Supplementary Material

MHC I Stabilizing Potential of Computer-Designed Octapeptides

Joanna M. Wisniewska, Natalie Jäger, Anja Freier, Florian O. Losch, Karl-Heinz Wiesmüller, Peter Walden, Paul Wrede, Gisbert Schneider, Jan A. Hiss

Supplementary 1: Descriptor list and specifications. Five defined 'descriptor spaces' (sets a.-e.):

a. 40-dimensional (8 residues × 5 properties): hydrophobicity [1]; hydrophilicity [2]; bulkiness [3]; refractivity [4]; H-2K^b stabilization values [5].
b. 40-dimensional (8 × 5): flexibility [6]; number of possible hydrogen bonds [7]; molecular weight [8]; isoelectricity [9]; hydrophobicity [10].
c. 48-dimensional (8 × 6): six pharmacophoric features of the 20 standard amino acids [11].
d. 264-dimensional (8 × 33): 33 descriptors selected from a set of 184 descriptors that were computed by the software package MOE (Molecular Operation Environment, v2006.08, Chemical Computing Group, Inc., Montreal, Canada). The 33 descriptors were selected based on Kolmogorov-Smirnov statistics [12], so that they separate stabilizing from non-stabilizing octapeptides in at least five residue positions at the 5% significance level. The descriptor space d. consists of following 33 descriptors: diameter, VDistEq, VDistMa, einerPath, weinerPol, BCUT_PEOE_2, CUT_SLOGP_2, BCUT_SMR_1, BCUT_SMR_2, BCUT_SMR_3, GCUT_PEOE_3, GCUT_SLOGP_3, a_count, a_IC, a_ICM, b_count, chi0v, chi1v, a_heavy, a_nC, b_heavy, chi0, chi1_C, VAdjEq, zagreb, PEOE_RPC, PEOE_VSA-0, PEOE_VSA_POS, Kier2, KierA1, KierA2, KierA3, vsa_hyd.
e. 904-dimensional (8 × 113): 113 descriptors selected from a set of more than 1600 descriptors of the software package DragonX (DRAGON for Linux (software for Molecular Descriptor Calculations) Version 1.2, 2006) using Kolmogorov-Smirnov statistics which separate stabilizing from non-stabilizing octapeptides in at least six residue positions. The descriptor space e. includes following 113 descriptors: MW, Sv, Se, Sp, Ss, nAT, nBT, ZM1V, ZMV, Dz, SMTIV, GMTIV, WhetZ, Whetm, Whetv, Whete, Whetp, Jhetp, DEL, S0K, S1K, S2K, X0v, X1v, XMO, ISIZ, IAC, TIC0, BIC0, SIC1, CIC1, BIC1, IC3, IC4, TIC4, IC5, TIC5, SIC5, BIC5, ATS1m, ATS2m, ATS3m, ATS1e, ATS2e, ATS3e, ATS4e, MATS1m, MATS2m, MATS1v, MATS2v, MATS1e, MATS1p, MATS2p, EEig02d, EEig03d, EEig04d, EEig06d, EEig09d, ESpm02d, ESpm03d, ESpm04d, ESpm05d, ESpm06d, ESpm03r, ESpm04r, ESpm05r, BEHm3, BEHm4, BEHm5, BEHm6, BELm1, BELm2, BELm4, BELm6, BEHv2, BEHv3, BEHv4, BEHv5, BEHv6, BEHv7, BELv5, BEHe2, BEHe3, BEHe4, BEHe5, BEHe6, BELe2, BELe4, BEHp2, BEHp4, BEHp5, BEHp6, BEHp7, BELp5, Eig1Z, Eig1m, Eig1v, Eig1e, Eig1p, AEigZ, AEigm, AEigv, AEige, AEigp, VEZ1, VEZ2, VEm1, VEm2, VRv1, VEe1, VEe2, VRp1, AMR.

Supplementary 2: Experimentally tested octapeptides. Sequences are given in single-letter code, motif anchor positions fulfilling the SYFPEITHI motif are underlined. *S*-score: SYFPEITHI score [13]. IEDB-ANN score: predicted IC_{50} values by IEDB-ANN method [14]. < 500 nM: recommended cutoff for IEDB-ANN score [14] for binders predicted with intermediate affinity; *active*: IEDB-ANN score < 500 nM; *0*: IEDB-ANN score > 500 nM. EC_{50} values correspond to the peptide concentration required for 50% of maximal MHC I protein stabilization; values in brackets are standard deviations ($N = 3$). n.d.: not detectable.

	Peptide number	Sequence	*S*-score	*IEDB*-ANN score [µM]	< 500 nM	Experimental EC_{50} [µM]
stabilizing peptides	1	FRYPYKTL	27	0.5	active	24.0 (± 11)
	2	FRYIYHTL	27	0.2	active	20.0 (± 4.0)
	3	FHWDYRGL	22	0.4	active	25.0 (± 11)
	4	WRFKYDNL	22	2.3	0	7.0 (± 4.0)
	5	WRFVYWRL	20	1.2	0	8.0 (± 4.0)
	6	WRFIYFNL	21	0.4	active	3.6 (± 1.2)

	7	WDFKFDSV	17	4.7	0	9.0 ($\pm$ 4.0)
	8	WKFIFDPV	16	3.0	0	0.4 ($\pm$ 0.1)
	9	FHHAHRTV	8	10.5	0	9.0 ($\pm$ 4.0)
non-stabilizing peptides of category i	10	FRYEYRSL	28	0.1	active	n.d.
	11	WRYIYHSI	22	2.5	0	n.d.
	12	HRYVYRNI	24	0.4	active	n.d.
	13	FHYAYRSV	23	0.1	active	n.d.
	14	YRYKYDRL	27	0.5	0	n.d.
	15	WRYQYDNL	27	0.7	0	n.d.
	16	WRYRYWSL	27	0.6	0	n.d.
	17	WRYNYDPL	26	1.7	0	n.d.
	18	WRYHYDPL	26	2.4	0	n.d.
	19	WKYQYDNL	27	0.3	active	n.d.
	20	WKYIFDPV	22	1.2	0	n.d.
	21	WKYPFDPV	22	2.0	0	n.d.
non-stabilizing peptides of category ii	22	FRYLYKNA	17	1.2	0	n.d.
	23	FRYVWRTL	18	1.7	0	n.d.
	24	FHYLYHTA	17	0.9	0	n.d.
	25	FRYPYHTP	17	17.8	0	n.d.
	26	FRWEYRGL	22	0.7	0	n.d.
	27	FRHIYRTI	18	9.3	0	n.d.
	28	FRHGYRQI	18	12.6	0	n.d.
	29	FHWAYHTV	17	1.2	0	n.d.
	30	WRWLYKGV	16	4.3	0	n.d.
	31	WRFPYDQL	21	6.6	0	n.d.
	32	WRFKYDPL	21	3.3	0	n.d.
	33	WRFPYDKL	21	6.8	0	n.d.
	34	WRFVYDNL	21	1.9	0	n.d.
	35	WKFKFDPV	17	2.8	0	n.d.
	36	WKINFDPV	17	8.2	0	n.d.
	37	TTEWYTKI	18	3.4	0	n.d.
5 of 23 non-stabilizing peptides of	38	RGEVFTAT	13	24.0	0	n.d.
	39	FHYDHRNA	9	6.5	0	n.d.
	40	HRWVFWQP	11	26.0	0	n.d.
	41	SIKNFHYY	12	13.6	0	n.d.
	42	FRHDYHSP	11	29.2	0	n.d.
7 of 120 non-stabilizing peptides of category iv	43	TVEQGVTQ	1	37.5	0	n.d.
	44	FHHGHNVP	0	36.8	0	n.d.
	45	QDGHEIHR	1	38.4	0	n.d.
	46	TVEQGVTQ	1	37.5	0	n.d.
	47	GVDQSYLK	0	38.1	0	n.d.
	48	SVENPILR	2	33.7	0	n.d.
	49	NEGWTIHR	1	38.8	0	n.d.
	50	SVDHSIFK	2	36.1	0	n.d.
18 of 23 non-stabilizing peptides of category iii	51	FHYGHKTA	7	16.2	0	n.d.
	52	FHYIHRSA	7	0.42	0	n.d.
	53	WKFRVDRV	7	15.9	0	n.d.
	54	HGEFYPLP	11	30.0	0	n.d.
	55	HGEWYPFD	11	32.2	0	n.d.
	56	TIKNDDSI	7	31.8	0	n.d.
	57	GIKQYRYK	12	28.4	0	n.d.
	58	RGDYYATE	13	23.3	0	n.d.
	59	GIENFHKD	12	34.5	0	n.d.
	60	VVDHFRSP	11	23.2	0	n.d.
	61	VVEHFRLW	11	17.3	0	n.d.
	62	VIENFSAY	12	9.5	0	n.d.
	63	HGEYYRQH	13	32.9	0	n.d.
	64	HGDFYIVS	12	29.4	0	n.d.
	65	RGEPNSLI	9	33.0	0	n.d.

66	HGEWYHFA	11	25.7	0	n.d.
67	HGEWYPGP	11	29.6	0	n.d.
68	RGKWEKRI	8	30.3	0	n.d.

113 of 120 non-stabilizing peptides of category iv

69	SDGHQPHR	1	36.0	0	n.d.
70	QEGWDGRR	0	38.5	0	n.d.
71	NEGHGIHR	1	39.0	0	n.d.
72	NGGWGVHR	1	36.6	0	n.d.
73	NEGYDPHR	0	38.8	0	n.d.
74	NEGHTIHR	1	39.2	0	n.d.
75	QDGHQAHR	0	37.7	0	n.d.
76	NEGWQAKR	1	39.0	0	n.d.
77	NDGWDPHR	0	38.9	0	n.d.
78	QDGHSVHR	0	38.1	0	n.d.
79	QDGYNPHR	0	37.0	0	n.d.
80	SDGHQPKR	2	36.9	0	n.d.
81	QDGWGPHR	0	37.1	0	n.d.
82	NEVHGARN	0	38.2	0	n.d.
83	NDPHEIHN	2	38.1	0	n.d.
84	NEVYGPHQ	0	37.4	0	n.d.
85	QGIYGVHN	2	33.0	0	n.d.
86	QDEHGIRN	1	39.0	0	n.d.
87	QDIHGARQ	1	35.8	0	n.d.
88	NEPFTIKN	3	37.8	0	n.d.
89	NDAWGLHQ	0	39.2	0	n.d.
90	QEVHGIHQ	1	37.5	0	n.d.
91	GVETDADN	0	39.6	0	n.d.
92	GVEWKADN	0	39.6	0	n.d.
93	GVEADRDN	1	40.0	0	n.d.
94	GAEGERDN	2	40.0	0	n.d.
95	GAEADAWN	1	39.6	0	n.d.
96	GVEEKADN	1	39.7	0	n.d.
97	GVEGEADN	0	39.7	0	n.d.
98	GAEDERDN	3	39.9	0	n.d.
99	SVEGEADN	1	38.6	0	n.d.
100	GVEDEPDN	1	39.6	0	n.d.
101	GVEGDPDN	0	39.7	0	n.d.
102	GVEGDRVN	1	40.0	0	n.d.
103	GVEGDADN	0	39.8	0	n.d.
104	GAEEERVN	3	39.6	0	n.d.
105	SVDGERTK	3	38.6	0	n.d.
106	GADAKADK	1	39.3	0	n.d.
107	GVDGEADK	0	39.3	0	n.d.
108	GADDEPDK	2	39.5	0	n.d.
109	GVDFEADK	0	38.9	0	n.d.
110	GVDSDAWK	0	39.1	0	n.d.
111	GVKGKADK	0	29.6	0	n.d.
112	GVDDEADK	1	30.5	0	n.d.
113	GVDQSWKG	1	38.4	0	n.d.
114	SIKQSNYY	2	33.5	0	n.d.
115	SVDNSVVF	1	33.4	0	n.d.
116	TPENPTYW	1	35.1	0	n.d.
117	SVEKGVVH	2	37.6	0	n.d.
118	SVETVTSP	2	35.7	0	n.d.
119	SIDNGVVW	2	30.8	0	n.d.
120	SVENGVLR	1	33.9	0	n.d.
121	TVDNVILR	1	33.2	0	n.d.
122	TPENSVVR	0	36.9	0	n.d.
123	VVEWPVHN	0	34.1	0	n.d.
124	SVEQSVVH	1	38.0	0	n.d.
125	GIFQQNLP	1	32.0	0	n.d.
126	SVELSIHP	2	35.5	0	n.d.
127	GIDQGVFH	1	38.3	0	n.d.
128	SVEQGILR	2	34.3	0	n.d.
129	TVENSIHK	1	36.0	0	n.d.
130	SIWQVTKD	4	24.2	0	n.d.
131	SVDQGLLQ	1	36.5	0	n.d.

132	TVESVSSP	1	36.4	0	n.d.
133	GIRQSLKE	2	37.5	0	n.d.
134	TPDNGLFR	0	36.6	0	n.d.
135	SVDQGLKE	2	36.9	0	n.d.
136	TVENSVPQ	0	35.3	0	n.d.
137	TVENLSSY	1	33.0	0	n.d.
138	SVEPESHG	2	37.2	0	n.d.
139	SVENGILW	2	32.0	0	n.d.
140	SIDQGVAE	2	32.7	0	n.d.
141	SIEWGLFK	2	34.3	0	n.d.
142	SVEPDSHG	2	37.6	0	n.d.
143	TVENGLVQ	0	38.2	0	n.d.
144	SVENGVPQ	1	34.1	0	n.d.
145	SIKQSNEP	2	36.9	0	n.d.
146	STEPDVAW	1	34.9	0	n.d.
147	TVDTTVSW	0	33.1	0	n.d.
148	IVDHGTAP	2	37.0	0	n.d.
149	TVEQGILH	1	37.9	0	n.d.
150	TVEPWRSW	1	32.9	0	n.d.
151	GLDPDRHW	1	38.7	0	n.d.
152	WTEWVSPP	1	34.8	0	n.d.
153	KVEQSSPE	1	36.2	0	n.d.
154	IVEWGVHW	1	23.8	0	n.d.
155	IVEHSDVP	1	38.7	0	n.d.
156	IVKQGTKW	3	28.2	0	n.d.
157	VVEQSIRY	3	30.6	0	n.d.
158	HVEQGSLY	1	38.2	0	n.d.
159	IVEWITAG	2	35.2	0	n.d.
160	HPEQGVRW	0	36.6	0	n.d.
161	TVDNSVAY	0	35.8	0	n.d.
162	VVDPVLHY	0	29.3	0	n.d.
163	HPEHGVRW	0	37.1	0	n.d.
164	SVEQPLDP	1	37.4	0	n.d.
165	TLDWGVRW	0	31.2	0	n.d.
166	TLKNSKNW	1	31.6	0	n.d.
167	TFAHGVRW	0	24.7	0	n.d.
168	IVELTSAY	2	30.8	0	n.d.
169	SVENPIAQ	2	35.0	0	n.d.
170	RGTYHRTK	4	34.4	0	n.d.
171	RGKYGPFE	2	34.8	0	n.d.
172	RGDREPPA	3	33.7	0	n.d.
173	VVEWVEHY	1	31.8	0	n.d.
174	RGDNPSYT	3	34.7	0	n.d.
175	RGEPLALA	2	34.0	0	n.d.
176	RGEHWIDQ	3	36.2	0	n.d.
177	RGEIWIFW	3	27.1	0	n.d.
178	TVDISTVH	1	39.0	0	n.d.
179	RGEIWHLQ	2	36.4	0	n.d.
180	SVEISIAQ	2	35.8	0	n.d.

Supplementary References:

[1] D. M. Engelman, T. A. Steitz, A. Goldman, Annu. Rev. Biophys. Biophys. Chem. 1986, 15, 321–353.
[2] T. P. Hopp, K. R. Woods, Proc. Natl. Acad. Sci. USA 1981, 8, 3824–3828.
[3] D. D. Jones, J. Theor. Biol. 1975, 50, 167–184.
[4] K. Udaka, K. H. Wiesmuller, S. Kienle, G. Jung, P. Walden, J. Biol. Chem. 1995, 270, 24130–24134.
[5] R. Bhaskaran, P. K. Ponnuswamy, Int. J. Pept. Prot. Res. 1988, 32, 241–255.
[6] D. S. Dwyer, J. Biomol. Struct. Dyn. 2001, 18, 881–892.
[7] G. D. Fasman, Handbook of Biochemistry and Molecular Biology, Proteins CRC Press, Cleveland, 1976.
[8] G. D. Fasman, Handbook of Biochemistry and Molecular Biology, Proteins CRC Press, Cleveland, 1976.
[9] J. M. Zimmerman, N. Eliezer, R. Simha, J. Theor. Biol. 1968, 21, 170–201.
[10] S. D. Black, D. R. Mould, Anal. Biochem. 1991, 193, 72–82.

[11] G. Schneider, K.-H. Baringhaus, Molecular Design – Concepts and Applications, Wiley-VCH, Weinheim, 2008.
[12] D. N. Rassokhin, D. K. Agrafiotis, J. Mol. Graph. Model. 2000, 18,368–382.
[13] H. G. Rammensee, J. Bachmann, N. N. Emmerich, O. A. Bachor and S. Stevanovic, "SYFPEITHI: database for MHC ligands and peptide motifs", *Immunogenetics*, vol. 50, pp. 213-219, 1999.
[14] M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemøller, K. Lamberth, S. Buus, S. Brunak, O. Lund, "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations", *Protein Science*, vol. 12, pp. 1007-1017, 2003.