SUPPLEMENTARY INFORMATION

A combination of Cα–H hydrogen bonds and van der Waals packing modulates the stability of GxxxG-mediated dimers in membranes

Samantha M. Anderson‡, Benjamin K. Mueller‡, Evan J. Lange, Alessandro Senes*

Department of Biochemistry
University of Wisconsin-Madison
433 Babcock Dr.
Madison WI 53706

‡These authors contributed equally

*Corresponding author:
phone +1-608-890-2584
senes@wisc.edu

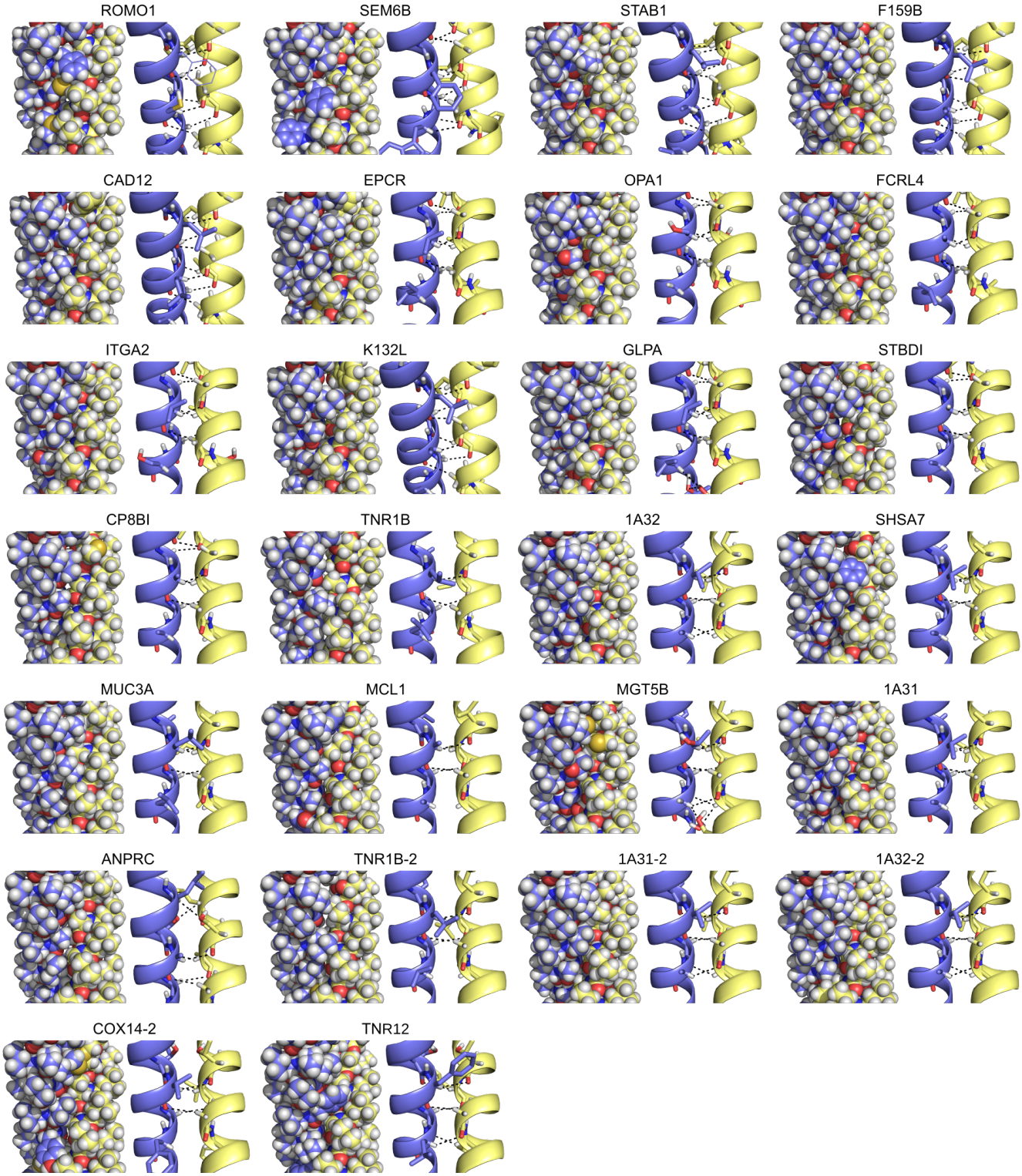

Figure S1. Structural models of the 26 final experimental constructs. The constructs are sorted left to right by CATM energy score. For each construct, a space filling representation is shown on the left to illustrate the packing, and a cartoon representation is shown on the right to illustrate the Cα–H hydrogen bond network.

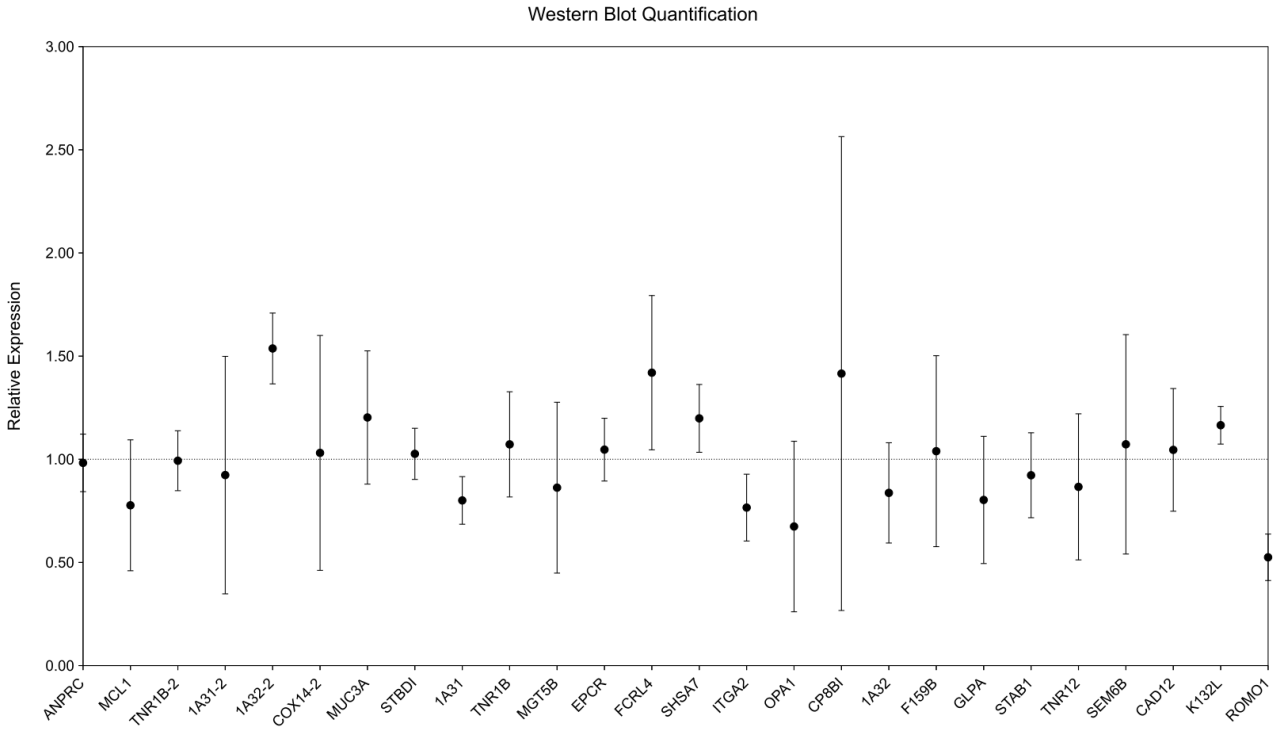

Figure S2. Quantification of expression of the TOXCAT construct using immunoblotting. Normalized expression. The individual construct expression range is 0.52-1.54 fold of the average expression (dashed line) with a standard deviation of the relative expression of the 26 samples of 0.22. The error bars represent the standard deviation of four biological replicates of each construct.

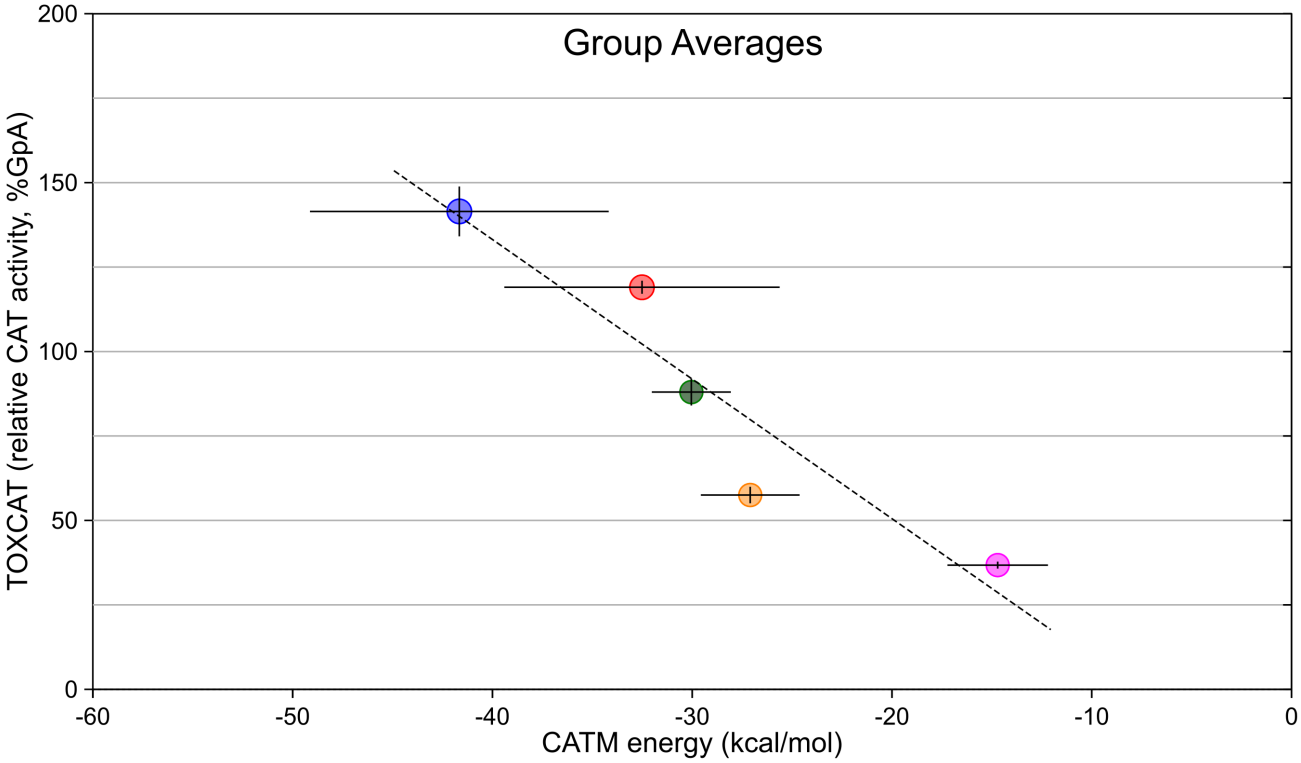

Figure S3. Group averages containing the TNR12 outlier. Same data as in Fig. 4a, with the inclusion the outlier point included in the >100% group. Each point corresponds to the group averaged of five bins based on CAT activity from weak (>25%, magenta) to very strong (>125%, blue), in 25% intervals. The error bars represent the standard error of the average. The dashed line is the linear regression of the data (R^2 = 0.883, p < 0.05).

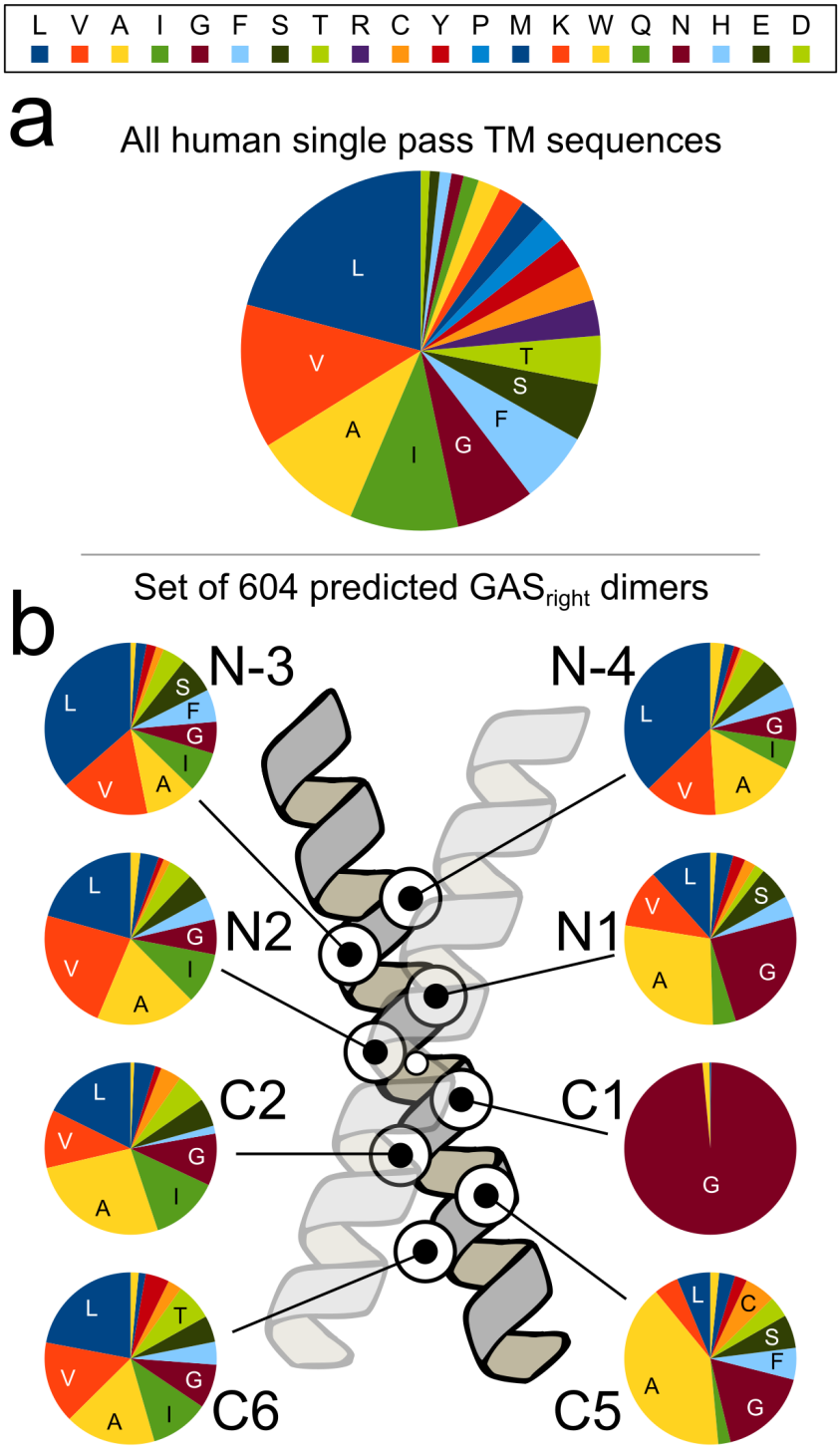

Figure S4. Composition of the interfacial positions of the 604 predicted GAS$_{right}$ dimers. a) For reference, overall composition of all single pass human TM domains. b) Composition of the interfacial position of the final 604 predicted GAS$_{right}$ sequences. These sequences exclude any sequence that contained polar amino acids or proline residues at the eight interfacial positons. The observed biases are consistent with those expected for GAS$_{right}$ motifs, with Gly almost invariably present at position C1, and small amino acids frequently present at positions N1 and C5.

Table S1. Final set of TOXCAT constructs and their C1 mutants

Uniprot AC[1]	Name	Wild-type Sequence	Poly-Leu Construct	TOXCAT[2]	CATM (kcal/mol)[3]
P60602	ROMO1[4]	VKM**GF**VM**GC**AV**GM**AA**GA**LFGTFSCLRIGM	RASLLL**GF**LL**GC**LL**GM**LL**GA**LILI	155%±28%	-56.3±0.1
P60602	ROMO1-G12I		RASLLL**GF**LL**GC**LL**IM**LL**GA**LILI	80%±42%	-24.3±0.1
A8MWY0	K132L	**WL**KV**GA**GV**GA**FT**AV**LLVALTCYFWKKN	RASLLL**WL**LL**GA**LL**GA**LL**AV**LILI	139%±19%	-31.8±0.3
A8MWY0	K132L_G12I		RASLLL**WL**LL**GA**LL**IA**LL**AV**LILI	22%±10%	N.M.
P55289	CAD12	**FL**AV**GL**ST**GA**LI**AI**LLCIVILLAIVVLYVALRRQ	RASLLL**FL**LL**GL**LL**GA**LL**AI**LILI	130%±12%	-36.9±0.5
P55289	CAD12_G12I		RASLLL**FL**LL**GL**LL**IA**LL**AI**LILI	34%±9%	N.M.
P02724	GLPA[4]	EAEIT**LI**IF**GV**MA**GV**IG**TI**LLISYGIRRL	RASLLL**LI**LL**GV**LL**GV**LL**TI**LILI	124%±46%	-31.3±0.3
P02724	GLPA_G12I		RASLLL**LI**LL**GV**LL**IV**LL**TI**LILI	32%±4%	N.M.
Q9H3T3	SEM6B	TSSVAAFVVG**AV**VS**GF**SV**GW**FV**GL**RER	RASLLL**AV**LL**GF**LL**GW**LL**GL**LILI	122%±10%	-44.5±0.3
Q9H3T3	SEM6B_G12I		RASLLL**AV**LL**GF**LL**IW**LL**GL**LILI	27%±5%	N.M.
Q9NY15	STAB1	VAAGVGA**VL**AA**GA**LL**GL**VA**GA**LYLRAR	RASLLL**VL**LL**GA**LL**GL**LL**GA**LILI	119%±18%	-40.6±0.0
Q9NY15	STAB1_G12I		RASLLL**VL**LL**GA**LL**IL**LL**GA**LILI	22%±4%	N.M.
Q9NP84	TNR12	LWAILGG**AL**SL**TF**VL**GL**LS**GF**LVWRRC	RASLLL**AL**LL**TF**LL**GL**LL**GF**LILI	119%±14%	-6.3±1.4
Q9NP84	TNR12_G12I		RASLLL**AL**LL**TF**LL**IL**LL**GF**LILI	29%±5%	N.M.
A6NKW6	F159B[4]	**SL**SI**GA**LI**GL**GI**AA**LVLLAFVISVCVL	RASLLL**SL**LL**GA**LL**GL**LL**AA**LILI	112%±20%	-39.9±0.0
A6NKW6	F159B_G12I		RASLLL**SL**LL**GA**LL**IL**LL**AA**LILI	32%±7%	-5.2±0.9
Q9UNU6	CP8B1	**MV**LW**GA**VL**GA**LL**VV**IAGYLCLAGM	RASLLL**MV**LL**GA**LL**GA**LL**VV**LILI	95%±15%	-28.7±0.3
Q9UNU6	CP8B1_G12I		RASLLL**MV**LL**GA**LL**IA**LL**VV**LILI	27%±6%	N.M.
P10314	1A32	IAIVGIIAGL**VL**FG**AM**FA**GA**VV**AA**VRWRRK	RASLLL**VL**LL**AM**LL**GA**LL**AA**LILI	94%±16%	-25.1±0.2
P10314	1A32_G12I		RASLLL**VL**LL**AM**LL**IA**LL**AA**LILI	20%±5%	N.M.
O60313	OPA1	ATRLLKLR**YL**IL**GS**AV**GG**GY**TA**K	RASLLL**YL**LL**GS**LL**GG**LL**TA**LILI	85%±20%	-34.0±0.1
O60313	OPA1_G12I		RASLLL**YL**LL**GS**LL**IG**LL**TA**LILI	21%±3%	-15.0±0.6
P17301	ITA2	**V**AT**GV**II**GS**II**AG**ILLLLALVAILWKLG	RASLLL**LV**LL**GV**LL**GS**LL**AG**LILI	78%±8%	-32.3±0.1
P17301	ITA2_G12I		RASLLL**LV**LL**GV**LL**IS**LL**AG**LILI	22%±4%	N.M.
A6NL88	SHSA7	STAYVVCGVI**SF**AL**AV**GV**GA**KV**AF**SKA	RASLLL**SF**LL**AV**LL**GA**LL**AF**LILI	68%±13%	-23.2±0.4
A6NL88	SHSA7_G12I		RASLLL**SF**LL**AV**LL**IA**LL**AF**LILI	17%±4%	N.M.
Q9UNN8	EPCR	YTSLVLGVLVGS**FI**IA**GV**AV**GI**FL**CT**G	RASLLL**FI**LL**GV**LL**GI**LL**CT**LILI	61%±11%	-36.7±0.2
Q9UNN8	EPCR_G12I		RASLLL**FI**LL**GV**LL**II**LL**CT**LILI	12%±8%	N.M.
Q96PJ5	FCRL4	DG**LV**AA**GA**TG**GL**LS**AL**LLAVALLFICW	RASLLL**LV**LL**GA**LL**GL**LL**AL**LILI	61%±12%	-33.0±0.1
Q96PJ5	FCRL4_G12I		RASLLL**LV**LL**GA**LL**IL**LL**AL**LILI	25%±5%	N.M.
Q3V5L5	MGT5B	FRLFVLGIGFFT**LC**FL**MT**SL**GG**QF**SA**R	RASLLL**LC**LL**MT**LL**GG**LL**SA**LILI	58%±12%	-19.9±0.4
Q3V5L5	MGT5B_G12I		RASLLL**LC**LL**MT**LL**IG**LL**SA**LILI	29%±6%	-10.8±0.1
P16189	1A31	IAIVGIIAGL**VL**FG**AV**FA**GA**VV**AA**VRWRRK	RASLLL**VL**LL**AV**LL**GA**LL**AA**LILI	52%±12%	-19.5±0.1
P16189	1A31_G12I		RASLLL**VL**LL**AV**LL**IA**LL**AA**LILI	25%±6%	N.M.
P20333	TNR1B	**TG**DF**AL**AV**GL**IV**GV**TALGLLIIGVVNCVIMTQVKKK	RASLLL**TG**LL**AL**LL**GL**LL**GV**LILI	51%±5%	-28.4±0.2
P20333	TNR1B G12I		RASLLL**TG**LL**AL**LL**IL**LL**GV**LILI	18%±8%	N.M.
O95210	STBD1	VWSA**LL**VG**GG**LA**GA**LF**VW**LLRGG	RASLLL**LL**LL**GG**LL**GA**LL**VW**LILI	51%±9%	-28.9±0.0
O95210	STBD1_G12I		RASLLL**LL**LL**GG**LL**IA**LL**VW**LILI	21%±4%	N.M.
Q02505	MUC3A	WR**AL**VG**GL**TA**GA**ALLVLLLLALGVRAV	RASLLL**LL**LL**AL**LL**GL**LL**GA**LILI	40%±10%	-22.8±1.1
Q02505	MUC3A_G12I		RASLLL**LL**LL**AL**LL**IL**LL**GA**LILI	16%±5%	N.M.
P10314	1A32-2	IAIVG**II**AG**LV**LF**GA**MFAGAVVAAVRWRRK	RASLLL**II**LL**LV**LL**GA**LL**FA**LILI	39%±4%	-8.2±0.4
P10314	1A32-2_G12I		RASLLL**II**LL**LV**LL**IA**LL**FA**LILI	4%±9%	N.M.
Q96I36	COX14-2	YKTFSTS**MM**LL**TV**YG**GY**LC**SV**RVYHYFQW	RASLLL**MM**LL**TV**LL**GY**LL**SC**LILI	39%±4%	-7.9±0.1
Q96I36	COX14-2_G12I		RASLLL**MM**LL**TV**LL**IY**LL**SC**LILI	28%±2%	N.M.
P16189	1A31-2	IAIVG**II**AG**LV**LF**GA**VF**AG**AVVAAVRWRRK	RASLLL**II**LL**LV**LL**GA**LL**AG**LILI	36%±10%	-10.7±0.0
P16189	1A31-2_G12I		RASLLL**II**LL**LV**LL**IA**LL**AG**LILI	19%±13%	N.M.
P20333	TNR1B-2	TGDFALAVGLIVGV**TA**LG**LL**II**GV**VN**CV**IMTQVKKK	RASLLL**TA**LL**LL**LL**GV**LL**CV**LILI	35%±12%	-11.3±0.1
P20333	TNR1B-2_G12I		RASLLL**TA**LL**LL**LL**IV**LL**CV**LILI	15%±2%	N.M.
Q07820	MCL1	EGGIRNVLL**AF**AG**VA**GV**GA**GL**AY**LIR	RASLLL**AF**LL**VA**LL**GA**LL**AY**LILI	35%±8%	-22.6±0.1
Q07820	MCL1-G12I		RASLLL**AF**LL**VA**LL**IA**LL**AY**LILI	19%±4%	N.M.
P17342	ANPRC	**L**EE**SA**VT**GI**VV**GA**LLGAGLLMAFYFFRKK	RASLLL**LL**LL**SA**LL**GI**LL**GA**LILI	34%±4%	-19.4±0.1
P17342	ANPRC-G12I		RASLLL**LL**LL**SA**LL**II**LL**GA**LILI	24%±1%	N.M.

1) Uniprot Accession
2) Relative CAT activity compared to GpA (average ± standard deviation)
3) CATM energy of the poly-Leu construct (average ± standard deviation). N.M. = no model predicted.
4) Constructs with C1 values >30% but at least a 75% reduction

Table S2. Constructs removed because of growth defect on maltose media

Uniprot AC[1]	Name	Wild-type Sequence	Poly-Leu Construct	TOXCAT[2]	CATM (kcal/mol)[3]
Q15116	PDCD1	TLVVGVVGGLLGSLVLLVWVLAVICSR	RASLLLTLLLGVLLGLLLSLLILI	215%±5%	-35.6±0.6
Q15116	PDCD1_G12I		RASLLLTLLLGVLLILLLSLLILI	44%±10%	N.M.
Q96LC7	SIG10	FSNGAFLGIGITALLFLCLALIIMKIL	RASLLLLFLLGALLGILLTALILI	214%±10%	-38.9±0.9
Q96LC7	SIG10_G12I		RASLLLLFLLGALLIILLTALILI	36%±2%	N.M.
Q15904	VAS1	DCASFFSAGIWMGLLTSLFMLFIFTYG	RASLLLFFLLGILLGILLSLLILI	165%±19%	-44.5±0.8
Q15904	VAS1-G12I		RASLLLFFLLGILLIILLSLLILI	20%±9%	N.M.
Q96D42	HAVCR1	TKGIYAGVCISVLVLLALLGVIIAKKY	RASLLLLLLLGILLGVLLSVLILI	160%±5%	-37.5±0.9
Q96D42	HAVCR1_G12I		RASLLLLLLLGILLIVLLSVLILI	20%±5%	N.M.
Q6P7N7	TMM81	VASALGIGIAIGVVGGVLVRIVLCALR	RASLLLALLLGILLGVLLGVLILI	145%±6%	-44.8±0.7
Q6P7N7	TMM8I_G12I		RASLLLALLLGILLIVLLGVLILI	29%±2%	N.M.
Q8NCU8	YB039	ERTLQLSVLVAFASGVLLGWQAN	RASLLLSVLLAFLLGVLLGWLILI	135%±45%	-23.2±0.5
Q8NCU8	YB039_G12I		RASLLLSVLLAFLLIVLLGWLILI	-2%±1%	N.M.
Q9Y6N1	COX11	KTTLTYVAAVAVGMLGASYAAVALY	RASLLLTYLLAVLLGMLLASLILI	128%±9%	-33.3±0.3
Q9Y6N1	COX11_G12I		RASLLLTYLLAVLLIMLLASLILI	25%±5%	N.M.
P13591	NCAM1	TSGLSTGAIVGILIVIFVLLLVVVDIT	RASLLLLLLLGLLLGALLGILILI	127%±1%	-32.8±0.6
P13591	NCAM1-G12I		RASLLLLLLLGLLLIALLGILILI	34%±2%	N.M.
O75354	ENTP6	SLRVAKVAYALGLCVGVFIYVAYIKWH	RASLLLVALLAYLLGLLLGVLILI	106%±7%	-25.3±0.1
O75354	ENTP6-G12I		RASLLLVALLAYLLILLLGVLILI	73%±2%	N.M.
Q9NVM1	EVA1B	ESFGLYFVLGVCFGLLLTLCLLVISIS	RASLLLYFLLGVLLGLLLTLLILI	105%±10%	-38.0±0.5
Q9NVM1	EVA1B_G12I		RASLLLYFLLGVLLILLLTLLILI	26%±6%	N.M.
Q86X52	CHSS1	GRRAWLSVLLGLVLGFVLASRLVLARA	RASLLLSVLLGLLLGFLLASLILI	93%±7%	-35.8±0.1
Q86X52	CHSS1-G12I		RASLLLSVLLGLLLIFLLASLILI	17%±5%	N.M.
Q9UKU0	ACSL6	FRSLSATTLVSMGALAAILAYWFTHRA	RASLLLSALLLVLLGALLAILILI	59%±17%	-13.3±0.1
Q9UKU0	ACSL6-G12I		RASLLLSALLLVLLIALLAILILI	24%±4%	N.M.
Q93038	TNR25	WRQMFWVQVLLAGLVVALLLGATLTYT	RASLLLFWLLVLLLGLLLALLILI	56%±9%	-17.5±0.3
Q93038	TNR25_G12I		RASLLLFWLLVLLLILLLALLILI	56%±8%	N.M.
P05026	AT1B1	WFKILLFYVIFYGCLAGIFIGTIQVMLLTISEFK	RASLLLVILLGCLLGILLGTLILI	39%±3%	-33.7±0.3
P05026	AT1B1_G12I		RASLLLVILLGCLLIILLGTLILI	24%±8%	N.M.
Q8N6P7	I22R1	TWTYSFSGAFLFSMGFLVAVLCYLSYR	RASLLLLTLLYSLLGALLFSLILI	31%±9%	-21.9±0.1
Q8N6P7	I22R1_G12I		RASLLLLTLLYSLLIALLFSLILI	39%±11%	N.M.

1) Uniprot Accession
2) Relative CAT activity compared to GpA (average ± standard deviation)
3) CATM energy of the poly-Leu construct (average ± standard deviation). N.M. = no model predicted.

Table S3. Constructs removed because the TOXCAT signal of the WT sequence was <30%

Uniprot AC[1]	Name	Wild-type Sequence	Poly-Leu Construct	TOXCAT[2]	CATM (kcal/mol)[3]
Q6UXN7	TO201	LLR**LL**AA**AA**AC**GA**FA**FL**GYCIYLNRK	RASLLL**LL**LL**AA**LL**GA**LL**FL**LILI	28%±4%	-26.6±1.3
Q6UXN7	TO20L_G12I		RASLLL**LL**LL**AA**LL**IA**LL**FL**LILI	29%±2%	N.M.
Q8TEM1	PO210	SYQVMF**FT**LF**AL**LA**GT**AV**MI**IAYHTVC	RASLLL**FT**LL**AL**LL**GT**LL**MI**LILI	28%±5%	-25.7±0.2
Q8TEM1	PO210_G12I		RASLLL**FT**LL**AL**LL**IT**LL**MI**LILI	40%±11%	N.M.
Q9BRQ8	AIFM2	QVSVESGALHV**VI**VG**GG**FG**GI**AA**AS**QL	RASLLL**VI**LL**GG**LL**GI**LL**AS**LILI	26%±8%	-29.0±0.1
Q9BRQ8	AIFM2_G12I		RASLLL**VI**LL**GG**LL**II**LL**AS**LILI	22%±10%	N.M.
Q8IVU1	IGDCC3	TT**GI**VI**GI**HI**GV**TCIIFCVLFLLFGQR	RASLLL**LL**LL**GI**LL**GI**LL**GV**LILI	25%±6%	-41.1±0.5
Q8IVU1	IGDCC3_G12I		RASLLL**LL**LL**GI**LL**II**LL**GV**LILI	22%±4%	N.M.
O15197	EPHB6	RL**SL**VI**GS**IL**GA**LAFLLLAAITVLAVV	RASLLL**LL**LL**SL**LL**GS**LL**GA**LILI	25%±3%	-21.6±0.2
O15197	EPHB6_G12I		RASLLL**LL**LL**SL**LL**IS**LL**GA**LILI	18%±5%	N.M.
Q13586	STIM1	LKDFMLV**VS**IV**IG**VG**GC**WF**AY**IQNRYS	RASLLL**VS**LL**IG**LL**GC**LL**AY**LILI	24%±6%	-20.4±0.4
Q13586	STIM1_G12I		RASLLL**VS**LL**IG**LL**IC**LL**AY**LILI	27%±8%	-20.2±1.9
B6SEH8	ERVV1	KRA**LG**LI**LA**GM**GA**AI**GM**IAAWGGFTYH	RASLLL**LG**LL**LA**LL**GA**LL**GM**LILI	24%±8%	-19.5±0.2
B6SEH8	ERVV1_G12I		RASLLL**LG**LL**LA**LL**IA**LL**GM**LILI	22%±5%	-9.2±0.4
Q8N3T1	GLT15	HRACRLQ**FL**LL**LL**ML**GC**VL**MM**VAMLHAAH	RASLLL**FL**LL**LL**LL**GC**LL**MM**LILI	19%±1%	-18.1±0.4
Q8N3T1	GLT15_G12I		RASLLL**FL**LL**LL**LL**IC**LL**MM**LILI	37%±7%	N.M.
P05067	A4	**S**NK**GA**II**GL**MV**GG**VVIATVIVITLVMLKKK	RASLLL**LS**LL**GA**LL**GL**LL**GG**LILI	18%±4%	-35.2±0.1
P05067	A4_G12I		RASLLL**LS**LL**GA**LL**IL**LL**GG**LILI	16%±3%	N.M.
Q96I36	COX14	YKTFST**SM**ML**LT**VY**GG**YL**CS**VRVYHYFQW	RASLLL**SM**LL**LT**LL**GG**LL**CS**LILI	14%±2%	-21.1±0.1
Q96I36	COX14_G12I		RASLLL**SM**LL**LT**LL**IG**LL**CS**LILI	27%±7%	-5.1±0.7

[1]Uniprot Accession
[2]Relative CAT activity, compared to GpA (average ± standard deviation)
[3]CATM energy of the poly-Leu construct (average ± standard deviation). N.M. = no model predicted.

Table S4. Constructs discarded because of TOXCAT signal >30% for the $C1_{Gly \rightarrow Ile}$ variant

Uniprot AC[1]	Name	Wild-type Sequence	Poly-Leu Construct	TOXCAT[2]	CATM (kcal/mol)[3]
P43489	TNR4	GRAVAA**IL**GL**GL**VL**GL**LG**AL**AILLALY	RASLLL**IL**LL**GL**LL**GL**LL**AL**LILI	102%±11%	-39.7±0.1
P43489	TNR4_G12I		RASLLL**IL**LL**GL**LL**IL**LL**AL**LILI	109%±13%	N.M.
Q9UPZ6	THS7A	K**TW**VY**GV**AA**GA**FVLLIFIVSMIYLACK	RASLLL**LL**LL**TW**LL**GV**LL**GA**LILI	87%±37%	-3.2±0.1
Q9UPZ6	THS7A_G12I		RASLLL**LL**LL**TW**LL**IV**LL**GA**LILI	76%±9%	N.M.
Q7L8C5	STY13	SVA**VI**AL**GA**TL**GT**AT**SI**LALCGVTCLCRH	RASLLL**VI**LL**GA**LL**GT**LL**SI**LILI	79%±5%	-31.0±0.1
Q7L8C5	STY13_G12I		RASLLL**VI**LL**GA**LL**IT**LL**SI**LILI	45%±6%	N.M.
Q08ET2	SIG14	LVL**TL**IR**GA**LM**GA**GF**LL**TYGLTWIYYTRC	RASLLL**TL**LL**GA**LL**GA**LL**LL**LILI	77%±7%	-26.3±0.1
Q08ET2	SIG14_G12I		RASLLL**TL**LL**GA**LL**IA**LL**LL**LILI	39%±14%	N.M.
Q8TDF5	NETO1	S**GT**VI**GV**TS**CI**VIILIIISVIVQIKQA	RASLLL**LL**LL**GT**LL**GV**LL**CI**LILI	72%±5%	-32.9±0.3
Q8TDF5	NETO1_G12I		RASLLL**LL**LL**GT**LL**IV**LL**CI**LILI	38%±%	N.M.
Q99795	GPA33	**M**NV**AL**YV**GI**AV**GV**VAALIIIGIIIY**C**	RASLLL**LM**LL**AL**LL**GI**LL**GV**LILI	68%±17%	-23.3±0.0
Q99795	GPA33_G12I		RASLLL**LM**LL**AL**LL**II**LL**GV**LILI	38%±7%	N.M.
Q9BY71	LRRC3	TTD**VA**ML**VT**MF**GW**FA**MV**IAYVVYYVRH	RASLLL**VA**LL**VT**LL**GW**LL**MV**LILI	68%±10%	-8.3±0.4
Q9BY71	LRRC3_G12I		RASLLL**VA**LL**VT**LL**IW**LL**MV**LILI	57%±12%	N.M.
Q30201	HFE	**TL**VI**GV**IS**GI**AV**FV**VILFIGILFIILRKRQ	RASLLL**TL**LL**GV**LL**GI**LL**FV**LILI	64%±9%	-23.3±0.2
Q30201	HFE-G12I		RASLLL**TL**LL**GV**LL**II**LL**FV**LILI	53%±5%	N.M.
Q5SSG8	MUC21	AWEIFLITLVSV**VA**AV**GL**FA**GL**FF**CV**R	RASLLL**VA**LL**GL**LL**GL**LL**CV**LILI	60%±6%	-32.3±0.4
Q5SSG8	MUC21_G12I		RASLLL**VA**LL**GL**LL**IL**LL**CV**LILI	43%±%	N.M.
Q9P2S2	NRXN2	GMVVGIVAAAALCILILLYAM	RASLLL**LS**LL**GM**LL**GI**LL**AA**LILI	54%±2%	-33.7±0.2
Q9P2S2	NRXN2_G12I		RASLLL**LS**LL**GM**LL**II**LL**AA**LILI	39%±8%	N.M.
Q2M385	MPEG1	SG**GA**AA**GV**TV**GV**TTILAVVITLAIYGT	RASLLL**LL**LL**GA**LL**GV**LL**GV**LILI	34%±6%	-33.9±0.9
Q2M385	MPEG1_G12I		RASLLL**LL**LL**GA**LL**IV**LL**GV**LILI	33%±1%	N.M.
Q8N967	LRTM2	MGT**VI**IA**GV**VC**GV**VC**IM**MVVAAAYGCI	RASLLL**VI**LL**GV**LL**GV**LL**IM**LILI	32%±5%	-17.1±0.3
Q8N967	LRTM2_G12I		RASLLL**VI**LL**GV**LL**IV**LL**IM**LILI	31%±13%	N.M.
Q9H3N1	TMX1	WGSYTV**FA**LA**TL**FS**GL**LL**GL**CMIFVADCL	RASLLL**FA**LL**TL**LL**GL**LL**GL**LILI	31%±3%	-9.0±0.4
Q9H3N1	TMX1_G12I		RASLLL**FA**LL**TL**LL**IL**LL**GL**LILI	32%±5%	N.M.
Q9H6B4	CLMP	**M**VA**GA**VT**GI**VA**GA**LLIFLLVWLLIRRK	RASLLL**LM**LL**GA**LL**GI**LL**GA**LILI	30%±7%	-33.8±0.1
Q9H6B4	CLMP_G12I		RASLLL**LM**LL**GA**LL**II**LL**GA**LILI	33%±4%	N.M.

[1]Uniprot Accession
[2]Relative CAT activity, compared to GpA (average ± standard deviation)
[3]CATM energy of the poly-Leu construct (average ± standard deviation). N.M. = no model predicted.

Table S5. Progression of number of constructs from the computational analysis to the final set of experimental constructs

Human Single Pass Proteins in Uniprot	2,383
Wild-type with energy score below 0 kcal/mol	1,141
Poly-Leu sequences with energy score below 0 kcal/mol	1,020
Subset of sequences with non-polar interface	668
Poly-Leu sequences with energy score below -5 kcal/mol	604
Sequences selected for experimental analysis	65
Sequences that passed the maltose growth test (-15)	50
Sequences with TOXCAT >30% (-10)	40
Sequences with C1 mutation <30% and final set (-14)	26

Table S6. Uniprot Accession and description of the 26 final constructs applied for the analysis

Uniprot Accession	Uniprot Entry Name	Protein Name
A6NKW6	F159B_HUMAN	Membrane protein FAM159B
A6NL88	SHSA7_HUMAN	Protein shisa-7
A8MWY0	K132L_HUMAN	UPF0577 protein KIAA1324-like
O60313	OPA1_HUMAN	Dynamin-like 120 kDa protein, mitochondrial
O95210	STBD1_HUMAN	Starch-binding domain-containing protein 1
P02724	GLPA_HUMAN	Glycophorin-A
P10314	1A32_HUMAN	HLA class I histocompatibility antigen, A-32 alpha chain
P16189	1A31_HUMAN	HLA class I histocompatibility antigen, A-31 alpha chain
P17301	ITA2_HUMAN	Integrin alpha-2
P17342	ANPRC_HUMAN	Atrial natriuretic peptide receptor 3
P20333	TNR1B_HUMAN	Tumor necrosis factor receptor superfamily member 1B
P55289	CAD12_HUMAN	Cadherin-12
P60602	ROMO1_HUMAN	Reactive oxygen species modulator 1
Q02505	MUC3A_HUMAN	Mucin-3A
Q07820	MCL1_HUMAN	Induced myeloid leukemia cell differentiation protein Mcl-1
Q3V5L5	MGT5B_HUMAN	Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase B
Q96I36	COX14_HUMAN	Cytochrome c oxidase assembly protein COX14
Q96PJ5	FCRL4_HUMAN	Fc receptor-like protein 4
Q9H3T3	SEM6B_HUMAN	Semaphorin-6B
Q9NP84	TNR12_HUMAN	Tumor necrosis factor receptor superfamily member 12A
Q9NY15	STAB1_HUMAN	Stabilin-1
Q9UNN8	EPCR_HUMAN	Endothelial protein C receptor
Q9UNU6	CP8B1_HUMAN	7-alpha-hydroxycholest-4-en-3-one 12-alpha-hydroxylase

Table S7. Energetic and geometric properties of groups of CATM scores.

CATM Score Range[1]	-5 to -15	-15 to-25	-25 to -35	-35 to -45	-45 and below
Number of models	210	254	89	45	6
CATM energy score (kcal/mol)	-10.1±3.0	-19.5±2.7	-29.4±3.1	-39.0±2.9	-47.6±4.0
Van der Waals (kcal/mol)	-24.2±4.2	-29.9±3.7	-35.2±3.2	-39.7±3.7	-39.9±5.3
Cα-H hydrogen bonding (kcal/mol)	-3.4±4.9	-4.9±3.9	-7.6±3.2	-10.0±2.9	-13.2±0.8
Solvation (kcal/mol)	18.2±2.1	16.1±2.5	14.1±2.1	11.1±2.8	5.5±1.1
Crossing angle (°)	-52.3±8.1	-49.2±8.1	-47.3±4.4	-40.4±5.5	-31.6±2.1
Number of Cα-H bonds	4.7±1.0	4.3±0.9	5.6±1.3	7.3±1.1	8.0±0.0
Interface surface area (Å^2)	4730±910	4790±910	4730±710	4400±580	4070±490
Inter-helical distance (Å)	7.1±0.2	7.0±0.2	6.6±0.3	6.4±0.1	6.4±0.1
Van der Waals/Interface surface area (kcal/(mol Å^2))	-0.0051±0.0013	-0.0063±0.0014	-0.0076±0.0013	-0.0091±0.0010	-0.0098±0.0006
Sequences with GxxxG	17%	28%	75%	100%	100%
Sequences with Sm-xxx-Sm	93%	96%	100%	100%	100%
Sequences with Gly at N1	2%	13%	67%	100%	100%
Sequences with Gly at C1	99%	97%	100%	98%	100%
Sequences with Gly at C5	17%	14%	16%	36%	50%

[1]All values are reported as averages ± standard deviation.
[2]Sm-xxx-Sm are defined by any combinations of Gly, Ala, Ser and Cys at the first and last position.