Published in final edited form as:

Proteomics. 2011 February ; 11(3): 395–405. doi:10.1002/pmic.201000487.

Resurrection of a Clinical Antibody: Template ProteoGenomic *de novo* Proteomic Sequencing and Reverse Engineering of an Anti-Lymphotoxin Alpha Antibody

Natalie E. Castellana^{1,#}, Krista McCutcheon^{2,#}, Victoria C. Pham³, Kristin Harden³, Allen Nguyen⁴, Judy Young⁴, Camellia Adams², Kurt Schroeder², David Arnott³, Vineet Bafna¹, Jane L. Grogan^{5,*}, and Jennie R. Lill^{3,*}

¹ Department of Computer Science, University of California-San Diego, 9500 Gilman Drive, San Diego, CA 92101

² Department of Antibody Engineering, Genentech Inc. South San Francisco, CA 94044

³ Department of Protein Chemistry, Genentech Inc. South San Francisco, CA 94044

⁴ Department of Assay Technology, Genentech Inc. South San Francisco, CA 94044

⁵ Department of Immunology, Genentech Inc. South San Francisco, CA 94044

Abstract

A mouse hybridoma antibody directed against a member of the TNF-superfamily, lymphotoxin alpha (LT- α), was isolated from stored mouse ascites and purified to homogeneity. After more than a decade of storage the genetic material was not available for cloning, however biochemical assays with the ascites showed this antibody against LT- α (LT-3F12) to be a pre-clinical candidate for the treatment of several inflammatory pathologies. We have successfully rescued the LT-3F12 antibody by performing mass spectrometric analysis, primary amino acid sequence determination by template proteogenomics, and synthesis of the corresponding recombinant DNA by reverse engineering. The resurrected antibody was expressed, purified and shown to demonstrate the desired specificity and binding properties in a panel of immuno-biochemical tests. The work described herein demonstrates the powerful combination of high throughput informatic proteomic *de novo* sequencing with reverse engineering to re-establish monoclonal antibody expressing cells from archived protein sample, exemplifying the development of novel therapeutics from cryptic protein sources.

Keywords

Antibody; *De Novo* Sequencing; Lymphotoxin-alpha; Reverse Engineering

*Correspondence: Jennie Lill, Department of Protein Chemistry, Genentech, 1 DNA Way, MS413a, South San Francisco, CA 94080, USA, lill.jennie@gene.com. Jane L. Grogan, Department of Immunology, Genentech, 1 DNA Way, MS229, South San Francisco, CA 94080, USA, jgrogan@gene.com.

#These authors contributed equally to this work.

Competing Interests

The authors declare competing financial interests. Work was partly performed at, and funded by Genentech Inc. a wholly owned subsidiary of Roche Pharmaceuticals which develops and markets drugs for profit.

1.0 Introduction

Monoclonal antibody (mAb) therapies have been shown to be effective in the treatment of several inflammatory and autoimmune diseases. The range of therapies includes those that target and neutralize inflammatory cytokines, disrupt T cell activation, or deplete specific immune cell types associated with disease [1,2,3,4]. Lymphotoxin (LT)- α belongs to the tumor necrosis factor (TNF) super-family and is secreted as a homo-trimer (LT- α_3), or is expressed on the cell surface in complex with LT- β , predominantly as LT- $\alpha_1\beta_2$ hetero-trimers [5]. Secreted LT- α_3 binds the TNF receptors TNFRI and TNFRII and shares many functions with TNF- α . In contrast, LT- $\alpha_1\beta_2$ binds uniquely to the LT- β R receptor. LT- β R-mediated signals are thought to be crucial for robust immune responses, by orchestrating germinal center architecture, and regulating normal development of secondary lymph nodes. Clinically, LT- β R has been implicated in the development of tertiary lymphoid structures in chronically inflamed tissue associated with autoimmune disease [6]. In rheumatoid arthritis (RA) patients, LT- α and LT- β expression are elevated in the synovium [7]. Moreover, CD4⁺ subsets of T helper cells (Th1 and Th17), implicated in the pathology of RA, express surface and soluble LT- α [8]. Recently, an anti-murine LT- α mAb was reported to neutralize soluble LT- α_3 and deplete specific pathogenic subsets of immune cells expressing LT- α on their surface [8]. The demonstrated efficacy of this mechanism in inhibiting T cell-mediated diseases in a murine model of arthritis prompted us to look for monoclonal antibodies directed against human LT- α with similar properties.

Lymphotoxin was explored over 10 years ago at Genentech as a cancer adjuvant, and multiple antibodies were generated in mice to human LT- α_3 at that time [9]. Recent research scanning the Immune Response *In Silico* (IRIS) database for specific genes involved in the depletion of T_H1 and T_H17 cellular subsets identified LT- α as an attractive therapeutic target, reviving interest in this archived mAb collection [8]. One mAb (LT-3F12) exhibited excellent binding and blocking of LT- α to its different receptors and was deemed of interest as a tool for the pre-clinical investigation of the mechanism of blocking LT- α binding. The DNA, RNA and original cell line were no longer available (as these had not been maintained); therefore *de novo* sequencing was employed to recover the amino acid sequence identity of the purified antibody.

The *de novo* sequencing of a monoclonal antibody has previously been described whereby overlapping proteolytic digestions were performed, followed by mass spectrometric analysis and sequence determination using an automated suite of bioinformatic tools using Comparative Shotgun Proteomic Sequencing (CSPS) [10]. An alternative approach, termed template proteogenomics and implemented in a tool called GenoMS takes as input a collection of spectra (acquired from multiple protease digests), and a collection of imperfect templates and constraints (see methods section). At the heart of the approach is a novel method of extending a target amino acid sequence, by recruiting and aligning spectra that partially match, and has previously been used to sequence two antibodies in under 24 hours [11]. By using template proteogenomics applied to spectral datasets with multiple protease digests, 100% of the light chain and 94% of the heavy chain for the anti-LT- α antibody LT-3F12 were identified.

Complete sequence information for the variable domain of the LT-3F12 antibody was initially determined using GenoMS and then reconstructed into its full mouse IgG2b kappa form by reverse engineering of the protein sequence into rDNA for expression in CHO cells. *In vitro* testing of this resurrected molecule exhibited the desired binding and blocking capabilities.

To our knowledge, this is the first example of combining automated proteomic sequencing with reverse DNA engineering to produce a fully functional recombinant pre-clinical monoclonal antibody.

2.0 Materials & Methods

2.1 Mouse Ascites Production, Isolation/Purification of Antibody

The anti-LT antibody LT-3F12 from the cell line LT-3F12.2D3 was purified on Protein A Sepharose (RepliGen Corp., USA), after adjusting the salt concentration and pH. Specifically, glycine was added to the ascites to a final concentration of 1.5M, NaCl was added to 3M, and pH was adjusted to 8.6 with NaOH. After filtering the solution through a 0.45 μ m filter, it was loaded onto ~ 40 mL Protein A column at 2mL/min on a Pharmacia FPLC system using Load buffer (1.5 M Glycine, 3 M NaCl, pH 8.6). After washing with 5 column volumes of Load buffer, the antibody was eluted with 3 column volumes of 0.1M glycine, pH 2.9, immediately neutralized with 1 M Tris pH 8.0, and then dialyzed against phosphate buffered saline (PBS). Dialyzed material was sterile filtered (0.2 μ m) and stored at 4°C.

2.2 Fab Generation

The antibody was reconstituted in digestion buffer (20 mM Phosphate Buffer, pH 6.8 containing 20 mM Cysteine-HCl). The antibody (1 mg/mL) was incubated with immobilized papain resin 0.1 mL settled gel/mg IgG overnight at 37°C with rotation. After incubation the sample was centrifuged to separate the immobilized papain from the Fab digested mixture. The Fab fragment was then purified using an SP sepharose, strong cation exchange resin (1ml hitrap column), load buffer 50 mM NaOAc pH 5.0 and eluted with load buffer from 0–0.5M NaCl over 20 column volumes after which the sample was dialyzed against 1X PBS, pH 7.4

2.3 Proteolytic Digestions and Sample Preparation

The LT-3F12 antibody was diluted to a concentration of 0.5 mg/mL and 100 μ L was added to a 3 kD MWCO Centricon® (Millipore, Milford, MA). Samples were reduced in 5 mM DTT for 15 min. at 50°C, alkylated in n-isopropyl iodoacetamide (NIPIA; made in-house) in the dark for 30 min. and then centrifuged at 8,000 rpm for 10 min., to achieve a final volume of 10 μ L. NIPIA is routinely employed in our laboratory as it is an alkylated reagent which gives a good diagnostic peak for Cys in Edman degradation cycles (if required). Six separate vials containing 0.5 μ g antibody in 50 μ L were set up for digestion with trypsin (in 25 mM ammonium bicarbonate pH 8); pepsin (in 3 M acetic acid); Glu-C (in 25 mM ammonium bicarbonate pH 8); Chymotrypsin (in 25 mM ammonium bicarbonate pH 8), or Asp-N (in 50 mM Sodium Phosphate pH 6.5). All digestions were incubated for 4 hours at 37°C, except for pepsin, which was left at room temperature for 1 hour. After incubation, samples were spun through a 3 kDa MWCO filter, peptides were collected and dried to near completion in a speed-vacuum. Peptides were reconstituted into 20 μ L of 0.1% TFA and placed into an auto-sample vial.

2.4. Mass Spectrometric Analysis

Samples were injected (2 μ L) onto a reverse phase column (1.7 μ m BEH-130, 100 $\times$ 100 mm C18 column, Waters, Dublin, CA) via a Nano-acquity UPLC (Waters, Milford, MA). A flow rate of 1 μ L/min. was employed with a gradient of 2% to 90% solvent B (where solvent A is 0.1% formic acid in water and solvent B is 0.1% formic acid in ACN) applied over 40 min. with a total run time of 55 min. Samples were ionized using an Advance Nanospray ionization source (Michrom Bioresources, CA) and analyzed in the LTQ-Orbitrap mass spectrometer (ThermoFisher, San Jose, CA). Samples were either analyzed

using data dependent mode with full MS analysis performed at 60,000 FMHW in the Orbitrap with the top 8 most abundant ions fragmented by CID in the LTQ, or alternatively a full MS scan in the Orbitrap with 5 subsequent rounds of CID in the Orbitrap at 15,000 FMHW.

In addition to LC/MS/MS for analysis of proteolytic peptides from LT-3F12, the intact molecular weight analysis of both the heavy and light chain and also the separated fragments from the Fab were also performed. Samples were injected onto a reverse phase column (8 μ m 50 $\times$ 2.1mm PLRPS column, Varian, Lake Forest, CA) via an Agilent 1200 LC system (Agilent, Santa Clara, CA). The LC eluant was electro-sprayed into an Agilent TOF mass spectrometer where mass spectral data was acquired in the 200–6000 m/z range. Peaks from the total ion chromatogram were then integrated (Agilent Qualitative Analysis software) to obtain the raw charge state spectrum. Protein deconvolution of the m/z raw spectrum was generated using the mathematical Maximum Entropy algorithm. For determining the isotype of the antibody and for gaining constant region characterization a Mascot search (Matrix Science, London, UK) was performed with no enzyme selected and a tolerance of 25 ppm for the full MS and either 10 ppm for Orbitrap MS/MS data or 0.5 Da for LTQ MS/MS data respectively in the search parameters. All RAW data files generated can be found at <http://proteomics.ucsd.edu/3F12>.

2.5 Template Proteogenomic Analysis

GenoMS [11] is designed to predict a target protein sequence using a related sequence as a template. A database of all known V, D, J, and C segments for Mouse was constructed from the International Immunogenetics Information System (IMGT) GENE-DB [12]. These sequences were not identical to the anti-LT- α antibody heavy and light chains, but were used as guides for the template proteogenomic methods. The closest V segment to the true sequence had 91% identity to the anti-LT antibody heavy chain, and an additional 16 amino acids were completely missing from the database.

The template proteogenomic algorithm proceeds in three steps; *template-chain selection*, *anchor extension*, and *sequence construction*. In template-chain selection, templates from the database are scored, based on coverage by identified spectra. The highest scoring set of templates that obey user-defined order and mutual exclusion is selected for anchor extension. Anchors are identified on each selected template, using the set of peptides identified on the template. These are contiguous regions of the template that are identical to the target sequence. GenoMS extends anchors by recruiting spectra from peptides that partially overlap the anchors. The recruited spectra are aligned to produce a consensus spectrum. The sequence of the protein is determined by *de novo* sequencing the consensus spectra, and merging overlapping anchor extensions.

The template-chain selection method was modified from its original version to handle two features of the LT- α antibody data set that, although not uncommon, had not been identified on previous data sets to which the template proteogenomics approach had been applied. First, the pattern of mutations between the template sequence and the target antibody sequence were distributed across the antibody, and not clustered at hypervariable regions. This affected the ability of GenoMS to identify anchors and select high scoring templates. To address this challenge, we performed a mutation tolerant database search [13] on the templates, enabling the identification of 20 additional amino acids in the final sequence. InsPecT [14] is a database search tool which enables the identification of peptides with mutations and post-translational modifications. For mutation-tolerant searches, InsPecT allows a single mutation per peptide. The likelihood of the mutation according to the BLOSUM62 [15] matrix contributes to the score of the spectrum-peptide match.

The second challenge posed by the anti-LT- α antibody dataset was the mixture of spectra from the heavy and light chains. In previous work, GenoMS, was able to identify a single protein sequence at a time. Two independent runs of the algorithm were used to identify the two chains. We extended GenoMS to identify multiple template chains within the same database. See Supplemental Figure 1 (S1).

Once the two template chains have been determined, the algorithm proceeds as described above. The sequences of the two immunoglobulin chains are determined simultaneously.

2.6 Reconstructing the Antibody

The reconstructed LT-3F12 antibody amino acid sequence was aligned with a database of in-house recombinant monoclonal antibodies. Mouse monoclonal anti-TGF-beta antibody 2G7 corresponded well with the LT-3F12 V_H (IGHV1-80*01) and V_L (IGKV8-30*01) mouse germ-line sequences. To become LT-3F12, 2G7 required changes in 13 and 4 framework residues and 16 and 8 CDR residues in the V_H and V_L domains, respectively. The variable regions of the 2G7 mAb were previously sub-cloned into mammalian IgG expression pRK-based plasmids [16] containing either the constant region for mouse kappa or mouse IgG2a. Oligonucleotides with optimized expression codons (i.e. codons used in highly expressed genes) were designed to mutate the variable domains of 2G7 to the LT-3F12 sequence during 2 rounds of Kunkel mutagenesis on single stranded DNA preparations of 2G7 light chain and heavy chain pRK plasmids [17]. The first round of oligos included:

V_L

CA1845_GCTCATAGTGACCTTTTCTCCAACAGACACAGCCAGAGATGATGGC
GACTGTGACATCACGATATCTGAATGTACTCC,
CA1847_GCGATCAGGGACACCAGATTCCCTAGTGGATGCCC,
CA1848_GCCACGTCTTCAGCTTTTACACTGC TGATGG; and

V_H

CA1871_CCAGACTGCTGCAGCTGACCTTGTGAATGTACTCCAGTTGC,
CA1852_CCCACTCTATCCAGTAACTAGAGAAGGTGTATCCAGTAGCCTTGCA
GG, CA1854_GCCCTTGAACCTCTCATTGTAATTAGTACTACCACTTCCAGG,
CA1856_GGCTGAGGAGACGGTGACTGTGGTGCCTTGGCCCCAGTAGCCAT
GGTACCCGTCTGCACAGTAATAGACCGC.

The second round of oligos included:

V_L

CA1846_GCTGGTACCAGGCCAAGAAGTTCTTCTGATTGGTACTGTATAAAAG
ACTTTGACTGGACTTACAGC, CA1849_GGTCCCCCTCCGAACGTGCGCGGG
TAGGAGTAGTATTGCTGACAGTAATAAACTGCCAGGTC; and

V_H

CA1851_GCCTTGCAGGAGATCTTCACTGAAGCCCCAGGCTTCATCAGCTCAG
CTCC,
CA1853_CCACTTCCAGGACTAATCTCTCCAATCCACTCAAGGCCATGTCCAG
GCC, CA1855_CCGCAGAGTCCTCAGATGTGATCAGGCTGCTGAGCTGGATG
TAGGCAGTGTTGGAGGATTTGTCTGCATGAATGTTGCCTTGCC.

The final *in vitro* synthesized vectors were transformed into chemically competent XL-1 Blue cells (Stratagene, Agilent Technologies, USA) and mini-prep DNA from colonies submitted for sequencing. Plasmids for the heavy and light chains were co-transfected using

Fugene®6 (Roche Applied Science, Germany) into an adenovirus-transformed human embryonic kidney cell line, (293T, ATCC, USA), for transient expression. IgG was purified from conditioned media after 4–6 days, using recombinant Protein A Sepharose (RepliGen Corp., USA) affinity chromatography. Since the original LT-3F12 murine mAb was known to be of the IgG2b isotype, the V_H domain of LT-3F12 was also sub-cloned into an in-house pRK mouse IgG2b expression vector and expressed with the common LT-3F12 kappa light chain. The work flow employed for the de novo sequencing and reverse engineering of the 3F12 antibody is shown in Figure 1.

2.7 LT- α Biochemical and Biophysical Assays

For measurement of binding to LT α 3, MaxiSorp™ ELISA 96 well plates (Nunc, Thermo Fisher Scientific, NY, USA), were coated with 1 μ g/mL of human LT- α 3 in 50mM carbonate buffer (pH 9.6) (CB) (100 μ L/well) overnight. The unabsorbed solution was aspirated from the wells, before they were blocked with 150 μ L PBS containing 0.5 mg/mL bovine serum albumin (PBS-BSA) for 1–2 hours, and then washed with PBS containing 0.05% TWEEN™-20 (WB). Titrations of anti-LT α (purified from ascites vs. chimeric recloned antibody), TNFRII-Fc fusion protein, or isotype control IgG2a, diluted in 100 μ L assay diluent (AD, PBS-BSA, 0.05% TWEEN™-20, 10 ppm ProClin 300) were added to each well, incubated for 2 hours, and washed with WB. 100 μ L of 80 ng/mL horseradish peroxidase conjugated goat anti-mouse IgG (all subtypes) (Jackson ImmunoResearch Laboratories Inc., PA, USA), diluted in AD was then added to each well and incubated for 1 hour. Plates were then washed with WB and bound HRP was measured with a solution of tetramethylbenzidine (TMB)/H₂O₂ (KPL Inc., MD, USA). After 15 min., the reaction was quenched by the addition of 100 μ L of 1M phosphoric acid. The absorbance at 450 nm was read with a reference wavelength of 650 nm.

For measurement of antibody blocking human LT- α 3 ligand binding to TNFRII, microtiter wells were coated with 0.7 μ g/mL human TNFRII-huIgG1 (ENBREL™) in CB (100 μ L/well) overnight and unabsorbed solution was aspirated from the wells. LT- α 3-biotin (1.4 nM) was neutralized by preincubating with increasing concentrations of anti-LT α 3 3F12.ascites vs. ENBREL™ for 1 hour, before adding mixtures to coated microtiter plates. Plates were further incubated for 30 minutes, washed with WB, and then incubated for 30 minutes with streptavidin-horse radish peroxidase polymer (Sigma, USA) and developed as above for the binding ELISA.

FACS analysis was performed on a 293-LT- $\alpha\beta$ stable cell-line as previously described [18] Surface LT expression on 293-LT- $\alpha\beta$ cells was detected with LT-3F12.ascites, LT-3F12.chimera, and human LT β R-Fc [8] and detected using APC-conjugated anti-IgG (eBioscience). Samples were acquired on a FACSCalibur flow cytometer using CellQuest Pro v.5.1.1 software (BD Biosciences) and data analysis was conducted using FlowJo v6.4.2 software (Tree Star, Inc.).

The biomolecular interaction of LT- α 3 and LT-3F12 recombinant chimeric or ascites purified IgG were characterized by surface plasmon resonance on a BIACORE 3000 instrument. Amine chemistry was used to covalently capture 8000 response units (RU) of a goat anti-mouse or anti-human Fc polyclonal Fab' (Jackson Immunoresearch Laboratories Inc. USA) on three flow cells of a CM5 sensor chip (GE Healthcare, USA). All flow cells were blocked with ethanolamine, including the fourth flow cell, which was used as a reference. Between 200–500 RU of IgG was captured over the anti-Fab sensor chip using 5 μ g/mL IgG solutions flowing at a rate of 10 μ L/min in HBST (10 Mm hepes, pH 7.2, 150 mM NaCl, 3 mM EDTA, 0.005% polysorbate-20). After IgG capture, 1–25 nM of LT- α 3 was captured at a rate of 10 μ L/min in HBST. Regeneration between samples was done using 10 mM glycine, pH 1.5.

2.8 Functional Assays on Human Umbilical Endothelial Cells (HUVEC)

HUVEC (passages 3–7) were grown in Clonetics Endothelial Basal Media (EBM-2, Walkersville, MD) supplemented with EGM-2 SingleQuots (Clonetics, CA, USA). Cells were removed from sub-confluent flasks with Accutase (ICT, San Diego, CA), seeded at 15,000 cells/well in 50 μ L culture media in 96-well MSDTM High Bind plates (Meso Scale Discovery, Gaithersburg, MD), and allowed to attach for 1–2 hours in a humidified 37°C, 5% CO₂ incubator. Titrations of anti-LT- α (LT-3F12; ascites or chimeric) TNFRII-Fc fusion protein, or isotype control IgG₁ were pre-incubated with 0.1 nM LT- α 3 for 1 hour at RT with gentle shaking. Mixtures were added (50 μ L per well) to the wells containing HUVEC to give a total volume of 100 μ L per well. Cultures were incubated for 16 hours at 37°C in a humidified 5% CO₂ incubator to allow ICAM-1 expression. Plates were washed with 100 μ L cold HUVEC Assay Diluent (HAD; 2% fetal bovine serum (FBS), PBS). All subsequent steps were performed at 4°C on a shaker in the cold room and washed as described above between steps. Plates were blocked with 150 μ L of 30% FBS in PBS for 1–2 hours, incubated 3 hours with a 1:250 dilution of anti-human ICAM-1-biotin mAb (Bender MedSystems, CA, USA) in HAD, then incubated with MSDTM streptavidin-Sulfo TAG (1 μ g/mL) in HAD for 1 hour. Plates were read immediately after addition of 150 μ L 1x surfactant-free MSDTM Read Buffer T, on an MSDTM Sector Imager MA6000. The antibody concentration causing 50% inhibition of ICAM-1 expression (IC₅₀) was calculated using a four-parameter curve fitting program (Kaleidagraph, PA, USA).

3.0 Results

3.1 Analysis of antibodies from mouse ascites for binding and blocking of Lymphotoxin alpha

Archived mouse ascites generated from anti-LT- α hybridoma cells were screened for binding both soluble and membrane forms of LT- α and blocking the corresponding receptor binding activities. LT-3F12 was selected as an antibody that fulfilled these criteria. LT-3F12 bound LT- α 3 with an EC₅₀ of 18 pM, blocked LT- α 3 binding TNFRII-Fc at an IC₅₀ of 47 nM (similar to TNFRII-Fc itself; IC₅₀ 44 nM) (Figure 2a) and bound surface LT- α on 293-LT- α β cells with a 2-log shift in fluorescence intensity by FACS (Figure 2b). Biacore analysis showed LT-3F12 bound LT- α 3 with an overall KD of 0.3 nM (k_{on} -6.6×10^5 /M/s and k_{off} -1.7×10^4 /s).

3.2 Proteomic Analysis and Template ProteoGenomic De Novo Sequencing

Aliquots of the antibody were digested with five different proteolytic enzymes and each sample analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Database searching of the resulting product ion spectra resulted in approximately 75% sequence coverage of both the heavy and light chain constant regions, including a small portion of CDR L3 and H1. To determine the full sequence of the LT-3F12 antibody we employed the template proteogenomic techniques of GenoMS. Spectra were preprocessed by clustering related spectra and converted to Prefix Residue Mass Spectra prior to template proteogenomic analysis [11]. For the heavy chain, a template for the V gene (IGHV1-9*01) and C gene (IGHG2B*02) were identified, however, no template could be identified for the D and J genes and for these areas traditional techniques had to be performed to ensure correct amino acid assignment in this region[19]. 421 amino acids were predicted with 98% identity as compared to the final reverse-engineered LT-3F12 antibody sequence. The first pass of the algorithm resulted in missing sequence due to low spectral coverage. However, these regions could be inferred from the templates, increasing coverage to 434 amino acids with no loss in accuracy as compared to the final heavy chain sequence. The resulting sequence is shown in Figure 3a. Figure 3c shows a detailed view of the N-terminal anchor of the V region with recruited spectra and the resulting de novo sequenced consensus spectrum.

Light chain templates were selected for the V gene (IGKV8-30*01), J gene (IGKJ1*01), and C gene (IGKC*01). The template sequences were nearly identical to the target antibody (only 3 mutations in the V region). The extension was able to determine the exact germ line rearrangement boundaries. The predicted light chain sequence, consisting of 220 amino acids at 99% identity is shown in Figure 3b. All data can be found at <http://proteomics.ucsd.edu/3F12>.

To ensure that the sequences were correct for the data generated by template proteogenomics the Fab was generated and the theoretical intact molecular weights were calculated for the fully reduced and deglycosylated Fab light chain and heavy chain and these were compared to the masses observed from high resolution LC-MS intact molecular weight analysis. The intact MW for the Fab light chain was determined to be 24,372 Da, and the MW for the Fab HC 23,024 Da (Supplemental Figure S2). Both the light chain and heavy chain Fab masses matched the template proteogenomic generated sequence theoretical masses within 4 Daltons which are accounted for by incomplete reduction of the intra-disulfide bonds. Once it was determined that the masses were determined to be the same within the limits of the instrument resolution, the amino acid sequences were reverse translated into the corresponding DNA sequence.

3.3 Reverse Engineering & Chimerization of the LT-3F12 Antibody

Once the primary sequences of the antibody heavy and light chain variable domains were determined by template proteogenomics, we looked for the closest match to any of our in-house recombinant monoclonal antibodies (irrespective of antigen specificity) at the amino acid level. A previously cloned in-house mouse antibody, anti-TGF- β 2G7¹⁶ was found to be a reasonable match, requiring a total of 29 changes in the V_H region and 12 changes in the V_L region to match LT-3F12 (differences between 2G7 and LT-3F12 amino acids are indicated in purple boxes on the 2G7 sequence line in Figure 4). Although up to four oligonucleotides were simultaneously used for mutagenesis, the number and distribution of changes required two rounds of Kunkel mutagenesis on the 2G7 light chain or heavy chain plasmid templates to synthesize the final LT-3F12 sequence. Variable domain regions targeted by mutagenic oligonucleotides in rounds 1 and 2 to convert 2G7 recombinant DNA to the desired LT-3F12 DNA sequence are translated and shown in Figure 4, and the corresponding oligomeric DNA sequences are provided in section 2.6. An alignment to the mouse germ line V_H, IGKV_H1-80*01, and V_L, IGKV8-30*01, amino acid sequences from which LT-3F12 was derived *in vivo* is also shown in Figure 4. The perfect match of the obtained LT-3F12 V_L framework sequence to the corresponding mouse germ line IGKV8-30*01 framework sequence, lent confidence to the amino acid sequence obtained by *de novo* sequencing. The LT-3F12 V_H framework sequence contained only 5 divergences from the mouse germ line IGKV_H1-80*01 framework sequence, which likely represent junctional editing or somatic diversification *in vivo* [20]. Several residues in the LT-3F12 sequence are either leucine or isoleucine, amino acid structures that cannot be distinguished by mass spectrometric analysis using low energy collision induced dissociation. Only one of these isobaric amino acid residues, an isoleucine, at position 34 of CDR-H1, was not the germ-line mouse sequence (Figure 4), indicating somatic mutation from methionine during *in vivo* affinity maturation. With the exception of this one CDR isoleucine residue, the remaining residues were made to match the germ line sequence.

Because mouse antibody 2G7 was already in an IgG2a kappa expression plasmid, we did not need to generate or alter the constant regions of the antibody. However the V_H region from LT-3F12, once made in IgG2a, was sub-cloned into an in-house IgG2b vector to match the original antibody isotype, for expression and binding comparisons. ELISA binding comparisons showed no significant difference between the IgG2a or IgG2b format of LT-3F12 (data not shown).

3.4 Testing of the Reconstructed and Second Generation Antibodies

The original and synthetic LT-3F12 mAbs were compared in LT binding and blocking studies and found to be equivalent, indicating that the correct sequence had been recovered (and that any possible Ile/Leu substitutions were not significant) (Figure 5). The rescued LT-3F12 chimeric antibody bound LT- α 3 by ELISA with an EC₅₀ of 37 pM, comparable to the original ascites mAb (EC₅₀ 18 pM), while TNFR2-Fc fusion protein bound to LT- α 3 more weakly than the antibodies (EC₅₀ 98 pM estimated) (Figure 5a). LT β R-Fc did not bind to LT α 3, as expected. Moreover, the chimeric antibody bound surface LT- α β on 293-LT- α β with a comparable mean fluorescent intensity to the LT-3F12 ascites and LT β R-Fc (Figure 5b). To ensure the rescued LT-3F12.chimeric antibody retained the functional activity of the original ascites, the two antibodies were compared in a cell-based assay that measures blocking LT- α 3 induction of surface ICAM-1 expression on HUVEC (Figure 4c). LT- α 3 signaling induces target cells to up-regulate many chemokines and cytokines in an NF κ B-dependent manner [21]. LT-3F12.ascites and LT-3F12.chimera were equivalent in their ability to block ICAM-1 up-regulation on HUVEC (IC₅₀ 0.04 nM). These IC₅₀ values agree well with a 1:1 molar ratio of antibody inhibiting LT- α 3 trimers present in HUVEC culture medium, given that the antibodies are bivalent and each arm can bind and sequester one trimer. The mAbs were approximately five-fold more potent than TNFR2-Fc in this assay (IC₅₀ 0.19 nM).

BIACORE analysis demonstrated excellent LT-3F12 LT- α 3 binding kinetics (Figure 5d) with an overall KD of 0.3 nM, equivalent to the ascites-purified LT-3F12 IgG. Although overall KD were equivalent, the on-rate kinetics of the recombinant chimeric antibody was improved significantly over the ascites material and off-rate appeared moderately faster (about 2–4 fold). The BIACORE determinations were made over a year apart and 2-fold accuracy is anticipated. BIACORE analysis is quantitative and may be affected by degradation, aggregation or denaturation of the IgG and there was insufficient ascites to characterize this IgG and make side-by-side comparisons to the recombinant material. Both kinetic measurements of ascites and chimeric IgG would be subject to avidity effects as the LT- α 3 is a homodimer, allowing two arms of one IgG to bind simultaneously. Monovalent measurements with LT-3F12 Fab and LT- α 3 could be done to obtain 1:1 binding characteristics.

4.0 Discussion

Monoclonal antibodies have been widely demonstrated to be excellent tools for the treatment of pathologies ranging from cancer, inflammation and auto-immunity to infectious disease. Antibodies are typically generated against a target antigen of interest using one of several well-established techniques [22,23]. In some instances however, an antibody with ideal characteristics for pre-clinical or clinical consideration may be present only at the protein level and cannot be fabricated by means other than amino acid level sequencing and reverse engineering. Historically, this has been a very time consuming endeavor, but recent technology and informatic advances have reduced the analysis time to weeks or even days [10,11,19]. In this example we identified an antibody isolated from mouse ascites showing excellent binding and blocking activities to our target molecule of interest, LT- α , in both its soluble homotrimeric and membrane heterotrimeric forms. Although other efforts were exerted to duplicate the functionality of this antibody including generation of new antibodies, the LT-3F12 antibody remained a key molecule of interest for the LT- α pre-clinical investigations. In order to resurrect this antibody we first purified the antibody to homogeneity and then performed multiple enzymatic digestions followed by high resolution tandem mass spectrometric analysis and database searching and then employed a novel template proteogenomics approach to obtain the sequence of the antibody in its entirety at the protein level.

The first part of the challenge in any type of *de novo* sequencing analysis is the quantity of material available for analysis and the quality of the data produced. In this case < 3 mg of purified material allowed successful sequencing of the LT-3F12 antibody with the majority of this material (2mg) employed for Fab generation. To ensure quality tandem mass spectrometric data, high resolution MS/MS spectra were collected which allowed confident assignment of the amino acid residues and allowed us to distinguish between the near isobaric masses of lysine and glutamine.

By using the computational tool GenoMS, we were able to recover the majority of the heavy and light chain sequences of the LT-3F12 antibody in a matter of hours. The successful reverse engineering of the antibody demonstrates the sequencing accuracy of the method. Template proteogenomics offers a dramatic reduction in the time to antibody resurrection when used as a first pass method followed by low-throughput manual evaluation. The complexity of the mixture of heavy and light chain mass spectra, and the lack of strong template candidates due to mutations outside of the CDRs posed significant challenges to GenoMS. However, we extended the tool to accommodate the presence of multiple distinct proteins in the same mass spectrometry run with negligible increase in the running time of the algorithm. In addition, by allowing the inclusion of point mutations in peptides identified in the first phase of the algorithm, we are able to achieve greater coverage of the protein sequence.

Challenges in immunoglobulin sequencing by template proteogenomics persist beyond those that were encountered in the LT-3F12 antibody data set. Presently, GenoMS is able to identify at most one each of V, D, J, and C segments for use as templates. Polyclonal antibodies may contain protein sequence similar to multiple distinct germline segments. Chemical post-translational modifications pose another challenge. In the case of labile modifications, a single template sequence should produce multiple modified versions of an antibody sequence. The template proteogenomic algorithm must be expanded to handle the increased complexity of these samples.

The sequence was deemed to be “correct” based upon the agreement of both observed and theoretical masses for the Fab. In the majority of cases we see a deficit of 4 Da in the observed molecular weight compared to the theoretical molecular weight due to the presence of intra-chain disulphide bonds which are not reduced under the mild conditions used for reduction of inter-chain disulphides. Glycosylation of the heavy chain FC, usually occurring at a well-conserved motif (NXS/T) is a further complication that can be eliminated by Fab generation. Other sources of mass discrepancies include cation adducts such as Na and K and deamidation and oxidation events. The latter two in particular can be expected in long-archived samples, but did not prevent correlation of observed to theoretical molecular weights in the analysis of the LT-3F12 antibody.

After determining that that sequence was indeed correct by comparison of the theoretical to the observed intact molecular weights of the heavy and light chain of the Fab, recombinant DNA was constructed representing the genomic sequence encoding this protein. Importantly, the rescued chimeric antibody retained full potency compared to its parental ascites, in its ability to bind and block LT- α 3, as well as to bind surface LT α on cells.

There is no way to for the current approach to distinguish between the isobaric amino acid residues isoleucine and leucine, so in the CDRH1 sequence we decided to assign an isoleucine rather than leucine based on homology with other murine antibodies. It is uncertain how much difference in affinity of an antibody to an antigen is observed by switching these two isobaric amino acid residues and if this was a concern in future projects one could have both constructs synthesized, expressed and tested for affinity.

Although we synthesized our DNA by mutagenesis of the variable domain of an existing in-house recombinant antibody sequence, advances in DNA synthesis allows the direct construction of genetic material starting from information and raw chemicals. Improvements in, and affordability of custom DNA synthesis technology services are accelerating innovation across many areas of research and complement the advances in the proteomics arena. This workflow can be advantageous in several scenarios; For example if a Hybridoma cell line has become infected with a parasite of other agent not safe to work with, if one has an antibody producing B-cell that cannot be clonally grown or fused to an immortal cell line to generate a Hybridoma. This platform also allows for fast humanization compared to traditional techniques and is a robust method for the identification and resurrection of high affinity monoclonal antibodies.

Overall we have demonstrated through the powerful combination of high resolution mass spectrometry, Template ProteoGenomic *de novo* sequencing and reverse engineering and chimerization strategies that a purified antibody derived from ascites could be identified and synthesized into a recombinant DNA form, enabling pre-clinical studies. This accomplishment opens up the possibility of resurrecting other antibodies from obscure sources.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to Wilson Phung & Qui Phung for intact molecular weight analysis; and Devavani Chatterjea, Eugene Chiang and Michelle Francesco for early antibody characterization. Thanks to Sophia Lee, Richard Vandlen and Bob Kelley for their antibody engineering advice and expertise.

V.B. was supported by a grant from the NIH (P41-RR024851). N.E.C. was supported in part by the NSF IGERT Plant Systems Biology training grant # DGE-0504645.

Abbreviations

LT	Lymphotoxin
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
LC	Liquid chromatography

References

1. Waldmann TA. Immunotherapy: past, present and future. *Nature Medicine*. 2003; 9:269–277.
2. Martin F, Chan AC. B cell immunobiology: evolving concepts from the clinic. *Annu Rev Immunol*. 2006; 24:467–496.
3. Taylor PC, Feldmann M. Anti-TNF biologic agents: still the therapy of choice for rheumatoid arthritis. *Nat Rev Rheumatol*. 2009; 5:578–82. [PubMed: 19798034]
4. Ouyang W, Filvaroff E, Hu Y, Grogan J. Novel therapeutic targets along the Th17 pathway. *Eur J Immunol*. 2009; 39:670–5. [PubMed: 19283720]
5. Ware CF. Network communications: lymphotoxins, LIGHT and TNF. *Annu Rev Immunol*. 2005; 23:787–819. [PubMed: 15771586]
6. Weyand CM. Immunopathologic aspects of rheumatoid arthritis: who is the conductor and who plays the immunologic instrument? *J Rheumatol Suppl*. 2007; 79:9–14. [PubMed: 17611973]

7. Takemura S. Lymphoid neogenesis in rheumatoid synovitis. *J Immunol.* 2001; 167:1072–1080. [PubMed: 11441118]
8. Chiang EY, Kolumam GA, Yu X, Francesco M, Ivelja S, Peng I, Gribbling P, Shu J, Lee WP, Refino CJ, Balazs M, Paler-Martinez A, Nguyen A, Young J, Barck KH, Carano RA, Ferrando R, Diehl L, Chatterjea D, Grogan JL. Targeted depletion of lymphotoxin-alpha-expressing TH1 and TH17 cells inhibits autoimmune disease. *Nat Med.* 2009; 15:766–73. [PubMed: 19561618]
9. Wong GH, Kaspar RL, Zweiger G, Carlson C, Fong SE, Ehsani N, Vehar G. Strategies for manipulating apoptosis for cancer therapy with tumor necrosis factor and lymphotoxin. *J Cell Biochem.* 1996; 60:56–60. [PubMed: 8825416]
10. Bandeira N, Pham V, Pevzner P, Arnott D, Lill JR. Automated de novo protein sequencing of monoclonal antibodies. *Nat Biotechnol.* 2008; 26:1336–8. [PubMed: 19060866]
11. Castellana N, Pham V, Arnott A, Lill JR, Bafna V. Template Proteogenomics: A novel method for sequencing monoclonal antibodies. *Mol Cell Proteomics.* 2010; 9:1260–270. [PubMed: 20164058]
12. Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Muller W, Bontrop R, Lemaitre M, Malik A, Barbie V, Chaume D. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* 1999; 27:209–212. [PubMed: 9847182]
13. Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA. Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol.* 2005; 23:1562–1567. [PubMed: 16311586]
14. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem.* 2005; 77:4626–4639. [PubMed: 16013882]
15. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *PNAS.* 1992; 89:10915–10919. [PubMed: 1438297]
16. Gonzalez LC, Loyet KM, Calemme-Fenau J, Chauhan V, Wranik B, Ouyang W, Eaton DL. A coreceptor interaction between the CD28 and TNF receptor family members B and T lymphocyte attenuator and herpesvirus entry mediator. *PNAS.* 2005; 102:1116–1121. [PubMed: 15647361]
17. Kunkel TA, Bebenek K, McClary J. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* 1991; 204:125–139. [PubMed: 1943776]
18. Young J, Yu X, Wolslegel K, Nguyen A, Kung C, Chiang EY, Kolumam GA, Wei N, Wong WL, DeForge L, Townsend MJ, Grogan JL. Lymphotoxin-ab heterotrimers are cleaved by metalloproteinases and contribute to synovitis in rheumatoid arthritis. *Cytokine.* 2010 in press.
19. Pham V, Henzel WJ, Arnott D, Hymowitz S, Sandoval W, Truong BT, Lowman H, Lill JR. De novo proteomic sequencing of a monoclonal antibody raised against OX40 Ligand. *Analytical Biochemistry.* 2006; 352:77–86. [PubMed: 16545334]
20. Neuberger MS. Antibody diversification by somatic mutation: from Burnet onwards. *Immunology and Cell Biology.* 2008; 86:124–132. [PubMed: 18180793]
21. Hochman PS, Majeau GR, Mackay F, Browning JL. Pro-inflammatory responses are efficiently induced by homotrimeric but not heterotrimeric lymphotoxin ligands. *J Inflamm.* 1995; 46:220–34. [PubMed: 8878796]
22. Nissim A, Chernajovsky Y. Historical development of monoclonal antibody therapeutics. *Handb Exp Pharmacol.* 2008; 181:3–18. [PubMed: 18071939]
23. Bostrom J, Fuh G. Design and construction of synthetic phage-displayed Fab libraries. *Methods Mol Biol.* 2009; 562:17–35. [PubMed: 19554284]

Figure 1. A flow chart of the strategy undertaken to sequence and reverse engineer the LT-3F12 monoclonal antibody.

Figure 2.

LT-3F12.ascites specifically binds LT- α 3. (a) LT-3F12.ascites blocked binding of LT- α 3 to microtiter plates coated with TNFRII-Fc with an IC-50 similar to TNFRII-Fc in solution. Human LT- α 3-biotin was preincubated with titrations of 3F12.ascites or TNFRII-Fc, and then added to microtiter plates coated with TNFRII-Fc. LT α 3-biotin bound to coated TNFRII-Fc was detected with streptavidin-HRP followed by TMB substrate. (b). FACS of 293-LT α β cells stained with anti-LT-3F12.ascites (blue histogram) or isotype control antibody (grey histogram) and detected using APC-conjugated anti-IgG.

Figure 3.

The LT-3F12 mAb heavy chain and light chain reconstructed from protein template databases. The grey rectangles are anchors while the arrows, annotated with sequence, are the extended and merged sequence. Text above the anchors indicates the GI number of the template used and coordinates within the anchors indicate their position within the template. Red amino acids or sections of anchors were incorrectly predicted. Green segments of anchors are mutations recovered by the mutation tolerant database search. Blue sections indicate where and how many amino acids can be inferred from the template. The templates were constructed from a genomic accession; therefore several splice junctions are covered by anchors and are shown as dotted grey arcs. A: The heavy chain full sequence

reconstruction, shown with the predicted sequence for the variable region aligned to the called sequence of the variable region. B: The light chain full sequence reconstruction, shown with the predicted sequence for the variable region aligned to the actual sequence of the variable region. C: A detailed view of spectra recruited to the N-terminal anchor of the heavy chain. The recruited spectra are aligned, and the resulting de novo sequenced consensus spectrum is shown. Prior to recruitment the spectra are clustered and converted to prefix residue mass spectra as described previously [11]. Spectra from three rounds of anchor extension are shown with the final consensus spectrum.

Light chain			Kabat - CDR L1		
			Chothia - CDR L1		
			Contact - CDR L1		
Kabat number	1	2	3	4	5
IGHV8-30*01	D	I	V	M	S
GNE-LT2002	D	I	V	M	S
2G7	D	I	M	T	S
CA1845_round 1	D	I	V	M	S
CA1846_round 2	D	I	V	M	S
			Kabat - CDR L2		
			Chothia - CDR L2		
			Contact CDR L2		
Kabat number	37	38	39	40	41
IGHV8-30*01	Q	Q	K	P	G
GNE-LT2002	Q	Q	K	P	G
2G7	Q	Q	K	P	G
CA1847_round 1	Q	Q	K	P	G
CA1848_round 1	Q	Q	K	P	G
CA1849_round 2	Q	Q	K	P	G
			Kabat - CDR L3		
			Chothia - CDR L3		
			Contact - CDR L3		
Kabat number	72	73	74	75	76
IGHV8-30*01	T	L	T	I	S
GNE-LT2002	T	L	T	I	S
2G7	T	L	T	I	S
CA1848_round 1	T	L	T	I	S
CA1849_round 2	T	L	T	I	S
			Kabat - CDR H1		
			Chothia - CDR H1		
			Contact - CDR H1		
Kabat number	1	2	3	4	5
IGHV1-80*01	Q	V	Q	L	Q
GNE-LT2002	Q	V	Q	L	Q
2G7	Q	V	Q	L	Q
CA1871_round 1	Q	V	Q	L	Q
CA1882_round 1	Q	V	Q	L	Q
CA1851_round 2	Q	V	Q	L	Q
CA1853_round 2	Q	V	Q	L	Q
			Kabat - CDR H2		
			Chothia - CDR H2		
			Contact - CDR H2		
Kabat number	44	45	46	47	48
IGHV1-80*01	G	L	E	W	I
GNE-LT2002	G	L	E	W	I
2G7	G	L	E	W	I
CA1854_round 1	G	L	E	W	I
CA1853_round 2	G	L	E	W	I
CA1855_round 2	G	L	E	W	I
			Kabat - CDR H3		
			Chothia - CDR H3		
			Contact - CDR H3		
Kabat number	83	84	85	86	87
IGHV1-80*01	T	S	E	D	S
GNE-LT2002	T	S	E	D	S
2G7	T	S	E	D	S
CA1856_round 1	T	S	E	D	S
CA1855_round 2	T	S	E	D	S

Figure 4.

The amino acids employed for light and heavy chain synthesis. Variable domain regions targeted by mutagenic oligonucleotides in rounds 1 and 2 to convert 2G7 recombinant DNA to the desired LT-3F12 DNA sequence.

Figure 5.

Equivalent potency of LT-3F12 ascites vs. rescued chimera in binding and blocking LT- α 3. (a). Binding of anti-LT-3F12.chim, anti-LT-3F12.ascites, TNFRII-Fc and LT β R-Fc to plates coated with LT- α 3. EC₅₀ values for each molecule are shown. (b) FACS of 293-LT α β cells stained with anti-LT-3F12.ascites (blue histogram), anti-LT-3F12.chim (green histogram), LT β R-Fc (black line) or isotype control antibody (grey histogram) and detected using APC-conjugated anti-IgG. (c) HUVEC grown in microtiter plates were stimulated with LT- α 3 pre-incubated with titrations of anti-LT-3F12.ascites, anti-LT-3F12.chim, TNFRII-Fc, or isotype control mouse IgG2a antibody. Blocking of LT- α 3 induction of ICAM expression on HUVEC and IC₅₀ values are shown. (D) Surface plasmon resonance (BIACORE) analyses of the LT-3F12 chimeric and ascites IgG binding kinetics to LT- α 3 demonstrate sub-nanomolar binding (KD 0.3 Nm. IgG was captured by the Fc region on a CM5 sensor chip covalently coupled with goat anti-species Fc Fab'2 pAb. Traces are shown illustrating

the on-rate and off-rate kinetic profile for a titration of LT- α 3 from 1.25 to 10 nM flowed over LT-3F12 chimeric IgG.