# Characterization of the MHC I immunopeptidome of HaCaT cells

*Alistair Bailey*

*May 09 2018*

# Contents

# List of Tables

# List of Figures

# Summary

This report details objective 1 of the ImmunoSense project: Characterization of the MHC I immunopeptidome of HaCaT cells ± DNCB.

The key findings are:

- The identification of 7611 unique peptides under control conditions.
- The identification of 6687 unique peptides under DNCB treatment conditions.
- 9mer peptides represent around 55% of the total peptidome across all six replicates.
- 60% of unique peptides are shared between treatment and control peptidomes, with 25% and 14% of peptides seen only in control and treatment peptidomes respectively.
- Two DNCB modified peptides were identified in two biological replicates (four technical replicates).
- One of the modified peptides derived from a Keratin, type I cytoskeletal 13 protein (P13646)
- The second modified peptide derived from the active site of Glutathione S-transferase omega-1 (P78417)

# Chapter 1

# Introduction

This report details objective 1 of the ImmunoSense project: Characterization of the MHC I immunopeptidome of HaCaT cells (Boukamp et al., 1988) ± DNCB. The HaCaT cell line is a spontaneously transformed keratinocyte line.

Here we seek to capture class I major histocompatibility complex molecules (MHC I) presented on the surface of cells. MHC I sample the intracellular proteome in the form of short peptides, predominantly 9 amino acids in length. The collective name for the peptides presented is the immunopeptidome. For a recent summary of MHC I antigen processing and presentation see (van Hateren et al., 2017).

We used HaCaT cells to test two non-exclusive hypotheses:

- Sensitiser-modified peptides presented by MHC I directly stimulate the immune system.
- Sensitiser modification of proteins alters the MHC I presented repertoire of peptides.

2,4-Dinitrochlorobenzene (DNCB) is our model sensitiser. In cell lysates DNCB modifies approximately 0.12% Lys, 0.03% Cys and < 0.03% Tyr, His and Arg (Parkinson et al., 2018).

For an overview of immunopeptidomics goals and challenges see this short article: (Caron et al., 2017).

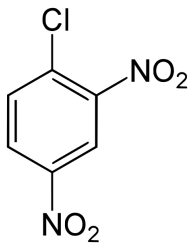


Figure 1.1: The structure of 2,4-Dinitrochlorobenzene (DNCB).

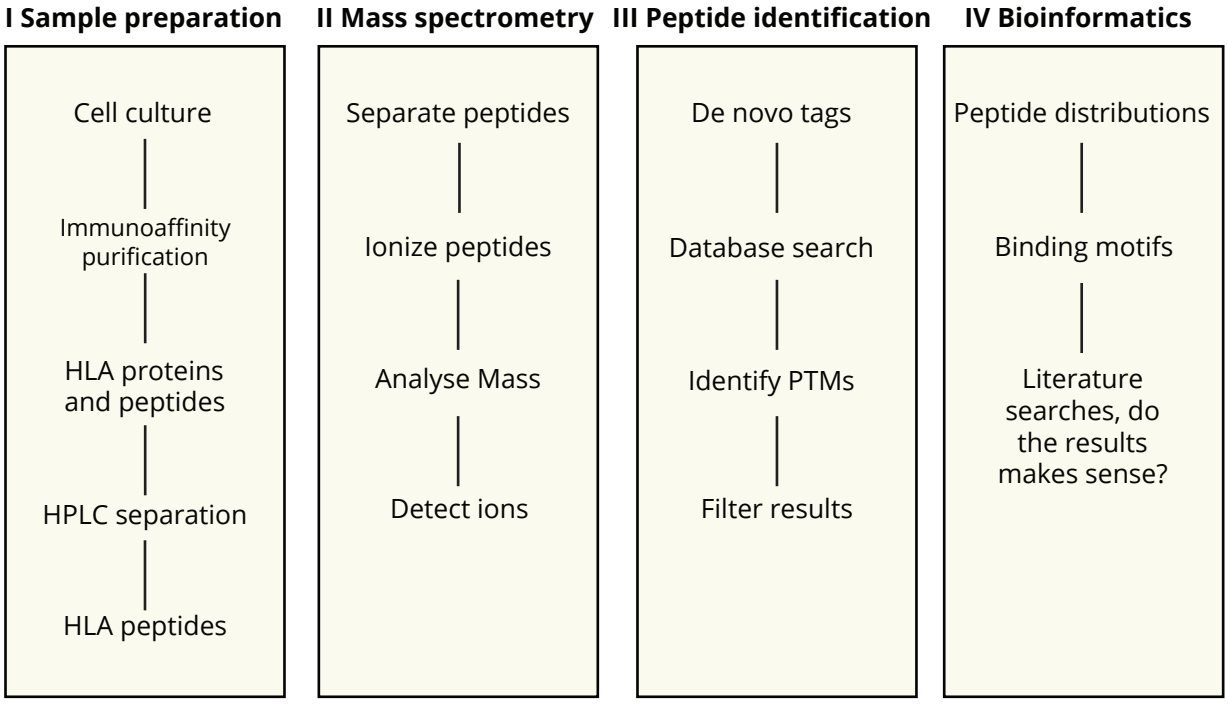


Figure 1.2: The immunopeptidome workflow.

Figure 1.2 shows the immunopeptidome workflow used in the generation of this work.

The following sections detail the methodology, summary tables, histograms and venn diagrams, biochemical properties of the peptidomes and properties of DNCB modified peptides.

# Chapter 2

# Methods

## 2.1 Cell culture

HaCaT cells (Boukamp et al., 1988) purchased from Cell lines service were cultured in DMEM medium supplemented with 4.5g/L glucose and 2mM L-glutamine and 10% fetal bovine serum.

For each replicate two populations were grown to 500 x $10^6$ cells each, and then the medium changed to DMEM/F-12 without serum. One population was used for the control and the other population treated with a 10 $\mu$M 2,4-Dinitrochlorobenzene (DNCB) mixture of DNCB and DNCB-$D_3$ for 24 hours. Cells were harvested with 0.25% Trypsin-EDTA and then washed with DPBS and pelleted prior to freezing and storage at -80°C.

Three biological replicates of control and treatment pellets were cultured, these were then split in half to create six technical replicates:

- 1A and 1B
- 2A and 2B
- 3A and 3B

## 2.2 HLA typing of HaCaT cells

DNA extraction performed with QIAGEN DNAeasy kit under the supervision of Dr. Emma Reeves, and typing done by Southampton General Hospital tissue typing unit.

Confirmation of these results was made using genomics analysis using publicly available datasets and HISAT2 (Kim et al., 2015).

Together this yielded HLA types:

- Homozygous HLA-A*31:01
- HLA-B*40:01, HLA-B*51:01
- HLA-C*03, HLA-C*15

There is uncertainty of the C alleles beyond two digit identification.

The literature had indicated that HaCaTs carried HLA-A*02 (Lemaitre et al., 2004), but our results disconfirm this observation.

## 2.3 Immunoaffinity purification and HPLC fractionation

Cells were lysed in a buffer of 20 mM Tris-HCL (pH 8.0), 150 mM NaCl, 0.5% IGEPAL, 0.25% Sodium deoxycholate, 1 mM EDTA, 0.2 mM Iodoacetamide and protease inhibitors at 4°C on a roller for 30 minutes.

Cell nuclei were removed by centrifugation at 2,000g for 10 minutes. The remaining lysate was then clarified by centrifugation at 15,000g (13,000 rpm) for 1 hour.

The lysate was then added to a tube containing pan-MHC I binding antibody W6/32 crosslinked to Protein A sepahrose beads at a concentration of 2 mg antibody per mL of beads and incubated on a roller at 4°C for 2 hours.

The lysate and beads were then transferred to a column for washing and elution of the MHC I molecules using 5 mL 10% acetic acid, denaturing the MHC I molecules into a heavy chain, $\beta_2$m and peptide mixture.

The eluate was then dried in a speed-vac evaporator and re-suspended in 500 $\mu$L of 0.1% TFA/1% acetonitrile prior to injection into the high performance liquid chromatography (HPLC) system.

A Thermo UltiMate 3000 system using a Chromolith HighResolution RP-18 endcapped 100-4.6 HPLC column were used to separate and collect the peptides for mass spectrometry analysis. 0.5 mL peptide fractions were collected over 8 minutes. These fractions were pooled as odd and even fractions and then dried in a speed-vac evaporator.

These fractions were then re-suspended in 20 $\mu$L of 1% Formic acid and split into four samples, two odd and two even, for mass spectrometry analysis.

## 2.4 Mass spectrometry

For each sample a total of 10 $\mu$L was injected onto a LC system connected to a Thermo Fusion Tribrid mass spectrometer in Orbitrap-Orbitrap mode for maximum sensitivity.

Using 0.1 % formic acid (buffer A) and 0.1% formic acid in acetonitrile (buffer B) each sample was run on gradient of to 30% buffer B over 110 minutes, followed by a five minute gradient to 80% buffer B.

## 2.5 Data analysis

The raw spectrum files were analysed using Peaks Studio (Zhang et al., 2012).

Variable modifications were set for N-terminal acetylation (42.01 Da) and methionine oxidation (15.99 Da) and for DNCB (166.00 Da). As the lysis buffer contains iodoacetamide cysteine carbamidomethylation (57.02 Da) was also set as a variable modification.

De-novo peptide identification confidence settings were set at 80%, and a false discovery rate (FDR) of 1% was used for the database searches.

The human proteome UP000005640, UniProt release 2017_05 of 42,186 proteins, was used for database searches.

Downstream analysis of the Peaks Studio identifications was performed in R (R Core Team, 2018).

# Chapter 3

# Peptidome size and proportions

Three biological replicates of control and treatment cells were cultured, these were then split in half to create six technical replicates for purification and analysis:

- 1A and 1B
- 2A and 2B
- 3A and 3B

The raw spectrum files were analysed using Peaks Studio (Zhang et al., 2012) and then processed further in R (R Core Team, 2018).

Contaminant peptides were removed and I consider peptides only of length 8-15 amino acids in length as real MHC I peptides. Longer peptides may be real peptides, but this conservative approach increases the confidence of my observations.

## 3.1 Summary tables

Table 3.1 shows the number of peptides from each replicate of peptide length 8 to 15 amino acids.

Table 3.1: Peptidome 8-15mer totals

| Replicate | Number of peptides |
|---|---|
| Control 1A | 1677 |
| Control 1B | 2400 |
| Control 2A | 2051 |
| Control 2B | 1018 |
| Control 3A | 5130 |
| Control 3B | 5738 |
| DNCB 1A | 1427 |
| DNCB 1B | 2042 |
| DNCB 2A | 1482 |
| DNCB 2B | 1817 |
| DNCB 3A | 4727 |
| DNCB 3B | 4831 |

Table 3.2 shows the proportions of 9-mer peptides in each replicate. As 9-mer is the predominant peptide length for MHC I molecules, we use the proportion of 9-mers as a benchmark for peptidome purity and replication.

Table 3.3 shows the number of unique peptides found across all replicates.

## 3.2 Histograms

Figure 3.1 shows the distributions of peptide lengths in each peptidome in absolute numbers.

Figure 3.2 shows these distributions in terms of proportions of peptide lengths in each peptidome.

The variation in peptidome size between replicates we believe is primarily due to improvements in the purification and analysis methodology during the generation of this data. Hence the final replicate is yields the largest number of peptides.

Table 3.2: Peptidome 9-mer proportions

| Replicate | 9-mer percent |
|---|---|
| Control 1A | 53 |
| Control 1B | 52 |
| Control 2A | 54 |
| Control 2B | 52 |
| Control 3A | 55 |
| Control 3B | 51 |
| DNCB 1A | 56 |
| DNCB 1B | 57 |
| DNCB 2A | 47 |
| DNCB 2B | 53 |
| DNCB 3A | 55 |
| DNCB 3B | 55 |

Table 3.3: Number of unique peptides

| Treatment | Number of peptides |
|---|---|
| Control | 7611 |
| DNCB | 6687 |

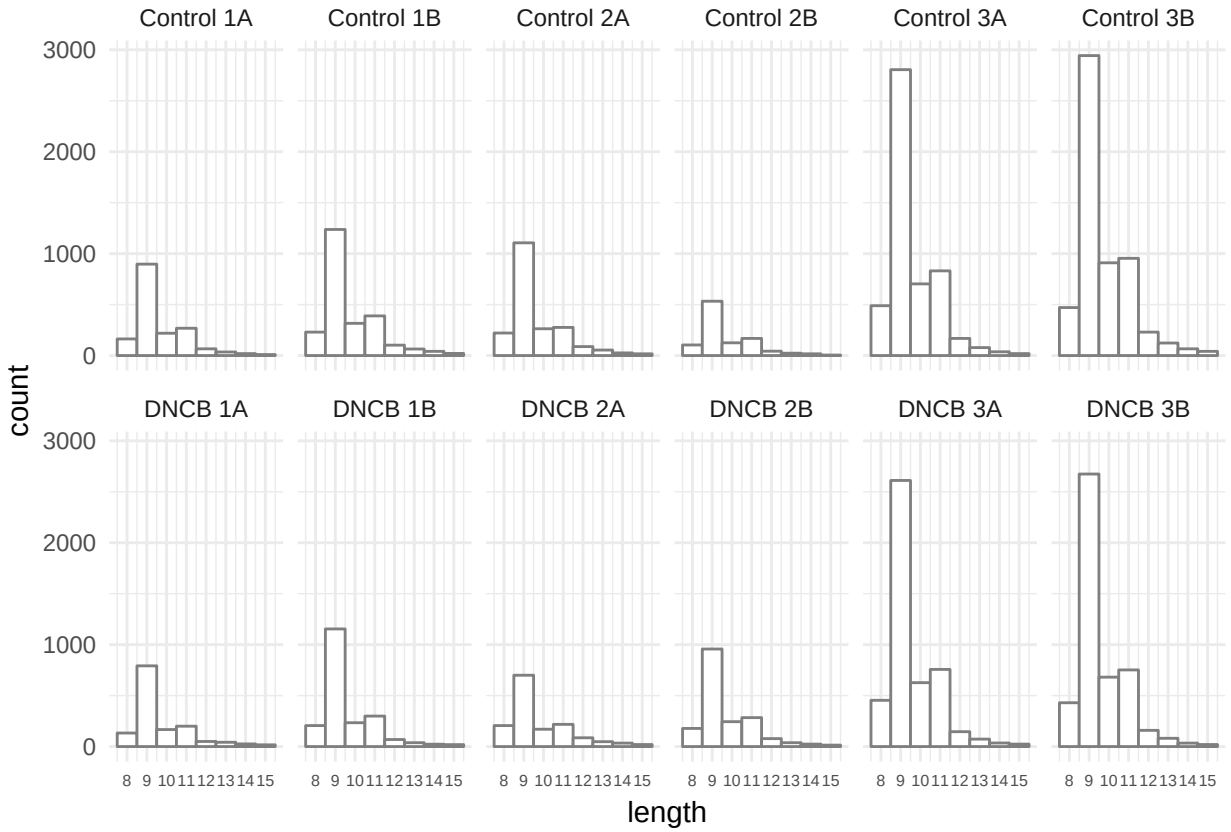


Figure 3.1: Histogram of peptide numbers for peptides of 9-15 amino acids. This plot shows a large degree of variation in total numbers of peptides between replicates.

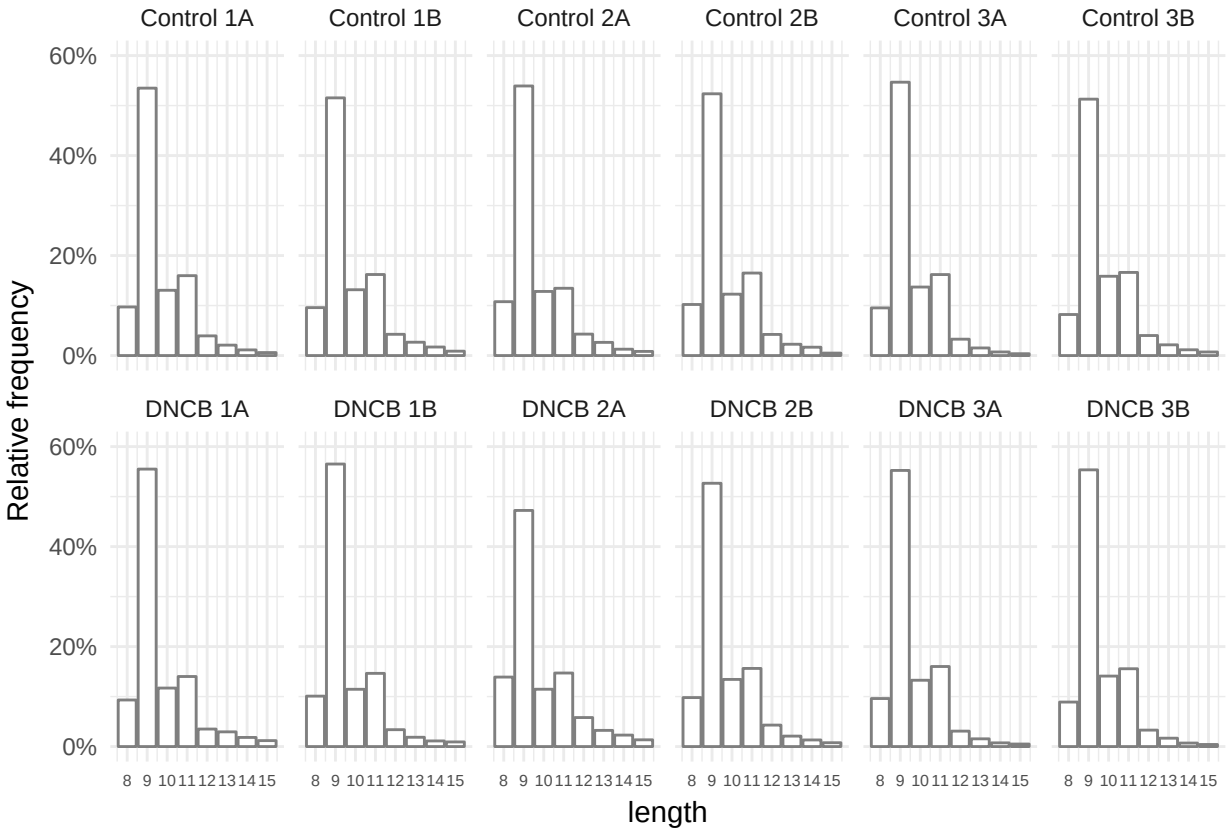


Figure 3.2: Histogram of proportion of peptide lengths, 9-15 amino acids. This plot shows good replication in terms of relative peptide length frequency between replicates.

## 3.3 Venn diagrams

To compare the peptidomes of control and treated cells, I first removed all modifications from the peptides such that I compare peptides in their unmodified form. I then selected the unique peptides across all replicates to create combined peptidomes for control and treatment cases.

Figure 3.3 indicates that the majority of peptides are the same under treatment and control. Further analysis is required to establish whether there is any significance to the populations not shared between treatment and control, but this data suggests that the core peptidome repertoire is shared between control and DNCB treatment.

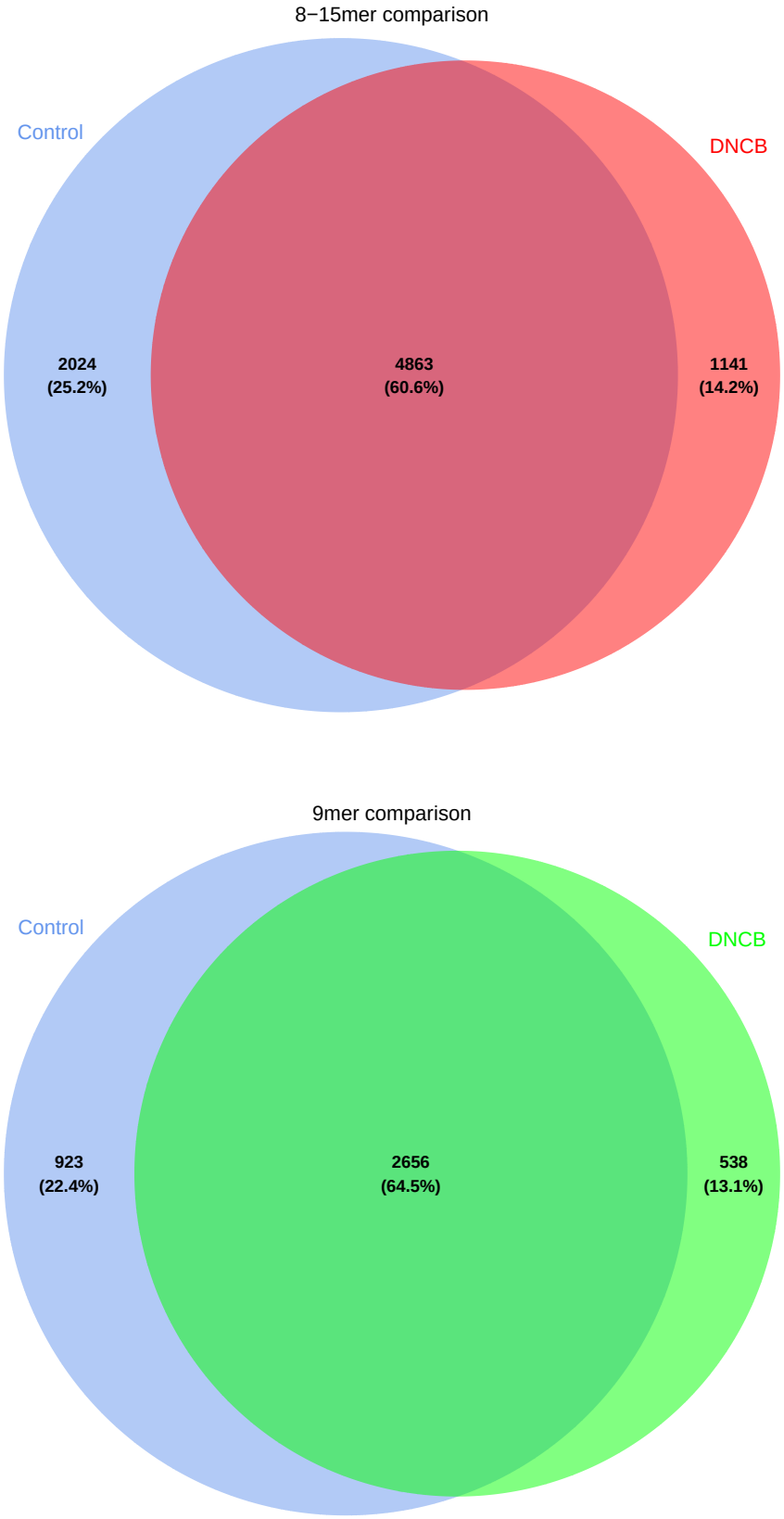


Figure 3.3: Venn diagrams of unique Control and DNCB peptides

# Chapter 4

# Peptide binding motifs

Having quantified the peptidome, to further confirm that these peptides are MHC I peptides I did two analysis:

1. I used the GibbsCluster tool (Andreatta et al., 2017) to group the 9mer peptides according to their amino acid sequence to identify motifs and compared these groupings to motifs identified for peptides matching the HaCat MHC allotypes in the curated immune epitope database (IEDB) (Vita et al., 2015).
2. Using netMHC (Nielsen et al., 2003; Andreatta and Nielsen, 2016) predictions for binding affinities of clustered 9mer peptides to their identified MHC I allotypes were made to cross validate the clustering.

## 4.1 Clustering of peptides

The unique 9mer peptides for treatment and control with post-translational and/or DNCB modifications were clustered using GibbsCluster (Andreatta et al., 2017).

Figure 4.1 indicates that three clusters of peptides. This is indicated by the size of the bar scoring the highest Kullbach-Leibler distance, a measure of information entropy. By

comparison to motifs of peptides from the IEDB in Figure 4.2 we can see that these three clusters correspond with motifs for HLA-A*31, HLA-B*40 and HLA-B*51.

The logos corresponding with each cluster measure strength of the preference for each amino acid at each peptide position, as indicated by the size the letters.

Figure 4.2 also shows that HLA-C*03 and HLA-C*15 have less clear motifs than the A and B allotypes, sharing biochemically similar preferences at potion 9 with HLA-B*40 and HLA-B*51, and also at position 2 with HLA-B*51. This means that the HLA-C peptides are mixed into these clusters.

## 4.2 Peptide binding predictions

Figure 4.3 shows the results of using NetMHC to predict whether the peptides from each cluster bound to their identified motif.

As suggested by the peptide logos, for HLA-A*31:01 which has a clear motif 92% of peptides from this cluster were predicted as binders. For the B allotypes this number was reduced, presumably due to the difficulty in separating these peptides from each other and from the C allotypes.

However, what this does confirm is that the peptides we observe are indeed MHC I peptides and not non-specific 9mers.

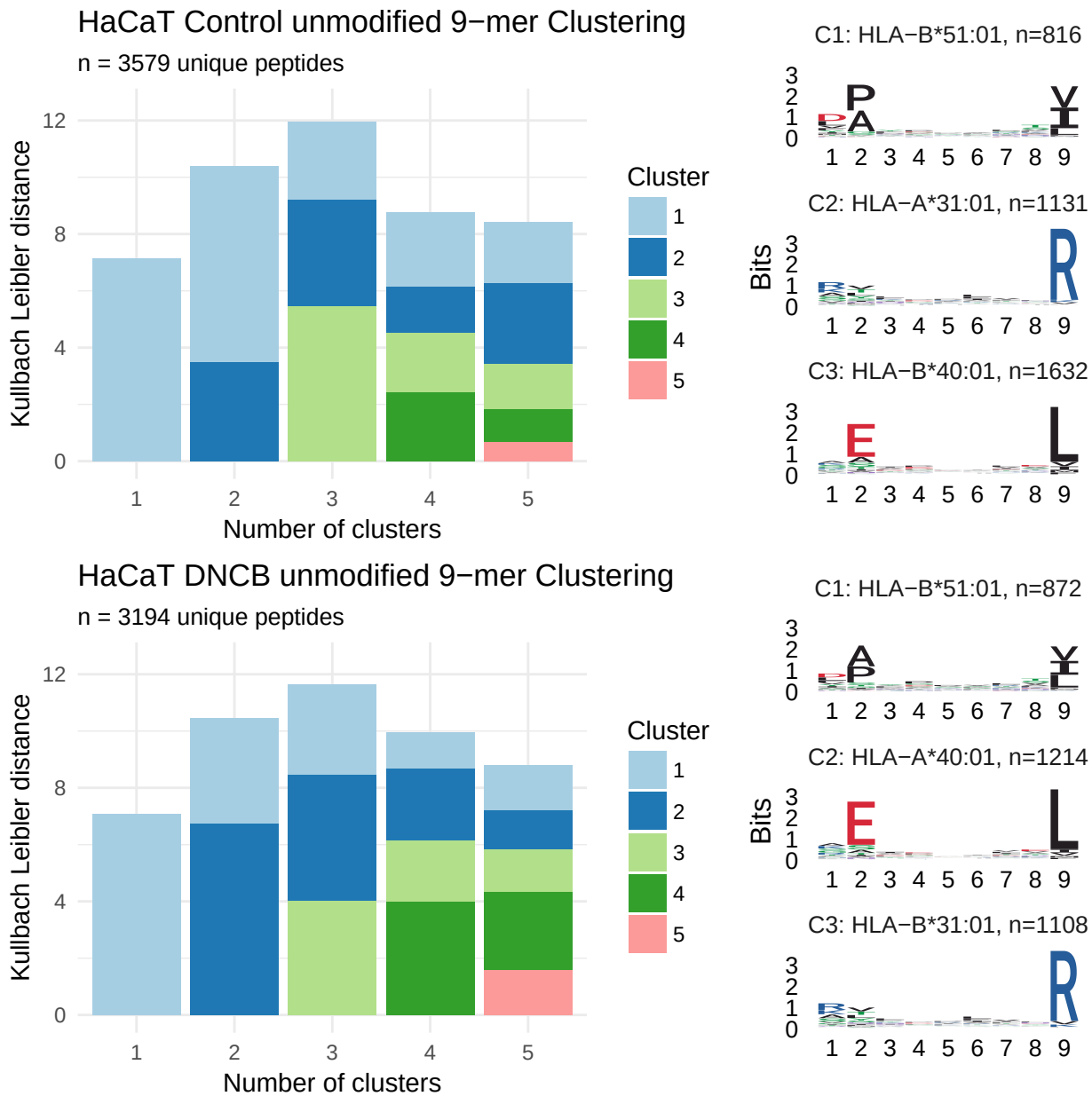


Figure 4.1: Gibbs clustering of 9mer peptides.

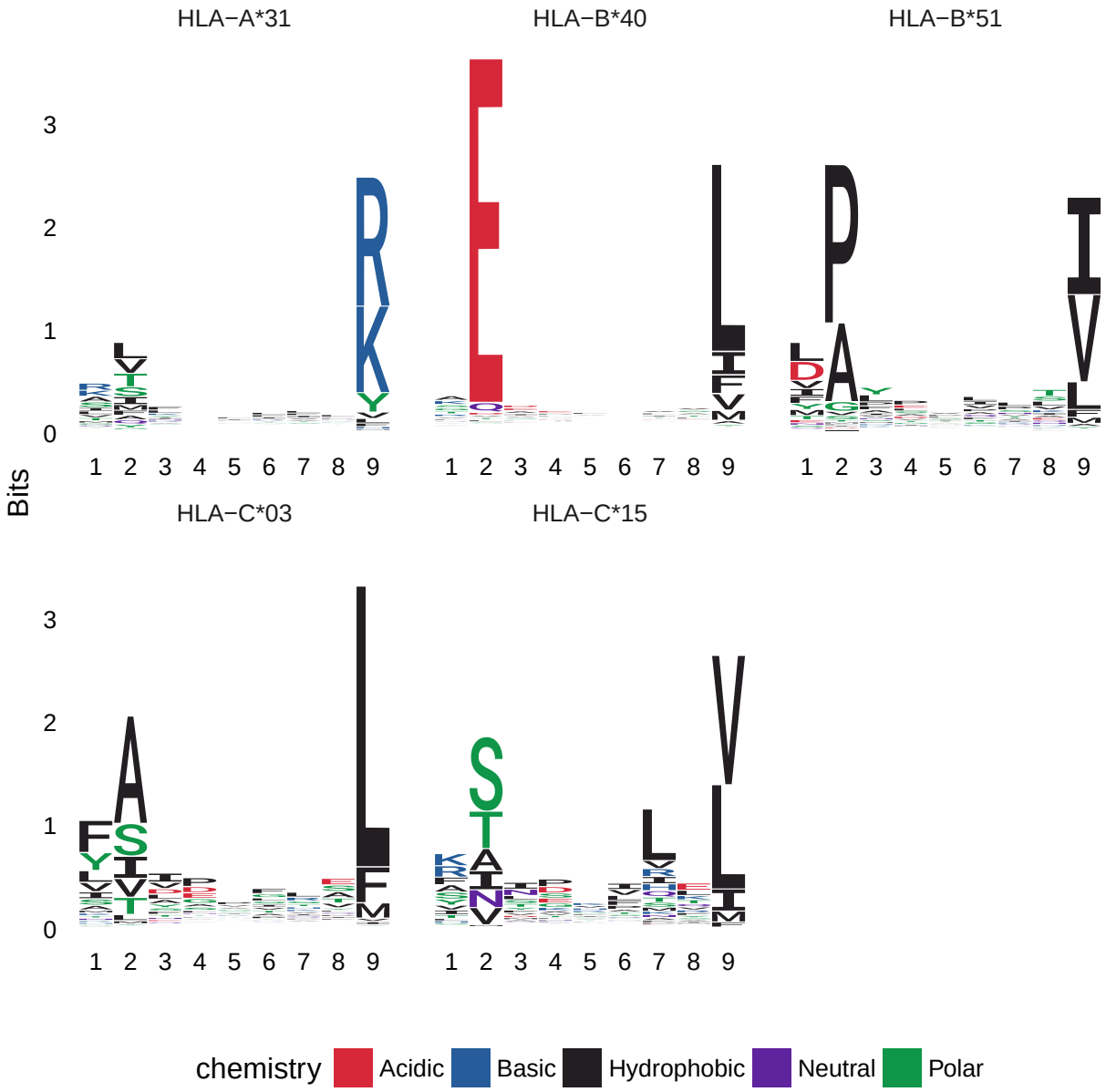


Figure 4.2: Immune epitope database peptide motifs.

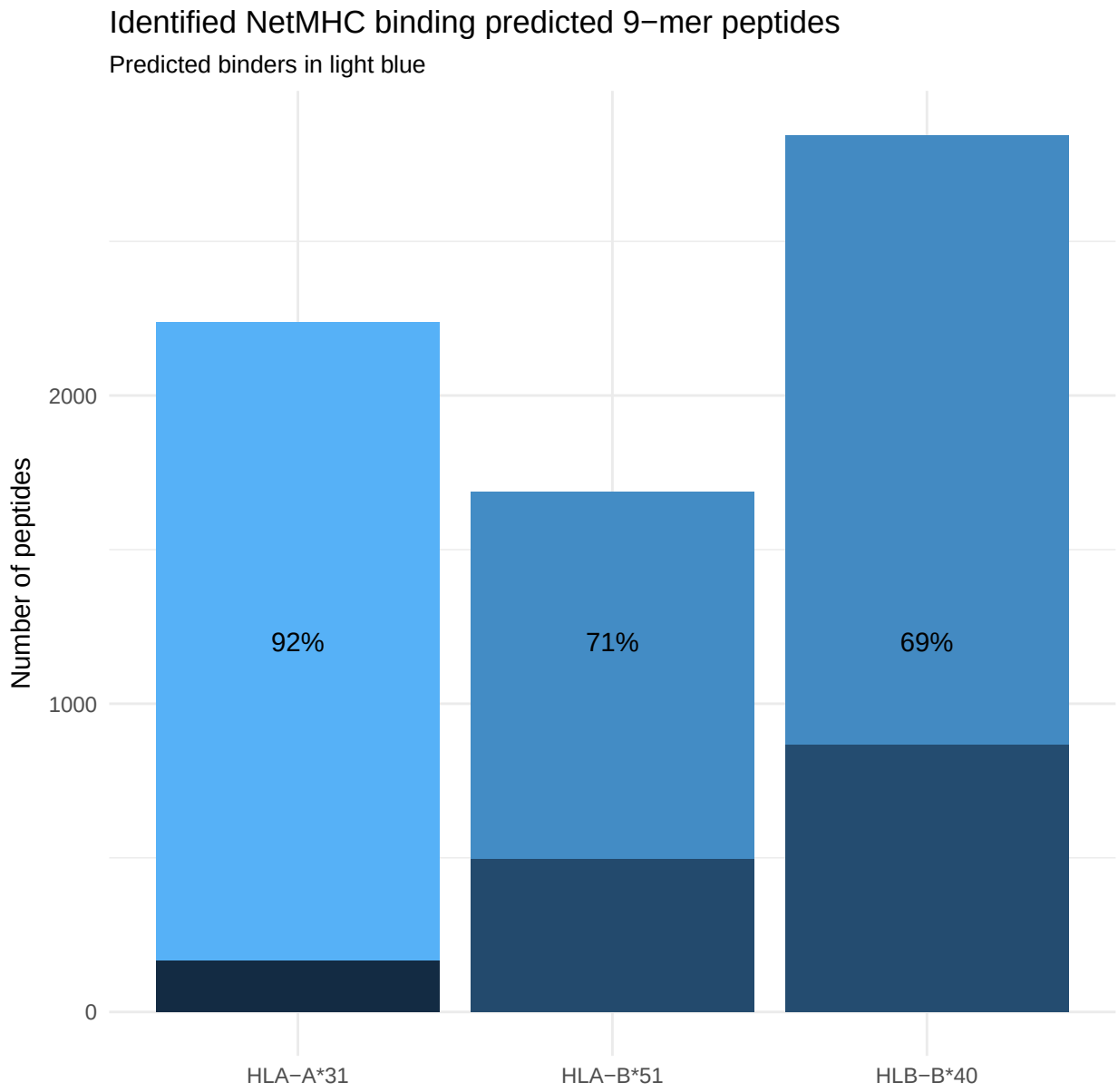


Figure 4.3: NetMHC predicted binders

# Chapter 5

# DNCB modifications

As noted in Section 2.1 one HaCaT population was treated with a 10 $\mu$M 2,4-Dinitrochlorobenzene (DNCB) mixture of DNCB and DNCB-DNCB-$D_3$ for 24 hours.

Searches were performed for DNCB (166.00 Da) and then for any peptide identified the corresponding DNCB-DNCB-$D_3$ peak in the mass spectrum was used to exclude false positive identifications.

## 5.1 Identified modified peptides

Table 5.1 shows that two peptides, one 9mer and one 10mer were observed in two biological replicates and multiple technical replicates.

## 5.2 MS spectra confirmed modifications

Figure 5.1 show the MS spectra for the Keratin, type I cytoskeletal 13 9mer peptide AETECRYAL with a cysteine modification. The unmodified peptide mass to charge ratio m/z = 611.25. The peptide was doubly charged meaning we expect the DNCB-DNCB-$D_3$ modified peptide peak at 3/2 = 1.5 m/z unit higher, 611.75.

Table 5.1: DNCB modified peptides

| Peptide | Length | Replicate | UNIPROT ID |
|---|---|---|---|
| AETEC(+166.00)RYAL | 9 | DNCB 1A | P13646 |
| RFC(+166.00)PFAERTR | 10 | DNCB 1A | P78417 |
| RFC(+166.00)PFAERTR | 10 | DNCB 1B | P78417 |
| AETEC(+166.00)RYAL | 9 | DNCB 1B | P13646 |
| AETEC(+166.00)RYAL | 9 | DNCB 3A | P13646 |
| RFC(+166.00)PFAERTR | 10 | DNCB 3A | P78417 |
| AETEC(+166.00)RYAL | 9 | DNCB 3B | P13646 |

Ideally the DNCB-DNCB-$D_3$ modified peptide should have the same intensity as the DNCB modified peptide, but the difference is likely as result of unequal concentrations of the two forms of DNCB in the treatment.

This peptide was identified by GibbsCluster as binding to HLA-A*40:01 and confirmed using netMHC.

Figure 5.2 show the MS spectra for the Glutathione S-transferase omega-1 10mer peptide RFCPFAERTR, also with a cysteine modification. The unmodified peptide mass to charge ratio m/z = 483.55. The peptide was triply charged meaning we expect the DNCB-DNCB-$D_3$ modified peptide peak at 3/3 = 1 m/z unit higher, 484.55.

This peptide was identified by GibbsCluster as binding to HLA-A*31:01 and confirmed using netMHC.

Glutathione S-transferase omega-1 is important for detoxification of cells (Spriggs et al., 2016, Jacquoilleot et al. (2015)), and therefore this particular observation may also be indicative of the mechanism of DNCB sensitisation.

## 5.3 Putative MHC-peptide structures

To test whether these modifications were likely to be visible to T-cell receptors, I created putative structures of the DNCB modified peptides using USCF Chimera (Pettersen et al., 2004).

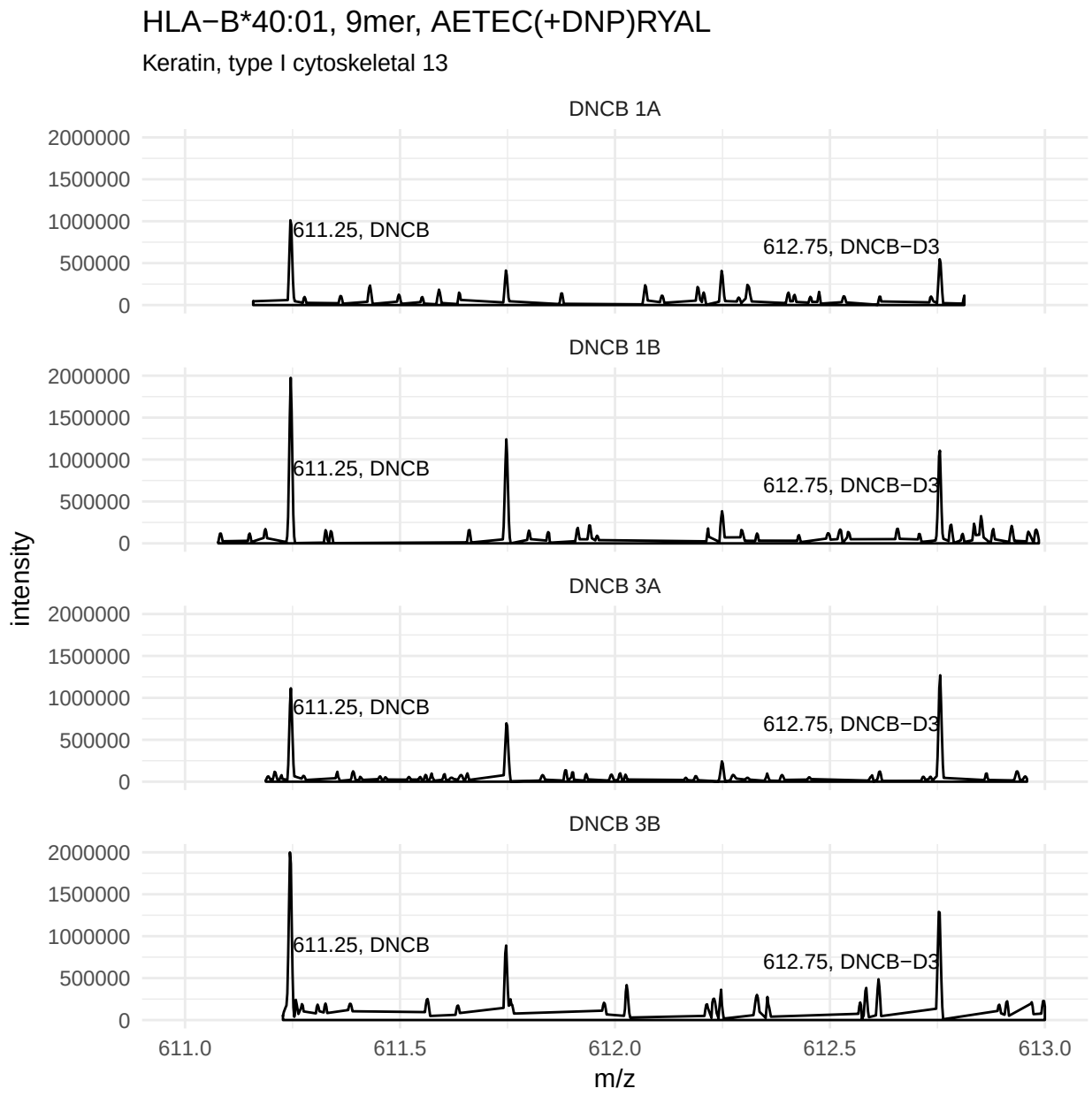


Figure 5.1: Keratin, type I cytoskeletal 13 survey scans

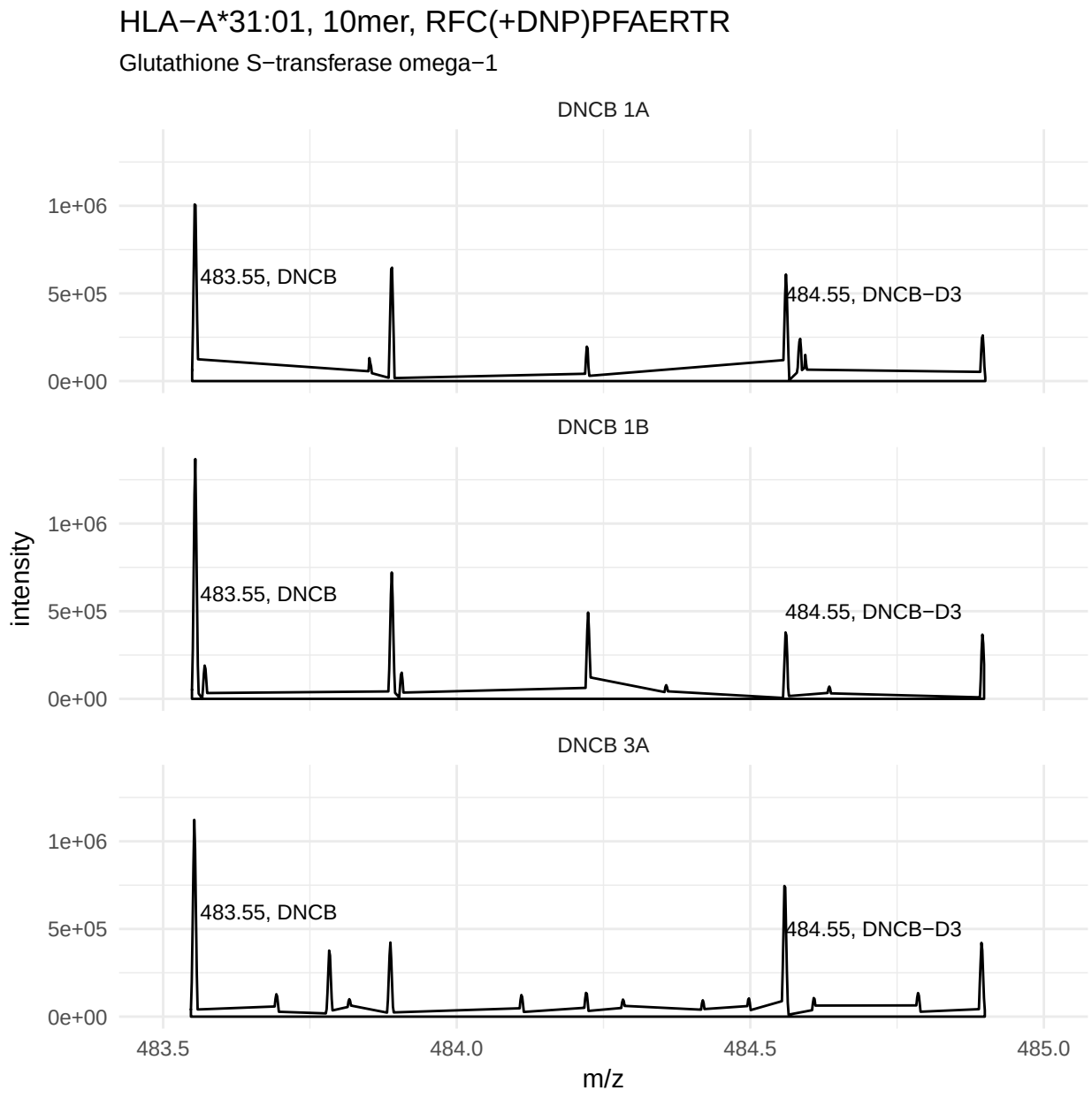


Figure 5.2: Glutathione S-transferase omega-1 survey scans

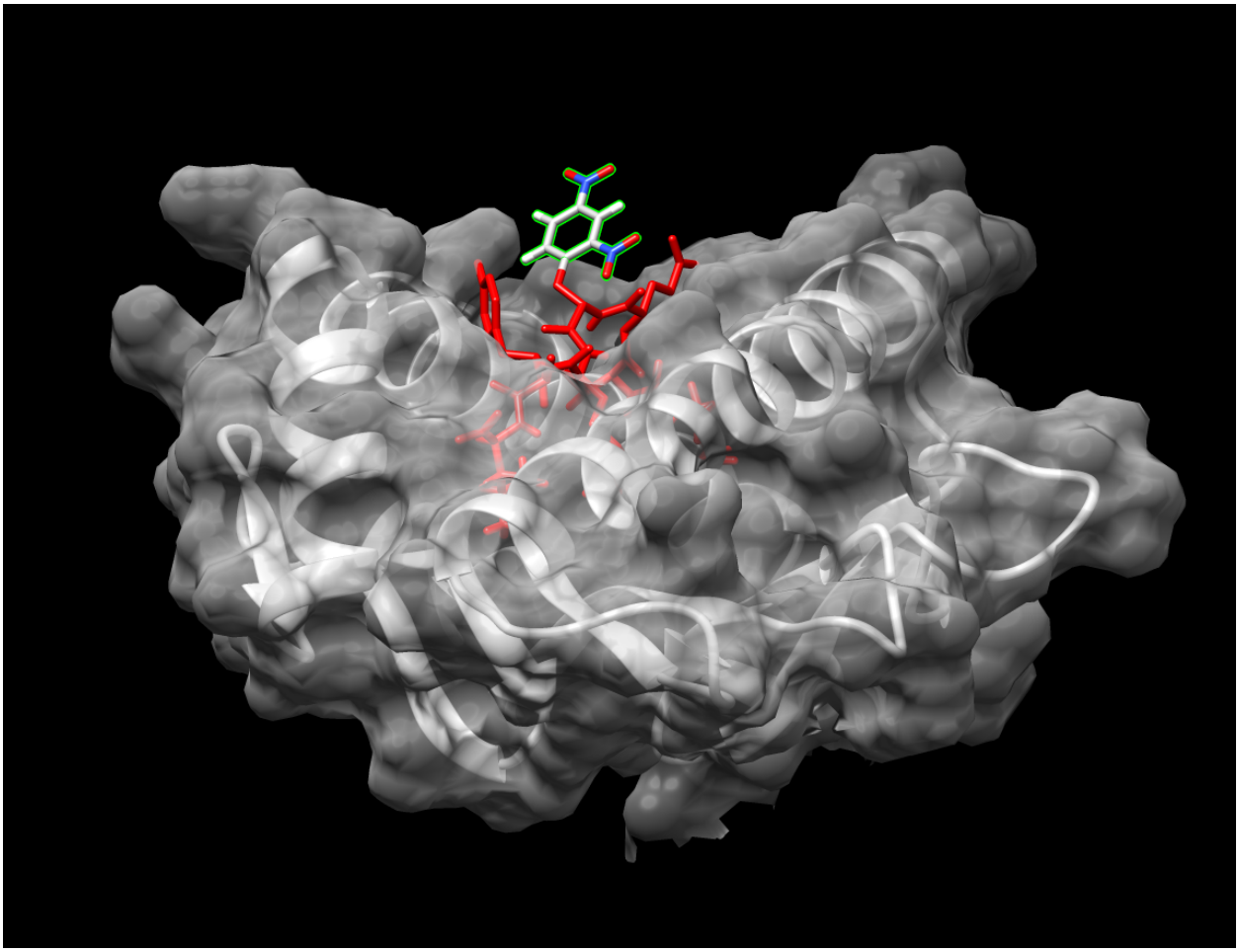

Figure 5.3: Putative structure of HLA-B*40:01 with 9mer AETEC(+DNP)RYAL

Figure 5.3 shows the DNCB modified cysteine AETEC(+DNP)RYAL from Keratin, type I cytoskeletal 13 is solvent exposed and oriented toward TCR interaction.

Figure 5.4 shows that the modified cysteine in the 10mer RFC(+DNP)PFAERTR derived from Glutathione S-transferase omega-1 and bound to HLA-A*31:01 is less clearly exposed, but could still interact with a TCR.

Interestingly, the RFC(+DNP)PFAERTR peptide derives from the active site of Glutathione S-transferase omega-1, highlighted in Figure 5.5

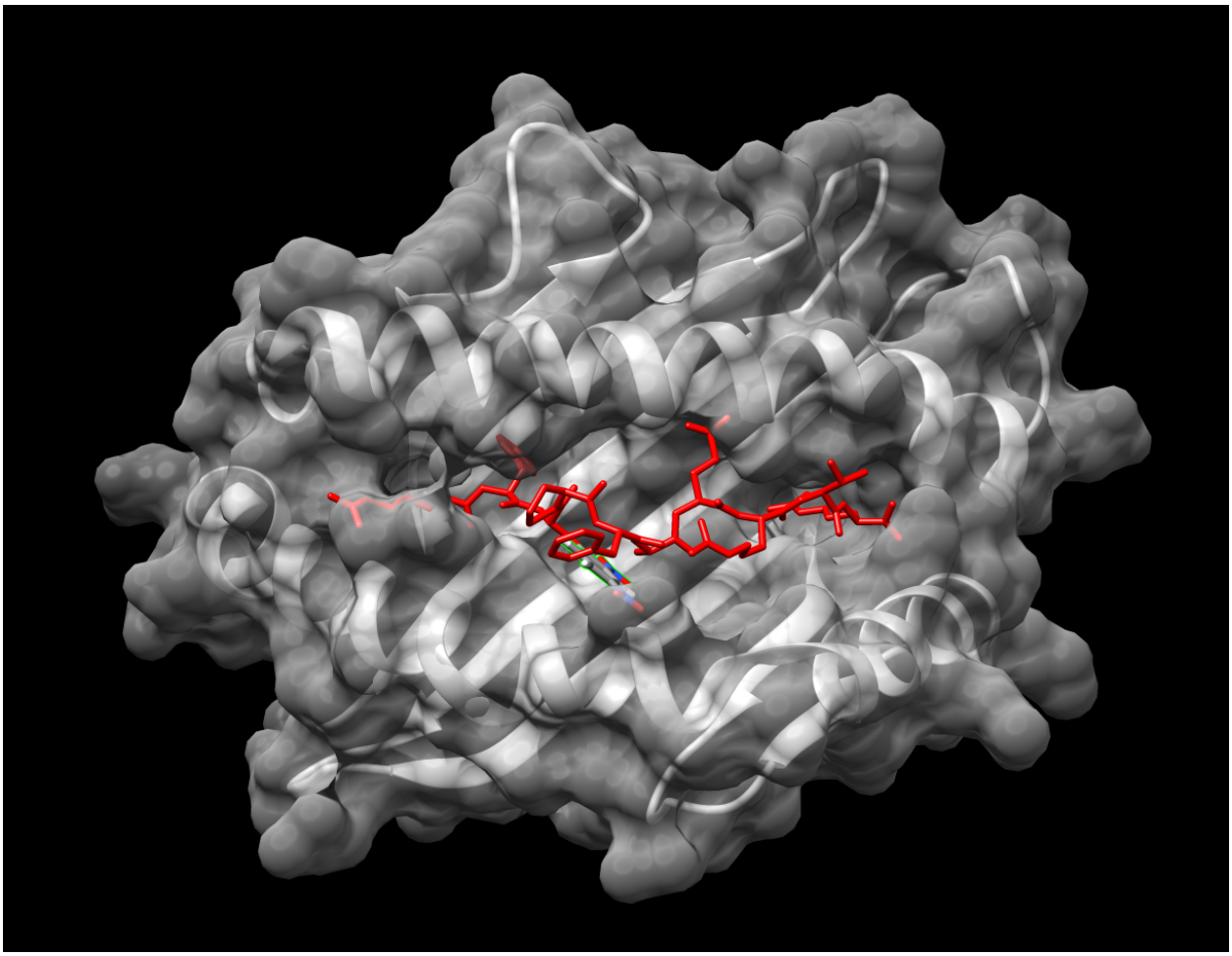

Figure 5.4: Putative structure of HLA-A*31:01 with 10mer RFC(+DNP)PFAERTR

Figure 5.5: Glutathione S-transferase omega-1. The glutathione ligand is shown in red, with the active site cysteine that is modified by DNCB highlighted in green.

## 5.4 Peptide predictions for Keratin type I cytoskeletal 13

To check for the possibility of other DNCB modified peptides from Keratin type I cytoskeletal 13, I used netMHC to find 9mer and 10mer peptides from the full protein sequences with C, K, Y or H in solvent exposed positions 3 to 8 of the peptide.

HLA-A*31:01, HLA-B*40:01 and HLA-B*51:01 have frequencies in the English population of about 6%, 11% and 9% respectively. Whereas for HLA-A*02:01 the frequency is approximately 50%.

Therefore I also used netMHC to predict peptides that would bind to HLA-A*02:01 that could be synthesised and modified for immunological testing, and in advance of the generation of HaCaT A2 transfected peptidome identification.

Table 5.2 shows a list of the predicted binders for these with potential for DNCB modification deriving from Keratin type I cytoskeletal 13.

## 5.5 Peptide predictions for Glutathione S-transferase omega-1

As for Keratin type I cytoskeletal 13, I did the same thing for Glutathione S-transferase omega-1.

Table 5.3 shows a list of the predicted binders for these with potential for DNCB modification deriving from Glutathione S-transferase omega-1 and binding to HLA-A*31:01, A*02:01, B*40:01 and B*51:01.

Table 5.2: NetMHC predicted Keratin type I cytoskeletal 13 peptides

| Peptide | Length | Allotype | Rank |
|---|---|---|---|
| RLASYLEKV | 9 | HLA-A*02:01 | 0.090 |
| SLNEELAYM | 9 | HLA-A*02:01 | 0.300 |
| MLLDIKTRL | 9 | HLA-A*02:01 | 0.175 |
| DRLASYLEKV | 10 | HLA-A*02:01 | 0.900 |
| RLKYENELAL | 10 | HLA-A*02:01 | 1.100 |
| KMLLDIKTRL | 10 | HLA-A*02:01 | 1.700 |
| RGVSTCSTR | 9 | HLA-A*31:01 | 0.600 |
| YYKTIEELR | 9 | HLA-A*31:01 | 0.600 |
| KMLLDIKTR | 9 | HLA-A*31:01 | 0.900 |
| RLASYLEKVR | 10 | HLA-A*31:01 | 1.000 |
| REQYEAMAER | 10 | HLA-A*31:01 | 0.800 |
| TSKTEITELR | 10 | HLA-A*31:01 | 1.500 |
| NEKITMQNL | 9 | HLA-B*40:01 | 1.400 |
| IEELRDKIL | 9 | HLA-B*40:01 | 1.400 |
| REQYEAMAE | 9 | HLA-B*40:01 | 1.100 |
| AEEWFHTKS | 9 | HLA-B*40:01 | 1.500 |
| AETECRYAL | 9 | HLA-B*40:01 | 0.015 |
| TECRYALQL | 9 | HLA-B*40:01 | 0.150 |
| QEYKMLLDI | 9 | HLA-B*40:01 | 0.800 |
| QEIATYRSL | 9 | HLA-B*40:01 | 0.050 |
| GNEKITMQNL | 10 | HLA-B*40:01 | 1.200 |
| LEVKIRDWHL | 10 | HLA-B*40:01 | 0.700 |
| AEMREQYEAM | 10 | HLA-B*40:01 | 0.400 |
| AEEWFHTKSA | 10 | HLA-B*40:01 | 1.600 |
| VAETECRYAL | 10 | HLA-B*40:01 | 0.300 |
| AETECRYALQ | 10 | HLA-B*40:01 | 0.600 |
| ETECRYALQL | 10 | HLA-B*40:01 | 1.200 |
| EQEIATYRSL | 10 | HLA-B*40:01 | 0.500 |
| QEIATYRSLL | 10 | HLA-B*40:01 | 0.250 |

Table 5.3: NetMHC predicted Glutathione S-transferase omega-1 peptides

| Peptide | Length | Allotype | Rank |
|---|---|---|---|
| YSMRFCPFA | 9 | HLA-A*02:01 | 0.70 |
| MILELFSKV | 9 | HLA-A*02:01 | 0.15 |
| ELFSKVPSL | 9 | HLA-A*02:01 | 1.20 |
| SISMIDYLI | 9 | HLA-A*02:01 | 1.70 |
| SMIDYLIWP | 9 | HLA-A*02:01 | 1.50 |
| LELFSKVPSL | 10 | HLA-A*02:01 | 0.60 |
| ELFSKVPSLV | 10 | HLA-A*02:01 | 2.00 |
| GSIRIYSMR | 9 | HLA-A*31:01 | 0.20 |
| MRFCPFAER | 9 | HLA-A*31:01 | 0.90 |
| RTRLVLKAK | 9 | HLA-A*31:01 | 0.50 |
| KNKPEWFFK | 9 | HLA-A*31:01 | 0.40 |
| EGSIRIYSMR | 10 | HLA-A*31:01 | 1.10 |
| SMRFCPFAER | 10 | HLA-A*31:01 | 0.09 |
| RFCPFAERTR | 10 | HLA-A*31:01 | 0.40 |
| KNKPEWFFKK | 10 | HLA-A*31:01 | 0.60 |
| KLKLWMAAMK | 10 | HLA-A*31:01 | 1.00 |
| GQLIYESAI | 9 | HLA-B*40:01 | 1.90 |
| KEDYAGLKE | 9 | HLA-B*40:01 | 1.40 |
| KEFTKLEEV | 9 | HLA-B*40:01 | 0.80 |
| FERLEAMKL | 9 | HLA-B*40:01 | 0.50 |
| SEKDWQGFL | 9 | HLA-B*40:01 | 0.90 |
| YESAITCEYL | 10 | HLA-B*40:01 | 0.15 |
| LDEAYPGKKL | 10 | HLA-B*40:01 | 1.50 |
| YEKACQKMIL | 10 | HLA-B*40:01 | 0.20 |
| LELFSKVPSL | 10 | HLA-B*40:01 | 0.25 |
| EEFRKEFTKL | 10 | HLA-B*40:01 | 1.00 |
| RKEFTKLEEV | 10 | HLA-B*40:01 | 1.20 |
| KEFTKLEEVL | 10 | HLA-B*40:01 | 0.50 |
| NECVDHTPKL | 10 | HLA-B*40:01 | 0.80 |
| EAYPGKKLL | 9 | HLA-B*51:01 | 2.00 |
| MILELFSKV | 9 | HLA-B*51:01 | 1.00 |
| DPYEKACQKM | 10 | HLA-B*51:01 | 1.20 |
| MAAMKEDPTV | 10 | HLA-B*51:01 | 0.60 |

# Chapter 6

# Next steps

Following on from the data presented here are the next steps:

1. Complete the protein turnover analysis to look for correspondence with the peptidome data.
2. Synthesise DNCB modified peptides to confirm the spectra and use in immunological assays to test for relevance.
3. Complete and compare HLA-A*02:01 transfectant peptidome and protein turnover assays.

# References


Andreatta, M., Alvarez, B., and Nielsen, M. (2017). Gibbscluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res*.

Andreatta, M. and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the mhc class i system. *Bioinformatics (Oxford, England)*, 32:511–517.

Boukamp, P., Petrussevska, R. T., Breitkreutz, D., Hornung, J., Markham, A., and Fusenig, N. E. (1988). Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J Cell Biol*, 106(3):761–71.

Caron, E., Aebersold, R., Banaei-Esfahani, A., Chong, C., and Bassani-Sternberg, M. (2017). A case for a human immuno-peptidome project consortium. *Immunity*, 47:203–208.

Jacquoilleot, S., Sheffield, D., Olayanju, A., Sison-Young, R., Kitteringham, N. R., Naisbitt, D. J., and Aleksic, M. (2015). Glutathione metabolism in the hacat cell line as a model for the detoxification of the model sensitisers 2,4-dinitrohalobenzenes in human skin. *Toxicology letters*, 237:11–20.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4):357.

Lemaitre, G., Lamartine, J., Pitaval, A., Vaigot, P., Garin, J., Bouet, S., Petat, C., Soularue, P., Gidrol, X., Martin, M. T., and Waksman, G. (2004). Expression profiling of genes and proteins in hacat keratinocytes: proliferating versus differentiated state. *J Cell Biochem*, 93(5):1048–62.

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S. L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2003). Reliable prediction of t-cell epitopes using neural networks with novel sequence representations. *Protein science : a publication of the Protein Society*, 12:1007–1017.

Parkinson, E., Aleksic, M., Cubberley, R., Kaur-Atwal, G., Vissers, J. P. C., and Skipp, P. (2018). Determination of protein haptenation by chemical sensitizers within the complexity of the human skin proteome. *Toxicological sciences : an official journal of the Society of Toxicology*, 162:429–438.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). Ucsf chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Spriggs, S., Sheffield, D., Olayanju, A., Kitteringham, N. R., Naisbitt, D. J., and Aleksic, M. (2016). Effect of repeated daily dosing with 2,4-dinitrochlorobenzene on glutathione biosynthesis and nrf2 activation in reconstructed human epidermis. *Toxicological sciences : an official journal of the Society of Toxicology*, 154:5–15.

van Hateren, A., Bailey, A., and Elliott, T. (2017). Recent advances in major histocompatibility complex (mhc) class i antigen presentation: Plastic mhc molecules and tapbpr-mediated quality control. *F1000Research*, 6:158.

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., Wheeler, D. K., Gabbard, J. L., Hix, D., Sette, A., and Peters, B. (2015). The immune epitope database (iedb) 3.0. *Nucleic acids research*, 43:D405–D412.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012). Peaks db: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*, 11(4):M111–010587.