Structure 17

Supplemental Data

Structural Determinants

of Transmembrane Helical Proteins

Susan E. Harrington and Nir Ben-Tal

S1. Rigid body geometry background

Please see Figure S1 for an illustration of the way in which approximate knowledge of the positions of three given noncollinear points on a rigid body (given in terms of three associated regions of space where the points must lie) allows us to build an ensemble of positions of the body consistent with that information.

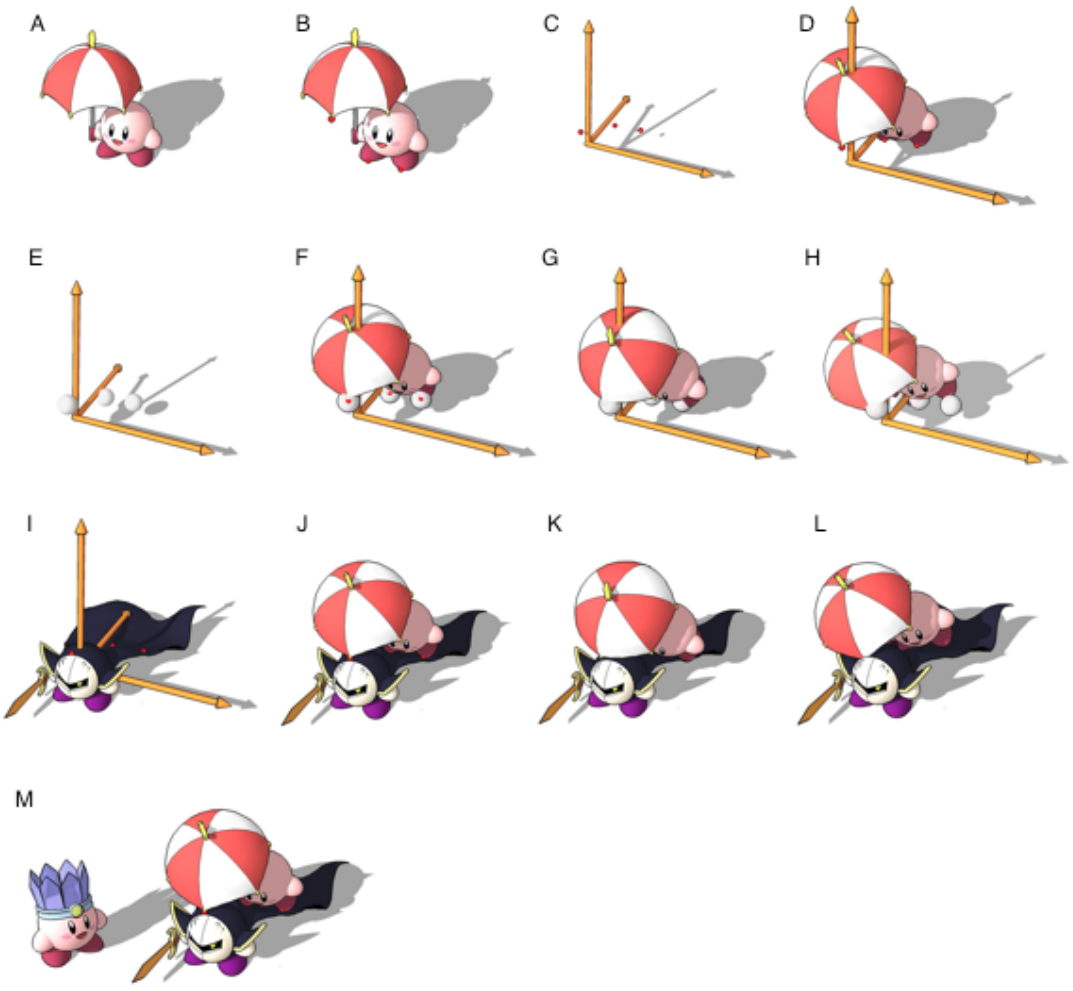

Figure S1. Rigid body positioning

Each row should be read left-to-right.

First row of panels: (A) The rigid body to be positioned. (B) The rigid body with three noncollinear points marked in red: both tips of his feet and the point of the front rib of his umbrella. (C) The desired position of those three points in space. (D) The rigid body position given by those three points.

Second row of panels: Here we do not know precisely where the three marked points on the rigid body are in space, but rather regions where the marked points must lie. By placing a grid on each region we obtain an ensemble of positions of the rigid body with

the marked points in their appropriate regions. Note that if there were any additional marked points on the body that needed to lie in given regions in space, we could check our ensemble to make sure those points did lie within those regions. (E) The regions where the three marked points must lie are shown as spheres. By placing a grid on each region, we obtain an ensemble of positions. (F) One position in space of the body with the marked points in the appropriate regions. (G) Another position with the marked points in the appropriate regions. (H) Yet another position.
Third row of panels: The same procedure can be used when the regions are given by interactions with another rigid body. (I) The three marked points in red on the caped rigid body. (J) The first rigid body with the marked points positioned relative to the caped body. (K) If we know only regions where the marked points on the first body must lie we obtain an ensemble of positions; compare these to those in the second row. (L) Yet another position.
Fourth row of panels: (M) The procedure can be repeated with a new third body, any conformation from the ensemble of the first two bodies, and a given set of at least three interactions between them.
(Kirby figures: nico_sketcher, Kirby Update 3, Google 3D Warehouse, accessed July 12, 2008)

Grid point restrictions
Recall that we place grids over the three associated regions to specify the precise positions of the three given noncollinear points. These three positions then determine the position of the body in space. However, the rigid body imposes conditions on the grid points because the distance between any two grid points must match the distance between the associated points on the body. We do the following to make these match. We place the first point on the body on a grid point position in the first region. For the second point, the dimension of its associated grid drops by one because the distance between the second grid point and the first grid point must match the distance between the two points on the body. This is done by keeping the first two coordinates of each point of the grid and selecting the third (by solving the associated quadratic) so that the distance matches correctly and the resulting grid point is still inside the correct region. Likewise, the dimension of the associated third grid drops again once the first two grid points are selected (since two distances need to be matched). In this case, we keep one coordinate of the original grid point and we have two simultaneous quadratics to solve to match the two distances and obtain the two remaining coordinates. Once this is done, the resulting grid points are checked to make sure they lie within the original region.

S2. Assembly order
The helices are sequentially numbered including chain identifier, e.g., 1-A is the first membrane helix in chain A. The parentheses indicate the order of assembly, decomposition of the determining set of the interactions of the five types, and pieces used. For 1ORS, (((1-C 2-C) 3-a-C) (3-b-C 4-C)) means that 1-C and 2-C were assembled, then the pieces (1-C 2-C) and 3-a-C, then 3-b-C and 4-C, and finally ((1-C 2-C) 3-a-C) and (3-b-C 4-C) to complete the structure.

Table S2.

Protein	Type	Assembly Order
1ORS	voltage sensor	(((1-C 2-C) 3-a-C) (3-b-C 4-C))
1AFO	glycophorin	(1-A 1-B)
3B9W	Rh protein, poss. ammonia channel	((((((1-A (2-A 3-A)) 4-A) 5-A) 6-A) ((7-A (8-A 9-A)) 10-A)) 11-A)
2BS2	fumarate reductase	(((1-C 2-C) (3-C 4-C)) 5-C)
2OAR	mechanosensitive channel, MscL	((1-A 2-A) (1-B 2-B)), sub-dimer
2Z73	rhodopsin, GPCR	(((((1-A 2-A) 3-A) 4-A) (6-A 7-A)) 5-A)
2RH1	β2-andrenergic receptor, GPCR	(((((1-A 2-A) 3-A) 4-A) 5-A) (6-A 7-A))
1C3W	bacteriorhodopsin	((((1-A 2-A) 3-A) ((4-A 5-A) 6-A)) 7-A)
2QTS	acid-sensing ion channel	((1-A 2-A) ((1-B 2-B) (1-C 2-C)))
2H88	succinate oxidoreductase	((1-C ((1-D 2-D) 3-D)) (2-C 3-C))
2UUH	leukotriene LTC4 synthase	(1-A ((2-A 3-A) 4-A))
1BL8	potassium channel	((1-A 2-A) 3-A) ((1-B 2-B) 3-B)) sub-dimer, symmetrized
2BL2	rotor of V-type ATPase	((((1-A 2-A) 3-A) 4-A) (((1-B 2-B) 3-B) 4-B)) sub-dimer, symmetrized
1OKC	mitochondrial ADP/ATP carrier	((((1-A 2-A) 3-A) (4-A 5-A)) 6-A)
2B6O	aquaporin	((((((1-A 2-A) (3-A 4-A)) 6-A) 8-A) 5-A) 7-A)

S3. Statistics of the determining sets of interactions of the five types

For all but the last column in Table ST3, we see the fraction of a type of interaction in the determining set. The Hbond column gives the fraction of determining set interactions that are hydrogen bonds, the Arom column gives the fraction of aromatic interactions, and the Salt bridge column gives the fraction of salt bridges. These three kinds of interactions make up the polar group. The next two columns give the fractions of small (Gly/Ala/Ser) knob-in-hole interactions, and Ile/Val/Leu knob-in-hole interactions in Ile/Val/Leu patches as defined in Experimental Procedures. The packing group consists of these two kinds of interactions. Together, the polar and packing group percentages add to 100% outside of rounding errors. The last column lists the fraction of helix residues in the determining set of interactions.

Table S3.

		Polar			Packing		
Protein	**Type**	**H-bond**	**Arom**	**Salt bridge**	**Small**	**I,V,L patch**	**Helix res in det. set**
1ORS	voltage sensor	40%	8%	4%	20%	28%	20%
1AFO	glycophorin	0%	0%	0%	25%	75%	11%
3B9W	Rh protein, poss. ammonia channel	24%	11%	0%	53%	12%	25%
2BS2	fumarate reductase	48%	11%	0%	19%	22%	19%
2OAR	mechanosensitive channel, MscL	27%	0%	0%	36%	36%	18%
2Z73	rhodopsin	26%	37%	0%	32%	5%	16%
2RH1	β2-adrenergic receptor	36%	8%	0%	28%	28%	18%
1C3W	bacteriorhodopsin	24%	35%	0%	22%	19%	20%
2QTS	acid-sensing ion channel	26%	41%	7%	19%	7%	17%
2H88	succinate oxidoreductase	61%	9%	0%	18%	12%	21%
2UUH	leukotriene LTC4 synthase	18%	25%	0%	36%	21%	20%
1BL8	potassium channel	25%	8%	0%	47%	20%	20%
2BL2	rotor of V-type ATPase	30%	6%	0%	49%	15%	41%
1OKC	mitochondrial ADP/ATP carrier	42%	6%	6%	29%	16%	20%
2B6O	aquaporin	19%	0%	0%	50%	31%	28%

S4. Determining sets of interactions of the five types in homologous proteins

Comparison of the determining sets of the first three helices of bacteriorhodopsin (1C3W) and halorhodopsin (1E12), and the determining sets of the first three helices of sensory rhodopsins II (1XIO and 1H68).

The notation for the donor/acceptor pair in polar interactions is donresnum-donchain-donatomtype-accresnum-accchain-accatomtype. Donor hydrogens, which are added during our analysis, are not listed here since they are not present in the pdb files. Here, in most cases, there is only one possible donor hydrogen.

The notation for the packing interactions is knobresnum-knobchain-holesresnum-holechain-holeposition. A hole is defined by the residues j, j+3, j+4, j+7, and the holeresnum = j, by definition. The hole position is ‘l’, ‘m’, or ‘u’, as defined in Methods. It tells us where the knob is in the hole: closest to j (‘1’), closest to j+3 (‘m’) or closest to j+7 (‘u’). The knob-in-hole interactions listed can be of either the small or I/V/L patch variety.

Table S4.

Bacteriorhodopsin, 1C3W	**Determining set interactions**
between 1-A 2-A	30-A-NZ-43-A-OH, 27-A-CD2-46-A-O, 54-A-CD1-17-A-O, 57-A-CD2-13-A-O, 61-A-10-A-m
between 2-A 3-A	46-A-OG1-96-A-OD2, 42-A-CD1-96-A-OD1, 42-A-CE1-96-A-O, 81-A-56-A-l, 49-A-89-A-l, 92-A-45-A-l
Halorhodopsin, 1E12	
between 1-A 2-A	36-A-ND2-77-A-OH, 63-A-NE1-47-A-O, 47-A-CD2-66-A-O, 33-A-77-A-l, 81-A-30-A-m, 44-A-67-A-l, 40-A-70-A-l, 74-A-33-A-u
between 2-A 3-A	108-A-NH1-77-A-OH, 65-A-118-A-l, 76-A-104-A-u, 115-A-69-A-l, 62-A-122-A-l, 125-A-58-A-m
Common or closely substituted	**1C3W/1E12**
between 1-A 2-A	27-A-CD2-46-A-O/47-A-CD2-66-A-O, 61-A-10-A-m/81-A-30-A-m
between 2-A 3-A	none

Sensory rhodopsin II, 1XIO	**Determining set interactions**
between 1-A and 2-A	11-A-OH-47-A-OG, 22-A-CD2-40-A-O, 34-A-NE2-26-A-O, 19-A-41-A-l, 48-A-8-A-u, 55-A-5-A-l
between 2-A and 3-A	46-A-NE1-74-A-O, 47-A-OG-75-A-OD2, 51-A-CE1-75-A-OD2, 89-A-CZ3-35-A-O, 39-A-82-A-l, 50-A-68-A-u, 71-A-50-A-l
Sensory rhodopsin II, 1H68	
between 1-A and 2-A	44-A-OG-19-A-OG1,19-A-OG1-44-A-OG, 37-A-CD2-22-A-O, 44-A-15-A-l, 55-A-5-A-l, 48-A-12-A-l, 12-A-48-A-l
between 2-A and 3-A	86-A-CE1-36-A-O, 82-A-39-A-l, 50-A-68-A-u, 43-A-79-A-l
Common or closely substituted	**1XIO/1H68**
between 1-A and 2-A	55-A-5-A-l/55-A-5-A-l
between 2-A and 3-A	50-A-68-A-u/50-A-68-A-u

S5. Helices
For each protein, the residue intervals of the helices are shown in sequence order.

Table S5.

Protein	Helices
1ORS	(25-C, 47-C); (53-C, 79-C); (86-C, 97-C); (100-C, 112-C); (116-C, 148-C)
1AFO	(72-A, 98-A); (72-B, 98-B)
3B9W	(10-A, 37-A); (42-A, 66-A); (77-A, 97-A); (104-A, 129-A); (145-A, 160-A); (174-A, 197-A); (205-A, 229-A); (236-A, 250-A); (260-A, 285-A); (295-A, 315-A); (319-A, 345-A)
2BS2	(21-C, 53-C); (76-C, 99-C); (122-C, 149-C); (169-C, 194-C); (203-C, 237-C)
2OAR	(13-A, 46-A); (69-A, 99-A) (the same for chains B, C, D, and E)
2Z73	(31-A, 61-A); (70-A, 96-A); (105-A, 135-A); (149-A, 172-A); (195-A, 235-A); (247-A, 286-A); (294-A, 317-A)
2RH1	(30-A, 60-A); (67-A 96-A); (103-A, 135-A); (147-A, 171-A); (198-A, 228-A); (268-A, 298-A); (307-A, 328-A)
1C3W	(9-A, 31-A); (38-A, 62-A); (80-A, 100-A); (105-A, 127-A); (131-A, 155-A); (165-A, 191-A); (201-A, 224-A)
2QTS	(44-A, 70-A); (427-A, 455-A); (44-B, 68-B); (426-B, 460-B); (41-C, 68-C); (426-C, 453-C)
2H88	(41-C, 62-C); (82-C, 109-C); (116-C, 139-C); (5-D, 29-D); (33-D, 57-D); (62-D, 88-D)
2UUH	(5-A, 32-A); (44-A, 72-A); (76-A, 99-A); (101-A, 142-A)
1BL8	(23-A, 51-A); (62-A, 74-A); (86-A, 114-A) (the same for chains B, C, and D)
2BL2	(11-A, 45-A); (52-A, 78-A); (86-A, 122-A); (128-A, 156-A) (the same for chains B, C,..., J)
1OKC	(3-A, 38-A); (73-A, 97-A); (109-A, 142-A); (176-A, 199-A); (210-A, 238-A); (272-A, 292-A)
2B6O	(8-A, 31-A); (40-A, 59-A); (69-A, 78-A); (82-A, 107-A); (126-A, 148-A); (159-A, 175-A); (185-A, 194-A); (203-A, 220-A)

S6. Determining sets of interactions of the five types
Below we list the determining sets used in assembly. Each line describes how two pieces were assembled using the listed subset of interactions of the five types. The helices are numbered from 0, so that 0-A is the first helix of chain A. The interaction notation is described in S3. A donor/acceptor pair followed by –SB indicates a salt bridge. The interactions are inferred geometrically and should not be regarded as definitive since there are invariably ambiguous cases. In particular, we tried to avoid protonation ambiguities and some bonds were omitted because of them.

1ORS
pieceA 0-C pieceB 1-C assembly (0-C 1-C) 2-C 3-C 4-C subnetwork 59-C-OH-45-C-OE1 41-C-58-C-l 34-C-65-C-l 58-C-41-C-l 65-C-34-C-l

pieceA 0-Cj1-C pieceB 2-C assembly ((0-C 1-C) 2-C) 3-C 4-C subnetwork 73-C-OH-90-C-O 73-C-CE2-90-C-OG1 76-C-NE-86-C-OH 76-C-NH2-93-C-OE1 76-C-NH2-93-C-OE2 73-C-CD2-93-C-OE2
pieceA 3-C pieceB 4-C assembly ((0-C 1-C) 2-C) (3-C 4-C) subnetwork 123-C-NE-107-C-OE2-SB 110-C-115-C-m 122-C-103-C-l 103-C-122-C-m 119-C-107-C-l
pieceA 0-Cj1-Cj2-C pieceB 3-Cj4-C assembly (((0-C 1-C) 2-C) (3-C 4-C)) subnetwork 133-C-NH1-38-C-O 133-C-NH1-62-C-OD1 133-C-NH2-62-C-OD1 133-C-NH2-62-C-OD2 132-C-OG-92-C-O 140-C-32-C-l 35-C-133-C-m 132-C-89-C-u 96-C-129-C-l

1AFO
pieceA 0-A pieceB 0-B assembly (0-A 0-B) subnetwork 80-A-76-B-m 76-B-72-A-m 80-B-76-A-m 79-A-76-B-l 79-B-76-A-l 76-A-72-B-m 83-B-80-A-l 83-A-80-B-l

3B9W
pieceA 1-A pieceB 2-A assembly 0-A (1-A 2-A) 3-A 4-A 5-A 6-A 7-A 8-A 9-A 10-A subnetwork 60-A-CE2-83-A-O 57-A-81-A-u 84-A-53-A-u 52-A-88-A-m 88-A-49-A-u
pieceA 0-A pieceB 1-Aj2-A assembly (0-A (1-A 2-A)) 3-A 4-A 5-A 6-A 7-A 8-A 9-A 10-A subnetwork 32-A-CA-49-A-OG1 16-A-NE2-79-A-OD1 32-A-42-A-u 45-A-32-A-l 27-A-85-A-l
pieceA 0-Aj1-Aj2-A pieceB 3-A assembly ((0-A (1-A 2-A)) 3-A) 4-A 5-A 6-A 7-A 8-A 9-A 10-A subnetwork 60-A-OH-122-A-OE1 83-A-CD1-122-A-OE2 59-A-108-A-u 115-A-56-A-l 111-A-48-A-u 108-A-48-A-u 91-A-107-A-u
pieceA 0-Aj1-Aj2-Aj3-A pieceB 4-A assembly (((0-A (1-A 2-A)) 3-A) 4-A) 5-A 6-A 7-A 8-A 9-A 10-A subnetwork 149-A-90-A-l 153-A-90-A-u 97-A-150-A-m 94-A-149-A-l
pieceA 0-Aj1-Aj2-Aj3-Aj4-A pieceB 5-A assembly ((((0-A (1-A 2-A)) 3-A) 4-A) 5-A) 6-A 7-A 8-A 9-A 10-A subnetwork 30-A-CA-183-A-O 190-A-CE1-22-A-OD1 190-A-CD1-22-A-OD1 22-A-ND2-193-A-OG 183-A-26-A-u 30-A-183-A-l 186-A-26-A-m
pieceA 7-A pieceB 8-A assembly ((((0-A (1-A 2-A)) 3-A) 4-A) 5-A) 6-A (7-A 8-A) 9-A 10-A subnetwork 249-A-262-A-u 245-A-266-A-m 242-A-270-A-l 265-A-246-A-m 269-A-242-A-u
pieceA 6-A pieceB 7-Aj8-A assembly ((((0-A (1-A 2-A)) 3-A) 4-A) 5-A) (6-A (7-A 8-A)) 9-A 10-A subnetwork 226-A-OG-238-A-OD2 266-A-CA-214-A-O 263-A-CD1-213-A-O 246-A-211-A-m 243-A-215-A-m 266-A-210-A-u 270-A-218-A-l 262-A-206-A-u 225-A-267-A-u 221-A-263-A-u 274-A-218-A-u 218-A-263-A-u 214-A-259-A-u 217-A-263-A-l
pieceA 6-Aj7-Aj8-A pieceB 9-A assembly ((((0-A (1-A 2-A)) 3-A) 4-A) 5-A) ((6-A (7-A 8-A)) 9-A) 10-A subnetwork 273-A-OG-298-A-O 276-A-CA-298-A-OD1 298-A-ND2-281-A-OE1 299-A-241-A-l
pieceA 0-Aj1-Aj2-Aj3-Aj4-Aj5-A pieceB 6-Aj7-Aj8-Aj9-A assembly (((((0-A (1-A 2-A)) 3-A) 4-A) 5-A) ((6-A (7-A 8-A)) 9-A)) 10-A subnetwork 146-A-ND1-300-A-ND1 300-A-ND1-146-A-ND1 151-A-CE1-297-A-ND1 188-A-NE1-219-A-OG1 184-A-OG-240-A-O 296-A-93-A-l 147-A-297-A-u 301-A-143-A-u 150-A-293-A-m 240-A-177-A-m
pieceA 0-Aj1-Aj2-Aj3-Aj4-Aj5-Aj6-Aj7-Aj8-Aj9-A pieceB 10-A assembly ((((((0-A (1-A 2-A)) 3-A) 4-A) 5-A) ((6-A (7-A 8-A)) 9-A)) 10-A) subnetwork 152-A-CD1-333-A-O 329-A-OG1-121-A-OD1 322-A-NE2-308-A-O 121-A-ND2-329-A-OG1 152-A-CE1-336-A-O 341-A-152-A-u 160-A-341-A-l 337-A-152-A-l 333-A-144-A-u 155-A-334-A-l 159-A-334-A-u

2BS2
pieceA 0-C pieceB 1-C assembly (0-C 1-C) 2-C 3-C 4-C subnetwork 33-C-OG1-93-C-ND1 93-C-ND1-33-C-OG1 99-C-NH2-30-C-OE1 37-C-86-C-m
pieceA 2-C pieceB 3-C assembly (0-C 1-C) (2-C 3-C) 4-C subnetwork 172-C-CE1-143-C-ND1 182-C-ND1-129-C-O 143-C-ND1-172-C-OH 189-C-NE-129-C-OE1 189-C-NH2-129-C-OE1 132-C-OG1-182-C-ND1 178-C-132-C-l 175-C-136-C-m 139-C-171-C-l 171-C-135-C-u 185-C-125-C-l

pieceA 0-Cj1-C pieceB 2-Cj3-C assembly ((0-C 1-C) (2-C 3-C)) 4-C subnetwork 194-C-NE1-31-C-OG 90-C-CZ-136-C-O 187-C-35-C-l 190-C-31-C-l 83-C-144-C-l 144-C-83-C-m
pieceA 0-Cj1-Cj2-Cj3-C pieceB 4-C assembly (((0-C 1-C) (2-C 3-C)) 4-C) subnetwork 228-C-CE1-169-C-O 184-C-OG-217-C-OG 217-C-OG-184-C-OG 231-C-OH-48-C-O 206-C-NE-188-C-OH 176-C-221-C-l

2OAR (subdimer)
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 0-B 1-B subnetwork 35-A-OG1-78-A-OD1 40-A-OG1-74-A-OG 74-A-OG-40-A-OG1 31-A-78-A-m
pieceA 0-B pieceB 1-B assembly (0-A 1-A) (0-B 1-B) subnetwork 35-B-OG1-78-B-OD1 40-B-OG1-74-B-OG 74-B-OG-40-B-OG1 31-B-78-B-m
pieceA 0-Aj1-A pieceB 0-Bj1-B assembly ((0-A 1-A) (0-B 1-B)) subnetwork 17-A-17-B-l 18-A-20-B-l 21-A-17-B-u 24-B-18-A-l 20-B-14-A-l 14-A-16-B-l 75-B-26-A-u

2Z73
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 3-A 4-A 5-A 6-A subnetwork 93-A-OG1-38-A-O 59-A-CE2-70-A-OD1 81-A-CD1-49-A-O 38-A-89-A-u 77-A-52-A-l 48-A-77-A-u 45-A-81-A-u 56-A-74-A-l 74-A-52-A-u 88-A-38-A-m
pieceA 0-Aj1-A pieceB 2-A assembly ((0-A 1-A) 2-A) 3-A 4-A 5-A 6-A subnetwork 72-A-CE2-128-A-O 118-A-CD2-79-A-OG 83-A-CE1-115-A-O 122-A-OG-80-A-OD2 111-A-CE1-86-A-O 79-A-OG-118-A-O 75-A-ND2-125-A-OG1 111-A-OH-87-A-O 86-A-107-A-u
pieceA 0-Aj1-Aj2-A pieceB 3-A assembly (((0-A 1-A) 2-A) 3-A) 4-A 5-A 6-A subnetwork 113-A-CD1-167-A-O 120-A-CE1-166-A-O 160-A-NE1-75-A-OD1 167-A-113-A-l 117-A-160-A-u
pieceA 5-A pieceB 6-A assembly (((0-A 1-A) 2-A) 3-A) 4-A (5-A 6-A) subnetwork 269-A-NE2-307-A-OG 303-A-CD2-273-A-O 303-A-CE2-272-A-O 304-A-270-A-u
pieceA 0-Aj1-Aj2-Aj3-A pieceB 5-Aj6-A assembly ((((0-A 1-A) 2-A) 3-A) (5-A 6-A)) 4-A subnetwork 52-A-ND2-308-A-O 83-A-CE2-305-A-O 270-A-CZ-119-A-O 305-A-NZ-111-A-OH 309-A-44-A-l 308-A-76-A-u
pieceA 0-Aj1-Aj2-Aj3-Aj5-Aj6-A pieceB 4-A assembly (((((0-A 1-A) 2-A) 3-A) (5-A 6-A)) 4-A) subnetwork 219-A-SG-130-A-OG 162-A-CZ2-207-A-O 205-A-CZ-278-A-O 127-A-212-A-l

2RH1
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 3-A 4-A 5-A 6-A subnetwork 87-A-40-A-m 48-A-80-A-l 83-A-44-A-l 44-A-80-A-u 95-A-33-A-l 80-A-48-A-l 76-A-51-A-l
pieceA 0-Aj1-A pieceB 2-A assembly ((0-A 1-A) 2-A) 3-A 4-A 5-A 6-A subnetwork 109-A-NE1-86-A-O 116-A-SG-78-A-O 71-A-CE2-130-A-OD2 78-A-116-A-l
pieceA 0-Aj1-Aj2-A pieceB 3-A assembly (((0-A 1-A) 2-A) 3-A) 4-A 5-A 6-A subnetwork 158-A-NE1-74-A-OG 158-A-CZ2-74-A-O 161-A-OG-122-A-OE1 119-A-154-A-u 111-A-162-A-u 165-A-111-A-l 161-A-115-A-l
pieceA 0-Aj1-Aj2-Aj3-A pieceB 4-A assembly ((((0-A 1-A) 2-A) 3-A) 4-A) 5-A 6-A subnetwork 125-A-SG-211-A-O 222-A-128-A-m 128-A-211-A-u 135-A-222-A-l 218-A-125-A-m
pieceA 5-A pieceB 6-A assembly ((((0-A 1-A) 2-A) 3-A) 4-A) (5-A 6-A) subnetwork 328-A-NE-274-A-OG1 308-A-OH-293-A-OD1 285-A-SG-318-A-OD1 315-A-282-A-m 325-A-274-A-l
pieceA 0-Aj1-Aj2-Aj3-Aj4-A pieceB 5-Aj6-A assembly (((((0-A 1-A) 2-A) 3-A) 4-A) (5-A 6-A)) subnetwork 51-A-ND2-319-A-O 316-A-OH-113-A-OD2 209-A-OH-290-A-O 316-A-CE2-113-A-OD1 219-A-OH-272-A-O 319-A-OG-79-A-OD1 320-A-43-A-m 319-A-75-A-u

1C3W

pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 3-A 4-A 5-A 6-A subnetwork 30-A-NZ-43-A-OH 27-A-CD2-46-A-O 54-A-CD1-17-A-O 57-A-CD2-13-A-O 61-A-10-A-m
pieceA 0-Aj1-A pieceB 2-A assembly ((0-A 1-A) 2-A) 3-A 4-A 5-A 6-A subnetwork 46-A-OG1-96-A-OD2 42-A-CD1-96-A-OD1 42-A-CE1-96-A-O 81-A-56-A-l 49-A-89-A-l 45-A-92-A-l 92-A-45-A-l
pieceA 3-A pieceB 4-A assembly ((0-A 1-A) 2-A) (3-A 4-A) 5-A 6-A subnetwork 141-A-OG-118-A-O 141-A-114-A-u 114-A-144-A-l 144-A-114-A-l 125-A-130-A-u 148-A-111-A-l 122-A-134-A-u
pieceA 3-Aj4-A pieceB 5-A assembly ((0-A 1-A) 2-A) ((3-A 4-A) 5-A) 6-A subnetwork 182-A-CZ3-115-A-OD2 182-A-CH2-115-A-OD2 138-A-NE1-186-A-O 175-A-NH1-152-A-O
pieceA 0-Aj1-Aj2-A pieceB 3-Aj4-Aj5-A assembly (((0-A 1-A) 2-A) ((3-A 4-A) 5-A)) 6-A subnetwork 90-A-OG1-115-A-OD1 189-A-NE1-83-A-OH 182-A-CH2-90-A-O 94-A-108-A-u
pieceA 0-Aj1-Aj2-Aj3-Aj4-Aj5-A pieceB 6-A assembly ((((0-A 1-A) 2-A) ((3-A 4-A) 5-A)) 6-A) subnetwork 27-A-CZ-217-A-O 219-A-CE1-96-A-OD2 57-A-CE1-212-A-OD2 57-A-OH-212-A-OD2 185-A-CE2-212-A-OD1 185-A-OH-212-A-OD1 219-A-CE2-178-A-OG1 181-A-211-A-m 174-A-219-A-l 211-A-181-A-l

2QTS
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 0-B 1-B 0-C 1-C subnetwork 441-A-CD2-54-A-OG 65-A-NH2-433-A-OD2 47-A-NE1-448-A-OG1 47-A-CZ3-451-A-OE1 47-A-CZ2-447-A-O 51-A-CE2-445-A-OG 434-A-58-A-m 444-A-47-A-u
pieceA 0-B pieceB 1-B assembly (0-A 1-A) (0-B 1-B) 0-C 1-C subnetwork 47-B-NE1-451-B-OE2 65-B-NE-433-B-OD2-SB 51-B-CD2-448-B-OG1 51-B-CZ-451-B-OE2 441-B-CE2-55-B-O
pieceA 0-C pieceB 1-C assembly (0-A 1-A) (0-B 1-B) (0-C 1-C) subnetwork 44-C-NH1-451-C-O 448-C-OG1-47-C-O 441-C-CD2-54-C-OG 65-C-NH2-433-C-OD2 47-C-CZ3-447-C-O
pieceA 0-Bj1-B pieceB 0-Cj1-C assembly (0-A 1-A) ((0-B 1-B) (0-C 1-C)) subnetwork 43-C-NZ-451-B-OE1 65-C-NH1-433-B-OD1-SB 450-B-39-C-m 46-C-443-B-l
pieceA 0-Aj1-A pieceB 0-Bj1-Bj0-Cj1-C assembly ((0-A 1-A) ((0-B 1-B) (0-C 1-C))) subnetwork 47-A-CZ3-450-C-O 47-B-CH2-450-A-O 57-A-439-C-l 57-B-435-A-u 447-C-440-A-l

2H88
pieceA 1-C pieceB 2-C assembly 0-C (1-C 2-C) 0-D 1-D 2-D subnetwork 104-C-NE-122-C-OG 96-C-OG-130-C-O 100-C-NE1-133-C-OG 104-C-NH2-119-C-O 85-C-OH-135-C-O 138-C-85-C-l
pieceA 0-D pieceB 1-D assembly 0-C (1-C 2-C) (0-D 1-D) 2-D subnetwork 17-D-OG-46-D-ND1 53-D-NE2-13-D-OE2 49-D-CZ3-12-D-O 23-D-35-D-m 39-D-20-D-u 42-D-16-D-l 16-D-42-D-m 20-D-39-D-u
pieceA 0-Dj1-D pieceB 2-D assembly 0-C (1-C 2-C) ((0-D 1-D) 2-D) subnetwork 48-D-ND1-76-D-OG 76-D-OG-48-D-ND1 87-D-ND2-36-D-O 80-D-CD1-44-D-OG1 40-D-80-D-m 72-D-48-D-l
pieceA 0-C pieceB 0-Dj1-Dj2-D assembly (0-C ((0-D 1-D) 2-D)) (1-C 2-C) subnetwork 54-C-OG-79-D-OG1 82-D-CA-57-C-OG 43-C-NH1-57-D-OD2 43-C-NH2-57-D-OD2 79-D-OG1-54-C-OG 57-C-OG-79-D-O 82-D-57-C-l 60-C-82-D-l
pieceA 0-Cj0-Dj1-Dj2-D pieceB 1-Cj2-C assembly ((0-C ((0-D 1-D) 2-D)) (1-C 2-C)) subnetwork 45-C-OG1-98-C-ND1 98-C-ND1-45-C-OG1 98-C-ND1-42-C-O 14-D-NE-109-C-OD2 10-D-CZ2-109-C-OD1

2UUH
pieceA 1-A pieceB 2-A assembly 0-A (1-A 2-A) 3-A subnetwork 48-A-NE-97-A-O 90-A-NH1-55-A-OD1 90-A-NH1-56-A-SG 90-A-NH2-55-A-OD1 94-A-CE1-49-A-O 52-A-90-A-u 98-A-45-A-m 79-A-67-A-l

pieceA 1-Aj2-A pieceB 3-A assembly 0-A ((1-A 2-A) 3-A) subnetwork 86-A-CD1-111-A-O 93-A-CD1-104-A-O 70-A-118-A-l 122-A-66-A-l 96-A-100-A-m 89-A-107-A-l 114-A-82-A-l
pieceA 0-A pieceB 1-Aj2-Aj3-A assembly (0-A ((1-A 2-A) 3-A)) subnetwork 68-A-NE1-12-A-OG1 22-A-CE2-56-A-SG 60-A-CE1-15-A-O 68-A-CH2-5-A-O 22-A-CZ-90-A-O 8-A-64-A-u 29-A-42-A-u 67-A-8-A-l 49-A-22-A-u 87-A-11-A-l 11-A-80-A-l 80-A-4-A-m 18-A-87-A-l

1BL8 (subdimer)
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 0-B 1-B 2-B 0-C 1-C 2-C 0-D 1-D 2-D subnetwork 44-A-OG-69-A-OG 69-A-OG-40-A-O 65-A-44-A-l 40-A-66-A-m
pieceA 0-Aj1-A pieceB 2-A assembly ((0-A 1-A) 2-A) 0-B 1-B 2-B 0-C 1-C 2-C 0-D 1-D 2-D subnetwork 25-A-ND1-109-A-O 112-A-OG1-25-A-ND1 39-A-91-A-u 91-A-43-A-l 29-A-102-A-m 50-A-84-A-m 73-A-92-A-u
pieceA 0-B pieceB 1-B assembly ((0-A 1-A) 2-A) (0-B 1-B) 2-B 0-C 1-C 2-C 0-D 1-D 2-D subnetwork 44-B-OG-69-B-OG 69-B-OG-40-B-O 65-B-44-B-l 40-B-66-B-m
pieceA 0-Bj1-B pieceB 2-B assembly ((0-A 1-A) 2-A) ((0-B 1-B) 2-B) 0-C 1-C 2-C 0-D 1-D 2-D subnetwork 25-B-ND1-109-B-O 112-B-OG1-25-B-ND1 91-B-43-B-l 39-B-91-B-u 50-B-84-B-m 29-B-102-B-m 73-B-92-B-u

2BL2 (subdimer)
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 3-A 0-B 1-B 2-B 3-B subnetwork 59-A-NE2-34-A-O 30-A-OG-62-A-O 69-A-CA-26-A-OG 73-A-19-A-l 41-A-52-A-u 66-A-23-A-m 22-A-69-A-l 69-A-19-A-u
pieceA 0-Aj1-A pieceB 2-A assembly ((0-A 1-A) 2-A) 3-A 0-B 1-B 2-B 3-B subnetwork 37-A-OG1-111-A-O 107-A-CA-29-A-O 65-A-NE2-106-A-OG 65-A-NE2-110-A-OE1 33-A-104-A-u 29-A-100-A-u 111-A-33-A-l 104-A-25-A-l 107-A-26-A-m 44-A-118-A-l 40-A-111-A-u 118-A-37-A-m 100-A-18-A-u 89-A-8-A-u 15-A-85-A-u 115-A-33-A-m 18-A-89-A-m
pieceA 0-Aj1-Aj2-A pieceB 3-A assembly (((0-A 1-A) 2-A) 3-A) 0-B 1-B 2-B 3-B subnetwork 146-A-CE1-68-A-O 68-A-OH-139-A-OE2 64-A-OG1-139-A-OE2 110-A-NE2-135-A-O 149-A-OG-95-A-O 145-A-CA-102-A-OG1 110-A-NE2-139-A-OE1 153-A-72-A-m 61-A-132-A-l 95-A-149-A-l 142-A-99-A-u 145-A-95-A-u 135-A-110-A-l 156-A-88-A-l
pieceA 0-B pieceB 1-B assembly (((0-A 1-A) 2-A) 3-A) (0-B 1-B) 2-B 3-B subnetwork 59-B-NE2-34-B-O 69-B-CA-26-B-OG 30-B-OG-62-B-O 69-B-19-B-u 73-B-19-B-l 66-B-23-B-m 22-B-69-B-l 41-B-52-B-u
pieceA 0-Bj1-B pieceB 2-B assembly (((0-A 1-A) 2-A) 3-A) ((0-B 1-B) 2-B) 3-B subnetwork 37-B-OG1-111-B-O 65-B-NE2-106-B-OG 65-B-NE2-110-B-OE1 107-B-26-B-m 15-B-85-B-u 115-B-33-B-m 111-B-33-B-l 29-B-100-B-u 40-B-111-B-u 33-B-104-B-u 44-B-118-B-l 100-B-18-B-u 89-B-11-B-l 18-B-89-B-m 118-B-37-B-m 104-B-25-B-l
pieceA 0-Bj1-Bj2-B pieceB 3-B assembly (((0-A 1-A) 2-A) 3-A) (((0-B 1-B) 2-B) 3-B) subnetwork 149-B-OG-95-B-O 110-B-NE2-135-B-O 110-B-NE2-139-B-OE1 145-B-CA-102-B-OG1 64-B-OG1-139-B-OE2 68-B-OH-139-B-OE2 146-B-CE1-68-B-O 153-B-72-B-m 61-B-132-B-l 145-B-95-B-u 135-B-110-B-l 156-B-88-B-l 113-B-131-B-m 95-B-149-B-l 142-B-99-B-u
pieceA 0-Aj1-Aj2-Aj3-A pieceB 0-Bj1-Bj2-Bj3-B assembly ((((0-A 1-A) 2-A) 3-A) (((0-B 1-B) 2-B) 3-B)) subnetwork 134-A-CE1-45-B-OG1 141-A-OH-63-B-O 63-B-CA-141-A-OH 134-A-CD2-59-B-OE1 59-B-NE2-134-A-O 13-B-87-A-l 109-A-28-B-m 94-A-13-B-m 42-B-113-A-m 116-A-35-B-u 31-B-102-A-m 20-B-94-A-l 35-B-106-A-u 144-A-63-B-u

1OKC
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 3-A 4-A 5-A subnetwork 14-A-CA-80-A-OH 23-A-OG1-73-A-OD1 76-A-ND2-29-A-OE1 80-A-OH-11-A-O 11-A-CZ-81-A-O

pieceA 0-Aj1-A pieceB 2-A assembly ((0-A 1-A) 2-A) 3-A 4-A 5-A subnetwork 79-A-NH1-134-A-OD1 79-A-NH1-134-A-OD2-SB 79-A-NE-134-A-OD1-SB 94-A-OH-125-A-OG1 125-A-OG1-94-A-OH 141-A-26-A-u 133-A-75-A-l 75-A-130-A-u 122-A-87-A-u 118-A-91-A-m
pieceA 3-A pieceB 4-A assembly ((0-A 1-A) 2-A) (3-A 4-A) 5-A subnetwork 227-A-182-A-l 193-A-215-A-u 223-A-186-A-l 189-A-219-A-u
pieceA 0-Aj1-Aj2-A pieceB 3-Aj4-A assembly (((0-A 1-A) 2-A) (3-A 4-A)) 5-A subnetwork 32-A-NZ-231-A-OD2 187-A-NH2-115-A-OD1 187-A-NE-115-A-O 131-A-CE2-179-A-OG 234-A-NH2-134-A-OD2 234-A-NE-131-A-OH 116-A-184-A-l
pieceA 0-Aj1-Aj2-Aj3-Aj4-A pieceB 5-A assembly ((((0-A 1-A) 2-A) (3-A 4-A)) 5-A) subnetwork 272-A-CA-228-A-OH 21-A-283-A-l 272-A-32-A-l 275-A-28-A-l 283-A-18-A-u

2B6O
pieceA 0-A pieceB 1-A assembly (0-A 1-A) 2-A 3-A 4-A 5-A 6-A 7-A subnetwork 51-A-23-A-l 47-A-23-A-u 30-A-40-A-u 27-A-44-A-m
pieceA 2-A pieceB 3-A assembly (0-A 1-A) (2-A 3-A) 4-A 5-A 6-A 7-A subnetwork 72-A-OG1-93-A-OE1 86-A-69-A-u 76-A-79-A-u
pieceA 0-Aj1-A pieceB 2-Aj3-A assembly ((0-A 1-A) (2-A 3-A)) 4-A 5-A 6-A 7-A subnetwork 100-A-CA-21-A-O 96-A-17-A-l 100-A-21-A-l 92-A-13-A-l
pieceA 0-Aj1-Aj2-Aj3-A pieceB 5-A assembly (((0-A 1-A) (2-A 3-A)) 5-A) 4-A 6-A 7-A subnetwork 49-A-CA-169-A-OG1 159-A-OG-57-A-OE1 169-A-OG1-45-A-O 165-A-49-A-l 45-A-169-A-l 168-A-45-A-m 49-A-162-A-u 53-A-158-A-m 161-A-49-A-u 56-A-157-A-m 52-A-161-A-u
pieceA 0-Aj1-Aj2-Aj3-Aj5-A pieceB 7-A assembly ((((0-A 1-A) (2-A 3-A)) 5-A) 7-A) 4-A 6-A subnetwork 78-A-CA-220-A-OD1 213-A-70-A-u 74-A-209-A-u 70-A-205-A-u 216-A-71-A-u 212-A-67-A-m
pieceA 0-Aj1-Aj2-Aj3-Aj5-Aj7-A pieceB 4-A assembly (((((0-A 1-A) (2-A 3-A)) 5-A) 7-A) 4-A) 6-A subnetwork 140-A-NE2-167-A-O 144-A-SG-167-A-OG 167-A-137-A-u 135-A-207-A-l 203-A-127-A-l 207-A-131-A-l 211-A-131-A-u 143-A-215-A-l 139-A-211-A-l 215-A-139-A-l
pieceA 0-Aj1-Aj2-Aj3-Aj5-Aj7-Aj4-A pieceB 6-A assembly ((((((0-A 1-A) (2-A 3-A)) 5-A) 7-A) 4-A) 6-A) subnetwork 190-A-94-A-m 97-A-183-A-u 98-A-186-A-u 186-A-90-A-m

S7. Structures without determining sets: what the failure of our hypothesis looks like in a retracted structure for EmrE

The fact that most native structures have determining sets can paradoxically create the impression that such sets of interactions occur easily and are of little significance. To help illustrate the falsity of that idea, we analyze a retracted structure.

The 2F2M structure for EmrE has numerous structural anomalies and is incompatible with other experimental data. It was eventually retracted [33]. It is of interest here because it was considered plausible by some experts, yet it is completely wrong due to some unfortunate computational errors discovered later. We will apply our algorithms and analysis to this unquestionably incorrect structure.

It is easy to see from Fig. SD2 (A) that this structure probably does not have a determining set of our interhelical interactions. This is because of the red helix at the bottom of the panel A: It appears to be too far from the rest of the protein to make any contacts of the five types. This is indeed the case. But the rest of the structure, which is more plausibly packed, also lacks these kinds of interhelical interactions. Here is the set of interhelical interactions of the five types for 2F2M according to our algorithm:

Between helices 0-B and 2-B: 75-B-OG-24-B-OG, 78-B-CD1-24-B-O, 78-B-CE1-24-B-O, 75-B-20-B-l

Between helices 1-B and 2-B: 42-B-67-B-u

Between helices 3-A and 0-B: 93-A-7-B-m

For the entire structure, there are just six interhelical interactions of our five types: there is one hydrogen bond, two aromatic interactions, and three packing interactions (two I/V/L patch packings and one small residue packing). See the lower panels of Fig. SD2 for two views of these interactions. For the average interface in our test set, the number of interactions was 6.94, so for this protein with eight helices (and hence seven interfaces), a comparable total number of interactions would be 6.94*7 = 48.58. For every interface the minimum number of interactions in this case would be four, so the minimum possible total of interactions for our assembly algorithm would be 4*7 = 28. The six interhelical interactions of the five types that exist are completely inadequate to reassemble the backbone of the structure. When we apply our algorithm to decompose the set of interactions to obtain the reassembly order, the best we can do is: 0-A 1-A 2-A 3-A (0-B 2-B) 1-B 3-B. So we can build only one helix pair, and a non-sequence-adjacent one at that. This is completely different from what we found for the high-resolution native structures in our test set.

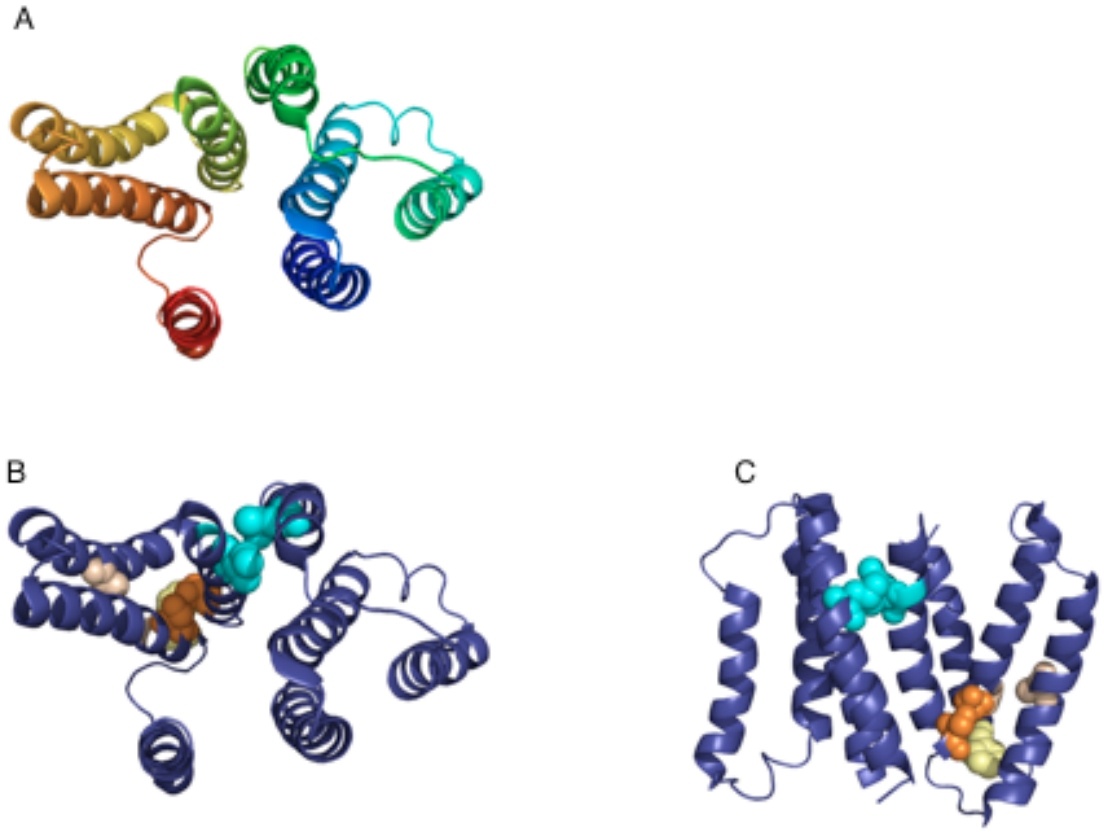

Figure S2. Insufficient set of interactions of the five types in a retracted structure for EmrE
(A) A retracted structure for EmrE (PDB id: 2F2M) shown colored according to sequence order with the colors of the spectrum.
(B) The interhelical interactions of the five types displayed; there are only six in the entire structure. They are displayed as was explained in Fig. 7. Hydrogen bonds are in orange, aromatic interactions in yellow, small residue close knob-in-hole packings in wheat, and I/V/L knob-in-hole packings in I/V/L patches in cyan.
(C) A different view of the interactions in the structure.

S8. Interaction cutoff details
Two of our five interactions have standard definitions as explained in Experimental Procedures. Here we provide more details about the interaction geometries for the other three, and a sketch of the standard definitions used for hydrogen bonds and salt bridges. But it should be noted that many definitions are possible, and many variations would make little difference to the results.

The quadratic penalty terms described below all have unit coefficients unless otherwise specified. The distance cutoffs describe where these quadratic penalties begin rather than absolute cutoffs.

The two packing interactions: small and I/V/L patch close knob-in-hole

As was outlined in Experimental Procedures, both of the packing interactions depend on the definition of close interhelical knob-in-hole packing. Recall that the sidechains (i.e. knobs) have a reduced representation described in Experimental Procedures. Any residue type can pack in a close knob-in-hole fashion, but the only sidechain knobs that appear in these two types of interactions are those of Gly/Ala/Ser and Ile/Val/Leu, and the knob atoms are Cβ except for Gly and Leu, which are Cα and Cγ. The hole is defined by a set of helix residues i, i+3, i+4, i+7 and atoms in those residues: O in i, C in i+3, Cα in i+4, Cβ (Cα for Gly) in i+7. Each hole is divided into three regions: lower (closest to i, denoted 'l'), middle (closest to i+3 and i+4, denoted 'm') or upper (closest to i+7, denoted 'u'). The knob atom must lie within a set of distance cutoffs to the surrounding hole atoms.

For the lower region, the penalty function is based on distances from the knob position to the following atoms: to O (in i), to C (in i+3) and to Cα (in i+4). The penalty is quadratic and based on these three distances; additionally the sum of the three distances and the sum of the distances to C (in i+3) and to Cα (in i+4) are also used. The penalties begin at about the following distances: 5 Å for the knob to i-O, 6 Å for the knob to (i+3)-C and the knob to (i+4)-Cα, 11.5 Å for the sum of the latter two, and 16.5 Å for the full total of the three distances.

For the upper region, the penalty function is identical to that described for the lower region except that O (in i) is replaced by Cβ (Cα for Gly) (in i+7).

For the middle region, we use the same three distances as for the lower region unless i-O does not lie on the helix in which case we use the same three distances as for the upper region. The distance cutoffs are different: 5 Å for distance from the knob to C (in i+3) and to Cα (in i+4); 6.5 Å for the distance from the knob to O (in i) (unless residue i is not on the helix, in which case Cβ (Cα for Gly) (in i+7) is used); and 10.75 Å for the sum of the distances from the knob position to C (in i+3) and to Cα (in i+4).

The overlap penalty is also computed from the knob atom and the residues that surround the hole and added to the total penalty. The total penalty should be less than 0.5, except for the case when the knob is Val/Ile/Leu. In this case, an additional error term is added because the geometry of these sidechains means that their knob atoms (Cβ for V/I; Cγ for L) tend to sit farther out. The additional error term is about 2.9.

Aromatic interactions

As was outlined in Experimental Procedures, the ring atoms can act as weak donors to acceptors. (The center of the ring can act as an acceptor, but does not for this test set.) The distances used to define the geometry of these interactions are the distance from the acceptor to the donor atom and the distances from the acceptor atom to the donor-adjacent atoms in the ring. The cutoffs for these distances are 3.6 Å for the donor-acceptor atoms and 4.65 Å for the donor-adjacent atoms. Additionally, the acceptor atom is meant to lie near the plane of the ring, so the absolute value of the cosine of angle between the vector difference of the donor and acceptor positions and the normal to the plane of the ring is also used for the penalties. Also, if the distance from the acceptor to a donor-adjacent atom is less than that to the donor atom itself, a quadratic penalty is assessed with the donor-acceptor distance cutoff. Together the sum of these penalty terms has an absolute cutoff of about 0.4 for the donor/acceptor pair to be considered part of an aromatic interaction.

Hydrogen bonds

There are five geometric criteria given for a hydrogen bond in McDonald and Thornton [38]. Two are given as distance cutoffs: 3.9 Å for the maximum distance between the donor atom and the acceptor atom, and 2.5 Å for the maximum distance between the

donor hydrogen and the acceptor atom. Three are given as angle restrictions: the angle formed by the donor atom, the hydrogen, and the acceptor atom, D_H_A, must be larger than 90 degrees; the angle formed by the hydrogen, the acceptor atom, and the acceptor antecedent atom, H_A_AA, must be larger than ninety degrees; and the angle formed by the donor atom, the acceptor atom, and the acceptor antecedent, D_A_AA, must be larger than ninety degrees.
The quadratic penalty function based on these geometric criteria uses the above cutoffs, except that the donor-acceptor distance cutoff is 3.8 Å. The angle conditions are all converted to distance conditions. The coefficient for the D_H_A derived cutoff is 20 and that for the H_A_AA derived cutoff is 10; as mentioned before, all others are equal to 1.
The structures do not initially have hydrogens, and potential bonds are checked for the two conditions applicable without hydrogens. If they are met, then hydrogens are added as described in [38], and the cutoff for the full penalty is 0.05.

Salt bridges
Salt bridges are defined by the distance between the presumed positively and negatively charged atoms. We took 4 Å with an error term added, which came to just under 4.25 Å.

S9. Coordinates for reduced sidechain representatives: bumps
The bumps are intended to give the obstruction created by a sidechain of a given amino acid type irrespective of its rotameric state. They were made by averaging the sidechain conformations of the residues of a particular type in a test set of solved soluble structures, and removing any atoms from the averaged sidechain that did not fall within a given cutoff of some atom (not necessarily the same one) for every sidechain conformation of the residues of that type in the test set. See Fig. 2 for examples.
The overlap penalties associated with bumps are lower than those associated with collisions of other atoms as was described in Experimental Procedures. This is designed to take into account the approximate nature of these representations.
The basis vectors for the coordinates below are obtained by applying the Gram-Schmidt algorithm to the vectors Cβ-Cα, N-Cα, and C-Cα, where Cβ, N, C and Cα give the positions of these atoms in a residue. (There are some slight deviations from orthogonality caused by rounding errors.) The order of the atoms is the conventional one for each residue type, and most of the residues have chopped representations, i.e., atoms missing at the end. The number of atoms in a representative can be misleading since some do not at all obey usual VDW conditions and overlap each other.
The bumps for Ala and Gly are Cβ and Cα, and are not shown below. Below are the coordinates for the bumps of the other residues types in terms of the basis described above. The atoms are given in the conventional order for that residue type so that, e.g., the first three numbers give the position of Cβ, etc.

Arg: 1.53 0.00 0.00 2.09 -0.87 -0.24 3.17 -1.13 -0.59
Asn: 1.53 0.00 0.00 2.08 -1.02 -0.09 2.54 -1.41 0.03
Asp: 1.53 0.00 0.00 2.10 -0.81 -0.33
Cys: 1.53 0.00 0.00 2.20 -1.01 -0.18
Gln: 1.53 0.00 0.00 2.11 -0.79 -0.16
Glu: 1.53 -0.01 0.01 2.12 -0.73 -0.10
His: 1.54 0.00 0.00 2.09 -0.36 -0.49 2.54 -0.58 -0.69 2.24 -0.52 -0.81
Ile: 1.54 0.00 0.00 2.05 -0.97 0.31 2.10 0.22 -0.97 2.97 -1.28 0.34
Leu: 1.53 0.00 0.00 2.16 -0.90 -0.26 2.77 -1.28 -0.32 2.57 -1.39 -0.70
Lys: 1.53 -0.01 -0.03 2.11 -0.86 -0.29 3.07 -1.15 -0.45
Met: 1.53 0.00 0.00 2.07 -0.73 -0.26
Phe: 1.53 0.00 0.00 2.11 -0.42 -0.77 2.76 0.14 -0.90
Pro: 1.53 0.00 0.00 1.92 0.11 1.35 0.81 0.00 2.29
Ser: 1.53 0.00 0.00 2.01 0.00 -0.10
Thr: 1.54 0.00 0.00 2.01 -0.70 0.54 2.09 -0.11 -0.90

Trp: 1.53 0.00 0.00 2.03 -0.18 -0.78 2.50 -0.80 -0.84
Tyr: 1.56 -0.03 0.01 2.13 -0.59 -0.41 2.66 -0.27 -0.32
Val: 1.54 0.00 0.00 2.10 0.11 -0.98 2.07 -0.92 0.43

S10. Scoring for decomposition of the set of interhelical interactions of the five types

The algorithm for decomposition was outlined in Experimental Procedures. An essential step in the iterative procedure is to take a triple of pieces in sequence order H_j, H_j+1, H_j+2 and to decide which, if either, of the two pairs to assemble together: H_j, H_j+1 or H_j+1, H_j+2. If both of the possible set of interface interhelical interactions of the five types contain at least four (three if one of the pieces is a half-helix) interactions, we have to decide which is the better pair. There are many possible decompositions of a determining set of interactions, and one was selected automatically according to the scoring outlined below, but many others could correspond to equally good or better physical pathways. The resulting orders of assembly are given in S2.

The total score for the combined piece is based on the following terms: a connecting loop length score (with shorter loops preferred), the sum of the number of helices in the two pieces(with smaller pieces preferred), a bias score for the pair earlier in the sequence, and a score for the set of interhelical interactions of the five types. The score for the set of interactions is the sum of the scores of the interactions in the set, so that more interactions are preferred. The score for each interhelical interaction depends on the type of interaction and the side chains involved. For hydrogen bonds between shorter side chains or with a backbone oxygen acceptor, a bonus was added to the default score for a hydrogen bond, so that these interactions between shorter side chains are preferred. In the case of oligomers, a bias was added to force the monomers to be assembled first.