Computational Design of Self-Assembling Cyclic Protein Homooligomers

Jorge A. Fallas\textsuperscript{1,2*}, George Ueda\textsuperscript{1,2*}, William Sheffler\textsuperscript{1,2*}, Vanessa Nguyen\textsuperscript{2}, Dan E. McNamara\textsuperscript{4,7}, Banumathi Sankaran\textsuperscript{5}, Jose Henrique Pereira\textsuperscript{5,6}, Fabio Parmeggiani\textsuperscript{1,2}, TJ Brunette\textsuperscript{1,2}, Duilio Cascio\textsuperscript{4}, Todd R. Yeates\textsuperscript{4}, Peter Zwart\textsuperscript{5}, and David Baker\textsuperscript{1,2,3*}

Table of Contents

| Supplementary Tables and Figures | 2 |
|----------------------------------|---|
| Supplementary Table 1           | 2-12 |
| Supplementary Table 2           | 13-14 |
| Supplementary Table 3           | 15 |
| Supplementary Table 4           | 16 |
| Supplementary Figure 1          | 16 |
| Supplementary Figure 2          | 17 |
| Supplementary Figure 3          | 18-19 |
| Supplementary Figure 4          | 20-21 |
| Supplementary Figure 5          | 22-23 |
| Supplementary Figure 6          | 23 |
| Supplementary Figure 7          | 24 |
| Supplementary Figure 8          | 24 |
| Supplementary Figure 9          | 25 |

| Supplementary Methods           | 26 |
|---------------------------------|---|
| Repeat protein scaffolds        | 27 |
| Cyclic symmetry definition file | 27 |
| Motif database construction     | 27 |
| Symmetric interface design      | 27-28 |
| Two-body asymmetric docking     | 28 |
| Size exclusion chromatography   | 28 |
| Protein Expression and Purification | 28-29 |
| Size exclusion chromatography with multiple angle light scattering | 29-30 |
| Generation of extension ensemble and determination of SAXS-suggested model | 30 |
| Crystallography, data acquisition, structure determination and refinement | 31-34 |

| Supplementary References       | 35 |
### Supplementary Tables and Figures

| Design      | Sequence                                                                 |
|-------------|-------------------------------------------------------------------------|
| 1na0C2 (SEC)| MHHHHHHGHGNSAEAWYNLGNAAYKQGDYDEAIYYQKALELDPNPNAE AWYNLGNAAYKQGDYDEAIYYKTKKLRDLDNAAEAWYNLGNAAYKQGDYEQAILAYVIALALDPNPNAEAKQNLNGNAEQEK |
| ank1C2_1 (SEC-MALS, SAXS) | MHHHHHHSWGSSELGKRLEIAAENGNKDRVKDLIENGADVNASDSD GRTPHLHHAENGHEVALLIESGADVNAKDSDGRTPLHHAENGHEVKKLLISGADV NAKDSGRTPLHHAENGHEVVLLASSGADVNAKDSDGRTPLPLHHAENGHKRVVLNLGADNTSDSDGRTPLDLAREHGNEEVVKALEKQ |
| ank1C2_1_r5 (SEC-MALS, SAXS) | MHHHHHHSWGSSELGKRLEIAAENGNKDRVKDLIENGADVNASDSD GRTPHLHHAENGHEVLLIESGADVNAKDSDGRTPLHHAENGHEVKKLLISGADV NAKDSGRTPLHHAENGHEVVLLASSGADVNAKDSDGRTPLPLHHAENGHKRVVLNLGADNTSDSDGRTPLDLAREHGNEEVVKALEKQQGWW |
| ank1C2_2 (SEC-MALS) | MSELGKRLEAAENGNKDRVKDLIENGADVNASDSDGRTPHLHHAENG HKEVVNLISGADVNASKDSGRTPLHHAENGHLRVLKVLLISGADV NAKDSGRTPLHHAENGHEVVLLSAGADVNAKDSDGRTPLPLHHAENGHKRVVLNLGADNTSDSDGRTPLDLAREHGNEEVVKALEDLEH|
| ank3C2_1 (SEC-MALS, SAXS) | MSELGKRLEAAENGNKDRVKDLLENGADVNASDSDGRTPHLHHAENG HAKVVLLLLLQQAGDPNAKDSGRTPLHHAENGHVAVVVALLLMHGAD PNAKDSGRTPLHHAENGHEEVVLLLAMGADPNTSDSDGRTPLDLA REHGNEEVVKALEDLEH|
| ank3C2_2 (SEC-MALS, SAXS) | MSELGKRLEAAENGNKDRVKDLLENGADVNASDSDGRTPHLHHAENG HAKVVLLLLLQQAGDPNAKDSGRTPLHHAENGHVAVVVALLLMHGAD PNAKDSGRTPLHHAENGHEEVVLLLAMGADPNTSDSDGRTPLDLA REHGNEEVVKALEDLEH|
| HR10C2_1 (SEC-MALS, SAXS) | MSSTKEELRLLVKKVVENAKRKGDDEEAREAAREAFELVREAEERA GIDSTVVVLAAAAALIMSVVAAAGSAGYDIEAARAAAEAFKRVAEAKR AGITSSSVSLAIALISLVSNAQSEGIESEADAAAEEAFKRVAEAKR AGITKEETLMIAAIELARVAAEAEERNDIEAARQAAEERKKAELK GSWLEHHHH |
| HR10C2_2 (SEC-MALS, SAXS) | MSSEKEELRRLVKIVVENAKRKGDDEEAREAARAAFEIVRAAALKLA GIDSSVEVLEAIRLIKEVVENQREGYDIAVAAIAAAAVAFVAVVAAAAA DITSSEVLEAIRLIKEVVENQREGYVILLAALAAAFAVVEVAAAKRA GITSETKLRAIEIRKRVEEAAQREGNDIEAARQAAEERKKAELKGS WLEHHHH |
| HR10C2_3 (SEC-MALS, SAXS) | MSAKETLREILVTVLAAKIGDDEEAREAARAAFEIVRAAALKLA GIDSSVEVLEAIRLIKEVVENQREGYDIAVAAIAAAAVAFVAVVAAAAA DITSSEVLEAIRLIKEVVENQREGYVILLAALAAAFAVVEVAAAKRA GITSETKLRAIEIRKRVEEAAQREGNDIEAARQAAEERKKAELKGS WLEHHHH |

2
| Protein ID | Sequence Details |
|------------|-----------------|
| HR07C2     | MHHHHHHHGSWGSTKEDARSTCEKAARKAAESNDEEVAKQAAKDCLESVAKQAGMPTKEAARSFCEAAARAAAESNDEEVIKIAAKACLEVAKQAGMPTKTAALAFCAALRAALESNDEEVAKIAMKACREVAQAGMD |
| Protein | Sequence |
|---------|----------|
| tpr1C3_2 | MAKIAMLGRVAGMQQGQLEAAAKAYKIAIELDPNDAEAWKELGKVEKL GRLDEAAEAYKKAIELDPNDAEAWKELGKVEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH |
| tpr1C3_3 | MAELAYLDGLKEAKGDKELRLLLAVDPNDAEAWKELGKVEK KQGRDLKAAAAYKKIAIELDPNDAEAWKELGKVEKLGRLDEAAEAYK KAIELDPNDAEAWKELGKVEKLGRLDEAAEAYKKAIELDPNDLEHHHH |
| tpr1C3_4 | MAQKAKRIGKAEEKGQYLLMLAYIQALHEDPNDAEAWKELGKVAE KDGLDEAAEAYKIAIELDPNDAEAWKELGKVEKLGRLDEAAEAYKKAIEL DPNDAEAWKELGKVEKLGRLDEAAEAYKKAIELDPNDLEHHHH |
| tpr1C3_5 | MAKLAMLVAGMVAQQAGRLMAAKLYKIAIELDPNDAEAWKELGKVEKL GRLDEAAEAYKKAIELDPNDAEAWKELGKVEKLGRLDEAAEAYKKAIEL DPNDAEAWKELGKVEKLGRLDEAAEAYKKAIELDPNDLEHHHH |
| ank1C3 | MSDLGLENLAMAAALGKDRVKDILENGADVNASIDGLTPLHMAAEMG HKEVVKILLISGADVNAKDSGTMPLHHAARNGHKEVVKILLISGADV NAKDSGRTPLHHAATENRHVAKEVKKILLISGADVNTSDDGRTPLDLAR EHQTIVVLLRLQHHHH |
| HL1C3 | MYENEAIIWAMQGQDYTEAAKAAEAKEGARYMTATAWEKSGDYTEAAK AKWAGDYTEAAKAWEKSGDYTEAAKAWEKAGDYTEAAKAWEKSGD YTEAAKAWEKAGDYTEA AADEAWKELGKVEKLGRLDEAAEAYKKAIEL DPNDAEAWKELGKVEKLGRLDEAAEAYKKAIELDPNDLEHHHH |
| HR10C3_1 | MSSEKEELEERLVKIVVENAKRKGDDTEEAARIIAIIFALVALAAALMA IDSSVELEAIRLIKEVVENAQREGYDISEAAALAAAEMARVAAAKRAG ITSSVELEAIRLIKEVVENAQREGQDISFIIAAAATAFKLVALAARKAG ITSETLKAIIEREIIKRVEEAEQREGNAIAAALAKAIAEAKAVKAKLGSW LEHHHH |
| HR10C3_2 | MHHHHHHGWSGEKKEELREILVAVVANAKEGKDDTEEAAREA FELVREAERAGIDSEVVLALLIIAVVILAAAMGIDSYEAAARREAAEAF KRVAEEAKRAGITSSEVELEAIRLIKEVVENAQREGYDISEAAARREAAEAF KRVAEEAKRAGITSSTTLMIAIIIRLAVEEAQAGNDEAARRREAAEAF REAAE |
| HR04C3_1 | MHTCETEARLVAEMVLEKRLGVSEDEEAIVAILISLVISTEKRSGGSYEV ICECVARVAIEVALKRSGTSEDEIAEVARISEVIRLKESSGSSYEVICE CVARVAIEVALKRSGTSEDEIAEVARISEVIRLKESSGSSYEVIC |
| HR04C3_2 | MHHHHHHGWSDEKEEKARRVAEKVERLRKSATNEAIISVREAEIS EVIRLKESSGSSYEVICVARVAIEVALKRSGTSEDEIAEVARISEVIRL KRSGTSEDEIAEVARISEVIRLKESSGSSYEVICVARVAIEVALKRSGMKCV QRIVVECVEALKRSGTSEDEIEVRKVSEVERTLKESSGWSLEHHHH |
| HR04C3_2 | MHHHHHHGWSDECEKARRVAEKSVERKRSNTANAEIAEAVREIS EVIRLKESSGSSYEVICVARVAIEVALKRSGTSEDEIAEVARISEVIRL KRSGTSEDEIAEVARISEVIRLKESSGSSYEVICVARVAIEVALKRS |
| HR04C3_2 | MHHHHHHGWSDECEEKARRVAEKVERLRKSGTSANAEIAEAVREIS EVIRLKESSGSSYEVICVARVAIEVALKRSGTSEDEIAEVARISEVIRL KRSGTSEDEIAEVARISEVIRLKESSGSSYEVICVARVAIEVALKRSG |
| HR04C3_2 | MHHHHHHGWSDECEEKARRVAEKVERLRKSANAEIAEAEVREIS EVIRLKESSGSSYEVICVARVAIEVALKRSGTSEDEIAEVARISEVIRL KRSGTSEDEIAEVARISEVIRLKESSGSSYEVICVARVAIEVALKRSMKCV QRIVVECVEALKRSGTSEDEIEVRKVSEVERTLKESSGWSLEHHHH |
| HR04C3_2 | MHHHHHHGWSDECEEKARRVAEKSVERKRSNTANAEIAEAVREIS EVIRLKESSGSSYEVICVARVAIEVALKRSGTSEDEIAEVARISEVIRL KRSGTSEDEIAEVARISEVIRLKESSGSSYEVICVARVAIEVALKRSG |
| HR04C3_2 | MHHHHHHGWSDECEEKARRVAEKVERLRKSGTSANAEIAEIAEAVREI
K64sC3 (SEC-MALS) MSEKESATLLALLRALALKAKDPEAQKSFREALGEALKKLGAASPKA IEAFAEALEALGIALEGATPDPEAIKAFAEALGAALKRLGATDPVAIVAFAL ALGLALEELGATDPEAIKAFAEALGAALKRLGATDPEAIQAFFALTALGKA LKELGATDPEAIKAFAEALGAALKRLGATDPEAIKAFAEALGKALKELG ATDPEAIKAFAEALGALRKGATDPEAIKAFAEALGKALKELGATDPE AIKAFAEALGALKRLGATSPEAIKAFAEALGALKELGATDPEAIKR F AERLGDDELREKAGTDPERIKAERERERKQEGKTDGSWLEHHHHHHH

KP16C3_1 MAKNELVDIAAQAIERQDRAVALVALSVVAMNSDKEVLEIIEAIKVAIDKQDE NIVAVAVKLVAASNDKEVLEIIEAIKVAIDKQDENIVQQVVKVAESNDKEVLE IAKVAIDKQDENIVTSVVKIAESNDKEVLEIIEAIKVAIDKQDENIVASA VKIVA ESNDKEVLEIIEAIKVAIDKQDENIVQQVVKVAESNDKEVLEIIEAIKVAIDK QDENIVTSVVKVAESNDDEEEVIQIEYKVAEREKQDENIAQIEEVLKDNGSWE LHHHHHHHHH

KP16C3_2 MHHHHHHGGSGWSEKNKLVEEEAEAIENQDENKVQQIVEDVARSNDKEV LIEIIEAIKVAIDKQDENIVASVVKVAESNDKEVLEIIEAIKVAIDKQDENIVQQV VKIVA ESNDKEVLEIIEAIKVAIDKQDENIVTSVVKIAESNDKEVLEIIEAIKVAIDK QDENIVASVVKVAESNDREVIIEAIKVAIDKQDENIVQQVVKVAESNDKVK LIQIAQVAIDKQDENIVTSVVKVAESNDDEEVIQIEYKVAEREKQDENIAQIEE VLSEDE

KP17C3_1 MSNELALDIVALSSTETESIREIEKLYKRDKELIKQAVAQAEALSVDTE VIRVIIEVLYKEDKELIKQAVAQAEALKRVDTEVIRVIIEVLYKEDKELIKQAVA EALTSVTDTEVIRVIIEVLYKEDKELIKQAVAQAEALSVTDTEVIRVIIEVLYK EDKELIKQAVAQAEALKRVDTEVIRVIIEVLYKEDKELIKQAVAQAEALT SVTDTEVIRVIIEVLYKEDKELIKQAVAQAEALKRVDTEVIRVIIEVLYKEDK ELIKQAVAQAEALT SVTDSEVIRVIIEVLYKEDKELIDEAVREALREWTDPEVQRRIIKEVRNQQG SWLEHHHHH

KP17C3_2 MHHHHHHGGSGWSEQLQCEVAERLREEDTDERIRQIIIEQLYKEDKELIK QAVAQAEALSVDTEVIRVIIEVLYKEDKELIKQAVAQAEALKRVDTEVIRVIIEV LKEDVELIAQAEALTSVTDTEVIRVIIEVLYKEAESALAVVAVAQAEASVT DTEVIRVIIEVLYKEDKELIKQAVAQAEALSVTDTEVIRVIIEVLYKEDKELIKQ AVEARDSVTDSEVIRVIIEVLYKEDKELIDEAVREALREWTDPEVQRRIIKEV RNRQGWSWLEHHHHHHHH

KP17C3_3 MRNEANLSLVALLLSVATDTERIRQIIIEQLYKEDKELIKQAVARALASV DTEVIRVIIEVLYKEDKELIKQAVAQAEALKRVDTEVIRVIIEVLYKEDKELIK QA AEALTSTDEVRIIEQVLTDKEDKELIKQAVAQAEALSVTDTEQVIMAVLFS LDKELIKQAVAQAEALKRVDTEVIRVIIEVLYSLDKELEKQAVAQAEALT STVDTEVIRVIIEVLYKEDKELIDEAVREALREWTDPEVQRIIKEVRNQQQGS WLEHHHHHHHHHHHH

KP17C3_4 (SEC-MALS) MHHHHHHGGSGWSEQLQCEVAERLREEDTDERIRQIIIEQLYKEDKELIK QAVAQAEALSVDTEVIRVIIEVLYKEDKELIKQAVAQAEALKRVDTEVIRVI EIYKEDKELIKQAVAQAEALTSTDEVRIIEVLYKEDKELIKQAVAQAEAL ASVTDEVRIIEVLYKEDKDLIAEVAEALKRVDTEVIRVIIEVLYKEDK ALIVALVAAEALSVDTEVIRVIIEVLYKEDKELIRATLAVVLAVQTDPVR MVIIIIVSVALQQ

KP17C3_5 MSNSVSLAVVIVILLSIETDTERIRIIIEKLYSREDKELIKQAVATALALVEDTE VIRVIIEVLYKEDKELIKQAVAQAEALSVSTEVEQVIRVIIEVLYKEDKELIKQAVA EALTSTDEVRIIEVLYKEDKELIKQAVAQAEALSVDTEVIRVIIEVLYKE DKELEKQAVAQAEALKRVDTEVIRVIIEVLYKEDKELIKQAVAQAEALTSTDVTEV
IRVIIEVLYKEDERLIDEAVREALREVTDPEVQRIIKEVVRNQQGWSLLEHHH

HR00C3_1
MIEEVVAEMIDILAESSSKSIEELAQADNKTTEQVQAQSIEQIANNATTIIQLIENLAKLASEEFMARAIASIAELAKKAEIYRLADNHTDTFMANAINAIAI\NATAILAIANLASHTEEMMARAIASIAELAKKAEIYRLADNHDTDKFMAQAIEAIALLATAILAIALLASNHTTEEMFAIKAISIAIAELAKKAI\EYRLADNHTSPTYIEKAIKAEIIKARKAIKAIKAIEMLAKNITEYEKFAIKSAIDEIREKAKEAIKRLEDNRTLEHHHHHH

HR00C3_2
MIEEVVAEMIDILAESSSKSIEELARAADNKTTEKAVAEAIIEIARLATAAIQLIIEAIKELSIEEFMARAIASIAELAKKAEIYRLADNHTDTFMARAI\AAIYRLADNHTSPTYIEKAIKAEIIKARKAIKAIKAIEMLAKNITEYEKAKSAIDEIREKAKEAIKRLEDNRTLEHHHHHH

prxC3
MGDEMRKVMLALAIYRALLNEDIEVAKEIARAADIEEALRENNSDEMAKFMLAKAALAVLLAALKNNDDEVAKEIARAAAMIIVIALRAENSDEMAKKMILEL\AKRVLDAAKNNDATEIREEQAEAAELEAWLEHHHHHH

HR08C3
MGHHHHHHWDEECEEKARRVAEKVERLKRSGTSDEAIIEEVAREISEVIPRLKESSGSSYEVICEVARVIAEVALKRSGTSGSKSKEIIEVARIVISEVIRL\LKESSGSSYEVICEVARVIAEVEALKRSGTSDEAIIEEAVARESVEVIRLLESGSEEEVILKVCARILLEALERSGTKKLIIALMLIVLIVLITIRS

HR04C3_3
MGHHHHHHWDEECEEKARRVAEKVERLKRSGTSDEAIIEEVAREISEVIPRLKESSGSSYEVICEVARVIAEVALKRSGTSGSKSKEIIEVARIVISEVIRL\LKESSGSSYEVICEVARVIAEVEALKRSGTSDEAIIEEAVARESVEVIRLLESGSEEEVILKVCARILLEALERSGTKKLIIALMLIVLIVLITIRS

HR79C3
MGESDEILAMLVIILALLAIALMMAAETGDPRVEEALASELVEAAAEVEEDPSRDVLKALLLIIIVALVALLAALITGDPVERELARELVRL\AVEAAEVEQRNPSSDVEAILKIVALAAVRLAARAGTDPVRELARELVRLAVEAAEVQRNPSSDEVEANEALKVIIKAVEAVERAEVASLREAESG\DPEKREKARERVEAVERAEVEQQRPDPSLWLEHHHHHH

3ltjC3
MGHREHTDPLKVLLYIVILEAELYLRRAAAALGKIDDEEAVEPLIKAL\KDVEDALVRRAAADALGQIGDERAVEPLIKALKVEDGLRASAVALGQIG\DERAVEPLIKALKDERVRAAAALGIQDERAVEPLIKALKDERGK\VRKAAARALGEIGGERVERAAAMEKLAETGTGFRKAVNLYLTHKWLEHHHHHH

ank1C4_1
MSMLGKLILLLAELGLLLVVVMLLISNGADVNASDSDGRTPLHHAANENG\HKMVVMILLIKGADVKNAKDSGDGRTPLHHAENGHKEVKKELIEMGADV\NAKDSGDGRTPLHHAANENGKKEVKKLISGADVNTSDGRTPLDLAR
| Code     | Sequence                                                                                           |
|----------|----------------------------------------------------------------------------------------------------|
| ank1C4_2| **MSEDGELLILAAELGIAEAVRMLIEQGADVNASDDDGRTPLHHAENGHLAVVLLLLKGDADVNAKDSGDRTPLHHAENGHTVVLLLILMGADV** |
| ank3C4_2| **MTELGIALAIALVGDKDRVKDLLENGADVNASARAGMTPHLAALLLGHKVEVKLLLSQGADPNAKDKDGKTPHLHAAENGHWAVHLLEQGADPNTSDSGRTPLDLAKEHGEVVVTLLLKGGEIAHHDDLEHHHHH** |
| ank3C4_1| **MNDLGMLLIMAAMEGKIVVVLLEKGADPNASKDGKTPHLHAAENGGLIIVLLLEKGADPNASKDGKTPHLHAAENGHEIKEVEALLEHGPNAKDSDGRTPLHYAAENGHKEIVKLLLSQGADPNTSDSGRTPLDLAREHGENEEIVKLLEKQLEHHHHH** |
| ank4C4  | **MSTEGKMLIIAAREGMIVVIVLLEKGADPNASDKDGRTPLHAAENGHLIIVLLLEKGADPNASKDGKTPHLHAAENGHEIKEVEALLEHGPNAKDSDGRTPLHYAAENGHKEIVKLLLSQGADPNTSDSGRTPLDLAREHГНЕЕIVKLLEKQLEHHHHH** |
| 1na0C4_1| **MTLARVAYILGAIAYAQGPEYDIAITAYQVALSDLDPNNAEAWYNLGNAYYKQGDYDEAEIYQQKALELDPNNAEAWYNLGNAYYKQGDYDEAEIYQQKALELDPNNAEAKQNGLNAPQKQGLEHHHHH** |
| 1na0C4_2| **MTAAEIAYNMGAAYKEGDYMAMITAYQLAEELDPNNAEAAANLGNAYYKQGTYLMAILFYLIALLDDPNNAEAWYNLGNAYYKQGDYDEAEIYQQKALELDPNNAEAKQNGLNAPQKQGLEHHHHH** |
| 1na0C4_3| **MHHHHHHGSNNAEAWYNLGNAYYKQGDASEAIVYYLLAVLDPNNAEAWYNLGNAYYKQGDYDEAEIYQQKALELDPNNAEAWYNLGNAYYKQGDYDEAEIYQQKALELDPNNAEAKQNGLNAPQKQGLEHHHHH** |
| 2fo7C4  | **MASMAIYNSFYQGDYTMAMLMYILALLLDPRSAAAYNLGNAYGKYDEEAEIYQQKALELDPNNAEAWYNLGNAYYKQGDYDEAEIYQQKALELDPNNAEAKQNGLNAPQKQGLEHHHHH** |
| HR10C4_1| **MTVLAVILALILIIVANAKRKGDITTEAAALAAEEAFALVLVAAAERAGIDSSEVLELAIRLIKEKVENAQPREGYDISEAAARAAAEEAFKRVAEEAAKRAGIT** |
| HR10C4_2| **MHHHHHHGSWGSSEKEELRLLLAVIMIAAIKGGDSEEAREAAAREAFELVREAERAGIDSSEVLELAIRLIEEVENAEEEYGYDISEAAARAAAEEAFKRVAEAAKRAGITSSEVLKMAIEIRKVEECQRGENDISEAAARQAEEEFKKAEELEKGSWLEHHHHH** |
| HR04C4_1| **MHHHHHHGSWGSDECEKARRVAEKLKRSGETSEIAEAEVAREISEVRILKESGGSSYEVICEVCVARIAEVEALKRSGTSAEVIAKIVARVISEVIRTLEKESGGSSYDEVICEVARIAEVEALKKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECIRVILEIIIEALKRSGETSEQDVMLIVMAVLVVLATLQLSGS** |
| Sequence ID | Description | Sequence |
|-------------|-------------|----------|
| SEC-MALS   | ALKDERKVATAAAALGAGDRAVEPLIALKDEEGAVRSLAVALGKGDRAVEPLIALKDERKVVRVAAAFALGEIGDERAVEPLIALKDE EGMVRQSAAADALGGEIGERVRAAMEKLAETGTGFARvKAVNVYEHTK |
| HL1C5      | MYQLIAMLMVLGAYKLAIAAEKAGLYLAAAVAWELSGDYTEAAKA WEKAGDYTEAAKAWESGDGTYTEAAKAWEKAGDYTEAAKAWEKSGD YTEAAKAWEKAASYEAAKWAWEKSGDYTEAAKAWEKAGDYTEAAKAW EKSGDYTEAAKAWEKAGDLEHHHHHH |
| tpr1C5     | MAEAWKELYLGVKLGLDEAEAYGAIEDPNDAAEAWKELGVLEKL GRLDEAEAYGAIEDPNDAAEAWKELGVLEKLGLDEAEAYRMMALL EDVTDAEAAMLLGRLVGLTGLRALLAMILAVLLKPNALIEHHHH |
| ank1C5     | MSILGLMLVAARNGKDKLVRLIENGADVNASDSGRTPLHAAENGA EVVEILLISGADVNAXAKDSGRTPLHAAENGAHILVLLLSKADVNKS DGRTPHLHAAENGEHKVEVIALSAGADVNTSDSGRTPDLAREHGNIEE VVKLLEKQGWSLEHHHHHH |
| arm8C5_1   | MNTVERAVKLLTSTSDRTQIAAAALLALIASGPASAIVLVIAGGVEVLV KLTTSTDSEVQKEAARALANIASGPDVDAIRAVEAGGVEVLVKTLLTSTDSE VQKEAARALANIASGPDDEAIKAAVADGVEVLVKTLLTSTDSEVQKEAAR ALANIASGPDDEAIKAAVADGVEVLVKTLLTSTDSEVQKEAARALANIASG PTSAIKVDAVGVEVLQKLLTSTDSEVQKEAQRALENIKSGWSLEHHH HH |
| arm8C5_2   | MNDVEQLVKALTSTDSTLQMAAMMLAEIASGPARAIALIVYAAGGVEVLV KLTTSTDSEVQKEAARALANIASGPDVDAIRAVEAGGVEVLVKTLLTSTDSE VQKEAARALANIASGPDDEAIKAAVADGVEVLVKTLLTSTDSEVQKEAARAL ANIASGPDDEAIKAAVADGVEVLQKLLTSTDSEVQKEAARALANIASGPTS AIKAIKVDAVGVEVLQKLLTSTDSEVQKEAQRALENIKSGWSLEHHH HH |
| HR10C5_1   | MSNTQLLVEILVIIVANARRKDDEAKLAAIALLVIAAERAGIDS SEVLELAIRLIKEVVENAQREGYDISEEAARRAAAEAFKRVAAEAKRAGITS SEVLELAIRLIKEVVENAQREGYDISEEAARRAAAEAFKRVAAEAKRAGITS SETLKRÄIEEKRVERVIAEQREGNDISEEAAQAAEERFKNKAEELKGWSLE HHHHH |
| HR10C5_2   | MSAEKLMLMAKLIIVAENAKRKGDDTLIAAMALFEIVRIAEEAGID SSEVLELAIRLIKEVVENAQREGYDISEIAALAAAMAFALVAIAEAKRAGITS SEVLELAIRLIKEVVENAQREGYDISEEAARRAAAEAFKRVAAEAKRAGITS SETLKRÄIEEKRVERVIAEQREGNDISEEAAQAAEERFKNKAEELKGWSLE HHHHH |
| HR07C5_1   | MTKSTARITCMIAAIAAARENDEMVAVMAALVCLMVAEQAGMPTEAA RSFCEAAARAAASINDDEEVKIAAACKACLEVAAAGMPTKEAARSFCEAA AARAAAESNDDEEVKIAAACKACLEVAKQAGMPTEAARSFCEAAAKRAA KESNDDEEVEKIAAACKACLEVAKQAGMPGWSLEHHHHHH |
| HR07C5_2   | MHHHHHHGSHWSGTKEARDSTCEAARKAAAESENDDEVKAAKDCLE VAQAGMPTEAARSFCEAAARAAAESNDDEVKIAAACKACLEVAKQA GMPTKEAARSFCEAAARAAAESNDDEVKIAAACKACLEVAKQAGMPMTR EAAAFCAVARRALAMESNDDEEVEKIAECAACLVALQAGMP |
| HR07C5_3   | MHHHHHHGSHWSGTKEARDSTCEAARKAAAESENDMEAAIAALLCALVA KEAGMPTKEAARSFCEAAARAAAESNDDEVKIAAACKACLVAKAAGMP |
| **HR04C5_1** (SEC) | **MRACEAEAMLIKAIVMMLKENGTEDEIAAEVAREISEVIRTLKESGSSYLVICECVARIVAMVEALKLSGTEDEIAEIVARVISEVIRTLKESGSSYKCI CVCVADIVAEIVEALKRNTSEDEIAEIVARVISEVIRTLKESGSSYEVIAA CVIAIVLAAIKLRKSGTEDEINEIVRRVKSEVERTLKESGSWLEHHHHH H |
|-------------------|----------------------------------------------------------|
| **K64sC5**        | **MKEEAAERTALLALLALSELRRDPLAQLLFDVALGVALDRLGAASPEAIKA FAEALGKALKELGATGSPVAAAFALALGALKRALGATPDPEAIKAFAEALGK ALKELGATPDPEAIKAFAEALGALKRALGATPDPEAIKAFAEALGKALKNGA AALKRLGATPDPEAIKAFAEALGKALKELGATPDPEAIKAFAEALGALKRALGATPDPEAIKAFAEALGKALKELGATPDPEAIKAFAEALG |
| **KP16C5_1**      | **MHHHHHHGSGSWSGSEKNKLVVEAAKEAIENQDKKKKVSQIVEDVARSNDKEV LIEIAKVAIDKQDENIVAVVVVLVASNDKEVLIIEAIKVAIDKQDENIVQVLLV SIVAISNDKEVLIIEAIKVAIDKQDENIVTSVRIVARSNDKEVLIIEAIKVAIDK QDENIVASVVKVIAESNDKEVLIIEAIKVAIDKQDENIVQQVVKVIAESNDKEV LIEIAKVAIDKQDSNVSTSVKIVAESNDEVIQIEYKVAREKQDRIAVSII VVLLDD** |
| **KP16C5_2**      | **MHHHHHHGSGSWSGSEKNKLVVEAAKEAIENQDENKVQIVEDVARSNDKEV LIEIAKVAIDKQDENIVAAVVIASNDKEVLIIEAIKVAIDKQDENIVTVVAV VAVSEDKEVLIIEAIKVAIDKQDENIVTSVRIVARSNDKEVLIIEAIKVAIDKQ DENIVASVVKVIAESNDKEVLIIEAIKVAIDKQDENIVQQVVKVIAESNDKEV LIEIAKVAIDKQDRIVRSVVKVIAESNDEVIQIEYKVAREKQDRLTIAVSM VVAIRED** |
| **2fo7C6** (SEC)  | **MAEALYNMGKYYKQGDYEVAIIAYQQALELDPRSAEAWYNLGNAYY KQGDYDEAEAYYQKALELDPRSAEAYNALGNAYYKQGQGDYQAEEAIAYM ALALDPRSAEAWYNLGNAMYKMGIYDASIEYYQKALELDPRSAEAWYNLGNAYY** |
| **arm8C6** (SEC)  | **MNEVEKLVKLLTTSTDSVLMMAALANIASGPAAAIARIILAGGVKVL VKLTTSTDSAVQKLARALANIASGPDDLAILAIVDAGGVEVLKVLLTTSTD SEVQKEAARALANIASGPDEAIKAIVDAGGVEVLKLLTTSTDSEVQKEAARALANIASGPDEAIKAIVDAGGVEVLKLLTTSTDSEVQKEAARALANIASGPDEAIKAIVDAGGVEVLKLLTTSTDSEVQKEAARALANIASGPDEAIKAIVDAGGVEVLKLLTTSTDSEVQKEAARALANIASGPDEAIKAIVDAGGVEVLKLLTTSTDSEVQKEAARALANIA** |
| **HR10C6_1**      | **MSKEKEALLRLIAIIVIAALRKGDDAQAAMAARRVAFLLVRMAAIAGIDSS EVLLEALRILLEKVENAQQREGYDIAAAALAAALAFMRAAEAAKRAGITSSE VELEAIKLEKVVENAQREGYDIAAAALAAALAFMRAAEAAKRAGITSSETL KRAIEEIRKRVEEAAQREGNIDEAARQAAEEFRRKKAEEKLGSWLEHHHHH H** |
| **HR10C6_2** (SEC) | **MSAVKQALMRMLIMIENAKRKGDTRLAEEKAAEAFIEIVREAARAG IDSSEVLEALRILLEKVENAQQREGYDISKAALAAASAFMRAEAAKRAGITSSE EVLLEALRILLEKVENAQQREGYDIAAAALAAALAFMRAAEAAKRAGITSSETL KRAIEEIRKRVEEAAQREGNIDEAARQAAEEFRRKKAEEKLGSWLEHHHHH H** |
**Supplementary Table 1.** List of all designs tested and their corresponding amino acid sequences including initiating methionine and (His)$_6$ tag. Designs that expressed solubly are denoted in bold and the experimental methods used to characterize them are listed under their name.
| Design      | MW design (kDa) | MW mon (kDa) | MW MALS (kDa) | Oligomerization State | Elution Volume (mL) |
|------------|----------------|-------------|---------------|-----------------------|---------------------|
| ank1C2_1   | 36.2           | 18.1        | 31.5          | 1.7                   | 16.0                |
| ank1C2_2   | 35.8           | 17.9        | 25.5          | 1.4                   | 11.9*               |
| ank1C3     | 53.4           | 17.8        | 28.2          | 1.6                   | 17.7                |
| ank1C4_1   | 71.6           | 17.9        | 38.8          | 2.2                   | 14.1                |
| ank1C4_2   | 71.6           | 17.9        | 68.9          | 3.8                   | 14.1                |
| ank3C2_1   | 35.6           | 17.8        | 32.0          | 2.0                   | 11.2                |
| ank3C2_2   | 35.8           | 17.9        | 30.1          | 1.7                   | 16.1                |
| ank3C4_1   | 35.8           | 17.9        | 75.0          | 4.2                   | 12.0                |
| ank3C4_2   | 35.4           | 17.7        | 55.0          | 3.1                   | 14.7                |
| ank4C4     | 72.4           | 18.1        | 72.4          | 4.0                   | 14.1                |
| 1na0C3_1   | 44.4           | 14.8        | 46.1          | 3.1                   | 15.2                |
| 1na0C3_3   | 44.1           | 14.7        | 45.6          | 3.0                   | 15.3                |
| 1na0C3_5   | 44.2           | 14.7        | 44.0          | 3.0                   | 15.3                |
| 1na0C3_7   | 44.1           | 14.7        | 46.5          | 3.2                   | 14.6                |
| 1na0C4_1   | 58.0           | 14.5        | 55.1          | 3.8                   | 13.3                |
| tpr1C3_2   | 48.6           | 16.2        | 16.3          | 1.0                   | 16.7                |
| tpr1C3_3   | 48.9           | 16.3        | 57.4          | 3.5                   | 11.0*               |
| tpr1C3_4   | 48.6           | 16.2        | 30.0          | 1.9                   | 10.0*               |
| tpr1C4_2   | 64.8           | 16.2        | 65.5          | 4.0                   | 14.4                |
| 3ltjC3     | 65.4           | 21.8        | 237.7         | 10.9                  | 12.2                |
| 3ltjC5     | 103.9          | 20.8        | ----          | ----                  | ----                |
| KP16C6     | 176            | 29.3        | 270           | 92.2                  | 8.0                 |
| KP17C3_4   | 90.10          | 30.3        | 87.0          | 2.9                   | 13.0                |
| K64sC3     | 105.9          | 35.3        | 38.2          | 1.1                   | 14.0                |
|                   | MW   | Mw   | MW MALS | MW Mon | MW Design |
|-------------------|------|------|---------|--------|-----------|
| HL1C5             | 90.5 | 18.1 | 66.0    | 3.6    | 14.4      |
| HR00C3_2          | 93.6 | 31.2 | 81.0    | 2.6    | 13.9      |
| HR00C6            | 190.2| 31.7 | 237.0   | 7.5    | 12.2      |
| HR04C3_3          | 66.9 | 22.3 | 108.4   | 4.9    | 13.4      |
| HR04C4_1          | 90.4 | 22.6 | 88.2    | 3.9    | 13.5      |
| HR04C4_3          | 91.2 | 22.8 | 43.3    | 1.9    | 15.3      |
| HR04C4_4          | 90.8 | 22.7 | 119.5   | 5.3    | 13.5      |
| HR04C6_3          | 140.4| 23.4 | 112.8   | 4.8    | 13.5      |
| HR08C3            | 60.9 | 20.3 | 56.5    | 2.8    | 15.0      |
| HR10C2_1          | 43.4 | 21.7 | 34.2    | 1.6    | 15.9      |
| HR10C2_2          | 44.9 | 22.4 | 42.6    | 1.9    | 15.6      |
| HR10C2_3          | 43.0 | 21.5 | 42.8    | 2.0    | 15.5      |
| HR10C3_2          | 64.5 | 21.5 | 44.2    | 2.1    | 15.5      |
| HR10C5_1          | 113.7| 22.7 | 370.0   | 165.9  | 8.1       |
| HR10C5_2          | 113.4| 22.7 | 120.4   | 1.1    | 13.5      |
| HR14C6            | 116.4| 19.4 | 53.8    | 2.8    | 15.0      |
| HR54C6            | 119.4| 19.9 | 117.1   | 5.9    | 13.0      |
| HR71C6            | 139.2| 23.2 | 111.9   | 4.8    | 13.3      |
| HR79C2            | 53.4 | 26.7 | 47.9    | 1.8    | 15.2      |
| HR81C2            | 51.8 | 25.9 | 46.8    | 1.8    | 15.2      |
| ank1C2_1_r5       | 50.7 | 25.4 | 48.0    | 1.9    | 15.7      |
| HR04C4_1_r4       | 134.7| 33.4 | 120.9   | 3.6    | 13.0      |

* Elution volume reported for a Superdex 75 10/300 GL gel filtration column.

**Supplementary Table 2.** Summary of molecular weights used to assess the oligomeric state of the proteins in solution. MW design refers to the expected molecular weight for designed oligomer, MW mon is the molecular weight of the protomer, MW MALS is the experimentally determined molecular weight by multi-angle light scattering. Designs in bold have the expected oligomerization state with discrepancies between the experimental and computational quantities <= 13%.
| Design          | Rg model (Å) | Rg SAXS (Å) | χ SAXS | r.m.s.d. Crystal Struct. | Closest Homolog (Sequence Identity % / r.m.s.d.) |
|-----------------|--------------|-------------|--------|--------------------------|-----------------------------------------------|
| ank3C2_1        | 21.9         | 20.2        | 1.9    | 1.0 Å                    | ank1C2_1 (79 / 5.3 Å)                      |
| ank1C2_1        | 25           | 24.0        | 2.7    | 0.9 Å                    | ank3C2_1 (79 / 5.3 Å)                      |
| HR10C2_2        | 24.2         | 25.7        | 2.1    | ----                     | HR10C5_2 (71 / 10.4 Å)                     |
| HR79C2          | 22.5         | 23.4        | 2.4    | ----                     | HR81C2 (26 / 18.1 Å)                       |
| HR81C2          | 20.8         | 23.1        | 2.5    | ----                     | HR79C2 (26 / 18.1 Å)                       |
| 1na0C3_1        | 27.9         | 28.0        | 1.2    | ----                     | 1na0C4_1 (84 / 7.9 Å)                      |
| 1na0C3_3        | 29.3         | 29.0        | 1.0    | 1.0 Å                    | 1na0C3_7 (88 / 1.9 Å)                      |
| 1na0C3_5        | 29.9         | 29.5        | 2.6    | ----                     | 1na0C3_7 (84 / 2.5 Å)                      |
| 1na0C3_7        | 26.2         | 25.4        | 1.0    | ----                     | 1na0C3_3 (88 / 1.9 Å)                      |
| HR00C3_2        | 31.5         | 35.0        | 3.1    | 0.9 Å                    | HR10C2_2 (26 / 21.6 Å)                     |
| HR08C3          | 24.2         | 27.0        | 2.1    | ----                     | HR79C2 (20 / 21.9 Å)                       |
| 1na0C4_1        | 31.2         | 30.7        | 2.3    | ----                     | 1na0C3_1 (84 / 7.9 Å)                      |
| ank1C4_2        | 32.2         | 29.2        | 2.9    | 1.1 Å                    | ank1C2_1 (81 / 13.0 Å)                     |
| HR04C4_1        | 33.2         | 36.0        | 2.8    | ----                     | HR81C2 (18 / 17.8 Å)                       |
| HR10C5_2        | 38.9         | 43.3        | 3.1    | ----                     | HR10C2_2 (71 / 10.4 Å)                     |
| ank1C2_1_r5     | 33.2         | 35.8        | 2.9    | ----                     |       | ----|
| HR04C4_1_r4     | 40.9         | 41.9        | 1.0    | ----                     |       | ----|

**Supplementary Table 3.** Summary of experimental and computed quantities used to assess the supramolecular configuration of the designed oligomers in solution using SAXS and their closest homolog in the set determined by global multiple sequence alignment. The designs in this table are considered successful and exhibit discrepancies between the experimental and computational quantities < 13% for the molecular weight, < 11 % for the radius of gyration and a χ <= 3.1.
| Design     | Rg model (Å) | Rg SAXS (Å) | χ SAXS | RMSD xtal |
|------------|-------------|-------------|-------|----------|
| ank3C2_2  | 24.9        | 21.6        | 5.1   | -----    |
| HR10C2_1  | 20.14       | 23.0        | 2.7   | -----    |
| HR10C2_3  | 24.5        | 30.0        | 22.5  | -----    |
| HR10C3_2  | 24.5        | 29.1        | 6.6   | -----    |
| ank3C4_1  | 31.1        | 31.1        | 23.0  | -----    |
| ank4C4    | 32.2        | 36.4        | 3.8   | > 10.0 Å |
| HR04C4_3  | 35.6        | 28.4        | 26.1  | -----    |
| HR04C4_4  | 35.5        | 42.5        | 29.1  | -----    |
| tpr1C4_2  | 21          | 23.1        | 8.8   | -----    |
| HR00C6    | 50.2        | 46.1        | 2.2   | -----    |
| HR54C6    | 33.4        | 48.9        | 15.6  | -----    |

**Supplementary Table 4.** Summary of experimental and computed quantities used to assess the supramolecular configuration of the designed oligomers in solution using SAXS. The designs in this table exhibit large discrepancies between the experimental and computational values for molecular weight and/or X-ray scattering profiles.

**Supplementary Figure 1.** Comparison of the full atom and coarse-grained computed binding energies for a set of designed repeat protein oligomers. **a**, Rosetta full atom energy function and a sequence-agnostic Rosetta centroid score function. **b**, Rosetta full atom energy function and RPX score as utilized to design oligomers in the main text.
Supplementary Figure 2. Histograms of computational metrics for designs that expressed solubly. Structurally validated designs are shown in red and unsuccessful designs in blue.
Supplementary Figure 3. SEC profiles for failed designs. Primary size exclusion chromatograms obtained from a Superdex 200 gel filtration column for expressed proteins directly after purification by immobilized Ni$^{2+}$ affinity chromatography. Designs listed here either obviously formed non-specific assemblies and/or did not appear to form the target oligomeric species based on the column retention time of their corresponding monomeric scaffolds. 5 designs presented here had monodisperse profiles with retention times that were estimated to form off-target oligomeric species, whereas 15 designs had polydisperse or aggregated profiles.
* Size-exclusion chromatogram produced using a Superdex 75 10/300 GL gel filtration column

**Supplementary Figure 4.** SEC-MALS chromatograms for failed designs. Predominant oligomeric species for each design were collected by fractionation from a primary size exclusion run, and 29 sizing profiles are presented here from the subsequent round of high-performance size exclusion chromatography from a Superdex 200 gel filtration column.
Supplementary Figure 5. Assessment of the solution conformation of selected cyclic oligomers. From left to right: computational model, symmetric docking funnel, SEC chromatogram used for molecular weight determination and SAXS scattering profiles, measured (black dots) and computed from the model (red line). a, ank1C2_1. b, HR10C2_2. c, ank1C4_2. d, 1na0C4_1. e, 1na0C3_1. f, 1na0C3_3. g, 1na0C3_5. h, 1na0C3_7. i, HR81C2.

Supplementary Figure 6. Solution conformation of design ank1C4_2. a, Comparison between the experimental SAXS scattering profile (black dots) and C4 symmetric design model (red line). b, Comparison between the experimental SAXS scattering profile (black dots) and C2 tetramer found in the X-ray structure.
Supplementary Figure 7. SEC-MALS of selected designs in 1M and 2M GuHCl. Secondary size exclusion profiles obtained from a Superdex 200 5/150 GL gel filtration column (3 mL volume) coupled to multi-angle light scattering in 25mM Tris 150mM NaCl pH 8.2 1M GuHCl (top row) and 2M GuHCl (bottom row). “MW (design)” refers to the molecular weight of the target oligomer, whereas “MW (MALS)” refers to the experimentally determined molecular weight for the predominant species. a, HR81C2. b, HR00C3_2. c, HR04C4_1. d, HR10C5_2.

Supplementary Figure 8. Solution-state SAXS profile comparison for ank4C4 design model and corresponding crystal structure. a, Comparison between the experimental SAXS scattering profile (black dots) and C4 symmetric design model (red line). b, Comparison between the experimental SAXS scattering profile (black dots) and D2 symmetric crystal structure (cyan line).
Supplementary Figure 9. Structural comparison of the 15 validated designs. Designs are grouped by repeat scaffold family and oligomerization state; β-propellers (PDB ID from left to right 3ww9, 3ww8, 3wwb) are shown for structural diversity comparison. Boxed designs share < 2.5 Å r.m.s.d.
Supplementary Methods

Repeat protein scaffolds. Repeat proteins, comprised of recurring 20-50 residue stretches, are ideal for use in protein-based material design due to their high stability and capability to have altered lengths and curvatures by varying the number of repeating modules. Listed below are the RCSB Protein Data Bank entries for selected scaffolds. An additional a set of scaffolds is provided in which experimental small-angle X-ray scattering data agreed with the computational model.

| Crystal Structures (PDB ID) | SAXS Validated Models |
|-----------------------------|-----------------------|
| ank1 (4GPM)                 | tpr1                  |
| ank3 (4GMR)                 | KP16                  |
| ank4 (4HB5)                 | KP17                  |
| 3ltj (3LTJ)                 | K64s                  |
| 1na0 (1NA0)                 | hl1                   |
| 2fo7 (2FO7)                 | HR00                  |
| arm8 (4HXT)                 |                       |
| HR04 (5CWB)                 |                       |
| HR07 (5CWD)                 |                       |
| HR08 (5CWF)                 |                       |
| HR10 (5CWG)                 |                       |
| HR14 (5CWH)                 |                       |
| HR54 (5CWL)                 |                       |
| HR64 (5CWM)                 |                       |
| HR71 (5CWN)                 |                       |
| HR79 (5CWP)                 |                       |
| HR81 (5CWQ)                 |                       |
Cyclic symmetry definition file. In order to model cyclic homooligomers within the Rosetta framework, we implemented symmetry definition files to generate the specified symmetry starting from a single subunit of each docked configuration. These files are needed to properly calculate all of the score terms in symmetric poses. The cyclic symmetry definition files used in this study are attached to the manuscript.

Motif database construction. The database of pairwise motifs used in this study was constructed from a culled set of deposited proteins from the RCSB Protein Data Bank as of August 2012, comprised of 23162 structures with a reported resolution of 2.5 Å or less. Because each independently crystallized protein in this set was thought to contain new structural information and afford high-resolution features, no sequence redundancy cut was used. Residue pair contact areas were filtered with a 1.7 Å² probe, only highest occupancy state rotamers were used, and pairs without at least one partner within the first chain were ignored. Weights for each residue pair were introduced using the Rosetta Score12 score function³ and the Rosetta implementation of the DSSP secondary structure prediction server was used to assign a backbone type to each residue in a given pair. Residue pair energies greater than -0.001 energy units were discarded.

Symmetric interface design. The top ten non-redundant cyclic docked configurations in order of motif score were chosen as the input set for a Rosetta interface design protocol. This protocol took an input .pdb file containing a single subunit of each docked configuration with the cyclic axis aligned to the vector [0, 0, 1], as well as a cyclic symmetry definition file. In each design trajectory, the subunit was initially perturbed by a translation perpendicular to the axis of symmetry, as well as a random rotation in three-dimensional space. The applied perturbation was selected from a gaussian distribution bounded by user-defined distances and angles. An oligomer with the specified cyclic symmetry was then generated using the information stored in the symmetry definition file to cyclize the monomeric subunit. For the interface design protocol, designable positions were designated as residues that met the following criteria: beta carbon within 10 Å of at least one beta carbon from another subunit, at least one atom within 5 Å of any other atom from the same subunit, a non-zero surface area accessible to solvent, identity of neither proline nor glycine. While keeping other subunit residues fixed, designable positions were optimized using the Rosetta packing
algorithm with the default Talaris2013 score function\textsuperscript{3} and the extended rotamer library available in Rosetta. Initially, packing was executed with a modified score function using a full atom repulsive term weight of 0.05 (as opposed to a standard weight of 1.0) in order to sample more of the sequence space for the particular configuration. Once a sequence was converged upon, designable positions were allowed to minimize side chain torsion angles using the same reduced repulsive term weight. A subsequent round of packing and minimization was conducted, but with the repulsive term weight reset to 1.0 in order to converge on a local minimum of standard Rosetta energy. Individual design trajectories were filtered and refined by single point reversions for mutations that were deemed non-contributory to stabilizing the bound state of the interface. The design with the best overall scores for each unique docked configuration was then added to a set of finalized proteins to be experimentally validated.

**Two-body Asymmetric Docking.** Asymmetric docking was performed for chains A and B of each oligomer. Patchdock\textsuperscript{4} was used to generate 7 alternative starting configurations that were each used together with the designed binding mode as a starting points for a RosettaDock\textsuperscript{5}. Each oligomer was characterized by a value ($\Delta E = \min(E_{\text{local}}) - \min(E_{\text{global}})$) corresponding to the difference in energy of the lowest energy state sampled during the local section of the algorithm minus the lowest energy state sampled during the global section of the algorithm. Negative values indicate that the starting conformation corresponds to the lowest energy state found during docking while positive values indicate that alternative lower energy conformations were found.

**Size exclusion chromatography.** Elution samples for each designed protein were concentrated down using a 10,000 MWCO protein concentrator (Novagen) and fractionated by size on an AKTA pure chromatography system using a Superdex 200 10/300 GL gel filtration column (GE Life Sciences) in 25mM Tris 150mM NaCl pH 8 (TBS) unless otherwise noted. Sizing profiles were noted based on absorption at 220 nm and 280 nm wavelength light for each fraction. Molecular weights for predominant species in each protein trace were estimated by comparison to the corresponding monomeric profile.

**Protein Expression and Purification.** Synthetic genes for these designed proteins were optimized for *E. coli* expression and assembled from purchased genes (Genscript)
ligated into the pET21-NESG vector at restriction sites NdeI and XhoI. These plasmids were cloned into BL21 (DE3) E. coli competent cells. Transformants were inoculated and grown in either LB or TB medium with either 100 mg L\(^{-1}\) carbenicillin or 150 mg L\(^{-1}\) ampicillin at 37 °C until an OD\(_{600}\) of 0.7. Isopropyl-thio-β-D-galactopyranoside was then added at a concentration of 1 mM to induce protein expression. Expression proceeded for 20 hours at 18 °C until the cell cultures were harvested by centrifugation. Cell pellets were resuspended in TBS and lysed by sonication. Each filtered lysate was then purified by Ni\(^{2+}\) immobilized metal affinity chromatography with Ni-NTA Superflow resin (Qiagen). Resin with bound cell lysate was washed with five column volumes of 25 mM imidazole and five column volumes of 50mM imidazole. The desired proteins were then eluted with five column volumes of 400 mM imidazole and further purified by size exclusion chromatography.

**Size exclusion chromatography with multi-angle light scattering.** Fractions containing single predominant species from the initial round of size exclusion chromatography were concentrated down with 10,000 MWCO protein concentrators (Novagen) to a concentration of 1.0–2.0 mg mL\(^{-1}\). 100 uL of each sample was then run through a high-performance liquid chromatography system (Agilent) using (unless otherwise noted) a Superdex 200 10/300 GL gel filtration column (GE Life Sciences) at an elution rate of 0.50 mL min\(^{-1}\) in TBS. These fractionation runs were coupled to a multi-angle light scattering detector (Wyatt) in order to determine the absolute molecular weights for each designed protein. The following equation\(^6\) derived from the Rayleigh-Debye-Gans theory of light scattering\(^7\) was used in the ASTRA software to calculate the molecular weight of the major species present in each sample:

\[
\frac{K^*c}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2c
\]

where:

- \(R(\theta,c)\) is the excess Rayleigh ratio of the solution as a function of scattering angle \(\theta\) and concentration \(c\). It is directly proportional to the intensity of the scattered light in excess of the light scattered by the pure solvent.
- \(c\) is the solute concentration.
- \(M_w\) is the weight-averaged solute molar mass.
- \(A_2\) is the second virial coefficient in the virial expansion of the osmotic pressure.
- \(K^*\) is the constant \(4\pi^2(dn/dc)^2 n_0^2 / N_a \lambda_0^4\).
- \(N_a\) is Avogadro's number. This number always appears when concentration is measured in g/mL and molar mass in g/mol.
- \(P(\theta)\) describes the angular dependence of the scattered light, and can be related to the rms radius.
● $n_0$ is the index of refraction of the solvent
● $\lambda_0$ is the vacuum wavelength of the laser

Accounting for error in light scattering data acquisition, species with calculated molecular weights within 13% of the expected target molecular weight for each design were considered to be forming the anticipated oligomeric state.

**Small-angle X-ray Scattering.** Designed proteins that predominantly formed the target oligomeric species were re-expressed and purified for low-resolution structure determination while in solution by small-angle X-ray scattering (SAXS). A purified elution sample and concentrated sample of each protein were sent for data collection at the SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California\(^8\). A beam exposure time of between 0.5-2.0 seconds was used to obtain diffraction data, which we represent in plots of log intensity ($I$) vs. $q$.

where:

● $q = (4\pi \sin \theta)/\lambda$
● $2\theta$ is angle of diffraction from detector origin
● $\lambda$ is wavelength of the incident X-ray beam

Experimental diffraction data was then analyzed with the java-based application, Scatter. Minimum $q$ values ($q_{\text{min}}$) and experimental radii of gyration ($R_g$) were determined by Guinier analysis\(^4\). Data resolution, reflected by maximum $q$ value ($q_{\text{max}}$), was determined by a characteristic asymptote in signal intensity described by Porod’s Law\(^9\). Refined data sets and corresponding designed model .pdb files were input to the FoXS web server to compute the agreement (evaluated as X) between the experimental and model-computed profiles\(^10\).

**Generation of extension ensemble and determination of SAXS-suggested model.** A set of designed homooligomers, one each of C2 and C4 symmetry, that had been structurally validated by X-ray diffraction crystallography and/or SAXS were selected as candidates for extension. Because the repeating units of the initial scaffolds were not perfectly superimposable, each unique repeat unit (aside from N- and C- capping repeats) was propagated to generate several models with two additional repeat units (three for C2 oligomer, two for C4 oligomer). 100 trajectories of a Rosetta protocol that previously showed to sample the local energy landscape\(^11\) was then performed on each extended model. The total extension set was then input to FoXS with an experimentally-
obtained profile to determine an ensemble of models that agreed within a threshold to the data.

**Crystallography, data acquisition, structure determination and refinement.**
Selected designs were expressed as above and purified by IMAC and SEC on a Superdex 200 10/300 GL gel filtration column using a buffer containing 25 mM Tris pH 8.0 and 50 mM NaCl. Fractions corresponding to the designed oligomers were combined and concentrated for screening.

Crystallization trials for ank3C2_1 were performed at 16.5 mg/ml. The protein crystallized readily in a variety of conditions and optimization was performed using 100 mM Tris pH 8.5, 200 mM magnesium chloride and 30% (v/v) PEG 400. Initial crystallization for 1na0C3_3 trials were performed at 15 mg/ml and produced crystals in 2.4 M sodium malonate pH 7.0 that did not yield a diffraction pattern. Upon concentration crystals that diffracted up to 2.1 Angstroms grew in 2 months.

Crystallization trials for ank1C4_2 were performed at 12 mg/ml and pyramidal crystals were observed within 2 weeks in 100 mM sodium acetate pH 4.6 and 2.0 M ammonium sulfate. Diffraction data were collected at the Advanced Photon Source at Argonne National Laboratory in Lemont, Illinois. Data reduction was carried out using XDS/SCALEDS\(^\text{12}\). Molecular Replacement was performed in the program PHASER\(^\text{13}\) using the design models as search models. Solutions were refined using the program PHENIX\(^\text{14}\) or BUSTER\(^\text{15}\). MR solutions were initially subjected to rigid body refinement and subsequently coordinate refinement. Individual atomic displacement parameter (ADP) refinement and automated water picking were also performed. Refinement protocols were run iteratively while the quality of the model was assessed by the R/R*-free values. Finally, alternating cycles of refinement and model building in COOT were performed using the using the 2mFo-DFc map to obtain the final coordinates\(^\text{16}\).

HR00C3_2 and ank1C2_1 were dialyzed against 25 mM Tris buffer pH 8.0 and 150 mM NaCl. The final concentration of HR00C3_2 and ank1C2_1 used for crystallization trials were 12 mg ml\(^{-1}\). The HR00C3_2 and ank1C2_1 protein were screened with a Phoenix Robot (Art Robbins Instruments) using the following crystallization screens: Crystal Screen, Natrix, PEG/Ion, Index and PEGRx (Hampton Research, Aliso Viejo, CA) and Berkeley Screen (Lawrence Berkeley National Laboratory). Crystals of HR00C3_2 and ank1C2_1 were found in Berkeley Screen conditions. HR00C3_2 was found in condition of 0.3 M Sodium Citrate, 0.1 M Hepes pH
7.5 and 15 % PEG 3,350 and ank1C2_1 was found in 0.4 M Sodium Chloride, 0.1 M Tris-HCl pH 8.5 and 30 % PEG 3,350. HR00C3_2 and ank1C2_1 crystals were obtained after 4 days by the sitting-drop vapor-diffusion method with the drops consisting of a mixture of 0.2 µl of protein solution and 0.2 µl of reservoir solution. Crystallization trials for ank4C4 were performed with a stock protein concentration of 15 mg/ml with three sample to condition ratios in the following crystallization screens: PEG/Ion, Index (Hampton Research, Aliso Viejo, CA), Morpheus (Molecular Dimensions). Hanging-drop optimization was performed with an evenly distributed pH and concentration gradient, and the protein produced crystals within 3 days in a mixture of 1 µl protein solution and 1 µl reservoir solution of 2.1 M DL-Malic Acid pH 7.0. Diffraction data were collected at the Advanced Light Source (at Beamline 8.2.1) at Lawrence Berkeley National Laboratory in Berkeley, California. Integration, scaling and merging of the X-ray diffraction data were carried out with the HKL2000 package\textsuperscript{17} for HR00C3_2 and ank1C2_1 while iMosflm/Scala\textsuperscript{18} was used for ank4C4. An analysis of the intensity statistics carried out on HR00C3_2 by Phenix.xtriage program indicated that the data was merohedrally twinned with twin law (-h, -k, l) with an estimated twin fraction of 46%. Molecular replacement was carried out using PHASER\textsuperscript{13} in PHENIX suite\textsuperscript{14} (using using a monomer predicted by Rosetta ab initio structure prediction as the initial search model. Refinement was carried out with phenix.refine\textsuperscript{14}, using a twin-based target for HR00C3_2 and a maximum likelihood target for ank1C2_1 and ank4C4. Reciprocal space refinement was complemented by rounds of manual model adjustment in COOT\textsuperscript{16}. Root-mean-square deviation differences from ideal geometries for bond lengths, angles and dihedrals were calculated with Phenix\textsuperscript{14}. The overall stereochemical quality of all final models was assessed using the program MOLPROBITYPRO\textsuperscript{19}. 
### Data Acquisition

|                         | ank3C2_1 (PDB ID 5HRY) | 1na0C3_3 (PDB ID 5HRZ) | ank1C4_2 (PDB ID 5HS0) |
|-------------------------|-------------------------|-------------------------|-------------------------|
| **Space group**         | P2₁2₁2₁                 | R32                     | P6₅22                   |
| **Cell dimensions**     |                         |                         |                         |
| a, b, c (Å)             | 106.3, 106.2, 106.6    | 83.6, 83.6, 141.9       | 110.5, 110.5, 182.8     |
| α, β, γ (°)             | 90.0, 90.0, 90.0        | 90.0, 90.0, 120.0       | 90.0, 90.0, 120.0       |
| **Resolution (Å)**      | 74.24 – 2.00 (2.07-2.00) | 64.48 – 2.15 (2.22-2.15) | 84.75 – 2.40 (2.49-2.40) |
| **R_{merge}**           | 6.2 (46.0)              | 6.7 (67.6)              | 7.5 (67.7)              |
| **CC_{1/2}**            | 0.999 (0.867)           | 0.999 (0.875)           | 0.999 (0.885)           |
| **<I/σ_I>**             | 11.9 (2.2)              | 20.9 (3.3)              | 17.6 (3.0)              |
| **Completeness (%)**    | 94.0 (95.0)             | 99.7 (97.8)             | 99.8 (100)              |
| **Multiplicity**        | 3.9 (3.7)               | 9.6 (9.2)               | 9.4 (10.0)              |
| **Wilson B-factor (Å²)** | 34.1                   | 45.9                    | 54.1                    |

### Refinement

|                         |                        |                        |                        |
|-------------------------|-------------------------|-------------------------|-------------------------|
| **Resolution range (Å)** | 75.24 - 2.00           | 64.48 - 2.15           | 84.75 - 2.40           |
| **No. of reflections**  | 77065                   | 10640                   | 264243                 |
| **R_{work} (%) / R_{free} (%)** | 0.21 / 0.25       | 0.18 / 0.21            | 0.18 / 0.21            |
| **Average B-factors (Å²)** |                        |                        |                        |
| Protein                 | 53.8                    | 53.8                    | 48.3                   |
| Water                   | 28.2                    | 55.9                    | 66.5                   |
| **R.m.s.d. deviations** |                        |                        |                        |
| Bond length (Å)         | 0.010                   | 0.010                   | 0.004                  |
| Bond angles (°)         | 1.15                    | 0.94                    | 0.76                   |
| **Ramachandran favored (%)** | 99.7                 | 99.2                    | 98.1                   |
| **Ramachandran outliers (%)** | 0.0                | 0.0                     | 0.0                    |

**Supplementary Table 5.** Data collection and refinement statistics for ank3C2_1, 1na0C3_3, and ank1C4_2. Statistics for the highest-resolution shell are shown in parentheses.
# Data Acquisition

|                      | ank1C2_1 (PDB ID 5KBA) | HR00C3_2 (PDB ID 5K7V) | ank4C4 (PDB ID 5KWD) |
|----------------------|------------------------|------------------------|----------------------|
| **Space group**      | C 1 2 1                | P 3 2 1                | P2;2;2;1             |
| **Cell dimensions**  |                        |                        |                      |
| a, b, c (Å)          | 93.37, 48.82, 139.27   | 159.00, 159.00, 94.98  | 77.7, 89.47, 99.80   |
| α, β, γ (°)          | 90.0, 90.0, 98.9       | 90.0, 90.0, 120.0      | 90.0, 90.0, 90.0     |
| **Resolution (Å)**   | 19.91 - 2.601          | 45.9 - 3.166           | 66.4-2.38            |
|                     | (2.694 - 2.601)        | (3.28 - 3.166)         |                      |
| **R<sub>merge</sub>**| 0.139 (0.5507)         | 0.1486 (0.8645)        | 0.141 (0.643)        |
| **CC<sub>1/2</sub>**| 0.994 (0.846)          | 0.997 (0.583)          | .886(0.532)          |
| **<i>/σ<sub>i</sub>**| 14.52 (2.61)           | 13.32 (2.41)           | 3.0 (3.0)            |
| **Completeness (%)** | 98 (97)                | 98 (85)                | 97 (92)              |
| **Multiplicty**      | 7.2 (5.9)              | 9.8 (7.0)              | 2.0 (2.0)            |
| **Wilson B-factor (Å<sup>2</sup>)** | 41.38 | 38.2 | |

# Refinement

|                      |                      |                      |                      |
|----------------------|----------------------|----------------------|----------------------|
| **Resolution range (Å)** | 19.90 - 2.60       | 47.49 - 3.16          | 66.40-2.75           |
| **No. of reflections** | 18970 (1564)        | 23539 (1994)          | 19498 (1248)         |
| **R<sub>work</sub> (%) / R<sub>free</sub> (%)** | 0.20 / 0.24        | 0.20 / 0.22           | 0.23/0.28            |
| **Average B-factors (Å<sup>2</sup>)** | 59.51 | 80.67 | 50.5 |
|                      | 59.96 | 80.89 | 50.6 |
|                      | 36.93 | 46.96 | 40.0 |
| **R.m.s.d. deviations** |                      |                      |                      |
| Bond length (Å)      | 0.003              | 0.003                | 0.003                |
| Bond angles (°)      | 0.738              | 0.702                | 0.52                 |
| **Ramachandran favored (%)** | 95.4 | 99.5 | 97.1 |
| **Ramachandran outliers (%)** | 0.5 | 0.2 | 0.3 |

**Supplementary Table 6.** Data collection and refinement statistics for ank1C2_1, HR00C3_2 and ank4C4. Statistics for the highest-resolution shell are shown in parentheses.
Supplementary References

1. Chenna, R. et al. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **31**, 3497-3500, doi:10.1093/nar/gkg500 (2003).

2. Arnout R. D. Voet, H. N., Christine Addy, David Simoncini, Daiki Terada, Satoru Unzai, Sam-Yong Park, Kam Y. J. Zhang, and Jeremy R. H. Tame. Computational design of a self-assembling symmetrical β-propeller protein. *PNAS* **111**, 15102-15107 (2014).

3. Leaver-Fay, A. et al. in *Methods in Enzymology*, Vol 487: Computer Methods, Pt C Methods in Enzymology (eds M. L. Johnson & L. Brand) 545-574 (Elsevier Academic Press Inc, 2011).

4. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research* **33**, W363-W367, doi:10.1093/nar/gki481 (2005).

5. Gray, J. J. & Baker, D. Protein-protein docking predictions with RosettaDock. *Biophysical Journal* **86**, 306A-306A (2004).

6. Zimm, B. H. Apparatus and Methods for Measurement and interpretation of the Angular Variation of Light Scattering. *Journal of Chemical Physics* **16**, 1099-1116, doi:10.1063/1.1746740 (1948).

7. Wyatt, P. J. LIGHT-SCATTERING AND THE ABSOLUTE CHARACTERIZATION OF MACROMOLECULES. *Analytica Chimica Acta* **272**, 1-40, doi:10.1016/0003-2670(93)80373-s (1993).

8. Dyer, K. N. et al. High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. *Structural Genomics: General Applications* **1091**, 245-258, doi:10.1007/978-1-62703-691-7_18 (2014).

9. Sinha, S. K., Sirota, E. B., Garoff, S. & Stanley, H. B. X-RAY AND NEUTRON-SCATTERING FROM ROUGH SURFACES. *Physical Review B* **38**, 2297-2311, doi:10.1103/PhysRevB.38.2297 (1988).

10. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research* **38**, W540-W544, doi:10.1093/nar/gkq461 (2010).

11. Tyka, M. D. et al. Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *Journal of Molecular Biology* **405**, 607-618, doi:10.1016/j.jmb.2010.11.008 (2011).

12. Kabsch, W. XDS. *Acta Crystallographica Section D-Biological Crystallography* **66**, 125-132, doi:10.1107/s0907444909047337 (2010).

13. McCoy, A. J. et al. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674, doi:10.1107/s0021889807021206 (2007).

14. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**, 213-221, doi:10.1107/s0907444909052925 (2010).

15. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallographica Section D-Biological Crystallography* **68**, 368-380, doi:10.1107/s0907444911056058 (2012).

16. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2126-2132, doi:10.1107/s0907444904019158 (2004).

17. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276**, 307-326, doi:10.1016/s0076-6879(97)76066-x (1997).

18. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.
Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**, W375-W383, doi:10.1093/nar/gkm216 (2007).