Predicting pMHC-I Binding from LC-MS/MS Data
Using Hidden Markov Models

Jordan Force

August 23, 2017

Table of Contents

Predicting pMHC-I Binding from LC-MS/MS Data

Jordan Force

Background
Data
Prediction Methods
Future Work

1 Background

2 Data

3 Prediction Methods

4 Future Work

Table of Contents

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

❶ Background

❷ Data

❸ Prediction Methods

❹ Future Work

MHC-I Function

Predicting
pMHC-I Binding
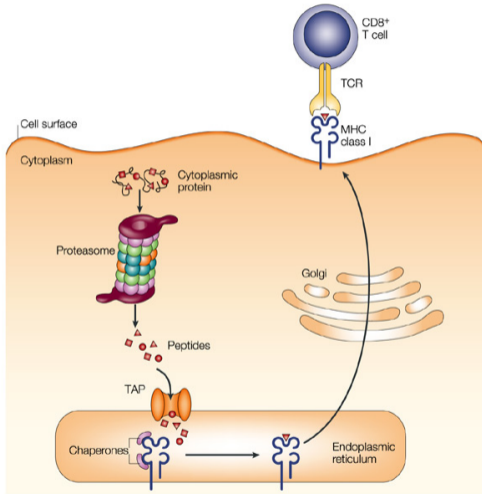from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

Nature Reviews | Immunology

Figure: The function of the MHC-I

A More Memorable Diagram

Predicting
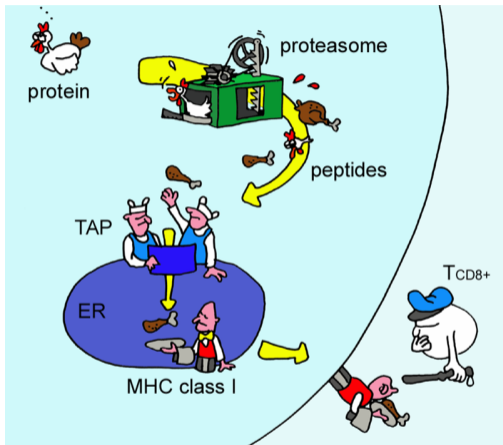pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Figure: The function of the MHC-I, with chicken

Measuring Affinity

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

1. Produce MHC-I Protein
2. Fold it *in-vitro* with peptide
3. Measure stability of peptide-MHC complex
4. Or competition with a standard known binder

Disadvantages to this Approach

Predicting pMHC-I Binding from LC-MS/MS Data

Jordan Force

Background

Data

Prediction Methods

Future Work

1. Binding not done in normal environment

Note: This is a work in progress!

Table of Contents

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

1 Background

2 Data

3 Prediction Methods

4 Future Work

MS/MS Data Source

1. Data came from Abelin et al.
2. Used cell lines expressing a single HLA allele
3. Sequenced presented peptides (from endogenous proteins) using LC-MS/MS

Improvements

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work
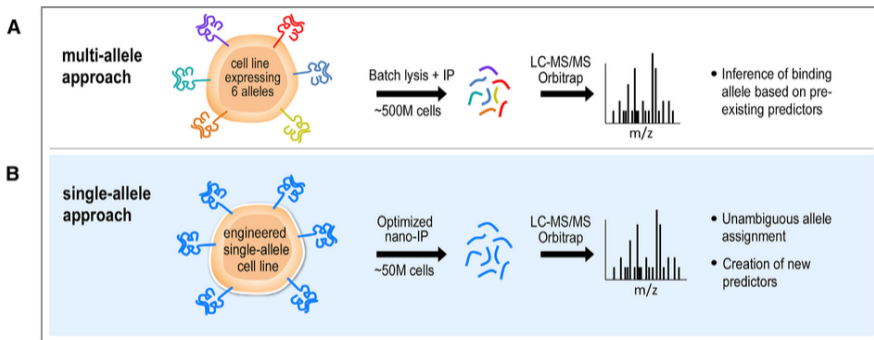
Figure: Improvement

How MS/MS Works

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

Figure: How Mass Spec Works

Table of Contents

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

1 Background

2 Data

3 Prediction Methods

4 Future Work

PWM

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

1. Construct a matrix with frequency of amino acid at each position
2. Compute likelihood of a given peptide under this matrix

HMM Structure
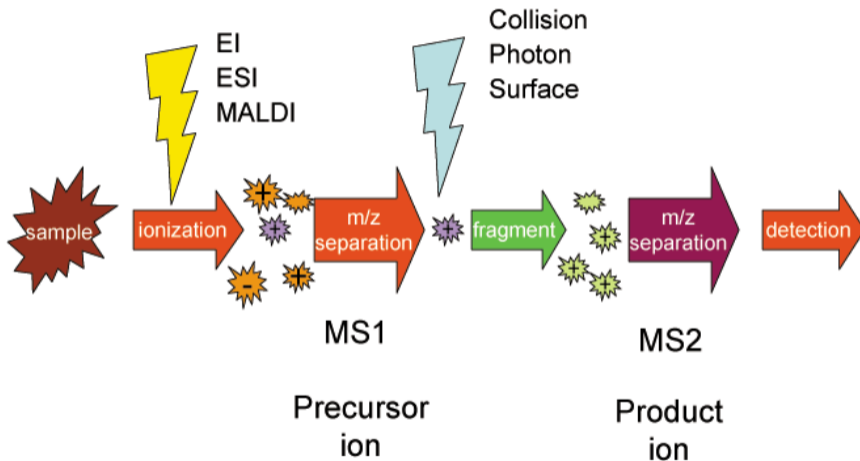
Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

Left-to-right HMM, with n rows, and w columns. w is length of peptides.
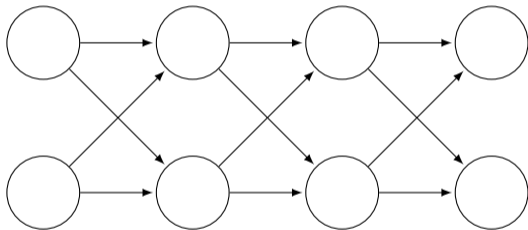
Figure: HMM with $w = 4$ and $n = 2$

HMM Training

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

1. Train HMM *only* on the binder dataset
2. Used EM algorithm included in *hmmlearn* package
3. Ran EM algorithm 10 times, picked parameters that gave highest likelihood.

HMM Scoring

1. Compute probability $\log_2(P(x_1 x_2 \dots x_w))$ of sequence under HMM model
2. Adjust classification threshold and generate ROC curve

(some) Results

① Generated random peptides, uniform distribution of amino acids

Result for peptides of length 9, HLA-A0101

HLA: HLA-A0101, Num positives: 447
length: 9

(some) Results

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

1. Selected random peptides from human proteome

Result for peptides of length 9, HLA-A0101

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

HLA: HLA-A0101, Num positives: 447
length: 9

Legend:
- Fold 0, AUC: 1.00
- Fold 1, AUC: 1.00
- Fold 2, AUC: 1.00
- Fold 3, AUC: 0.99
- Fold 4, AUC: 1.00
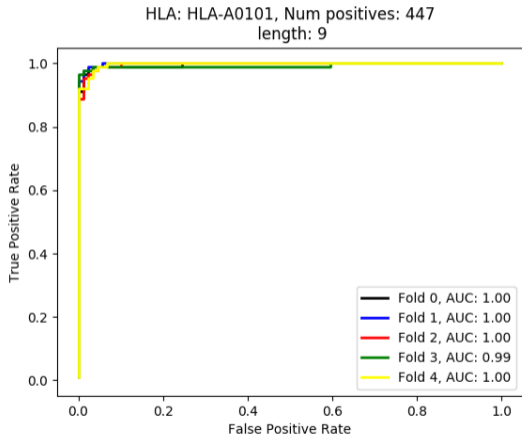
(some) Results

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

1. Used peptides that were presented by HLA-A0201, HLA-A0203, HLA-A0204 and HLA-A0207, but not HLA-A0101, as negative
2. Still performed very well

HLA: HLA-A0101, Num positives: 447
length: 9

Table of Contents

Predicting
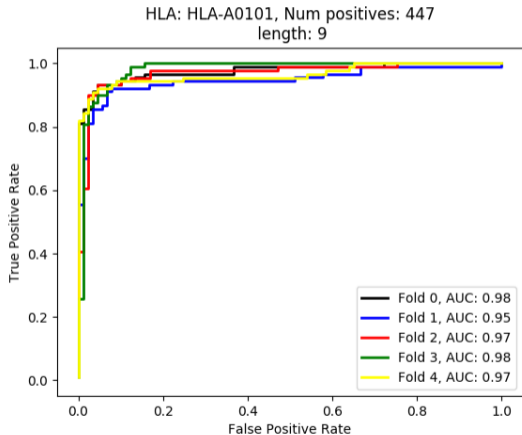pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background
Data
Prediction
Methods
Future Work

Generate data for mouse MHC

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background
Data
Prediction
Methods
Future Work

UConn recently bought an MS/MS system appropriate for protein work; will generate data for mouse MHC alleles

Flanking Residues

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

Upstream

	U10	U9	U8	U7	U6	U5	U4	U3	U2	U1	
	+4	-4	-1	-7	-1	+2	+3	+12	+18	-10	Q
	+8	+7	+7	+2	+2	+1	-4	+13	+24	+9	N
	-6	-1	-6	-2	-8	-5	-2	+4	+10	-16	T
	-10	-10	-10	-9	-6	-10	-11	-9	+8	+8	S
	-4	-5	-3	-2	-8	+2	+2	+25	+2	+28	R
	+11	+13	+6	+11	+7	+12	+7	+13	+8	+30	K
	-11	-1	-3	+0	+1	-3	0	+7	+3	-8	H
	-13	-8	-6	-11	-7	-20	-22	-45	-62	-70	P
	+6	+3	-1	-2	+4	+4	+14	+5	+21	+31	A
	-1	+1	-1	-2	+2	+0	-13	-22	-1	-1	G
	+5	+3	+5	+2	+0	+3	+2	+1	+5	-5	E
	+11	+13	+8	+7	+2	-4	-9	-1	+3	+12	D
	-11	-13	-24	-9	-11	-14	-12	+6	+9	+7	C
	-3	+10	+4	-3	+13	+9	+9	+9	+11	+9	M
	-1	0	+1	+4	-1	+6	+4	+3	-4	+12	L
	+1	+4	+6	+8	+9	+6	+10	+9	-7	-24	V
	+12	+5	+9	+12	+2	+5	+18	-1	-29	-25	I
	0	-2	+10	-2	+7	+9	+13	+10	+14	+14	Y
	+6	-6	+4	+4	+6	+2	+3	-6	-5	-6	F
	-11	-10	-1	-15	-5	-13	-11	-22	-11	-5	W
	-13	-13	-12	-10	-8	-10	-12	-11	-10	-31	-

Flanking Residues (continued)

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

Downstream

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Q	+11	+2	0	+0	+1	−2	0	−1	+3	+0
N	+17	+16	+13	+9	+9	−4	+4	+7	+5	+4
T	−5	−8	+1	0	−4	+1	−4	−6	−8	−4
S	+18	−6	−4	−10	−7	−10	−7	−10	−7	−10
R	+27	+13	+0	+0	+1	+1	−2	−5	−2	−4
K	+51	+29	+13	+13	+11	+8	+10	+10	+9	+8
H	+21	−17	+4	+5	−1	−2	+0	0	−12	−3
P	−41	−24	−18	+1	−3	−6	−4	−7	−6	−8
A	+42	+2	+8	+6	−2	−1	−1	−1	−2	0
G	+6	+17	+8	+2	+4	+7	+5	−2	0	−2
E	−20	−10	+9	−1	+9	+2	+7	+6	+4	+7
D	−18	+7	+13	+6	+8	+15	+12	+9	+1	+12
C	+1	−12	−13	−14	−18	−17	−24	−17	−26	−17
M	+17	+16	−3	−1	−5	+3	+0	+5	+10	−2
L	−23	+3	−14	−12	−10	−5	−6	−5	+0	−2
V	−10	−8	−8	+7	0	−1	+4	+4	+6	+7
I	−41	−2	−15	−2	−1	+7	+2	+10	+5	+2
Y	−19	−20	+6	+9	+3	−3	+3	+4	−2	+2
F	−27	−23	+3	−13	−6	−5	−8	+1	+3	−4
W	−55	−24	−13	−22	−19	−21	−23	−18	−5	−8
−	+271	+114	+79	+61	+47	+39	+30	+26	+21	+20

Higher Order Models

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

1. Allow for relationships between residues that are non-adjacent

Predicting
pMHC-I Binding
from
LC-MS/MS
Data

Jordan Force

Background

Data

Prediction
Methods

Future Work

Jennifer G Abelin, Derin B Keskin, Siranush Sarkizova, Christina R Hartigan, Wandi Zhang, John Sidney, Jonathan Stevens, William Lane, Guang Lan Zhang, Thomas M Eisenhaure, et al.
Mass spectrometry profiling of hla-associated peptidomes in mono-allelic cells enables more accurate epitope prediction.
Immunity, 46(2):315–326, 2017.