MS Bioworks
Protein Mass Spectrometry Services

3950 Varsity Drive
Ann Arbor, MI 48108
734-929-5083
www.msbioworks.com
info@msbioworks.com

Project Report

Information

Client:	
Institute:	
MSB Project Number:	MSB-
Date submitted:	
Date completed:	

Samples

Client identifier	MSB identifier	Notes
Biosimilar	A	Trypsin
	B	Chymotrypsin
	C	Elastase

Objective

Confirmation of sequence and identification of single site amino acid substitutions using MSB-03 and error tolerant database searching.

Experimental Methods

Sample Preparation

Trypsin digestion was performed using a ProGest robot (DigiLab) with the following protocol:

- Washed with 25mM ammonium bicarbonate followed by acetonitrile.
- Reduced with 10mM dithiothreitol at 60°C followed by alkylation with 50mM iodoacetamide at RT.
- Digested with sequencing grade trypsin (Promega) at 37°C for 4h.
- Quenched with formic acid and the supernatant was analyzed directly without further processing.

Chymotrypsin and elastase digests were performed manually with the following protocol:

- Washed with 25mM ammonium bicarbonate followed by acetonitrile.

- Reduced with 10mM dithiothreitol at 60°C followed by alkylation with 50mM iodoacetamide at RT.
- Digested with chymotrypsin/elastase (Promega) at 37°C overnight.
- Quenched with formic acid and the supernatant was analyzed directly without further processing.

Mass Spectrometry

Each gel digest was analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher LTQ Orbitrap Velos. Peptides were loaded on a trapping column and eluted over a 75µm analytical column at 350nL/min; both columns were packed with Jupiter Proteo resin (Phenomenex). The mass spectrometer was operated in data-dependent mode, with MS performed in the Orbitrap at 60,000 FWHM resolution and MS/MS performed in the LTQ. The fifteen most abundant ions were selected for MS/MS.

Data Processing

Data were searched using a local copy of Mascot with the following parameters:

Enzyme: Trypsin/P or None (Chymotrypsin and Elastase)
Database: Custom biotech drug database appended with *Biosimilar Heavy Chain (concatenated forward and reverse plus common contaminants)
Fixed modification: Carbamidomethyl (C)
Variable modifications: Oxidation (M), Acetyl (N-term), Pyro-Glu (N-term Q), Deamidation (N,Q),
Mass values: Monoisotopic
Peptide Mass Tolerance: 10 ppm
Fragment Mass Tolerance: 0.8 Da
Max Missed Cleavages: 2

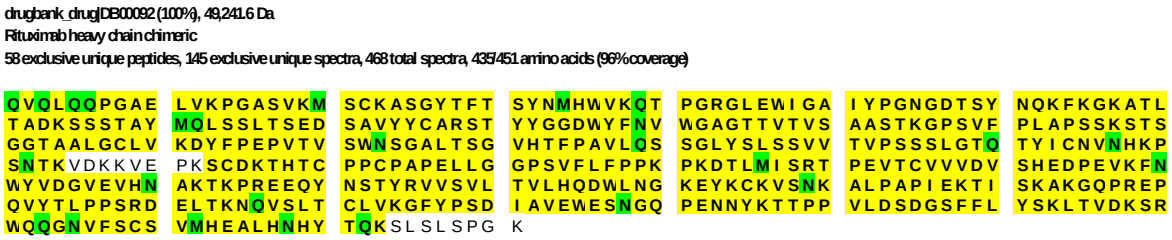
*
>Biosimilar Heavy Chain
QVQLQQPGAELVKPGASVKMSCKASGYTFTSYNMHWVKQTPGRGLEWIGAIYPGNGDTSYNQKFKGKATLTADKSS
STAYMQLSSLTSEDSAVYYCARSTYYGGDWYFNVWGAGTTVTVSAASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFP
EPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKvepkSCDKTHTCPPCPA
PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH
QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPEN
NYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGk

Mascot DAT files were parsed into the Scaffold algorithm for validation, filtering and to create a non-redundant list per sample. Data were filtered using a minimum protein value of 90%, a minimum peptide value of 50% (Prophet scores) and requiring at least two unique peptides per protein.

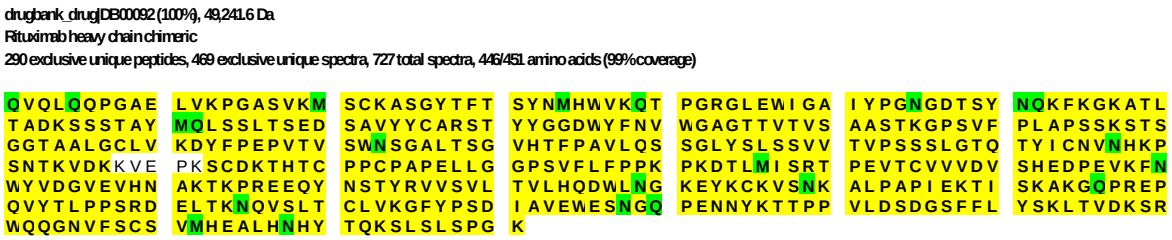
Results

Detection and Sequence Coverage

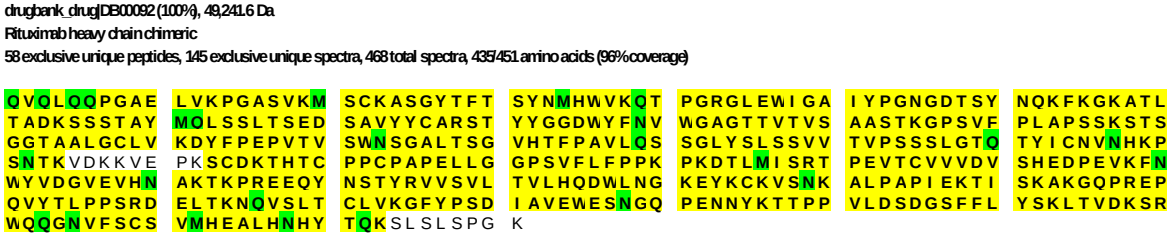
The target protein was detected with 98% sequence coverage when combining all three enzyme datasets. The combined sequence coverage map is below.



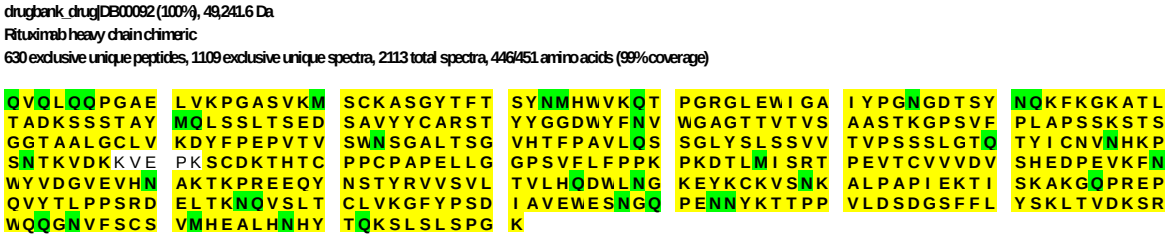
Trypsin sequence coverage



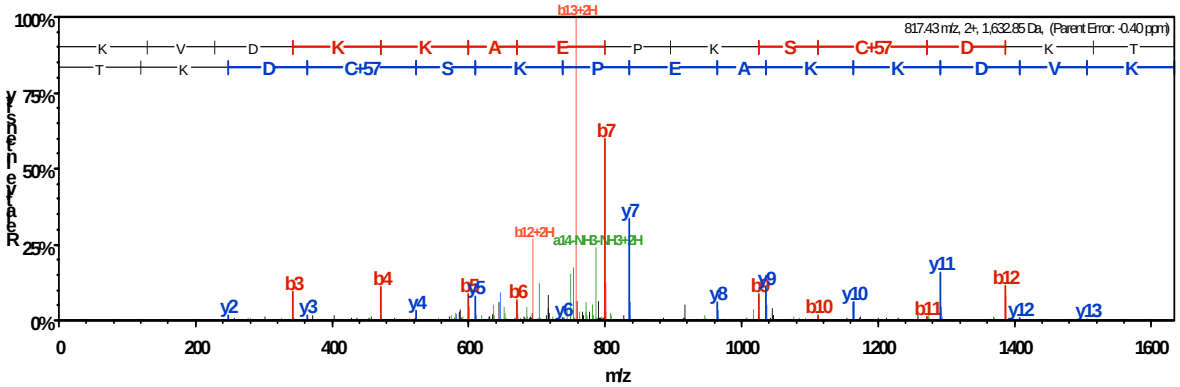
Chymotrypsin sequence coverage



Elastase sequence coverage



The unconfirmed portion of the sequence was suspected to be the site of an error in the suggested sequence. An error tolerant search identified a V →A at position 219. The spectrum confirming the presence of the alanine is below.



The database was modified to include the sequence variant and the search repeated. The target protein was identified with 100% sequence coverage.

drugbank_drug|DB00092 (100%), 49,213.6 Da
Rituximab heavy chain chimeric
659 exclusive unique peptides, 1165 exclusive unique spectra, 2207 total spectra, 451/451 amino acids (100% coverage)

QVQLQQPGAE LVKPGASVKM SCKASGYTFT SYNMHWVKQT PGRGLEWIGA IYPGNGDTSY NQKFKGKATL
TADKSSSTAY MQLSSLTSED SAVYYCARST YYGGDWYFNV WGAGTTVTVS AASTKGPSVF PLAPSSKSTS
GGTAALGCLV KDYFPEPVTV SWNSGALTSG VHTFPAVLQS SGLYSLSSVV TVPSSSLGTQ TYICNVNHKP
SNTKVDKKAE PKSCDKTHTC PPCPAPELLG GPSVFLFPPK PKDTLMISRT PEVTCVVVDV SHEDPEVKFN
WYVDGVEVHN AKTKPREEQY NSTYRVVSVL TVLHQDWLNG KEYKCKVSNK ALPAPIEKTI SKAKGQPREP
QVYTLPPSRD ELTKNQVSLT CLVKGFYPSD IAVEWESNGQ PENNYKTTPP VLDSDGSFFL YSKLTVDKSR
WQQGNVFSCS VMHEALHNHY TQKSLSLSPG K