCCTCAAAACCCCGACCAAATTCGACGTTGATGCGACCTTCTCCGATGGCACCAA<br>ATCTACTGGTACTGTTCACCGCACCTCT |
| HBI_b_26 | MGAEVAQVLPGKWQVHMTN<br>EDGTTSTGTMTVQPRSPYTFD<br>VKFKGTMSDGRPITGNGKVT<br>MKTPDTLDVDLTYSDGKKVT<br>GKVTMRSPTQLDWDLTTSDGS<br>KVTGTSKRQE | ATGGGCGCGGAAGTTGCGCAGGTTCTGCCGGGTAAATGGCAAGTTCACATGACCAACGAA<br>GACGGTACCACCTCTACCGGCACCATGACCGTTCAGCCGCGTTCCCCGTACACCTTCGACG<br>TTAAATTCAAAGGTACCATGTCCGACGGTCGTCCGATCACCGGCAACGGTAAGGTTACCAT<br>GAAAACCCCGGACACCCTGGACGTTGATCTGACCTACTCTGACGGTAAAAAGGTGACTGG<br>TAAAGTGACCATGCGTTCTCCGACCCAGCTCGACTGGGACCTGACCACTTCTGATGGTTCT<br>AAAGTTACTGGCACTTCTAAACGTCAGGAA |

| | | |
|---|---|---|
| HBI_b_27 | MGPEAVNMLPGDWDIQLTSE<br>DGSQFRGTFRVRPTTPTTVQV<br>TMRGTVSDGRPSQGQGYVTV<br>DSPTDMQVKMTMSDGSQAQG<br>TIKLDSPTTLKIKLTYADGFTA<br>QGTFTRDG | ATGGGCCCGGAAGCGGTTAACATGCTGCCGGGTGACTGGGACATCCAGCTGACCTCTGAA<br>GACGGTTCTCAGTTCCGCGGTACCTTTCGTGTTCGTCCAACCACCCCGACCACTGTTCAGGT<br>TACCATGCGTGGTACTGTTTCTGACGGTCGTCCGTCTCAGGGTCAGGGTTACGTTACCGTTG<br>ACTCTCCGACTGACATGCAGGTTAAAATGACCATGTCCGACGGCTCTCAGGCGCAAGGTAC<br>TATCAAGCTGGACTCCCCGACTACCCTGAAAATCAAACTCACCTACGCGGACGGTTTCACC<br>GCGCAGGGTACTTTTACCCGTGACGGT |
| HBI_b_28 | MGPAAVQMLPGTWNITSTYE<br>DGTTMQGTVHVTPESPDTVKV<br>TVQGQASDGKPMRGQGHLTM<br>QSPHQVQVRLTSEDGTQAQGV<br>VKVDSPTQFQWDLTASDGVR<br>LTGTTHRTS | ATGGGCCCGGCTGCTGTTCAGATGCTGCCGGGTACCTGGAACATCACCTCTACCTACGAAG<br>ACGGTACCACCATGCAGGGTACCGTTCACGTTACCCCGGAATCTCCGGACACCGTTAAAGT<br>TACCGTTCAGGGTCAGGCTTCTGACGGTAAACCGATGCGTGGTCAGGGTCACCTGACCATG<br>CAGTCTCCGCACCAGGTTCAGGTTCGTCTGACCTCTGAAGACGGTACCCAGGCTCAGGGTG<br>TTGTTAAAGTTGACTCTCCGACCCAGTTCCAGTGGGACCTGACCGCTTCTGACGGTGTTCGT<br>CTGACCGGTACCACCCACCGTACCTCT |
| HBI_b_29 | MGPAAIQVLPGTWQFQVTFSD<br>GQTSRGHVHVQPRTPTTVQVH<br>FTGRSSDGRPFNGKGHMTMK<br>TPTTVDVNVTQSDGATSTGKF<br>TMKSPTTLDVRFTIEDGQTAQ<br>GQLHRQS | ATGGGCCCAGCGGCGATCCAAGTTCTGCCGGGTACCTGGCAGTTCCAGGTTACCTTCTCTG<br>ACGGCCAGACCTCTCGTGGTCATGTTCACGTGCAGCCGCGTACCCCGACCACCGTTCAGGT<br>TCACTTCACCGGTCGTTCTTCTGATGGTCGTCCGTTCAACGGTAAAGGTCACATGACTATG<br>AAAACCCCAACTACCGTGGATGTTAACGTGACCCAGTCCGACGGTGCGACCTCTACTGGTA<br>AATTCACCATGAAATCTCCGACTACCCTGGACGTTCGTTTCACTATTGAAGACGGTCAGAC<br>CGCGCAGGGTCAGCTCCACCGTCAATCT |
| HBI_b_30 | MGPKCVSQLPGDWDVTFHSE<br>DGSTFRGTVRMQPRTPDTVQV<br>KMTGTVSDGRPSTGTGYVHV<br>DSPHDVKFQMTMSDGSTATG<br>TLKCDSPTQFTVRLTYADGFT<br>AQGHLQRDG | ATGGGCCCGAAGTGCGTTTCTCAGCTGCCGGGTGACTGGGACGTTACCTTCCACTCTGAAG<br>ACGGTTCCACCTTCCGTGGTACCGTTCGTATGCAGCCGCGTACCCCGGACACCGTTCAGGT<br>TAAAATGACCGGCACCGTGTCCGACGGTCGTCCGTCCACTGGTACTGGTTACGTTCACGTT<br>GACTCTCCGCACGACGTTAAATTCCAGATGACCATGTCTGATGGTTCTACCGCGACCGGTA<br>CCCTGAAATGCGACTCCCCGACCCAGTTCACCGTGCGCCTGACCTACGCGGACGGTTTCAC<br>CGCGCAGGGTCACCTCCAGCGTGACGGT |
| HBI_b_31 | MGAASQVYLPGKWDLNFTSS<br>DGTTFTGKMTVQPKTPDTVDV<br>TIKGKQSDGNPTNGQGQLHVE<br>SPTTFTWDVTYADGKTFKGKT<br>QLKTPTTLQVDVRAADGSKST<br>GYLTRKD | ATGGGCGCGGCGTCTCAGGTTTATCTGCCGGGTAAATGGGACCTGAACTTCACCTCCTCTG<br>ACGGTACCACTTTCACCGGTAAAATGACCGTTCAGCCGAAGACCCCGGACACCGTGGACG<br>TTACTATCAAAGGTAAACAGTCCGATGGCAACCCGACCAACGGCCAAGGTCAGCTGCACG<br>TTGAATCTCCGACCACCTTCACTTGGGATGTTACCTACGCTGACGGCAAAACCTTCAAAGG<br>CAAGACCCAGCTGAAAACCCCAACCACCCTCCAGGTTGACGTGCGTGCGGCGGACGGCTC<br>TAAATCTACCGGTTACCTGACCCGTAAAGAC |
| HBI_b_32 | MGQEVAQVLPGDWQVHMTN<br>EDGQTSTGTVTFQPRSPYTFDV<br>KFKGTMSDGRPITGKGKMTM<br>KTPDTMDIDVTYSDGKKVTGK<br>VTMKSPTQLDWDLTTSDGSK<br>VTGTSHRQE | ATGGGCCAGGAAGTTGCTCAAGTTCTGCCGGGTGACTGGCAGGTTCACATGACCAACGAA<br>GACGGTCAGACCTCTACTGGTACCGTTACCTTCCAGCCGCGTTCTCCGTACACCTTCGACGT<br>TAAATTCAAAGGTACCATGTCCGACGGTCGTCCGATCACCGGTAAGGGTAAAATGACTAT<br>GAAAACCCCGGACACTATGGACATCGACGTGACCTACTCTGACGGTAAAAAAGTGACCGG<br>CAAAGTTACCATGAAATCTCCGACCCAGCTCGACTGGGACCTGACCACCTCTGATGGTTCC<br>AAGGTTACTGGCACCTCTCACCGTCAAGAA |
| HBI_b_33 | MGPEAINVLPGDWDVTFHSED<br>GSTFRGTVRMQPRTPDTVQVK<br>MTGTVSDGRPSTGTGYVHVDS<br>PHDVKFQMTMSDGSTATGTL<br>KMDSPTQFTVRLTYADGFTAQ<br>GHLQRDG | ATGGGCCCGGAAGCGATCAATGTTCTGCCGGGTGACTGGGACGTTACCTTCCACTCTGAAG<br>ATGGTTCTACTTTCCGTGGTACCGTTCGTATGCAGCCGCGTACCCCGGACACCGTTCAGGTT<br>AAAATGACCGGCACCGTGTCCGACGGTCGTCCGTCTACTGGTACTGGTTACGTTCACGTTG<br>ATTCTCCGCACGACGTTAAATTCCAGATGACCATGTCTGACGGCTCTACCGCGACCGGTAC<br>CCTGAAAATGGACTCTCCAACCCAGTTCACCGTGCGCCTGACCTACGCGGACGGTTTCACC<br>GCGCAGGGTCACCTCCAGCGTGACGGT |
| HBI_b_34 | MGADAQSVLPGTWDINVTFSD<br>GSTFTGTLHFTPRTPTTFDVTL<br>KGHSSDGQPATGQGKVTLKTP<br>TTLDIDIQNSDGSKSTGQITMD<br>TPYDLKFTAKLSDGVTFTGTL<br>KRRE | ATGGGCGCGGACGCGCAGTCTGTTCTGCCGGGTACCTGGGACATCAACGTTACCTTCTCCG<br>ACGGCTCTACCTTCACCGGCACCCTGCACTTCACCCCGCGTACCCCAACCACCTTCGACGT<br>GACCCTCAAAGGTCACTCCTCTGACGGTCAGCCGGCGACTGGTCAAGGTAAGGTTACCCTG<br>AAAACCCCGACCACCCTGGACATCGACATCCAGAACTCTGATGGCTCCAAATCTACCGGTC<br>AGATCACTATGGACACCCCGTACGACCTGAAATTCACCGCGAAACTGTCCGATGGCGTTAC<br>TTTCACTGGTACTCTCAAGCGTCGTGAA |
| HBI_b_35 | MGSNAAQFLPGTWDVTFTAE<br>DGSTFQGKVHIKPETPDRVRV<br>TFKGTQSDGKPATGNGSLQVD<br>TPTTVKVQVHYADGKDAKGT<br>VTLRTPTTFTFDAQLADGAKS<br>TGQVTRKE | ATGGGCTCTAATGCGGCGCAGTTCCTGCCGGGTACCTGGGACGTTACTTTCACCGCGGAAG<br>ACGGTTCTACCTTCCAGGGTAAAGTGCACATCAAACCGGAAACCCCGGACCGTGTTCGTGT<br>TACCTTCAAAGGCACCCAGTCTGATGGTAAACCGGCGACCGGTAATGGTTCTCTGCAGGTT<br>GACACCCCAACCACCGTTAAAGTTCAGGTTCACTACGCTGACGGCAAGGACGCGAAAGGT<br>ACCGTTACTCTGCGTACCCCGACCACCTTCACCTTCGACGCGCAGCTGGCGGACGGTGCGA<br>AATCCACCGGTCAAGTTACCCGTAAAGAA |

| | | |
|---|---|---|
| HBI_b_36 | MGTSCSSGAPGNWDLQMTSS<br>DGSQSRGTITMKPQTPDQLQV<br>TVKGHFSDGKPFKGTGYVQVT<br>TPTQLTVHLTYSDGRQATGQF<br>NCDTPTDFKVTFTFSDGSTAQ<br>GTVKRTE | ATGGGCACCTCTTGCTCCTCTGGTGCTCCGGGTAACTGGGACCTGCAGATGACTTCTTCCG<br>ACGGCTCCCAGTCTCGTGGCACCATCACCATGAAACCGCAGACCCCGGACCAGCTCCAGG<br>TGACCGTTAAAGGCCACTTCTCTGACGGCAAACCGTTCAAAGGTACCGGTTACGTTCAGGT<br>TACCACCCCGACCCAGCTGACCGTGCATCTGACCTATTCTGATGGTCGTCAGGCGACTGGT<br>CAGTTCAACTGCGATACCCCAACTGACTTCAAAGTTACCTTCACCTTCTCCGATGGTTCTAC<br>CGCGCAGGGTACCGTGAAGCGTACCGAA |
| HBI_b_37 | MGASASQALPGTWKVTFNSE<br>DGLQTHGVMTVQPDTPYTFK<br>VTFQGTHADGKPIRGQGKMTI<br>DTPTTVTFTVTAEDGRQQTGQ<br>VTVKSPDTMDVTAQAADGTT<br>YTGQVHRQK | ATGGGCGCTTCTGCTTCTCAGGCTCTGCCGGGTACCTGGAAAGTTACCTTCAACTCTGAAG<br>ACGGTCTGCAGACCCACGGTGTTATGACCGTTCAGCCGGACACCCCGTACACCTTCAAAGT<br>TACCTTCCAGGGTACCCACGCTGACGGTAAACCGATCCGTGGTCAGGGTAAAATGACCATC<br>GACACCCCGACCACCGTTACCTTCACCGTTACCGCTGAAGACGGTCGTCAGCAGACCGGTC<br>AGGTTACCGTTAAATCTCCGGACACTATGGACGTTACCGCTCAGGCTGCTGACGGTACCAC<br>CTACACCGGTCAGGTTCACCGTCAGAAA |
| HBI_b_38 | MGQKVAQVLPGTWQVTMTN<br>EDGQTSQGQMHFQPRSPYTLD<br>VKAQGTMSDGRPIQGKGKVT<br>MKTPDTMDVDITYSDGKQVQ<br>GQVTLDSPTQFKFDVTTSDGS<br>KVTGTLQRQE | ATGGGCCAGAAAGTTGCGCAGGTTCTGCCGGGTACTTGGCAGGTTACCATGACCAACGAA<br>GATGGTCAGACCTCTCAGGGTCAGATGCACTTCCAACCGCGTTCTCCGTACACTCTGGACG<br>TTAAAGCGCAAGGTACCATGTCTGATGGTCGTCCGATCCAGGGTAAAGGTAAAGTGACCA<br>TGAAAACCCCGGACACTATGGACGTGGACATCACCTACTCTGACGGTAAACAGGTTCAGG<br>GCCAAGTTACTCTCGACTCTCCGACCCAGTTCAAATTCGACGTTACCACCTCCGACGGTTCT<br>AAGGTTACTGGCACCCTGCAACGTCAGGAA |
| HBI_b_39 | MGPEAAEAAPGNWDLQMTSS<br>DGSQSRGTITMKPQTPDQLQV<br>TVKGHFSDGKPFKGTGYVQVT<br>TPTQLTVHLTYSDGRQATGQF<br>NLDTPTDFKVTFTFSDGSTAQ<br>GTVKRTE | ATGGGCCCGGAAGCTGCTGAAGCTGCTCCGGGTAACTGGGACCTGCAGATGACCTCTTCTG<br>ACGGTTCTCAGTCTCGTGGTACCATCACCATGAAACCGCAGACCCCGGACCAGCTGCAGGT<br>TACCGTTAAAGGTCACTTCTCTGACGGTAAACCGTTCAAAGGTACCGGTTACGTTCAGGTT<br>ACCACCCCGACCCAGCTGACCGTTCACCTGACCTACTCTGACGGTCGTCAGGCTACCGGTC<br>AGTTCAACCTGGACACCCCGACCGACTTCAAAGTTACCTTCACCTTCTCTGACGGTTCTACC<br>GCTCAGGGTACCGTTAAACGTACCGAA |
| HBI_b_40 | MGNAAAQYLPGKWKFTTTAE<br>DGVTITGKVTIQPRTPTTLDITV<br>TGTQSDGRPTTGTGKFHVKTP<br>TTVDSKLQLSDGRTFTGQMTV<br>DSPDTVTVTWTMQDGTTQQG<br>QITRQE | ATGGGCAACGCTGCGGCTCAGTACCTGCCGGGTAAATGGAAATTCACCACCACCGCGGAA<br>GACGGTGTTACCATCACCGGTAAAGTTACTATCCAGCCGCGTACCCCAACCACCCTGGACA<br>TCACTGTTACCGGCACCCAGTCCGACGGTCGTCCGACCACTGGTACTGGTAAGTTCCACGT<br>TAAAACCCCGACTACCGTTGACTCCAAACTGCAGCTGTCTGATGGCCGTACCTTCACCGGC<br>CAGATGACTGTTGATTCTCCGGACACCGTTACCGTGACCTGGACCATGCAGGACGGTACCA<br>CCCAGCAGGGTCAGATTACCCGTCAGGAA |
| HBI_b_41 | MGTYAAQFLPGTWDIDTTAED<br>GSKFTGKLTVQPDTPTQLKVT<br>TNGKASDGKPATGQGTVTVET<br>PTTVKFQAKASDGNDITGKFT<br>VRTPTTLDVDYQAADGVKST<br>GKLTRRD | ATGGGCACCTACGCGGCTCAGTTCCTGCCGGGTACCTGGGACATCGACACCACCGCGGAA<br>GACGGTTCTAAGTTCACTGGTAAACTCACCGTTCAGCCGGACACTCCAACCCAGCTGAAAG<br>TTACCACCAACGGTAAAGCGTCTGACGGCAAGCCGGCGACCGGCCAGGGTACCGTGACCG<br>TTGAAACTCCGACTACCGTTAAATTCCAGGCGAAAGCTTCCGACGGTAACGACATCACCGG<br>TAAATTCACTGTTCGCACCCCGACCACTCTGGACGTTGACTACCAAGCGGCGGACGGTGTT<br>AAATCTACTGGCAAACTGACTCGTCGTGAC |
| HBI_b_42 | MGCDARTVLPGTWDINVTFSD<br>GSTFTGTLHFTPRTPTTFDVTL<br>KGHSSDGQPATGQGKVTLKTP<br>TTLDIDIQNSDGSKSTGQITCD<br>TPYDLKFTAKLSDGVTFTGTL<br>KRRE | ATGGGCTGCGATGCGCGTACTGTTCTGCCGGGTACCTGGGACATCAATGTTACTTTCTCTG<br>ACGGTTCTACTTTCACTGGCACCCTGCACTTCACCCCGCGTACCCCAACCACTTTCGACGTT<br>ACTCTGAAAGGTCATTCTTCCGACGGTCAGCCGGCGACCGGCCAGGGTAAAGTGACCCTG<br>AAAACCCCGACCACCCTGGACATCGACATCCAGAACTCCGATGGTTCTAAATCTACCGGTC<br>AAATCACTTGCGATACCCCGTACGACCTGAAATTCACCGCGAAACTGTCTGATGGCGTTAC<br>CTTCACCGGTACCCTCAAGCGTCGTGAA |
| HBI_b_43 | MGASATQALPGTWTLTFNSED<br>GLQTHGQWTMQPKTPTTVDV<br>TVQGTHADGKPIRGQGKMTV<br>DTPTTVTFTVTAEDGRQQTGQ<br>VTVKSPTTMDVTAQAADGTTF<br>TGKVHRQS | ATGGGCGCTTCTGCTACCCAGGCTCTGCCGGGTACCTGGACCCTGACCTTCAACTCTGAAG<br>ACGGTCTGCAGACCCACGGTCAGTGGACCATGCAGCCGAAAACCCCGACCACCGTTGACG<br>TTACCGTTCAGGGTACCCACGCTGACGGTAAACCGATCCGTGGTCAGGGTAAAATGACCGT<br>TGACACCCCGACCACCGTTACCTTCACCGTTACCGCTGAAGACGGTCGTCAGCAGACCGGT<br>CAGGTTACCGTTAAATCTCCGACCACTATGGACGTTACCGCTCAGGCTGCTGACGGTACCA<br>CCTTCACCGGTAAAGTTCACCGTCAGTCT |
| HBI_b_44 | MGPAAAQVLPGTWDVKFTSK<br>DGTTITGKMTIKPRSPETFDVT<br>MTGNMSDGKPYQGQGQVTVR<br>TPDTVDVQVTAKDGRTFRGKI<br>TLRSPTKMTLTSTASDGQTAT<br>GHFRRQP | ATGGGCCCGGCTGCTGCTCAGGTTCTGCCGGGTACCTGGGACGTTAAATTCACCTCTAAAG<br>ACGGTACCACCATCACCGGTAAAATGACCATCAAACCGCGTTCTCCGGAAACCTTCGACGT<br>TACCATGACCGGTAACATGTCTGACGGTAAACCGTACCAGGGTCAGGGTCAGGTTACCGTT<br>CGTACCCCGGACACCGTTGACGTTCAGGTTACCGCTAAAGACGGTCGTACCTTCCGTGGTA<br>AAATCACCCTGCGTTCTCCGACCAAAATGACCCTGACCTCTACCGCTTCTGACGGTCAGAC<br>CGCTACCGGTCACTTCCGTCGTCAGCCG |

| | | |
|---|---|---|
| HBI_b_45 | MGALAQEAAPGTWDVQMDSS<br>DGSKSRGKLHLKPTTPTQFTV<br>TLTGHFSDGKPFQGNGYVDVT<br>TPTTLTLTVTYKDGSQAQGKL<br>DFETPTTLKFTLTFSDGSTAKG<br>DVTRTE | ATGGGCGCTCTGGCTCAGGAAGCTGCTCCGGGTACCTGGGACGTTCAGATGGACTCTTCTG<br>ACGGTTCTAAATCTCGTGGTAAACTGCACCTGAAACCGACCACCCCGACCCAGTTCACCGT<br>TACCCTGACCGGTCACTTCTCTGACGGTAAACCGTTCCAGGGTAACGGTTACGTTGACGTT<br>ACCACCCCGACCACCCTGACCCTGACCGTTACCTACAAAGACGGTTCTCAGGCTCAGGGTA<br>AACTGGACTTCGAAACCCCGACCACCCTGAAATTCACCCTGACCTTCTCTGACGGTTCTAC<br>CGCTAAAGGTGACGTTACCCGTACCGAA |
| HBI_b_46 | MGNAAAQFLPGKWKFQTTAE<br>DGVTITGTFTVQPRTPTRVDVT<br>VTGTQSDGRPTTGTGQMQVR<br>TPTTVDVRLTLSDGRTVQGKL<br>TVDSPDTVTITLTMQDGTTQQ<br>GQLTRQE | ATGGGCAACGCTGCTGCTCAGTTCCTGCCGGGTAAATGGAAATTCCAGACCACCGCTGAA<br>GACGGTGTTACCATCACCGGTACCTTCACCGTTCAGCCGCGTACCCCGACCCGTGTTGACG<br>TTACCGTTACCGGTACCCAGTCTGACGGTCGTCCGACCACCGGTACCGGTCAGATGCAGGT<br>TCGTACCCCGACCACCGTTGACGTTCGTCTGACCCTGTCTGACGGTCGTACCGTTCAGGGT<br>AAACTGACCGTTGACTCTCCGGACACCGTTACCATCACCCTGACCATGCAGGACGGTACCA<br>CCCAGCAGGGTCAGCTGACCCGTCAGGAA |
| HBI_b_47 | MGPECHKVLPGTWDFNATNE<br>DGATITGQLDMQPRTPDRVQV<br>HSNGQYSDGKPVQGTGHVQC<br>DTPTTVKFQVDLSDGKTIRGQ<br>LDLKTPTTVDITVTASDGSTST<br>GQLTRKE | ATGGGCCCGGAATGCCACAAAGTTCTGCCGGGTACCTGGGACTTCAACGCTACCAACGAA<br>GACGGTGCTACCATCACCGGTCAGCTGGACATGCAGCCGCGTACCCCGGACCGTGTTCAG<br>GTTCACTCTAACGGTCAGTACTCTGACGGTAAACCGGTTCAGGGTACCGGTCACGTTCAGT<br>GCGACACCCCGACCACCGTTAAATTCCAGGTTGACCTGTCTGACGGTAAAACCATCCGTGG<br>TCAGCTGGACCTGAAAACCCCGACCACCGTTGACATCACCGTTACCGCTTCTGACGGTTCT<br>ACCTCTACCGGTCAGCTGACCCGTAAAGAA |
| HBI_b_48 | MGQLTCQNLPGKWKFNTTAE<br>DGVTITGTLQVQPRTPTTFDVN<br>LKGHQSDGRPTKGNGKMTCK<br>TPDQVDSRFQLSDGRTFQGKL<br>QVDSPDTLTVNWTMQDGTTQ<br>TGQLTRQE | ATGGGCCAGCTGACTTGCCAGAACCTGCCGGGTAAATGGAAATTCAACACCACCGCGGAA<br>GACGGTGTTACCATCACCGGTACCCTGCAAGTGCAGCCGCGTACCCCGACCACCTTCGACG<br>TTAACCTGAAAGGTCACCAGTCTGATGGTCGTCCGACCAAAGGTAACGGTAAAATGACCT<br>GCAAAACCCCGGACCAAGTTGACTCCCGTTTCCAGCTGTCTGACGGTCGTACCTTCCAGGG<br>TAAACTCCAGGTGGACTCTCCGGACACCCTGACCGTTAACTGGACCATGCAGGACGGTACC<br>ACCCAGACCGGCCAGCTCACTCGTCAGGAA |
| HBI_b_49 | MGALAQEAAPGTWDVTMTSS<br>DGSQSQGQFHMQPTTPTRFTV<br>TVRGHFSDGKPFKGQGYVDV<br>ETPTTLRINVTYSDGSQATGKL<br>QFDTPTDVKVTLTFSDGSTAQ<br>GTMKRTE | ATGGGCGCGCTGGCGCAGGAAGCGGCTCCGGGTACCTGGGACGTGACCATGACCTCCTCT<br>GATGGCTCTCAGTCTCAGGGTCAGTTCCACATGCAGCCGACCACTCCGACCCGTTTCACCG<br>TTACCGTTCGTGGTCACTTTTCTGACGGTAAACCGTTCAAAGGCCAGGGTTACGTTGACGT<br>TGAAACTCCAACCACCCTCCGTATCAACGTTACCTACTCCGATGGTTCTCAGGCGACCGGT<br>AAGCTGCAATTTGACACCCCGACCGACGTTAAAGTTACCCTGACCTTCTCCGACGGCTCTA<br>CCGCGCAGGGTACTATGAAACGTACCGAA |
| HBI_b_50 | MGADVQQVLPGTWDLRVTQE<br>DGTTTQGKVTVKPETPTTLRF<br>TSKGTMSDGKPFHGQGHFTVK<br>SPTTVQIQQTASDGKTATGSLT<br>MKTPTTLDVTMTNQDGTTAQ<br>GKMTRKS | ATGGGCGCGGACGTTCAGCAGGTTCTGCCGGGTACCTGGGACCTGCGTGTTACCCAGGAA<br>GATGGTACTACCACCCAGGGTAAAGTTACCGTTAAACCGGAAACTCCAACCACTCTGCGTT<br>TCACCTCTAAAGGTACCATGTCTGATGGTAAACCGTTCCACGGTCAGGGTCACTTCACTGT<br>TAAATCTCCGACCACTGTTCAGATCCAGCAGACCGCGTCTGACGGTAAAACCGCGACCGGT<br>TCTCTGACCATGAAAACCCCGACTACTCTGGACGTTACCATGACCAACCAGGACGGTACCA<br>CCGCGCAGGGCAAGATGACCCGTAAATCT |
| HBI_b_51 | MGCDKANQLPGTWKFDVTAE<br>DGSQFTGKVTVQPDSPDTVKI<br>TFNGTQSDGKPATGQGTLTCT<br>SPETVKITVTYSDGKTFTGYVT<br>LRTPTQFQVDATANDGTKSTG<br>YMRRTE | ATGGGCTGCGACAAAGCGAACCAGCTGCCGGGTACTTGGAAATTCGACGTTACTGCGGAA<br>GACGGTTCTCAGTTTACTGGCAAAGTTACCGTTCAGCCGGACTCTCCGGACACTGTTAAGA<br>TCACCTTCAACGGTACCCAGTCCGACGGCAAACCGGCGACTGGTCAGGGTACTCTCACTTG<br>CACCTCTCCAGAAACCGTGAAAATCACCGTGACCTACTCTGACGGTAAAACCTTCACTGGT<br>TACGTGACCCTGCGTACCCCGACCCAGTTCCAGGTTGACGCGACCGCGAACGACGGTACTA<br>AATCTACCGGTTATATGCGTCGTACCGAA |
| HBI_b_52 | MGPCSTNVLPGTWKFTVTWS<br>DGQTSTGQVHFQPRTPTTLQV<br>HFRGRSSDGRPFNGKGHVTM<br>KTPTTFDVNVTQSDGATSTGKI<br>TCKSPTTIDVTFTIEDGQTATG<br>QMHRQS | ATGGGCCCGTGCTCTACCAACGTTCTGCCGGGTACCTGGAAATTCACCGTTACTTGGTCTG<br>ACGGCCAGACCTCTACCGGCCAAGTTCATTTTCAGCCGCGTACTCCGACTACTCTGCAGGT<br>TCACTTCCGTGGTCGTTCTTCCGATGGTCGTCCGTTCAACGGTAAAGGTCACGTTACTATGA<br>AAACCCCAACCACTTTCGATGTTAATGTTACCCAGTCTGATGGTGCGACTTCTACTGGTAA<br>AATCACTTGCAAATCTCCGACCACCATCGACGTTACCTTCACCATTGAAGACGGTCAAACT<br>GCTACCGGTCAGATGCACCGTCAATCT |
| HBI_b_53 | MGPDLSKVLPGTWDFNATNE<br>DGATITGQLDMQPRTPDRVQV<br>HSNGQYSDGKPVQGTGHVQV<br>DTPTTVKFQVDLSDGKTIRGQ<br>LDLKTPTTVDITVTASDGSTST<br>GQLTRKE | ATGGGCCCGGATCTGTCTAAAGTTCTGCCGGGTACCTGGGACTTCAACGCGACCAACGAA<br>GACGGTGCGACCATCACTGGTCAGCTCGACATGCAGCCGCGTACCCCGGACCGTGTTCAG<br>GTTCACTCTAACGGTCAGTACTCTGATGGCAAGCCGGTTCAGGGTACCGGTCATGTTCAAG<br>TTGACACTCCGACCACTGTGAAATTCCAGGTTGACCTGTCTGACGGCAAAACCATCCGCGG<br>TCAACTGGACCTGAAAACCCCAACCACCGTTGACATCACCGTTACCGCTTCCGACGGTTCT<br>ACCTCTACCGGCCAGCTCACTCGTAAAGAA |

| | | |
|---|---|---|
| HBI_b_54 | MGCDQWRILPGTWDINVTFSD<br>GSTFTGTLHFTPRTPTTFDVTL<br>KGHSSDGQPATGQGKVTLKTC<br>TTLDIDIQNSDGSKSTGQITMD<br>TPYDLKFTAKLSDGVTFTGTL<br>KRRE | ATGGGCTGCGACCAGTGGCGTATCCTGCCGGGTACCTGGGACATCAACGTTACCTTCTCTG<br>ACGGTTCTACCTTCACCGGTACCCTGCACTTCACCCCGCGTACCCCGACCACCTTCGACGTT<br>ACCCTGAAAGGTCACTCTTCTGACGGTCAGCCGGCTACCGGTCAGGGTAAAGTTACCCTGA<br>AAACCTGCACCACCCTGGACATCGACATCCAGAACTCTGACGGTTCTAAATCTACCGGTCA<br>GATCACTATGGACACCCCGTACGACCTGAAATTCACCGCTAAACTGTCTGACGGTGTTACC<br>TTCACCGGTACCCTGAAACGTCGTGAA |
| HBI_b_55 | MGKDAANVLPGKWKFNTTAE<br>DGVTITGTLQVQPRTPTTFDVN<br>LKGHQSDGRPTKGNGKMTMK<br>TPDQVDSRFQLSDGRTFQGKL<br>QVDSPDTLTVNWTMQDGTTQ<br>TGQLTRQE | ATGGGCAAAGACGCGGCGAACGTGCTGCCGGGTAAATGGAAATTCAACACCACCGCGGAA<br>GACGGTGTTACCATCACCGGTACTCTGCAAGTTCAGCCGCGTACCCCGACCACCTTCGACG<br>TTAACCTGAAAGGTCACCAGTCCGACGGTCGTCCAACCAAAGGTAACGGTAAAATGACCA<br>TGAAAACCCCGGACCAGGTTGACTCCCGTTTCCAGCTGTCTGATGGTCGTACCTTCCAGGG<br>TAAACTCCAGGTGGACTCTCCGGACACCCTGACCGTTAACTGGACCATGCAGGACGGTACC<br>ACCCAGACTGGTCAGCTCACTCGTCAGGAA |
| HBI_b_56 | MGSALAQQLPGTWKVDVTSE<br>DGVRTTGQVHFQPRTPTTMDV<br>TLTGTHADGKPFTGQGKVTIT<br>TPTTVKVTVTYEDGSTATGQF<br>TVDSPTQLKFDMTASDGTRFT<br>GTVQRQS | ATGGGCTCTGCGCTGGCGCAGCAACTGCCGGGTACTTGGAAAGTTGATGTTACCTCTGAAG<br>ACGGTGTTCGTACCACTGGTCAGGTTCACTTCCAGCCGCGTACCCCAACCACTATGGACGT<br>GACCCTGACTGGCACCCACGCGGACGGTAAACCGTTCACTGGCCAAGGTAAAGTGACCAT<br>CACCACCCCGACCACCGTTAAAGTTACCGTTACCTACGAGGACGGCTCTACCGCGACCGGC<br>CAGTTCACCGTTGACTCTCCGACCCAGCTGAAATTCGACATGACCGCGTCTGACGGTACCC<br>GTTTTACCGGCACCGTGCAGCGTCAATCT |

**Supplementary Table 5: List of oligos used for making mutants of DFHBI-binding designs.** DNA oligo sequences and the purpose of usage are listed in this table.

| Name | Sequence | Purpose |
|---|---|---|
| b32_Y71F_For | GACATCGACGTGACCTTCTCTGACGGTAAAAAAG | making b32_Y71F mutant |
| b32_Y71F_Rev | CTTTTTTACCGTCAGAGAAGGTCACGTCGATGTC | |
| b32_S23V_For | GAAGACGGTCAGACCGTTACTGGTACCGTTACC | making b32_S23V mutant |
| b32_S23V_Rev | GGTAACGGTACCAGTAACGGTCTGACCGTCTTC | |
| b32_T95V_For | CTGGGACCTGACCGTTTCTGATGGTTCCAAG | making b32_T95V mutant |
| b32_T95V_Rev | CTTGGAACCATCAGAAACGGTCAGGTCCCAG | |
| b32_N17S_For | GGTTCACATGACCTCTGAAGACGGTCAGAC | making b32_N17S mutant |
| b32_N17S_Rev | GTCTGACCGTCTTCAGAGGTCATGTGAACC | |
| b32_Q21T_For | CCAACGAAGACGGTACCACCTCTACTGGTACC | making b32_Q21T mutant |
| b32_Q21T_Rev | GGTACCAGTAGAGGTGGTACCGTCTTCGTTGG | |
| b32_M15V_For | GACTGGCAGGTTCACGTTACCAACGAAGACGGTC | making b32_M15V mutant |
| b32_M15V_Rev | GACCGTCTTCGTTGGTAACGTGAACCTGCCAGTC | |
| b32_M15F_For | GACTGGCAGGTTCACTTCACCAACGAAGACGGTC | making b32_M15F mutant |
| b32_M15F_Rev | GACCGTCTTCGTTGGTGAAGTGAACCTGCCAGTC | |
| b32_V27M_For | CCTCTACTGGTACCATGACCTTCCAGCCGCGTTC | making b32_V23M mutant |
| b32_V27M_Rev | GAACGCGGCTGGAAGGTCATGGTACCAGTAGAGG | |
| b32_F37Y_For | CGTTCTCCGTACACCTACGACGTTAAATTCAAAG | making b32_F37Y mutant |
| b32_F37Y_Rev | CTTTGAATTTAACGTCGTAGGTGTACGGAGAACG | |
| b32_K75I_For | CCTACTCTGACGGTATCAAAGTGACCGGCAAAG | making b32_K75I mutant |
| b32_K75I_Rev | CTTTGCCGGTCACTTTGATACCGTCAGAGTAGG | |
| b32_M59C_For | GTAAGGGTAAAATGACTTGCAAAACCCCGGACACTATG | making M59C/Q1C double mutant |
| b32_M59C_Rev | CATAGTGTCCGGGGTTTTGCAAGTCATTTTACCCTTAC | |
| b32_Q1C_For | GAAGGAGATATACCATGGGCTGCGAAGTTGCTCAAGTTC | |
| b11_Y71F_For | GATGTTGACATCACCTTCTCTGACGGCAAACAG | making b11_Y71F mutant |
| b11_Y71F_Rev | CTGTTTGCCGTCAGAGAAGGTGATGTCAACATC | |
| b11_S23V_For | GAAGACGGTCAGACCGTTCAGGGTCAGATGCAC | making b11_S23V mutant |
| b11_S23V_Rev | GTGCATCTGACCCTGAACGGTCTGACCGTCTTC | |
| b11_T95V_For | GTTTAAGTTTGACGTTACCGTTTCTGATGGTTCTAAAG | making b11_T95V mutant |
| b11_T95V_Rev | CTTTAGAACCATCAGAAACGGTAACGTCAAACTTAAAC | |
| b11_N17S_For | GCAAGTTACTATGACCTCTGAAGACGGTCAGAC | making b11_N17S mutant |

| | | |
|---|---|---|
| b11_N17S_Rev | GTCTGACCGTCTTCAGAGGTCATAGTAACTTGC | |
| b11_Q21T_For | CCAACGAAGACGGTACCACCTCTCAGGGTCAG | making b11_Q21T mutant |
| b11_Q21T_Rev | CTGACCCTGAGAGGTGGTACCGTCTTCGTTGG | |
| b11_M15F_For | CCTGGCAAGTTACTTTCACCAACGAAGACGGTC | making b11_M15F mutant |
| b11_M15F_Rev | GACCGTCTTCGTTGGTGAAAGTAACTTGCCAGG | |
| b11_M15V_For | CCTGGCAAGTTACTGTTACCAACGAAGACGGTC | making b11_M15V mutant |
| b11_M15V_Rev | GACCGTCTTCGTTGGTAACAGTAACTTGCCAGG | |
| b11_K75V_For | CACCTACTCTGACGGCGTGCAGGTTCAGGGCCAG | making b11_K75V mutant |
| b11_K75V_Rev | CTGGCCCTGAACCTGCACGCCGTCAGAGTAGGTG | |
| b11_L105F_For | CTAAAGTTACCGGCACCTTCCAACGTCAGGAAC | making b11_L105F mutant |
| b11_L105F_Rev | GTTCCTGACGTTGGAAGGTGCCGGTAACTTTAG | |
| b26_T21Q_For | CCAACGAAGACGGTCAGACCTCTACCGGCAC | making b26_T21Q mutant |
| b26_T21Q_Rev | GTGCCGGTAGAGGTCTGACCGTCTTCGTTGG | |

**Supplementary Table 6: List of sequences for b11 loop designs.** Protein sequences and DNA encoding sequences (optimized for *Saccharomyces cerevisae* yeast codon usage) of thirty-five designs are provided in this table.

| Design ID | Protein Sequence | DNA sequence |
|---|---|---|
| b11_L3B_11c | CRAASLLPGTWQVTMTNEDG GTSQGQMHFQPRSPYTLDVKA QGTTSGSLSSTESYQGKGKVT CKTPDTMDVDITYSDGLQVQG QVTLDSPTQFKFDVTTSDGSK VTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA GATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT TGGATGTTAAAGCTCAAGGTACTACATCTGGTTCATTATCTTCAACTGAATCATA CCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATAT CACATACTCTGATGGTTTGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACT CAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTGC AAAGACAAGAA |
| b11_L3B_2c | CRAASLLPGTWQVTMTNEDG GTSQGQMHFQPRSPYTLDVKA QGHTDGNMQDTESIQGKGKV TCKTPDTMDVDITYSDGLQVQ GQVTLDSPTQFKFDVTTSDGS KVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA GATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT TGGATGTTAAAGCTCAAGGTCATACAGATGGTAATATGCAAGATACTGAATCAA TCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATA TTACTTACTCTGATGGTTTGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAAC TCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTA CAAAGACAAGAA |
| b11_L3C_6c | CRAASLLPGTWQVTMTNEDG VTSQGQMHFQPRSPYTLDVKA QGEYKKSDANPSLNGKPIQGK GKVTCKTPDTMDVDITYSDG MQVQGQVTLDSPTQFKFDVTT SDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACTTGGCAAGTTACTATGACAAATGAAG ATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATT GGATGTTAAAGCTCAGGGTGAATACAAGAAATCTGATGCAAACCCATCATTGAA CGGTAAACCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTAT GGATGTTGATATCACTTACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATTG GATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTA CTGGTACATTACAAAGACAAGAA |
| b11_L3C_8c | CRAASLLPGTWQVTMTNEDG VTSQGQMHFQPRSPYTLDVKA QGSFDQTNSAPDLDGAPIQGK GKVTCKTPDTMDVDITYSHG MQVQGQVTLDSPTQFKFDVTT SDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA GATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT TGGATGTTAAAGCTCAAGGTTCTTTTGATCAAACTAATTCAGCTCCAGATTTGGA TGGTGCACCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTAT GGATGTTGATATTACATACTCTCATGGTATGCAAGTTCAAGGTCAAGTTACATTG GATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTA CTGGTACATTACAAAGACAAGAA |
| b11_L3D_1c | CRAASLLPGTWQVTMTNEDG GTSQGQMHFQPRSPYTLDVKA QGHGTGSLKGIPYQGKGKVTC KTPDTMDVDITYSDGMAVQG QVTLDSPTQFKFDVTTSDGSK VTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA GATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT TGGATGTTAAAGCTCAAGGTCATGGTACTGGTTCATTAAAAGGTATTCCATACCA GGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCAC ATACTCTGATGGTATGGCAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAA TTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAA GACAAGAA |

| | | |
|---|---|---|
| b11_L3D_3c | CRAASLLPGTWQVTMTNEDGGTSQGQMHFQPRSPYTLDVKAQGSGSGNLSGVPIQGKGKVTCKTPDTMDVDITYSDGMAVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAAGCTCAAGGTTCTGGTTCTGGTAATTTGTCAGGTGTTCCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTACTTACTCTGATGGTATGGCAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3F_1c | CRAASLLPGTWQVTMTNEDGATSQGQMHFQPRSPYTLDVKAQGTLHGFGNTIDSSIQGKGKVTCKTPDTMDVDITYSDGMQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTGCTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAAGCACAAGGTACATTGCATGGTTTCGGTAACACTATCGATTCTTCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTACTTACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3F_2c | CRAASLLPGTWQVTMTNEDGVTSQGQMHFQPRSPYTLDVKAQGTASGSGKDANKSYQGKGKVTCKTPDTMDVDITYSDGMQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAAGCTCAAGGTACTGCATCTGGTTCTGGTAAAGATGCTAATAAGTCATACCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTACATACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3G_1c | CRAASLLPGTWQVTMTNEDGMTSQGQMHFQPRSPYTLDVKAQGQAKSSSSPNQGSPYQGKGKVTCKTPDTMDVDITYSDGLQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTATGACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAAGCTCAAGGTCAAGCAAAATCTTCATCTTCACCAAATCAAGGTTCACCATACCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTACTTACTCTGATGGTTTGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3G_2c | CRAASLLPGTWQVTMTNEDGMTSQGQMHFQPRSPYTLDVKAQGQAKSSSSPYSGSPIQGKGKVTCKTPDTMDVDITYSDGLQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTATGACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAAGCTCAAGGTCAAGCAAAATCTTCATCTTCACCATACTCTGGTTCACCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCACTTACTCTGATGGTTTGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3G_3c | CRAASLLPGTWQVTMTNEDGVTSQGQMHFQPRSPYTLDVKAQGQTKNSNSPYSGSPWQGKGKVTCKTPDTMDVDITYSDGLQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAGGCTCAAGGTCAAACTAAAAATTCTAACTCACCATACTCTGGTTCACCATGGCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCACATACTCTGATGGTTTGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3IA_1c | CRAASLLPGTWQVTMTNEDGATSQGQMHFQPRSPYTLDVKAQGEADSQSEDVRKKLGSSPTYQGKGKVTCKTPDTMDVDITYSSGMQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTGCTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACATTGGATGTTAAAGCTCAAGGTGAAGCAGATTCTCAATCAGAAGATGTTAGAAAGAAATTGGGTTCTTCACCAACATACCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCACTTACTCTTCAGGCATGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGACAAGAA |
| b11_L3IA_3c | CRAASLLPGTWQVTMTNEDGVTSQGQMHFQPRSPYTLDVKAQGEDKSSSEKSRDDIGASPTYQGKGKVTCKTPDTMDVDITYSSGMQVQGQVTLDSPTQFKFDVTTSDGSKVTGTLQRQE | TGTAGAGCTGCATCATTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAAGATGGTGTTACTTCACAAGGTCAAATGCATTTTCAACCAAGATCTCCATATACATTGGATGTTAAAGCTCAAGGTGAGGATAAGTCTTCATCTGAAAAATCAAGAGATGATATTGGTGCATCTCCAACATACCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCACTTACTCATCTGGTATGCAAGTTCAAGGTCAAGTTACATTGGATTCTCCAACTCAATTCAAATTCGATGTTACTACATCAGATGGTTCTAAAGTTACTGGTACATTACAAAGACAAGAA |

| | | |
|---|---|---|
| b11_L3Jm_1c | CRAASLLPGTWQVTMTNEDG<br>GTSQGQMHFQPRSPYTLDVKA<br>QGQANPNSDDPTFRGTPIQGK<br>GKVTCKTPDTMDVDITYSDGL<br>QVQGQVTLDSPTQFKFDVTTS<br>DGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTCAAGCAAATCCAAATTCAGATGATCCAACTTTTA<br>GAGGTACTCCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTA<br>TGGATGTTGATATTACTTACTCTGATGGTTTGCAAGTTCAAGGTCAAGTTACATT<br>GGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTT<br>ACTGGTACATTACAAAGACAAGAA |
| b11_L3Jm_2c | CRAASLLPGTWQVTMTNEDG<br>GTSQGQMHFQPRSPYTLDVKA<br>QGQYSPSSDDPSLRGTPIQGKG<br>KVTCKTPDTMDVDITYSDGM<br>QVQGQVTLDSPTQFKFDVTTS<br>DGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTCAATACTCTCCATCTTCAGATGATCCATCATTAAG<br>AGGTACTCCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTAT<br>GGATGTTGATATCACATACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATTG<br>GATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTA<br>CTGGTACATTACAAAGACAAGAA |
| b11_L3Jnm_2c | CRAASLLPGTWQVTMTNEDG<br>VTSQGQMHFQPRSPYTLDVKA<br>QGQYDPRTEDSQLSGTPIQGK<br>GKVTCKTPDTMDVDITYSDG<br>MQVQGQVTLDSPTQFKFDVTT<br>SDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTCAATACGATCCAAGAACAGAAGATTCTCAATTAT<br>CAGGTACTCCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTA<br>TGGATGTTGATATCACTTACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATT<br>GGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTT<br>ACTGGTACATTACAAAGACAAGAA |
| b11_L3Jnm_3c | CRAASLLPGTWQVTMTNEDG<br>VTSQGQMHFQPRSPYTLDVKA<br>QGQVDPNSNDSKLRGSPIQGK<br>GKVTCKTPDTMDVDITYSHG<br>MQVQGQVTLDSPTQFKFDVTT<br>SDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTCAAGTTGATCCAAACTCTAACGATTCAAAGTTGA<br>GAGGTTCACCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTA<br>TGGATGTTGATATTACTTACTCTCATGGTATGCAAGTTCAAGGTCAAGTTACATT<br>GGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTT<br>ACTGGTACATTACAAAGACAAGAA |
| b11_L3K_1c | CRAASLLPGTWQVTMTNEDG<br>LTSQGQMHFQPRSPYTLDVKA<br>QGKVRSSDSRPDLNTEYQGKG<br>KVTCKTPDTMDVDITYNNGM<br>QVQGQVTLDSPTQFKFDVTTS<br>DGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTTTGACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATACACAT<br>TGGATGTTAAGGCTCAGGGTAAAGTTAGATCTTCAGATTCTAGACCAGATTTGA<br>ACACTGAATACCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGG<br>ATGTTGATATCACATACAACAATGGTATGCAAGTTCAAGGTCAAGTTACATTGG<br>ATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTAC<br>TGGTACATTACAAAGACAAGAA |
| b11_L3K_4c | CRAASLLPGTWQVTMTNEDG<br>VTSQGQMHFQPRSPYTLDVKA<br>QGSFDSTDSNPDLDSSIQGKGK<br>VTCKTPDTMDVDITYSHGMQ<br>VQGQVTLDSPTQFKFDVTTSD<br>GSKVTGTLQRQE | TGTAGAGCTGCATCATTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGTTACTTCACAAGGTCAAATGCATTTTCAACCAAGATCTCCATATACAT<br>TGGATGTTAAAGCTCAAGGTTCTTTCGATTCAACTGATTCTAACCCAGATTTGGA<br>TTCTTCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGA<br>TGTTGATATTACATACTCACATGGTATGCAAGTTCAAGGTCAAGTTACATTGGAT<br>TCTCCAACTCAATTCAAATTCGATGTTACTACATCAGATGGTTCTAAAGTTACTG<br>GTACATTACAAAGACAAGAA |
| b11_L3K_5c | CRAASLLPGTWQVTMTNEDG<br>VTSQGQMHFQPRSPYTLDVKA<br>QGQWRTTDSAPKLTTTLQGKG<br>KVTCKTPDTMDVDITYSDGM<br>QVQGQVTLDSPTQFKFDVTTS<br>DGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACTTGGCAAGTTACAATGACTAATGAAG<br>ATGGTGTTACATCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACTTT<br>GGATGTTAAAGCTCAAGGTCAATGGAGAACTACAGATTCAGCACCAAAATTGAC<br>TACAACTTTGCAGGGTAAAGGTAAAGTTACTTGTAAGACTCCAGATACAATGGA<br>TGTTGATATTACATACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACTTTGGAT<br>TCACCAACACAATTCAAATTCGATGTTACAACTTCTGATGGTTCAAAAGTTACAG<br>GTACTTTACAAAGACAAGAA |
| b11_L3L_2c | CRAASLLPGTWQVTMTNEDG<br>TTSQGQMHFQPRSPYTLDVKA<br>QGSYSPSTPTPSGEDSSISGKGK<br>VTCKTPDTMDVDIKYSDGMQ<br>VQGQVTLDSPTQFKFDVTTSD<br>GSKVTGTLQRQE | TGTAGAGCTGCATCATTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTACTACATCACAAGGTCAAATGCATTTTCAACCAAGATCTCCATATACTT<br>TGGATGTTAAAGCTCAAGGTTCTTACTCACCATCTACTCCAACACCATCAGGTGA<br>AGATTCTTCAATCTCTGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATG<br>GATGTTGATATCAAGTACTCAGATGGTATGCAAGTTCAAGGTCAAGTTACATTG<br>GATTCTCCAACTCAATTCAAATTCGATGTTACTACATCAGATGGTTCTAAAGTTA<br>CTGGTACATTACAAAGACAAGAA |

| | | |
|---|---|---|
| b11_L3L_4c | CRAASLLPGTWQVTMTNEDG<br>LTSEGQMHFQPRSPYTLDVKA<br>QGEFNPNSPSALNINDSISGKG<br>KVTCKTPDTMDVDIKYSNGLQ<br>VQGQVTLDSPTQFKFDVTTSD<br>GSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTTTGACTTCTGAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TAGATGTTAAAGCTCAGGGTGAATTCAATCCAAACTCTCCATCAGCATTGAACA<br>TCAACGATTCTATCTCTGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTAT<br>GGATGTTGATATCAAGTACTCTAATGGTTTGCAAGTTCAAGGTCAAGTTACATTG<br>GATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTA<br>CTGGTACATTACAAAGACAAGAA |
| b11_L3m_2c | CRAASLLPGTWQVTMTNEDG<br>GTSQGQMHFQPRSPYTLDVKA<br>QGQATDGSNKGTPYQGKGKV<br>TCKTPDTMDVDITYSDGIQVQ<br>GQVTLDSPTQFKFDVTTSDGS<br>KVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGGTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTCAAGCAACAGATGGTTCAAATAAGGGTACTCCAT<br>ACCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATA<br>TCACTTACTCTGATGGTATCCAAGTTCAAGGTCAAGTTACATTGGATTCACCAAC<br>TCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTA<br>CAAAGACAAGAA |
| b11_L3Mm_2c | CRAASLLPGTWQVTMTNEDG<br>LTSQGQMHFQPRSPYTLDVKA<br>QGELKDNSDPSWDSSIQGKGK<br>VTCKTPDTMDVDITYSDGMQ<br>VQGQVTLDSPTQFKFDVTTSD<br>GSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTTTGACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAGGCTCAGGGTGAATTGAAGGATAACTCTGATCCATCATGGGATT<br>CTTCAATCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATG<br>TTGATATTACTTACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATTGGATTC<br>ACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGT<br>ACATTACAAAGACAAGAA |
| b11_L3N_1c | CRAASLLPGTWQVTMTNEDG<br>VTSRGQMHFQPRSPYTLDVKA<br>QGSFDSNNDPAISGSTSISGKG<br>KVTCKTPDTMDVDITYSDGAQ<br>VQGQVTLDSPTQFKFDVTTSD<br>GSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGTTACTTCTAGAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAGGCTCAAGGTTCTTTCGATTCAAACAACGATCCAGCAATTTCTGG<br>TTCAACTTCTATCTCTGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATG<br>GATGTTGATATTACATACTCTGATGGTGCTCAAGTTCAAGGTCAAGTTACATTGG<br>ATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTAC<br>TGGTACATTACAAAGACAAGAA |
| b11_L3N_3c | CRAASLLPGTWQVTMTNEDG<br>VTSEGQMHFQPRSPYTLDVKA<br>QGTIKTNNDPNFKGTTDISGKG<br>KVTCKTPDTMDVDITYSDGIQ<br>VQGQVTLDSPTQFKFDVTTSD<br>GSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGTTACTTCTGAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTATTAAAACAAACAACGATCCAAACTTCAAAG<br>GTACTACAGATATCTCTGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTA<br>TGGATGTTGATATTACTTACTCTGATGGTATTCAAGTTCAAGGTCAAGTTACATT<br>GGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTT<br>ACTGGTACATTACAAAGACAAGAA |
| b11_L3N_4c | CRAASLLPGTWQVTMTNEDG<br>LTSQGQMHFQPRSPYTLDVKA<br>QGKLKSQSVPALRGSTSISGKG<br>KVTCKTPDTMDVDITYSDGM<br>QVQGQVTLDSPTQFKFDVTTS<br>DGSKVTGTLQRQE | TGTAGAGCTGCATCATTGTTACCAGGTACATGGCAAGTTACTATGACAAACGAA<br>GATGGTTTAACTTCACAAGGTCAAATGCATTTTCAACCAAGATCTCCATATACAT<br>TGGATGTTAAGGCTCAGGGTAAATTGAAGTCTCAATCAGTTCCAGCATTGAGAG<br>GTTCTACTTCAATCTCTGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTAT<br>GGATGTTGATATTACATACTCAGATGGTATGCAAGTTCAAGGTCAAGTTACATTG<br>GATTCTCCAACTCAATTCAAATTCGATGTTACTACATCAGATGGTTCTAAAGTTA<br>CTGGTACATTGCAAAGACAAGAA |
| b11_L3N_7c | CRAASLLPGTWQVTMTNEDG<br>ATSSGQMHFQPRSPYTLDVKA<br>QGTWESQDTPNAQGNQSISGK<br>GKVTCKTPDTMDVDITYSDG<br>MQVQGQVTLDSPTQFKFDVTT<br>SDGSKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGCTACTTCTTCAGGTCAAATGCATTTTCAACCAAGATCTCCATATACAT<br>TGGATGTTAAAGCTCAAGGTACATGGGAATCACAAGATACTCCAAATGCACAAG<br>GTAACCAATCTATCTCTGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTA<br>TGGATGTTGATATTACTTACTCTGATGGTATGCAAGTTCAAGGTCAAGTTACATT<br>GGATTCACCAACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTT<br>ACTGGTACATTACAAAGACAAGAA |
| b11_L3nm_4c | CRAASLLPGTWQVTMTNEDG<br>VTSQGQMHFQPRSPYTLDVKA<br>QGQGRSGKLKGNPIQGKGKVT<br>CKTPDTMDVDITYSHGMQVQ<br>GQVTLDSPTQFKFDVTTSDGS<br>KVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTCAAGGTAGATCTGGTAAATTGAAGGGTAACCCAA<br>TCCAGGGTAAAGGTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATA<br>TTACTTACTCTCATGGTATGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAAC<br>TCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTA<br>CAAAGACAAGAA |

| | | |
|---|---|---|
| b11_L5C_1c | CRAASLLPGTWQVTMTNEDG<br>QTSQGQMHFQPRSPYTLDVKA<br>QGTMSDGRPIQGKGKVTCKTP<br>DTMDVDIQYNINNGLRVQGQ<br>VTLDSPTQFKFDVTTSDGSKVT<br>GTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTCAAACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTATGTCAGATGGTAGACCAATCCAGGGTAAAG<br>GTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCCAATACAACA<br>TCAACAACGGTTTGAGAGTTCAAGGTCAAGTTACATTGGATTCTCCAACTCAATT<br>CAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGA<br>CAAGAA |
| b11_L5D_1c | CRAASLLPGTWQVTMTNEDG<br>QTSQGQMHFQPRSPYTLDVKA<br>QGTMSDGRPISGSGKVTCKTP<br>DTMDVDIQYGSALNGASVQG<br>QVTLDSPTQFKFDVTTSDGSK<br>VTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTCAAACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTATGTCAGATGGTAGACCAATTTCTGGTTCTGG<br>TAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTCAATACGGTTCT<br>GCTTTGAATGGTGCATCAGTTCAAGGTCAAGTTACATTGGATTCTCCAACTCAAT<br>TCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAG<br>ACAAGAA |
| b11_L5E_1c | CRAASLLPGTWQVTMTNEDG<br>QTSQGQMHFQPRSPYTLDVKA<br>QGTLSDGRPIQGSGKVTCKTP<br>DTMDVDIQYDPTAFKGKAKV<br>QGQVTLDSPTQFKFDVTTSDG<br>SKVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTCAAACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTTTATCTGATGGTAGACCAATTCAAGGTTCTGG<br>TAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTCAATACGATCCA<br>ACAGCTTTTAAAGGTAAAGCAAAAGTTCAAGGTCAAGTTACATTGGATTCACCA<br>ACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACAT<br>TACAAAGACAAGAA |
| b11_L5E_3c | CRAASLLPGTWQVTMTNEDG<br>QTSQGQMHFQPRSPYTLDVKA<br>QGTLSDGRPIKGSGKVTCKTP<br>DTMDVDIKYDPPAFNGTLRVQ<br>GQVTLDSPTQFKFDVTTSDGS<br>KVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTCAAACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTTTATCTGATGGTAGACCAATTAAAGGTTCTGG<br>TAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCAAGTACGATCC<br>ACCAGCTTTTAATGGTACATTGAGAGTTCAAGGTCAAGTTACATTGGATTCACCA<br>ACTCAATTCAAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACAT<br>TGCAAAGACAAGAA |
| b11_L5F_4c<br>(b11L5F) | CRAASLLPGTWQVTMTNEDG<br>QTSQGQMHFQPRSPYTLDVKA<br>QGTISDGRPISGKGKVTCKTPD<br>TMDVDITYPSLGNMKVQGQV<br>TLDSPTQFKFDVTTSDGSKVTG<br>TLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTCAAACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTATTTCTGATGGTAGACCAATCTCTGGTAAAG<br>GTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTACATACCCATC<br>TTTGGGTAACATGAAGGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTC<br>AAATTCGATGTTACTACATCTGATGGTTCAAAAGTTACTGGTACATTACAAAGAC<br>AAGAA |
| b11_L7_2c | CRAASLLPGTWQVTMTNEDG<br>TTSQGQMHFQPRSPYTLDVKA<br>QGTLSDGRPIQGKGKVTCKTP<br>DTMDVDITYSHGVQVQGQVT<br>LDSPTQFKFDVRSDGTGNTMT<br>GRVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAATGAA<br>GATGGTACTACATCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTTTATCTGATGGTAGACCAATCCAGGGTAAAG<br>GTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATTACATACTCACA<br>TGGTGTTCAAGTTCAAGGTCAAGTTACATTGGATTCTCCAACTCAATTCAAATTC<br>GATGTTAGATCAGATGGTACTGGTAATACTATGACAGGTAGAGTTACTGGTACA<br>TTACAAAGACAAGAA |
| b11_L7_5c | CRAASLLPGTWQVTMTNENG<br>VTSQGQMHFQPRSPYTLDVKA<br>QGTLSDGRPIQGKGKVTCKTP<br>DTMDVDITYSNGMQVQGQVT<br>LDSPTQFKFDVTTKGAGNTHT<br>GRVTGTLQRQE | TGTAGAGCTGCATCTTTGTTACCAGGTACATGGCAAGTTACTATGACAAACGAA<br>AACGGTGTTACTTCTCAAGGTCAAATGCATTTTCAACCAAGATCACCATATACAT<br>TGGATGTTAAAGCTCAAGGTACTTTATCAGATGGTAGACCAATCCAGGGTAAAG<br>GTAAAGTTACTTGTAAGACACCAGATACTATGGATGTTGATATCACATACTCTAA<br>TGGTATGCAAGTTCAAGGTCAAGTTACATTGGATTCACCAACTCAATTCAAATTC<br>GATGTTACTACAAAAGGTGCAGGTAATACTCATACAGGTAGAGTTACTGGTACA<br>TTACAAAGACAAGAA |

**Supplementary Table 7: List of oligos used for constructing b11L5F deep mutational scanning library.** Two oligos (R and F) were synthesized for each of 111 site-directed

mutagenesis PCR to construct the b11L5F single mutational scanning library. DNA oligo sequences are listed in this table.

| Name | Sequence |
|---|---|
| L5F_1R | CATATGGCTAGCCGACCCT |
| L5F_2R | ACACATATGGCTAGCCGACC |
| L5F_3R | TCTACACATATGGCTAGCCGAC |
| L5F_4R | AGCTCTACACATATGGCTAGCC |
| L5F_5R | TGCAGCTCTACACATATGGCT |
| L5F_6R | AGATGCAGCTCTACACATATGGC |
| L5F_7R | CAAAGATGCAGCTCTACACATATGG |
| L5F_8R | TAACAAAGATGCAGCTCTACACAT |
| L5F_9R | TGGTAACAAAGATGCAGCTCTA |
| L5F_10R | ACCTGGTAACAAAGATGCAGCT |
| L5F_11R | TGTACCTGGTAACAAAGATGCAG |
| L5F_12R | CCATGTACCTGGTAACAAAGATGC |
| L5F_13R | TTGCCATGTACCTGGTAACAAA |
| L5F_14R | AACTTGCCATGTACCTGGTAAC |
| L5F_15R | AGTAACTTGCCATGTACCTGGT |
| L5F_16R | CATAGTAACTTGCCATGTACCTGG |
| L5F_17R | TGTCATAGTAACTTGCCATGTACCT |
| L5F_18R | ATTTGTCATAGTAACTTGCCATGTACC |
| L5F_19R | TTCATTTGTCATAGTAACTTGCCATG |
| L5F_20R | ATCTTCATTTGTCATAGTAACTTGCCA |
| L5F_21R | ACCATCTTCATTTGTCATAGTAACTTGC |
| L5F_22R | TTGACCATCTTCATTTGTCATAGTAACT |
| L5F_23R | AGTTTGACCATCTTCATTTGTCATA |
| L5F_24R | AGAAGTTTGACCATCTTCATTTGTC |
| L5F_25R | TTGAGAAGTTTGACCATCTTCATTTG |
| L5F_26R | ACCTTGAGAAGTTTGACCATCTT |
| L5F_27R | TTGACCTTGAGAAGTTTGACCAT |
| L5F_28R | CATTTGACCTTGAGAAGTTTGACCA |
| L5F_29R | ATGCATTTGACCTTGAGAAGTTTG |
| L5F_30R | AAAATGCATTTGACCTTGAGAAGTT |
| L5F_31R | TTGAAAATGCATTTGACCTTGAGAA |
| L5F_32R | TGGTTGAAAATGCATTTGACCTT |
| L5F_33R | TCTTGGTTGAAAATGCATTTGACCT |
| L5F_34R | TGATCTTGGTTGAAAATGCATTTGA |
| L5F_35R | TGGTGATCTTGGTTGAAAATGCA |
| L5F_36R | ATATGGTGATCTTGGTTGAAAATGCA |
| L5F_37R | TGTATATGGTGATCTTGGTTGAAAATG |
| L5F_38R | CAATGTATATGGTGATCTTGGTTGAA |

| | |
|---|---|
| L5F_39R | ATCCAATGTATATGGTGATCTTGGTT |
| L5F_40R | AACATCCAATGTATATGGTGATCTTGG |
| L5F_41R | TTTAACATCCAATGTATATGGTGATCTTG |
| L5F_42R | AGCTTTAACATCCAATGTATATGGTGA |
| L5F_43R | TTGAGCTTTAACATCCAATGTATATGGT |
| L5F_44R | ACCTTGAGCTTTAACATCCAATGT |
| L5F_45R | AGTACCTTGAGCTTTAACATCCAA |
| L5F_46R | AATAGTACCTTGAGCTTTAACATCCAA |
| L5F_47R | AGAAATAGTACCTTGAGCTTTAACATCC |
| L5F_48R | ATCAGAAATAGTACCTTGAGCTTTAACA |
| L5F_49R | ACCATCAGAAATAGTACCTTGAGCT |
| L5F_50R | TCTACCATCAGAAATAGTACCTTGAGC |
| L5F_51R | TGGTCTACCATCAGAAATAGTACCT |
| L5F_52R | GATTGGTCTACCATCAGAAATAGTACC |
| L5F_53R | AGAGATTGGTCTACCATCAGAAATA |
| L5F_54R | ACCAGAGATTGGTCTACCATCAG |
| L5F_55R | TTTACCAGAGATTGGTCTACCATCA |
| L5F_56R | ACCTTTACCAGAGATTGGTCTACC |
| L5F_57R | TTTACCTTTACCAGAGATTGGTCTA |
| L5F_58R | AACTTTACCTTTACCAGAGATTGGTC |
| L5F_59R | AGTAACTTTACCTTTACCAGAGATTGGT |
| L5F_60R | ACAAGTAACTTTACCTTTACCAGAGAT |
| L5F_61R | CTTACAAGTAACTTTACCTTTACCAGAG |
| L5F_62R | TGTCTTACAAGTAACTTTACCTTTACCA |
| L5F_63R | TGGTGTCTTACAAGTAACTTTACCTT |
| L5F_64R | ATCTGGTGTCTTACAAGTAACTTTACC |
| L5F_65R | AGTATCTGGTGTCTTACAAGTAACTT |
| L5F_66R | CATAGTATCTGGTGTCTTACAAGTAACT |
| L5F_67R | ATCCATAGTATCTGGTGTCTTACAAG |
| L5F_68R | AACATCCATAGTATCTGGTGTCTTA |
| L5F_69R | ATCAACATCCATAGTATCTGGTGTC |
| L5F_70R | AATATCAACATCCATAGTATCTGGTGTC |
| L5F_71R | TGTAATATCAACATCCATAGTATCTGGTG |
| L5F_72R | GTATGTAATATCAACATCCATAGTATCTGGT |
| L5F_73R | TGGGTATGTAATATCAACATCCATAGTA |
| L5F_74R | AGATGGGTATGTAATATCAACATCCAT |
| L5F_75R | CAAAGATGGGTATGTAATATCAACATCCA |
| L5F_76R | ACCCAAAGATGGGTATGTAATATCAA |
| L5F_77R | GTTACCCAAAGATGGGTATGTAATATCA |
| L5F_78R | CATGTTACCCAAAGATGGGTATGT |
| L5F_79R | CTTCATGTTACCCAAAGATGGGT |

| | |
|---|---|
| L5F_80R | AACCTTCATGTTACCCAAAGATGG |
| L5F_81R | TTGAACCTTCATGTTACCCAAAGA |
| L5F_82R | ACCTTGAACCTTCATGTTACCCA |
| L5F_83R | TTGACCTTGAACCTTCATGTTACC |
| L5F_84R | AACTTGACCTTGAACCTTCATGTT |
| L5F_85R | TGTAACTTGACCTTGAACCTTCAT |
| L5F_86R | CAATGTAACTTGACCTTGAACCTT |
| L5F_87R | ATCCAATGTAACTTGACCTTGAACC |
| L5F_88R | TGAATCCAATGTAACTTGACCTTGA |
| L5F_89R | TGGTGAATCCAATGTAACTTGACC |
| L5F_90R | AGTTGGTGAATCCAATGTAACTTGA |
| L5F_91R | TTGAGTTGGTGAATCCAATGTAACT |
| L5F_92R | GAATTGAGTTGGTGAATCCAATGTA |
| L5F_93R | TTTGAATTGAGTTGGTGAATCCAAT |
| L5F_94R | GAATTTGAATTGAGTTGGTGAATCCA |
| L5F_95R | ATCGAATTTGAATTGAGTTGGTGAA |
| L5F_96R | AACATCGAATTTGAATTGAGTTGGTG |
| L5F_97R | AGTAACATCGAATTTGAATTGAGTTGG |
| L5F_98R | TGTAGTAACATCGAATTTGAATTGAGTT |
| L5F_99R | AGATGTAGTAACATCGAATTTGAATTGAG |
| L5F_100R | ATCAGATGTAGTAACATCGAATTTGAAT |
| L5F_101R | ACCATCAGATGTAGTAACATCGAAT |
| L5F_102R | TGAACCATCAGATGTAGTAACATCGA |
| L5F_103R | TTTTGAACCATCAGATGTAGTAACATC |
| L5F_104R | AACTTTTGAACCATCAGATGTAGTAAC |
| L5F_105R | AGTAACTTTTGAACCATCAGATGTAGT |
| L5F_106R | ACCAGTAACTTTTGAACCATCAGA |
| L5F_107R | TGTACCAGTAACTTTTGAACCATCA |
| L5F_108R | TAATGTACCAGTAACTTTTGAACCATC |
| L5F_109R | TTGTAATGTACCAGTAACTTTTGAACCA |
| L5F_110R | TCTTTGTAATGTACCAGTAACTTTTGAA |
| L5F_111R | TTGTCTTTGTAATGTACCAGTAACTTT |
| L5F_1 | AGGGTCGGCTAGCCATATGNNKAGAGCTGCATCTTTGTTACCA |
| L5F_2 | GGTCGGCTAGCCATATGTGTNNKGCTGCATCTTTGTTACCAGG |
| L5F_3 | GTCGGCTAGCCATATGTGTAGANNKGCATCTTTGTTACCAGGTACATG |
| L5F_4 | GGCTAGCCATATGTGTAGAGCTNNKTCTTTGTTACCAGGTACATGGC |
| L5F_5 | AGCCATATGTGTAGAGCTGCANNKTTGTTACCAGGTACATGGCAA |
| L5F_6 | GCCATATGTGTAGAGCTGCATCTNNKTTACCAGGTACATGGCAAGTT |
| L5F_7 | CCATATGTGTAGAGCTGCATCTTTGNNKCCAGGTACATGGCAAGTTACT |
| L5F_8 | ATGTGTAGAGCTGCATCTTTGTTANNKGGTACATGGCAAGTTACTATGAC |
| L5F_9 | TAGAGCTGCATCTTTGTTACCANNKACATGGCAAGTTACTATGACAAATG |

| | |
|---|---|
| L5F_10 | AGCTGCATCTTTGTTACCAGGTNNKTGGCAAGTTACTATGACAAATGAAG |
| L5F_11 | CTGCATCTTTGTTACCAGGTACANNKCAAGTTACTATGACAAATGAAGATGGT |
| L5F_12 | GCATCTTTGTTACCAGGTACATGGNNKGTTACTATGACAAATGAAGATGGTCA |
| L5F_13 | TTTGTTACCAGGTACATGGCAANNKACTATGACAAATGAAGATGGTCAAA |
| L5F_14 | GTTACCAGGTACATGGCAAGTTNNKATGACAAATGAAGATGGTCAAACT |
| L5F_15 | ACCAGGTACATGGCAAGTTACTNNKACAAATGAAGATGGTCAAACTTCT |
| L5F_16 | CCAGGTACATGGCAAGTTACTATGNNKAATGAAGATGGTCAAACTTCTCAAG |
| L5F_17 | AGGTACATGGCAAGTTACTATGACANNKGAAGATGGTCAAACTTCTCAAGG |
| L5F_18 | GGTACATGGCAAGTTACTATGACAAATNNKGATGGTCAAACTTCTCAAGGTC |
| L5F_19 | CATGGCAAGTTACTATGACAAATGAANNKGGTCAAACTTCTCAAGGTCAAAT |
| L5F_20 | TGGCAAGTTACTATGACAAATGAAGATNNKCAAACTTCTCAAGGTCAAATGCA |
| L5F_21 | GCAAGTTACTATGACAAATGAAGATGGTNNKACTTCTCAAGGTCAAATGCATTT |
| L5F_22 | AGTTACTATGACAAATGAAGATGGTCAANNKTCTCAAGGTCAAATGCATTTTCA |
| L5F_23 | TATGACAAATGAAGATGGTCAAACTNNKCAAGGTCAAATGCATTTTCAACC |
| L5F_24 | GACAAATGAAGATGGTCAAACTTCTNNKGGTCAAATGCATTTTCAACCAAG |
| L5F_25 | CAAATGAAGATGGTCAAACTTCTCAANNKCAAATGCATTTTCAACCAAGATCA |
| L5F_26 | AAGATGGTCAAACTTCTCAAGGTNNKATGCATTTTCAACCAAGATCACC |
| L5F_27 | ATGGTCAAACTTCTCAAGGTCAANNKCATTTTCAACCAAGATCACCATATAC |
| L5F_28 | TGGTCAAACTTCTCAAGGTCAAATGNNKTTTCAACCAAGATCACCATATACATT |
| L5F_29 | CAAACTTCTCAAGGTCAAATGCATNNKCAACCAAGATCACCATATACATTGG |
| L5F_30 | AACTTCTCAAGGTCAAATGCATTTTNNKCCAAGATCACCATATACATTGGATG |
| L5F_31 | TTCTCAAGGTCAAATGCATTTTCAANNKAGATCACCATATACATTGGATGTTAAAG |
| L5F_32 | AAGGTCAAATGCATTTTCAACCANNKTCACCATATACATTGGATGTTAAAGC |
| L5F_33 | AGGTCAAATGCATTTTCAACCAAGANNKCCATATACATTGGATGTTAAAGCTCA |
| L5F_34 | TCAAATGCATTTTCAACCAAGATCANNKTATACATTGGATGTTAAAGCTCAAGG |
| L5F_35 | TGCATTTTCAACCAAGATCACCANNKACATTGGATGTTAAAGCTCAAGG |
| L5F_36 | TGCATTTTCAACCAAGATCACCATATNNKTTGGATGTTAAAGCTCAAGGTACT |
| L5F_37 | CATTTTCAACCAAGATCACCATATACANNKGATGTTAAAGCTCAAGGTACTATTTCT |
| L5F_38 | TTCAACCAAGATCACCATATACATTGNNKGTTAAAGCTCAAGGTACTATTTCTGA |
| L5F_39 | AACCAAGATCACCATATACATTGGATNNKAAAGCTCAAGGTACTATTTCTGATG |
| L5F_40 | CCAAGATCACCATATACATTGGATGTTNNKGCTCAAGGTACTATTTCTGATGGT |
| L5F_41 | CAAGATCACCATATACATTGGATGTTAAANNKCAAGGTACTATTTCTGATGGTAGACC |
| L5F_42 | TCACCATATACATTGGATGTTAAAGCTNNKGGTACTATTTCTGATGGTAGACCA |
| L5F_43 | ACCATATACATTGGATGTTAAAGCTCAANNKACTATTTCTGATGGTAGACCAATCT |
| L5F_44 | ACATTGGATGTTAAAGCTCAAGGTNNKATTTCTGATGGTAGACCAATCTCT |
| L5F_45 | TTGGATGTTAAAGCTCAAGGTACTNNKTCTGATGGTAGACCAATCTCTGG |
| L5F_46 | TTGGATGTTAAAGCTCAAGGTACTATTNNKGATGGTAGACCAATCTCTGGTAA |
| L5F_47 | GGATGTTAAAGCTCAAGGTACTATTTCTNNKGGTAGACCAATCTCTGGTAAAGG |
| L5F_48 | TGTTAAAGCTCAAGGTACTATTTCTGATNNKAGACCAATCTCTGGTAAAGGTAAA |
| L5F_49 | AGCTCAAGGTACTATTTCTGATGGTNNKCCAATCTCTGGTAAAGGTAAAGTTAC |
| L5F_50 | GCTCAAGGTACTATTTCTGATGGTAGANNKATCTCTGGTAAAGGTAAAGTTACTTGT |

| | |
|---|---|
| L5F_51 | AGGTACTATTTCTGATGGTAGACCANNKTCTGGTAAAGGTAAAGTTACTTGTAAG |
| L5F_52 | GGTACTATTTCTGATGGTAGACCAATCNNKGGTAAAGGTAAAGTTACTTGTAAGACA |
| L5F_53 | TATTTCTGATGGTAGACCAATCTCTNNKAAAGGTAAAGTTACTTGTAAGACACC |
| L5F_54 | CTGATGGTAGACCAATCTCTGGTNNKGGTAAAGTTACTTGTAAGACACCAG |
| L5F_55 | TGATGGTAGACCAATCTCTGGTAAANNKAAAGTTACTTGTAAGACACCAGATAC |
| L5F_56 | GGTAGACCAATCTCTGGTAAAGGTNNKGTTACTTGTAAGACACCAGATACTATG |
| L5F_57 | TAGACCAATCTCTGGTAAAGGTAAANNKACTTGTAAGACACCAGATACTATGG |
| L5F_58 | GACCAATCTCTGGTAAAGGTAAAGTTNNKTGTAAGACACCAGATACTATGGATG |
| L5F_59 | ACCAATCTCTGGTAAAGGTAAAGTTACTNNKAAGACACCAGATACTATGGATGTT |
| L5F_60 | ATCTCTGGTAAAGGTAAAGTTACTTGTNNKACACCAGATACTATGGATGTTGAT |
| L5F_61 | CTCTGGTAAAGGTAAAGTTACTTGTAAGNNKCCAGATACTATGGATGTTGATATTACATAC |
| L5F_62 | TGGTAAAGGTAAAGTTACTTGTAAGACANNKGATACTATGGATGTTGATATTACATACCC |
| L5F_63 | AAGGTAAAGTTACTTGTAAGACACCANNKACTATGGATGTTGATATTACATACCCA |
| L5F_64 | GGTAAAGTTACTTGTAAGACACCAGATNNKATGGATGTTGATATTACATACCCATCT |
| L5F_65 | AAGTTACTTGTAAGACACCAGATACTNNKGATGTTGATATTACATACCCATCTTTGG |
| L5F_66 | AGTTACTTGTAAGACACCAGATACTATGNNKGTTGATATTACATACCCATCTTTGGG |
| L5F_67 | CTTGTAAGACACCAGATACTATGGATNNKGATATTACATACCCATCTTTGGGTAAC |
| L5F_68 | TAAGACACCAGATACTATGGATGTTNNKATTACATACCCATCTTTGGGTAACA |
| L5F_69 | GACACCAGATACTATGGATGTTGATNNKACATACCCATCTTTGGGTAACA |
| L5F_70 | GACACCAGATACTATGGATGTTGATATTNNKTACCCATCTTTGGGTAACATGAA |
| L5F_71 | CACCAGATACTATGGATGTTGATATTACANNKCCATCTTTGGGTAACATGAAGG |
| L5F_72 | ACCAGATACTATGGATGTTGATATTACATACNNKTCTTTGGGTAACATGAAGGTTCA |
| L5F_73 | TACTATGGATGTTGATATTACATACCCANNKTTGGGTAACATGAAGGTTCAAGG |
| L5F_74 | ATGGATGTTGATATTACATACCCATCTNNKGGTAACATGAAGGTTCAAGGTC |
| L5F_75 | TGGATGTTGATATTACATACCCATCTTTGNNKAACATGAAGGTTCAAGGTCAAGT |
| L5F_76 | TTGATATTACATACCCATCTTTGGGTNNKATGAAGGTTCAAGGTCAAGTTACA |
| L5F_77 | TGATATTACATACCCATCTTTGGGTAACNNKAAGGTTCAAGGTCAAGTTACATTG |
| L5F_78 | ACATACCCATCTTTGGGTAACATGNNKGTTCAAGGTCAAGTTACATTGGA |
| L5F_79 | ACCCATCTTTGGGTAACATGAAGNNKCAAGGTCAAGTTACATTGGATTCA |
| L5F_80 | CCATCTTTGGGTAACATGAAGGTTNNKGGTCAAGTTACATTGGATTCACC |
| L5F_81 | TCTTTGGGTAACATGAAGGTTCAANNKCAAGTTACATTGGATTCACCAACT |
| L5F_82 | TGGGTAACATGAAGGTTCAAGGTNNKGTTACATTGGATTCACCAACTCA |
| L5F_83 | GGTAACATGAAGGTTCAAGGTCAANNKACATTGGATTCACCAACTCAATTC |
| L5F_84 | AACATGAAGGTTCAAGGTCAAGTTNNKTTGGATTCACCAACTCAATTCAAA |
| L5F_85 | ATGAAGGTTCAAGGTCAAGTTACANNKGATTCACCAACTCAATTCAAATTCG |
| L5F_86 | AAGGTTCAAGGTCAAGTTACATTGNNKTCACCAACTCAATTCAAATTCGA |
| L5F_87 | GGTTCAAGGTCAAGTTACATTGGATNNKCCAACTCAATTCAAATTCGATGTTAC |
| L5F_88 | TCAAGGTCAAGTTACATTGGATTCANNKACTCAATTCAAATTCGATGTTACTACA |
| L5F_89 | GGTCAAGTTACATTGGATTCACCANNKCAATTCAAATTCGATGTTACTACATCTG |
| L5F_90 | TCAAGTTACATTGGATTCACCAACTNNKTTCAAATTCGATGTTACTACATCTGA |
| L5F_91 | AGTTACATTGGATTCACCAACTCAANNKAAATTCGATGTTACTACATCTGATGG |

| | |
|---|---|
| L5F_92 | TACATTGGATTCACCAACTCAATTCNNKTTCGATGTTACTACATCTGATGGT |
| L5F_93 | ATTGGATTCACCAACTCAATTCAAANNKGATGTTACTACATCTGATGGTTCAA |
| L5F_94 | TGGATTCACCAACTCAATTCAAATTCNNKGTTACTACATCTGATGGTTCAAAAGT |
| L5F_95 | TTCACCAACTCAATTCAAATTCGATNNKACTACATCTGATGGTTCAAAAGTTAC |
| L5F_96 | CACCAACTCAATTCAAATTCGATGTTNNKACATCTGATGGTTCAAAAGTTACTG |
| L5F_97 | CCAACTCAATTCAAATTCGATGTTACTNNKTCTGATGGTTCAAAAGTTACTGGT |
| L5F_98 | AACTCAATTCAAATTCGATGTTACTACANNKGATGGTTCAAAAGTTACTGGTACA |
| L5F_99 | CTCAATTCAAATTCGATGTTACTACATCTNNKGGTTCAAAAGTTACTGGTACATTACA |
| L5F_100 | ATTCAAATTCGATGTTACTACATCTGATNNKTCAAAAGTTACTGGTACATTACAAAGA |
| L5F_101 | ATTCGATGTTACTACATCTGATGGTNNKAAAGTTACTGGTACATTACAAAGACA |
| L5F_102 | TCGATGTTACTACATCTGATGGTTCANNKGTTACTGGTACATTACAAAGACAAGA |
| L5F_103 | GATGTTACTACATCTGATGGTTCAAAANNKACTGGTACATTACAAAGACAAGAAC |
| L5F_104 | GTTACTACATCTGATGGTTCAAAAGTTNNKGGTACATTACAAAGACAAGAACTCG |
| L5F_105 | ACTACATCTGATGGTTCAAAAGTTACTNNKACATTACAAAGACAAGAACTCGAG |
| L5F_106 | TCTGATGGTTCAAAAGTTACTGGTNNKTTACAAAGACAAGAACTCGAGGG |
| L5F_107 | TGATGGTTCAAAAGTTACTGGTACANNKCAAAGACAAGAACTCGAGGGA |
| L5F_108 | GATGGTTCAAAAGTTACTGGTACATTANNKAGACAAGAACTCGAGGGAGG |
| L5F_109 | TGGTTCAAAAGTTACTGGTACATTACAANNKCAAGAACTCGAGGGAGGC |
| L5F_110 | TTCAAAAGTTACTGGTACATTACAAAGANNKGAACTCGAGGGAGGCGG |
| L5F_111 | AAAGTTACTGGTACATTACAAAGACAANNKCTCGAGGGAGGCGGAT |

**Supplementary Table 8: List of primers used for sequencing b11L5F libraires by Illumina Miseq sequencer.** Two rounds of PCR were performed to amplify genes prior to illumina chip sequencing. DNA primer sequences and the purpose of usage are summarized in this table. Barcode sequences are lower case.

| **Name** | **Sequence** | **Purpose** |
|---|---|---|
| pETCON_miseq_offset0_f | TCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGGGTCGGCTAGCCATATG | amplifying genes from pECTON2 open reading frame |
| pETCON_miseq_offset1_f | TCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGGGTCGGCTAGCCATATG | |
| pETCON_miseq_offset2_f | TCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGGGTCGGCTAGCCATATG | |
| pETCON_miseq_offset3_f | TCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGGGTCGGCTAGCCATATG | |
| pETCON_miseq_offset0_r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGATCCGCCCCCCTCGAG | |
| pETCON_miseq_offset1_r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGATCCGCCCCCCTCGAG | |
| pETCON_miseq_offset2_r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGATCCGCCCCCCTCGAG | |
| pETCON_miseq_offset3_r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGATCCGCCCCCCTCGAG | |
| miseq_start_adapt | AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCT | attaching Illumina adapter to the 5'-end |
| TSBC_26 | CAAGCAGAAGACGGCATACGAGAT gctcat GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Naïve_1 |
| TSBC_27 | CAAGCAGAAGACGGCATACGAGAT aggaat GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library FITC_1 |

| | | |
|---|---|---|
| TSBC_28 | CAAGCAGAAGACGGCATACGAGAT cttttg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Bind_1 |
| TSBC_33 | CAAGCAGAAGACGGCATACGAGAT cgcctg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T9_1 |
| TSBC_34 | CAAGCAGAAGACGGCATACGAGAT gccatg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T27_1 |
| TSBC_35 | CAAGCAGAAGACGGCATACGAGAT aaaatg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T81_1 |
| TSBC_36 | CAAGCAGAAGACGGCATACGAGAT tgttgg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T243_1 |
| TSBC_37 | CAAGCAGAAGACGGCATACGAGAT attccg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch9_1 |
| TSBC_38 | CAAGCAGAAGACGGCATACGAGAT agctag GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch27_1 |
| TSBC_39 | CAAGCAGAAGACGGCATACGAGAT gtatag GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch81_1 |
| TSBC_40 | CAAGCAGAAGACGGCATACGAGAT tctgag GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch243_1 |
| TSBC_29 | CAAGCAGAAGACGGCATACGAGAT tagttg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library naïve_2 |
| TSBC_30 | CAAGCAGAAGACGGCATACGAGAT ccggtg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library FITC_2 |
| TSBC_31 | CAAGCAGAAGACGGCATACGAGAT atcgtg GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Bind_2 |
| TSBC_41 | CAAGCAGAAGACGGCATACGAGAT gtcgtc GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T9_2 |
| TSBC_42 | CAAGCAGAAGACGGCATACGAGAT cgatta GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T27_2 |
| TSBC_43 | CAAGCAGAAGACGGCATACGAGAT gctgta GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T81_2 |
| TSBC_44 | CAAGCAGAAGACGGCATACGAGAT attata GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library T243_2 |
| TSBC_45 | CAAGCAGAAGACGGCATACGAGAT gaatga GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch9_2 |
| TSBC_46 | CAAGCAGAAGACGGCATACGAGAT tcggga GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch27_2 |
| TSBC_47 | CAAGCAGAAGACGGCATACGAGAT cttcga GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch81_2 |
| TSBC_48 | CAAGCAGAAGACGGCATACGAGAT tgccga GTGACTGGAGTTCAGACGTGTGCTCTTCCG | barcoding library Ch243_2 |

**Supplementary Table 9: List of sequences for designs from the 2nd round of design calculation based on the lowest-energy ligand docking model of b11L5F.1.** Protein sequences and DNA encoding sequences (optimized for *E.coli* codon usage) of five designs (nC1-5) are provided in this table.

| **Design ID** | **Protein Sequence** | **DNA sequence** |
|---|---|---|
| b11L5F_M10_noCC1 (nC1) | NRAYRMLPGTWQVTMTNED GQTSQGQMHIQPRSPYTLDV VAQGTISDGRPISGYGKVTVK TPDTLQVHITYPSLGNIKVQG QITLDSPTQFTFNATTSDGKN LTGTLQRQE | AACCGTGCGTACCGTATGCTGCCGGGTACCTGGCAGGTTACCATGACCAACGAAGAC GGTCAGACCTCTCAAGGTCAGATGCACATCCAGCCGCGTTCTCCGTATACCCTGGAC GTTGTTGCGCAGGGTACTATCTCCGATGGTCGTCCGATCTCTGGTTACGGTAAAGTTA CCGTTAAAACCCCGGACACCCTCCAGGTTCACATCACCTACCCGTCTCTGGGTAACAT CAAAGTTCAGGGCCAGATCACTCTGGACTCTCCGACCCAGTTCACCTTCAACGCGAC CACCTCTGATGGTAAAAACCTGACCGGTACTCTCCAACGTCAGGAA |
| b11L5F_M10_noCC2 (nC2) | NRAASLLPGTWQVTMTNEDG QTSQGQMHIQPRSPYTLDVV AQGTISDGRPISGYGKVTVKT PDTMQVHITYPSLGNIKVQGQ ITLDSPTQFTFNATTSDGKNLT GTLQRQE | AACCGTGCGGCTTCTCTGCTGCCGGGTACCTGGCAAGTTACTATGACCAACGAAGAC GGTCAGACCTCTCAAGGTCAGATGCACATCCAGCCGCGTTCTCCGTATACCCTGGAC GTTGTTGCGCAGGGTACCATCTCTGATGGTCGTCCGATCTCTGGTTACGGTAAGGTTA CCGTTAAAACCCCGGACACCATGCAAGTTCACATCACCTACCCGTCTCTGGGTAACA TCAAAGTTCAGGGCCAGATCACCCTCGACTCTCCGACCCAGTTCACCTTCAACGCGA CCACCTCTGACGGTAAAAACCTGACCGGTACTCTGCAACGTCAGGAA |
| b11L5F_M10_noCC3 (nC3) | NRAAANLPGTWQVTMTNED GQTSQGQMHFQPRSPYTLDV VAQGTISDGRPISGYGKVTVK TPDTMNVDITYPSLGNIKVQG QITLDSPTQFTFNATTSDGKK LTGTIQRQE | AACCGTGCGGCTGCGAATCTGCCGGGTACCTGGCAGGTTACCATGACCAACGAAGAC GGTCAGACCTCTCAGGGCCAGATGCACTTCCAGCCGCGTTCTCCGTACACCCTGGAT GTTGTTGCGCAGGGCACTATCTCTGACGGTCGTCCGATCTCTGGTTACGGTAAAGTTA CCGTTAAAACCCCGGACACCATGAACGTTGACATCACCTACCCGTCTCTGGGTAACA TCAAAGTTCAGGGTCAGATCACCCTCGACTCTCCGACCCAGTTCACCTTCAACGCGA CCACCTCTGATGGCAAAAAACTGACCGGCACCATTCAGCGTCAGGAA |
| b11L5F_M10_noCC4 (nC4) | NTAIANLPGTWQVTMTNEDG QTSQGQMHIQPRSPYTADVV AQGTISDGRPISGYGKLTAKT PDTVNVQITYPSLGNINVQGQI TNDSPTQAHFNATTSDGKKLT GTMQRQE | AACACCGCGATTGCGAATCTGCCGGGTACCTGGCAAGTGACCATGACCAATGAGGAC GGTCAGACCTCTCAGGGCCAGATGCACATCCAGCCGCGTTCTCCGTACACCGCTGAC GTTGTTGCGCAGGGTACCATCTCTGATGGTCGTCCGATCTCTGGTTACGGTAAGCTGA CCGCGAAAACCCCGGACACCGTGAACGTGCAGATTACTTACCCGTCTCTGGGTAACA TCAATGTGCAAGGTCAGATCACCAACGACTCTCCGACCCAGGCGCACTTCAACGCGA CCACCTCTGACGGTAAAAAACTCACCGGTACTATGCAGCGTCAGGAA |

| | | |
|---|---|---|
| b11L5F_M10_noCC5 (nC5 -> **b11L5F.2**) | NRAAQLLPGTWQVTMTNED<br>GQTSQGQMHFQPRSPYTLDIV<br>AQGTISDGRPISGYGKVTVKT<br>PDTMHVNITYPSLGNIKVQGQ<br>ITLDSPTQFTWNSTTSDGKKL<br>TGTLQRQE | AACCGTGCGGCGCAGCTCCTGCCGGGTACCTGGCAGGTTACCATGACCAACGAAGAC<br>GGTCAGACCTCCCAGGGCCAGATGCACTTCCAGCCGCGTTCTCCGTACACCCTGGAC<br>ATCGTTGCGCAGGGCACCATTTCTGACGGTCGTCCGATCTCTGGTTACGGTAAAGTTA<br>CCGTTAAAACCCCGGACACCATGCACGTTAACATTACCTACCCGTCTCTGGGTAACA<br>TCAAAGTTCAGGGTCAAATCACCCTCGACTCTCCGACCCAGTTCACCTGGAACTCTAC<br>TACCTCTGATGGTAAAAAACTGACCGGTACTCTGCAACGTCAGGAA |

**Supplementary Table 10: List of oligos used for constructing b11L5F.1 and b11L5F.2 combinatorial libraries.** DNA assembly method was used to construct the combinatorial libraries (see Methods). DNA oligo sequences and the purpose of usage are summarized in this table. Mutations and doping ratios are highlighted in bold using the degenerate letters defined on IDT website (https://www.idtdna.com).

| Name | Sequence | Purpose |
|---|---|---|
| oligo1_rev | CAGGTACCCGGCAGGAGCTGCGC**S(R1:01009900)(H1:01980001)**ACGGCTCATATGGCTAGCCGACCCTC | assembling b11L5F.1 and b11L5F.2 combinatorial library |
| oligo2_for | CAGCTCCTGCCGGGTACCTGGCAG**(D1:01009801)(Y2:00010099)S**ACC**(D3:98000101)(Y2:00010099)(S3:00019900)**ACCAACGAAGACGGTCAGACCTCCCAG | assembling b11L5F.1 and b11L5F.2 combinatorial library |
| oligo3_rev | GGTGTACGGAGAACGCGGCTG**(S3:00019900)(R2:99000100)(H2:98010001)**GTG(**S2:00990100)(R2:99000100)(H3:01010098)**CTGGCCCTGGGAGGTCTGACCGTCTTC | assembling b11L5F.1 and b11L5F.2 combinatorial library |
| oligo4_for | AGCCGCGTTCTCCGTACACC**(D2:01000198)(H3:01010098)(S3:00019900)**GAC**RTA**GTTGCG**(S2:00990100)(W1:99000001)A**GGCACCATTTCTGACGGTCGTC | assembling b11L5F.2 combinatorial library |
| oligo5_rev | TCGGATCCGCCTCCCTCGAGTTCCTGACGTTG(**S2:00990100)(R2:99000100)(H2:98010001)**AGTACCGGTCAGTTTTTTACCATC | assembling b11L5F.2 combinatorial library |
| noCC5_ultramer1-1_rev | TTTGATGTTACCCAGAGACGGGTAGGTTABGTTTRCGTGYAKGGTGTC**KTC**GGTTTTAACGGTTRCTTTACCGTAACCGGTGATCGGACGACCGTCAGAAATGGTGCC | assembling b11L5F.2 combinatorial library |
| noCC5_ultramer1-2_rev | TTTGATGTTACCCAGAGACGGGTAGGTTABGTTTRCGTGYAKGGTGTC**CGG**GGTTTTAACGGTTRCTTTACCGTAACCGGTGATCGGACGACCGTCAGAAATGGTGCC | assembling b11L5F.2 combinatorial library |
| noCC5_ultramer2-1_for | CGTCTCTGGGTAACATCAAA**RYA**CAGGGTCAA**RTA**ACC**MTG**GACTCTCCGACCCAG**KYC**ACCTGGAACTCTACTACC**KCA**GATGGTAAAAAACTGACCGGTAC | assembling b11L5F.2 combinatorial library |
| noCC5_ultramer2-2_for | CGTCTCTGGGTAACATCAAA**RYA**CAGGGTCAA**RTA**ACC**TAC**GACTCTCCGACCCAG**KYC**ACCTGGAACTCTACTACC**KCA**GATGGTAAAAAACTGACCGGTAC | assembling b11L5F.2 combinatorial library |
| sv_oligo4_for | CAGCCGCGTTCTCCGTACACC**MTG**GACGTTGTTGCTCAGGGTACCATC | assembling b11L5F.1 combinatorial library |
| loop3_wt_rev | GTAACTTTACCGTAACCAGAGATCGGACGACCGTCAGAGATGGTACCCTGAGCAACAACG | assembling b11L5F.1 combinatorial library |
| sv_loop5_ultramer-1_for | CACCGGTTACGGTAAAGTTACCGTTAAAACC**GAM**GACACC**MTR**GAC**GYA**GACATCACCTACCCGTCTCTGGGTAAC**ATC**AAAGTTCAGGGTCAG**RTA**ACC**MTG**GACTCTCCGACCCAGTTCAAATTCGACGCAACC | assembling b11L5F.1 combinatorial library |
| sv_loop5_ultramer-2_for | CACCGGTTACGGTAAAGTTACCGTTAAAACC**CCG**GACACC**MTR**GAC**GYA**GACATCACCTACCCGTCTCTGGGTAACATCAAAGTTCAGGGTCAG**RTA**ACC**MTG**GACTCTCCGACCCAGTTCAAATTCGACGCAACC | assembling b11L5F.1 combinatorial library |
| sv_loop7_wt_rev | CGCTGCAGGGTACCGGTCAGTTTTTTACCGTCAGAGGTGGTTGCGTCGAATTTGAACTG | assembling b11L5F.1 combinatorial library |

| | | |
|---|---|---|
| sv_loop7_variants_rev | **GCTGCAGGGTACCGGTCAGACGACCGGTSAHGTTACCCGCACCTTTGS TGGTTGCGTCGAATTTGAACTG** | assembling b11L5F.1 combinatorial library |
| sv_end_for | CTGACCGGTACCCTGCAGCGTCAGGAA***CTCGAG***GGAGGCGGATCCGAACA AAAGC | assembling b11L5F.1 combinatorial library |
| forward_amp | TGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATG | amplifying pETCON2 gene reading frame |
| reverse_amp | CTTCAGAAATAAGCTTTTGTTCGGATCCGCCTCCCTCGAG | amplifying pETCON2 gene reading frame |

**Supplementary Table 11: List of primers for b11L5F.1-based error prone library.** Two adjacent primers were synthesized for constructing b11L5F.1 error prone library. DNA primer sequences and the purpose of usage are summarized in this table.

| Name | Sequence | Purpose |
|---|---|---|
| Up_ATG_ptc_for | GAGGCGGAGGGTCGGCTAGCCATATG | error-prone PCR |
| Down_Xhol_ptc_rev | GCTTTTGTTCGGATCCGCCCCCCTCGAG | error-prone PCR |

**Supplementary Table 12: List of primers for cloning mammalian subcellular targeting tags.** Site-directed primers were synthesized for fusing mammalian subcellular targeting tags to mFAP genes. DNA primer sequences and the purpose of usage are summarized in this table.

| Name | Sequence | Purpose |
|---|---|---|
| Nt_Tom20_for | CTTAAGCTTGGTACCGAGCTCGCCACCATGGTAGGCCG GAACAGTGCAATC | cloning Tom20 to N-term of mFAP |
| Nt_Tom20_mFAP_rev | CAGCAGCTGGGCGGCGCGGCTACCAGAGCCGAAGTTG GGGTCTG | cloning Tom20 to N-term of mFAP |
| Nt_3NLS_for | CTTAAGCTTGGTACCGAGCTCGCCACCATGGATCCTAA GAAAAAGCGCAAG | cloning nucleus-targeting sequence to N-term of mFAP |
| Nt_3NLS_mFAP_rev | CAGCAGCTGGGCGGCGCGGCTACCAGAGCCTACCTTCC GCTTCTTC | cloning nucleus-targeting sequence to N-term of mFAP |
| Nt_mts_for | CTTAAGCTTGGTACCGAGCTCGCCACCATGTCCGTCCT GACGCCGCTGCTG | cloning mitochondrial-targeting sequence to N-term of mFAP |
| Nt_mts_mFAP_rev | CAGCAGCTGGGCGGCGCGGCTACCAGAGCCGACCGGT GGATCCCCCAACGAATG | cloning mitochondrial-targeting sequence to N-term of mFAP |
| Ct_CAAX_mFAP_for | GACCGGCACCCTGCAGCGCCAGGAGCGAAAACATAAA GAAAAGATGAGCAA | cloning membrane-targeting sequence to C-term of mFAP |
| Ct_CAAX_rev | GAGCGGCCGCCACTGTGCTGGATTTATCACATAATTAC ACACTTTGTCTTTG | cloning membrane-targeting sequence to C-term of mFAP |
| Ct_Sec61b_mFAP_for | GACCGGCACCCTGCAGCGCCAGGAGATGCCTGGTCCG ACCCCCAG | cloning ER-targeting protein to C-term of mFAP |
| Ct_Sec61b_rev | GAGCGGCCGCCACTGTGCTGGATTTATCACGAACGAGT GTACTTGCCCCAA | cloning ER-targeting protein to C-term of mFAP |
| pcDNA5_start_for | CTTAAGCTTGGTACCGAGCTCGCCACCATG | amplifying pcDNA5/FRT/TO reading frame gene |
| mFAP_for | AGCCGCGCCGCCCAGCTGCTG | amplifying mFAP1 and mFAP2 |
| mFAP_rev | CTCCTGGCGCTGCAGGGTGCCGGTC | amplifying mFAP1 and mFAP2 |
| pcDNA5_stop_rev | GAGCGGCCGCCACTGTGCTGGATTTATCA | Amplifying pcDNA5/FRT/TO reading frame gene |

**Supplementary Table 13: List of oligos used for general cloning and site-directed mutagenesis.** Additional DNA oligos were synthesized for subcloning and mutagenesis purposes. DNA sequences and the purpose of usage were summarized in this table. Mutation site are highlighted in bold.

| Name | Sequence | Purpose |
|---|---|---|
| pETCON_SV_for | GAGGCGGAGGGTCGGCTAGCCATATGAGCCGTGCTGCTTCTCTG | cloning b11L5F.1 from e.coli plasmids to yeast plasmid |
| pETCON_SV_rev | GCTTTTGTTCGGATCCGCCCCCCTCGAGTTCCTGACGCTGCAGGGTAC | cloning b11L5F.1 from e.coli plasmids to yeast plasmid |
| SV_M77I_rev | GTTACCCAGAGACGGGTAG | |
| SV_M77I_for | CTACCCGTCTCTGGGTAAC**ATC**AAAGTTCAGGGTCAGATCAC | making b11L5F.1_M77I mutant |
| SV_S101K_rev | ACCGTCAGAGGTGGTTGCGTC | |
| SV_S101K_for | GACGCAACCACCTCTGACGGT**AAG**AAACTGACCGGTACCCTGCAG | making b11L5F.1_S101K mutant |
| nCC5_N1S_for | CTGGTGCCGCGCGGCAGCTCC**TCT**CGTGCGGCGCAGCTCCTG | |
| nCC5_N1S_rev | GGAGCTGCCGCGCGGCACCAG | making b11L5F.2_N1S mutant |
| nCC5_I39V_for | GTTCTCCGTACACCCTGGAC**GTT**GTTGCGCAGGGCACCATTTCTG | |
| nCC5_I39V_rev | GTCCAGGGTGTACGGAGAAC | making b11L5F.2_I39V mutant |
| nCC5_T96V_for | GACCCAGTTCACCTGGAACTCT**GTT**ACCTCTGATGGTAAAAAACTGAC | |
| nCC5_T96V_rev | AGAGTTCCAGGTGAACTGGGTC | making b11L5F.2_T96V mutant |
| pCDB24_for | CATTGAAGCCCACCGTGAACAGATTG | amplifiying the gene with N-term sumo tag |
| b11L5F_before83_rev | CTGACCCTGAACTTTCATGTTAC | |
| b11L5F_before103_rev | TTTAGAACCGTCAGAGGTGGTAAC | making b11L5F mutations |
| b11L5F_83I_95G_for | GTAACATGAAAGTTCAGGGTCAG**ATC**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GGT**ACCACCTCTGACGGTTCTAAAG | making b11L5F_83I_95G and b11L5F_83I_95G_103L variants |
| b11L5F_83I_95A_for | GTAACATGAAAGTTCAGGGTCAG**ATC**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GCA**ACCACCTCTGACGGTTCTAAAG | making b11L5F_83I_95A and b11L5F_83I_95A_103L variants |
| b11L5F_83L_95G_for | GTAACATGAAAGTTCAGGGTCAG**CTG**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GGT**ACCACCTCTGACGGTTCTAAAG | making b11L5F_83L_95G variant |
| b11L5F_83L_95A_for | GTAACATGAAAGTTCAGGGTCAG**CTG**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GCA**ACCACCTCTGACGGTTCTAAAG | making b11L5F_83L_95A and b11L5F_83L_95G_103L variants |
| b11L5F_83M_95G_for | GTAACATGAAAGTTCAGGGTCAG**ATG**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GGT**ACCACCTCTGACGGTTCTAAAG | making b11L5F_83M_95G and b11L5F_83M_95G_103L variants |
| b11L5F_83M_95A_for | GTAACATGAAAGTTCAGGGTCAG**ATG**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GCA**ACCACCTCTGACGGTTCTAAAG | making b11L5F_83M_95A and b11L5F_83M_95G_103L variants |
| b11L5F_83M_95G_103L_for | GTAACATGAAAGTTCAGGGTCAG**ATG**ACCCTGGACTCTCCGACCCAGTTCAAATTCGAC**GGT**ACCACCTCTGACGGTTCTAAA**CTG**ACCGGTACCCTGCAGCGTCAG | making b11L5F_83I_95G_103L variant |
| b11L5F_103L_for | ACCACCTCTGACGGTTCTAAACTGACCGGTACCCTGCAGCGTCAG | making b11L5F_103L and all the triple variants |
| pET15_b11_for | CTGGTGCCGCGCGGCAGCTCC**ATG**TGCCGTGCTGCTTCTCTGCTG | subcloning to pET15 for thrombin cleavage |
| M10_P62ED_for | GTAAAGTTACCTGCAAAACC**GAM**GACACCATGGACGTTGACATC | |
| M10_P62ED_rev | GGTTTTGCAGGTAACTTTAC | making M10_P62E and P62D mutants |
| M10_P8NT_for | TGCCGTGCTGCTTCTCTGCTG**AMC**GGTACCTGGCAGGTTACCATG | |
| M10_P8NTQ_rev | CAGCAGAGAAGCAGCACG | |
| M10_P8Q_for | TGCCGTGCTGCTTCTCTGCTG**CAG**GGTACCTGGCAGGTTACCATG | making M10_P8T, P8N, P8Q mutants* |
| M10_K40V_Rev | GATCGGACGACCGTCAGAGATGGTACCCTGAGCAACAACGTCCAGGGTGTACGGAG | making M10_K40V mutant |
| M10_K54Y_For | CCATCTCTGACGGTCGTCCGATCTCTGGT**TAC**GGTAAAGTTACCTGCAAAAC | making M10_K54Y mutant |
| M10_D68N_Rev | GTTACCCAGAGACGGGTAGGTGATGTTAACGTCCATGGTGTCCGGGGTTTTG | making M10_D68N mutant |
| M10_K78T_For | CACCTACCCGTCTCTGGGTAACATG**ACC**GTTCAGGGTCAGATCACCCTG | making M10_K78T mutant |

| | | |
|---|---|---|
| M10_D86H_K92T_for | TTCAGGGTCAGATCACCCTG**CAC**TCTCCGACCCAGTTC**ACC**TTCGACGCAACCACCCTCTG | |
| M10_D86H_K92T_rev | CAGGGTGATCTGACCCTGAAC | making M10_D86H_K92T double mutant |
| M10_D94V_for | TCTCCGACCCAGTTCAAATTC**GT**TGCAACCACCTCTGACGGTTC | |
| M10_D94V_rev | GAATTTGAACTGGGTCGGAG | making M10_D94V mutant |
| M10_V57A_Rev | GGTGTCCGGGGTTTTGCAGGTTRCTTTACCTTTACCAGAGATCGGAC | making M10_V57A mutant |
| M10_M65IL_For | ACCTGCAAAACCCCGGACACC**MTR**GACGTTGACATCACCTACCCGTC | making M10_M65I, M65L mutants |
| M10_M77VIL_for | CTACCCGTCTCTGGGTAAC**VTA**AAAGTTCAGGGTCAGATCAC | making M10_M77V, M77I, M77L mutants |
| M10_M77VIL_rev | GTTACCCAGAGACGGGTAG | |
| M10_L85YF_for | GTTCAGGGTCAGATCACC**TWC**GACTCTCCGACCCAGTTCAAATTC | |
| M10_L85YF_rev | GGTGATCTGACCCTGAAC | making M10_L85Y, L85F mutants |
| p24_b11_addG_for | CACCGTGAACAGATTGGCGGC**GGT**TGCCGTGCTGCTTCTCTGCTG | adding G for sumo cleavage |
| M10sm1_C1SND_for | GTGCCGCGCGGCAGCTCCATG**RRC**CGTGCTGCTTCTCTGCTGC | making C1S_C59V, C1N_C59V and C1D_C59V with two surface mutants (K40V, K54Y) for M10 |
| M10sm1_C59V_for | CTCTGGTTACGGTAAAGTTACC**GTT**AAAACCCCGGACACCATG | |
| M10sm1_C1SND_rev | CATGGAGCTGCCGCGCGGCAC | |
| M10sm1_C59V_rev | GGTAACTTTACCGTAACCAG | |

*M10 = b11L5F_83I_95A_103L

**Supplementary Table 14: List of genes optimized for mammalian expression.** DNA encoding sequences of mFAP1 and mFAP2 optimized for mammalian codon usage were synthesized and fused to various subcellular targeting tags. Full-length genes are listed in this table with subcellular targeting tag in bold and fusion linker sequence in Italic.

| Name | DNA sequence |
|---|---|
| mFAP1 | ATGAGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGTGGCACTTCCAGCCCCGCAGCCCCTACACCCTGGACATCGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCACCGGCTACGGCAAGGCCACCGTGAAGACCGACGACACCCTGCACGCCAACCTGACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAGATCACCTACGACAGCCCCACCCAGTTCACCTGGAACAGCACCACCAGCGACGGCAAGAAGCTGACCGGCACCCTGCAGCGCCAGGAG |
| mFAP2 | ATGAGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGATGCACTTCCAGCCCCGCAGCCCCTACACCATGGACGTGGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCAGCGGCTACGGCAAGGTGACCGTGAAGACCCCCGACACCCTGGACGTGGACATCACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAGATCACCATGGACAGCCCCACCCAGTTCAAGTTCGACGCCACCACCAAGGGCGCCGGCAACTTCACCGGCCGCCTGACCGGCACCCTGCAGCGCCAGGAG |
| tom20-mFAP1 | **ATGGTAGGCCGGAACAGTGCAATCGCGGCGGGAGTATGTGGTGCGCTGTTCATCGGCTATTGCATTTACTTTGATAGAAAGCGGAGATCAGACCCCAACTTC***GGCTCTGGT*AGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGTGGCACTTCCAGCCCCGCAGCCCCTACACCCTGGACATCGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCACCGGCTACGGCAAGGCCACCGTGAAGACCGACGACACCCTGCACGCCAACCTGACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAGATCACCTACGACAGCCCCACCCAGTTCACCTGGAACAGCACCACCAGCGACGGCAAGAAGCTGACCGGCACCCTGCAGCGCCAGGAG |
| tom20-mFAP2 | **ATGGTAGGCCGGAACAGTGCAATCGCGGCGGGAGTATGTGGTGCGCTGTTCATCGGCTATTGCATTTACTTTGATAGAAAGCGGAGATCAGACCCCAACTTC***GGCTCTGGT*AGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGATGCACTTCCAGCCCCGCAGCCCCTACACCATGGACGTGGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCAGCGGCTACGGCAAGGTGACCGTGAAGACCCCCGACACCCTGGACGTGGACATCACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAGATCACCATGGACAGCCCCACCCAGTTCAAGTTCGACGCCACCACCAAGGGCGCCGGCAACTTCACCGGCCGCCTGACCGGCACCCTGCAGCGCCAGGAG |
| 3nls-mFAP1 | **ATGGATCCTAAGAAAAAGCGCAAGGTTGACCCCAAAAAAAAGAGGAAGGTGGACCCTAAGAAGAAGCGGAAGGTA***GGCTCTGGTAGC*CGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGTGGCACTTCCAGCCCCGCAGCCCCTACACCCTGGACATCGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCACCGGCTACGGCAAGGCCACCGTGAAGACCGACGACACCCTGCACGCCAACCTGACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCC |

| | |
|---|---|
| | AGATCACCTACGACAGCCCCACCCAGTTCACCTGGAACAGCACCACCAGCGACGGCAAGAAGCTGACCGGCACCCTGCAGCGC CAGGAG |
| 3nls-mFAP2 | **ATGGATCCTAAGAAAAAGCGCAAGGTTGACCCCAAAAAAAAGAGGAAGGTGGACCCTAAGAAGAAGCGGAAGGTA***GGC TCTGGTAGC*CGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCA GATGCACTTCCAGCCCCGCAGCCCCTACACCATGGACGTGGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCAGCGGCT ACGGCAAGGTGACCGTGAAGACCCCCGACACCCTGGACGTGGACATCACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGC CAGATCACCATGGACAGCCCCACCCAGTTCAAGTTCGACGCCACCACCAAGGGCGCCGGCAACTTCACCGGCCGCCTGACCGG CACCCTGCAGCGCCAGGAG |
| mts-mFAP1 | **ATGTCCGTCCTGACGCCGCTGCTGCTGCGGGGCTTGACAGGCTCGGCCCGGCGGCTCCCAGTGCCGCGCGCCAAGATC CATTCGTTGGGGGATCCACCGGTC***GGCTCTGGT*AGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCA ACGAGGACGGCCAGACCAGCCAGGGCCAGTGGCACTTCCAGCCCCGCAGCCCCTACACCCTGGACATCGTGGCCCAGGGCACC ATCAGCGACGGCCGCCCCATCACCGGCTACGGCAAGGCCACCGTGAAGACCGACGACACCCTGCACGCCAACCTGACCTACCC CAGCCTGGGCAACATCAAGGCCCAGGGCCAGATCACCTACGACAGCCCCACCCAGTTCACCTGGAACAGCACCACCAGCGACG GCAAGAAGCTGACCGGCACCCTGCAGCGCCAGGAG |
| mts-mFAP2 | **ATGTCCGTCCTGACGCCGCTGCTGCTGCGGGGCTTGACAGGCTCGGCCCGGCGGCTCCCAGTGCCGCGCGCCAAGATC CATTCGTTGGGGGATCCACCGGTC***GGCTCTGGT*AGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCA ACGAGGACGGCCAGACCAGCCAGGGCCAGATGCACTTCCAGCCCCGCAGCCCCTACACCATGGACGTGGTGGCCCAGGGCACC ATCAGCGACGGCCGCCCCATCAGCGGCTACGGCAAGGTGACCGTGAAGACCCCCGACACCCTGGACGTGGACATCACCTACCC CAGCCTGGGCAACATCAAGGCCCAGGGCCAGATCACCATGGACAGCCCCACCCAGTTCAAGTTCGACGCCACCACCAAGGGCG CCGGCAACTTCACCGGCCGCCTGACCGGCACCCTGCAGCGCCAGGAG |
| mFAP1-caax | ATGAGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGTG GCACTTCCAGCCCCGCAGCCCCTACACCCTGGACATCGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCACCGGCTACGG CAAGGCCACCGTGAAGACCGACGACACCCTGCACGCCAACCTGACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAGA TCACCTACGACAGCCCCACCCAGTTCACCTGGAACAGCACCACCAGCGACGGCAAGAAGCTGACCGGCACCCTGCAGCGCCAG GAG**CGAAAACATAAAGAAAAGATGAGCAAAGATGGTAAAAAGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAATTATGT** |
| mFAP2-caax | ATGAGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGAT GCACTTCCAGCCCCGCAGCCCCTACACCATGGACGTGGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCAGCGGCTACG GCAAGGTGACCGTGAAGACCCCCGACACCCTGGACGTGGACATCACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAG ATCACCATGGACAGCCCCACCCAGTTCAAGTTCGACGCCACCACCAAGGGCGCCGGCAACTTCACCGGCCGCCTGACCGGCAC CCTGCAGCGCCAGGAG**CGAAAACATAAAGAAAAGATGAGCAAAGATGGTAAAAAGAAGAAAAAGAAGTCAAAGACAAAG TGTGTAATTATGT** |
| mFAP1-sec61b | ATGAGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGTG GCACTTCCAGCCCCGCAGCCCCTACACCCTGGACATCGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCACCGGCTACGG CAAGGCCACCGTGAAGACCGACGACACCCTGCACGCCAACCTGACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAGA TCACCTACGACAGCCCCACCCAGTTCACCTGGAACAGCACCACCAGCGACGGCAAGAAGCTGACCGGCACCCTGCAGCGCCAG GAG**ATGCCTGGTCCGACCCCCAGTGGCACTAACGTGGGATCCTCAGGGCGCTCTCCCAGCAAAGCAGTGGCCGCCCGG GCGGCGGGATCCACTGTCCGGCAGAGGAAAAATGCCAGCTGTGGGACAAGGAGTGCAGGCCGCACAACCTCGGCAGGC ACCGGGGGGGATGTGGCGATTCTACACAGAAGATTCACCTGGGCTCAAAGTTGGCCCTGTTCCAGTATTGGTTATGAGTC TTCTGTTCATCGCTTCTGTATTTATGTTGCACATTTGGGGCAAGTACACTCGTTCG** |
| mFAP2-sec61b | ATGAGCCGCGCCGCCCAGCTGCTGCCCGGCACCTGGCAGGTGACCATGACCAACGAGGACGGCCAGACCAGCCAGGGCCAGAT GCACTTCCAGCCCCGCAGCCCCTACACCATGGACGTGGTGGCCCAGGGCACCATCAGCGACGGCCGCCCCATCAGCGGCTACG GCAAGGTGACCGTGAAGACCCCCGACACCCTGGACGTGGACATCACCTACCCCAGCCTGGGCAACATCAAGGCCCAGGGCCAG ATCACCATGGACAGCCCCACCCAGTTCAAGTTCGACGCCACCACCAAGGGCGCCGGCAACTTCACCGGCCGCCTGACCGGCAC CCTGCAGCGCCAGGAG**ATGCCTGGTCCGACCCCCAGTGGCACTAACGTGGGATCCTCAGGGCGCTCTCCCAGCAAAGCA GTGGCCGCCCGGGCGGCGGGATCCACTGTCCGGCAGAGGAAAAATGCCAGCTGTGGGACAAGGAGTGCAGGCCGCAC AACCTCGGCAGGCACCGGGGGGATGTGGCGATTCTACACAGAAGATTCACCTGGGCTCAAAGTTGGCCCTGTTCCAGTA TTGGTTATGAGTCTTCTGTTCATCGCTTCTGTATTTATGTTGCACATTTGGGGCAAGTACACTCGTTCG** |

**Supplementary Table 15: List of amino acid sequences of relevant DFHBI-binding fluorescence-activating proteins.** Amino acid sequences were used to compare the mutations (Extended Data Fig. 10).

| Name | Protein sequence |
|---|---|

| | |
|---|---|
| b11 | CRAASLLPGTWQVTMTNEDGQTSQGQMHFQPRSPYTLDVKAQGTMSDGRPIQGKGKVTCKTPDTMDVDITYSDGKQVQGQVTLDSPT QFKFDVTTSDGSKVTGTLQRQELE |
| b11L5F | CRAASLLPGTWQVTMTNEDGQTSQGQMHFQPRSPYTLDVKAQGTISDGRPISGKGKVTCKTPDTMDVDITYPSLGNMKVQGQVTLDSPT QFKFDVTTSDGSKVTGTLQRQELE |
| b11L5F.1 | SRAASLLPGTWQVTMTNEDGQTSQGQMHFQPRSPYTLDVVAQGTISDGRPISGYGKVTVKTPDTMDVDITYPSLGNMKVQGQITLDSPT QFKFDATTSDGSKLTGTLQRQE |
| b11L5F.2 | SRAAQLLPGTWQVTMTNEDGQTSQGQMHFQPRSPYTLDVVAQGTISDGRPISGYGKVTVKTPDTMHVNITYPSLGNIKVQGQITLDSPTQ FTWNSVTSDGKKLTGTLQRQE |
| mFAP0 | SRAAQLLPGTWQVTMTNEDGQTSQGQMHFQPRSPYTLDIVAQGTISDGRPITGYGKVTVKTDDTLHVNITYPSLGNIKVQGQITMDSPTQ ATWNSTTSDGKKLTGTLQRQE |
| mFAP1 | SRAAQLLPGTWQVTMTNEDGQTSQGQWHFQPRSPYTLDIVAQGTISDGRPITGYGKATVKTDDTLHANLTYPSLGNIKAQGQITYDSPTQ FTWNSTTSDGKKLTGTLQRQE |
| mFAP2 | SRAAQLLPGTWQVTMTNEDGQTSQGQMHFQPRSPYTMDVVAQGTISDGRPISGYGKVTVKTPDTLDVDITYPSLGNIKAQGQITMDSPT QFKFDATTKGAGNFTGRLTGTLQRQE |

**Supplementary Table 16: Flow cytometry statistics.** Five yeast libraries were constructed, sorted and sequenced in this study with their description and sorting statistics listed in this table. * indicates that library was deep sequenced by Illumina Miseq sequencer; ** indicates that library was sampled by Sanger sequencing. Sorting mode and sorting efficiency were those provided by Sony SH800 sorter.

| Library | Library Size | Parent Library | Child Library | Cell Treatment | Fluorescence Label | # of Cells Analyzed | # of Cells Collected | % Cells Collected | Sorting Mode | Sorting Efficiency |
|---|---|---|---|---|---|---|---|---|---|---|
| b11L5F_SSM_library_rep_1 | 2,091 | Naïve | FITC_1 * | no special treatment | 1:50 dilution FITC-conjugated anti-c-Myc antibody | 2,026,233 | 294,842 | 17.11% | Purity | 85.03% |
| | | Naïve | T243_1 * | 0.07μM trypsin | | 2,022,656 | 248,270 | 14.68% | Purity | 83.61% |
| | | Naïve | T81_1 * | 0.21μM trypsin | | 2,019,873 | 166,580 | 9.95% | Purity | 82.86% |
| | | Naïve | T27_1 * | 0.63μM trypsin | | 2,028,557 | 23,645 | 1.42% | Purity | 82.12% |
| | | Naïve | T9_1 * | 1.89μM trypsin | | 2,019,247 | 23,792 | 1.45% | Purity | 81.14% |
| | | Naïve | Ch243_1 * | 0.08μM chymotrypsin | | 1,834,048 | 274,542 | 17.00% | Purity | 88.04% |
| | | Naïve | Ch81_1 * | 0.24μM chymotrypsin | | 2,054,811 | 133,695 | 7.80% | Purity | 83.45% |
| | | Naïve | Ch27_1 * | 0.72μM chymotrypsin | | 2,044,317 | 65,756 | 3.79% | Purity | 84.86% |
| | | Naïve | Ch9_1 * | 2.16μM chymotrypsin | | 2,039,805 | 11194 | 0.65% | Purity | 85.49% |
| | | Naïve | s1_1 | | 100μM DFHBI | 3,143,645 | 95,310 | 3.56% | Purity | 85.11% |
| | | s1_1 | Bind_1 * | no special treatment | | 2,727,878 | 15,041 | 0.69% | Purity | 80.43% |
| b11L5F_SSM_library_rep_2 | 2,091 | Naïve | FITC_2 * | no special treatment | 1:50 dilution FITC-conjugated anti-c-Myc antibody | 1,504,471 | 262,030 | 20.51% | Purity | 84.93% |
| | | Naïve | T243_2 * | 0.07μM trypsin | | 1,536,281 | 214,908 | 16.45% | Purity | 85.03% |
| | | Naïve | T81_2 * | 0.21μM trypsin | | 1,387,237 | 138,339 | 11.68% | Purity | 85.41% |
| | | Naïve | T27_2 * | 0.63μM trypsin | | 1,394,728 | 22,102 | 1.89% | Purity | 83.69% |
| | | Naïve | T9_2 * | 1.89μM trypsin | | 1,337,500 | 25,138 | 2.23% | Purity | 84.45% |
| | | Naïve | Ch243_2 * | 0.08μM chymotrypsin | | 2,105,152 | 164,295 | 8.99% | Purity | 86.79% |
| | | Naïve | Ch81_2 * | 0.24μM chymotrypsin | | 2,043,617 | 160,941 | 9.19% | Purity | 85.72% |
| | | Naïve | Ch27_2 * | 0.72μM chymotrypsin | | 2,457,433 | 94,532 | 4.57% | Purity | 84.23% |
| | | Naïve | Ch9_2 * | 2.16μM chymotrypsin | | 2,022,869 | 10391 | 0.63% | Purity | 82.09% |
| | | Naïve | s1_2 | | 100μM DFHBI | 2,816,735 | 37,488 | 1.55% | Purity | 85.67% |
| | | s1_2 | Bind_2 * | no special treatment | | 2,516,899 | 15,164 | 0.71% | Purity | 85.04% |
| b11L5F.2_combinatorial_library | 5.50E+07 | Naïve | b11L5F.2_exp | | 1:50 dilution FITC-conjugated anti-c-Myc antibody | 57,268,884 | 16,884,146 | 34.77% | Normal | 84.78% |
| | | Naïve | b11L5F.2_S1 | | 10μM DFHBI | 43,520,270 | 317,251 | 0.88% | Normal | 82.55% |
| | | b11L5F.2_exp | b11L5F.2_exp_S1 | no special treatment | 10μM DFHBI | 50,535,467 | 354,309 | 0.83% | Normal | 84.69% |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | b11L5F.2_S1 | b11L5F.2_S2 | | 10μM DFHBI | 9,390,040 | 168,311 | 2.16% | Normal | 82.91% |
| | | b11L5F.2_exp_S1 | b11L5F.2_S2 | | 10μM DFHBI | 11,669,754 | 110,864 | 1.16% | Normal | 82.16% |
| | | b11L5F.2_S2 | b11L5F.2_S3 | | 10μM DFHBI | 4,147,893 | 31,920 | 0.89% | Purity | 86.30% |
| | | b11L5F.2_S2 | b11L5F.2_pro_S3 | 0.07μM trypsin, 0.08μM chymotrypsin | 10μM DFHBI | 4,192,495 | 57,111 | 1.65% | Purity | 82.40% |
| | | b11L5F.2_S3 | b11L5F.2_S4 | no special treatment | 5μM DFHBI | 3,339,138 | 24,133 | 0.82% | Purity | 88.10% |
| | | b11L5F.2_pro_S3 | b11L5F.2_pro_S4 | 0.21μM trypsin, 0.24μM chymotrypsin | 10μM DFHBI | 2,963,184 | 24,111 | 0.92% | Purity | 88.54% |
| | | b11L5F.2_S4 | | no special treatment | 5μM DFHBI | 4,134,767 | 20,299 | 0.57% | Purity | 86.06% |
| | | b11L5F.2_pro_S4 | ** | 0.21μM trypsin, 0.24μM chymotrypsin | 10μM DFHBI | 4,976,531 | 30,097 | 0.72% | Purity | 83.72% |
| b11L5F.1_combinatorial_library | 8.808E+07 | Naïve | b11L5F.1_exp | | 1:50 dilution FITC-conjugated anti-c-Myc antibody | 57,945,735 | 16,309,530 | 31.91% | Normal | 88.20% |
| | | Naïve | b11L5F.1_S1 | | 10μM DFHBI | 55,905,778 | 1,325,730 | 2.98% | Normal | 79.65% |
| | | b11L5F.1_exp | b11L5F.1_exp_S1 | | 10μM DFHBI | 42,803,239 | 1,578,029 | 4.63% | Normal | 79.61% |
| | | b11L5F.1_S1 | b11L5F.1_S2 | | 10μM DFHBI | 16,615,788 | 500,395 | 3.69% | Normal | 81.69% |
| | | b11L5F.1_exp_S1 | b11L5F.1_S2 | | 10μM DFHBI | 15,554,114 | 195,104 | 1.52% | Normal | 82.50% |
| | | b11L5F.1_S2 | b11L5F.1_S3 | no special treatment | 10μM DFHBI | 6,005,774 | 40,727 | 0.85% | Purity | 79.85% |
| | | b11L5F.1_S2 | b11L5F.1_pro_S3 | 0.07μM trypsin, 0.08μM chymotrypsin | 10μM DFHBI | 4,427,412 | 56,627 | 1.48% | Purity | 86.60% |
| | | b11L5F.1_S3 | b11L5F.1_S4 | no special treatment | 5 μM DFHBI | 2,509,911 | 22,088 | 0.99% | Purity | 88.90% |
| | | b11L5F.1_pro_S3 | b11L5F.1_pro_S4 | 0.21μM trypsin, 0.24μM chymotrypsin | 10μM DFHBI | 2,924,047 | 41,170 | 1.59% | Purity | 88.35% |
| | | b11L5F.1_S4 | b11L5F.1_S5 | no special treatment | 5 μM DFHBI | 3,945,143 | 34,923 | 1.05% | Purity | 84.47% |
| | | b11L5F.1_pro_S4 | b11L5F.1_pro_S5 | 0.21μM trypsin, 0.24μM chymotrypsin | 5μM DFHBI | 3,370,242 | 21,185 | 0.73% | Purity | 85.53% |
| | | b11L5F.1_S5 | | | 5 μM DFHBI | 2,885,994 | 10,113 | 0.38% | Purity | 91.84% |
| | | b11L5F.1_pro_S5 | ** | no special treatment | 5 μM DFHBI | 6,282,601 | 13,790 | 0.28% | Purity | 78.31% |
| b11L5F.1_error_prone_library | 6.10E+07 | Naïve | ep_exp | | 1:50 dilution FITC-conjugated anti-c-Myc antibody | 61,035,243 | 13,651,919 | 26.58% | Normal | 84.16% |
| | | Naïve | ep_S1 | | 10μM DFHBI | 46,493,935 | 1,679,927 | 4.46% | Normal | 81.06% |
| | | ep_exp | ep_exp_S1 | | 10μM DFHBI | 46,122,324 | 1,246,377 | 3.27% | Normal | 82.68% |
| | | ep_S1 | ep_S2 | | 10μM DFHBI | 16,467,524 | 275,358 | 2.10% | Normal | 79.48% |
| | | ep_exp_S1 | ep_S2 | | 10μM DFHBI | 13,634,466 | 201,031 | 1.77% | Normal | 83.51% |
| | | ep_S2 | ep_S3 | no special treatment | 10μM DFHBI | 4,493,258 | 35,222 | 0.93% | Purity | 84.68% |
| | | ep_S2 | ep_pro_S3 | 0.07μM trypsin, 0.08μM chymotrypsin | 10μM DFHBI | 3,695,738 | 56,674 | 1.74% | Purity | 87.93% |
| | | ep_S3 | ep_S4 | no special treatment | 5 μM DFHBI | 2,836,705 | 20,763 | 0.82% | Purity | 89.56% |
| | | ep_pro_S3 | ep_pro_S4 | 0.21μM trypsin, 0.24μM chymotrypsin | 10μM DFHBI | 2,521,459 | 24,824 | 1.10% | Purity | 89.66% |
| | | ep_S4 | | no special treatment | 5μM DFHBI | 4,695,960 | 25,038 | 0.64% | Purity | 83.29% |
| | | ep_pro_S4 | ** | 0.21μM trypsin, 0.24μM chymotrypsin | 10μM DFHBI | 4,257,680 | 25,170 | 0.72% | Purity | 82.24% |

**Supplementary Table 17: Amino acid propensities categorized by protein depths and secondary structures.** Natural amino acid frequencies in different parts of a protein were analyzed and used for defining the sequence design space (see Supplementary Methods).

| Protein Depth | Amino Acid | Count | | | Propensity | | |
|---|---|---|---|---|---|---|---|
| | | C (random coil) | H (α helix) | E (β sheet) | C | H | E |
| (all) | A | 63723 | 123182 | 40008 | -0.40 | 0.45 | -0.40 |
| | C | 13643 | 12396 | 11689 | -0.04 | -0.27 | 0.41 |
| | G | 127043 | 36623 | 30714 | 0.82 | -1.07 | -0.56 |
| | I | 35770 | 64736 | 63749 | -0.77 | -0.01 | 0.74 |
| | L | 64253 | 137273 | 68219 | -0.64 | 0.36 | 0.12 |
| | M | 17883 | 29490 | 14524 | -0.36 | 0.26 | 0.01 |
| | F | 32391 | 44954 | 37848 | -0.40 | -0.02 | 0.50 |
| | P | 88031 | 25856 | 14041 | 0.89 | -0.97 | -1.08 |
| | W | 11105 | 16130 | 11877 | -0.38 | 0.06 | 0.38 |
| | Y | 28590 | 39100 | 32725 | -0.38 | -0.03 | 0.49 |
| | V | 45482 | 65823 | 84625 | -0.68 | -0.24 | 0.89 |
| | R | 46551 | 68488 | 31013 | -0.22 | 0.24 | -0.13 |
| | N | 64175 | 39287 | 18202 | 0.51 | -0.30 | -0.64 |
| | D | 86951 | 57796 | 22335 | 0.49 | -0.20 | -0.80 |
| | Q | 34122 | 55429 | 19444 | -0.24 | 0.36 | -0.38 |
| | E | 60432 | 103442 | 31537 | -0.26 | 0.42 | -0.53 |
| | H | 27427 | 23809 | 15246 | 0.15 | -0.15 | -0.02 |
| | K | 59644 | 76271 | 31190 | -0.05 | 0.20 | -0.32 |
| | S | 80629 | 57654 | 35541 | 0.32 | -0.26 | -0.19 |
| | T | 62094 | 46113 | 44480 | 0.13 | -0.39 | 0.32 |
| 0-5 (protein surface) | | C | H | E | C | H | E |
| | A | 36354 | 47186 | 7322 | -0.66 | -0.01 | -1.46 |
| | C | 5309 | 2015 | 1814 | -0.84 | -1.97 | -0.89 |
| | G | 77095 | 11317 | 5556 | 0.65 | -1.85 | -1.64 |
| | I | 16333 | 15515 | 13389 | -1.35 | -1.15 | -0.12 |
| | L | 32710 | 38136 | 14770 | -1.06 | -0.57 | -0.70 |
| | M | 10444 | 9389 | 3699 | -0.58 | -0.46 | -0.57 |
| | F | 15054 | 11312 | 8496 | -0.95 | -1.09 | -0.27 |
| | P | 59379 | 16359 | 6898 | 0.88 | -0.71 | -0.72 |
| | W | 6107 | 5908 | 4387 | -0.69 | -0.47 | 0.34 |
| | Y | 17698 | 18160 | 14140 | -0.52 | -0.21 | 0.66 |
| | V | 22934 | 17386 | 19055 | -1.11 | -1.24 | 0.13 |
| | R | 39472 | 56232 | 23564 | 0.10 | 0.88 | 0.86 |
| | N | 49890 | 27277 | 9787 | 0.70 | 0.10 | -0.14 |
| | D | 71111 | 45461 | 13136 | 0.75 | 0.38 | -0.18 |
| | Q | 28351 | 43233 | 12927 | 0.04 | 0.92 | 0.42 |
| | E | 53718 | 88301 | 23717 | 0.12 | 1.11 | 0.45 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | H | 19626 | 14689 | 7458 | 0.22 | 0.08 | 0.34 |
| | K | 55948 | 69036 | 26584 | 0.41 | 0.98 | 0.84 |
| | S | 54987 | 32141 | 13702 | 0.32 | -0.18 | -0.17 |
| | T | 43328 | 23568 | 21274 | 0.17 | -0.44 | 0.65 |
| | | **C** | **H** | **E** | **C** | **H** | **E** |
| | A | 25177 | 69343 | 27062 | 0.02 | 0.81 | -0.07 |
| | C | 7717 | 9569 | 8570 | 0.90 | 0.54 | 0.86 |
| | G | 45442 | 21759 | 20515 | 1.09 | -0.64 | -0.24 |
| | I | 18431 | 46458 | 43318 | 0.03 | 0.70 | 1.08 |
| | L | 30072 | 93791 | 46343 | 0.02 | 1.00 | 0.46 |
| | M | 6879 | 18612 | 9120 | 0.02 | 0.79 | 0.24 |
| | F | 16501 | 31913 | 26000 | 0.38 | 0.67 | 0.85 |
| | P | 26507 | 8508 | 6168 | 0.92 | -1.39 | -1.37 |
| | W | 4760 | 9771 | 6912 | 0.15 | 0.52 | 0.50 |
| | Y | 10411 | 20063 | 17367 | -0.08 | 0.20 | 0.47 |
| | V | 21327 | 45259 | 56587 | -0.01 | 0.41 | 1.21 |
| | R | 6722 | 11730 | 6971 | -1.25 | -1.12 | -1.39 |
| | N | 13206 | 11035 | 7367 | -0.02 | -0.94 | -1.04 |
| | D | 14819 | 11490 | 8252 | -0.31 | -1.34 | -1.34 |
| | Q | 5303 | 11330 | 5833 | -1.17 | -0.74 | -1.22 |
| | E | 6167 | 14192 | 7001 | -1.80 | -1.26 | -1.80 |
| | H | 7149 | 8443 | 6814 | -0.03 | -0.46 | -0.28 |
| | K | 3469 | 6882 | 4248 | -2.40 | -2.08 | -2.30 |
| **5-10(protein core)** | S | 23616 | 22929 | 18896 | 0.31 | -0.40 | -0.20 |
| | T | 17255 | 20235 | 20259 | 0.04 | -0.39 | 0.09 |

**Supplementary Table 18**: X-ray crystallography data collection and refinement statistics. X-ray diffraction data for each protein structure were collected on a single crystal and processed as described in Methods.

| | BB1 (PDB ID: 6D0T) | HBI_b_10 (PDB ID: 6CZJ) | b11L5F_LGL (PDB ID: 6CZG) | mFAP1 (PDB ID: 6CZH) | mFAP0 (PDB ID: 6CZI) |
|---|---|---|---|---|---|
| **Data collection** | | | | | |
| Space group | P 1 21 1 | P 1 21 1 | C 1 2 1 | P 1 21 1 | P 21 21 21 |
| Cell dimensions | | | | | |
| *a*, *b*, *c* (Å) | 55.76, 32.27, 81.65 | 42.9, 36.7, 61.9 | 84.9, 35.3, 59.6 | 40.1, 47.9, 52.8 | 48.1, 59.4, 72.8 |
| a, b, g (°) | 90, 100.64, 90 | 90, 91.1, 90 | 90, 90.9, 90 | 90, 91.5, 90 | 90, 90, 90 |
| Resolution (Å) | 27.81-1.63 (1.688-1.63) | 50-2.1 (2.14-2.1) | 50-2.2 (2.24-2.2) | 50-2.3 (2.34-2.3) | 40-1.8 (1.87-1.80) |
| $R_{sym}$ or $R_{merge}$ | 0.058 (0.653) | 0.020 (0.027) | 0.029 (0.064) | 0.067 (0.306) | 0.063 (0.434) |
| $I$ / s$I$ | 13.1 (1.9) | 72.0 (50.4) | 60.8 (29.4) | 20.9 (3.0) | 37.1 (5.0) |
| Completeness (%) | 99.8 (99.8) | 97.5 (81.7) | 98.5 | 98.2 (85.1) | 98.9 (86.9) |
| Redundancy | 4.0 (4.1) | 7.3 (6.0) | 7.4 (7.3) | 6.9 (4.6) | 11.7 (9.2) |
| **Refinement** | | | | | |
| Resolution (Å) | 1.63 | 2.1 | 2.2 | 2.3 | 1.8 |
| No. reflections | 36147 | 11219 | 9067 | 8895 | 18915 |
| $R_{work}$ / $R_{free}$ | 0.1515/0.1840 | 17.89/22.61 | 21.68/27.52 | 18.07/21.81 | 20.59/24.13 |
| No. atoms | | | | | |
| Protein | 1892 | 1589 | 1634 | 1631 | 1625 |
| Ligand/ion | 0 | 15 | 36 | 36 | 36 |
| Water | 101 | 208 | 84 | 82 | 129 |
| *B*-factors | | | | | |
| Protein | 26.46 | 15.83 | 26.48 | 27.94 | 21.36 |
| Ligand/ion | N/A | 35.59 | 39.59 | 22.89 | 13.68 |
| Water | 37.46 | 23.55 | 31.58 | 31.1 | 32.74 |
| R.m.s. deviations | | | | | |
| Bond lengths (Å) | 0.01 | 0.002 | 0.002 | 0.009 | 0.003 |
| Bond angles (°) | 1.059 | 0.52 | 0.67 | 1.1 | 0.7 |

*Values in parentheses are for highest-resolution shell.

# Supplementary Data

## EXAMPLE COMMAND LINES

**Example command line for selecting/designing the parametric models:**

PATH_TO_ROSETTA/Rosetta/main/source/bin/remodel.default.linuxgccrelease

-parser:protocol <protocol.xml> ***# see Supplementary Data: paramtetric_bb_minpackfilter.xml, parametric_bb_design.xml***
-database PATH_TO_ROSETTA/Rosetta/main/database
-nstruct 10
-linmem_ig 10
-use_bicubic_interpolation
-use_incorrect_hbond_deriv false
-score:weights trp_ala_mod.wts
-no_his_his_pairE
-hbond_sp2_correction
-score:weights sp2_correction
-analytic_stable_evaluation
-icoor_05_2009
-lj_hbond_hdis 1.75
-lj_hbond_OH_donor_dis 2.6
-hackelec_min_dis 2.0
-scale_d 1
-scale_theta 1
-holes:dalphaball PATH_TO_ROSETTA/Rosetta/main/database/DAlphaBall.icc

## Example command line for connecting beta-strands generated from parametric models:

PATH_TO_ROSETTA/Rosetta/main/source/bin/remodel.default.linuxgccrelease

-s <picked.pdb> # selected input pdb
-remodel:blueprint <2_2_2_2_2_1_2.bp> ***# see Supplementary Data: 2_2_2_2_2_1_2.bp***
-nstruct 5
-database PATH_TO_ROSETTA/Rosetta/main/database
-num_trajectory 1
-lh:db_path PATH_TO_FRAGMENT_DATABASE/3to25mer/
-lh_ex_limit 8
-lh:max_radius 10
-use_loop_hash
-out:user_tag 2_2_2_2_2_1_2
-out:suffix 2_2_2_2_2_1_2
-out:file:silent picked-2_2_2_2_2_1_2.silent
-ss_pair 1.0
-rsigma 1.0
-hb_lrbb 1.5
-remodel:use_cart_relax
-remodel:free_relax

## Example command line for constructing beta-barrel backbones based on the 2D map:

PATH_TO_ROSETTA/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease

-parser:protocol <bb_2D_assembly.xml> ***# see Supplementary Data: bb_2D_assembly.xml***
-database PATH_TO_ROSETTA/Rosetta/main/database/
-s <input_dipeptide.pdb> *# an arbitrary dipeptide to define the start and end of the protein chain*
-picking_old_max_score 1
-maxruntime 14400
-nstruct 100
-seed_offset 4
-holes:dalphaball PATH_TO_ROSETTA/Rosetta/main/database/DAlphaBall.icc

## Example command line for designing sequences after fragment assembly:

PATH_TO_ROSETTA/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease

-parser:protocol <bb_2D_design.xml> ***# see Supplementary Data: bb_2D_design.xml***
-database PATH_TO_ROSETTA/Rosetta/main/database/

-s <input_dipeptide.pdb>
-picking_old_max_score 1
-maxruntime 74000
-nstruct 5
-rama_prepro_steep
-beta

## Example command line for designing disulfide bonds:

PATH_TO_ROSETTA/Rosetta/main/source/bin/remodel.static.linuxgccrelease
-database PATH_TO_ROSETTA/Rosetta/main/database/
-s <input.pdb>
-remodel:blueprint <helix_remodel.bp> ***# see Supplementary Data: helix_remodel.bp***
-remodel:build_disulf
-save_top 20
-remodel:use_pose_relax
-num_trajectory 250
-bypass_closure -match_rt_limit 2.5

## Example command line for correcting small molecule partial charges:

PATH_TO_AMBER/amber12/AmberTools/bin/antechamber -i <input.mol2> -fi mol2 -o <output.mol2> -fo mol2 -c bcc -at sybyl -nc <charge>

## Example command line for generating ligand .param files:

PATH_TO_ROSETTA/Rosetta/main/source/scripts/python/public/molfile_to_params.py -n HBI <input.mol2>

## Example command line for running RIFgen:

PATH_TO_RIF/rifgen @rifgen_options

**rifgen_options:**

```
# target and params here
-rifgen:target PATH_TO_TARGET/HBIh_0001.pdb
-extra_res_fa  PATH_TO_TARGET/HBIh.params

# names of output file and directory
-rifgen:outdir        OUTPUT_FOLDER
-rifgen:data_cache_dir PATH_TO_RIF/data/scheme_data
-rifgen:outfile OUTPUT_FILE_NAME.gz
-rifgen:apores VAL ILE LEU MET PHE          # apolar residue types
-rifgen:donres SER THR TYR GLN ASN HIS HIS_D # donor residue types
-rifgen:accres HBI                          # acceptor residue types

# options for dumping pdbs condaining a small fraction of the RIF residues for inspection
-rifgen:rif_hbond_dump_fraction  0.00001
-rifgen:rif_apo_dump_fraction    0.00001

# mutiplier for default hydrophobic residue score cut
# this will largely determine how long rif generation takes and
# how large the resulting rif is. lowering this number will cause
# more possible hydrophobic interactions to be found, but will make
# the rif bigger and make the search take longer. if you don't care
# much about hydrophobic interactions, or only want a few good ones
# raise this number to maybe 1.2 (fine to play with it) OR if you
# aren't getting good hydrophobic packing, lower it to maybe 0.8 or 0.7.
-rifgen:score_cut_adjust 1.0
-rifgen:score_threshold -0.5       # max acceptable score to go in rif
-rifgen:hbond_weight 2.0           # max score per-hbond
-rifgen:upweight_multi_hbond 1.0   # extra score factor for bidentate hbonds
```

```
# params for super-fussy hbonder search
-rifgen:tip_tol_deg         30.0   # hbonds off-ideal by this much angle
-rifgen:rot_samp_resl       3.75  # angular sampling covering radius
-hbond_cart_sample_hack_range 0.50  # cart search for hbonders up to this far away
-hbond_cart_sample_hack_resl  0.25 # cart search for hbonders at this resl
-hash_cart_resl    0.7     # main rif hash table cart resolution
-hash_angle_resl  14.0     # main rif hash table angle resolution
-rifgen::rif_type RotScoreSat # if you don't want satisfaction constraints, use "RotScore"
-rifgen::rf_oversample 2                  # number of samples per base-grid cell
-rifgen::rosetta_field_resl 0.125          # base resolution of energy grids
#-rifgen::search_resolutions 4.0 2.0 1.0 0.5 # apo residue position 6D search resls
-rifgen::search_resolutions 3.0 1.5 0.75 # apo residue position 6D search resls

# memory request depends on the computers
-rif_accum_scratch_size_M 24000 # 250gb, jojo only!

-rifgen:beam_size_M 10000.0
-rifgen:hash_preallocate_mult 0.125
-rifgen:max_rf_bounding_ratio 4.0
-add_orbitals
-renumber_pdb
-database PATH_TO_ROSETTA_DATABASE/database

-rifgen:hash_cart_resls   16.0   8.0   4.0   2.0   1.0
-rifgen:hash_cart_bounds   512   512   512   512   512
-rifgen:lever_bounds      16.0   8.0   4.0   2.0   1.0
-rifgen:hash_ang_resls     38.8  24.4  17.2  13.6  11.8
-rifgen:lever_radii        23.6 18.785501 13.324600  8.425850  4.855575
```

## Example command line for running RIFdock:

```
PATH_TO_RIF/rif_dock_test
-scaffolds LIST_OF_SCAFFOLDS
-scaffold_res LIST_OF_POSITION_FILES
@rifdock_hbi.flags
```

**rifdock_hbi.flags**

```
# the block below comes from the bottom of the log file from rif generation, just copy it
# if you are running the docking in a different place than where you ran rif generation,
# you will have to adjust these paths!
############################################ what you need for docking ############################################
-rif_dock:target_pdb           ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz_target.pdb
-in:file:extra_res_fa          ./test_input/hbi/HBIh.params
-rif_dock:target_rf_resl       0.125
-rif_dock:target_rf_cache      ./rifgen_hbi/__RF_HBIh_0001.pdb_CEN_trhash54435770_resl0.125_osamp2_replonlybdry
-rif_dock:target_bounding_xmaps ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz_BOUNDING_RIF_16.xmap.gz
-rif_dock:target_bounding_xmaps ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz_BOUNDING_RIF_08.xmap.gz
-rif_dock:target_bounding_xmaps ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz_BOUNDING_RIF_04.xmap.gz
-rif_dock:target_bounding_xmaps ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz_BOUNDING_RIF_02.xmap.gz
-rif_dock:target_bounding_xmaps ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz_BOUNDING_RIF_01.xmap.gz
-rif_dock:target_rif           ./rifgen_hbi/hbi_resl0.7_sca0.8_hb30_6_hb2umh1_sat5.rif.gz
##########################################################################################################################
##
# this is where the output will go, and how much
-rif_dock:outdir OUTPUT_FOLDER
-rif_dock:dokfile all.dok
```

```
-rif_dock:n_pdb_out 20 # max number of output pdbs
# set this to the number of hbonds to the target which are required
-require_satisfaction 4
#############################################################################################
# these flags control the overall time the search will take a few alternate options are included
# setting the require_satisfaction flag above to a high vale will make search faster across the
# board, so experiment with that also

# reasonable defaults:
-beam_size_M 5
-hsearch_scale_factor 1.2
# # very fast search, probably with low quality results
        # -beam_size_M 1
        # -hsearch_scale_factor 1.6
# # slow thorough search
        # -beam_size_M 30
        # -hsearch_scale_factor 1.0
# score cut for the rosetta "score," which is kinda a ddg, but with hbond weighs highern
-rif_dock:rosetta_score_cut -10.0
# make this number higher to have less redundant results or lower to have more similar results
# this is NOT a proper rmsd (yet), unfortunately, so if you want to tweak it you'll have to experiment
-rif_dock:redundancy_filter_mag 1.5
# rotamer packing options
-rif_dock::pack_iter_mult 4.0
-rif_dock:hack_pack_frac 0.20
-hack_pack true
-rif_dock::rf_resl 0.5
-rif_dock::rf_oversample 2
-rif_dock:rotrf_resl   0.3
-rif_dock:rotrf_spread 0.0
-rif_dock:rotrf_scale_atr 1.0
-rif_dock:rotrf_cache_dir PATH_TO_ROTAMER_TABLE
-rif_dock:data_cache_dir  SCAFFOLD_CACHE_FOLDER
-rif_dock:use_scaffold_bounding_grids 0
-rif_dock:cache_scaffold_data true
-rif_dock:upweight_iface 1.3
-rif_dock:hbond_weight 3.0
# value of 1.0 could up to double hbscore if bidentate, triple if tridentate... best in conjunction with low-ish starting hbweight
-rif_dock:upweight_multi_hbond 1.0
-rif_dock:rosetta_score_fraction 0.01
-rif_dock:rosetta_min_fraction 0.14
-rif_dock:pdb_info_pikaa false
-rif_dock:align_output_to_scaffold true
-rif_dock:global_score_cut -10.0
-rif_dock:scaffold_to_ala true        # change output sequence backbone to poly-ala
-rif_dock:scaffold_to_ala_selonly false
-add_native_scaffold_rots_when_packing 0   # 1 if scaffold input sequence is meaningful (eg. natural proteins as scaffolds )
-bonus_to_native_scaffold_res         0          # -0.5 if scaffold input sequence is meaningful
-add_orbitals
-database PATH_TO_ROSETTA_DATABASE/database
#-rif_dock:target_tag conf01    # for multiple ligand conformers
-rif_dock:target_rf_oversample 2
-mute core.scoring.ScoreFunctionFactory
```

**Position file used for docking DFHBI into beta barrel:**

```
15  17  21  23  41  45  49  51  69  71  75  77  93  95  99 101
```

## Example command line for running sequence design after RIF docking:

PATH_TO_ROSETTA/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease
-parser:protocol <design_protocol>.xml ***#see Supplementary Data: hbi_p2_rectBarrel.xml, hbi_p2_rectBarrel_aacomp.xml, hbi_p2_rectBarrel_releaserif.xml, resfile_design.xml, pocket_loop_redesign.xml, 2ndround_design.xml***
-in:file:s <input>.pdb
@HBI_design.flags

**HBI_design.flags**

-beta
-rama_prepro_steep
-run::preserve_header
-packing::use_input_sc
-packing:extrachi_cutoff 18
-packing::ex1
-packing::ex2
-linmem_ig 10
-nblist_autoupdate
-ignore_unrecognized_res
-no_optH false #no_optH is to avoid optimizing proton position for His which should be set False.
-enzdes
-detect_design_interface
-enzdes::minimize_ligand_torsions 5.0
-flip_HNQ #Flip between different protonated states for amino acids H, N and Q
-no_his_his_pairE #Histine is always favored by Rosetta since it can be buried or exposed to solvate. This flag is set to avoid putting in a lot of paired His.
-chemical:exclude_patches LowerDNA UpperDNA Cterm_amidation VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated thr_phosphorylated tyr_phosphorylated tyr_sulfated lys_dimethylated lys_monomethylated lys_trimethylated lys_acetylated glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated MethylatedProteinCterm
-nstruct 5 # generating 5 outputs for each input
-jd2:ntrials 1
-extra_res_fa PATH_TO_PARAM_FILE/HBI_rch.params
-database PATH_TO_ROSETTA/Rosetta/main/database
-holes:dalphaball DAlphaBall.gcc # Rosetta hole definition file

## Example command lines for running MD-based model refinement:

## 1) Preparation & equilibration

***NOTE: Required input scripts highlighted with underlines are fully described at the bottom of each section***

### 1-1. Prepare & minimize protein

$AMBERHOME/bin/tleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff12SB -s prep.txt
$AMBERHOME/bin/sander -O -i minimize.in -p vacuum.prmtop -c vacuum.inpcrd -r min.rst -ref vacuum.inpcrd
$AMBERHOME/bin/ambpdb -p vacuum.prmtop < min.rst > minimized.pdb

### 1-2. Solvate & minimize whole system

$AMBERHOME/bin/tleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff12SB -f
mpiexec -np 12 $AMBERHOME/bin/pmemd.MPI -O -i minimize.in -p solv.prmtop -c solv.inpcrd -r solvent_minimized.rst -ref solv.inpcrd

### 1-3. Short MD to heat up whole system

mpiexec -np 12 $AMBERHOME/bin/pmemd.MPI -O -i heat.in -p solv.prmtop -r eq.rst -ref solvent_minimized.rst -c solvent_minimized.rst -ref solvent_minimized.rst

**<prep.txt>**

source leaprc.gaff
loadoff ions08.lib
prt = loadpdb [input.pdb] # pdb used for input

saveamberparm prt vacuum.prmtop vacuum.inpcrd
quit

**<solvate.txt>**

source leaprc.gaff

```
loadoff ions08.lib
loadamberparams frcmod.ionsjc_tip3p

PRT = loadpdb minimized.pdb
solvateBox PRT TIP3PBOX 10

addions PRT Na+ 0
addions PRT Cl- 0

saveamberparm PRT solv.prmtop solv.inpcrd
quit
```

**<minimize.in>**

```
&cntrl
imin   = 1,  maxcyc = 5000,  ncyc   = 2500,  ntr    = 1,  ntb    = 1,  cut    = 10.0
 /
500.0
RES 1 [nres] #nres = total number of residues in protein
END
END
```

**<heat.in>**

```
&cntrl
imin   = 0,irest  = 0,ntx    = 1,ntb    = 1,cut    = 10.0,ntr    = 1,restraint_wt = 0.1,
ntc    = 2,ntf    = 2,tempi  = 50.0,temp0  = 300.0,ntt    = 3,gamma_ln = 1.0,
nstlim = 25000, dt = 0.002,ntpr = 5000, ntwx = 5000, ntwr = 5000
/
1.0
RES 1 [nres] #nres = total number of residues in protein
END
END
```

## 2) Production run (repeat 5 times with X=1~5)

mpiexec -np 12 $AMBERHOME/bin/pmemd.MPI -O -i md.in -p solv.prmtop -c eq.rst -r md.rst -ref eq.rst -x mdcrd.[X]

**<md.in>**

```
&cntrl
imin = 0, irest = 1, ntx = 5, iwrap=1,ntb = 2, pres0 = 1.0, ntp = 1,
taup = 2.0,ig=-1,cut = 10.0, ntr = 1,ntc = 2, ntf = 2,
restraintmask = ':1-[nres]@CA',restraint_wt = 0.05, #nres = total number of residues in protein
tempi = 300, temp0 = 300, ntt = 3, gamma_ln = 1.0,
nstlim = 5000000, dt = 0.002,ntpr = 25000, ntwx = 25000, ntwr = 25000
/
```

## 3) Structural averaging and regularization

```
$AMBERHOME/bin/ptraj  solv.prmtop < trjavrg.in
$ROSETTA/source/bin/relax.linuxgccrelease -s trjavrg.pdb -score:weights ref2015_cart \
-set_weights coordinate_constraint 10.0 -constrain_relax_to_start_coords \
-relax:script cart2r.script -database $ROSETTA/database
```

**<trjavrg.in>**

```
trajin mdcrd.1 1 1000 1
trajin mdcrd.2 1 1000 1
trajin mdcrd.3 1 1000 1
trajin mdcrd.4 1 1000 1
trajin mdcrd.5 1 1000 1
strip :WAT
strip :Na+
strip :Cl-
average trjavrg.pdb pdb
```

**<cart2r.script>**

```
switch:cartesian
```

```
repeat 2
ramp_repack_min 0.02  0.01     1.0  50
ramp_repack_min 0.250 0.01     0.5  50
ramp_repack_min 0.550 0.01     0.1 100
ramp_repack_min 1     0.00001  0.1 200
Accept_to_best
endrepeat
```

### Example command line for running RosettaLigand docking:

*Please see examples in:*
*G Lemmon and J Meiler, RosettaLigand docking with flexible XML protocols, Methods Mol Biol 2012, 819: 143-155.*

### Example command line for building the new loop L5F for b11:

```
PATH_TO_ROSETTA/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease
        -parser:protocol remodel_L5F.xml  # see Supplementary data: remodel_L5F.xml
        -input b11.pdb
        -database PATH_TO_ROSETTA/Rosetta/main/database
        -save_top 20
        -remodel:use_pose_relax
        -num_trajectory 50
        -extra_res_fa HBI.fa.params
        -extra_res_cen HBI.cen.params  # see Supplementary data: HBI.cen.params and HBI.fa.params
        -vall pipette7s.vall.gz  # Supplementary data: custom fragment library
        -beta
        -cst_file L5F.cst   # see Supplementary data: L5F.cst
        -hb_lrbb 0.75
        -ex1
        -ex2
```

## EXAMPLE ROSETTASCRIPTS XML FILES

### RosettaScripts XML file used for minimizing, packing and filtering of disconnected parametric strands arrangements:

### paramtetric_bb_minpackfilter.xml

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
          <hard weights=trp_ala3_mod.wts />
          <soft weights=soft_rep_trp_ala/>
          <hard_ele weights=trp_ala2_ele2.wts />
          <hard_bb weights=bb_only.wts >
              <Reweight scoretype=hbond_lr_bb weight=5. />
          </hard_bb>
    </SCOREFXNS>
    <TASKOPERATIONS>
        <ReadResfile name=resfile filename=bb_param.res/>  # see Supplementary Data: bb_param.res
        <IncludeCurrent name=current/>
        <LimitAromaChi2 name=arochi />
        <ExtraRotamersGeneric name=ex1_ex2 ex1=1 ex2=1/>
        <ExtraRotamersGeneric name=ex1 ex1=1/>
        <LayerDesign name=all_layers layer=other make_pymol_script=1 >
            <CombinedTasks name=barrel_core>
                <SelectBySASA state=bound mode=mc core=1 probe_radius=2.0 core_asa=35 surface_asa=45 verbose=1/>
            </CombinedTasks>
```

```
                    <barrel_core>
                        <all copy_layer=core />
                     </barrel_core>
                     <CombinedTasks name=barrel_surface>
                          <SelectBySASA state=bound mode=mc surface=1 probe_radius=2.0 core_asa=35 surface_asa=45 verbose=1/>
                     </CombinedTasks>
                     <barrel_surface>
                          <all copy_layer=surface  />
                     </barrel_s urface>
                     <CombinedTasks name=barrel_boundary>
                          <SelectBySASA state=bound mode=mc boundary=1 probe_radius=2.0 core_asa=35 surface_asa=45 verbose=1/>
                     </CombinedTasks>
                     <barrel_boundary>
                          <all copy_layer=boundary  />
                    </barrel_boundary>
               </LayerDesign>

              <SelectBySASA name=select_core state=bound mode=mc core=1 probe_radius=2.0 core_asa=30 surface_asa=45 verbose=1/>
               <SelectBySASA name=select_boundary state=bound mode=mc boundary=1 probe_radius=2.0 core_asa=35 surface_asa=45
verbose=1/>
               <SelectBySASA name=select_surface state=bound mode=mc surface=1 probe_radius=2.0 core_asa=35 surface_asa=40 verbose=1/>
               <RestrictAbsentCanonicalAAS name=ala_only resnum=0 keep_aas="A" />
      </TASKOPERATIONS>
      <FILTERS>
               <Holes name=holes threshold=3.0 confidence=0/>
               <PackStat name=packstat threshold=0.65 confidence=0/>
      </FILTERS>
      <MOVERS>
               <PackRotamersMover name=softpack_core scorefxn=soft task_operations=all_layers,select_core,current,arochi/>
               <PackRotamersMover name=softpack_surface scorefxn=soft task_operations=all_layers,select_surface,current,arochi/>
               <PackRotamersMover name=hardpack_surface scorefxn=hard_ele task_operations=all_layers,select_surface,current,arochi,ex1/>
               <PackRotamersMover name=hardpack_core scorefxn=hard task_operations=all_layers,select_core,current,arochi,ex1_ex2/>
               <PackRotamersMover name=softpack_boundary scorefxn=soft task_operations=all_layers,select_boundary,current,arochi/>
               <PackRotamersMover name=hardpack_boundary  scorefxn=hard
task_operations=all_layers,select_boundary,current,arochi,ex1_ex2/>
               <MinMover name=hardmin_cart scorefxn=hard type=lbfgs_armijo_nonmonotone tolerance=0.0001  chi=1 bb=1 bondangle=1
bondlength=1 jump=all cartesian=1/>
               <AddConstraintsToCurrentConformationMover name=add_cst use_distance_cst=0 max_distance=12. coord_dev=5.0 min_seq_sep=8
/>
               <ClearConstraintsMover name="clearconstraints"/>
               <PDBReload name=reload />
               <MinMover name=hardmin_bb scorefxn=hard_bb type=lbfgs_armijo_nonmonotone tolerance=0.0001  chi=1 bb=1 bondangle=1
bondlength=1 jump=all cartesian=1/>
               <PackRotamersMover name=transform_sc scorefxn=hard task_operations=ala_only/>
   </MOVERS>
   <APPLY_TO_POSE>
   </APPLY_TO_POSE>
   <PROTOCOLS>
               ////////////////////////////////////////////////////////////////////////////
               // Minimization to enforce backbone interactions//
               ////////////////////////////////////////////////////////////////////////////
    <Add mover=transform_sc/>
    <Add mover=add_cst/>
    <Add mover=hardmin_bb/>
    <Add mover=clearconstraints/>
               ////////////////////////////////////
               // Sidechains design    //
```

```
        ////////////////////////////////////
    <Add mover=softpack_core/>
    <Add mover=softpack_boundary/>
    <Add mover=softpack_surface/>
    <Add mover=hardmin_sconly/>
    <Add mover=hardpack_core/>
    <Add mover=hardpack_boundary/>
    <Add mover=hardpack_surface/>
    <Add mover=hardmin_cart/>
        ////////////////////////////////////////////////////////////
        // Selection of well packed backbones//
        ////////////////////////////////////////////////////////////
    <Add filter=holes/>
    <Add filter=packstat/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML file used for sequence design of parametrically generated beta-barrels: parametric_bb_design.xml

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <hard weights=talaris2013.B.wts />
    </SCOREFXNS>
    <TASKOPERATIONS>
        <ReadResfile name=resfile filename=bb_param.res/>  # see Supplementary Data: bb_param.res
        <LimitAromaChi2 name=limitchi2 />
        <LayerDesign name=all_layers layer=other make_pymol_script=1 >
            <CombinedTasks name=barrel_core>
                <SelectBySASA state=bound mode=mc core=1 probe_radius=2.0 core_asa=35 surface_asa=45 verbose=1/>
            </CombinedTasks>
            <barrel_core>
                <all copy_layer=core />
            </barrel_core>
            <CombinedTasks name=barrel_surface>
                <SelectBySASA state=bound mode=mc surface=1 probe_radius=2.0 core_asa=35 surface_asa=45 verbose=1/>
            </CombinedTasks>
            <barrel_surface>
                <all copy_layer=surface  />
            </barrel_surface>
            <CombinedTasks name=barrel_boundary>
                <SelectBySASA state=bound mode=mc boundary=1 probe_radius=2.0 core_asa=35 surface_asa=45
verbose=1/>
            </CombinedTasks>
            <barrel_boundary>
                <all copy_layer=boundary  />
            </barrel_boundary>
        </LayerDesign>
    </TASKOPERATIONS>
    <FILTERS>
        <SSPrediction name="sspred" confidence="0" threshold=0.4 use_svm="1" use_probability="1"/>
        <Holes name=holes threshold=3.0 confidence=0/>
        <PackStat name=packstat threshold=0.65 confidence=0/>
        <ScoreType name="rama" score_type="rama" threshold=0.0 confidence="0" />
        <ScoreType name="score" score_type="total_score" threshold=0.0 confidence="0" />
        <Geometry name=geo omega=165 cart_bonded=35 confidence=1/>
    </FILTERS>
```

```
    <MOVERS>
        <Dssp name=dssp/>
        <FastRelax name=relax />
        <FastDesign name="fdesign" task_operations="resfile,limitchi2,all_layers" scorefxn="hard" allow_design="1"
only_design_worst_region="0" design_by_psipred="0" design_by_frag_qual="0" repeats="1" clear_designable_residues="0" dumpall="0"
max_redesigns="2000" ramp_design_constraints="0" />
        <ParsedProtocol name=design >
            <Add mover_name=fdesign />
            <Add mover_name=relax />
            <Add mover_name=dssp />
        </ParsedProtocol>
  </MOVERS>
  <APPLY_TO_POSE>
  </APPLY_TO_POSE>
  <PROTOCOLS>
        <Add mover=design />
        <Add filter=holes/>
        <Add filter=packstat/>
        <Add filter=geo/>
        <Add filter_name=score />
        <Add filter_name=rama />
        <Add filter_name=sspred />
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML file used for generating beta-barrel backbones based on the 2D map: bb_2D_assembly.xml

***# see Supplementary Data: bb_2d.bp, bb_2d.cst***

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
        <ScoreFunction name=”SFXN1” weights="fldsgn_cen_omega02.wts" >
            <Reweight scoretype=”atom_pair_constraint” weight=”1.0” />
            <Reweight scoretype=”angle_constraint” weight=”1.0” />
            <Reweight scoretype=”dihedral_constraint” weight=”1.0” />
            <Reweight scoretype=”coordinate_constraint” weight=”1.0” />
         </ScoreFunction>
        </SCOREFXNS>
        <FILTERS>
            <SecondaryStructure name=”ss1” use_abego=”0” blueprint="bb_2d.bp" confidence=”1” cutoff="0.9" />
            <SheetTopology  name=”st1”  topology="1-2.A.99;2-3.A.99;3-4.A.99;4-5.A.99;5-6.A.99;6-7.A.99;7-8.A.99;1-8.A.99"
blueprint="blueprint" confidence=”1”/>
            <CompoundStatement name=”secst1” >
                <AND filter_name=”ss1” />
                <AND filter_name=”st1” />
            </CompoundStatement>
            <ScoreType name="cen_total" scorefxn="SFXN1" score_type="total_score" threshold="1000000" />
            <ScoreType name="vdw" scorefxn="SFXN1" score_type="vdw" threshold="1000000" />
            <ScoreType name="rg" scorefxn="SFXN1" score_type="rg" threshold="1000000" />
            <ScoreType name="cen_rama" scorefxn="SFXN1" score_type="rama" threshold="1000000" />
            <ScoreType name="sspair" scorefxn="SFXN1" score_type="ss_pair" threshold="1000000" />
            <ScoreType name="rsigma" scorefxn="SFXN1" score_type="rsigma" threshold="1000000" />
        </FILTERS>
        <TASKOPERATIONS>
        </TASKOPERATIONS>
        <MOVERS>
            ////////////////////////////////////
```

```
                    // General movers//
                    ////////////////////////////////////
                    <Dssp name="dssp"/>
                    <SwitchResidueTypeSetMover name="fullatom" set="fa_standard"/>
                    <SwitchResidueTypeSetMover name="cent" set="centroid"/>
                    ////////////////////////////////////
                    // SHEET-BUILDING //
                    ////////////////////////////////////
                    <BluePrintBDR name="bdr1" scorefxn="SFXN1" use_abego_bias="1" blueprint="bb_2d.bp"
constraint_file="bb_2d.cst"/>
                    <ConstraintSetMover name="addcst1" add_constraints="1" cst_file="bb_2d.cst"/>
                    <MinMover name="min1" scorefxn="SFXN1" chi="1" bb="1" type="dfpmin_armijo_nonmonotone_atol"
tolerance="0.0001"/>
                    <ParsedProtocol name="cenmin1" >
                              <Add mover_name="cent" />
                              <Add mover_name="addcst1" />
                              <Add mover_name="min1" />
                              <Add mover_name="fullatom" />
                    </ParsedProtocol>
                    <ParsedProtocol name="bdr1ss" >
                              <Add mover_name="bdr1" />
                              <Add mover_name="cenmin1" />
                              <Add mover_name="dssp" />
                    </ParsedProtocol>
                    <LoopOver name="loop1" mover_name="bdr1ss" filter_name="secst1" drift="0" iterations="20"
ms_whenfail="FAIL_DO_NOT_RETRY"/>
          </MOVERS>
          <APPLY_TO_POSE>
          </APPLY_TO_POSE>
          <PROTOCOLS>
                    <Add mover_name="loop1" />
                    <Add mover_name="fullatom" />
                    <Add filter_name="cen_total" />
                    <Add filter_name="vdw" />
                    <Add filter_name="rg" />
                    <Add filter_name="cen_rama" />
                    <Add filter_name="sspair" />
                    <Add filter_name="rsigma" />
          </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML file used for sequence design of beta-barrel backbones from 2D map-based fragment assembly:
## bb_2D_design.xml

```
# see Supplementary Data: bb_2d.res,  cst_trp.cst
<ROSETTASCRIPTS>
          <SCOREFXNS>
                    <SFX2 weights=beta_nov15_cst.wts>
                    </SFX2>
```

```
	</SCOREFXNS>
	<TASKOPERATIONS>
		<LayerDesign name=all_layers layer=all pore_radius=2.0 use_sidechain_neighbors="True" make_pymol_script=1 core=2.1
surface=1.0>
			<core>
				<all append="M" specification="designable" operation="design" />
			</core>
			<surface>
				<all specification="designable" operation="design" />
			</surface>
			<boundary>
				<all specification="designable" operation="design" />
			</boundary>
			<OperateOnResidueSubset name="gly_pro" >
				<PreventRepackingRLT/>
				<Index resnums="8,9,20,25,31,34,43,48,50,53,55,62,74,79,86,98,103"/>
				<all specification="fixed" operation="omit" />
			</OperateOnResidueSubset>
			<OperateOnResidueSubset name="no_design">
				<PreventRepackingRLT/>
				<Index resnums="107,11" />
				<all specification="fixed" operation="omit" />
			</OperateOnResidueSubset>
		</LayerDesign>
		<OperateOnResidueSubset name="exclude_pro_gly">
			<RestrictToRepackingRLT/>
			<Index resnums="8,9,20,25,31,34,43,48,50,53,55,62,74,79,86,98,103"/>
		</OperateOnResidueSubset>
		<ReadResfile name="resfile" filename="bb_2d.res" />
	</TASKOPERATIONS>
	<FILTERS>
			<Geometry name=geo omega=165 cart_bonded=20 confidence=0 />
			<PackStat name=packstat threshold=0.4 confidence=0 />
			<SSPrediction name="sspred" confidence="0" threshold=0.4 use_svm="1" use_probability="1"/>
	</FILTERS>
	<MOVERS>
			<Dssp name=dssp/>
			<FastDesign name="fdesign" task_operations=all_layers,exclude_pro_gly,resfile scorefxn="SFX2"
max_redesigns="2000" cst_file="cst_trp.cst" ramp_down_constraints="1" />
			<MutateResidue name="Pro31" target="31A" new_res="PRO" />
			<MutateResidue name="Pro50" target="50A" new_res="PRO" />
			<MutateResidue name="Pro34" target="34A" new_res="PRO" />
			<MutateResidue name="Pro62" target="62A" new_res="PRO" />
			<MutateResidue name="Pro86" target="86A" new_res="PRO" />
			<MutateResidue name="Pro8" target="8A" new_res="PRO" />
	</MOVERS>
	<APPLY_TO_POSE>
	</APPLY_TO_POSE>
	<PROTOCOLS>
			<Add mover_name="Pro31" />
			<Add mover_name="Pro50" />
			<Add mover_name="Pro34" />
			<Add mover_name="Pro62" />
			<Add mover_name="Pro86" />
			<Add mover_name="Pro8" />
			<Add mover_name="fdesign" />
			<Add filter=packstat/>
```

```
                <Add filter=geo/>
                <Add filter_name=sspred />
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML files used for designing DFHBI-binding beta-barrels:

## hbi_p2_rectBarrel.xml *(used for performing 2-step iterative design calculations)*

***(see Supplementary Data for rectBarrel.resfile)***

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <beta weights="beta" />
        </SCOREFXNS>
        <RESIDUE_SELECTORS>
         ###### Basic Residue Selectors######
                # rif residues
                <ResiduePDBInfoHasLabel name="rifRes" property="RIFRES" />
                # ligand neighborhood
                <Neighborhood name="ligNeighborRes" distance="10.0" >
                <Chain chains="B" />
                </Neighborhood>
                # core, boundary, surface, cutoff values are specific for beta barrels
                <Layer name="coreRes" select_core="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
                <Layer name="boundRes" select_boundary="true" use_sidechain_neighbors="true" core_cutoff="2.1"
surface_cutoff="1.0"/>
                <Layer name="surfRes" select_surface="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
                # secondary structure, specific to your scaffold
                <SecondaryStructure name="all" ss="HEL"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEE
EELLEEEEEEELLLEEEEEEEEEL"/>
                <SecondaryStructure name="helix" ss="H"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEE
EELLEEEEEEELLLEEEEEEEEEL"/>
                <SecondaryStructure name="strand" ss="E"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEE
EELLEEEEEEELLLEEEEEEEEEL"/>
                <SecondaryStructure name="loop" ss="L"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEE
EELLEEEEEEELLLEEEEEEEEEL"/>
                # resfile residue
                <Index name="resfile_res"
resnums="8,9,11,18,19,20,25,29,31,32,33,34,35,43,46,47,48,50,53,55,59,60,61,62,63,72,73,74,75,79,83,84,85,86,87,97,98,103,107"/>
                <Not name="nonresfile_res" selector="resfile_res"/>
                # amino acids
                <ResidueName name="polarAA" residue_name3="ASP,GLU,ASN,GLN,LYS,ARG,SER,HIS,THR,TYR" />
        ##### Combinatotiral Selectors #####
                <And name="coreAll" selectors="coreRes,all,nonresfile_res" />
                <And name="coreH" selectors="coreRes,helix,nonresfile_res" />
                <And name="coreE" selectors="coreRes,strand,nonresfile_res" />
                <And name="coreL" selectors="coreRes,loop,nonresfile_res" />
                <And name="boundAll" selectors="boundRes,all,nonresfile_res" />
                <And name="boundH" selectors="boundRes,helix,nonresfile_res" />
                <And name="boundE" selectors="boundRes,strand,nonresfile_res" />
                <And name="boundL" selectors="boundRes,loop,nonresfile_res" />
                <And name="surfAll" selectors="surfRes,all,nonresfile_res" />
                <And name="surfH" selectors="surfRes,helix,nonresfile_res" />
```

```
        <And name="surfE" selectors="surfRes,strand,nonresfile_res" />
        <And name="surfL" selectors="surfRes,loop,nonresfile_res" />
        <And name="coreLigZone" selectors="coreRes,ligNeighborRes,nonresfile_res" />
        <And name="boundLigZone" selectors="boundRes,ligNeighborRes,nonresfile_res" />
        <And name="surfLigZone" selectors="surfRes,ligNeighborRes,nonresfile_res" />
        <Or name="pocket" selectors="coreLigZone,boundLigZone"/>
        <Not name="nonpocket" selector="pocket"/>
        <And name="nonpocket_nonres" selectors="nonpocket,nonresfile_res"/>
        <And name="polarRifRes" selectors="rifRes,polarAA"/>
</RESIDUE_SELECTORS>
<TASKOPERATIONS>
        <LimitAromaChi2 name="limchi2"/>
        <InitializeFromCommandline name="init"/>
        <ReadResfile name="resfile" filename="rectBarrel.resfile"/>
        <IncludeCurrent name="includeCurrent"/>
        <RestrictToRepacking name="repack_only" />
        # to test each selector
        <OperateOnResidueSubset name="test" selector="boundRes" >
                <PreventRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="fix_rifRes" selector="rifRes" >
                <PreventRepackingRLT/>
        </OperateOnResidueSubset>

        # for pocket
        <OperateOnResidueSubset name="design_coreLigZone_AA" selector="coreLigZone" >
                <RestrictAbsentCanonicalAASRLT aas="ASTFVYMILNQHW"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_boundLigZone_AA" selector="boundLigZone" >
                <RestrictAbsentCanonicalAASRLT aas="ASTFVYMILNQHW"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_pocketRes" selector="pocket" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_nonpocketRes" selector="nonpocket" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_nonpocket_nonres_Res" selector="nonpocket_nonres" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        # for structure elements
        <OperateOnResidueSubset name="design_allCore_AA" selector="coreAll" >
                <RestrictAbsentCanonicalAASRLT aas="AFILVWYM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_helixCore_AA" selector="coreH" >
                <RestrictAbsentCanonicalAASRLT aas="AFILVM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_strandCore_AA" selector="coreE" >
                <RestrictAbsentCanonicalAASRLT aas="FILVM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_loopCore_AA" selector="coreL" >
                <RestrictAbsentCanonicalAASRLT aas="AGFILPVWYM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_allBound_AA" selector="boundAll" >
                <RestrictAbsentCanonicalAASRLT aas="ADEFGIKLNPQRSTVY"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_helixBound_AA" selector="boundH" >
```

```
                        <RestrictAbsentCanonicalAASRLT aas="AILNQSTVY"/>
                    </OperateOnResidueSubset>
                    <OperateOnResidueSubset name="design_strandBound_AA" selector="boundE" >
                        <RestrictAbsentCanonicalAASRLT aas="ILNQSTVY"/>
                    </OperateOnResidueSubset>
                    <OperateOnResidueSubset name="design_loopBound_AA" selector="boundL" >
                        <RestrictAbsentCanonicalAASRLT aas="ADEFGIKLNPQRSTVY"/>
                    </OperateOnResidueSubset>
                    <OperateOnResidueSubset name="design_allSurf_AA" selector="surfAll" >
                        <RestrictAbsentCanonicalAASRLT aas="DEGHKNPQRST"/>
                    </OperateOnResidueSubset>
                    <OperateOnResidueSubset name="design_helixSurf_AA" selector="surfH" >
                        <RestrictAbsentCanonicalAASRLT aas="DEHKNQRST"/>
                    </OperateOnResidueSubset>
                    <OperateOnResidueSubset name="design_strandSurf_AA" selector="surfE" >
                        <RestrictAbsentCanonicalAASRLT aas="DHKNQRT"/>
                    </OperateOnResidueSubset>
                    <OperateOnResidueSubset name="design_loopSurf_AA" selector="surfL" >
                        <RestrictAbsentCanonicalAASRLT aas="DEGHKNPQRST"/>
                    </OperateOnResidueSubset>
            </TASKOPERATIONS>
            <FILTERS>
                    <LigInterfaceEnergy name="interfE" scorefxn="beta" energy_cutoff="9999"/>
                    <ShapeComplementarity name="SC" min_sc="0" min_interface="0" verbose="0" quick="0" jump="1"/>
                    <ScoreType name="totalscore" scorefxn="beta" threshold="9999" confidence="1"/>
                    <ResidueCount name="nres" confidence="1" />
                    <CalculatorFilter name="res_totalscore" confidence="1" equation="SCORE/NRES" threshold="0">
                        <SCORE name="SCORE" filter_name="totalscore" />
                        <NRES name="NRES" filter_name="nres" />
                    </CalculatorFilter>
                    <BuriedUnsatHbonds2 name="interf_uhb2" cutoff="200" scorefxn="beta" jump_number="1" />
                    <RepackWithoutLigand name="rwl" scorefxn="beta" target_res="all_repacked" rms_threshold="999"/>
                    #DFHBI specific atom hbond filters
                    <HbondsToAtom name="O1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O1" res_num="110"/>
                    <HbondsToAtom name="O2_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O2" res_num="110"/>
                    <HbondsToAtom name="N1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="N1" res_num="110"/>
            </FILTERS>
            <MOVERS>
                    <EnzRepackMinimize name="desmin_fixrif_pocket" design="1" repack_only="0" scorefxn_minimize="beta"
scorefxn_repack="beta" minimize_rb=1 minimize_sc=1 minimize_bb=0 cycles=1 minimize_lig=1 min_in_stages=0 backrub=0
task_operations="init,limchi2,fix_rifRes,resfile,design_coreLigZone_AA,design_boundLigZone_AA,repack_nonpocket_nonres_Res"/>
                    <FastDesign name="fdesign_nonpocket"
task_operations="init,limchi2,resfile,fix_rifRes,includeCurrent,repack_pocketRes,design_helixCore_AA,design_strandCore_AA,design_loopCor
e_AA,design_helixBound_AA,design_strandBound_AA,design_loopBound_AA,design_helixSurf_AA,design_strandSurf_AA,design_loopSurf_
AA"  scorefxn="beta" repeats="1" clear_designable_residues="0"  />
                    #calculate a myriad of ligand specific scores
                    <InterfaceScoreCalculator name="add_scores" chains="B" scorefxn="beta"/>
            </MOVERS>
            <PROTOCOLS>
                    <Add filter_name="SC"/>
                    <Add filter_name="interfE"/>
                    <Add mover_name="desmin_fixrif_pocket"/>
                    <Add mover_name="fdesign_nonpocket"/>
                    <Add filter_name="SC"/>
```

```
                <Add filter_name="interfE"/>
                <Add filter_name="res_totalscore"/>
                <Add mover_name="desmin_fixrif_pocket"/>
                <Add mover_name="fdesign_nonpocket"/>
                <Add filter_name="SC"/>
                <Add filter_name="interfE"/>
                <Add filter_name="res_totalscore"/>
                <Add mover_name="desmin_fixrif_pocket"/>
                <Add mover_name="fdesign_nonpocket"/>
                <Add filter_name="SC"/>
                <Add filter_name="interfE"/>
                <Add filter_name="interf_uhb2"/>
                <Add filter_name="rwl"/>
                <Add filter_name="res_totalscore"/>
                <Add filter_name="O1_hbond"/>
                <Add filter_name="O2_hbond"/>
                <Add filter_name="N1_hbond"/>
                <Add mover_name="add_scores"/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

## hbi_p2_rectBarrel_aacomp.xml *(added an additional mover to control amino acid composition in the packing core that favors aromatic residues and disfavors Methionine )*

***(see Supplementary Data for favour_core_aromatics.comp)***

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <ScoreFunction name="beta" weights="beta" />
                <ScoreFunction name="beta_aa" weights="beta" >
                <Reweight scoretype="aa_composition" weight="1.0" />
                <Set aa_composition_setup_file="favour_core_aromatics.comp" />
                </ScoreFunction>
        </SCOREFXNS>
        <RESIDUE_SELECTORS>
        ###### Basic Residue Selectors######
                # rif residues
                <ResiduePDBInfoHasLabel name="rifRes" property="RIFRES" />
                # ligand neighborhood
                <Neighborhood name="ligNeighborRes" distance="10.0" >
                        <Chain chains="B" />
                </Neighborhood>
                # core, boundary, surface, cutoff values are specific for beta barrels
                <Layer name="coreRes" select_core="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
                <Layer name="boundRes" select_boundary="true" use_sidechain_neighbors="true" core_cutoff="2.1"
surface_cutoff="1.0"/>
                <Layer name="surfRes" select_surface="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>

                # secondary structure, specific to your scaffold
                <SecondaryStructure name="all" ss="HEL"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL"/>
                <SecondaryStructure name="helix" ss="H"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL"/>
                <SecondaryStructure name="strand" ss="E"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL"/>
```

```
                <SecondaryStructure name="loop" ss="L"
pose_secstruct="LLHHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEELLLEEEEEEEEE
EELLEEEEEEELLLEEEEEEEEEL"/>
                # resfile residue
                <Index name="resfile_res"
resnums="8,9,11,18,19,20,25,29,31,32,33,34,35,43,46,47,48,50,53,55,59,60,61,62,63,72,73,74,75,79,83,84,85,86,87,97,98,103,107"/>
                <Not name="nonresfile_res" selector="resfile_res"/>
                 # amino acids
                <ResidueName name="polarAA" residue_name3="ASP,GLU,ASN,GLN,LYS,ARG,SER,HIS,THR,TYR" />
                # aa composition
                <Index name="packing_core" resnums="3,7,13,27,37,39,57,65,67,81,89,91,105"/>
        ##### Combinatorial Selectors #####
                <And name="coreAll" selectors="coreRes,all,nonresfile_res" />
                <And name="coreH" selectors="coreRes,helix,nonresfile_res" />
                <And name="coreE" selectors="coreRes,strand,nonresfile_res" />
                <And name="coreL" selectors="coreRes,loop,nonresfile_res" />
                <And name="boundAll" selectors="boundRes,all,nonresfile_res" />
                <And name="boundH" selectors="boundRes,helix,nonresfile_res" />
                <And name="boundE" selectors="boundRes,strand,nonresfile_res" />
                <And name="boundL" selectors="boundRes,loop,nonresfile_res" />
                <And name="surfAll" selectors="surfRes,all,nonresfile_res" />
                <And name="surfH" selectors="surfRes,helix,nonresfile_res" />
                <And name="surfE" selectors="surfRes,strand,nonresfile_res" />
                <And name="surfL" selectors="surfRes,loop,nonresfile_res" />
                <And name="coreLigZone" selectors="coreRes,ligNeighborRes,nonresfile_res" />
                <And name="boundLigZone" selectors="boundRes,ligNeighborRes,nonresfile_res" />
                <And name="surfLigZone" selectors="surfRes,ligNeighborRes,nonresfile_res" />
                <Or name="pocket" selectors="coreLigZone,boundLigZone"/>
                <Not name="nonpocket" selector="pocket"/>
                <And name="nonpocket_nonres" selectors="nonpocket,nonresfile_res"/>
                <And name="polarRifRes" selectors="rifRes,polarAA"/>
        </RESIDUE_SELECTORS>
        <TASKOPERATIONS>
                <LimitAromaChi2 name="limchi2"/>
                <InitializeFromCommandline name="init"/>
                <ReadResfile name="resfile" filename="rectBarrel.resfile"/>
                <IncludeCurrent name="includeCurrent"/>
                <RestrictToRepacking name="repack_only" />
                # to test each selector
                <OperateOnResidueSubset name="test" selector="boundRes" >
                        <PreventRepackingRLT/>
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="fix_rifRes" selector="rifRes" >
                        <PreventRepackingRLT/>
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="repack_rifRes" selector="rifRes" >
                        <RestrictToRepackingRLT/>
                </OperateOnResidueSubset>
                # for pocket
                <OperateOnResidueSubset name="design_coreLigZone_AA" selector="coreLigZone" >
                        <RestrictAbsentCanonicalAASRLT aas="ASTFVYMILNQHW"/>
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="design_boundLigZone_AA" selector="boundLigZone" >
                        <RestrictAbsentCanonicalAASRLT aas="ASTFVYMILNQHW"/>
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="repack_pocketRes" selector="pocket" >
                        <RestrictToRepackingRLT/>
```

```
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="repack_nonpocketRes" selector="nonpocket" >
			<RestrictToRepackingRLT/>
		</OperateOnResidueSubset>
		# for structure elements
		<OperateOnResidueSubset name="design_allCore_AA" selector="coreAll" >
			<RestrictAbsentCanonicalAASRLT aas="AFILVWYM"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_helixCore_AA" selector="coreH" >
			<RestrictAbsentCanonicalAASRLT aas="AFILVM"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_strandCore_AA" selector="coreE" >
			<RestrictAbsentCanonicalAASRLT aas="FILVM"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_loopCore_AA" selector="coreL" >
			<RestrictAbsentCanonicalAASRLT aas="AGFILPVWYM"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_allBound_AA" selector="boundAll" >
			<RestrictAbsentCanonicalAASRLT aas="ADEFGIKLNPQRSTVY"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_helixBound_AA" selector="boundH" >
			<RestrictAbsentCanonicalAASRLT aas="AILNQSTVY"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_strandBound_AA" selector="boundE" >
			<RestrictAbsentCanonicalAASRLT aas="ILNQSTVY"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_loopBound_AA" selector="boundL" >
			<RestrictAbsentCanonicalAASRLT aas="ADEFGIKLNPQRSTVY"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_allSurf_AA" selector="surfAll" >
			<RestrictAbsentCanonicalAASRLT aas="DEGHKNPQRST"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_helixSurf_AA" selector="surfH" >
			<RestrictAbsentCanonicalAASRLT aas="DEHKNQRST"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_strandSurf_AA" selector="surfE" >
			<RestrictAbsentCanonicalAASRLT aas="DHKNQRT"/>
		</OperateOnResidueSubset>
		<OperateOnResidueSubset name="design_loopSurf_AA" selector="surfL" >
			<RestrictAbsentCanonicalAASRLT aas="DEGHKNPQRST"/>
		</OperateOnResidueSubset>
	</TASKOPERATIONS>
	<FILTERS>
		<LigInterfaceEnergy name="interfE" scorefxn="beta" energy_cutoff="9999"/>
		<ShapeComplementarity name="SC" min_sc="0" min_interface="0" verbose="0" quick="0" jump="1"/>
		<ScoreType name="totalscore" scorefxn="beta" threshold="9999" confidence="1"/>
		<ResidueCount name="nres" confidence="1" />
		<CalculatorFilter name="res_totalscore" confidence="1" equation="SCORE/NRES" threshold="0">
			<Var name="SCORE" filter_name="totalscore" />
			<Var name="NRES" filter_name="nres" />
		</CalculatorFilter>
		<BuriedUnsatHbonds2 name="interf_uhb2" cutoff="200" scorefxn="beta" jump_number="1" />
		<RepackWithoutLigand name="rwl" scorefxn="beta" target_res="all_repacked" rms_threshold="999"/>
		#DFHBI specific atom hbond filters
		<HbondsToAtom name="O1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O1" res_num="110"/>
```

```
		<HbondsToAtom name="O2_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O2" res_num="110"/>
		<HbondsToAtom name="N1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="N1" res_num="110"/>
	</FILTERS>
	<MOVERS>
		<AddCompositionConstraintMover name="core_arom_AA" filename="favour_core_aromatics.comp"
selector="packing_core" />
		<ClearCompositionConstraintsMover name="clear_AA_constraints"/>
		<EnzRepackMinimize name="desmin_fixrif_pocket" design="1" repack_only="0" scorefxn_minimize="beta"
scorefxn_repack="beta" minimize_rb="1" minimize_sc="1" minimize_bb="0" cycles="1" minimize_lig="1" min_in_stages="0" backrub="0"
task_operations="init,limchi2,fix_rifRes,resfile,design_coreLigZone_AA,design_boundLigZone_AA,repack_nonpocket_nonres_Res"/>
		<FastDesign name="fdesign_nonpocket"
			task_operations="init,limchi2,resfile,fix_rifRes,includeCurrent,repack_pocketRes,design_helixCore_AA,design
			_strandCore_AA,design_loopCore_AA,design_helixBound_AA,design_strandBound_AA,design_loopBound_
			AA,design_helixSurf_AA,design_strandSurf_AA,design_loopSurf_AA"
			scorefxn="beta_aa"
			repeats="1"
			clear_designable_residues="0"  />
		#calculate a myriad of ligand specific scores
		<InterfaceScoreCalculator name="add_scores" chains="B" scorefxn="beta"/>
	</MOVERS>
	<PROTOCOLS>
		<Add mover_name="core_arom_AA"/>
		<Add mover_name="desmin_fixrif_pocket"/>
		<Add mover_name="fdesign_nonpocket"/>
			<Add filter_name="SC"/>
			<Add filter_name="interfE"/>
			<Add filter_name="res_totalscore"/>
		<Add mover_name="desmin_fixrif_pocket"/>
		<Add mover_name="fdesign_nonpocket"/>
			<Add filter_name="SC"/>
			<Add filter_name="interfE"/>
			<Add filter_name="res_totalscore"/>
			<Add mover_name="desmin_fixrif_pocket"/>
			<Add mover_name="fdesign_nonpocket"/>
			<Add mover_name="clear_AA_constraints"/>
			<Add filter_name="SC"/>
			<Add filter_name="interfE"/>
			<Add filter_name="interf_uhb2"/>
			<Add filter_name="res_totalscore"/>
			<Add filter_name="O1_hbond"/>
			<Add filter_name="O2_hbond"/>
			<Add filter_name="N1_hbond"/>
		<Add mover_name="add_scores"/>
	</PROTOCOLS>
</ROSETTASCRIPTS>
```

## hbi_p2_rectBarrel_releaserif.xml *(rotamer fixation constraints were released for RIF coordinating residues)*

## *(identical to hbi_p2_rectBarrel_aacomp.xml except for the definition of FastDesign)*

```
<FastDesign name="fdesign_nonpocket"
	task_operations="init,limchi2,resfile,repack_rifRes,includeCurrent,repack_pocketRes,design_helixCore_AA,design_strandCore_AA,d
	esign_loopCore_AA,design_helixBound_AA,design_strandBound_AA,design_loopBound_AA,design_helixSurf_AA,design_strandS
	urf_AA,design_loopSurf_AA"
	scorefxn="beta_aa"
```

```
        repeats="1"
        clear_designable_residues="0"  />
```

## resfile_design.xml *(used to perform profile-based sequence design for clustered designs)*

***(see Supplementary Data for favour_core_aromatics.comp  final.resfile)***

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <beta weights="beta" />
                <beta_aa weights="beta" >
                        <Reweight scoretype="aa_composition" weight="1.0" />
                        <Set aa_composition_setup_file="favour_core_aromatics.comp" />
                </beta_aa>
        </SCOREFXNS>
        <RESIDUE_SELECTORS>
                <ResidueName name="polarAA" residue_name3="ASP,GLU,ASN,GLN,LYS,ARG,SER,HIS" />
                <ResidueName name="Phe" residue_name3="PHE" />
                <ResidueName name="Met" residue_name3="MET" />
                <Index name="packing_core" resnums="3,7,13,27,37,39,57,65,67,81,89,91,105"/>
        </RESIDUE_SELECTORS>
        <TASKOPERATIONS>
                <LimitAromaChi2 name="limchi2"/>
                <InitializeFromCommandline name="init"/>
                <ReadResfile name="resfile" filename="final.resfile"/>
        </TASKOPERATIONS>
        <FILTERS>
                <LigInterfaceEnergy name="interfE" scorefxn="beta" energy_cutoff="0.0"/>
                <ShapeComplementarity name="SC" min_sc=0.1 min_interface=0 verbose=0 quick=0 jump=1/>
                <ScoreType name="totalscore" scorefxn="beta" threshold="0" confidence=1/>
                <ResidueCount name="nres" confidence="1" />
                <CalculatorFilter name="res_totalscore" confidence="1" equation="SCORE/NRES" threshold="0">
                        <SCORE name="SCORE" filter_name="totalscore" />
                        <NRES name="NRES" filter_name="nres" />
                </CalculatorFilter>
                 #buried unsatisfied polar
                <BuriedUnsatHbonds2 name="interf_uhb2" cutoff="200" scorefxn="beta" jump_number="1" />
                #DFHBI specific atom hbond filters
                <HbondsToAtom name="O1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O1" res_num="110"/>
                <HbondsToAtom name="O2_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O2" res_num="110"/>
                <HbondsToAtom name="N1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="N1" res_num="110"/>
                <ResidueCount name="packing_F" residue_types="PHE" residue_selector="packing_core" />
                <ResidueCount name="packing_M" residue_types="MET" residue_selector="packing_core" />
                 <CavityVolume name="cavity" />
                <InterfaceHoles name="interface_hole" jump="1"  threshold=200/>
                <PackStat name="packstat_complex" threshold="0" chain="0" repeats=1/>
                <PackStat name="packstat_apo" threshold="0" chain="1" repeats=1/>
        </FILTERS>
        <MOVERS>
                <AddCompositionConstraintMover name="core_arom_AA" filename="favour_core_aromatics.comp"
selector="packing_core" />
                <ClearCompositionConstraintsMover name="clear_AA_constraints" />
                <FastDesign name="fdesign_complex"
                        task_operations="init,limchi2,resfile"
                        scorefxn="beta_aa"
                        repeats="3"
```

```
                clear_designable_residues="0"  />
        <InterfaceScoreCalculator name="add_scores" chains="B" scorefxn="beta"/>
    </MOVERS>
     <PROTOCOLS>
        <Add mover_name="core_arom_AA"/>
        <Add mover_name="fdesign_complex"/>
        <Add mover_name="clear_AA_constraints"/>
            <Add filter_name="SC"/>
            <Add filter_name="interfE"/>
            <Add filter_name="res_totalscore"/>
            <Add filter_name="interf_uhb2"/>
            <Add filter_name="O1_hbond"/>
            <Add filter_name="N1_hbond"/>
            <Add filter_name="O2_hbond"/>
            <Add filter_name="packing_F"/>
            <Add filter_name="packing_M"/>
            <Add filter_name="cavity"/>
            <Add filter_name="interface_hole"/>
            <Add filter_name="packstat_complex"/>
            <Add filter_name="packstat_apo"/>
        <Add mover_name="add_scores"/>
     </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML file used for building loop 5F: remodel_L5F.xml

***(see Supplementary Data  for loop5F.bp)***

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <SFX1 weights="beta_nov15_cst.wts">
        </SFX1>
    </SCOREFXNS>
    <TASKOPERATIONS>
    </TASKOPERATIONS>
    <FILTERS>
        <LigInterfaceEnergy name="interfE" scorefxn="SFX1" energy_cutoff="9999"/>
        <ShapeComplementarity name="SC" min_sc="0" min_interface="0" verbose="0" quick="0" jump="1"/>
        <BuriedUnsatHbonds2 name="interf_uhb2" cutoff="200" scorefxn="SFX1" jump_number="1" />
        <HbondsToAtom  name="O1_hbond"  partners="0"  energy_cutoff="-0.5"  backbone="0"  bb_bb="0"  sidechain="1"
atomname="O1" res_num="114"/>
        <HbondsToAtom  name="O2_hbond"  partners="0"  energy_cutoff="-0.5"  backbone="0"  bb_bb="0"  sidechain="1"
atomname="O2" res_num="114"/>
        <HbondsToAtom  name="N1_hbond"  partners="0"  energy_cutoff="-0.5"  backbone="0"  bb_bb="0"  sidechain="1"
atomname="N1" res_num="114"/>
    </FILTERS>
    <MOVERS>
        <RemodelMover name="remodel" blueprint="loop5F.bp" quick_and_dirty="0" />
    </MOVERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
        <Add mover_name="remodel" />
        <Add filter_name="interfE" />
        <Add filter_name="SC" />
        <Add filter_name="interf_uhb2" />
        <Add filter_name="O1_hbond" />
```

```
                <Add filter_name="O2_hbond" />
                <Add filter_name="N1_hbond" />
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML file used for designing sequences for loop 5F: pocket_loop_redesign.xml

***(see Supplementary Data: loop5F.resfile)***

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <beta weights="beta" />
                <beta_soft weights="beta_nov15_soft" />
        </SCOREFXNS>
<RESIDUE_SELECTORS>
        # ligand neighborhood
        <Neighborhood name="ligNeighborRes" distance="9.0" >
                <Chain chains="B" />
        </Neighborhood>
        # core, boundary, surface, cutoff values are specific for beta barrels
        <Layer name="coreRes" select_core="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
        <Layer name="boundRes" select_boundary="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
        <Layer name="surfRes" select_surface="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
        <Index name="resfile_res" resnums="70,71,72,73,74,75,76,77,78,52,21,79,19,47,99,51,45"/>
        <Not name="nonresfile_res" selector="resfile_res"/>
        <And name="coreLigZone" selectors="coreRes,ligNeighborRes" />
        <And name="boundLigZone" selectors="boundRes,ligNeighborRes" />
        <And name="surfLigZone" selectors="surfRes,ligNeighborRes" />
        <Or name="pocket" selectors="coreLigZone,boundLigZone"/>
        <Not name="nonpocket" selector="pocket"/>
         # to be fixed during ligandbinding design, repackable during fast design
        <And name="nonpocket_nonresfile" selectors="nonpocket,nonresfile_res"/>
         # to be designed during ligandbinding design; repackable during fast design
        <And name="pocket_resfile" selectors="pocket,resfile_res"/>
        # to be repackage during ligand binding design; to be designed during fast design;
        <And name="nonpocket_resfile" selectors="nonpocket,resfile_res"/>
        # to be repackable during ligand binding design;
        <And name="pocket_nonresfile" selectors="pocket,nonresfile_res"/>
 </RESIDUE_SELECTORS>
 <TASKOPERATIONS>
        <LimitAromaChi2 name="limchi2"/>
        <InitializeFromCommandline name="init"/>
        <ReadResfile name="resfile" filename="./loop5F.resfile"/>
        <IncludeCurrent name="includeCurrent"/>
        <RestrictToRepacking name="repack_only" />
        # to test each selector
        <OperateOnResidueSubset name="test" selector="boundRes" >
                <PreventRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="fix_nonpocketnonresfileRes" selector="nonpocket_nonresfile" >
                <PreventRepackingRLT/>
        </OperateOnResidueSubset>
         <OperateOnResidueSubset name="repack_nonpocketnonresfileRes" selector="nonpocket_nonresfile" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_nonpocketresfileRes" selector="nonpocket_resfile" >
                <RestrictToRepackingRLT/>
```

```
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_pocketnonresfileRes" selector="pocket_nonresfile" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_pocketRes" selector="pocket" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_nonresfileRes" selector="nonresfile_res" >
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
  </TASKOPERATIONS>
  <FILTERS>
        <LigInterfaceEnergy name="interfE" scorefxn="beta" energy_cutoff="9999"/>
        <ShapeComplementarity name="SC" min_sc="0" min_interface="0" verbose="0" quick="0" jump="1"/>
        <ScoreType name="totalscore" scorefxn="beta" threshold="9999" confidence="1"/>
        <ResidueCount name="nres" confidence="1" />
        <CalculatorFilter name="res_totalscore" confidence="1" equation="SCORE/NRES" threshold="0">
                <SCORE name="SCORE" filter_name="totalscore" />
                <NRES name="NRES" filter_name="nres" />
        </CalculatorFilter>
        <BuriedUnsatHbonds2 name="interf_uhb2" cutoff="200" scorefxn="beta" jump_number="1" />
        <RepackWithoutLigand name="rwl" scorefxn="beta" target_res="all_repacked" rms_threshold="999"/>
        #DFHBI specific atom hbond filters
        <HbondsToAtom name="O1_hbond" partners="0" energy_cutoff="-0.25" backbone="0" bb_bb="0" sidechain="1" atomname="O1"
res_num="112"/>
        <HbondsToAtom name="O2_hbond" partners="0" energy_cutoff="-0.25" backbone="0" bb_bb="0" sidechain="1" atomname="O2"
res_num="112"/>
        <HbondsToAtom name="N1_hbond" partners="0" energy_cutoff="-0.25" backbone="0" bb_bb="0" sidechain="1" atomname="N1"
res_num="112"/>
         <CavityVolume name="cavity" />
        <InterfaceHoles name="interface_hole" jump="1"  threshold="200"/>
        <PackStat name="packstat_complex" threshold="0" chain="0" repeats="1"/>
  </FILTERS>
  <MOVERS>
        <EnzRepackMinimize name="desmin_loop_pocket" design="1" repack_only="0" scorefxn_minimize="beta"
scorefxn_repack="beta_soft" minimize_rb="1" minimize_sc="1" minimize_bb="0" cycles="1" minimize_lig="1" min_in_stages="0"
backrub="0"
task_operations="init,includeCurrent,resfile,repack_pocketnonresfileRes,repack_nonpocketresfileRes,fix_nonpocketnonresfileRes"/>
        <FastDesign name="fdesign_loop"
                task_operations="init,limchi2,resfile,includeCurrent,repack_pocketRes"
                scorefxn="beta"
                repeats="1"
                clear_designable_residues="0"  />
        #calculate a myriad of ligand specific scores
        <InterfaceScoreCalculator name="add_scores" chains="B" scorefxn="beta"/>
  </MOVERS>
  <PROTOCOLS>
        <Add mover_name="desmin_loop_pocket"/>
        <Add mover_name="fdesign_loop"/>
        <Add mover_name="desmin_loop_pocket"/>
        <Add mover_name="fdesign_loop"/>
        <Add filter_name="SC"/>
        <Add filter_name="interfE"/>
        <Add filter_name="interf_uhb2"/>
        <Add filter_name="res_totalscore"/>
        <Add filter_name="O1_hbond"/>
        <Add filter_name="O2_hbond"/>
```

```
        <Add filter_name="N1_hbond"/>
        <Add filter_name="cavity"/>
        <Add filter_name="interface_hole"/>
        <Add filter_name="packstat_complex"/>
        <Add mover_name="add_scores"/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

## RosettaScripts XML file used for re-designing b11L5F.1: fixed_hbi_p2_rectBarrel.xml

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <ScoreFunction name="beta" weights="beta"/>
        </SCOREFXNS>
        <RESIDUE_SELECTORS>
                <ResiduePDBInfoHasLabel name="rifRes" property="RIFRES"/>
                <Neighborhood distance="10.0" name="ligNeighborRes">
                        <Chain chains="B"/>
                </Neighborhood>
                <Layer core_cutoff="2.1" name="coreRes" select_core="true" surface_cutoff="1.0" use_sidechain_neighbors="true"/>
                <Layer core_cutoff="2.1" name="boundRes" select_boundary="true" surface_cutoff="1.0"
use_sidechain_neighbors="true"/>
                <Layer core_cutoff="2.1" name="surfRes" select_surface="true" surface_cutoff="1.0" use_sidechain_neighbors="true"/>
                <SecondaryStructure name="all"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL" ss="HEL"/>
                <SecondaryStructure name="helix"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL" ss="H"/>
                <SecondaryStructure name="strand"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL" ss="E"/>
                <SecondaryStructure name="loop"
pose_secstruct="LLHHHHHLEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEEEELLLEEEEEEEEEEEELLEEEEEEEELLLEEEEEEEEE
EELLEEEEEEEELLLEEEEEEEEEL" ss="L"/>
                <Index name="resfile_res"
resnums="8,9,11,18,19,20,25,29,31,32,33,34,35,43,46,47,48,50,53,55,59,60,61,62,63,70,71,72,73,81,86,87,88,89,99,100,105,109"/>
                <Not name="nonresfile_res" selector="resfile_res"/>
                <ResidueName name="polarAA" residue_name3="ASP,GLU,ASN,GLN,LYS,ARG,SER,HIS,THR,TYR"/>
                <And name="coreAll" selectors="coreRes,all,nonresfile_res"/>
                <And name="coreH" selectors="coreRes,helix,nonresfile_res"/>
                <And name="coreE" selectors="coreRes,strand,nonresfile_res"/>
                <And name="coreL" selectors="coreRes,loop,nonresfile_res"/>
                <And name="boundAll" selectors="boundRes,all,nonresfile_res"/>
                <And name="boundH" selectors="boundRes,helix,nonresfile_res"/>
                <And name="boundE" selectors="boundRes,strand,nonresfile_res"/>
                <And name="boundL" selectors="boundRes,loop,nonresfile_res"/>
                <And name="surfAll" selectors="surfRes,all,nonresfile_res"/>
                <And name="surfH" selectors="surfRes,helix,nonresfile_res"/>
                <And name="surfE" selectors="surfRes,strand,nonresfile_res"/>
                <And name="surfL" selectors="surfRes,loop,nonresfile_res"/>
                <And name="coreLigZone" selectors="coreRes,ligNeighborRes,nonresfile_res"/>
                <And name="boundLigZone" selectors="boundRes,ligNeighborRes,nonresfile_res"/>
                <And name="surfLigZone" selectors="surfRes,ligNeighborRes,nonresfile_res"/>
                <Or name="pocket" selectors="coreLigZone,boundLigZone"/>
```

```
        <Not name="nonpocket" selector="pocket"/>
        <And name="nonpocket_nonres" selectors="nonpocket,nonresfile_res"/>
        <And name="polarRifRes" selectors="rifRes,polarAA"/>
</RESIDUE_SELECTORS>
<TASKOPERATIONS>
        <LimitAromaChi2 name="limchi2"/>
        <InitializeFromCommandline name="init"/>
        <ReadResfile filename="./rectBarrel.resfile" name="resfile"/>
        <IncludeCurrent name="includeCurrent"/>
        <RestrictToRepacking name="repack_only"/>
        <OperateOnResidueSubset name="test" selector="boundRes">
                <PreventRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="fix_rifRes" selector="rifRes">
                <PreventRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_coreLigZone_AA" selector="coreLigZone">
                <RestrictAbsentCanonicalAASRLT aas="ASTFVYMILNQHW"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_boundLigZone_AA" selector="boundLigZone">
                <RestrictAbsentCanonicalAASRLT aas="ASTFVYMILNQHW"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_pocketRes" selector="pocket">
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_nonpocketRes" selector="nonpocket">
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="repack_nonpocket_nonres_Res" selector="nonpocket_nonres">
                <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_allCore_AA" selector="coreAll">
                <RestrictAbsentCanonicalAASRLT aas="AFILVWYM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_helixCore_AA" selector="coreH">
                <RestrictAbsentCanonicalAASRLT aas="AFILVM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_strandCore_AA" selector="coreE">
                <RestrictAbsentCanonicalAASRLT aas="FILVM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_loopCore_AA" selector="coreL">
                <RestrictAbsentCanonicalAASRLT aas="AGFILPVWYM"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_allBound_AA" selector="boundAll">
                <RestrictAbsentCanonicalAASRLT aas="ADEFGIKLNPQRSTVY"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_helixBound_AA" selector="boundH">
                <RestrictAbsentCanonicalAASRLT aas="AILNQSTVY"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_strandBound_AA" selector="boundE">
                <RestrictAbsentCanonicalAASRLT aas="ILNQSTVY"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_loopBound_AA" selector="boundL">
                <RestrictAbsentCanonicalAASRLT aas="ADEFGIKLNPQRSTVY"/>
        </OperateOnResidueSubset>
        <OperateOnResidueSubset name="design_allSurf_AA" selector="surfAll">
                <RestrictAbsentCanonicalAASRLT aas="DEGHKNPQRST"/>
```

```
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="design_helixSurf_AA" selector="surfH">
                        <RestrictAbsentCanonicalAASRLT aas="DEHKNQRST"/>
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="design_strandSurf_AA" selector="surfE">
                        <RestrictAbsentCanonicalAASRLT aas="HKNQRT"/>
                </OperateOnResidueSubset>
                <OperateOnResidueSubset name="design_loopSurf_AA" selector="surfL">
                        <RestrictAbsentCanonicalAASRLT aas="DEGHKNPQRST"/>
                </OperateOnResidueSubset>
        </TASKOPERATIONS>
        <FILTERS>
                <LigInterfaceEnergy energy_cutoff="9999" name="interfE" scorefxn="beta"/>
                <ShapeComplementarity jump="1" min_interface="0" min_sc="0" name="SC" quick="0" verbose="0"/>
                <ScoreType confidence="1" name="totalscore" scorefxn="beta" threshold="9999"/>
                <ResidueCount confidence="1" name="nres"/>
                <CalculatorFilter confidence="1" equation="SCORE/NRES" name="res_totalscore" threshold="0">
                        <Var filter_name="totalscore" name="SCORE"/>
                        <Var filter_name="nres" name="NRES"/>
                </CalculatorFilter>
                <BuriedUnsatHbonds2 cutoff="200" jump_number="1" name="interf_uhb2" scorefxn="beta"/>
                <RepackWithoutLigand name="rwl" rms_threshold="999" scorefxn="beta" target_res="all_repacked"/>
                <HbondsToAtom atomname="O1" backbone="0" bb_bb="0" energy_cutoff="-0.5" name="O1_hbond" partners="0"
res_num="112" sidechain="1"/>
                <HbondsToAtom atomname="O2" backbone="0" bb_bb="0" energy_cutoff="-0.5" name="O2_hbond" partners="0"
res_num="112" sidechain="1"/>
                <HbondsToAtom atomname="N1" backbone="0" bb_bb="0" energy_cutoff="-0.5" name="N1_hbond" partners="0"
res_num="112" sidechain="1"/>
        </FILTERS>
        <MOVERS>
                <EnzRepackMinimize backrub="0" cycles="1" design="1" min_in_stages="0" minimize_bb="0" minimize_lig="1"
minimize_rb="1" minimize_sc="1" name="desmin_fixrif_pocket" repack_only="0" scorefxn_minimize="beta" scorefxn_repack="beta"
task_operations="init,limchi2,fix_rifRes,resfile,design_coreLigZone_AA,design_boundLigZone_AA,repack_nonpocket_nonres_Res"/>
                <FastDesign clear_designable_residues="0" name="fdesign_nonpocket" repeats="1" scorefxn="beta"
task_operations="init,limchi2,resfile,fix_rifRes,includeCurrent,repack_pocketRes,design_helixCore_AA,design_strandCore_AA,design_loopCor
e_AA,design_helixBound_AA,design_strandBound_AA,design_loopBound_AA,design_helixSurf_AA,design_strandSurf_AA,design_loopSurf_
AA"/>
                <InterfaceScoreCalculator chains="B" name="add_scores" scorefxn="beta"/>
        </MOVERS>
        <PROTOCOLS>
                        <Add filter_name="SC"/>
                        <Add filter_name="interfE"/>
                <Add mover_name="desmin_fixrif_pocket"/>
                <Add mover_name="fdesign_nonpocket"/>
                        <Add filter_name="SC"/>
                        <Add filter_name="interfE"/>
                        <Add filter_name="res_totalscore"/>
                <Add mover_name="desmin_fixrif_pocket"/>
                <Add mover_name="fdesign_nonpocket"/>
                        <Add filter_name="SC"/>
                        <Add filter_name="interfE"/>
                        <Add filter_name="res_totalscore"/>
                <Add mover_name="desmin_fixrif_pocket"/>
                <Add mover_name="fdesign_nonpocket"/>
                        <Add filter_name="SC"/>
                        <Add filter_name="interfE"/>
                        <Add filter_name="interf_uhb2"/>
```

```
                <Add filter_name="rwl"/>
                <Add filter_name="res_totalscore"/>
                <Add filter_name="O1_hbond"/>
                <Add filter_name="O2_hbond"/>
                <Add filter_name="N1_hbond"/>
        <Add mover_name="add_scores"/>
    </PROTOCOLS>
```

## RosettaScripts XML file used for modeling mutations from library selection: make_mutations.xml

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <beta weights="beta" />
    </SCOREFXNS>
    <RESIDUE_SELECTORS>
        ###### Basic Residue Selectors######
        # rif residues
        <ResiduePDBInfoHasLabel name="rifRes" property="RIFRES" />
        # ligand neighborhood
        <Neighborhood name="ligNeighborRes" distance="10.0" >
            <Chain chains="B" />
            </Neighborhood>
            # core, boundary, surface, cutoff values are specific for beta barrels
        <Layer name="coreRes" select_core="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
        <Layer name="boundRes" select_boundary="true" use_sidechain_neighbors="true" core_cutoff="2.1"
surface_cutoff="1.0"/>
        <Layer name="surfRes" select_surface="true" use_sidechain_neighbors="true" core_cutoff="2.1" surface_cutoff="1.0"/>
    </RESIDUE_SELECTORS>
    <TASKOPERATIONS>
        <LimitAromaChi2 name="limchi2"/>
        <InitializeFromCommandline name="init"/>
        <IncludeCurrent name="includeCurrent"/>
        <RestrictToRepacking name="repack_only" />
        <OperateOnResidueSubset name="fix_rifRes" selector="rifRes" >
            <PreventRepackingRLT/>
        </OperateOnResidueSubset>
    </TASKOPERATIONS>
    <FILTERS>
        <LigInterfaceEnergy name="interfE" scorefxn="beta" energy_cutoff="0.0"/>
        <ShapeComplementarity name="SC" min_sc="0.1" min_interface="0" verbose="0" quick="0" jump="1"/>
        <ScoreType name="totalscore" scorefxn="beta" threshold="0" confidence="1"/>
        <ResidueCount name="nres" confidence="1" />
        # energy per residue
        <CalculatorFilter name="res_totalscore" confidence="1" equation="SCORE/NRES" threshold="0">
            <SCORE name="SCORE" filter_name="totalscore" />
            <NRES name="NRES" filter_name="nres" />
        </CalculatorFilter>
        #buried unsatisfied polar
        <BuriedUnsatHbonds2 name="interf_uhb2" cutoff="200" scorefxn="beta" jump_number="1" />
        #DFHBI specific atom hbond filters
        <HbondsToAtom name="O1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O1" res_num="110"/>
        <HbondsToAtom name="O2_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="O2" res_num="110"/>
        <HbondsToAtom name="N1_hbond" partners="0" energy_cutoff="-0.5" backbone="0" bb_bb="0" sidechain="1"
atomname="N1" res_num="110"/>
```

```
        </FILTERS>
        <MOVERS>
                <MutateResidue name="V103L" target="103A" new_res="LEU"/>
                <MutateResidue name="V83L" target="83A" new_res="LEU"/>
                <MutateResidue name="V83M" target="83A" new_res="MET"/>
                <MutateResidue name="F93W" target="93A" new_res="TRP"/>
                <MutateResidue name="V95A" target="95A" new_res="ALA"/>
                <MutateResidue name="V95G" target="95A" new_res="GLY"/>
                <FastRelax name="relax_holo"
                        task_operations="init,limchi2,includeCurrent"
                        scorefxn="beta"
                        repeats="3"  />
                #calculate a myriad of ligand specific scores
        <InterfaceScoreCalculator name="add_scores" chains="B" scorefxn="beta"/>
         </MOVERS>
        <PROTOCOLS>
                <Add mover_name="V83M"/>
                <Add mover_name="V95G"/>
                <Add mover_name="V103L"/>
                <Add mover_name="relax_holo"/>
                <Add filter_name="SC"/>
                <Add filter_name="interfE"/>
                <Add filter_name="res_totalscore"/>
                <Add filter_name="interf_uhb2"/>
                <Add filter_name="O1_hbond"/>
                <Add filter_name="N1_hbond"/>
                <Add filter_name="O2_hbond"/>
                <Add mover_name="add_scores"/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

# BLUEPRINT FILES

## Blueprint file for loop closure of parametrial beta-barrel backbones: 2_2_2_2_2_1_2.bp

```
1 G .
2 G .
3 G .
4 G .
5 G .
6 G .
7 G .
8 G E
9 G E
0 x L
0 x L
10 G E
11 G E
12 G .
13 G .
14 G .
15 G .
16 G .
17 G E
18 G E
```

0 x L
19 G E
20 G E
21 G .
22 G .
23 G .
24 G .
25 G .
26 G E
27 G E
0 x L
0 x L
28 G E
29 G E
30 G .
31 G .
32 G .
33 G .
34 G .
35 G E
36 G E
0 x L
0 x L
37 G E
38 G E
39 G .
40 G .
41 G .
42 G .
43 G .
44 G E
45 G E
0 x L
46 G E
47 G E
48 G .
49 G .
50 G .
51 G .
52 G .
53 G E
54 G E
0 x L
0 x L
55 G E
56 G E
57 G .
58 G .
59 G .
60 G .
61 G .
62 G E
63 G E
0 x L
0 x L
64 G E
65 G E

66 G .
67 G .
68 G .
69 G .
70 G .
71 G E
72 G E

## Blueprint file for building beta-barrel backbones based on the 2D map: bb_2d.bp

SSPAIR 1-2.A.-4;2-3.A.3
1 V L .
0 V L R
0 V H R
0 V H R
0 V H R
0 V H R
0 V L R
0 V L R
0 G EE R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V LA R
0 V LA R
0 V LG R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 G EE R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EA R
0 V EB R
0 V LA R
0 V LA R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 G EB R
0 V EB R
0 V EB R

0 V LA R
0 V LA R
0 V LG R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 G EE R
0 V EB R
0 G EE R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EA R
0 V EB R
0 V LA R
0 V LA R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V LA R
0 V LA R
0 V LG R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 G EE R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EA R
0 V EB R
0 V LA R
0 V LA R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V LA R
0 V LA R
0 V LG R
0 V EB R
0 V EB R
0 V EB R
0 V EB R

```
0   G   EE   R
0   V   EB   R
0   V   EB   R
0   V   EB   R
0   V   EB   R
0   V   EB   R
0   V   L    R
```

**Blueprint file for building N-terminal disulfide bonds:**
**helix_remodel.bp**

```
1 K L
2 N L
3 A L
4 A H
5 T H
6 A H
7 F .
8 P .
9 G .
10 T .
11 W .
12 D .
13 A .
14 T .
15 F .
16 T .
17 A .
18 E .
19 D .
20 G .
21 S .
22 T .
23 F .
24 Q .
25 G .
26 K .
27 L .
28 D .
29 I .
30 Q .
31 P .
32 T .
33 T .
34 P .
35 D .
36 R .
37 V .
38 T .
39 V .
40 T .
41 V .
42 K .
43 G .
44 T .
45 Q .
46 S .
```

47 D .
48 G .
49 K .
50 P .
51 A .
52 D .
53 G .
54 Q .
55 G .
56 T .
57 L .
58 Q .
59 L .
60 K .
61 T .
62 P .
63 T .
64 T .
65 M .
66 Q .
67 V .
68 T .
69 I .
70 R .
71 Y .
72 S .
73 D .
74 G .
75 K .
76 D .
77 A .
78 T .
79 G .
80 Y .
81 M .
82 T .
83 M .
84 T .
85 T .
86 P .
87 T .
88 T .
89 M .
90 T .
91 A .
92 D .
93 A .
94 Q .
95 L .
96 A .
97 D .
98 G .
99 A .
100 K .
101 S .
102 T .
103 G .

104 Q .
105 F .
106 T .
107 R .
108 K .
109 E .

### Blueprint file for inserting loop 5F:
### loop5F.bp

1 Q .
2 E .
3 V .
4 A .
5 Q .
6 V .
7 L .
8 P .
9 G .
10 D .
11 W .
12 Q .
13 V .
14 H .
15 M .
16 T .
17 N .
18 E .
19 D .
20 G .
21 Q .
22 T .
23 S .
24 T .
25 G .
26 T .
27 V .
28 T .
29 F .
30 Q .
31 P .
32 R .
33 S .
34 P .
35 Y .
36 T .
37 F .
38 D .
39 V .
40 K .
41 F .
42 K .
43 G .
44 T .
45 M .
46 S .
47 D .

48 G .
49 R .
50 P .
51 I .
52 T .
53 G .
54 K .
55 G .
56 K .
57 M .
58 T .
59 M .
60 K .
61 T .
62 P .
63 D .
64 T .
65 M .
66 D .
67 I .
68 D .
69 V .
70 T .
71 Y .
72 S L ALLAA
73 D L NOTAA PG
0 x L PIKAA LIVAYFW
0 x L PIKAA G
74 G L POLAR
75 K L NOTAA PG
76 K L PIKAA K
77 V .
78 T .
79 G .
80 K .
81 V .
82 T .
83 M .
84 K .
85 S .
86 P .
87 T .
88 Q .
89 L .
90 D .
91 W .
92 D .
93 L .
94 T .
95 T .
96 S .
97 D .
98 G .
99 S .
100 K .
101 V .
102 T .

103 G .
104 T .
105 S .
106 H .
107 R .
108 V .
109 E .

# CONSTRAIN FILES:

## Constrain file used for assembling beta barrels based on the 2D map: bb_2d.cst

**# all the hydrogen bond constraints**

***## angle N-H-O constraints were commented out for constructing the set 2 scaffolds for RIF docking***

AtomPair N 17 O 21 HARMONIC 3.0 0.5
Angle N 17 H 17 O 21 CIRCULARHARMONIC 3.1 0.3
Angle H 17 O 21 C 21 CIRCULARHARMONIC 3.1 0.3
AtomPair N 20 O 17 HARMONIC 3.0 0.5
Angle N 20 H 20 O 17 CIRCULARHARMONIC 3.1 0.3
Angle H 20 O 17 C 17 CIRCULARHARMONIC 3.1 0.3
AtomPair N 15 O 23 HARMONIC 3.0 0.5
Angle N 15 H 15 O 23 CIRCULARHARMONIC 3.1 0.3
Angle H 15 O 23 C 23 CIRCULARHARMONIC 3.1 0.3
AtomPair N 23 O 15 HARMONIC 3.0 0.5
Angle N 23 H 23 O 15 CIRCULARHARMONIC 3.1 0.3
Angle H 23 O 15 C 15 CIRCULARHARMONIC 3.1 0.3
AtomPair N 13 O 25 HARMONIC 3.0 0.5
Angle N 13 H 13 O 25 CIRCULARHARMONIC 3.1 0.3
Angle H 13 O 25 C 25 CIRCULARHARMONIC 3.1 0.3
AtomPair N 25 O 13 HARMONIC 3.0 0.5
Angle N 25 H 25 O 13 CIRCULARHARMONIC 3.1 0.3
Angle H 25 O 13 C 13 CIRCULARHARMONIC 3.1 0.3
AtomPair N 11 O 27 HARMONIC 3.0 0.5
Angle N 11 H 11 O 27 CIRCULARHARMONIC 3.1 0.3
Angle H 11 O 27 C 27 CIRCULARHARMONIC 3.1 0.3
AtomPair N 27 O 11 HARMONIC 3.0 0.5
Angle N 27 H 27 O 11 CIRCULARHARMONIC 3.1 0.3
Angle H 27 O 11 C 11 CIRCULARHARMONIC 3.1 0.3
AtomPair N 9 O 29 HARMONIC 3.0 0.5
Angle N 9 H 9 O 29 CIRCULARHARMONIC 3.1 0.3
Angle H 9 O 29 C 29 CIRCULARHARMONIC 3.1 0.3
AtomPair N 29 O 9 HARMONIC 3.0 0.5
Angle N 29 H 29 O 9 CIRCULARHARMONIC 3.1 0.3
Angle H 29 O 9 C 9 CIRCULARHARMONIC 3.1 0.3
AtomPair N 32 O 36 HARMONIC 3.0 0.5
Angle N 32 H 32 O 36 CIRCULARHARMONIC 3.1 0.3
Angle H 32 O 36 C 36 CIRCULARHARMONIC 3.1 0.3
AtomPair N 33 O 36 HARMONIC 3.0 0.5
Angle N 33 H 33 O 36 CIRCULARHARMONIC 3.1 0.3
Angle H 33 O 36 C 36 CIRCULARHARMONIC 3.1 0.3
AtomPair N 36 O 33 HARMONIC 3.0 0.5
Angle N 36 H 36 O 33 CIRCULARHARMONIC 3.1 0.3
Angle H 36 O 33 C 33 CIRCULARHARMONIC 3.1 0.3
AtomPair N 30 O 38 HARMONIC 3.0 0.5
Angle N 30 H 30 O 38 CIRCULARHARMONIC 3.1 0.3
Angle H 30 O 38 C 38 CIRCULARHARMONIC 3.1 0.3

AtomPair N 38 O 30 HARMONIC 3.0 0.5
Angle N 38 H 38 O 30 CIRCULARHARMONIC 3.1 0.3
Angle H 38 O 30 C 30 CIRCULARHARMONIC 3.1 0.3
AtomPair N 28 O 40 HARMONIC 3.0 0.5
Angle N 28 H 28 O 40 CIRCULARHARMONIC 3.1 0.3
Angle H 28 O 40 C 40 CIRCULARHARMONIC 3.1 0.3
AtomPair N 40 O 28 HARMONIC 3.0 0.5
Angle N 40 H 40 O 28 CIRCULARHARMONIC 3.1 0.3
Angle H 40 O 28 C 28 CIRCULARHARMONIC 3.1 0.3
AtomPair N 26 O 42 HARMONIC 3.0 0.5
Angle N 26 H 26 O 42 CIRCULARHARMONIC 3.1 0.3
Angle H 26 O 42 C 42 CIRCULARHARMONIC 3.1 0.3
AtomPair N 42 O 26 HARMONIC 3.0 0.5
Angle N 42 H 42 O 26 CIRCULARHARMONIC 3.1 0.3
Angle H 42 O 26 C 26 CIRCULARHARMONIC 3.1 0.3
AtomPair N 24 O 44 HARMONIC 3.0 0.5
Angle N 24 H 24 O 44 CIRCULARHARMONIC 3.1 0.3
Angle H 24 O 44 C 44 CIRCULARHARMONIC 3.1 0.3
AtomPair N 44 O 24 HARMONIC 3.0 0.5
Angle N 44 H 44 O 24 CIRCULARHARMONIC 3.1 0.3
Angle H 44 O 24 C 24 CIRCULARHARMONIC 3.1 0.3
AtomPair N 46 O 22 HARMONIC 3.0 0.5
Angle N 46 H 46 O 22 CIRCULARHARMONIC 3.1 0.3
Angle H 46 O 22 C 22 CIRCULARHARMONIC 3.1 0.3
AtomPair N 45 O 49 HARMONIC 3.0 0.5
Angle N 45 H 45 O 49 CIRCULARHARMONIC 3.1 0.3
Angle H 45 O 49 C 49 CIRCULARHARMONIC 3.1 0.3
AtomPair N 49 O 45 HARMONIC 3.0 0.5
Angle N 49 H 49 O 45 CIRCULARHARMONIC 3.1 0.3
Angle H 49 O 45 C 45 CIRCULARHARMONIC 3.1 0.3
AtomPair N 43 O 51 HARMONIC 3.0 0.5
Angle N 43 H 43 O 51 CIRCULARHARMONIC 3.1 0.3
Angle H 43 O 51 C 51 CIRCULARHARMONIC 3.1 0.3
AtomPair N 51 O 43 HARMONIC 3.0 0.5
Angle N 51 H 51 O 43 CIRCULARHARMONIC 3.1 0.3
Angle H 51 O 43 C 43 CIRCULARHARMONIC 3.1 0.3
AtomPair N 41 O 53 HARMONIC 3.0 0.5
Angle N 41 H 41 O 53 CIRCULARHARMONIC 3.1 0.3
Angle H 41 O 53 C 53 CIRCULARHARMONIC 3.1 0.3
AtomPair N 53 O 41 HARMONIC 3.0 0.5
Angle N 53 H 53 O 41 CIRCULARHARMONIC 3.1 0.3
Angle H 53 O 41 C 41 CIRCULARHARMONIC 3.1 0.3
AtomPair N 39 O 55 HARMONIC 3.0 0.5
Angle N 39 H 39 O 55 CIRCULARHARMONIC 3.1 0.3
Angle H 39 O 55 C 55 CIRCULARHARMONIC 3.1 0.3
AtomPair N 55 O 39 HARMONIC 3.0 0.5
Angle N 55 H 55 O 39 CIRCULARHARMONIC 3.1 0.3
Angle H 55 O 39 C 39 CIRCULARHARMONIC 3.1 0.3
AtomPair N 37 O 57 HARMONIC 3.0 0.5
Angle N 37 H 37 O 57 CIRCULARHARMONIC 3.1 0.3
Angle H 37 O 57 C 57 CIRCULARHARMONIC 3.1 0.3
AtomPair N 59 O 35 HARMONIC 3.0 0.5
Angle N 59 H 59 O 35 CIRCULARHARMONIC 3.1 0.3
Angle H 59 O 35 C 35 CIRCULARHARMONIC 3.1 0.3
AtomPair N 57 O 37 HARMONIC 3.0 0.5
Angle N 57 H 57 O 37 CIRCULARHARMONIC 3.1 0.3
Angle H 57 O 37 C 37 CIRCULARHARMONIC 3.1 0.3

AtomPair N 60 O 64 HARMONIC 3.0 0.5
Angle N 60 H 60 O 64 CIRCULARHARMONIC 3.1 0.3
Angle H 60 O 64 C 64 CIRCULARHARMONIC 3.1 0.3
AtomPair N 61 O 64 HARMONIC 3.0 0.5
Angle N 61 H 61 O 64 CIRCULARHARMONIC 3.1 0.3
Angle H 61 O 64 C 64 CIRCULARHARMONIC 3.1 0.3
AtomPair N 64 O 61 HARMONIC 3.0 0.5
Angle N 64 H 64 O 61 CIRCULARHARMONIC 3.1 0.3
Angle H 64 O 61 C 61 CIRCULARHARMONIC 3.1 0.3
AtomPair N 58 O 66 HARMONIC 3.0 0.5
Angle N 58 H 58 O 66 CIRCULARHARMONIC 3.1 0.3
Angle H 58 O 66 C 66 CIRCULARHARMONIC 3.1 0.3
AtomPair N 66 O 58 HARMONIC 3.0 0.5
Angle N 66 H 66 O 58 CIRCULARHARMONIC 3.1 0.3
Angle H 66 O 58 C 58 CIRCULARHARMONIC 3.1 0.3
AtomPair N 56 O 68 HARMONIC 3.0 0.5
Angle N 56 H 56 O 68 CIRCULARHARMONIC 3.1 0.3
Angle H 56 O 68 C 68 CIRCULARHARMONIC 3.1 0.3
AtomPair N 68 O 56 HARMONIC 3.0 0.5
Angle N 68 H 68 O 56 CIRCULARHARMONIC 3.1 0.3
Angle H 68 O 56 C 56 CIRCULARHARMONIC 3.1 0.3
AtomPair N 54 O 70 HARMONIC 3.0 0.5
Angle N 54 H 54 O 70 CIRCULARHARMONIC 3.1 0.3
Angle H 54 O 70 C 70 CIRCULARHARMONIC 3.1 0.3
AtomPair N 70 O 54 HARMONIC 3.0 0.5
Angle N 70 H 70 O 54 CIRCULARHARMONIC 3.1 0.3
Angle H 70 O 54 C 54 CIRCULARHARMONIC 3.1 0.3
AtomPair N 72 O 52 HARMONIC 3.0 0.5
Angle N 72 H 72 O 52 CIRCULARHARMONIC 3.1 0.3
Angle H 72 O 52 C 52 CIRCULARHARMONIC 3.1 0.3
AtomPair N 71 O 75 HARMONIC 3.0 0.5
Angle N 71 H 71 O 75 CIRCULARHARMONIC 3.1 0.3
Angle H 71 O 75 C 75 CIRCULARHARMONIC 3.1 0.3
AtomPair N 74 O 71 HARMONIC 3.0 0.5
Angle N 74 H 74 O 71 CIRCULARHARMONIC 3.1 0.3
Angle H 74 O 71 C 71 CIRCULARHARMONIC 3.1 0.3
AtomPair N 69 O 77 HARMONIC 3.0 0.5
Angle N 69 H 69 O 77 CIRCULARHARMONIC 3.1 0.3
Angle H 69 O 77 C 77 CIRCULARHARMONIC 3.1 0.3
AtomPair N 77 O 69 HARMONIC 3.0 0.5
Angle N 77 H 77 O 69 CIRCULARHARMONIC 3.1 0.3
Angle H 77 O 69 C 69 CIRCULARHARMONIC 3.1 0.3
AtomPair N 67 O 79 HARMONIC 3.0 0.5
Angle N 67 H 67 O 79 CIRCULARHARMONIC 3.1 0.3
Angle H 67 O 79 C 79 CIRCULARHARMONIC 3.1 0.3
AtomPair N 79 O 67 HARMONIC 3.0 0.5
Angle N 79 H 79 O 67 CIRCULARHARMONIC 3.1 0.3
Angle H 79 O 67 C 67 CIRCULARHARMONIC 3.1 0.3
AtomPair N 65 O 81 HARMONIC 3.0 0.5
Angle N 65 H 65 O 81 CIRCULARHARMONIC 3.1 0.3
Angle H 65 O 81 C 81 CIRCULARHARMONIC 3.1 0.3
AtomPair N 81 O 65 HARMONIC 3.0 0.5
Angle N 81 H 81 O 65 CIRCULARHARMONIC 3.1 0.3
Angle H 81 O 65 C 65 CIRCULARHARMONIC 3.1 0.3
AtomPair N 83 O 63 HARMONIC 3.0 0.5
Angle N 83 H 83 O 63 CIRCULARHARMONIC 3.1 0.3
Angle H 83 O 63 C 63 CIRCULARHARMONIC 3.1 0.3

AtomPair N 85 O 88 HARMONIC 3.0 0.5
Angle N 85 H 85 O 88 CIRCULARHARMONIC 3.1 0.3
Angle H 85 O 88 C 88 CIRCULARHARMONIC 3.1 0.3
AtomPair N 88 O 85 HARMONIC 3.0 0.5
Angle N 88 H 88 O 85 CIRCULARHARMONIC 3.1 0.3
Angle H 88 O 85 C 85 CIRCULARHARMONIC 3.1 0.3
AtomPair N 84 O 88 HARMONIC 3.0 0.5
Angle N 84 H 84 O 88 CIRCULARHARMONIC 3.1 0.3
Angle H 84 O 88 C 88 CIRCULARHARMONIC 3.1 0.3
AtomPair N 82 O 90 HARMONIC 3.0 0.5
Angle N 82 H 82 O 90 CIRCULARHARMONIC 3.1 0.3
Angle H 82 O 90 C 90 CIRCULARHARMONIC 3.1 0.3
AtomPair N 90 O 82 HARMONIC 3.0 0.5
Angle N 90 H 90 O 82 CIRCULARHARMONIC 3.1 0.3
Angle H 90 O 82 C 82 CIRCULARHARMONIC 3.1 0.3
AtomPair N 80 O 92 HARMONIC 3.0 0.5
Angle N 80 H 80 O 92 CIRCULARHARMONIC 3.1 0.3
Angle H 80 O 92 C 92 CIRCULARHARMONIC 3.1 0.3
AtomPair N 92 O 80 HARMONIC 3.0 0.5
Angle N 92 H 92 O 80 CIRCULARHARMONIC 3.1 0.3
Angle H 92 O 80 C 80 CIRCULARHARMONIC 3.1 0.3
AtomPair N 78 O 94 HARMONIC 3.0 0.5
Angle N 78 H 78 O 94 CIRCULARHARMONIC 3.1 0.3
Angle H 78 O 94 C 94 CIRCULARHARMONIC 3.1 0.3
AtomPair N 94 O 78 HARMONIC 3.0 0.5
Angle N 94 H 94 O 78 CIRCULARHARMONIC 3.1 0.3
Angle H 94 O 78 C 78 CIRCULARHARMONIC 3.1 0.3
AtomPair N 96 O 76 HARMONIC 3.0 0.5
Angle N 96 H 96 O 76 CIRCULARHARMONIC 3.1 0.3
Angle H 96 O 76 C 76 CIRCULARHARMONIC 3.1 0.3
AtomPair N 95 O 99 HARMONIC 3.0 0.5
Angle N 95 H 95 O 99 CIRCULARHARMONIC 3.1 0.3
Angle H 95 O 99 C 99 CIRCULARHARMONIC 3.1 0.3
AtomPair N 98 O 95 HARMONIC 3.0 0.5
Angle N 98 H 98 O 95 CIRCULARHARMONIC 3.1 0.3
Angle H 98 O 95 C 95 CIRCULARHARMONIC 3.1 0.3
AtomPair N 93 O 101 HARMONIC 3.0 0.5
Angle N 93 H 93 O 101 CIRCULARHARMONIC 3.1 0.3
Angle H 93 O 101 C 101 CIRCULARHARMONIC 3.1 0.3
AtomPair N 101 O 93 HARMONIC 3.0 0.5
Angle N 101 H 101 O 93 CIRCULARHARMONIC 3.1 0.3
Angle H 101 O 93 C 93 CIRCULARHARMONIC 3.1 0.3
AtomPair N 91 O 103 HARMONIC 3.0 0.5
Angle N 91 H 91 O 103 CIRCULARHARMONIC 3.1 0.3
Angle H 91 O 103 C 103 CIRCULARHARMONIC 3.1 0.3
AtomPair N 103 O 91 HARMONIC 3.0 0.5
Angle N 103 H 103 O 91 CIRCULARHARMONIC 3.1 0.3
Angle H 103 O 91 C 91 CIRCULARHARMONIC 3.1 0.3
AtomPair N 89 O 105 HARMONIC 3.0 0.5
Angle N 89 H 89 O 105 CIRCULARHARMONIC 3.1 0.3
Angle H 89 O 105 C 105 CIRCULARHARMONIC 3.1 0.3
AtomPair N 105 O 89 HARMONIC 3.0 0.5
Angle N 105 H 105 O 89 CIRCULARHARMONIC 3.1 0.3
Angle H 105 O 89 C 89 CIRCULARHARMONIC 3.1 0.3
AtomPair N 107 O 87 HARMONIC 3.0 0.5
Angle N 107 H 107 O 87 CIRCULARHARMONIC 3.1 0.3
Angle H 107 O 87 C 87 CIRCULARHARMONIC 3.1 0.3

AtomPair N 16 O 102 HARMONIC 3.0 0.5
Angle N 16 H 16 O 102 CIRCULARHARMONIC 3.1 0.3
Angle H 16 O 102 C 102 CIRCULARHARMONIC 3.1 0.3
AtomPair N 18 O 100 HARMONIC 3.0 0.5
Angle N 18 H 18 O 100 CIRCULARHARMONIC 3.1 0.3
Angle H 18 O 100 C 100 CIRCULARHARMONIC 3.1 0.3
AtomPair N 102 O 16 HARMONIC 3.0 0.5
Angle N 102 H 102 O 16 CIRCULARHARMONIC 3.1 0.3
Angle H 102 O 16 C 16 CIRCULARHARMONIC 3.1 0.3
AtomPair N 14 O 104 HARMONIC 3.0 0.5
Angle N 14 H 14 O 104 CIRCULARHARMONIC 3.1 0.3
Angle H 14 O 104 C 104 CIRCULARHARMONIC 3.1 0.3
AtomPair N 104 O 14 HARMONIC 3.0 0.5
Angle N 104 H 104 O 14 CIRCULARHARMONIC 3.1 0.3
Angle H 104 O 14 C 14 CIRCULARHARMONIC 3.1 0.3
AtomPair N 12 O 106 HARMONIC 3.0 0.5
Angle N 12 H 12 O 106 CIRCULARHARMONIC 3.1 0.3
Angle H 12 O 106 C 106 CIRCULARHARMONIC 3.1 0.3
AtomPair N 106 O 12 HARMONIC 3.0 0.5
Angle N 106 H 106 O 12 CIRCULARHARMONIC 3.1 0.3
Angle H 106 O 12 C 12 CIRCULARHARMONIC 3.1 0.3
AtomPair N 10 O 108 HARMONIC 3.0 0.5
Angle N 10 H 10 O 108 CIRCULARHARMONIC 3.1 0.3
Angle H 10 O 108 C 108 CIRCULARHARMONIC 3.1 0.3
AtomPair N 108 O 10 HARMONIC 3.0 0.5
Angle N 108 H 108 O 10 CIRCULARHARMONIC 3.1 0.3
Angle H 108 O 10 C 10 CIRCULARHARMONIC 3.1 0.3
**# The following constraints were used to define the backbone geometry of the tryptophan corner and capping N-terminal helix.**
**## Torsion angle constraints for residues Trp-4 and trp-3 (supplementary figure)**
Dihedral N 8 CA 8 C 8 N 9 CIRCULARHARMONIC 2.35 0.25
Dihedral C 7 N 8 CA 8 C 8 CIRCULARHARMONIC 5.20 0.25
Dihedral N 7 CA 7 C 7 N 8 CIRCULARHARMONIC 5.75 0.25
Dihedral C 6 N 7 CA 7 C 7 CIRCULARHARMONIC 4.90 0.25
**## Distance constraint mimicking the tryptophan to backbone carbonyl hydrogen bond**
AtomPair CA 11 O 8 BOUNDED 6.9 7.5 0.5
**## Distance constraints to place the residue in the center of the barrel to get good hydrophobic packing around it**
AtomPair CA 7 CA 62 BOUNDED 9.5 10.5 0.5
AtomPair CA 7 CA 86 BOUNDED 7.0 8.0 0.5
AtomPair CA 6 CA 34 HARMONIC 12.0 0.5
AtomPair CA 6 CA 62 HARMONIC 9.5 0.5
AtomPair CA 6 CA 86 HARMONIC 9.5 0.5
**## Distance constraints mimicking the Arginine to backbone hydrogen bonds**
AtomPair CA 7 CA 107 BOUNDED 8.5 10.0 0.5
AtomPair O 6 CA 107 HARMONIC 8.5 0.5

## Constrain file used for designing beta barrels based on the 2D map:

## cst_trp.cst

AtomPair NE1 11 O 7 HARMONIC 3.0 0.5

## Constrain file used for building the new loop for b11L5F:

**L5F.cst**

AtomPair O 77 N 71 BOUNDED 2.5 3.5 0.5 TAG
AtomPair O 71 N 74 BOUNDED 2.5 3.5 0.5 TAG
AtomPair O 71 N 76 BOUNDED 2.5 3.5 0.5 TAG
AtomPair O 74 N 77 BOUNDED 2.5 3.5 0.5 TAG

RESFILE FILES:

**Resfile used for designing nonfunctional beta barrels from parametric models:**

**bb_param.res**

```
ALLAA
start
8 A PIKAA ITVYWF
9 A PIKAA ITVYWF
10 A PIKAA GNDS
11 A PIKAA GNDS
12 A PIKAA ITVYWF
13 A PIKAA ITVYWF
16 A PIKAA ITVYWF
17 A PIKAA ITVYWF
18 A PIKAA G
19 A PIKAA NDS
20 A PIKAA ITVYWF
21 A PIKAA ITVYWF
44 A PIKAA ITVYWF
45 A PIKAA ITVYWF
46 A PIKAA G
47 A PIKAA NDS
48 A PIKAA ITVYWF
49 A PIKAA ITVYWF
52 A PIKAA ITVYWF
53 A PIKAA ITVYWF
54 A PIKAA GNDS
55 A PIKAA GNDS
56 A PIKAA ITVYWF
57 A PIKAA ITVYWF
61 A PIKAA ITVYWF
62 A PIKAA ITVYWF
63 A PIKAA GNDS
64 A PIKAA GNDS
65 A PIKAA GNDS
66 A PIKAA GNDS
67 A PIKAA ITVYWF
68 A PIKAA ITVYWF
72 A PIKAA ITVYWF
73 A PIKAA ITVYWF
74 A PIKAA G
75 A PIKAA NDS
76 A PIKAA ITVYWF
77 A PIKAA ITVYWF
```

**Resfile used for designing nonfunctional beta barrels from 2D-map assembly:**

**bb_2d.res**

```
ALLAA
start
8 A PIKAA P
9 A PIKAA G
11 A PIKAA W
17 A PIKAA ELSRT
18 A PIKAA AEKS
19 A PIKAA D
20 A PIKAA G
```

21 A PIKAA AKRS
25 A PIKAA G
29 A PIKAA ILMV
31 A PIKAA P
32 A POLAR
33 A PIKAA ST
34 A PIKAA P
35 A PIKAA DEHTY
43 A PIKAA G
45 A PIKAA ELRST
46 A PIKAA AEKS
47 A PIKAA D
48 A PIKAA G
49 A PIKAA AKRS
50 A PIKAA P
53 A PIKAA G
55 A PIKAA G
59 A PIKAA ILMV
60 A POLAR
61 A PIKAA ST
62 A PIKAA P
63 A PIKAA DEHTY
71 A PIKAA ELRST
72 A PIKAA AEKS
73 A PIKAA D
74 A PIKAA G
75 A PIKAA AKRS
79 A PIKAA G
83 A PIKAA ILMV
84 A POLAR
85 A PIKAA ST
86 A PIKAA P
87 A PIKAA DEHTY
97 A PIKAA D
98 A PIKAA G
103 A PIKAA G
107 A PIKAA R

## Resfile used for designing DFHBI-binding beta barrels:

## rectBarrel.resfile

ALLAA
start
8 A PIKAA P
9 A PIKAA G
11 A PIKAA W
18 A PIKAA AEKS
19 A PIKAA D
20 A PIKAA G
25 A PIKAA G
29 A PIKAA ILMV
31 A PIKAA P
32 A POLAR
33 A PIKAA ST
34 A PIKAA P
35 A PIKAA DEHTY
43 A PIKAA G

46 A PIKAA AEKS
47 A PIKAA D
48 A PIKAA G
50 A PIKAA P
53 A PIKAA G
55 A PIKAA G
59 A PIKAA ILMV
60 A POLAR
61 A PIKAA ST
62 A PIKAA P
63 A PIKAA DEHTY
72 A PIKAA AEKS
73 A PIKAA D
74 A PIKAA G
75 A PIKAA AKRS
79 A PIKAA G
83 A PIKAA ILMV
84 A POLAR
85 A PIKAA ST
86 A PIKAA P
87 A PIKAA DEHTY
97 A PIKAA D
98 A PIKAA G
103 A PIKAA G
107 A PIKAA R

**Resfile used for profile-based sequence design:**

**final.resfile** ***(final.resfile is different for each sequence cluster. This example is for designs based on 14_input_0065 input scaffold that yielded successful binders b11 and b32)***

ALLAA
start
1 A PIKAA QA
2 A PIKAA QEYK
3 A PIKAA V
4 A PIKAA AVF
5 A PIKAA Q
6 A PIKAA V
7 A PIKAA LMVIF
8 A PIKAA P
9 A PIKAA G
10 A PIKAA KRNTD
11 A PIKAA W
12 A PIKAA KDNQ
13 A PIKAA IV
14 A PIKAA TRHN
15 A PIKAA MF
16 A PIKAA TK
17 A PIKAA N
18 A PIKAA ES
19 A PIKAA D
20 A PIKAA G
21 A PIKAA TQVL
22 A PIKAA T
23 A PIKAA S
24 A PIKAA QT
25 A PIKAA G

26 A PIKAA TQH
27 A PIKAA MIFLV
28 A PIKAA TNHR
29 A PIKAA VIMF
30 A PIKAA Q
31 A PIKAA P
32 A PIKAA KR
33 A PIKAA S
34 A PIKAA P
35 A PIKAA Y
36 A PIKAA T
37 A PIKAA VLF
38 A PIKAA D
39 A PIKAA VIFL
40 A PIKAA QTKR
41 A PIKAA WAF
42 A PIKAA QTKR
43 A PIKAA G
44 A PIKAA T
45 A PIKAA LIM
46 A PIKAA S
47 A PIKAA D
48 A PIKAA G
49 A PIKAA R
50 A PIKAA P
51 A PIKAA I
52 A PIKAA QTKR
53 A PIKAA G
54 A PIKAA KTQN
55 A PIKAA G
56 A PIKAA QK
57 A PIKAA LVMF
58 A PIKAA T
59 A PIKAA M
60 A PIKAA RKHD
61 A PIKAA T
62 A PIKAA P
63 A PIKAA DTH
64 A PIKAA T
65 A PIKAA ML
66 A PIKAA QDT
67 A PIKAA VFLI
68 A PIKAA D
69 A PIKAA VIFL
70 A PIKAA TK
71 A PIKAA Y
72 A PIKAA S
73 A PIKAA D
74 A PIKAA G
75 A PIKAA K
76 A PIKAA KQ
77 A PIKAA VIMF
78 A PIKAA TKQ
79 A PIKAA G
80 A PIKAA QKH
81 A PIKAA VMF
82 A PIKAA T

```
83 A PIKAA LM
84 A PIKAA RKDHE
85 A PIKAA S
86 A PIKAA P
87 A PIKAA TE
88 A PIKAA KQ
89 A PIKAA LFI
90 A PIKAA TDQRK
91 A PIKAA IFLW
92 A PIKAA D
93 A PIKAA LFVI
94 A PIKAA T
95 A PIKAA T
96 A PIKAA IAS
97 A PIKAA D
98 A PIKAA G
99 A PIKAA LVTS
100 A PIKAA KQ
101 A PIKAA V
102 A PIKAA T
103 A PIKAA G
104 A PIKAA HT
105 A PIKAA LFVT  # backup Trp91
106 A PIKAA TRQHK
107 A PIKAA R
108 A PIKAA VLI
109 A PIKAA EK
```

### Resfile used to design sequences for loop5F:

### loop5F.resfile

```
NATAA
 start
70 A POLAR
71 A PIKAA Y            #(phi, psi)=(-117, 95) allowed for pre-pro
72 A PIKAA P
73 A PIKAA STAVIL                # new interacting residue but can be structural cross-strand interaction
74 A PIKAA LIVAYFWM          # new interacting residue
75 A PIKAA G
76 A POLAR
77 A PIKAA ASTCMLI
78 A POLAR
 # re-configure boundary interactions
52 A POLAR                                  # for potential cross-strand interaction
21 A NOTAA PGKR
79 A PIKAA VTSA                     # TS for hbonding 97T, VA for reconfigured pocket
 51 A PIKAA ILY
45 A PIKAA VLIFMST
 19 A PIKAA DNERKHST
47 A PIKAA DNERKHST
99 A PIKAA DNERKHST
```

## OTHER FILES USED IN THE DESIGN CALCULATION

### Parameters used to build disconnected beta-barrel backbones based on the hyperboloid model:

```
N 8 #number of strands
S 10 #shear number of the barrel
```

a 3.3 #default distance between Ca along a beta-strand (parameter d)
b 4.20-4.60:0.05 #distance between beta-strands in A and sampled around the ideal value derived from the PDB (parameter D)
nres 9 #number of residues per strand
topology 1,2,3,4,5,6,7,8 #topology - up-and-down beta-barrel
dr 0.9-1.1:0.05 #ratio applied on each of the elliptical radii, sampled around the ideal radius of a barrel of type (n=n; S=S)
dtw 0.9-1.1:0.05 #ratio applied on the staggering angle between the beta-strands and the Z axis, sampled around the ideal value for a barrel of type (N; S).

## Improved Rosetta energy function weights used to assemble beta-barrel backbones with near-native torsion angle distributions:

vdw 1.0
rg 1.0
rama 0.15
ss_pair 1.0
rsigma 1.0
omega 0.5
hbond_sr_bb 1.0
hbond_lr_bb 1.0
STRAND_STRAND_WEIGHTS 1 11

## Criteria used for selecting assembled backbones for sequence designs:

'vdw' < 1
'omega' < 14
'hbond_lr_bb' < -58
'rama' < 0

## Rosetta full-atom parameter file(.param) for DFHBI:
## (3-letter code for DFHBI is HBI; F atoms were replace by H for RIF docking calculation since its database was not supporting F atoms at the time)
## HBI_rch.param (HBI_fa.param)

NAME HBI
IO_STRING HBI Z
TYPE LIGAND
AA UNK

| | | | | |
|---|---|---|---|---|
| ATOM | C4 | aroC | X | 0.12 |
| ATOM | C2 | aroC | X | -0.16 |
| ATOM | N1 | Nhis | X | -0.54 |
| ATOM | C11 | aroC | X | 0.34 |
| ATOM | C1 | CH3 | X | -0.10 |
| ATOM | H2 | Hapo | X | 0.04 |
| ATOM | H3 | Hapo | X | 0.04 |
| ATOM | H1 | Hapo | X | 0.04 |
| ATOM | N2 | Npro | X | -0.44 |
| ATOM | C3 | CH3 | X | 0.11 |
| ATOM | H5 | Hapo | X | 0.03 |
| ATOM | H6 | Hapo | X | 0.03 |
| ATOM | H4 | Hapo | X | 0.03 |
| ATOM | C12 | CNH2 | X | 0.68 |
| ATOM | O2 | ONH2 | X | -0.68 |
| ATOM | C5 | aroC | X | -0.25 |
| ATOM | C6 | aroC | X | -0.05 |
| ATOM | C8 | aroC | X | -0.08 |
| ATOM | F1 | F | X | -0.17 |
| ATOM | C10 | aroC | X | 0.56 |

```
ATOM     O1         OOC     X       -0.67
ATOM     C9         aroC    X       -0.08
ATOM     C7         aroC    X       -0.05
ATOM     H7         Haro    X       0.13
ATOM     F2         F       X       -0.17
ATOM     H8         Haro    X       0.13
ATOM     H9         Haro    X       0.12
BOND_TYPE  C1   H2  1
BOND_TYPE  C1   H3  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C1   C11 1
BOND_TYPE  N1   C11 2
BOND_TYPE  N1   C2  1
BOND_TYPE  O1   C10 1
BOND_TYPE  C2   C4  2
BOND_TYPE  C2   C12 1
BOND_TYPE  N2   C11 1
BOND_TYPE  N2   C3  1
BOND_TYPE  N2   C12 4
BOND_TYPE  O2   C12 2
BOND_TYPE  C3   H5  1
BOND_TYPE  C3   H6  1
BOND_TYPE  C3   H4  1
BOND_TYPE  C4   C5  1
BOND_TYPE  C4   H9  1
BOND_TYPE  C5   C6  4
BOND_TYPE  C5   C7  4
BOND_TYPE  C6   C8  4
BOND_TYPE  C6   H8  1
BOND_TYPE  C7   C9  4
BOND_TYPE  C7   H7  1
BOND_TYPE  C8   F1  1
BOND_TYPE  C8   C10 4
BOND_TYPE  C9   C10 4
BOND_TYPE  C9   F2  1
#CHI 1  C2   C4   C5   C6
NBR_ATOM  C4
NBR_RADIUS 5.662129
ICOOR_INTERNAL      C4      0.000000        0.000000        0.000000  C4   C2   N1
ICOOR_INTERNAL      C2      0.000000        180.000000      1.349077  C4   C2   N1
ICOOR_INTERNAL      N1      -0.000001       52.329407       1.416959  C2   C4   N1
ICOOR_INTERNAL      C11     -179.910440     71.991456       1.292976  N1   C2   C4
ICOOR_INTERNAL      C1      -177.859132     57.151621       1.521362  C11  N1   C2
ICOOR_INTERNAL      H2       -0.565778      56.187873       1.209217  C1   C11  N1
ICOOR_INTERNAL      H3      -116.360238     67.723486       1.140332        C1   C11  H2
ICOOR_INTERNAL      H1      -125.915669     65.309896       1.136701  C1   C11  H3
ICOOR_INTERNAL      N2      177.947519      68.599173       1.419331  C11  N1   C1
ICOOR_INTERNAL      C3      179.906282      54.510752       1.445059  N2   C11  N1
ICOOR_INTERNAL      H5      -59.896679      70.508523       1.110663  C3   N2   C11
ICOOR_INTERNAL      H6      119.707222      70.408523       1.109885  C3   N2   H5
ICOOR_INTERNAL      H4      120.206443      68.909056       1.110946  C3   N2   H6
ICOOR_INTERNAL      C12     179.997516      72.740427       1.340786  N2   C11  C3
ICOOR_INTERNAL      O2      -179.974742     54.971911       1.221955  C12  N2   C11
ICOOR_INTERNAL      C5      -0.010405       52.749806       1.497857  C4   C2   N1
ICOOR_INTERNAL      C6      -0.385418       55.997554       1.413715  C5   C4   C2
ICOOR_INTERNAL      C8      179.949968      59.554822       1.398770  C6   C5   C4
ICOOR_INTERNAL      F1      -179.966601     60.244299       1.356498  C8   C6   C5
```

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ICOOR_INTERNAL | C10 | 179.977668 | 59.717550 | 1.397908 | C8 | C6 | F1 | |
| ICOOR_INTERNAL | O1 | 179.994297 | 59.945959 | 1.344216 | C10 | C8 | C6 | |
| ICOOR_INTERNAL | C9 | -179.932109 | 60.138841 | 1.396190 | C10 | C8 | O1 | |
| ICOOR_INTERNAL | C7 | -0.078832 | 59.911335 | 1.397451 | | C9 | C10 | C8 |
| ICOOR_INTERNAL | H7 | -179.951647 | 61.144133 | 1.083859 | C7 | C9 | C10 | |
| ICOOR_INTERNAL | F2 | -179.906656 | 59.913978 | 1.356367 | C9 | C10 | C7 | |
| ICOOR_INTERNAL | H8 | -179.989428 | 57.828087 | 1.072240 | C6 | C5 | C8 | |
| ICOOR_INTERNAL | H9 | 179.867077 | 63.647869 | 1.087977 | C4 | C2 | C5 | |

## Rosetta centroid parameter file(.param) for DFHBI:

## HBI.cen.param

```
AME HBI
IO_STRING HBI Z
TYPE LIGAND
AA UNK
ATOM  C4  CAbb  X   0.04
ATOM  C2  CAbb  X   0.03
ATOM  N1  OCbb  X   -0.79
ATOM  C11 CAbb  X   0.70
ATOM  C1  CAbb  X   -0.53
ATOM  N2  Nbb   X   -1.10
ATOM  C3  CAbb  X   -0.17
ATOM  C12 CAbb  X   1.08
ATOM  O2  OCbb  X   -0.83
ATOM  C5  CAbb  X   -0.26
ATOM  C6  CAbb  X   -0.15
ATOM  C8  CAbb  X   0.35
ATOM  F1  CAbb  X   -0.46
ATOM  C10 CAbb  X   0.38
ATOM  O1  OCbb  X   -0.83
ATOM  C9  CAbb  X   0.35
ATOM  C7  CAbb  X   -0.16
ATOM  F2  CAbb  X   -0.46
BOND_TYPE  C1   C11 1
BOND_TYPE  N1   C11 2
BOND_TYPE  N1   C2  1
BOND_TYPE  O1   C10 1
BOND_TYPE  C2   C4  2
BOND_TYPE  C2   C12 1
BOND_TYPE  N2   C11 1
BOND_TYPE  N2   C3  1
BOND_TYPE  N2   C12 4
BOND_TYPE  O2   C12 2
BOND_TYPE  C4   C5  1
BOND_TYPE  C5   C6  4
BOND_TYPE  C5   C7  4
BOND_TYPE  C6   C8  4
BOND_TYPE  C7   C9  4
BOND_TYPE  C8   F1  1
BOND_TYPE  C8   C10 4
BOND_TYPE  C9   C10 4
BOND_TYPE  C9   F2  1
#CHI 1  C2   C4   C5   C6
NBR_ATOM  C4
NBR_RADIUS 5.662663
ICOOR_INTERNAL  C4     0.000000         0.000000   0.000000   C4    C2     N1
```

```
ICOOR_INTERNAL C2     0.000000  180.000000       1.349492  C4         C2         N1
ICOOR_INTERNAL N1     0.000000   52.312127       1.416411  C2         C4         N1
ICOOR_INTERNAL C11 -179.876986   71.981668    1.292864  N1         C2    C4
ICOOR_INTERNAL C1  -177.823063   57.141926       1.521803  C11   N1   C2
ICOOR_INTERNAL N2   177.824660   68.608152       1.418583  C11   N1   C1
ICOOR_INTERNAL C3   179.909099   54.484510       1.445087  N2         C11   N1
ICOOR_INTERNAL C12 -179.933962   72.728853    1.340776  N2         C11   C3
ICOOR_INTERNAL O2   179.964406   54.938317       1.222126  C12   N2   C11
ICOOR_INTERNAL C5     0.003093   52.791761       1.497704  C4         C2         N1
ICOOR_INTERNAL C6    -0.352261   55.946592       1.413198  C5         C4         C2
ICOOR_INTERNAL C8   179.949796   59.531212       1.399355  C6         C5         C4
ICOOR_INTERNAL F1  -179.990868   60.282211       1.356833  C8         C6         C5
ICOOR_INTERNAL C10  179.985408   59.719418       1.396930  C8         C6         F1
ICOOR_INTERNAL O1  -179.980291   59.936638       1.344611  C10   C8   C6
ICOOR_INTERNAL C9  -179.990699   60.126809       1.396219  C10   C8   O1
ICOOR_INTERNAL C7    -0.021432   59.905469       1.397765  C9         C10   C8
ICOOR_INTERNAL    F2  -179.975061   59.878212    1.356616  C9         C10   C7
```

**Amino acids composition file used for biasing aromatic residues during sequence design for DFHBI-binding beta barrels**

**favour_core_aromatics.comp**

```
PENALTY_DEFINITION
# Define residue types to control
PROPERTIES AROMATIC
NOT_PROPERTIES POLAR CHARGED
# Declare desired quantity of these residues
FRACTION 0.15
# Set the penalty for having too few, at the desired number, and too many of the specified residues
PENALTIES 100 0 100
FRACT_DELTA_START -0.05
#DELTA_START -1
FRACT_DELTA_END 0.1
#DELTA_END 2
#set how the penalties are applied
BEFORE_FUNCTION CONSTANT
AFTER_FUNCTION CONSTANT
END_PENALTY_DEFINITION
```


57. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 (2014).

58. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491,** 222–227 (2012).

59. Craveur, P., Joseph, A. P., Rebehmed, J. & de Brevern, A. G. β-bulges: Extensive structural analyses of β-sheets irregularities. *Protein Sci.* (2013). doi:10.1002/pro.2324

60. Milner-White, E. J. Beta-bulges within loops as recurring features of protein structure. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **911,** 261–265 (1987).

61. Leader, D. P. & Milner-White, E. J. Motivated proteins: a web application for studying small three-dimensional protein motifs. *BMC Bioinformatics* **10,** 60 (2009).

62. James Milner-White, E. & Poet, R. Loops, bulges, turns and hairpins in proteins. *Trends Biochem. Sci.* **12,** 189–192 (1987).

63. Greene, L. H. *et al.* Role of conserved residues in structure and stability: Tryptophans of human serum retinol-binding protein, a model for the lipocalin superfamily. *Protein Sci.* **10,** 2301–2316 (2009).

64. Huang, P.-S. *et al.* De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12,** 29–34 (2015).

65. Wang, G. & Dunbrack, R. L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **19,** 1589–1591 (2003).

66. Kozma, D., Simon, I. & Tusnády, G. E. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41,** D524–9 (2013).

67. Burley, S. K. *et al.* RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* **27,** 316–330 (2018).

68. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32,** W500–W502 (2004).

69. Chakravarty, S. & Varadarajan, R. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* **7,** 723–732 (1999).

70. Lemmon, G. & Meiler, J. Rosetta Ligand docking with flexible XML protocols. *Methods Mol. Biol.*

**819,** 143–155 (2012).

71. Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6,** e20161 (2011).
72. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30,** e43 (2002).
73. Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332,** 816–821 (2011).
74. Bhardwaj, G. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **538,** 329–335 (2016).
75. Würth, C., Grabolle, M., Pauli, J., Spieles, M. & Resch-Genger, U. Relative and absolute determination of fluorescence quantum yields of transparent samples. *Nat. Protoc.* **8,** 1535–1550 (2013).
76. Olmsted, J. Calorimetric determinations of absolute fluorescence quantum yields. *J. Phys. Chem.* **83,** 2581–2584 (1979).
77. Sjöback, R., Nygren, J. & Kubista, M. Absorption and fluorescence properties of fluorescein. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **51,** L7–L21 (1995).
78. R Development Core Team. *The R Reference Manual: Base Package*. (Network Theory., 2003).