SUPPORTING INFORMATION

Toward A Soluble Model System for the Amyloid State

Nicole C. Thomas[⊥], Gail J. Bartlett[§], Derek N. Woolfson[§], and Samuel H. Gellman[⊥]*

[⊥]Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706

[§]School of Chemistry, Cantock's Close, University of Bristol, Bristol BS8 1TS, UK; School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1 TD, UK; BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

Note: The results documented below are derived from the Ph.D. thesis of Vanessa M. Kung, UW-Madison, 2016, who was supported in part by the Integrated Training for Physician-Scientists program, T32 GM008692.

β-Arch Consensus Sequences

Kajava and coworkers classified β-arches by the backbone dihedral angles of their loop regions ('β-arcs'), and identified several classes that produce 180° turns: three-residue (*ppl*, *xbl*), four-residue (*bepl*, *bebl*, *gbpl*, *bgpp*), five-residue (*blbbl*), and six-residue (*baepep*, *bllpbl*) β-arcs, where *a* = (right-handed) α-helix, *b* = β-strand, *d* = bridge (δ), *e* = extended chain (ε), *g* = γ', *l* = left-handed α-helix, *p* = polyproline-II, and *x* = "around Φ = -80° and Ψ = -140°" regions of the Ramachandran plot.[1] We mined the RCSB Protein Data Bank (PDB) for structures containing these backbone dihedral angles, using DSSP to assign dihedral angles.[2-4] The PDB was culled by sequence identity (30%), resolution (2.0 Å), and R-factor (0.25), on December 30, 2012.[5] After visual inspection of the culled set, we obtained the following sets of β-arches: 26 *ppl*, 0 *xbl*, 0 *bepl*, 8 *bebl*, 5 *gbpl*, 2 *bgpp*, 12 *blbbl*, 7 *bllpbl*, and 1 *baepep* β-arches. Of these, 3 *ppl*, 3 *bebl*, 5 *gbpl*, 1 *bgpp*, 12 *blbbl*, and 7 *bllpbl* β-arches produced 180° turns (**Figure S1**). Sequence logos in **Figure S1** were constructed with WebLogo.[6]

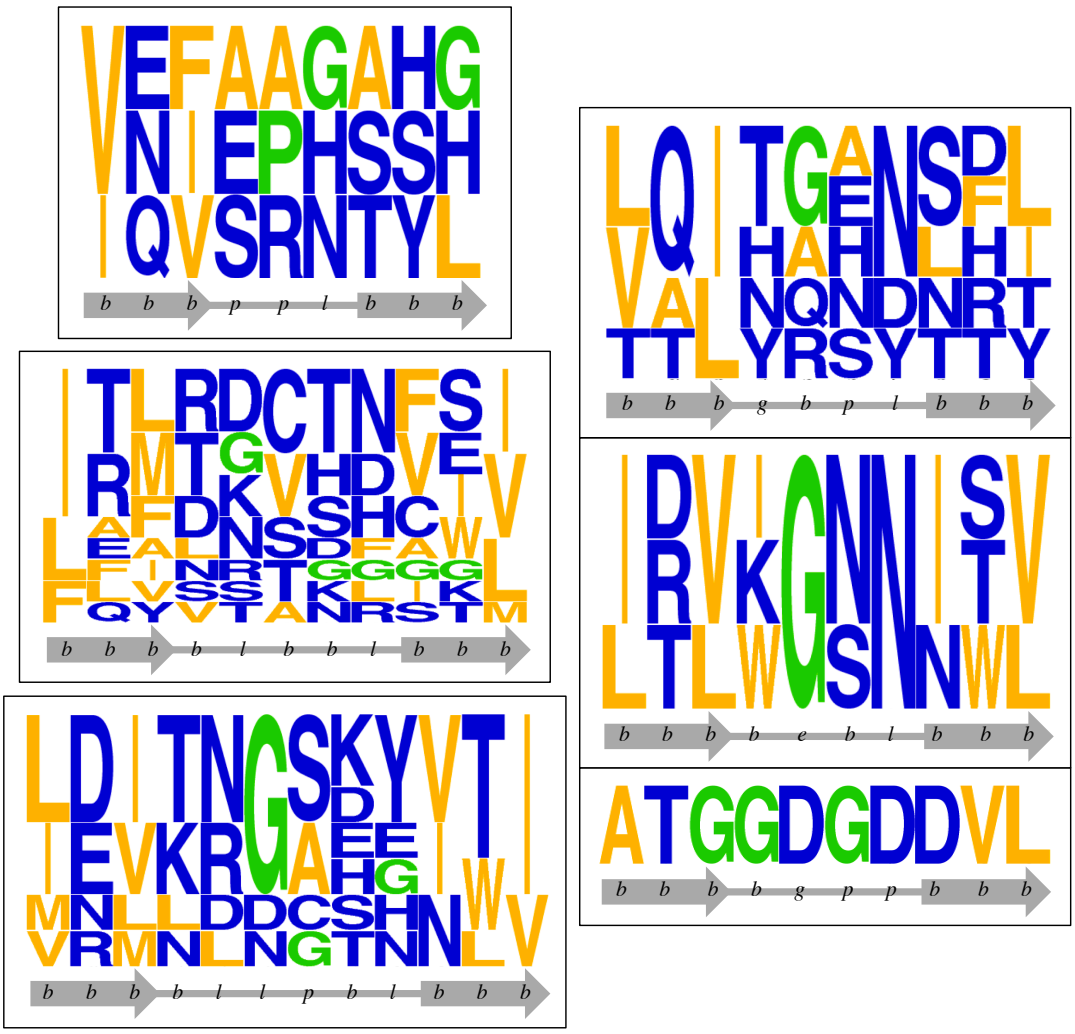


Figure S1. Sequence logos (residue frequency vs. position) for *ppl*, *gbpl*, *bebl*, *bgpp*, *blbbl*, and *bllpbl* β-arcs that produce 180° turns.

Using a set of 16 *blbbl* β-arcs – the 12 *blbbl* β-arcs in our culled dataset, plus 4 *blbbl* β-arcs that were retrieved from our culled PDB upon slight broadening of our backbone angle constraints for the β-strands that flank *blbbl* (i.e., slight relaxation of the definition of *b*, for positions "1B1"-"1B3" and "2B1"-"2B3") – a new sequence logo was generated (**Table S1 and Figure S2**). These data were normalized to account for the natural abundance of amino acids.[7]

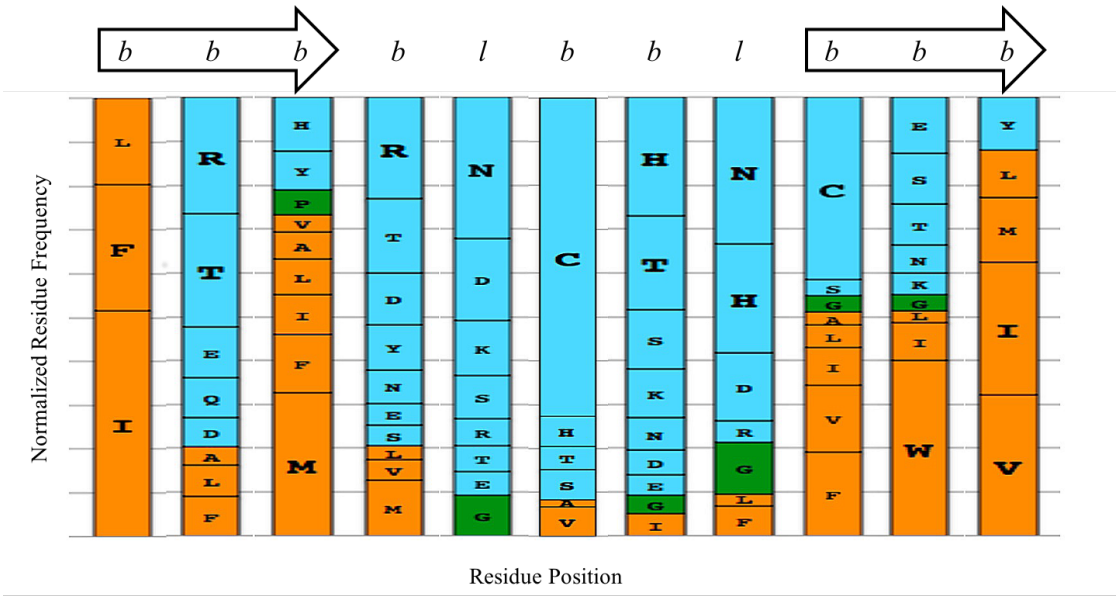


Figure S2. Sequence logo for *blbbl* β-arcs, where *b* = β-strand, and *l* = left-handed α-helix backbone dihedral angles. Residues of the β-arc are typically hydrophilic (blue), while residues of flanking β-strands are typically hydrophobic (orange).

Table S1. The redundant set (culled set in red font) of *blbbl*-containing sequences in December 30, 2012 PDB. The sequences that had a *blbbl* backbone conformation, but that were not 180° β-arcs on visual inspection, are in gray boxes.

PDB code	A1 res#	*bbb*	***blbbl***	*bbb*
1a9x	4171	---	SFTMG	---
1ag8	453	TVW	VNCYD	VFG
1bhe	106	ITA	VSTTN	SGI
"	157	IQI	NKSKN	FTL
"	207	---	MSSKN	---
1clw	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1cs0	171	---	SFTMG	---
1czf	185	FDV	GNSVG	VNI
1edq	452	---	NGYQN	---
1ehn	452	---	NGYQN	---
1eib	452	---	NGYQN	---
1ezg	29	---	TNSQH	---
"	41	---	TGSTD	---
"	53	---	TNSKD	---
"	65	---	TDSTN	---
"	77	---	TNSSG	---
1ffr	452	---	NGYQN	---
1go7	67	---	TNVFG	---
1go8	67	---	TNVFG	---
1gv4	578	---	WNVFN	---
1h54	490	---	NNVDN	---
1hg8	157	---	TGSSQ	---
"	196	FDI	SSSDH	VTL
1ia5	110	---	HSLTN	---
"	133	FSV	AGSDY	LTL
"	164	FDI	GTSTY	VTI
1idk	159	ITL	DDCDL	VWI
1iru	156	---	KNMQN	---
1ja9	170	---	TGIPN	---
1k4z	1404, 2404	IFI	GKCSQ	VLV
"	1423, 2423	ISL	SETES	CSV
"	1442, 2442	MDV	IKSNK	FGI
"	1482, 2482	IYT	SCSTA	INV
1k7g	67	---	TNVFG	---
1k7i	67	---	TNVFG	---
1k7q	67	---	TNVFG	---
1k8f	354	---	YKCVN	---
"	373	ITV	DNCKK	LGL
"	392	VEI	INSKD	VKV
"	432	IVS	AKSSE	MNV
1k9t	452	---	NGYQN	---
1kee	171	---	SFTMG	---
1kq5	1404, 2404	---	GKCSQ	---
"	1423, 2423	ISL	SETES	CSV
"	1442, 2442	MDV	IKSNK	FGI
"	1482, 2482	IYT	SCSTA	INV
1m6i	579	---	WNIFN	---
1mhs	818	---	FWSSI	---
1nh6	452	---	NGYQN	---
1nhc	138	---	HDVED	---
"	191	FDI	SESTG	VYI
1o04	453	TVW	VNCYD	VFG
1o88	258	---	EKAIN	---
1o9j	453	---	VNCYL	---
1p4k	80, 380	---	ACMEH	---
1pxz	111	LFM	RKVSH	VIL
"	156	ITM	RNVTN	AWI
1q1l	65	---	SGVRF	---
1qa1	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qa2	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qa3	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM

PDB code	A1 res#	*bbb*	***blbbl***	*bbb*
1qq1	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qrb	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qrc	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qxo	60	---	SGVRH	---
1r53	67	---	SGTEF	---
1rmg	96	IAV	TDTTD	FEL
"	133	LRL	TDVTH	FSV
"	156	FTM	DTCSD	GEV
1tyv	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1u79	104	---	AGCKG	---
1um0	67	---	SGVFE	---
1uun	152	---	TGAAG	---
1vbl	153	---	KNVDN	---
"	195	ISI	EGSSH	IWI
1vz2	169	---	ERVKF	---
1vz3	169	---	ERVKF	---
1whu	356	---	EYKRC	---
1x6l	452	---	NGYQN	---
1y4z	394	---	WGEKG	---
1y5i	394	---	WGEKG	---
1yo8	1155	---	FSQEX	---
1zum	453	TVW	VNCYD	VFG
2a8j	131	---	GALSG	---
2a8l	131	---	GALSG	---
2a8m	131	---	GALSG	---
2b0r	55	---	ENHVN	---
"	97	IFI	ENCVG	CIF
"	116	IEI	VNCDD	IKL
"	135	ISL	DKSNK	VNI
"	156	VYS	SKSSE	MNL
"	78	VSV	RCSQN	SRI
2b39	1355	---	VSCKK	---
2bx6	65	FLI	QDCEN	CNI
"	84	VTI	DDCTN	CII
"	103	VFF	RNCRD	CKC
"	120	FRV	RDCRK	LEV
"	138	---	ESSSN	---
2c1d	(118)	---	DGFVR	---
2cu2	284	HVA	LDTFG	CVV
"	301	VAT	LGVSG	LVV
2ewe	258	---	EKAIN	---
2f8o	30	---	FHASD	---
2g85	61	---	SGIRH	---
2gaw	80	---	ACMEH	---
2gl9	80, 380	---	ACMEH	---
2gp1	305	---	SNIMW	---
2inu	277	---	ENHEG	---
"	300	IEF	TGCNR	CSV
2iq7	129	---	HSLKS	---
"	152	FSI	NSATT	LGV
"	183	FDV	GSSTG	VYI
2o2p	860	---	INTYN	---
2onp	453	TVW	VNCYD	VFG
2pyg	178	---	DYLVD	---
"	255	VLL	KMTSD	ITL
"	278	VRV	YGAQD	VQI
2qhf	61	---	SGIRH	---
2qx3	92	---	KNKSD	---
"	270	---	ENIKN	---
2qxz	92	---	KNKSD	---
"	270	---	ENIKN	---

PDB code	A1 res#	*bbb*	***blbbl***	*bbb*
2qy1	92	---	KNKSD	---
"	270	---	ENIKN	---
2rhp	1155	LFV	FSQEM	VYF
2uvf	360	IMN	LENHN	VVA
"	386	IEF	GNSQN	VMV
"	570	---	SDLRD	---
"	594	WHF	SEVKN	VKV
2v5i	310	LDI	RSASR	VEI
"	330	YLF	RSCHH	CKV
2vbk	514	LRY	LRSGG	VSV
2vfm	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfn	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfo	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfp	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfq	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vle	453	TVW	VNCYD	VFG
2wk2	452	---	NGYQN	---
2wme	441	---	INTWG	---
2wox	441	---	INTWG	---
2x65	283	VIL	VDSDR	VFV
"	300	IAV	VGLSD	VIV
2xc1	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2yuh	14	---	SNLES	---
"	36	VLL	TELSN	CTV
"	55	LRL	TKAHS	CKL
"	74	VFL	EDCSD	CVL
"	91	---	HSTKD	---
"	109	---	EDCSG	---
2zah	367	---	GTVVA	---
3bh6	(43)	---	SGLKD	---
"	(65)	FLI	QDCEN	CNI
"	(84)	VTI	DDCTN	CII
"	(103)	VFF	RNCRD	CKC
"	(120)	FRV	RDCRK	LEV
"	(138)	---	ESSSN	---
3bh7	(65)	FLI	QDCEN	CNI
"	(84)	VTI	DDCTN	CII
"	(103)	VFF	RNCRD	CKC
"	(120)	FRV	RDCRK	LEV
"	(138)	---	ESSSN	---
3ddu	169	---	ERVKF	---
3cgw	394	---	WGEKG	---
3eqn	620	YFL	KNAAN	HYI
"	669	MTV	QNSHG	ILV
3fby	727	---	FSQEN	---
3gd3	578	---	WNVFN	---
3gd4	578	---	WNVFN	---
3gq8	222	LSI	RACHN	VYI
"	285	ITT	HHSQY	INI
"	587	SGV	LTSSS	SKT
"	602	SLI	AGSST	SEA
"	617	NTI	LGSLG	AVA
"	632	---	ISASQ	---
"	647	NLI	LSSYG	INT
3hb2	67	---	TNVFG	---
3hbu	67	---	TNVFG	---
3hbv	67	---	TNVFG	---
3hwc	3189	---	AVGTG	---
3inl	453	---	VNCYD	---

PDB code	A1 res#	*bbb*	***blbbl***	*bbb*
3ir5	394	---	WGEKG	---
3ir6	394	---	WGEKG	---
3ir7	394	---	WGEKG	---
3it5	80	---	YHMDQ	---
3iuj	170	---	KDVKF	---
3iur	170	---	KDVKF	---
3iwk	447	---	INCAQ	---
3jqy	48	---	VNSKE	---
3jur	120	VYA	LDCEN	VAI
"	196	VQF	YRCRN	VLV
"	219	---	VLSEN	---
"	244	---	ESCKY	---
3k5i	149	---	TMAYD	---
3km5	1315	---	YNCSD	---
3lj5	45	---	LSQYT	---
3ljq	80	---	ACMEH	---
"	380	---	ACMEH	---
3m1h	1584	---	FGCTD	---
3n80	453	TVW	VNCYD	VFG
3n81	453	TVW	VNCYD	VFG
3ogl	146	---	DKCSG	---
3qcw	627	---	ENKAG	---
3rhj	860	---	INTYN	---
3rhl	860	---	INTYN	---
3rho	860	---	INTYN	---
3rhq	860	---	INTYN	---
3riq	281	---	NSANN	---
"	301	---	TLCNH	---
3s2k	871	---	GHLDY	---
3s94	48	---	GGLED	---
"	357	---	EDIRH	---
3suc	222	LSI	RACHN	VYI
"	285	ITT	HHSQY	INI
"	571	---	MGSTS	---
"	587	SGV	LTSSS	SKT
"	617	NTI	LGSLG	AVA
"	632	---	ISASQ	---
"	647	NLI	LSSYG	INT
3th0	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
3ven	11	---	RDFHD	---
3ver	11	---	RDFHD	---
3vex	11	---	RDFND	---
3vez	11	---	RDFHD	---
3vf2	11	---	RDFHD	---
3vf4	11	---	RDFHD	---
3vmv	100	IRL	SNAHN	III
3zqa	441	---	INTWG	---
3zsc	88	LVI	KDAQN	VII
"	122	INV	ENSHH	IWI
4a0p	745	---	KDLDS	---
4a2a	43	---	RGLDE	---
4a2b	43	---	RGLDE	---
4ap6	97	---	IHQVR	---
4dg6	116	---	QELDQ	---
"	160	---	SEIYW	---
4ecd	61	---	TGVRH	---
4fdc	579	---	WNIFN	---
4fr8	453	TVW	VNCYD	VFG
4gdv	93	---	SAVQC	---
4h7n	415	---	INDAA	---
9gaa	80, 380	---	ACMEH	---
9gac	80	---	ACMEH	---
"	380	---	ACMEH	---
9gaf	80	---	ACMEH	---
"	380	---	ACMEH	---

Of note in our *blbbl* β-arc consensus sequence, is the occurrence of L-asparagine (Asn, N), and less commonly L-aspartate (Asp, D), at the two positions that adopt backbone dihedral angles in the left-handed α-helix (*l*) region of the Ramachandran plot. This result is consistent with the observation that asparaginyl residues adopt conformations in partially-allowed regions of the Ramachandran plot more readily than any other non-glycyl amino acids.[8] One theory is that the *l* conformation of an asparaginyl residue is stabilized by non-covalent attractive forces between the dipoles of the asparaginyl residue's side chain carbonyl, and the backbone carbonyl of either that residue or the previous residue.[9,10]

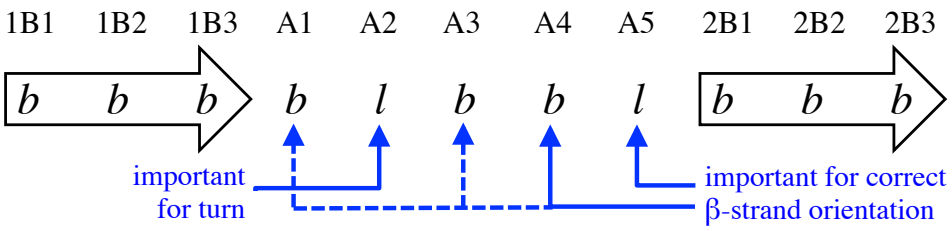


Figure S3. Contributions of backbone dihedral angles to *blbbl* β-arc.

To better characterize the influence of each element of the *blbbl* sequence on the β-arc's conformation, we mined the PDB (via DSSP) for sequences in which one element of *blbbl* was allowed to adopt any backbone conformation (*n*): *nlbbl*, *bnbbl*, *blnbl*, *blbnl*, and *blbbn*. In addition to the *blbbl* set, these searches returned culled sets of 1 *nlbbl*, 30 *bnbbl*, 2 *blnbl*, 3 *blbnl*, and 24 *blbbn* sequences. Of these, 1/1 *nlbbl*, 1/30 *bnbbl*, 2/2 *blnbl*, 1/3 *blbnl*, and 3/24 *blbbn* sequences were β-arcs (not necessarily 180°). For comparison, 12/13 of the *blbbl* sequences were 180° β-arcs. Examination of all these sequences revealed possible contributions of backbone angles to the β-arc (**Figure S3**). Many *bnbbl* sequences were parts of extended β-strands, which suggests that the *l* at position "A2" is important for creating a turn. Both *blbnl* sequences were β-turns, and many *blbbn* sequences created turns with flanking β-strands in orientations other than that of a β-arch; these results suggest that the *b* at position "A4" and the *l* at position "A5" are important for creating the distinctive β-strand orientation in a β-arch, which supports interactions between hydrophobic side chains on facing β-strands. With the importance of the *l* angles at positions "A2" and "A5" established, the culled PDB was searched for *nlnnl* sequences. Not counting *blbbl* sequences, the *nlnnl* search returned 20 sequences, 8/20 of which were 180° β-arcs. The *nlnnl* sequences that were not 180° β-arcs were often β-turns, or β-arcs with a < 180° change of chain direction; this suggested that the *b* angles at positions "A1," "A3," and "A4" contribute to the orientation of the β-strands in a 180° β-arch. However, the collective contribution of the *b* positions of *blbbl* to the orientation of the β-strands is probably not as important as the contribution of the "A5" *l* angle, given that 8/20 *nlnnl* sequences are still 180° β-arcs, while only 3/24 *blbbn* sequences are 180° β-arcs. Our set of 12 *blbbl* β-arches produced a consensus sequence with asparaginyl residues in both of the *l* positions. In contrast, in the consensus sequence for the 12 *blbbl*, 1 *bnbbl*, 3 *blbbn*, and 8 *nlnnl* (including 1 *nlbbl*, 1 *blnbl*, and 1 *blbnl*) β-arches, the most common amino acid at the *l* positions was glycine (Gly).

Kajava and coworkers identified single cases of non-*blbbl*, five-residue, 180° β-arcs: *blbbx*, *blbpl*, and *bldpl*.[1] The Ramachandran basins *p* and *d* neighbor the β-strand (*b*) basin. In this work we define a region *x* ($\Phi \in [30^\circ, 110^\circ]$, $\Psi \in [90^\circ, 120^\circ]$) that neighbors the left-handed α-helix (*l*) basin. Our searches identified no *blbbx* β-arches in our *blbbn* set, and one *blbpl* β-arch in our *blbnl* set. Mining the culled PDB for *blnnl* (including *bldpl*) sequences did not return any β-arches (besides the aforementioned two *blnbl* β-arches and one *blbnl* β-arch).

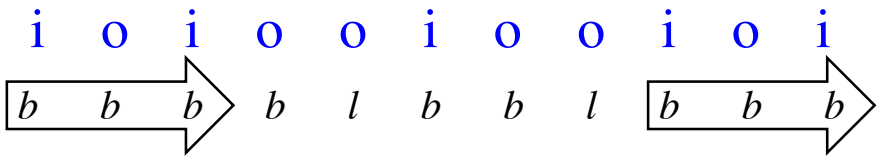


Figure S4. Solvent accessibility of residue side chains in the *blbbl* β-arc. Outside (o) residues are solvent-accessible; inside (i) residues are solvent-inaccessible.

For each β-arch in our culled dataset, we used established methods to determine the solvent-accessible surface area (ASA) of the residue side chains in the PDB structure;[11,12] assigned "outside (o)" or "inside (i)" to side chains that were solvent-accessible or inaccessible, respectively; and generated the o/i consensus sequence in **Figure S4**. A side-chain was deemed to be solvent-accessible, if its relative accessibility was >5%, as calculated using NACCESS.[12]

Synthesis and Characterization of Peptides

The OAllyl-Gly-(1*R*,2*S*)-CHDA-Val-OH and Alloc-Glu-Val-^{D}Pro-DADME-Fmoc linkers were synthesized via previously described routes.[13] Peptide **1** was synthesized via the route depicted in **Figure S5**, and all other peptides were synthesized by analogous routes. Peptides were purified by reversed-phase high-performance liquid chromatography (HPLC), using a C18 preparatory column, and a mobile phase of CH_3CN:H_2O:CF_3COOH. Peptide identities were confirmed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). MALDI-TOF MS showed the following m/z values for the $[M+H^+]^+$ species: peptide **1** calculated = 3488.93, observed = 3488.0; peptide **2** calculated = 2722.50, observed = 2722.1; peptide **3** calculated = 3488.93, observed = 3489.5; peptide **4** calculated = 3574.94, observed = 3574.8; peptide **5** calculated = 3460.90, observed = 3460.6. Purity was determined by analytical HPLC or ultra performance liquid chromatography (UPLC), and was >95% for all purified peptides. Analytical HPLC and UPLC runs were performed using a C18 column on a 10-60% B solvent gradient, where solvent A is 99.9:0.1 H_2O:TFA and solvent B is 99.9:0.1 acetonitrile:TFA. All chromatograms recorded at 220 nm. (**Figures S6**).

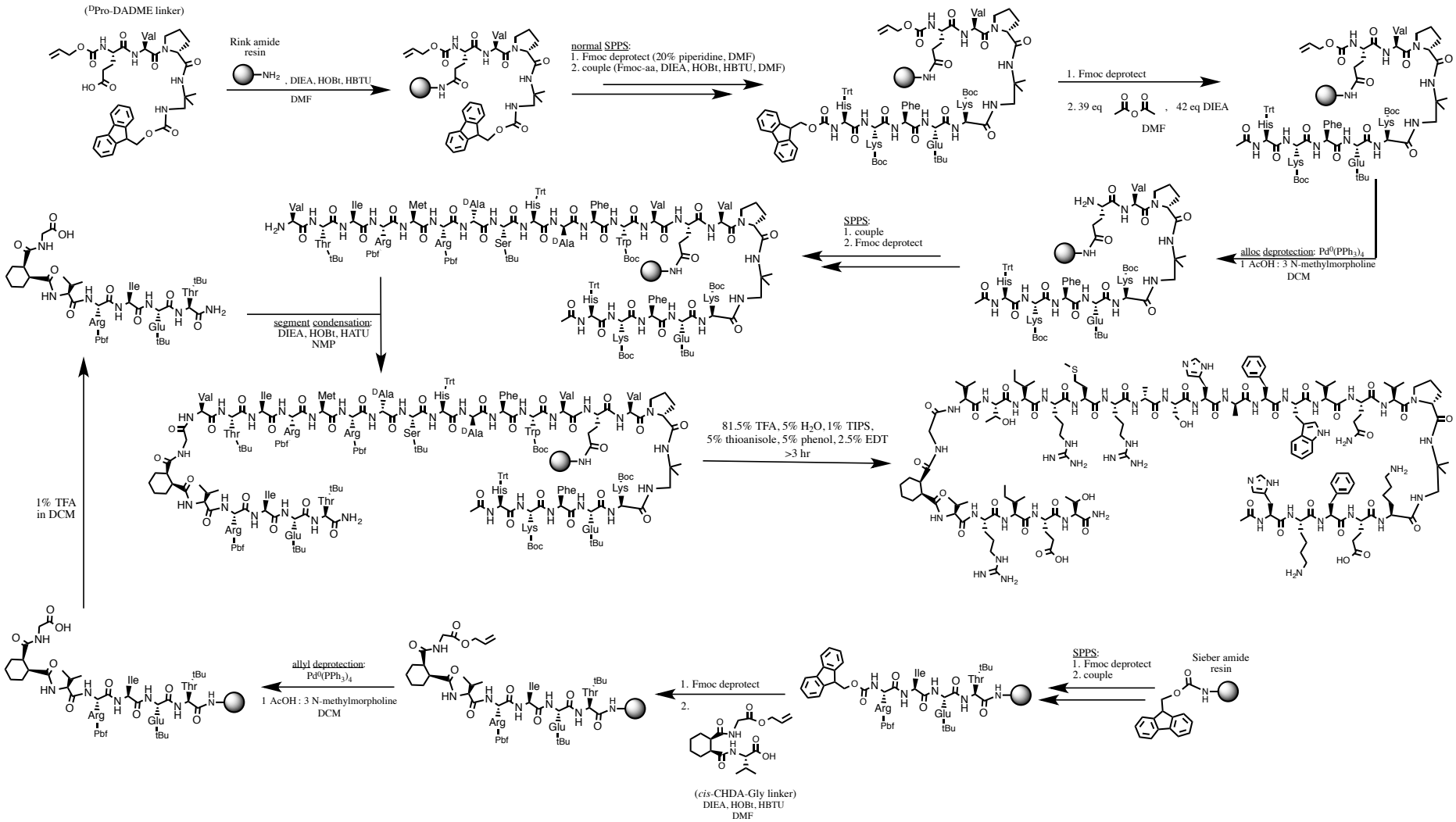
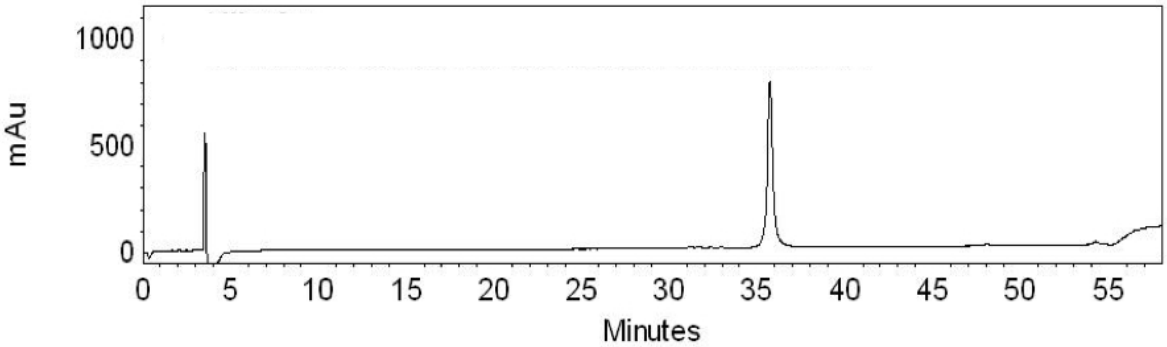


Figure S5. Synthetic route to peptide **1.** Analogous routes were employed to synthesize other peptides.



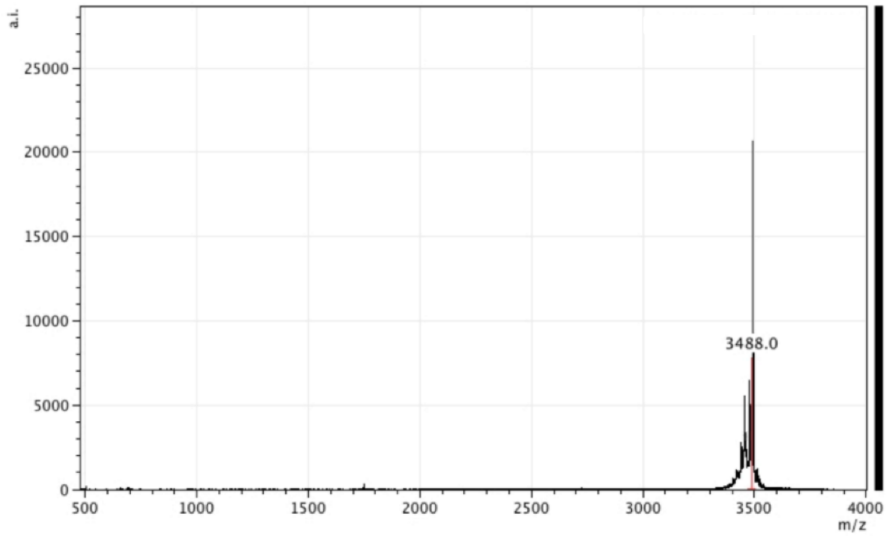
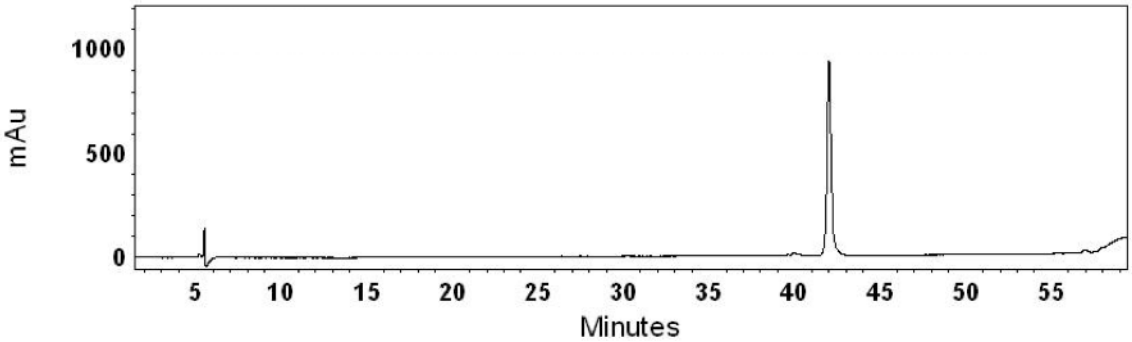


Figure S6A. Analytical HPLC trace and MALDI-TOF for peptide **1.** $[M+H]_{calc}$ = 3488.93



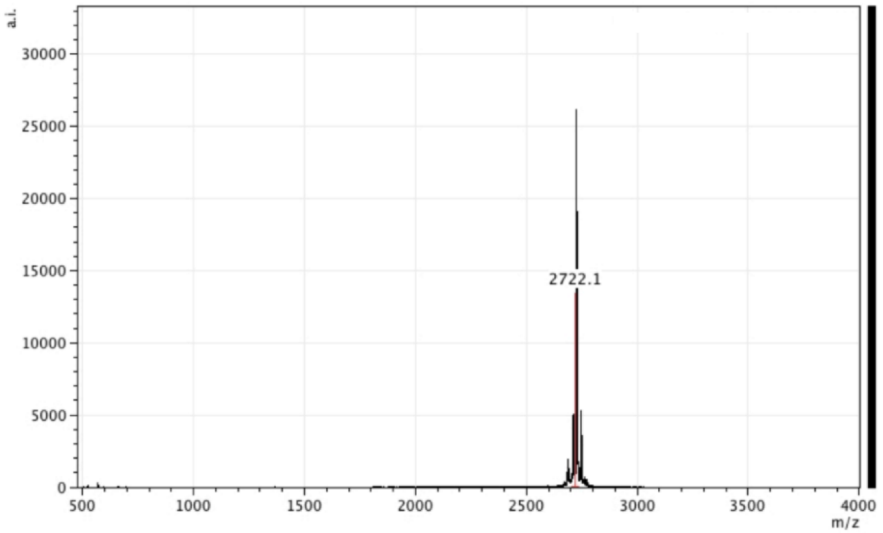
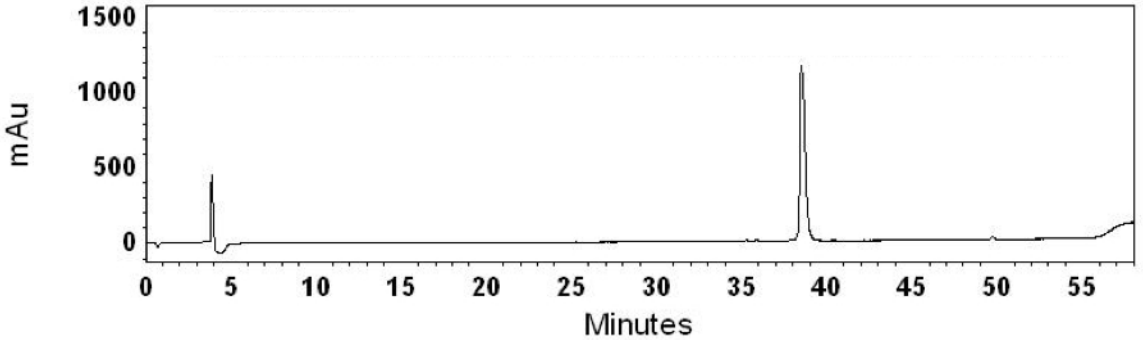


Figure S6B. Analytical HPLC trace and MALDI-TOF for peptide **2**. $[M+H]_{calc}$ = 2722.50



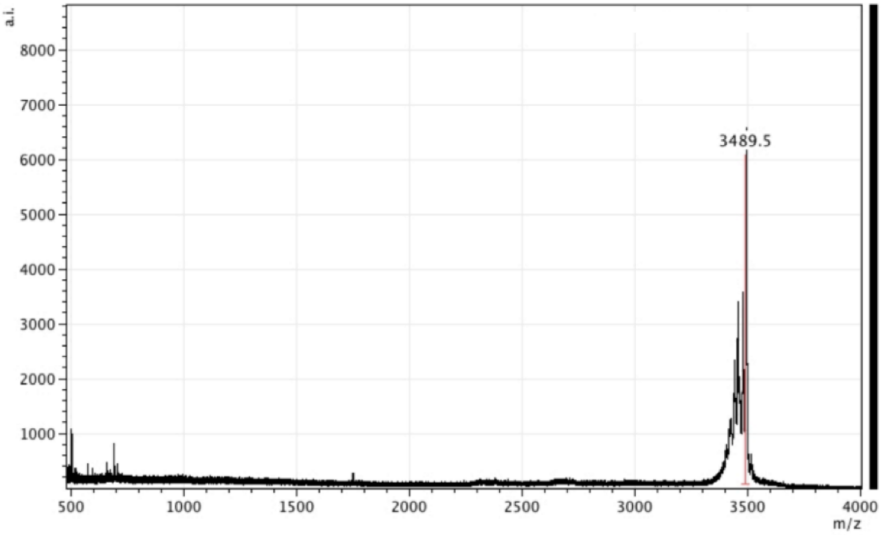


Figure S6C. Analytical HPLC trace and MALDI-TOF for peptide **3**. $[M+H]_{calc}$ = 3488.93

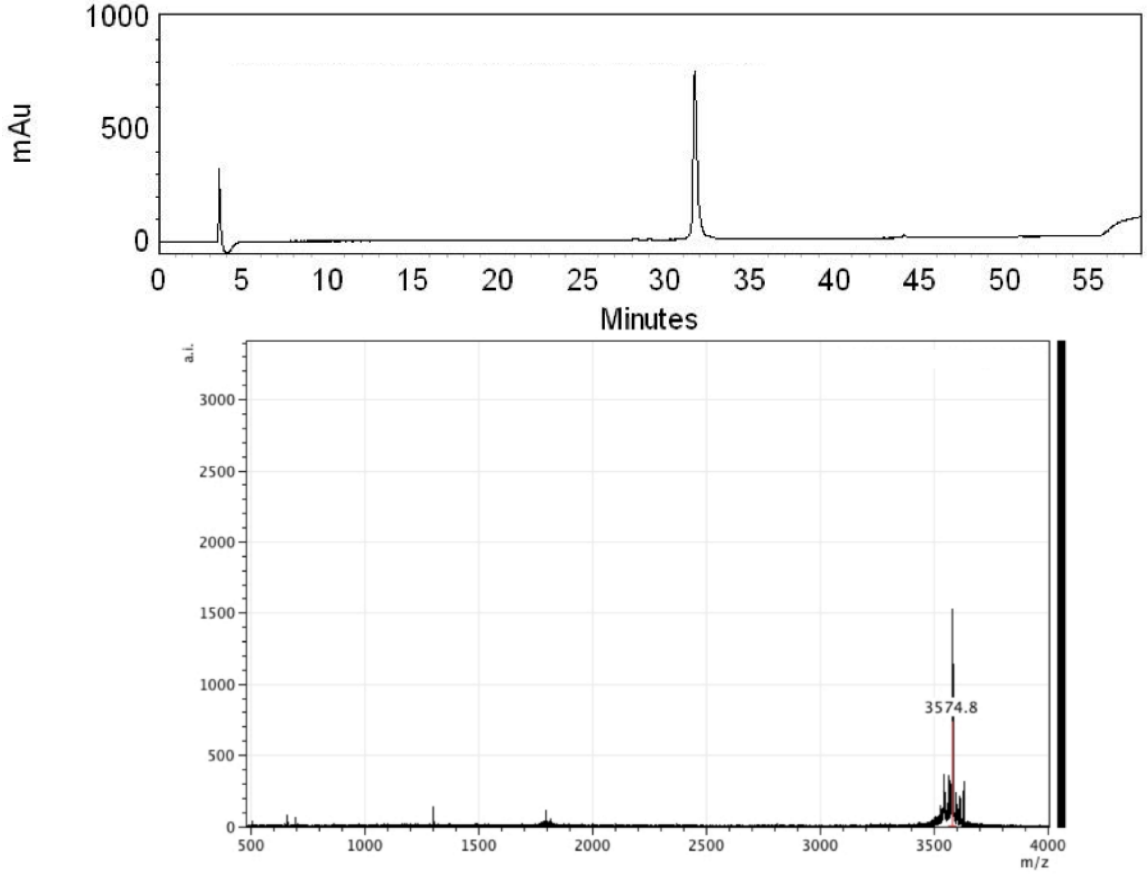


Figure S6D. Analytical HPLC trace and MALDI-TOF for peptide **4**. $[M+H]_{calc} = 3574.94$

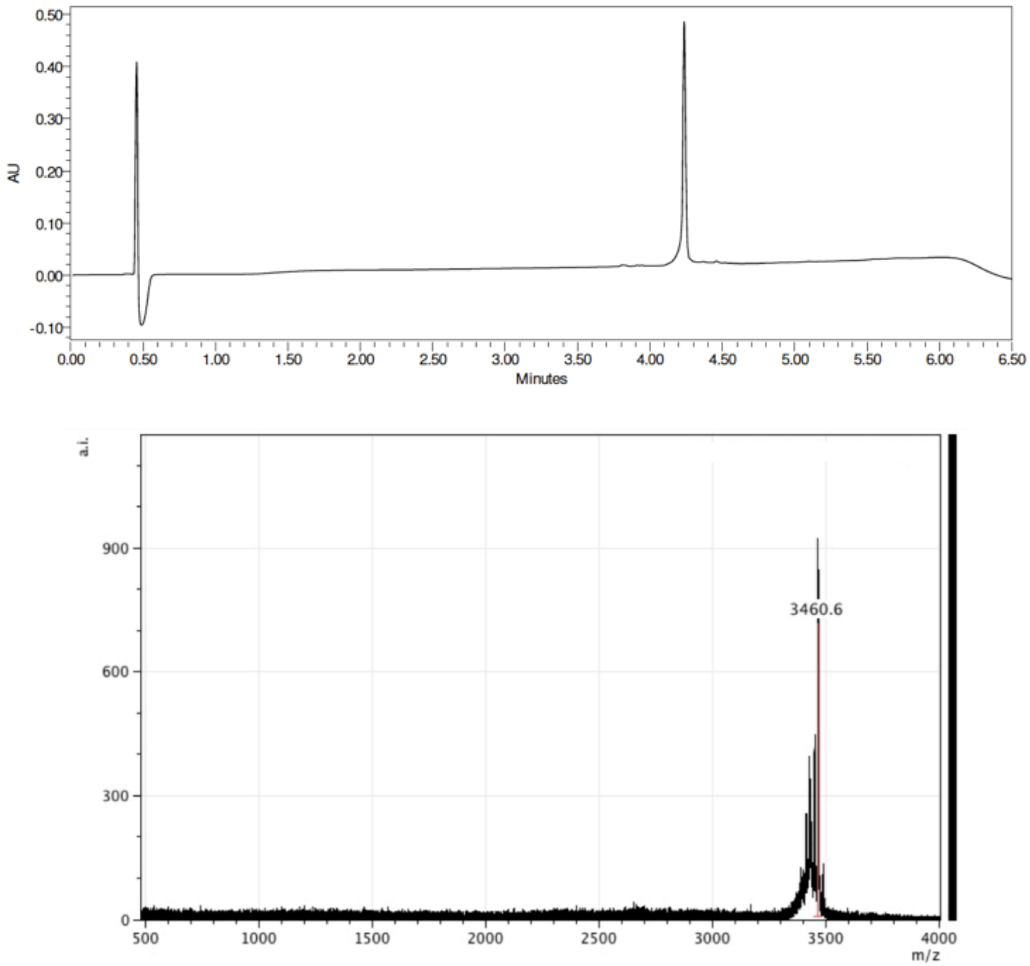


Figure S6E. UPLC trace and MALDI-TOF for peptide **5**. $[M+H]_{calc}$ = 3460.90

2D-NMR data acquisition and NMR structure determination

Each NMR sample was prepared by dissolving lyophilized peptide to 0.2 mM concentration in 9:1 H_2O:D_2O, pH 3.8, 2.5 mM NaOAc-d_3 buffer, with trace amounts of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) and/or dioxane as internal references. Peptide concentrations were determined by mass. Samples were prepared with total volumes of approximately 450 µL in 5 mm NMR tubes. Peptide samples were stable in solution for at least 1 week at 4-5 °C, showing no visible precipitation of peptide, NMR line broadening, or decrease in NMR signal intensity over the entire period of study. Low salt conditions were employed to facilitate observation of weaker NMR signals. For each peptide, the following sets of 1D ^{1}H spectra were superimposable: [1] before and after collection of 2D NMR spectra (COSY, TOCSY, NOESY), and [2] 2 mM peptide in low salt buffer (9:1 H_2O:D_2O, 2.5 mM NaOAc-d_3, pH 3.8) and 2 mM peptide in high salt buffer (9:1 H_2O:D_2O, 100 mM NaOAc-d_3, pH 3.8).

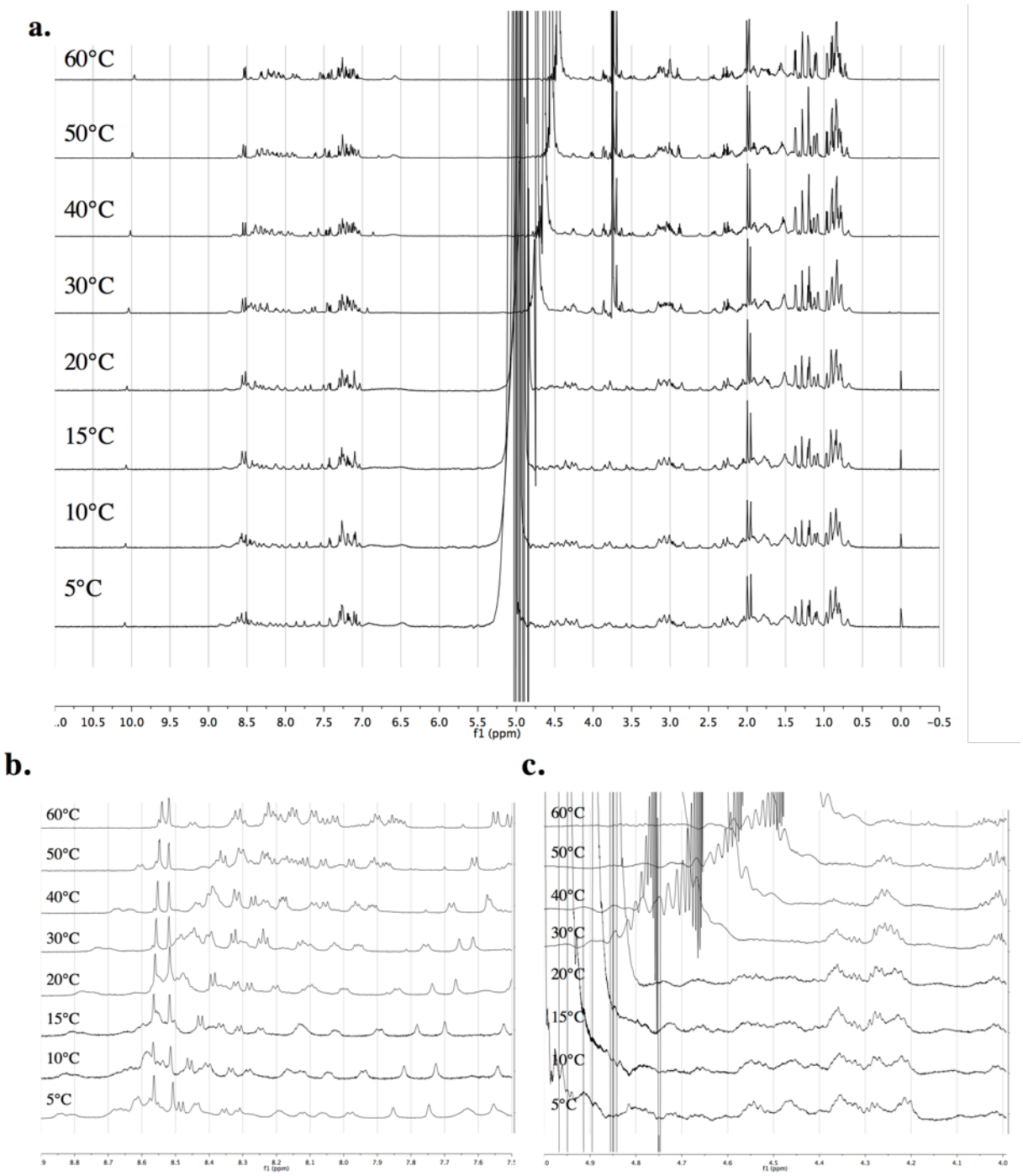


Figure S7. Increased ^{1}H NMR peak dispersion as temperature is decreased from 60 °C to 5 °C, for 0.2 mM peptide **1** in aqueous solution (2.5 mM NaOAc-d_3, pH 3.8, ± DSS or dioxane). Panels (b) and (c) show the Hα and HN regions, respectively.

Spectra were acquired at 5.0 °C on a Bruker Avance 600 MHz spectrometer, equipped with a 5 mm, z-axis gradient, triple resonance, cryogenic probe. The following standard Avance pulse programs were employed: 1D with water suppression using excitation sculpting (zgesgp),[14] phase-sensitive 2D COSY-DQF with WATERGATE using 3-9-19 (cosydfgpph19),[15-17] phase-

sensitive 2D TOCSY with WATERGATE 3-9-19 and mixing using DIPSI-2 (dipsi2gpph19),[16,17] and phase-sensitive 2D NOESY with WATERGATE using water flip-back and 3-9-19 (noesyfpgpph19)[16-18]. Spectral widths of 8013-8418 Hz were used, with collection of 2048 points in the direct (f2) and 256 points in the indirect (f1) dimensions. TOCSY experiments used a mixing time of 100 ms. NOESY experiments used a mixing time of 250 ms. For peptide **1**, increasing the NOESY mixing time to 500 ms did not increase the number of NOEs observed. Note also that variable temperature studies suggested peptide unfolding at higher temperatures, with less peak dispersion (**Figure S7**), less downfield HN and Hα chemical shifts, and fewer NOEs at higher temperatures. Data were processed using NMRPipe.[19] Data were analyzed using Sparky,[20] with employment of sequential assignment procedures to assign chemical shifts of protons (**Tables S2**).[21] Inter-residue NOEs are diagramed in **Figures S8**.

Table S2A. Proton resonances (ppm) for peptide **1.**

		NH	αH	βH		γH		δH		εH		others				
1	**Thr**	8.41	4.35	3.81		1.20										
2	**Glu**	8.51	4.34	1.65	1.37	2.82	2.82									
3	**Ile**	8.62	4.53	2.03		1.28	1.09					0.89 (γ^2H)				
4	**Arg**	8.69	4.62	1.75	1.75	1.55	1.55	3.09	3.09			7.21 (εH)				
5	**Val**	8.38	4.37			0.87	0.87									
6	**CHDA**	3.07 (αH), 2.93 (βH), 2.62 (αH), 2.07 (βH), 1.93 (βH), 1.83 (βH), 1.80 (γH), 1.62 (γH), 1.39 (γH), 1.27 (γH)														
7	**Gly**	8.20	4.84,4.04													
8	**Val**	8.12	4.22	2.19		0.96	0.96									
9	**Thr**	8.68	4.80	4.02		1.09										
10	**Ile**	8.82	4.36	1.79								0.81 (γ^2H)				
11	**Arg**	8.55	4.42	1.77	1.77	1.54	1.54	3.13	3.13			7.25 (εH)	7.11 (εH)			
12	**Met**	8.85	4.56	2.09	1.89	2.32	2.32									
13	**Arg**	8.36	4.20	1.17	1.17	1.45	1.45	2.84	2.84			7.54 (εH)				
14	**D-Ala**	8.45	4.38	1.21												
15	**Ser**	8.41	4.32	3.79	3.79											
16	**His**	8.66	4.67	3.31	3.16			7.27 (δ^2H)				8.56 (ε^1H)				
17	**D-Ala**	8.19	4.23	1.12												
18	**Phe**	8.32	4.78	3.09	2.88			7.29		7.75		7.61 (zH)				
19	**Trp**	8.45	4.72	3.12	2.94			7.11 (δ^1H)				7.56 (ε^3H)	7.17 (z^3H)	7.42 (h^2H)	7.05 (z^2H)	10.09 (ε^1H)
20	**Val**	8.24	4.10	1.91		0.87	0.87									
21	**Gln**	8.58	4.33	2.25	1.93	3.53	3.53			7.21	7.86					
22	**Val**	7.98	4.41	1.95		0.84	0.84									
23	**D-Pro**		4.28	2.24	2.24	2.02	2.02	3.79	3.79							
24	**DADME**	8.06 (NH), 3.56 (αH), 3.52 (αH), 1.31 (CH_3)														
25	**Lys**	8.62	4.49	2.42	2.02	1.53	1.53	1.78	1.78	3.02	3.02	7.63 (zH)				
26	**Glu**	8.56	4.51													
27	**Phe**	8.49	4.55	2.98	2.98			6.50		7.25		7.07 (zH)				
28	**Lys**	8.58	4.46	1.84	1.84	1.41	1.41	1.73	1.53	3.02	3.02	7.63 (zH)				
29	**His**	8.44	4.90	3.10	2.95			7.19 (δ^2H)				8.51 (ε^1H)				
30	**Ac**		1.78													

Table S2B. Proton resonances (ppm) for peptide **2**.

		NH	αH	βH		γH		δH		εH		others				
7	**Ac**		1.52													
8	**Val**	8.58	4.51	2.03		0.93	0.93									
9	**Thr**	8.33	4.24	3.78		1.16										
10	**Ile**	8.39	4.14	1.82		1.17						0.84 (γ^2H)				
11	**Arg**	8.33	4.34	1.18	1.18	1.48	1.48	2.83	2.83			7.54 (εH)				
12	**Met**	8.65	4.66	3.12	3.12	3.25	3.25									
13	**Arg**	8.55	4.56	1.85	1.85	1.61	1.61	2.93	2.93			7.58 (εH)				
14	**D-Ala**	8.37	4.40	1.17												
15	**Ser**	8.27	4.35	3.83	3.83											
16	**His**	8.35	4.69	3.12	3.12			7.25 (δ^2H)				8.55 (ε^1H)				
17	**D-Ala**	8.29	4.23	1.12												
18	**Phe**	8.45	4.56	3.02	3.02			7.11		7.25						
19	**Trp**	8.30	4.71	3.07	2.87			7.11 (δ^1H)						7.15 (h^2H)	7.43 (z^2H)	10.10 (ε^1H)
20	**Val**	8.18	4.08	1.90		0.83	0.83									
21	**Gln**	8.52	4.26	2.31	2.31	1.30	1.19			7.84	7.84					
22	**Val**	8.32	4.13	2.05		0.95	0.95									
23	**D-Pro**		4.37	2.24	2.24	2.05	1.93	3.78	3.78							
24	**DADME**	8.06 (NH), 3.52 (αH)														
25	**Lys**	8.56	4.47	2.24	2.24	1.53	1.43	1.80	1.73	3.01	3.01	7.64 (zH)				
26	**Glu**	8.53	4.45	1.50	1.50	1.80	1.71									
27	**Phe**	8.44	4.90	2.98	2.98			7.04		7.41		7.54 (zH)				
28	**Lys**	8.47	4.25	2.44	2.44	1.76	1.76	1.55	1.55	3.10	3.10	7.19 (zH)				
29	**His**	8.45	4.43	3.07	2.95			7.09 (δ^2H)				8.50 (ε^1H)				
30	**Ac**		1.36													

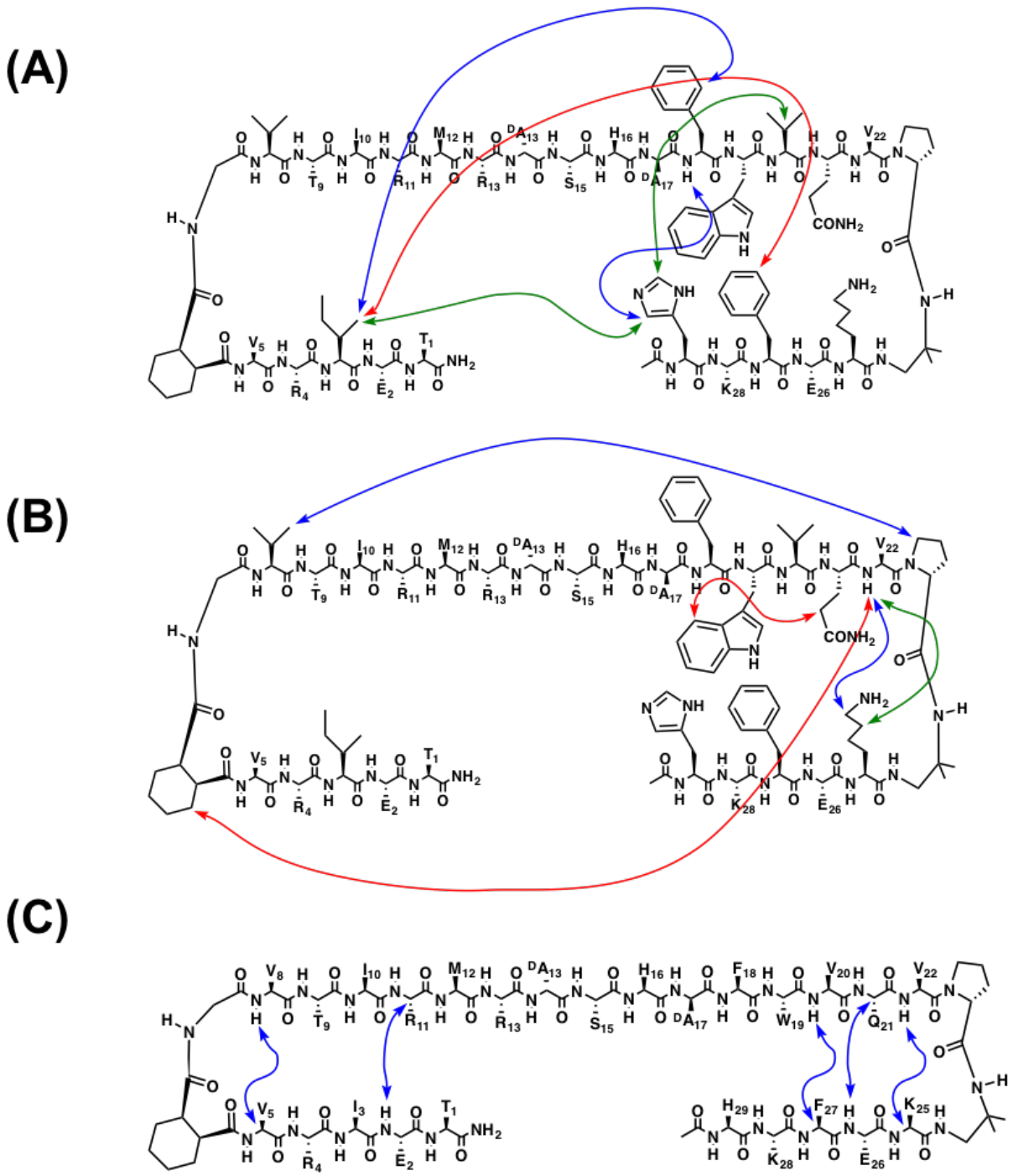


Figure S8. Graphical representation of NOEs observed for peptide **1** between protons on residues that are not adjacent in sequence. These medium- and long-range NOEs are highly diagnostic in terms of folding. (A) A subset of the NOEs observed for peptide **1** in which at least one proton is on a side chain. This subset features NOEs that include protons on the side chains of Ile-3, Phe-18, Val-20, Phe-27 and His-30. Different colors are used to help the reader follow individual arrows. (B) Another subset of the NOEs observed for peptide **1** in which at least one proton is on a side chain. This subset features NOEs that include protons on the side chains of the N-to-N linker (cyclohexyl ring), Val-8, Trp-19, Gln-21, Pro-23 and Lys-25. Different colors

are used to help the reader follow individual arrows. Collectively parts A and B of this figure show all NOEs between residues that are not adjacent in sequence and involving at least one side chain proton. (C) Inter-residue CαH - HN NOEs observed for peptide **1**. Collectively, parts A-C show all NOEs that could be assigned for **1** involving protons on residues that are not adjacent in sequence. All of these NOEs are consistent with a single compact conformation (or family of related conformations) that corresponds to the mini-amyloid design goal (V. Kung, Ph.D. thesis, University of Wisconsin-Madison, 2016).

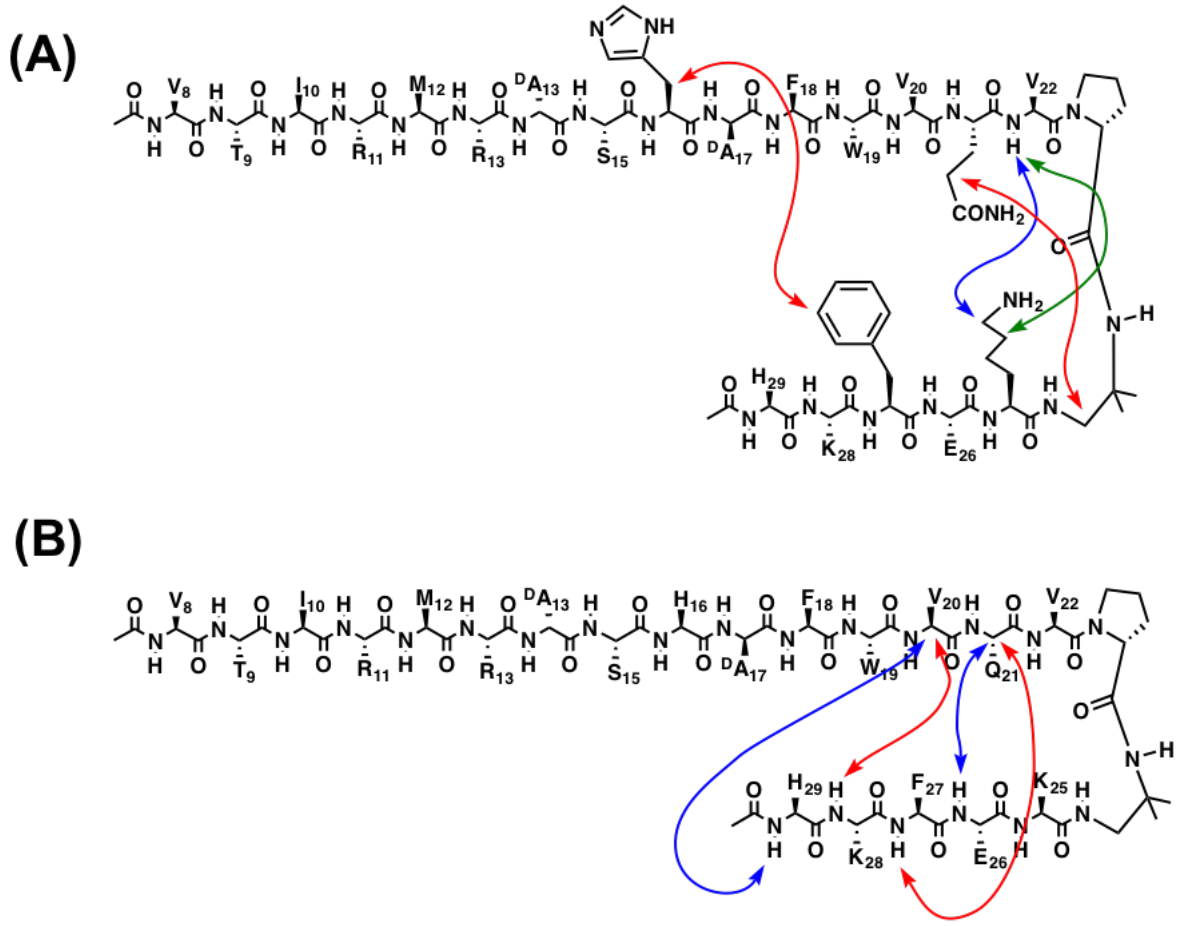


Figure S9. Graphical representation of NOEs observed for peptide **2** between protons on residues that are not adjacent in sequence. (A) NOEs observed for peptide **2** between protons from residues that are not adjacent in sequence, in which at least one proton is on a side chain. Different colors are used to help the reader follow individual arrows. (B) Inter-residue CαH - HN NOEs observed for peptide **2**. Different colors are used to help the reader follow individual arrows. It seems unlikely that all of these NOEs could arise from a single parallel β-sheet H-bonding pattern.

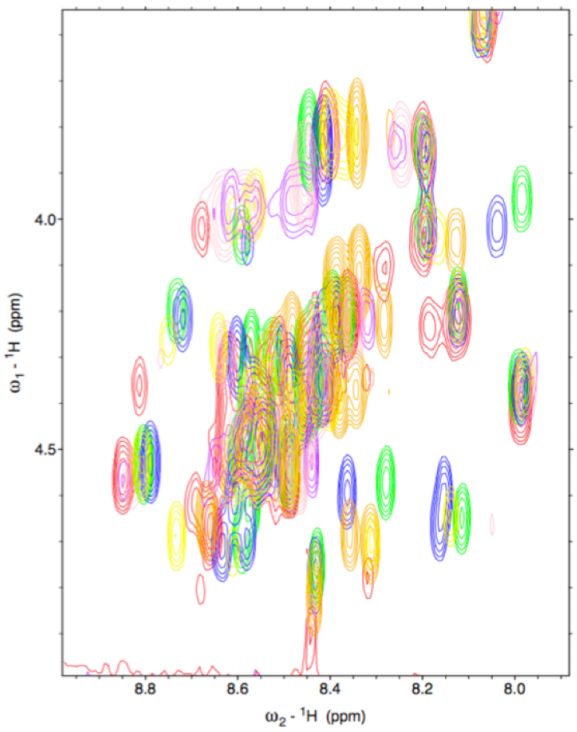


Figure S10. Overlay of Hα-HN regions of 2-D TOCSY spectra for peptide **1** (red), **2** (orange), **3** (green), **4** (blue) and **5** (yellow). Also shown are data for two other variants of **1**: replacement of β-arc residues ^{D}Ala-14, His-16 and ^{D}Ala-17 with Gly (pink), and replacement of all five β-arc residues, Arg-13 through ^{D}Ala-17, with Gly (purple). More peak dispersion and more downfield chemical shifts are observed for Hα and HN for peptide **1** relative to the other peptides. Spectra were taken at 5.0 °C, with 0.2 mM peptide, in 9:1 H_2O:D_2O, 2.5 mM NaOAc-d_3, pH 3.8.

Circular Dichroism (CD).

Circular dichroism can be used to assess protein folding in solution, with the association of α-helix (208 nm, 222 nm) and β-sheet (218 nm) secondary structures with characteristic minima in CD spectra.[21] CD measurements were taken with an AVIV Model 420 Circular Dichroism Spectrometer, using 1.00 nm bandwidths and 10.000 sec averaging times, in 1.00 mm quartz cuvettes. For each sample measurement, the corresponding buffer blank measurement was subtracted, and the baseline CD signal at 320 nm was set to zero. Peptide solutions were prepared in aqueous solution with low (2.5 mM NaOAc, pH 3.8) or high (2.5 mM NaOAc, 100 mM NaCl, pH 3.8) concentrations of a charge-shielding salt. Peptide concentrations were determined by mass, and verified with UV absorption measurements. CD spectra were obtained at 20.0 °C (293.2 K).

Effects of added NaCl.

Introduction of 100 mM NaCl caused a profound change in the CD spectrum of peptide **1** at 20°C (Figure S11A). The residue-normalized CD signal became significantly more intense in the presence of added salt, and the position of the minimum shifted slightly to higher wavelength. Diffusion-ordered NMR spectroscopy (DOSY) and analytical ultracentrifugation

(AUC) measurements in the presence of 100 mM NaCl (2.5 mM NaOAc, pH 3.8) indicated that the added salt causes peptide **1** to self-associate. However, the AUC data could not be interpreted in terms of a single species and instead suggested a mixture of aggregated forms. These solutions showed no sign of preciptation over the course of several days.

Addition of NaCl to aqueous solution is known to screen Coulombic repulsions between solutes of like charge, such as **1** (an oligocation at pH 3.8). In addition, NaCl enhances the hydrophobic effect. Both of these effects could underlie the ability of added salt to induce self-association of **1** in aqueous solution. Because the CD minimum near 220 nm grows more intense in the presence of 100 mM NaCl (Figure S11A), we conclude that aggregation of **1** leads to enhanced formation of β-sheet secondary structure, and we speculate that there is enhanced population of an amyloid-like state.

Figure S11B shows that peptides **4** (two ^{D}Ala→Asn modifications in the β-arc) and **5** (two ^{D}Ala→Gly modifications in the β-arc) develop strong CD minima near 220 nm in the presence of 100 mM NaCl, even though neither of these peptides displays a distinct minimum in this region in aqueous buffer lacking salt (Fig. 4). We speculate that the enhanced population of β-sheet secondary structure indicated by CD for both **4** and **5** in the presence of NaCl arises because the salt has induced aggregation and formation of an amyloid like state, analogous to the behavior of **1**.

Fig. S11C shows the effects of temperature on the per-residue CD signatures of peptides **1**, **4** and **5** in the presence of 100 mM NaCl in buffer was monitored at 220 nm. In each case, the CD intensity displays a cooperative transition over the observed temperature range (2° to 98°C); these transitions were fully reversible. The data for **4** and **5** display plateau regions at the lowest and highest temperatures, which suggest a fully folded/assembled state at the lowest temperatures and a fully unfolded/non-assembled state at the highest temperatures. The mid-points of these transitions ("melting temperatures") lie in the 45°-50°C range. The data for **1** show a low-temperature plateau similar to that seen for **4** and **5**; the data seem to approach a high-temperature plateau similar to that evident for **4** and **5**. Peptide **1** manifests a less cooperative transition (shallower slope at the midpoint) relative to **4** or **5**, but the data suggest that the assembly formed by **1** is more stable than those formed by **4** or **5** (higher mid-point temperature for **1**).

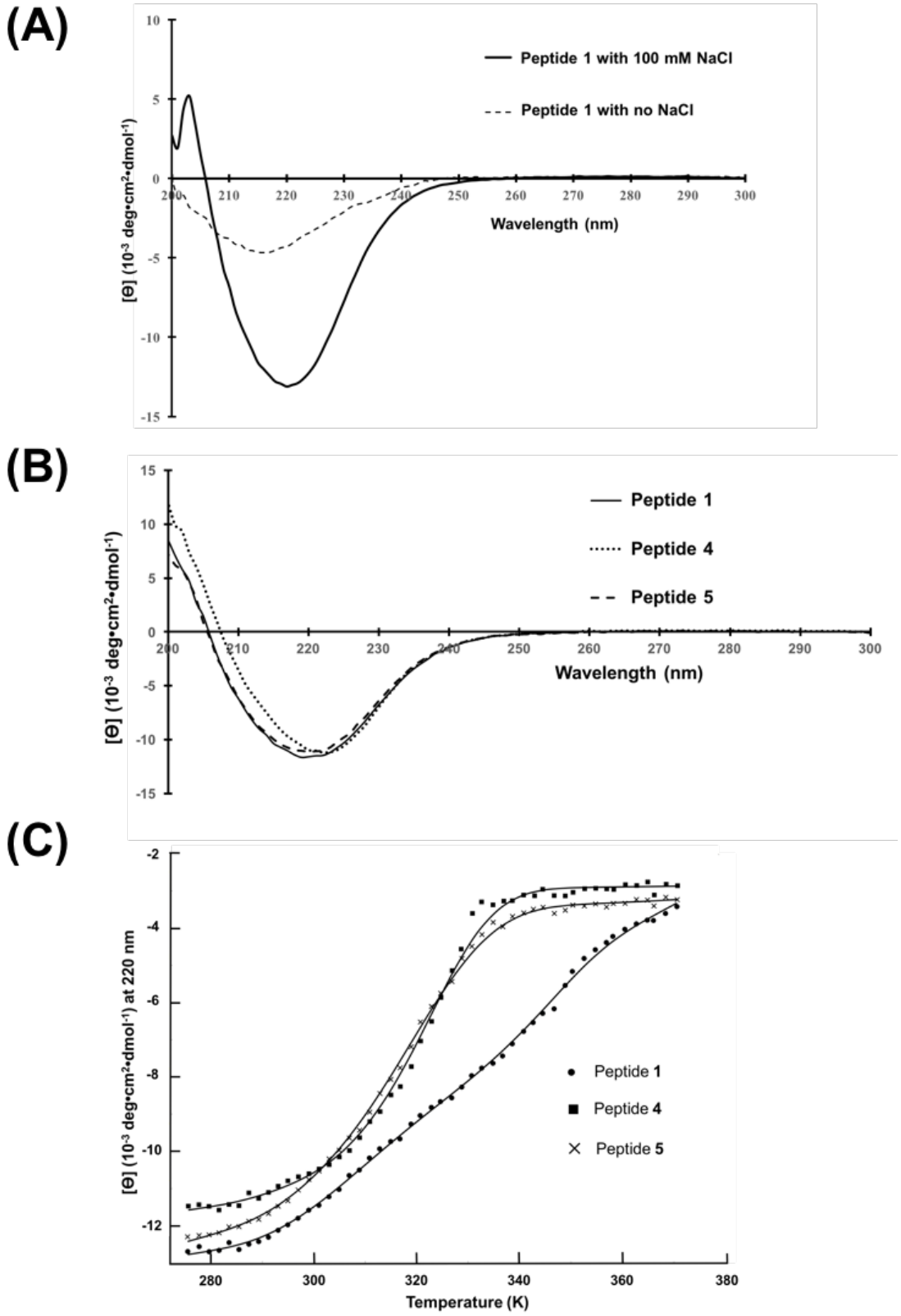


Figure S11. Effects of adding 100 mM NaCl on the per-residue circular dichroism of selected peptides. (A) Comparison of 0.2 mM **1** in aqueous 2.5 mM NaOAc, pH 3.8, 20º C (dashed line)

with the sample prepared under similar conditions but containing 100 mM NaCl (solid line). (B) Comparison of the per-residue circular dichroism of 0.1 mM peptide in aqueous 2.5 mM NaOAc, 100 mM NaCl, pH 3.8, 20º C: **1** (solid line), **4** (dotted line) and **5** (dashed line). (C) Mean residue ellipticity at 220 nm for 0.1 mM peptide in aqueous 2.5 mM NaOAc, 100 mM NaCl, pH 3.8, as a function of temperature: **1** (dots), **4** (squares) and **5** (crosses). Solid lines represent non-linear regression curve-fits for the thermal melt data.

Sedimentation Equilibrium Analytical Ultracentrifugation (AUC).

Our AUC experimental design and data analysis methods were reported in greater detail elsewhere.[13] Sedimentation equilibrium data were collected using a Beckman Coulter Model XL-A Analytical Ultracentrifuge at 4 °C, for peptide **1**. Peptide solutions were prepared by simple dissolution in buffer (2.5 mM NaOAc, pH 3.8, with peptide concentrations of 0.09, 0.06, and 0.03 mM) and used without further manipulation. Initial peptide concentrations were determined from UV absorption spectra of samples diluted into 6 M GdmCl, using an extinction coefficient (ε) of 5690 M^{-1} cm^{-1}.[22] Gradients were monitored at the wavelength of maximal absorbance (276 nm for peptide **1**). Data were collected at rotor speeds of 10k, 15k, 23k, 32k, 42k. For each rotor speed, we assumed samples had reached equilibrium when gradients collected at least 3 hours apart were superimposable. After data collection at the highest rotor speed, gradients at one of the lower speeds were measured again and found to be superimposable with original measurements, indicating that there had been no irreversible loss of material during the course of the experiment.

The equilibrium data were analyzed using software written by Darrell R. McCaslin, in a manner analogous to previously-described analysis methods.[23] In the analysis, the solvent density (ρ) at 4 °C was computed using density increment functions to be 1.001 g/mL.[24] The partial specific volume ($\bar{v}$) of peptide **1** (0.736 cm^3 g^{-1}) was calculated using consensus values for the natural amino acid residues, and corrections for acetyl, carboxamide, DADME, and CHDA groups.[25,26] The molecular weight (M_S) of peptide **1** (3489.2 Da) was calculated based on the peptide sequence, including non-natural linkers and termini.

Plots of the logarithm of the measured absorbance as a function of the squared radial distance from the center of rotation were linear, at all speeds and for all loading concentrations (**Figure S12**). Linearity in such plots is indicative of the presence of a single molecular weight entity.

The final estimate of the weight-average molecular weight (M_P, or MW_{obs}), reported in **Table S3**, was obtained from global fits of the data at the varied loading concentrations and speeds to the distribution of a single macromolecular species. Inclusion of a term for a second macromolecular species worsened fits.

The AUC experiments were designed to examine the peptide association state at low-salt buffer conditions similar to those employed for NMR structural studies. For peptide **1**, the weight-average molecular weight (MW_{obs}) was much smaller than the molecular weight based on the sequence (M_s). Furthermore, the apparent weight decreased with increasing concentration. These results suggest that at the low-salt conditions studied, the peptide is monomeric, and not forming dimers or larger molecular weight aggregates. We conclude that the observed $MW_{obs}/MW_{calc} \leq 1$ indicates a monomeric state, with electrostatic nonideality accounting for $MW_{obs}/MW_{calc} << 1$.

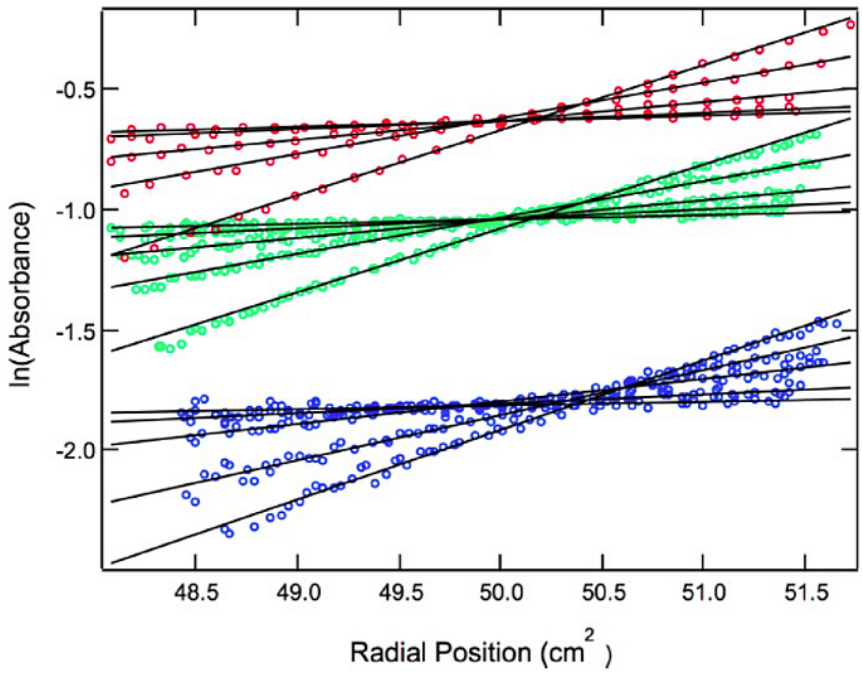


Figure S12. Equilibrium AUC data for peptide **1** in 2.5 mM NaOAc, pH 3.8. The data shown are for speeds of 10k, 15k, 23k, 32k, 42k, at initial loading concentrations of 0.03 mM (blue), 0.06 mM (green), and 0.09 mM (red).

Table S3. Sedimentation equilibrium results for peptide **1**. Weight-average molecular weight (MW_{obs}) from fits to a single-species model, reported with least-squares fitting error in the fitted parameter.

buffer	MW_{obs}/M_S	MW_{obs} (Da)
2.5 mM NaOAc, pH 3.8	0.693 ± 0.004	2417

NMR diffusion experiments

NMR experiments measuring peptide diffusion were performed. Samples were prepared by dissolving lyophilized peptide to ~2 mM concentration, with 0.6 mM dioxane, in aqueous buffer (9:1 H_2O:D_2O, 2.5 mM d_3-acetate, pH 3.8). Serial dilutions were made of these peptide solutions into buffer, to yield 0.2 mM and 0.02 mM peptide solutions. Spectra were acquired on a Bruker Avance 600 MHz spectrometer, equipped with a 5 mm, z-axis gradient, triple resonance, cryogenic probe. Self-diffusion rates were measured using a pulsed field gradient stimulated spin echo (PFGSTE) technique.[27] Spectra were acquired at 5.0 °C, with 75 pulsed field gradient (PFG) strengths in the 5-60 G/cm range, and acquisition of 16-240 transients per gradient experiment. Gradient pulses were applied for 5 ms, with a diffusion delay of 100 ms.

Data were analyzed using the variable gradient least square fitting routines in NMRPipe,[19] and peak intensities were fit with a single exponential decay function to determine translational self-diffusion constants. For each condition studied, the self-diffusion constant of the peptide represented an average of values obtained from the intensities of at least 4 separate peaks, including peaks in the aromatic and aliphatic regions. The hydrodynamic radius (R_H) of

the peptide was extracted by comparison of peptide diffusion data with that of an internal standard, 1,4-dioxane (R_H = 2.12 Å).[28]

The hydrodynamic radii derived from NMR diffusion experiments (obs R_H) are reported in **Table S4**. The predicted hydrodynamic radii of the monomeric peptides were calculated using the atomic-level (*a* = 2.84 Å) shell-model of HYDROPRO 10,[29] and are in good agreement with the hydrodynamic radii derived from NMR diffusion datas (**Table S4**). Furthermore, variable concentration NMR diffusion analysis of the peptides shows little to no increase in R_H values as peptide concentrations are increased from 0.02 mM to 0.2 mM. Also, no changes in chemical shifts, peak shapes, or splitting were observed on 1D ^{1}H NMR as peptide concentrations were increased from 0.02 mM to 0.2 mM. These results support the conclusion that peptides are monomeric (no peptide aggregation occurs) under the conditions used to obtain NMR structures of the peptides.

Table S4. Hydrodynamic radii of peptides, derived from NMR diffusion experiments (obs R_H) or HYDROPRO 10 calculations for monomeric peptides (calc R_H). All peptides are in 2.5 mM NaOAc, pH 3.8 buffer.

Peptide	**Obs R_H (Å)** At 0.2 mM peptide	**Obs R_H (Å)** At 0.02 mM peptide	**Calc R_H (Å)**
1	9.3 ± 0.6	8.3 ± 1.7	13.5
2	9.8 ± 0.4	9.5 ± 0.8	15.2
3	17.6 ± 0.4	---	---
4	14.0 ± 1.1	---	---
5	11.4 ± 0.4	8.4 ± 1.3	---

References

(1) Hennetin, J.; Jullian, B.; Steven, A. C.; Kajava, A. V. *J. Mol. Biol.* **2006**, *358*, 1094.

(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nuc. Acids Res.* **2000**, *28*, 235.

(3) Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G. *Nuc. Acids Res.* **2011**, *39*, D411.

(4) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.

(5) Wang, G. L.; Dunbrack, R. L. *Bioinformatics* **2003**, *19*, 1589.

(6) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188.

(7) Magrane, M.; Consortium, U. *Database : the journal of biological databases and curation* **2011**, *2011*, bar009.

(8) Hovmoller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr. D* **2002**, *58*, 768.

(9) Deane, C. M.; Allen, F. H.; Taylor, R.; Blundell, T. L. *Protein Eng.* **1999**, *12*, 1025.

(10) Srinivasan, N.; Anuradha, V. S.; Ramakrishnan, C.; Sowdhamini, R.; Balaram, P. *Int. J. Pept. Prot. Res.* **1994**, *44*, 112.

(11) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(12) Hubbard, S. J.; Campbell, S. F.; Thornton, J. M. *J. Mol. Biol.* **1991**, *220*, 507.

(13) Kung, V. M.; Cornilescu, G.; Gellman, S. H. *Angew. Chem. Int. Ed.* **2015**, *54*, 14336.
(14) Hwang, T. L.; Shaka, A. J. *J. Magn. Reson. Ser. A* **1995**, *112*, 275.
(15) Derome, A. E.; Williamson, M. P. *J. Magn. Reson.* **1990**, *88*, 177.
(16) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661.
(17) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *J. Magn. Reson. Ser. A* **1993**, *102*, 241.
(18) Lippens, G.; Dhalluin, C.; Wieruszeski, J. M. *J. Biomol. NMR* **1995**, *5*, 327.
(19) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(20) Goddard, T. D.; Kneller, D. G. Sparky 3, University of California - San Francisco.
(21) Greenfield, N. J. *Nature Protocols* **2006**, *1*, 2876.
(22) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.
(23) Laue, T. M. *Meth. Enzymol.* **1995**, *259*, 427.
(24) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; The Royal Society of Chemistry: Cambridge, U.K., 1992, p 90.
(25) Perkins, S. J. *Eur. J. Biochem.* **1986**, *157*, 169.
(26) Durchschlag, H.; Zipper, P. *Prog. Colloid Polym. Sci.* **1994**, *94*, 20.
(27) Tanner, J. E. *J Chem Phys* **1970**, *52*, 2523.
(28) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424.
(29) Ortega, A.; Amoros, D.; de la Torre, J. G. *Biophys. J.* **2011**, *101*, 892.
(30) Kuo, L. H.; Li, J. H.; Kuo, H. T.; Hung, C. Y.; Tsai, H. Y.; Chiu, W. C.; Wu, C. H.; Wang, W. R.; Yang, P. A.; Yao, Y. C.; Wong, T. W.; Huang, S. J.; Huang, S. L.; Cheng, R. P. *Biochemistry* **2013**, *52*, 7785.
(31) (a) Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan. S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. *Nature* **2015**, *525*, 486. (b) Laganowsky, A.; Liu, C.; Sawaya, M. R.; Whitelegge, J. P.; Park, J.; Zhao, M.; Pensalfini, A.; Soriaga, A. B.; Landau, M.; Teng, P.K.; Cascio, D.; Glabe, C.; Eisenberg, D. *Science* **2012**, *335*, 1228. (c) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, Ø. A.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447*, 453.