Supporting information for

Tertiary structural motif sequence statistics enable facile prediction and design of peptides that bind anti-apoptotic Bfl-1 and Mcl-1

Vincent Frappier[1¶], Justin M. Jenson[1¶], Jianfu Zhou[2], Gevorg Grigoryan[2,3,4*], Amy E. Keating[1,5,6*]

[1] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[2] Department of Computer Science, Dartmouth College, Hanover, New Hampshire, United States of America
[3] Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, New Hampshire, United States of America
[4] Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, United States of America
[5] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[6] Koch Center for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
* Corresponding author
Email: gevorg.grigoryan@dartmouth.edu (GG) or keating@mit.edu (AK)
[¶] These authors contributed equally and are listed alphabetically

Supporting Information Tables

Table S1. Natural BH3 sequences composing the sequence logo in Fig. 2A

	Sequence
Name	--|--2---|--3---|--4---
	efgabcdefgabcdefgabcdefg
PUMA	EQWAREIGAQLRRMADDLNAQYERRR
BIM	MRPEIWIAQELRRIGDEFNAYYARRV
NOXA	AELEVECATQLRRFGDKLNFRQKLLN
BAD	LWAAQRYGRELRRMSDEFVDSFKKGL
BAK	SSTMGQVGRQLAIIGDDINRRYDSEF
BAX	DASTKKLSECLKRIGDELDSNMELQR
HRK	SSAAQLTAARLKALGDELHQRTMWRR
BMF	HQAEVQIARKLQCIADQFHRLHVQQH
BIK	MEGSDALALRLACIGDEMDVSLRAPR
BID	EDIIRNIARHLAQVGDSMDRSIPPGL
MULE	GVMTQEVGQLLQDMGDDVYQQYRSLT
BECLIN	GGTMENLSRRLKVTGDLFDIMSGQTD
BOK	PGRLAEVCAVLLRLGDELEMIRPSVY

Table S2. Interaction prediction performance by template
See accompanying Excel file.

Table S3. Pairwise comparisons of binding interface structures (RMSD in Å)
See accompanying Excel file.

Table S4. Templates, dTERMen version and constraints for design calculations
See accompanying Excel file.

Table S5. X-ray data collection and refinement statistics

	Bfl-1:dF1 PDB ID 6MBB	**Bfl-1:d.F4 PDB ID 6MBC**	**Mcl-1:d.M1 PDB ID 6MBD**	**Mcl-1:dM7 PDB ID 6MBE**
Data Collection				
Space Group	P 1 21 1	P 1 21 1	P 21 21 21	P 32 2 1
Cell parameters				
a, b, c	43.223 42.92 47.718	43.466 42.905 46.666	64.792 69.733 84.853	80.758 80.758 57.95
α, β, γ	90 115.957 90	90 114.206 90	90 90 90	90 90 120
Rmeas	0.078 (0.399)	0.078 (0.43)	0.137 (0.981)	0.122 (.0698)
Rpim	0.029 (0.169)	0.036 (0.259)	0.047 (0.398)	0.045 (0.338)
Mean I/σ(I)	30.92 (2.9)	22.4 (2.0)	15.6 (0.655)	18.32 (1.49)
Completeness (%)	91.49 (77.84)	94.60 (74.42)	95.81 (67.33)	97.31 (78.96)
Redundancy	6.5 (4.3)	4.1 (1.9)	7.7 (3.9)	6.7 (3.4)
Refinement				
Resolution (Å)	38.86 - 1.587 (1.644 - 1.587)	42.56 - 1.752 (1.815 - 1.752)	24.3 - 1.945 (1.994 - 1.945)	29.94 - 2.247 (2.327 - 2.247)
Unique Reflections	19602 (1654)	13573 (1193)	25709 (1922)	9363 (833)
Rwork/Rfree	0.1651/0.1957 (0.2372/0.2628)	0.1832/0.2121 (0.2622/0.3308)	0.1957/0.2399 (0.3329/0.3700)	0.1919/0.2211 (0.2493/0.2586)
Number of non-hydrogen atoms	1623	1541	3047	1482
Average B-factors	30.2207	42.16	35.81	49.05
Rmsd				
Bond lengths (Å)	0.006	0.003	0.006	0.002
Bond angles (°)	0.77	0.535	0.742	0.409

Values in parentheses are for the highest-resolution shell.

Table S6. Physicochemical proprieties of designed peptides

Name	Seq	Helical Content (%)	Hydrophobicity	Z
M01	APKEKEVAETLRKIGEEINEALK	3.06	0.0313	-1
M05	APKEKEVARTLIKIGEEINEALK	0.65	0.13739	0
M06	APYLEQVARTLLHIGMEINEALR	2.56	0.48217	-1
M02	APYLEQVARTLRKIGEEINEALR	4.22	0.23435	0
M09	DIEQEIAEALKEVADELSKAIED	9.73	0.12478	-7
M03	DKTLEEIARELAKLAEEIDKEI	62.13	0.10591	-4
M07	DKTLEEIARELLKLALEIDKEI	69.68	0.27545	-3
M04	DKTLEEIARWLARLALEIDKEI	68.85	0.34273	-2
M10	DVVLSVAETLRELADRLYEEINT	28.21	0.33043	-4
F07	SLLEKLAEELAQLADELNKKFEK	47.03	0.17957	-2
F03	SLLEKLAEELRQLADELNKKFEK	59.08	0.12217	-1
F08	SLLEKLAEYLAQMGDEINKKYVK	14.5	0.26435	0
F04	SLLEKLAEYLRQMADEINKKYVK	54.9	0.22043	1
F06	SYIDKIADLIDKVVEEINSKLE	1.99	0.29227	-3
F02	SYIDKIADLIRKVAEEINSKLE	28.99	0.24	-1
F05	SYVDKIADLMKKVAEKINSDLT	16.08	0.22364	0
F01	SYVDKIADVMREVAEKINSDLT	6.83	0.21682	-2

Table S7. Median signals for FACS binding experiments of expression positive cells.

See accompanying Excel file.

Supporting Information Figures

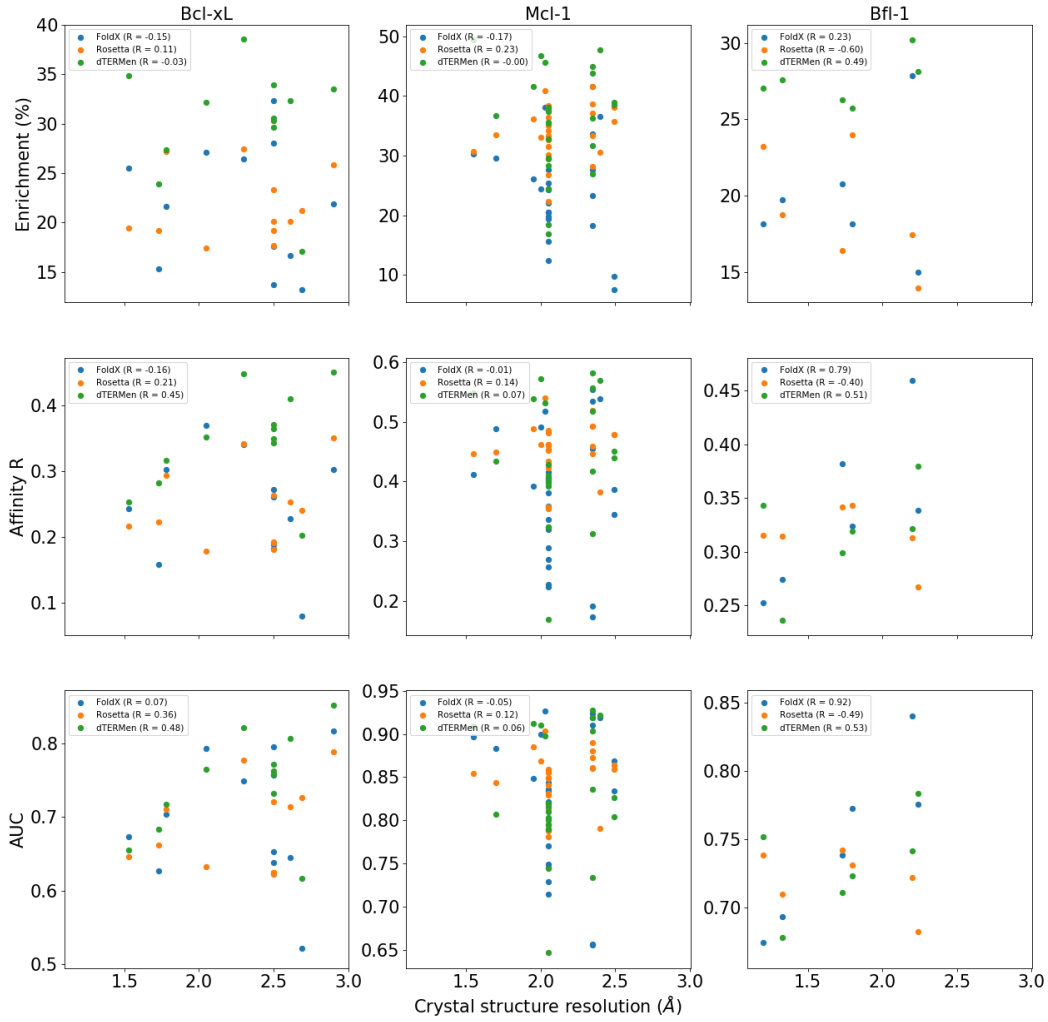

Figure S1. Relationship between crystal structure resolution and benchmark performance. Each point corresponds to a template used for prediction by FoldX (blue), Rosetta (orange) or dTERMen (green). Plots show results for different performance metrics and different target proteins, as described in the text. The Pearson R values describing the correlation between method performance and crystal-structure resolution (Å) are given in the inset and show that higher resolution structures don't provide better prediction performance.

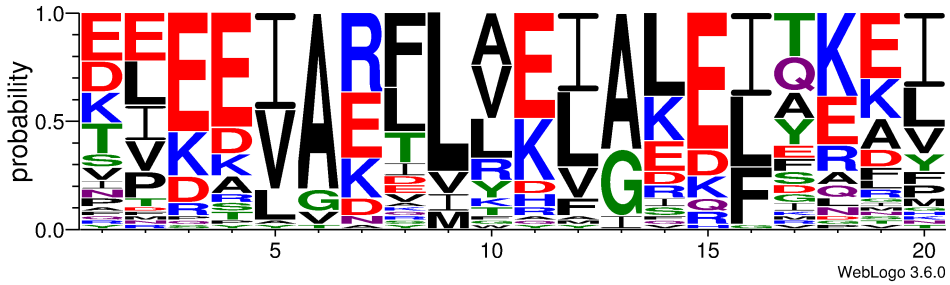

Figure S2. Diversity of sequences designed using different Bcl-x_L, Bfl-1 and Mcl-1 templates. Sequences were designed using same protocol described in the methods, without any constraints, on all templates used for the prediction benchmark. Positions 3a and 3f (conserved as Leu and Asp in native BH3 sequences) are positions 9 and 14, respectively, in the weblogo annotation.

Figure S3. Structure fragment used for analysis of helix-capping residue preferences. Purple color indicates the helix-capping fragment used to examine whether the preference for Asn/Asp at 4b in BH3 domains is echoed in similar structural environments of other proteins. Shown in green and cyan are the backbones of human Bfl-1 and peptide FS2, respectively (from PDB entry 5UUK). Magenta denotes the helix-capping motif used for the structure mining described in the main text, with the sidechain of the 4b position (Asn) shown making a helix-capping interaction. To define the motif, additional residues were selected on the Bcl-2 domain side of the interface so as to stipulate a match to the end of a helix (as opposed to, for example, matches emerging from center regions of helices, where a capping interaction would not be possible and where sequence preferences would be expected to be quite different). Equivalent motifs were defined in all other templates to cover the corresponding residues.

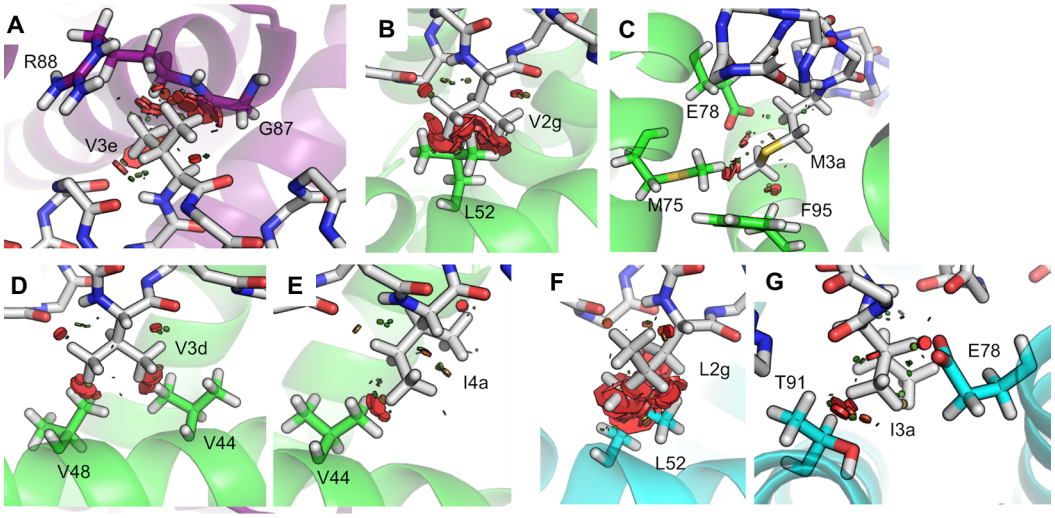

Figure S4. Modeled side-chain clashes for dTERMen designs on FS2:Bfl-1 template 5UUK. A representative sample of residues from the dTERMen designs that were predicted to clash in all rotamers in model structures. The backbone-dependent rotamers with the least predicted clashing, as assessed by PyMol, are shown. (A) Val at position 3e of dF6 (gray) is predicted to clash with Arg 88 and Gly 87 of Bfl-1 (purple). (B) Val at position 2g in dF1 (gray) is predicted to clash (red disks) with Leu 52 of Bfl-1 (green). (C) Met at position 3a in dF1 (gray) and dF5 is predicted to have minor clashes with Met 75, Glu 78, and Phe 95 of Bfl-1 (green). (D) Val at position 3d in dF1 (gray), dF2, dF5 and dF6 is predicted to clash with Val 44 and Val 48 of Bfl-1 (green). (E) Ile at position 3e in dF1 (gray), dF2, dF5 and dF6 is predicted to clash with Val 44 of Bfl-1 (green). (F) Leu at position 2g of dF2 (gray), dF5, and dF6 is predicted to clash with Leu 52 of Bfl-1 (blue). (G) Ile at position 3a of dF2 (gray) and dF6 is predicted to have minor clashes with Glu 78 and Thr 91 of Bfl-1 (blue).

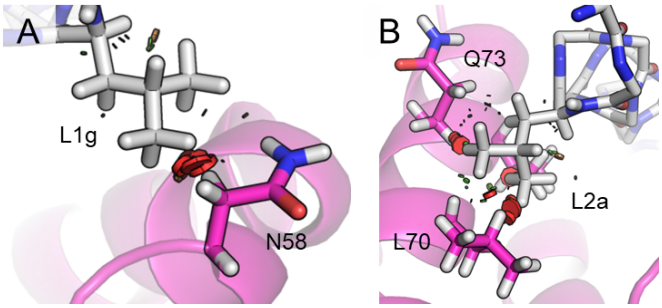

Figure S5. Modeled side-chain clashes for dTERMen designs on PUMA:Bfl-1 template 5UUL. A representative sample of residues from the dTERMen designs that were predicted to clash for all rotamers in model structures. The backbone-dependent rotamers with the least predicted clashing, as assessed by PyMol, are shown. (A) Leu at position 1g in dF3 (gray), dF4, dF7, and dF8 is predicted to clash (red disks) with Asn 58 of Bfl-1 (pink). (B) Leu at position 2a in dF3 (gray), dF4, dF7, and dF8 is predicted to clash with Gln 73 and Leu 70 of Bfl-1 (pink).

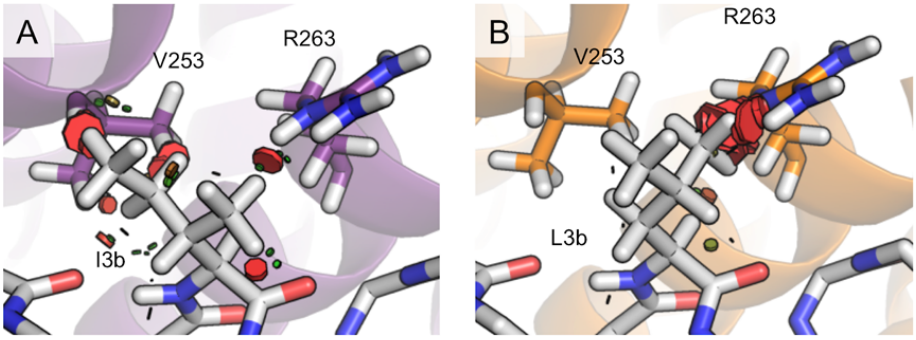

Figure S6. Modeled side-chain clashes for dTERMen designs on BIM:Mcl-1 template 2PQK. A representative sample of residues from the dTERMen designs that were predicted to clash for all rotamers in model structures. The backbone dependent rotamers with the least predicted clashing, as assessed by PyMol, are shown. (A) Ile at position 3b in dM5 (gray) is predicted to clash (red disks) with Val 253 and Arg 263 of Mcl-1(purple). (B) Leu at position 3b of dM5 (gray) is predicted to clash with Val 253 and Arg 263 of Mcl-1 (orange).

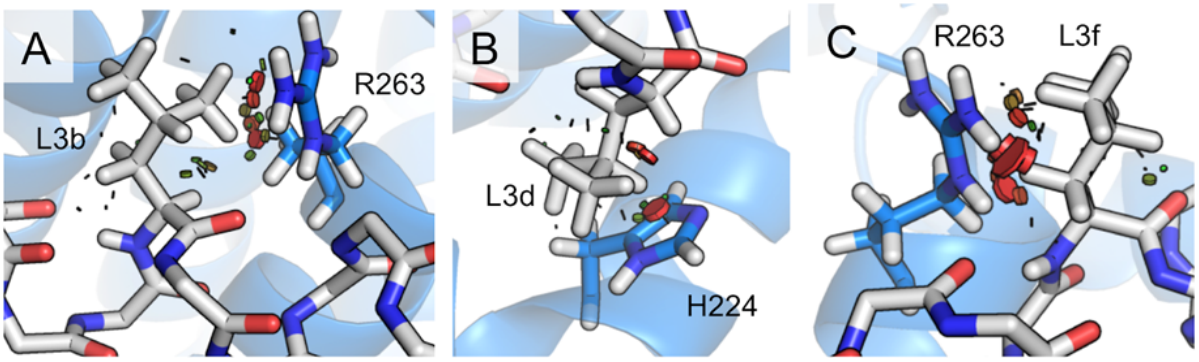

Figure S7. Modeled side-chain clashes for dTERMen designs on BID-MM:Mcl-1 template 5C3F. A representative sample of residues from the dTERMen designs that were predicted to clash for all rotamers in model structures. The backbone-dependent rotamers with the least predicted clashing, as assessed by PyMol, are shown. (A) Leucine at position 3b of dM7 (gray) is predicted to clash (red disks) with R263 of Mcl-1 (blue). (B) Leucine at position 3d of dM7 (gray), dM3, and dM4 is predicted to clash with H224 of Mcl-1 (blue). (C) Leucine at position 3f of dM7 (gray), dM3, and dM4 is predicted to clash with H224 of Mcl-1 (blue).

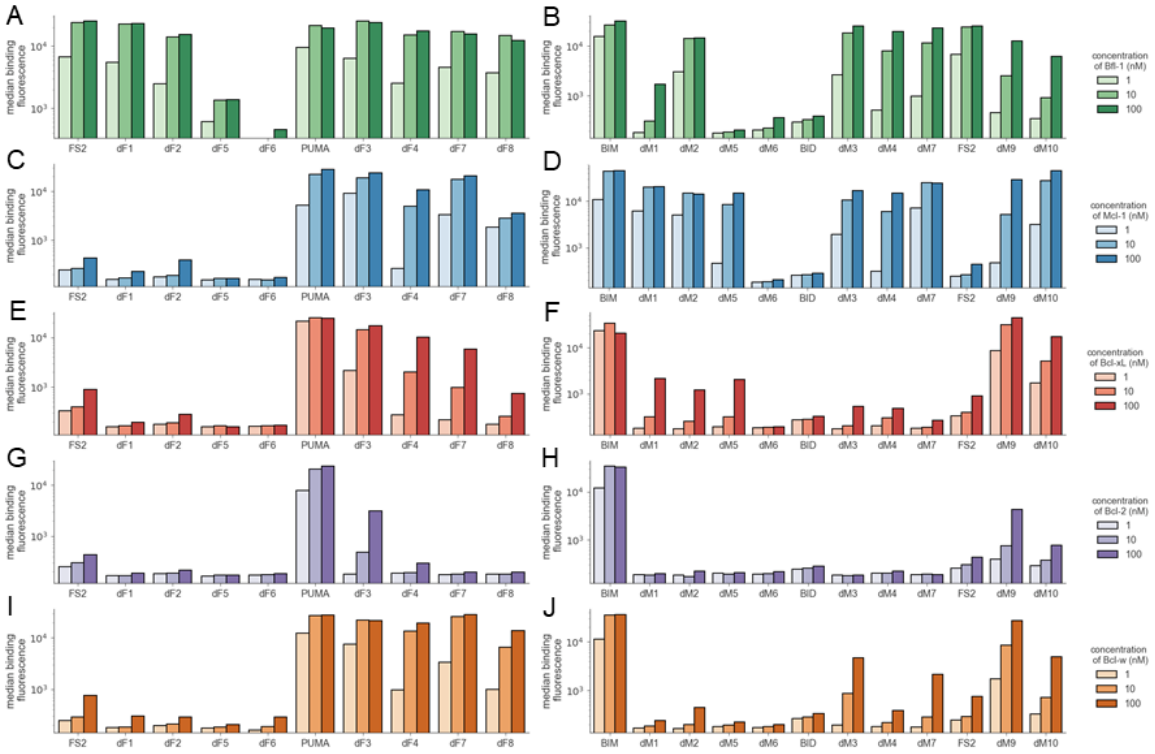

Figure S8. Binding of sequences to on- and off-target Bcl-2 family proteins. The designs from this study were assayed using FACS to quantify Bcl-2-protein binding to yeast-surface displayed peptides. Shown here is the median fluorescence binding signal of each peptide in the presence of 1, 10, or 100 nM of the target proteins Bfl-1 (A, B), Mcl-1 (C, D), Bcl-x_L (E, F), Bcl-2 (G, H), and Bcl-w (I, J). Plots in the left column (in panels A, C, E, G, I) show the binding profiles of peptides designed to target Bfl-1, and plots in the right column (panels B, D, F, H, J) show binding profiles of peptides designed to target Mcl-1. Data for replicate binding measurements are provided in Table S7.

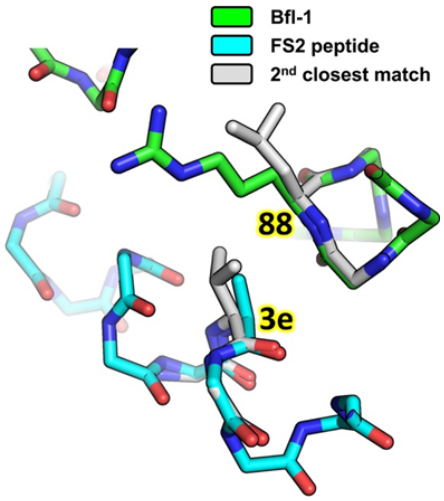

Figure S9. Analysis of close packing interactions at BH3 peptide position 3e. Modeling a valine in BH3 position 3e of template 5UUK reveals a clash with Bfl-1 position 88, but very closely matching helix-helix interfaces from the PDB have valines at the analogous site. Shown is a fragment of the complex between peptide FS2 and Bfl-1 from 5UUK (in cyan and green, respectively), with positions 3e and 88 indicated. The second best-matching fragment (by backbone RMSD) to the fragment of 5UUK that encompasses three residues around each of 3e and 88 is shown in grey (from PDB entry 3WDC). This fragment, with backbone RMSD of only 0.27 Å to the query, has valine at the position corresponding to 3e.

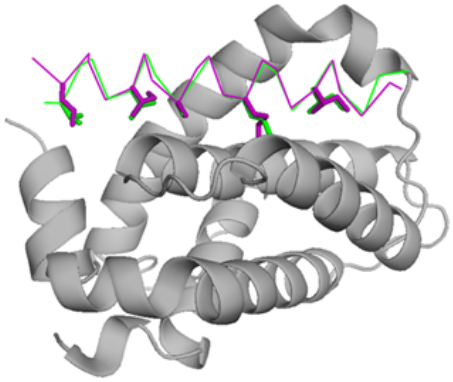

Figure S10. FS2 and dF1 bind similarly to Bfl-1. Superposition of structures 5UUK (FS2 bound to Bfl-1; FS2 in green) with the structure of dF1 bound to Bfl-1 (dF1 in purple).

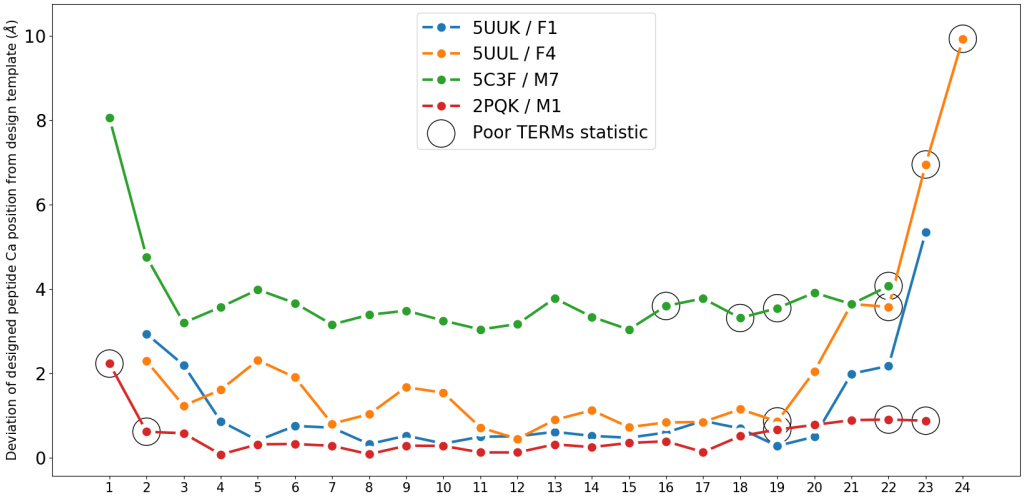

Figure S11. Residues with poor TERM statistics showed structural deviation from the input design template structure. The deviation (in Ångstroms) of every crystallized designed peptide residue from the position of the corresponding residue in the design template structure (C_α-C_α distance) is plotted as a function of residue number. The dTERMen scoring function is derived by identifying multiple TERMs and sub-TERMs around every peptide residue and then identifying matches to these TERMs in the PDB, to extract sequence statistics (see Methods). The average number of structural matches for all TERMs and sub-TERM was computed for each peptide position, and positions with fewer than 2500 matches on average are circled.

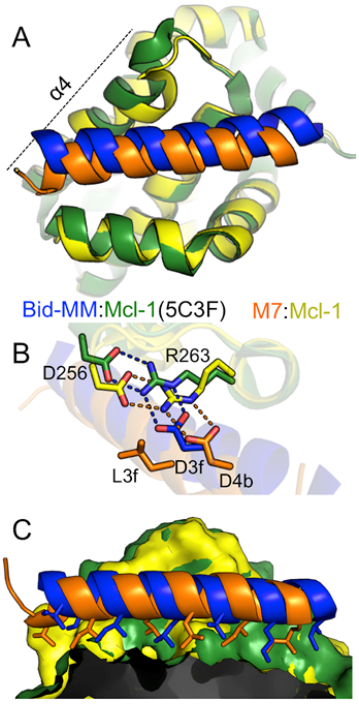

Figure S12. Comparison of the crystal structure of dM7 in complex with Mcl-1 with design template 5C3F. dM7 is in orange, Mcl-1 is in yellow; in structure 5C3F, BID-MM is in blue and Mcl-1 is in green. A) Structural alignment reveals rearrangement of helix 4 of Mcl-1. B) In 5C3F, as in most Mcl-1:BH3 structures, an aspartate in peptide position 3f forms a salt-bridge network with Arg 263 and Asp 256 of Bfl-1 (hydrogen bonds shown as blue dashes). A similar salt-bridge network is observed in the dM7:Mcl-1 complex, but with an aspartate one helical turn away, in peptide position 4b (hydrogen bonds show as orange dashes). C) The shifted binding mode of dM7 re-arranges hydrophobic contacts with Mcl-1 relative to those observed in the 5C3F structure.

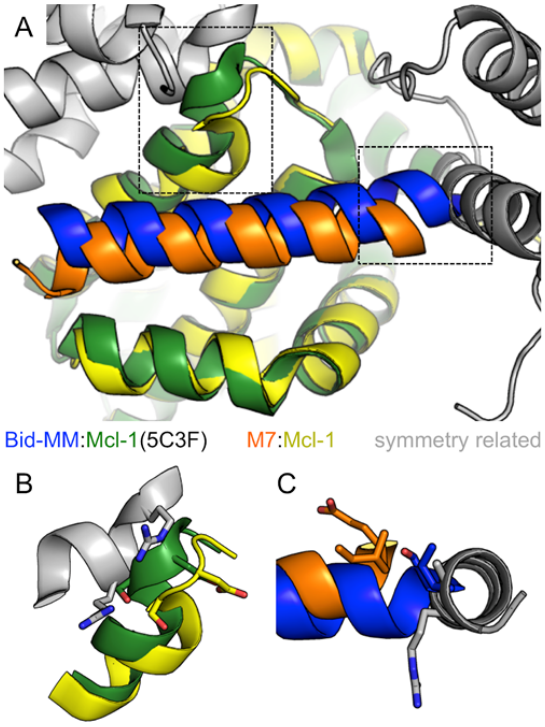

Figure S13. Crystal packing in the dM7:Mcl-1 structure. A) Symmetry related molecules (gray) in the dM7:Mcl-1 crystal would clash with Mcl-1 (as illustrated in B) and with the BH3 peptide (as illustrated in C), if the peptide bound in the same way as BID-MM in 5C3F.

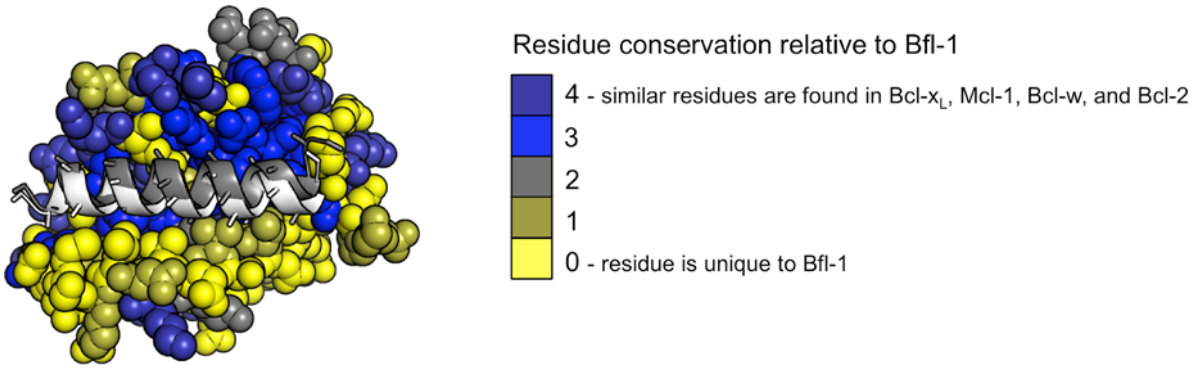

Figure S14. Analysis of Bcl-2 paralog similarity with respect to BH3 peptide binding. Superimposed structures show PUMA (dark gray, cartoon) and FS2 (white, cartoon) bound to Bfl-1 (spheres). Bfl-1 residues are colored according to the extent to which they are conserved in other Bcl-2 family paralogs Bcl-x_L, Mcl-1, Bcl-2 and Bcl-w, using a metric based on the Blosum62 matrix (see Methods).