# Analysis of the heterogeneity of immunoglobulins in human serum

**Kenneth C. Parker and Stephen J. Hattan** SimulTOF Systems, Sudbury, MA Poster

**Background**: Immunoglobulins are typically very heterogeneous in human sera. In certain disease states, the diversity becomes skewed such that certain clonal antibodies become dominant, or so that antibodies produced from certain immunoglobulin variable region families become enriched or depleted. It would be useful to get a snapshot of immunoglobulin diversity by using a protocol that is as simple as possible. This might be feasible by linear MALDI, but to identify heavy chain classes and variable region families, tryptic peptides provide relevant information. In MALDI of complex mixtures, arginine-containing (R) peptides are most readily detected, and there are relatively few R peptides from Ig constant regions. To test these ideas, we have developed methods to perform peptide mass fingerprinting and MS-MS database searching against immunoglobulin-specific protein databases (and ordinary human databases). The samples consist of unseparated protein digests, which yields information about the relative concentration of immunoglobulin variable reagion families.

## Experimental System:

- Buy Human Serum from Sigma, and Goat anti hemoglobin serum.
- Purify Immunoglobulins using ProteinA membrane (see Poster by Hattan).
- Reduce and digest with bovine trypsin (Worthington).
- Dilute into HCCA matrix, and spot.
- Collect reflectron spectra, and MS-MS Spectra.
- Look for pattern of patterns corresponding to Ig Constant and Variable regions.
- Compare scores of best Ig sequence to the rest of human proteome.
- Search against human SwissProt or human IMGT database.
- Proprietary software used for PMF and MSMS.
- For MSMS analyses, individual sequences are pooled, and for Ig, many distinct proteins likely can be digested to generate each peptide.

**Fig. 1A. Tryptic digest of whole serum (from Sigma), dominated by albumin.**

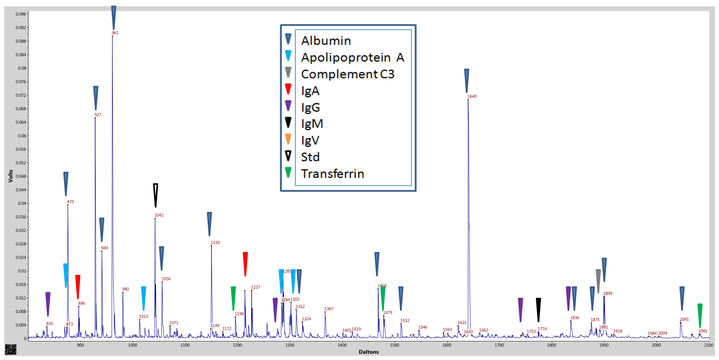
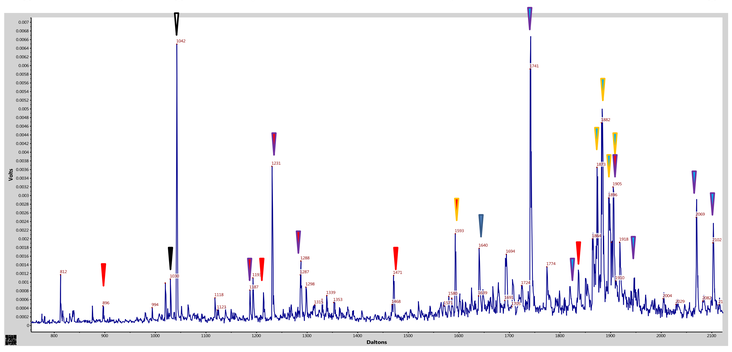
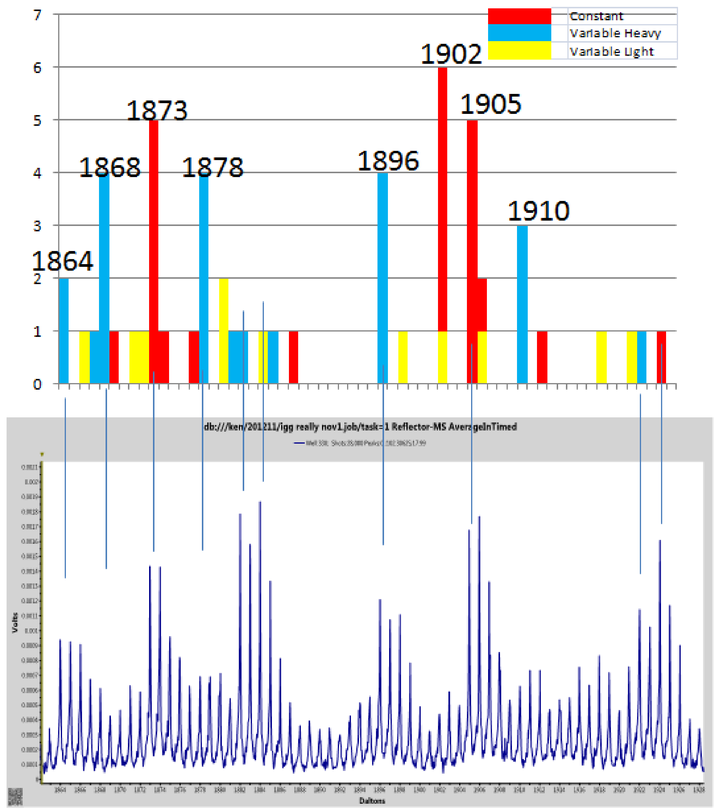



Fig 1**B. Trypic digest of proteins purified by Protein A from same digest.**




Cartoons of PMF identifications based on real mass spectra; mostly confirmed by multiple MS-MS spectra from the same (or similarly prepared digests). Note that many other peptides can be mapped at mz values higher than shown in the figure. Most strong signals correspond to Ig constant regions (including IgA constant regions), but many signals are clearly derived from variable region peptides. Some strong signals have not yet been explained by expected peptides from serum (or trypsin or keratin).

Fig. 2 Inset of parent spectrum between mz 1864 and 1924.




This region contains most theoretical N-terminal R-containing peptides from Ig H , IgK and IgL. The above histogram counts the number of distinct variable R-containing peptides that match within 20 ppm of the measured mass from a human germ line protein sequence library. (Many match < 5 ppm). Favorable constant region peptides were granted a histogram score of 5 to emphasize that they are expected to be present. Some measured masses apparently correspond to unique germ line encoded peptides. Other masses may well correspond to multiple variable region peptides with slightly different masses. In all case, it is expected that there are many other precursor peptides from additional N-temrinal variable region peptides with mutations, and also from variable region and constant peptides from other regions. The weak signal at 1902 is puzzling; additional work is needed to determine whether an expected constant region peptide is low in abundance, or perhaps incompleltely digested.

Table 1. PMF proposed IDs from whole serum digest.

| N | Symb | Length | Name | Matches | Score | %IM | ppw |
|---|---|---|---|---|---|---|---|
| 1 | ALB | 609 | Serum albumin | 18 | 408634 | 41.2 | 5.8 |
| 2 | IGHA1 | 353 | Ig alpha-1 chain C | 4 | 208877 | 4.7 | 2.3 |
| 3 | APOA1 | 267 | Apolipoprotein A-I | 11 | 35225 | 6.0 | 8.0 |
| 4 | A2M | 1474 | Alpha-2-macroglobulin | 11 | 30480 | 3.6 | 5.9 |
| 5 | IGHM | 452 | Ig mu chain C | 5 | 28292 | 1.4 | 3.8 |
| 6 | wrong | 140 | Putative | 4 | 16000 | 0.9 | 4.6 |
| 7 | C3 | 1663 | Complement C3 | 15 | 14097 | 3.2 | 6.7 |
| 8 | wrong | 70 | NADH dehydrogenase | 3 | 10163 | 3.6 | 4.5 |
| 9 | wrong | 204 | Rho GDP-dissociation | 2 | 6773 | 2.3 | 5.2 |
| 10 | TRFE_hu | 698 | serotransferrin | 7 | 4705 | 2.9 | 8.9 |

Proteins in green are abundant serum proteins. The others are wrong, and this demonstrates the limit of mass fingerprinting from complex mixtures like whole serum. Slightly different sample preparations typically yield similar numbers of well known serum proteins, but the exact proteins in these lists is variable. Of course, if albumin is not the top hit, something is very wrong! Immunoglobulins have a disadvantage in PMF because prominent peptides are listed separately in the database.

Inset of parent region

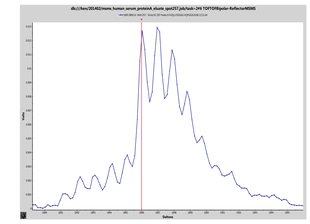


Fig. 3 MSMS spectrum of parent mass 1895.989

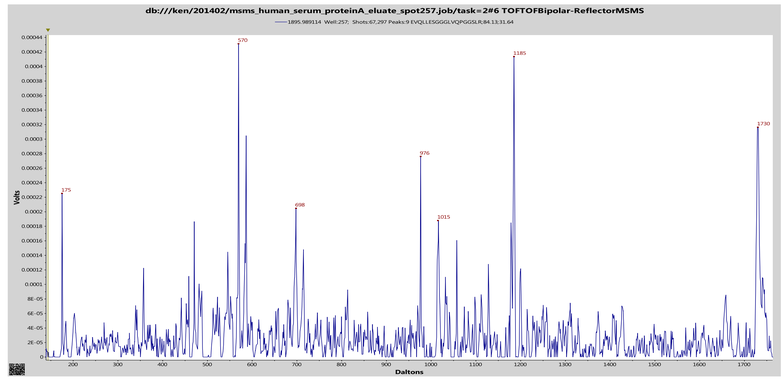



Table 2. Proposed peptides from MSMS analysis of above MSMS spectrum

Search human Ig database from IMGT with 25 amuy window -> 1865 peptides; top 40 are IgV N-termini

| I | p | MassTheo | MassExp | ppm | Sequence | M | TrS | Symb | Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1897.00 | 1895.99 | -535.0 | EVQLLESGGGLVEPGGSLR | 28 | 1200 | IGHV | IGHV3-23*01 |
| 2 | 0 | 1896.06 | 1895.99 | -35.2 | EVQLLESGGGLVKPGGSLR | 28 | 1179 | IGHV | IGHV3-21*01 |
| 3 | 1 | 1896.06 | 1895.99 | -35.2 | EVKLLESGGGLVQPGGSLR | 28 | 1157 | IGHV | IGHV3-66*01 |
| 4 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLLESGGGLVQPGGSLR | 28 | 1157 | IGHV | IGHV3-23*01 |
| 5 | 0 | 1895.04 | 1895.99 | 503.0 | EVQLLQSGGGLVQPGGSLR | 28 | 1093 | IGHV | IGHV3-23*01 |
| 6 | 0 | 1896.06 | 1895.99 | -35.2 | EVQLVETGGGLVKPGGSLR | 30 | 1015 | IGHV | IGHV3-30*02 |
| 7 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVETGGGLVQPGGSLR | 30 | 1015 | IGHV | IGHV3-53*02 |
| 8 | 0 | 1896.98 | 1895.99 | -521.7 | EVQQVESGGGLVQPGGSLR | 28 | 1007 | IGHV | IGHV3-23*04 |
| 9 | 0 | 1887.92 | 1895.99 | 4253.9 | EVQLSESGGGMVQPGGSLR | 29 | 981 | IGHV | IGHV3-23*01 |
| 10 | 0 | 1897.96 | 1895.99 | -1040.7 | EVQLVESGGGLDQPGGSLR | 26 | 968 | IGHV | IGHV3-7*01 |
| 11 | 0 | 1885.94 | 1895.99 | 5297.8 | EVQMVDSGGGLVQPGGSLR | 24 | 928 | IGHV | IGHV3-7*01 |
| 12 | 0 | 1896.06 | 1895.99 | -35.1 | EVQLVESGGALVKPGGSLR | 29 | 897 | IGHV | IGHV3-15*02 |
| 13 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVESGGALVQPGGSLR | 29 | 880 | IGHV | IGHV3-7*01 |
| 14 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVESGGAVIQPGGSLR | 29 | 871 | IGHV | IGHV3-43*01 |
| 15 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVESAGGLVQPGGSLR | 28 | 850 | IGHV | IGHV3-15*08 |
| 16 | 0 | 1896.06 | 1895.99 | -35.2 | EVQLVESGGGLLKPGGSLR | 28 | 787 | IGHV | IGHV3-15*01 |
| 17 | 0 | 1896.06 | 1895.99 | -35.2 | EVQLVESGGGLIKPGGSLR | 28 | 787 | IGHV | IGHV3-21*01 |
| 18 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVESGGGLLQPGGSLR | 28 | 756 | IGHV | IGHV3-74*01 |
| 19 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVESGGGLIQPGGSLR | 28 | 756 | IGHV | IGHV3-53*01 |
| 20 | 0 | 1887.95 | 1895.99 | 4240.7 | DIVMTQSPGTLSLSPGER | 24 | 750 | IGKV | IGKV3-20*01 |
| 21 | 0 | 1896.02 | 1895.99 | -16.0 | EVQLVESGGVLAQPGGSLR | 27 | 729 | IGHV | IGHV3-74*01 |
| 22 | 0 | 1896.00 | 1895.99 | -6.6 | qVQLMESGGGLVKPGGSLR | 23 | 727 | IGHV | IGHV3-11*01 |
| 23 | 0 | 1896.00 | 1895.99 | -6.6 | eVQLMESGGGLVKPGGSLR | 23 | 727 | IGHV | IGHV3-15*07 |
| 24 | 0 | 1905.02 | 1895.99 | -4763.0 | EVHLLESGGGLVQPGGSLR | 20 | 703 | IGHV | IGHV3-23*01 |
| 25 | 1 | 1896.06 | 1895.99 | -35.2 | VKLLEESGGGLVQPGGSLR | 23 | 695 | IGHV | IGHV3-23*01 |
| 26 | 0 | 1895.07 | 1895.99 | 483.8 | QVQLLESGGGLVKPGGSLR | 23 | 694 | IGHV | IGHV3-11*03 |
| 27 | 1 | 1898.00 | 1895.99 | -1059.9 | EVKVEESGGGLVQPGGSLR | 25 | 688 | IGHV | IGHV3-66*01 |
| 28 | 0 | 1896.03 | 1895.99 | -21.9 | EVQLVESGGGVVRPGGSLR | 26 | 687 | IGHV | IGHV3-20*01 |
| 29 | 0 | 1895.99 | 1895.99 | -2.7 | QVQLVQSGGGVEQPGGSLR | 25 | 687 | IGHV | IGHV3-48*03 |
| 30 | 1 | 1897.05 | 1895.99 | -560.1 | QITLKESGGGLVQPGGSLR | 23 | 679 | IGHV | IGHV3-48*03 |
| 31 | 1 | 1896.09 | 1895.99 | -54.4 | VKLLEESGGGLVKPGGSLR | 23 | 679 | IGHV | IGHV3-23*01 |
| 32 | 0 | 1895.97 | 1895.99 | 12.6 | eVQLMESGGGLVQPGGSLR | 23 | 679 | IGHV | IGHV3-15*07 |
| 33 | 0 | 1895.04 | 1895.99 | 503.0 | QVQLLESGGGLVQPGGSLR | 23 | 649 | IGHV | IGHV3-23*01 |
| 34 | 1 | 1895.07 | 1895.99 | 483.8 | VKLLEQSGGGLVQPGGSLR | 23 | 646 | IGHV | IGHV3-23*01 |
| 35 | 1 | 1895.07 | 1895.99 | 483.8 | QVKLLESGGGLVQPGGSLR | 23 | 633 | IGHV | IGHV3-23*01 |
| 36 | 1 | 1895.11 | 1895.99 | 464.6 | QVKLLESGGGLVKPGGSLR | 23 | 618 | IGHV | IGHV3-21*01 |
| 37 | 0 | 1896.00 | 1895.99 | -6.6 | qMQLVESGGGLVKPGGSLR | 21 | 606 | IGHV | IGHV3-11*03 |
| 38 | 0 | 1896.00 | 1895.99 | -6.6 | eMQLVESGGGLVKPGGSLR | 21 | 606 | IGHV | IGHV3-15*01 |
| 39 | 0 | 1895.07 | 1895.99 | 483.8 | VQLLEQSGGGLVKPGGSLR | 22 | 589 | IGHV | IGHV3-15*01 |
| 40 | 0 | 1896.02 | 1895.99 | -16.0 | VQLLEESGGGLVQPGGSLR | 22 | 589 | IGHV | IGHV3-7*01 |

Search human database with 25 amu window -> 19999 peptides; best hit score 656

| I | p | MassTheo | MassExp | ppmP | Sequence | M | TrS | Symb | Name |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1884.96 | 1895.99 | 5815.5 | GVCISPEAIVTDLPEDVK | 25 | 656 | wtf13 | Uncharacterized |
| 2 | 0 | 1889.96 | 1895.99 | 3180.5 | PRPPPPPPPPPEESCPAK | 54 | 584 | ABCI1 | ABC transporter I |
| 3 | 1 | 1887.96 | 1895.99 | 4232.7 | SDCPLGVTKEIAIWAER | 25 | 557 | ispF | 2-C-methyl-D- |
| 4 | 2 | 1903.04 | 1895.99 | -3716.8 | TPSTASLTRTPSRASLTR | 29 | 555 | rpsK | 30S ribosomal |
| 5 | 1 | 1902.89 | 1895.99 | -3638.1 | QDIMPEVDKQSGSPESR | 24 | 543 | purA | Adenylosuccinate |

32 distinct variable region peptides scored higher than any peptides from the human proteome, based on 40,000 human proteins in SwissProt, which does not contain all common Ig variable sequences from the germ line. Many of the strongest MSMS peaks can be explained by ions shared by many of these peptides. Note that each peptide above may be common to many germ line genes; other are specific to individual V region families, barring back mutation.

## Conclusions:

- Can see peptides from Ig in tryptic digest of whole serum.
- Can confirm some of them by MSMS without separation (not shown).
- Can readily see variable peptides in digests of Protein A purified protein.
- Can deduce something about IgG class usage from unseparated digests.
- Can deduce something about variable region family usage.
- Should be useful for characterizing monoclonal gammopathies (not shown here).
- In the future, study reproducibility of patterns from individuals.