

In [0]:

In [2]:

In [0]:

In [0]:

In [0]:

In [0]:

```
# Defines model
peptide_embedding_dim = 100
peptide_hidden_dim = 100
```

```

allele_embedding_dim = 100
allele_hidden_dim = 100

peptide_to_output_dim = peptide_hidden_dim*2*2
allele_to_output_dim = allele_hidden_dim*2*2

peptide_model = peptide_BiGRU(peptide_embedding_dim, peptide_hidden_dim).cuda()
allele_model = allele_BiGRU(allele_embedding_dim, allele_hidden_dim).cuda()
output_model = Output_Layer(peptide_model, allele_model, peptide_to_output_dim, allele_to_output_dim).cuda()
output_model.load_state_dict(torch.load('/content/drive/My Drive/Peptide MHC Project/Trained Models/trained_model.pth'))

```

Out[0]:

<All keys matched successfully>

In [0]:

```

affinities = []
for i in range(corona_peptides.shape[0]):
    peptide_affinities = []
    for j in range(alleles.shape[0]):
        peptide = torch.reshape(corona_embedding_tensor[i, :, :], (1, 15, 100))
        allele = torch.reshape(allele_embedding_tensor[j, :, :], (1, 106, 100))
        binding_pred = output_model(peptide, allele).item()
        peptide_affinities.append(binding_pred)
    affinities.append(peptide_affinities)
    if i % 100 == 0:
        print(i)

```

0
100
200
300
400
500
600

In [0]:

```

n_alleles = 30
all_affs = []
HLA_A_affs = []
HLA_B_affs = []
HLA_C_affs = []
[all_affs.extend(el) for el in affinities]
[HLA_A_affs.extend(el[:10]) for el in affinities]
[HLA_B_affs.extend(el[10:20]) for el in affinities]
[HLA_C_affs.extend(el[20:]) for el in affinities]

avg_aff = sum(all_affs) / len(all_affs)
meanHLA_A = sum(HLA_A_affs) / len(HLA_A_affs)
meanHLA_B = sum(HLA_B_affs) / len(HLA_B_affs)
meanHLA_C = sum(HLA_C_affs) / len(HLA_C_affs)

print("Mean Binding Affinity between peptides and 30 MHC-I alleles:", round(avg_aff, 3))
print("Mean Binding Affinity between peptides and 10 MHC-I alleles with gene:")
print("- HLA-A: {}".format(round(meanHLA_A, 3)))
print("- HLA-B: {}".format(round(meanHLA_B, 3)))
print("- HLA-C: {}\n".format(round(meanHLA_C, 3)))

```

Mean Binding Affinity between peptides and 30 MHC-I alleles: 0.714
Mean Binding Affinity between peptides and 10 MHC-I alleles with gene:
- HLA-A: 0.549
- HLA-B: 0.706
- HLA-C: 0.888

In [0]:

```

# Shows binding affinities by percentile
x = sorted(all_affs)
print("25th percentile:", round(x[int(len(x)*.25)], 3))
print("50th percentile:", round(x[int(len(x)*.5)], 3))
print("75th percentile: %f\n" % round(x[int(len(x)*.75)], 3))

# Almost all peptides will certainly bind to at least 1 MHC allele
n_peptides = len(affinities)
near_one_all = []
near_zero_all = []
either_all = []
for pep_affs in affinities:
    near_one = []
    near_zero = []

```

```

either = []
for aff in pep_affs:
    near_one.append(aff >= 0.9)
    near_zero.append(aff <= 0.1)
    either.append(aff >= 0.9 or aff <= 0.1)
near_one_all.append(near_one)
near_zero_all.append(near_zero)
either_all.append(either)

pct_near_one = sum([sum(near_one) / n_alleles for near_one in near_one_all]) / n_peptides
pct_near_zero = sum([sum(near_zero) / n_alleles for near_zero in near_zero_all]) / n_peptides
pct_either = sum([sum(either) / n_alleles for either in either_all]) / n_peptides
print("Percentage of binding affinities over 0.9: {}".format(round(pct_near_one*100, 1)))
print("Percentage of binding affinities under 0.1: {}".format(round(pct_near_zero*100, 1)))
print("Either (under 0.1 or over 0.9): {}".format(round(pct_either*100, 1)))

pctHLA_A = sum([sum(in_range[:10]) / 10 for in_range in near_one_all]) / n_peptides
pctHLA_B = sum([sum(in_range[10:20]) / 10 for in_range in near_one_all]) / n_peptides
pctHLA_C = sum([sum(in_range[20:]) / 10 for in_range in near_one_all]) / n_peptides
print("Percent over 0.9, by gene:")
print("- HLA-A: {}".format(round(pctHLA_A*100, 1)))
print("- HLA-B: {}".format(round(pctHLA_B*100, 1)))
print("- HLA-C: {}".format(round(pctHLA_C*100, 1)))

25th percentile: 0.29
50th percentile: 0.997
75th percentile: 1.000000

Percentage of binding affinities over 0.9: 63.4%
Percentage of binding affinities under 0.1: 21.0%
Either (under 0.1 or over 0.9): 84.4%
Percent over 0.9, by gene:
- HLA-A: 44.7%
- HLA-B: 61.8%
- HLA-C: 83.8%

```

In [0]:

```

# Finds average affinity by protein
# Extracts each protein's amino acid sequence
corona = pd.read_csv("coronavirus_netchop.csv")
aa_indices = corona.index[corona['pos'] == 1].tolist()
n_proteins = len(aa_indices)
protein_list = []
for i in range(n_proteins):
    if i < len(aa_indices) - 1:
        end_index = aa_indices[i + 1]
    else:
        end_index = corona.shape[0]
    protein = "".join(corona['AA'].iloc[aa_indices[i]:end_index].to_list())
    protein_list.append(protein)

# Finds which proteins each peptide is in
protein_affinities = [[] for _ in range(n_proteins)]
for i in range(n_peptides):
    peptide = corona_peptides.peptide.iloc[i]
    for j in range(n_proteins):
        if peptide in protein_list[j]:
            protein_affinities[j].append(affinities[i])

# Calculates average binding affinity per protein
print("Number of proteins:", n_proteins)
for i in range(n_proteins):
    count = len(protein_affinities[i])
    if count == 0:
        continue
    avg_aff = sum([sum(pep_affs) / len(pep_affs) for pep_affs in protein_affinities[i]]) / count
    print("Mean Binding Affinity between 30 MHC-I alleles and the {} peptides found in protein {}: {}".format(count, i+1, round(avg_aff, 3)))

```

```

Number of proteins: 14
Mean Binding Affinity between 30 MHC-I alleles and the 14 peptides found in protein 2: 0.75
Mean Binding Affinity between 30 MHC-I alleles and the 25 peptides found in protein 3: 0.83
Mean Binding Affinity between 30 MHC-I alleles and the 4 peptides found in protein 4: 0.728
Mean Binding Affinity between 30 MHC-I alleles and the 10 peptides found in protein 5: 0.68
Mean Binding Affinity between 30 MHC-I alleles and the 4 peptides found in protein 7: 0.669
Mean Binding Affinity between 30 MHC-I alleles and the 1 peptides found in protein 8: 0.451
Mean Binding Affinity between 30 MHC-I alleles and the 447 peptides found in protein 9: 0.712
Mean Binding Affinity between 30 MHC-I alleles and the 287 peptides found in protein 10: 0.696
Mean Binding Affinity between 30 MHC-I alleles and the 98 peptides found in protein 11: 0.711
Mean Binding Affinity between 30 MHC-I alleles and the 4 peptides found in protein 12: 0.814
Mean Binding Affinity between 30 MHC-I alleles and the 16 peptides found in protein 13: 0.612
Mean Binding Affinity between 30 MHC-I alleles and the 2 peptides found in protein 14: 0.721

```

In [0]:

```

# Protein 11 is the Spike Protein
count = len(protein_affinities[10])
spike_avg = sum([sum(pep_affs) / len(pep_affs) for pep_affs in protein_affinities[10]]) / count
print("Spike protein (11) mean binding affinity: {}".format(round(spike_avg, 3)))

meanHLA_A = sum([sum(pep_affs[:10]) / 10 for pep_affs in protein_affinities[10]]) / count
meanHLA_B = sum([sum(pep_affs[10:20]) / 10 for pep_affs in protein_affinities[10]]) / count
meanHLA_C = sum([sum(pep_affs[20:]) / 10 for pep_affs in protein_affinities[10]]) / count
print("By gene:")
print("- HLA-A: {}".format(round(meanHLA_A, 3)))
print("- HLA-B: {}".format(round(meanHLA_B, 3)))
print("- HLA-C: {}".format(round(meanHLA_C, 3)))

Spike protein (11) mean binding affinity: 0.711
By gene:
- HLA-A: 0.518
- HLA-B: 0.731
- HLA-C: 0.885

```

In [0]:

```

# Breaks peptides apart by length
length_affinities = [[] for _ in range(8)]
for i in range(n_peptides):
    peptide = corona_peptides.peptide.iloc[i]
    length_affinities[len(peptide)-8].append(affinities[i])

# Calculates mean binding affinity by peptide length
for i in range(8):
    all = []
    [all.extend(peps) for peps in length_affinities[i]]
    avg_aff = sum(all) / len(all)
    print("{} peptides of length {}".format(len(length_affinities[i]), i+8))
    print("- Mean binding affinity:", round(avg_aff, 2))
    sorted_affs = sorted(all)
    count = len(sorted_affs)
    first_idx = [aff >= 0.9 for aff in sorted_affs].index(True)
    pct_over = round((count - first_idx) / count * 100)
    print("- Percentage of binding affinities above 0.9: {}%".format(pct_over))

100 peptides of length 8
- Mean binding affinity: 0.82
- Percentage of binding affinities above 0.9: 76%
250 peptides of length 9
- Mean binding affinity: 0.57
- Percentage of binding affinities above 0.9: 45%
150 peptides of length 10
- Mean binding affinity: 0.7
- Percentage of binding affinities above 0.9: 61%
25 peptides of length 11
- Mean binding affinity: 0.89
- Percentage of binding affinities above 0.9: 86%
25 peptides of length 12
- Mean binding affinity: 0.94
- Percentage of binding affinities above 0.9: 90%
25 peptides of length 13
- Mean binding affinity: 0.98
- Percentage of binding affinities above 0.9: 96%
25 peptides of length 14
- Mean binding affinity: 0.98
- Percentage of binding affinities above 0.9: 97%
25 peptides of length 15
- Mean binding affinity: 0.94
- Percentage of binding affinities above 0.9: 93%

```

In [0]:

In [0]: