

Cluster-based comparison of peptide mass fingerprints obtained by MALDI-TOF mass spectrometry. Case study: long-term stability of rituximab

Pablo J. Villacorta¹, Antonio Salmerón-García², David A. Pelta¹, José Cabeza²,

Antonio Lario⁵, Natalia Navas^{2,✉}

(1) CITIC-UGR, Department of Computer Science and Artificial Intelligence, University of Granada, Avda. de Andalucía, 38 E-18071 Granada, Spain.

(2) UGC Intercentro Interniveles Farmacia Granada, “San Cecilio Hospital”, Biomedical Research Institute ibs.GRANADA. Hospitales Universitarios de Granada, E-18012 Granada, Spain.

(3) CSIC (Consejo Superior Investigaciones Científicas), ProteomicUnit, Parasitology and Biomedicine Institute “López Neira”, Parque Tecnológico Ciencias de la Salud, Avda. del Conocimiento, s/n. 18100 Armilla, Granada, 18005 Spain.

(4) Department of Analytical Chemistry, Science Faculty, Biomedical Research Institute ibis.GRANADA, University of Granada, Campus Fuentenueva s/n, E-18071 Granada, Spain.

✉Corresponding author: E-mail natalia@ugr.es, Phone +34 958243388, Fax. +34 958243328

Supplementary Information

Supplementary Figure 1:

Rituximab

► Light chain :

QIVLSQSPA	LSASPGEKVT	MTCRASSSVS	YIHWFQQKPG
SSPKPWYAT	SNLASGVPVR	FSGSGSGTSY	SLTISRVEAE
DAATYYCQQW	TSNPPTFGGG	TKLEIKRTVA	APSVFIFPPS
DEQLKSGTAS	VVCLLNNFYP	REAKVQWKVD	NALQSGNSQE
SVTEQDSKDS	TYSLSSTLTL	SKADYEKHKV	YACEVTHQGL
SSPVTKSFNR	GEC;		

► Heavy chain:

QVQLQQPGAE	LVKPGASVKM	SCKASGYTFT	SYNMHWVKQT
PGRGLEWIGA	IYPGNGDTSY	NQKFKGKATL	TADKSSSTAY
MQLSSLTSED	SAVYYCAR ST	YGGDWYFNV	WGAGTTVTVS
AASTKGPSVF	PLAPSSKSTS	GGTAALGCLV	DYFPEPVTV
SWNSGALTSG	VHTFPA VLQS	SGLYSLSSV	TVPSSSLGTQ
TYICNVNHKP	SNTKVDKKA	PKSCDKTHTC	PPCPAPELLG
GPSVFLFPPK	PKDTLMISRT	PEVTCVVVDV	SHEDPEVKFN
WYVDGVEVHN	AKTKPREEQY	N STYRVVSVL	TVLHQDWLNG
KEYKCKVSNK	ALPAPIEKTI	SKAKGQPREP	QVYTLPPSRD
ELTKNQVSLT	CLVKGFYPSD	IAVEWESNGQ	PENNYKTTTP
VLDSGDGFF L	YSKLTVDKSR	WQQGNVFSCS	VMHEALHNHY
TQKSLSLSPG	K		

Amino acid sequences of light and heavy chains of rituximab obtained from references [26-28]. Murine variable domains indicated in red and blue color; complementary determining regions indicated in blue letters. The N₃₀₁ glycosylation site is indicated in green.

Supplementary Table 1:

Table 1. Theoretical rituximab trypsin enzymatic digestion. Amino acid sequences of light and heavy chains taken from references [26-28]

<i>Sequence position</i>	<i>Missed cleavage</i>	<i>Theoretical mass¹</i>	<i>Sequence²</i>
[331-338]	0	838,50	<i>k.ALPAPIEK.t</i>
[61-76]	0	1606,77	<i>R.FSGSGSGTYSYSLTISR.v</i>
[279-292]	0	1677,80	<i>k.FNWWYVDGVEVHNAK.t</i>
[126-141]	0	1740,87	<i>k.SGTASVVCLLNNFYPR.e</i>
[24-38]	0	1791,81	<i>k.ASGYTFTSYNMHWVK.q</i>
[44-63]	0	2183,04	<i>r.GLEWIGAIYPGNGDTSYNQK.f</i>
[306-324]	1	2228,20	<i>r.VVSVLTVLHQDWLNGKEYK.c</i>
[145-168]	1	2677,27	<i>k.VQWKVDNALQSGNSQESVTEQDSK.d</i>
[306-321]	0	1808,00	<i>r.VVSVLTVLHQDWLNGK.e</i>
[260-278]	0	2082,00	<i>r.TPEVTCVVVDVSHEDPEVK.f</i>
[108-125]	0	1946,07	<i>r.TVAAPSVFIFPPSDEQLK.s</i>
[126-137]	0	1186,64	<i>k.GPSVFPLAPSSK.s</i>
[190-206]	0	1818,90	<i>k.VYACEVTHQGLSSPVTK.s</i>
[75-98]	0	2620,15	<i>k.SSSTAYMQLSSLTSEDSAVYYCAR.s</i>
[107-125]	1	2102,12	<i>k.RTVAAPSVFIFPPSDEQLK.s</i>
[253-259]	0	835,43	<i>k.DTLMISR.t</i>
[99-125]	0	2889,33	<i>r.STYYGGDWYFNWVWAGTTVTVSAASTK.g</i>
[68-98]	1	3320,52	<i>k.ATLTADKSSSTAYMQLSSLTSEDSAVYYCAR.s</i>
[39-43]	0	558,29	<i>k.QTPGR.g</i>
[39-63]	1	2722,32	<i>k.QTPGRGLEWIGAIYPGNGDTSYNQK.f</i>
[349-364]	1	1872,97	<i>r.EPQVYTLPPSRDELTK.n</i>
[293-305]	1	1671,80	<i>k.TKPREEQYNSTYR.v</i>
[145-148]	0	560,31	<i>k.VQWK.v</i>
[188-206]	1	2084,05	<i>k.HKVYACEVTHQGLSSPVTK.s</i>
[44-65]	1	2458,20	<i>r.GLEWIGAIYPGNGDTSYNQKFK.g</i>
[414-418]	0	575,33	<i>k.LTVDK.s</i>
[327-338]	1	1266,74	<i>k.VSNKALPAPIEK.t</i>
[1-24]	1	2516,32	<i>.QIVLSQSPAILSASPGEKVTMTCR.a</i>
[207-213]	1	812,33	<i>k.SFNRGEC.</i>
[421-443]	0	2744,24	<i>r.WQQGNVFCSCVMHEALHNHYTQK.s</i>
[207-210]	0	523,26	<i>k.SFNR.g</i>
[343-348]	1	656,38	<i>k.AKGQPR.e</i>
[19-24]	0	710,33	<i>k.VTMTCR.a</i>
[397-413]	0	1873,92	<i>k.TTPPVLDSDGSFFLYSK.l</i>
[218-222]	1	600,37	<i>k.KVEPK.s</i>
[183-189]	1	890,43	<i>k.ADYEKHK.v</i>

[66-74]	1	904,50	<i>k.GKATLTADK.s</i>
[331-342]	1	1267,76	<i>k.ALPAPIEKTISK.a</i>
[360-364]	0	605,31	<i>r.DELTK.n</i>
[339-344]	1	647,40	<i>k.TISKAK.g</i>
[68-74]	0	719,39	<i>k.ATLTADK.s</i>
[444-451]	0	788,45	<i>k.SLSLSPGK.</i>
[349-359]	0	1286,67	<i>r.EPQVYTLPPSR.d</i>
[149-168]	0	2135,96	<i>k.VDNALQSGNSQESVTEQDSK.d</i>
[24-43]	1	2331,09	<i>k.ASGYFTSYNMHWKQTPGR.g</i>
[375-396]	0	2544,13	<i>k.GFYPSDIAVEWESNGQPENNYK.t</i>
[293-296]	0	501,31	<i>k.TKPR.e</i>
[103-106]	0	502,32	<i>k.LEIK.r</i>
[183-187]	0	625,28	<i>k.ADYEK.h</i>
[103-107]	1	658,42	<i>k.LEIKR.t</i>
[322-326]	1	670,32	<i>k.EYKCK.v</i>
[325-330]	1	678,36	<i>k.CKVSNK.a</i>
[414-420]	1	818,47	<i>k.LTVDKSR.w</i>
[142-148]	1	888,49	<i>r.EAKVQWK.v</i>
[219-226]	1	905,43	<i>k.VEPKSCDK.t</i>
[365-374]	0	1104,60	<i>k.NQVSLTCLVK.g</i>
[297-305]	0	1189,51	<i>r.EEQYNSTYR.v</i>
[138-151]	0	1264,65	<i>k.STSGGTAALGCLVK.d</i>
[169-182]	0	1502,75	<i>k.DSTYLSSTLTLSK.a</i>
[360-374]	1	1690,90	<i>r.DELTKNQVSLTCLVK.g</i>
[345-359]	1	1724,90	<i>k.GQPREPQVYTLPPSR.d</i>
[1-18]	0	1825,00	<i>.QIVLSQSPAILSASPGEK.v</i>
[1-19]	0	1977,11	<i>.QVQLQQPGAELVKPGASVK.m</i>
[126-144]	1	2069,04	<i>k.SGTASVVCLLNNFYPREAK.v</i>
[169-187]	1	2109,02	<i>k.DSTYLSSTLTLSKADYEK.h</i>
[279-296]	1	2160,09	<i>k.FNWWYVDGVEVHNAKTKPR.e</i>
[20-38]	1	2240,99	<i>k.MSCKASGYFTSYNMHWVK.q</i>
[190-210]	1	2323,14	<i>k.VYACEVTHQGLSSPVTKSFNR.g</i>
[1-23]	1	2426,28	<i>.QVQLQQPGAELVKPGASVKMSCK.a</i>
[397-418]	1	2430,24	<i>k.TTPPVLDSDGSFFLYSKLTVDK.s</i>
[126-151]	1	2432,28	<i>k.GPSVFPLAPSSKSTSGGTAALGCLVK.d</i>
[227-252]	0	2730,41	<i>k.THTCPPCPAPELLGGPSVFLFPPKPK.d</i>
[77-102]	0	2821,24	<i>r.VEAEDAATYYCQQWTSNPPTFGGGTK.l</i>
[253-278]	1	2898,42	<i>k.DTLMISRTPEVTCVVVDVSHEDPEVK.f</i>
[297-321]	1	2978,50	<i>r.EEQYNSTYRVSVLTVLHQDWLNGK.e</i>
[419-443]	1	2987,37	<i>k.SRWQQGNVFSCSVMHEALHNHYTQK.s</i>
[223-252]	1	3163,57	<i>k.SCDKTHCPPCPAPELLGGPSVFLFPPKPK.d</i>
[77-106]	1	3304,54	<i>r.VEAEDAATYYCQQWTSNPPTFGGGTKLEIK.r</i>
[421-451]	1	3513,67	<i>r.WQQGNVFSCSVMHEALHNHYTQKSLSPGK.</i>
[227-259]	1	3546,83	<i>k.THTCPPCPAPELLGGPSVFLFPPKPKDTLMISR.t</i>

[149-182]	1	3619,70	<i>k.VDNALQSGNSQESVTEQDSKDYSLSSSTLTLTK.a</i>
[365-396]	1	3629,72	<i>k.NQVSLTCLVKGFYPSDIAVEWESNGQPENNYK.t</i>
[108-141]	1	3667,88	<i>r.TVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPR.e</i>
[260-292]	1	3740,79	<i>r.TPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAK.t</i>
[25-60]	0	3933,00	<i>r.ASSVSYIHWFQQKPGSSPKPWIYATSNLASGVPVR.f</i>
[99-137]	1	4056,96	<i>r.STYYGGDWYFNWGAGTTVTVSAASTKGPSVFPLAPSSK.s</i>
[375-413]	1	4399,03	<i>k.GFYPSDIAVEWESNGQPENNYKTTPPVLDSGSFFLYSK.l</i>
[61-102]	1	4408,99	<i>r.FSGSGSGTSYSLTISRVEAEDAATYYCQQWTSNPPTFGGGTK.l</i>
[19-60]	1	4624,32	<i>k.VTMTCRASSVSYIHWFQQKPGSSPKPWIYATSNLASGVPVR.f</i>

(1) Calculated in positive mode.

(2) Sequences indicated in blue are from peptides which were detected at least once throughout the long-term study.

(3) Amino acids sequences indicated in red are from the complementary determining regions.