An Radial Basis Network to Describe The Antibody Antigen Interactions

Liu Chuanxing

March 17, 2020

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Table of contents

Introduction

- Significance
- The Problem
- The Challenge and the solution
- The outline of our research

Data Extraction

- Complexes, CDRs, Contact Number
- Reduce Redundance
- Positive Cores, Negative Cores

Develop an Radial Basis Function Network(RBFN)

- Define the distance between two cores
- Cross Validation and testing

Applications of RBFN

- Antibody aa and antigen aa are different
- RBFN can predict the affinity

Significance

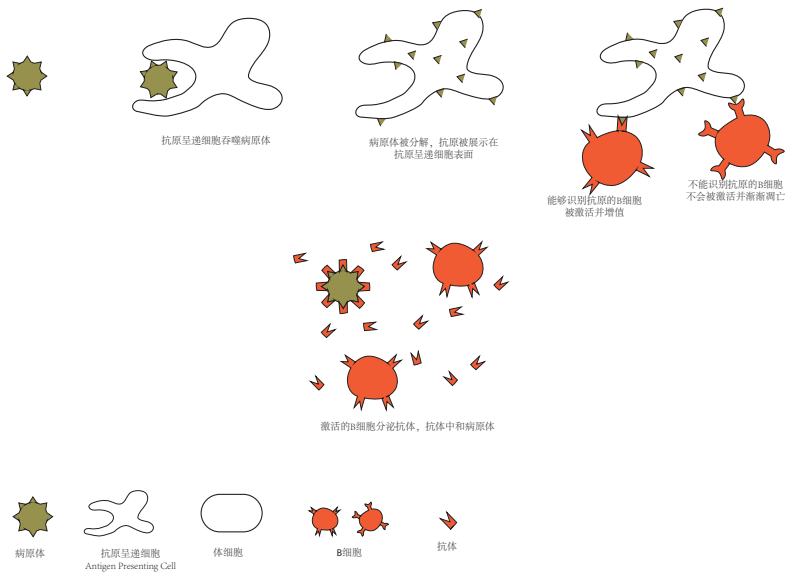


Figure: An illustration of serum immunity

- Design therapeutic antibodies
- Design effective vaccines

An Radial Basis Network to Describe The Antibody Antigen Interactions

Liu Chuanxing

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

The Problem

Do similar B-cell epitopes interact with similar paratopes?

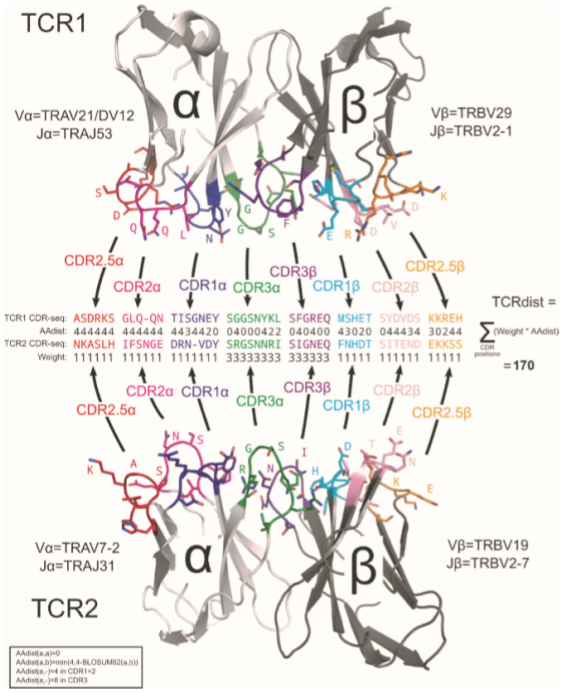
Here is a similar research about TCR and its epitopes

LETTER

doi:10.1038/nature22383

Quantifiable predictive features define epitope–specific T cell receptor repertoires

Pradyot Dash[1], Andrew J. Fiore-Gartland[2], Tomer Hertz[2,3], George C. Wang[4], Shalini Sharma[5], Aisha Souquette[1], Jeremy Chase Crawford[1], E. Bridie Clemens[6], Thi H. O. Nguyen[6], Katherine Kedzierska[6], Nicole L. La Gruta[6,7], Philip Bradley[8,9] & Paul G. Thomas[1]



An Radial Basis Network to Describe The Antibody Antigen Interactions

Liu Chuanxing

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

The Challenge

The interactions between B-cell epitopes and paratopes are conformational!

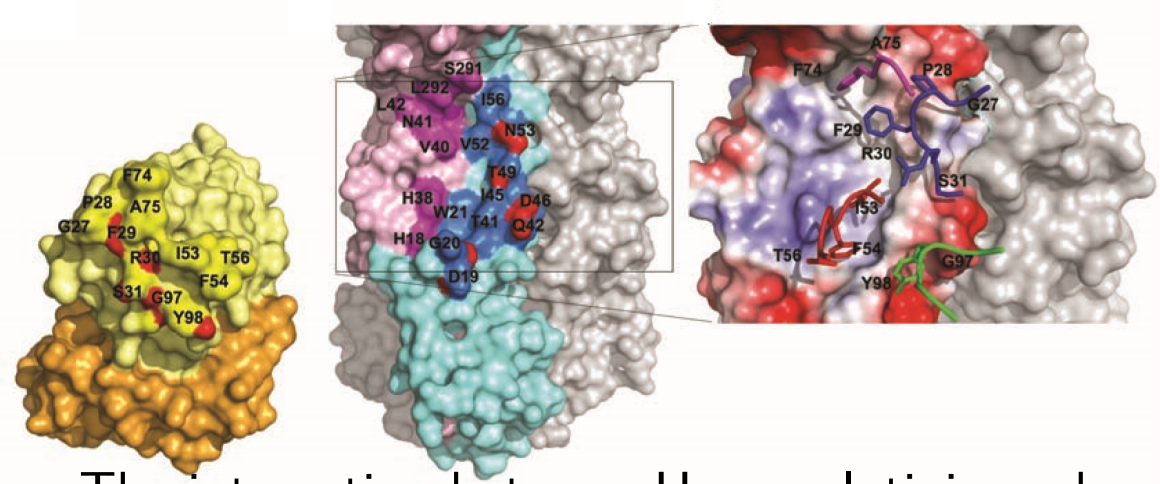


Figure: The interaction between Hemagglutinin and one of its antibodies

Solution of the challenge

We solved the challenge by focusing on the key amino acids(hot spots). As long as the sequence is short enough, the hot spots are composed of continuous sequences.

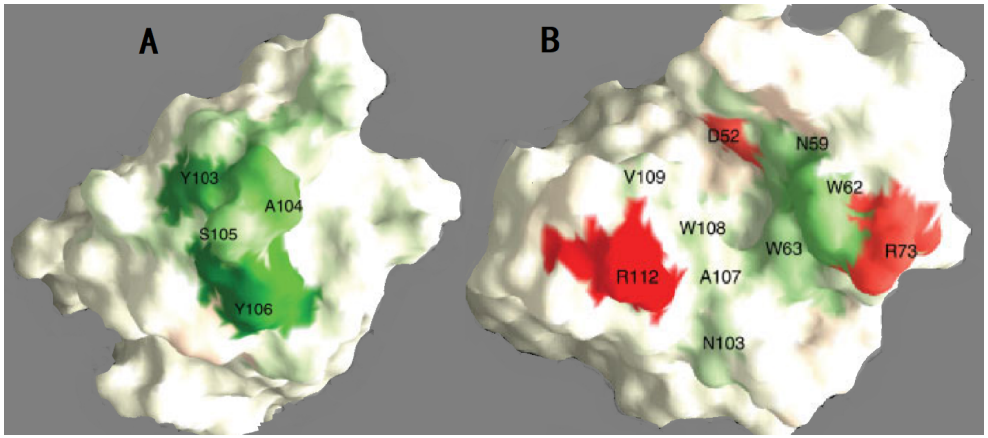


Figure: The hot spots between Hen Egg Lysozym(HEL) and one of its antibodies cAB-Lys3

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

The outline of our research

- Data extraction
- Develop an Radial Basis Function Network(RBFN)
- Use the RBFN to tell the difference between the antibody amino acids and the antigen amino acids.
- Use the RBFN to predict how a mutation can affect the affinity of an antibody-antigen complex.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Complex, CDR, Interacting Pairs

- 1624 antibody-antigen complexes with resolution $\leq 3A$.
- The CDRs are defined as follows

	CDR1	CDR2	CDR3
Max-CDRL	24 to 41	50 to 64	90 to 108
Max-CDRH	26 to 38	51 to 72	100 to 130

Table: Locations of the CDRs

- A and B are two amino acids, the contact number between A and B is defined as

$$\mathrm{CN}(A,B) = \sum_{a\in A}\sum_{b\in B}\chi\{d(a,b)\leq 4\}$$

$a \in A$ means a is an atom in A and a is not a hydrogen atom.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Define distance

The redundance reduction was based on the similarity of the CDRs. Each light/heavy chain was an individual.

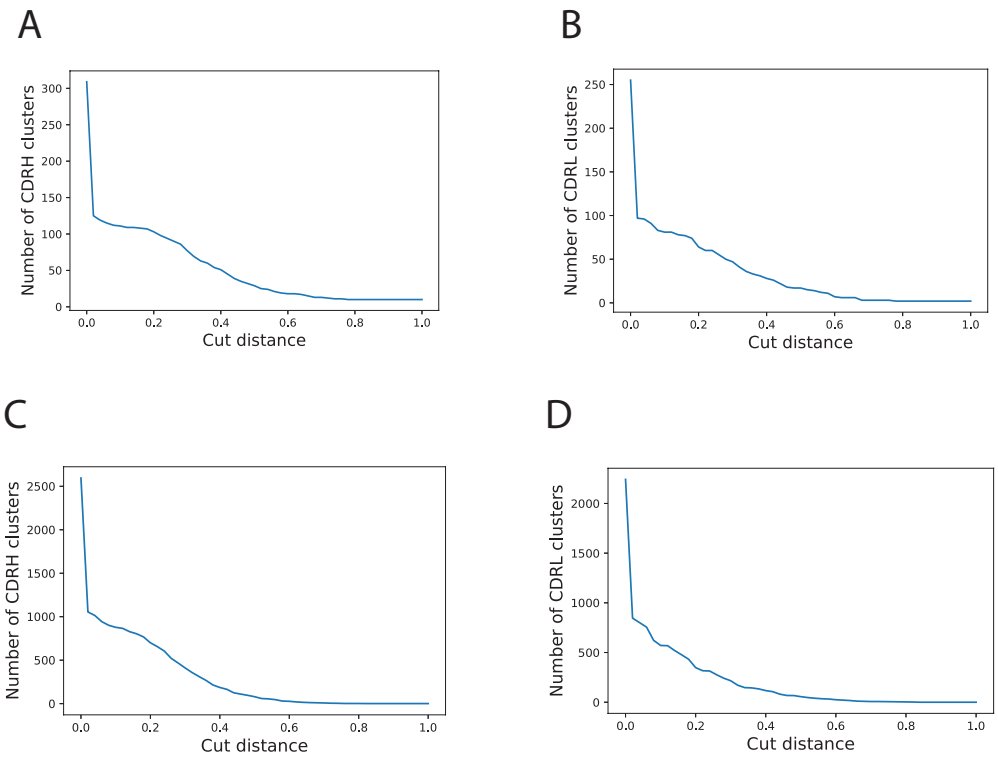
- **Scoring rules**

$$S(a,a) = 1; \quad S(a,b) = 0; \quad S(a,-) = 0$$

- **Calculate the distance** Concatenate the three CDRs. Suppose A and B are two concatenated CDRs of two light/heavy chains. $D(A,B) = 1 - S(A,B)/N$ where $N = min(Len(A), Len(B))$.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Select the cut-off distance



According to the Elbow Method, we chose 0.1 as the cut-off distance.

Select the representative from each cluster

Suppose A is the set of all the amino acids for a given light/heavy chain, and Ag is the set of all the amino acids in the corresponding antigen. The total contact number of A is defined as

$$TCN(A) = \sum_{a \in A} \sum_{b \in Ag} CN(a, b)$$

In each cluster, the chain with the largest TCN was selected as the representative.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Positive cores and Negative cores

- **Match-type** was defined as the lengths of the interacting sequences. Match-type(2,3) means 2 consecutive antibody amino acids interacting with 3 consecutive antigen amino acids.
- For a light/heavy chain and a match-type(m,n), the core is defined as the interacting sequences of match-type(m,n), with the largest contact number.
- The negative cores were randomly generated sequences pairs which were not in the training set.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

What do we have?

- The training set of different match-types. For each match-type, the training set consists of the positive cores and the randomly generated negative cores.
- The testing set of different match-types. For each match-type, there are 10 different testing sets, generated by combining the positive testing set with the 10 independently generated negative cores.
- The label for the positive cores is 1 and the label for the negative cores is -1.

Here, the range of the match-type is

$$\{(m, n) : m, n = 1, 2, 3\}$$

.

An Radial Basis Network to Describe The Antibody Antigen Interactions

Liu Chuanxing

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

The Substitution matrix

$$\text{Substitution matrix} = BLOSUM62$$
$$\text{gap} = Hp_1$$
$$\text{extended gap} = Hp_2$$

Here Hp_1 and Hp_2 were two hyperparameters. We use the complete BLOSUM62, not the truncated BLOSUM62 as Pradyot Dash did.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

The distance between two cores

Suppose (Ab1, Ag1) and (Ab2, Ag2) are two cores of match-type(m,n). This distance between them is defined by the following steps.

$$
\begin{aligned}
S_{Ab} &= \mathrm{Aln}(\mathrm{Ab}_1, \mathrm{Ab}_2) \\
S_{Ag} &= \mathrm{Aln}(\mathrm{Ag}_1, \mathrm{Ag}_2) \\
S^{+}_{Ab} &= \frac{\mathrm{S_{Ab}} + 4 \times m}{15 \times m} \\
S^{+}_{Ag} &= \frac{\mathrm{S_{Ag}} + 4 \times n}{15 \times n} \\
S &= S^{+}_{Ab} \times S^{+}_{Ag} \\
D &= 1 - S
\end{aligned}
$$

D is the distance defined.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Cross Validation

match-type	Hp_1	Hp_2	r	p	average AUC
(1,1)	-1	-1	0.0001	0.8	0.973
(1,2)	-1	-1	0.0001	0.8	0.860
(1,3)	-1	-1	0.0001	0.8	0.834
(2,1)	-1	-1	0.0001	0.8	0.870
(2,2)	-1	-1	0.0001	0.8	0.842
(2,3)	-1	-1	0.001	0.8	0.836
(3,1)	-1	-1	0.0001	0.8	0.867
(3,2)	-1	-1	0.001	0.8	0.862
(3,3)	-1	-1	0.001	0.8	0.862

Table: We did a 5 cross validation. The best parameter are the values corresponding to the highest average AUC.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Testing the model

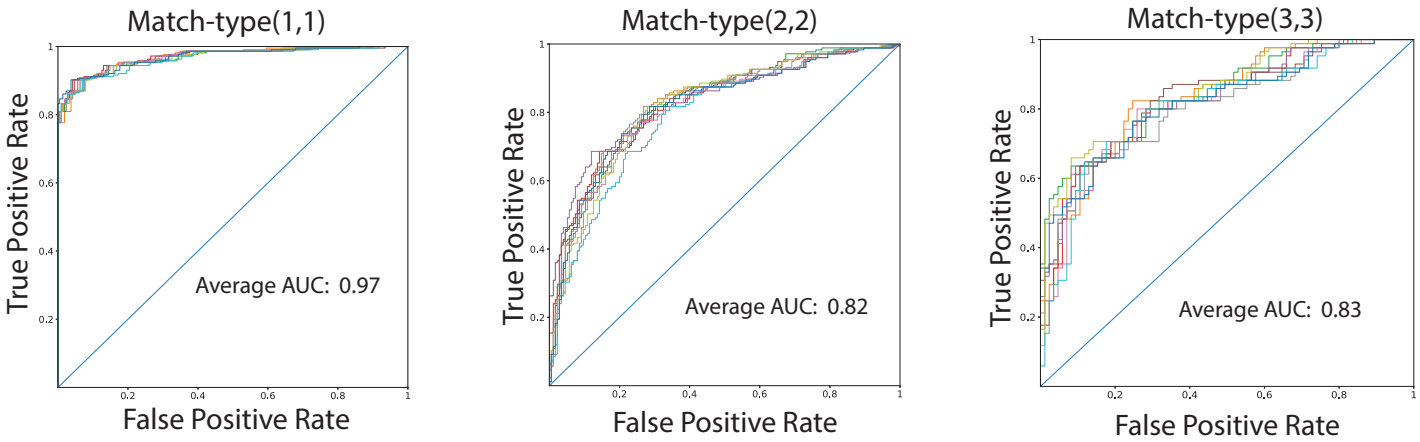
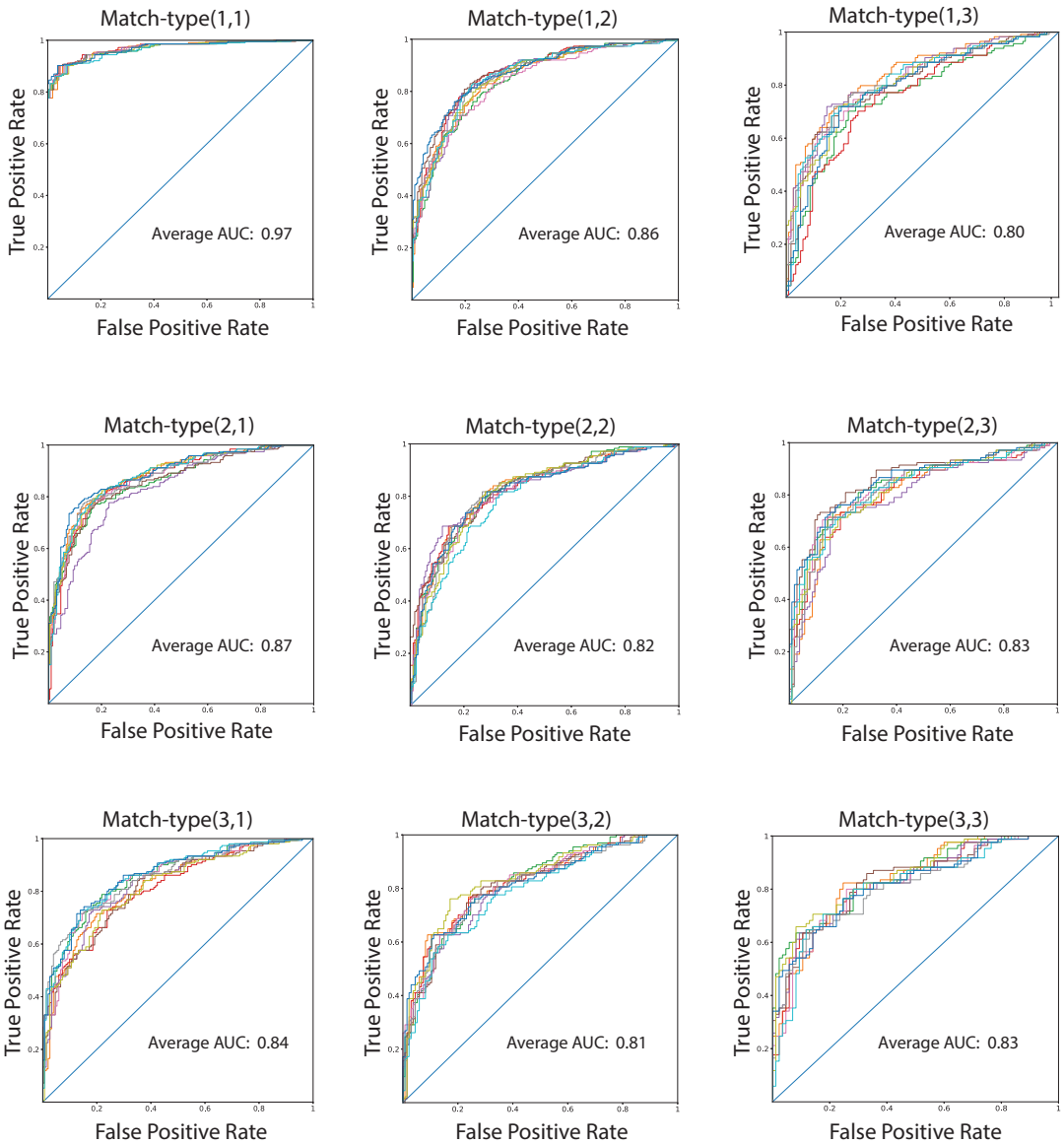


Figure: For each match-type, the testing were run on 10 independent testing set. The average AUC were calculated.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Testing the model



Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Antibody aa and antigen aa are different

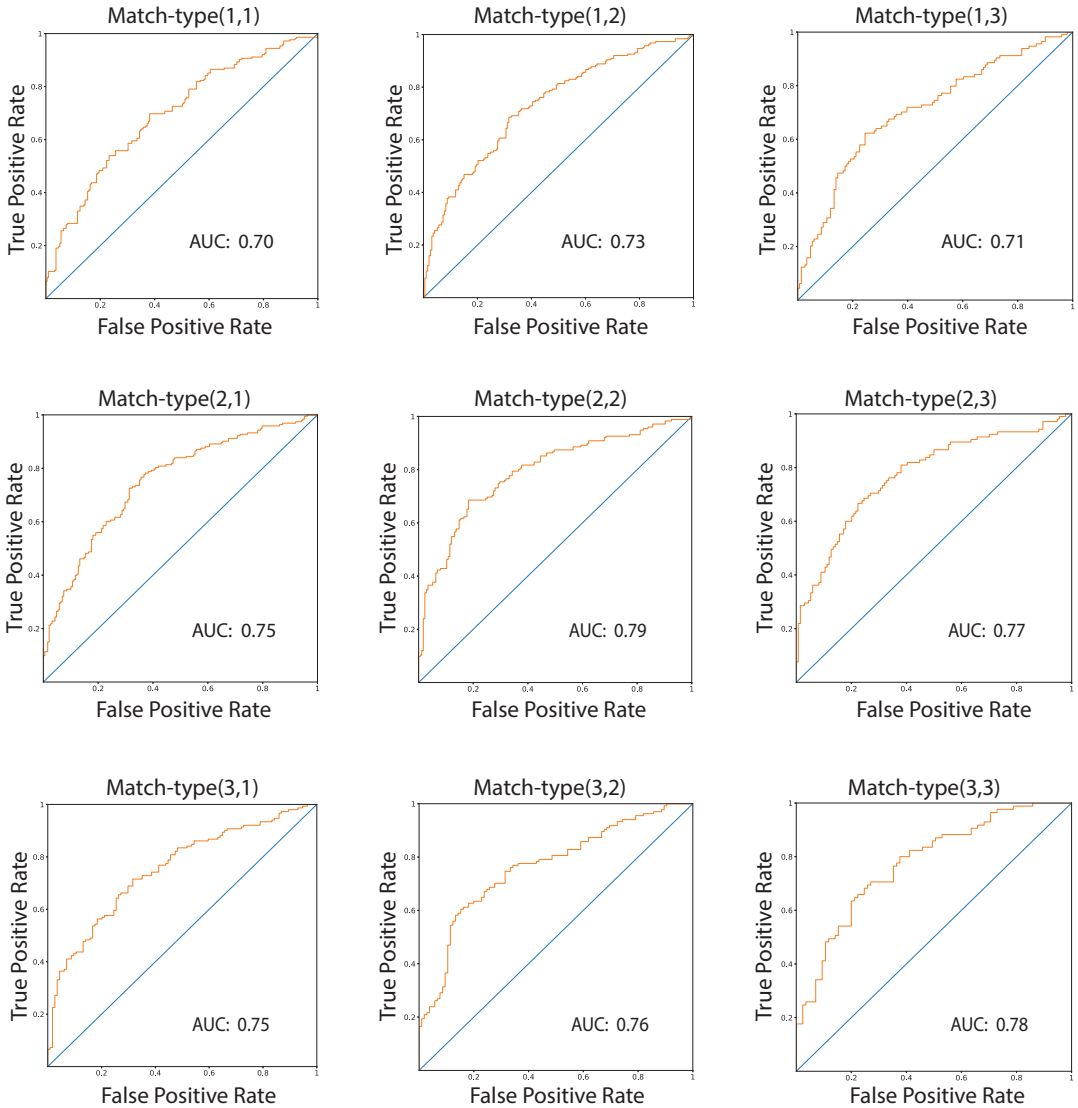
Suppose (AbSeq, AgSeq) is of match-type(m,n). If there is no difference between AbSeq and AgSeq, then our model will not be able to tell the difference between (AgSeq, AbSeq) and a positive core of match-type(n,m).

To prove the above statement, the testing set for each match-type was constructed as follows.

$$\mathrm{TR}_{(n,m)} = \{(AgSeq, AbSeq) : (AbSeq, AgSeq) \in \mathrm{T}_{(m,n)}\}$$
$$\mathrm{T} = \mathrm{TR}_{(n,m)} \cup \mathrm{T}_{(n,m)}$$

Here T is the testing set of match-type(n,m).

Antibody aa and antigen aa are different



Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

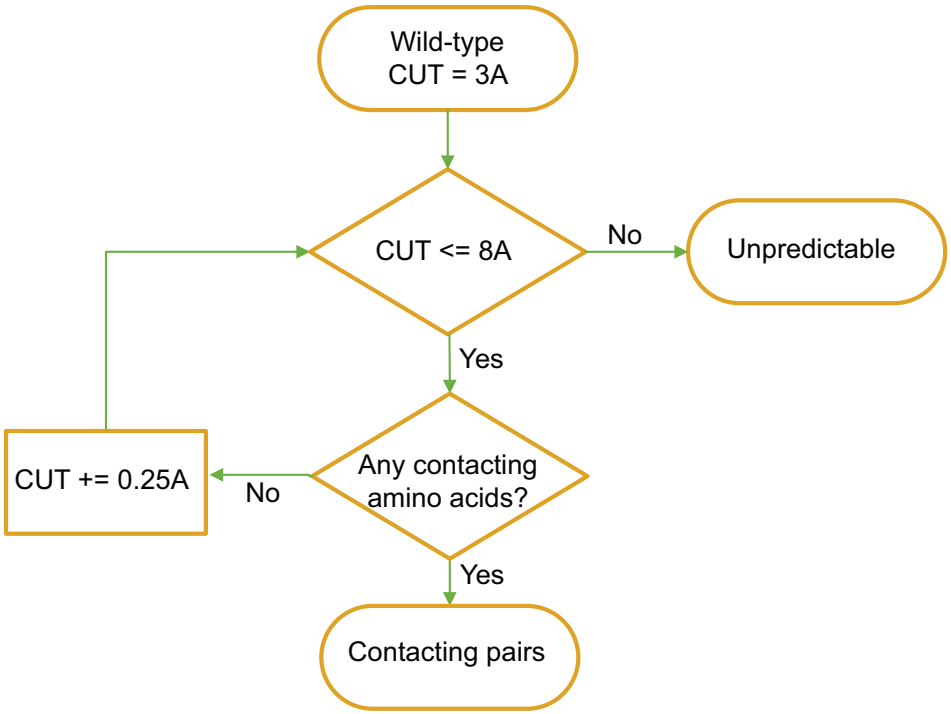
Basic assumptions

Assumption: if a mutation changes the interacting sequences towards the direction of positive cores, then it increases the affinity.

We uses the RBFN on match-type(1,1) to make predictions

Predictable Pairs

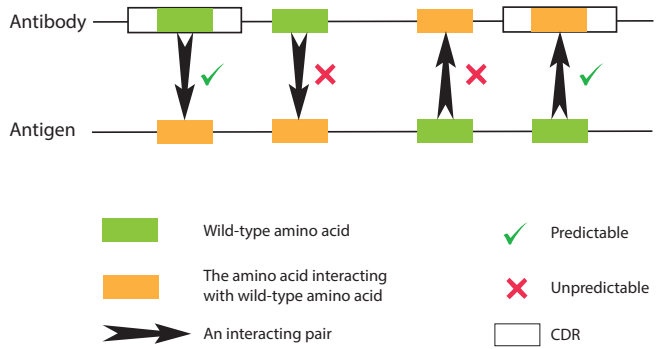
Step 1: Find the contacting pairs for each mutation.



Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Predictable Pairs

Step 2: Generate all predictable pairs



Step 3: For each mutation, pick the one with the largest contact number from the all the predictable pairs.

Make prediction

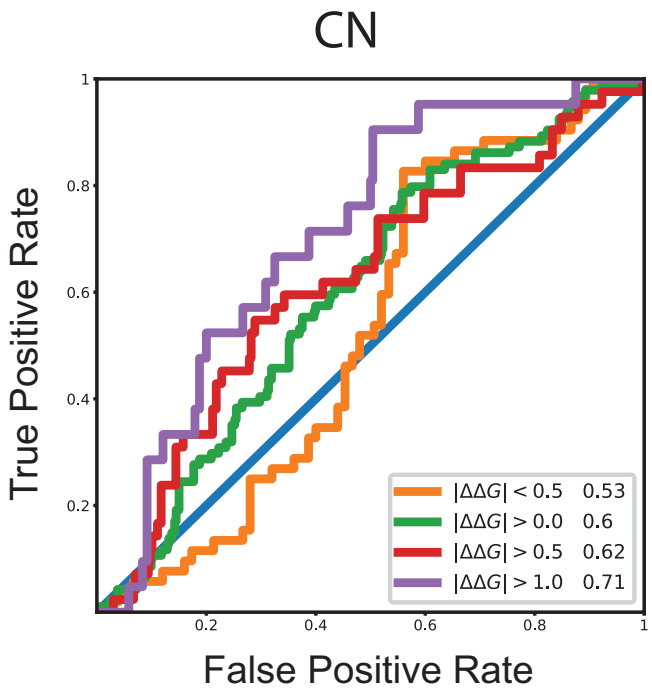
Suppose there are two mutations, Mut1, Mut2, in a antibody-antigen complex. (Mut1, Ag1) and (Mut2, Ag2) are two predicable pairs. (Wt1, Ag1) and (Wt2, Ag2) are corresponding wild-type pairs. Calculate the change of the returned values by our RBFN model:

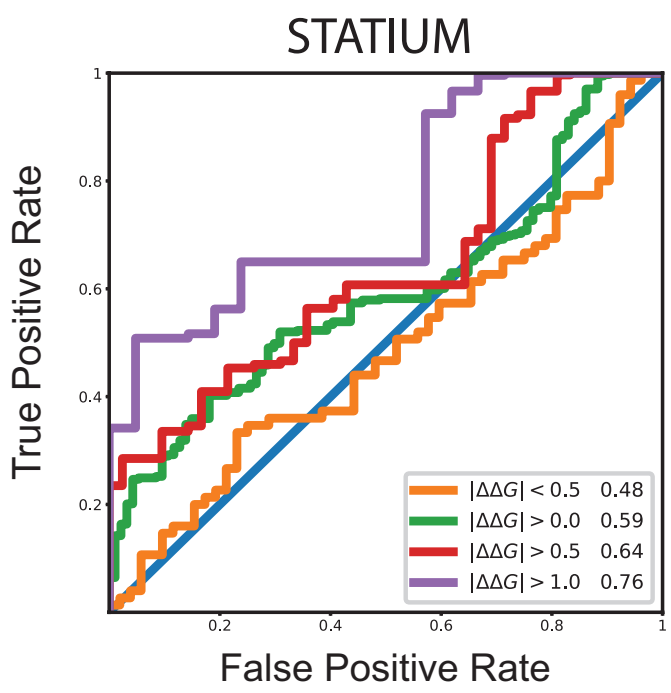
$$\Delta = \frac{1}{2} \sum_{i=1,2} (RBFN(Muti, Agi) - RBFN(Wti, Agi))$$

If $\Delta > 0$, the affinity increases. If $\Delta < 0$ the affinity decreases.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Results





Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity

Results

	$\Delta\Delta G < 0.5$	$\Delta\Delta G > 0$	$\Delta\Delta G > 0.5$	$\Delta\Delta G > 1$
CN	(0.43, 0.63)	(0.54, 0.66)	(0.53, 0.71)	(0.60, 0.80)
bASA	(0.44, 0.64)	(0.58, 0.69)	(0.57, 0.74)	(0.54, 0.80)
dfire	(0.45, 0.67)	(0.62, 0.73)	(0.63, 0.79)	(0.67, 0.84)
dDfire	(0.50, 0.71)	(0.57, 0.68)	(0.54, 0.70)	(0.57, 0.78)
Rosetta	(0.37, 0.68)	(0.57, 0.68)	(0.59, 0.76)	(0.67, 0.87)
STATIUM	(0.42, 0.63)	(0.57, 0.68)	(0.58, 0.75)	(0.68, 0.87)
D Studio	(0.48, 0.69)	(0.67, 0.77)	(0.70, 0.83)	(0.77, 0.92)
FoldX	(0.51, 0.71)	(0.69, 0.8)	(0.79, 0.91)	(0.86, 0.98)

Table: 95% confidence intervals constructed by Bootstrap. The iteration number is 10,000. D Studio is short for Discovery Studio.

Introduction
Significance
The Problem
The Challenge and the solution
The outline of our research
Data Extraction
Complexes, CDRs, Contact Number
Reduce Redundance
Positive Cores, Negative Cores
Develop an Radial Basis Function Network(RBFN)
Define the distance between two cores
Cross Validation and testing
Applications of RBFN
Antibody aa and antigen aa are different
RBFN can predict the affinity