



ИНСТИТУТ  
БИОИНФОРМАТИКИ



# Предсказывание силы связывания MHC I - пептид

**Проект выполнили:**

Мейстер Александр

Лайко Рудольф

Перцева Маргарита

Толстоухова Софья

Глазачев Владимир

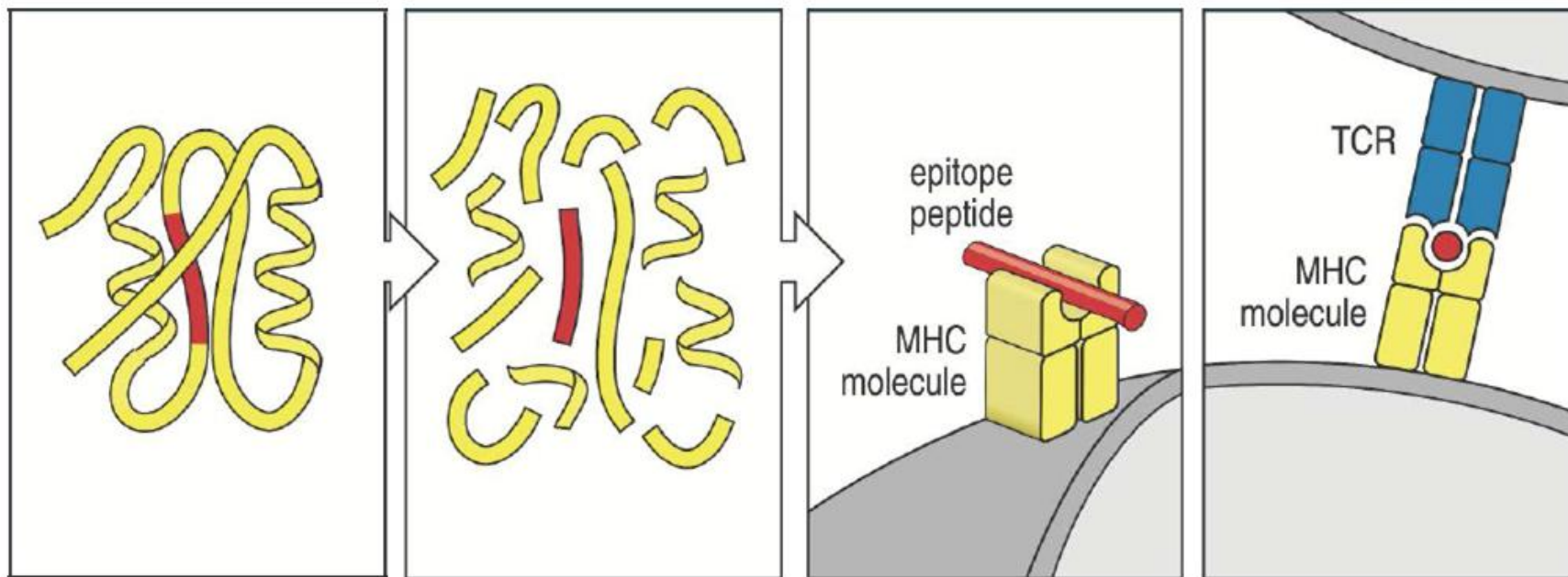
Мельников Эмиль

**Руководитель:**

Назаров Вадим

Июль 2016

# Биологическое обоснование



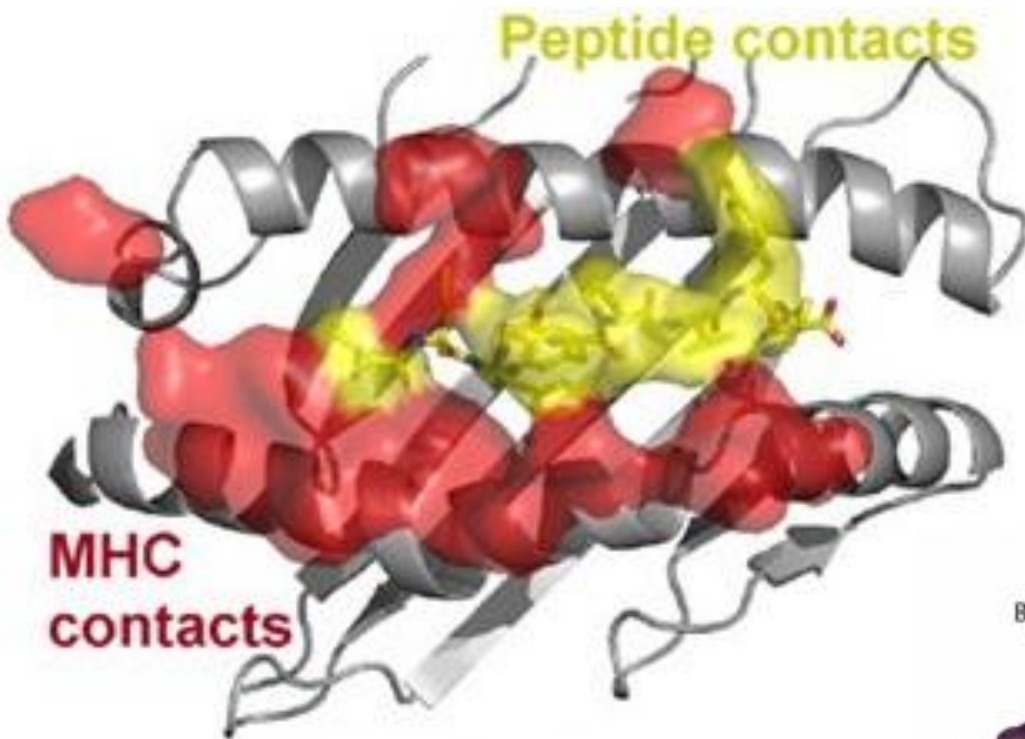
**Сила связывания  
MHC с пептидом –  
один из важных  
факторов  
иммуногенности !**

# Данные

## База данных: Immune Epitope Database and Analyse Resource (IEDB)

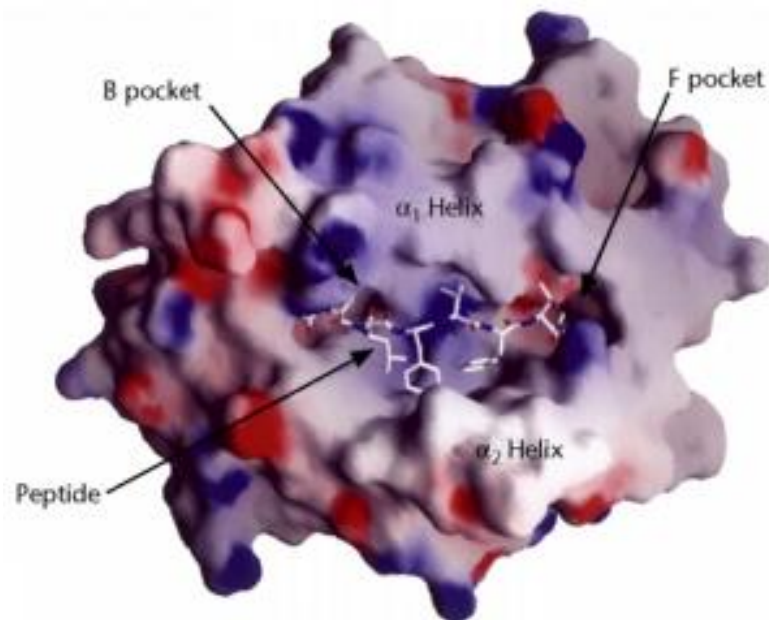
	species	mhc	peptide_length	cv	sequence	inequality	meas
179687	human	HLA-C*12:03	9	TBD	IMDEPTSSL	=	1.0
179688	human	HLA-C*12:03	9	TBD	DAMPGVLSY	=	1.0
179689	human	HLA-C*12:03	9	TBD	YYFQSGNEI	=	1.0
179690	human	HLA-C*12:03	9	TBD	LMDENTYAM	=	1.0
179691	human	HLA-C*12:03	9	TBD	FAAFYFVFI	=	1.0

# Подбор характеристик

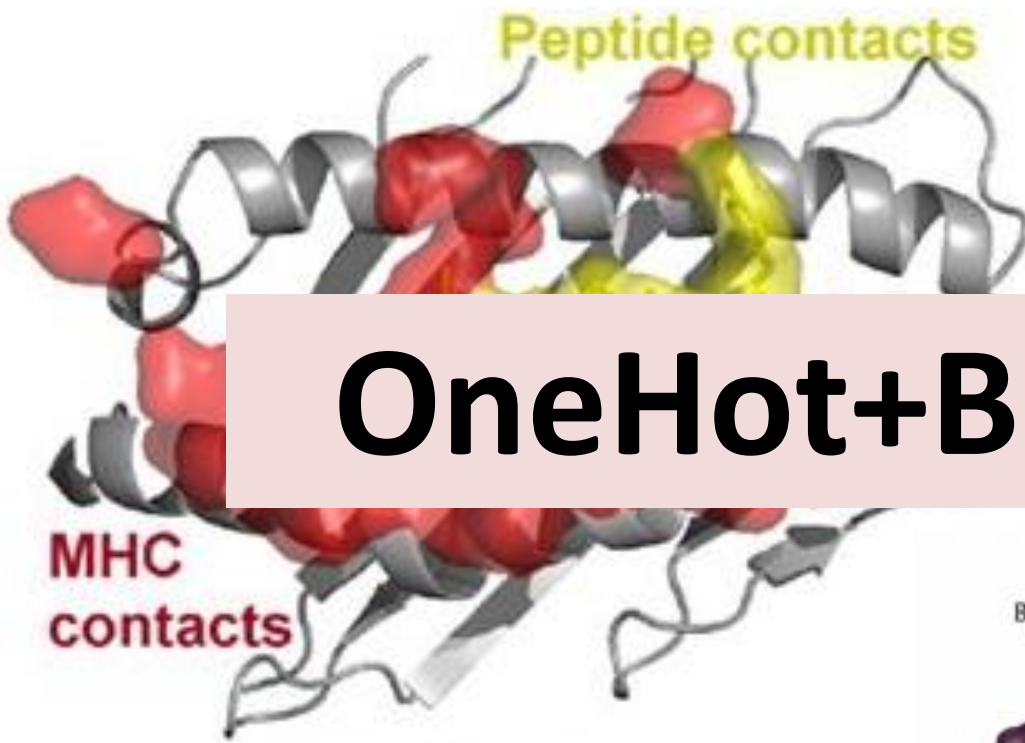


- Isoelectric point
- Molecular weight
- Word2Vec

- Kidera factors
- BLOSUM matrix
- Aromaticity



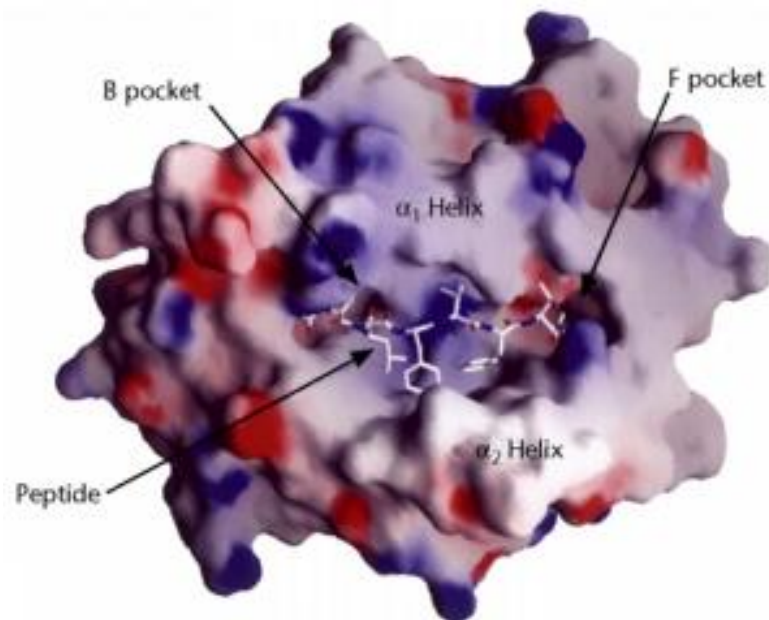
# Подбор характеристик



- Kidera factors
- BLOSUM matrix

## OneHot+BLOSSUM

- Isoelectric point
- Molecular weight
- Word2Vec





# Выбор модели



Naive Bayes



Linear Regression



SVM



Perceptron



Neural Networks



Decision Trees

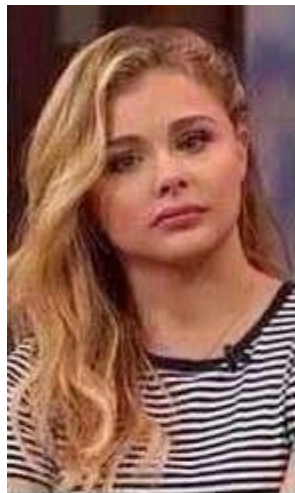


Random Forest



xgboost

# Выбор модели



Naive Bayes



Linear  
Regression



SVM



Perceptron



Neural  
Networks



Decision  
Trees



Random  
Forest



xgboost



# Выбор модели



Naive Bayes



Linear  
Regression



SVM



Perceptron



Neural  
Networks



Decision  
Trees



Random  
Forest



xgboost



# Выбор модели



Naive Bayes



Linear  
Regression



SVM



Perceptron



Neural  
Networks



Decision  
Trees

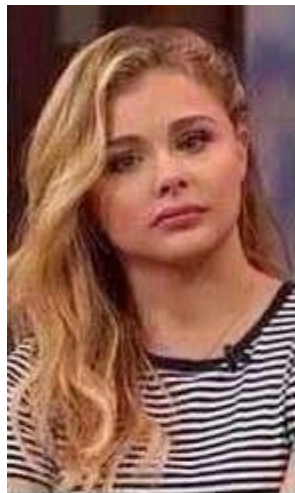


Random  
Forest



xgboost

# Выбор модели



Naive Bayes



Linear Regression



SVM



Perceptron



Neural Networks



Decision Trees



Random Forest



xgboost

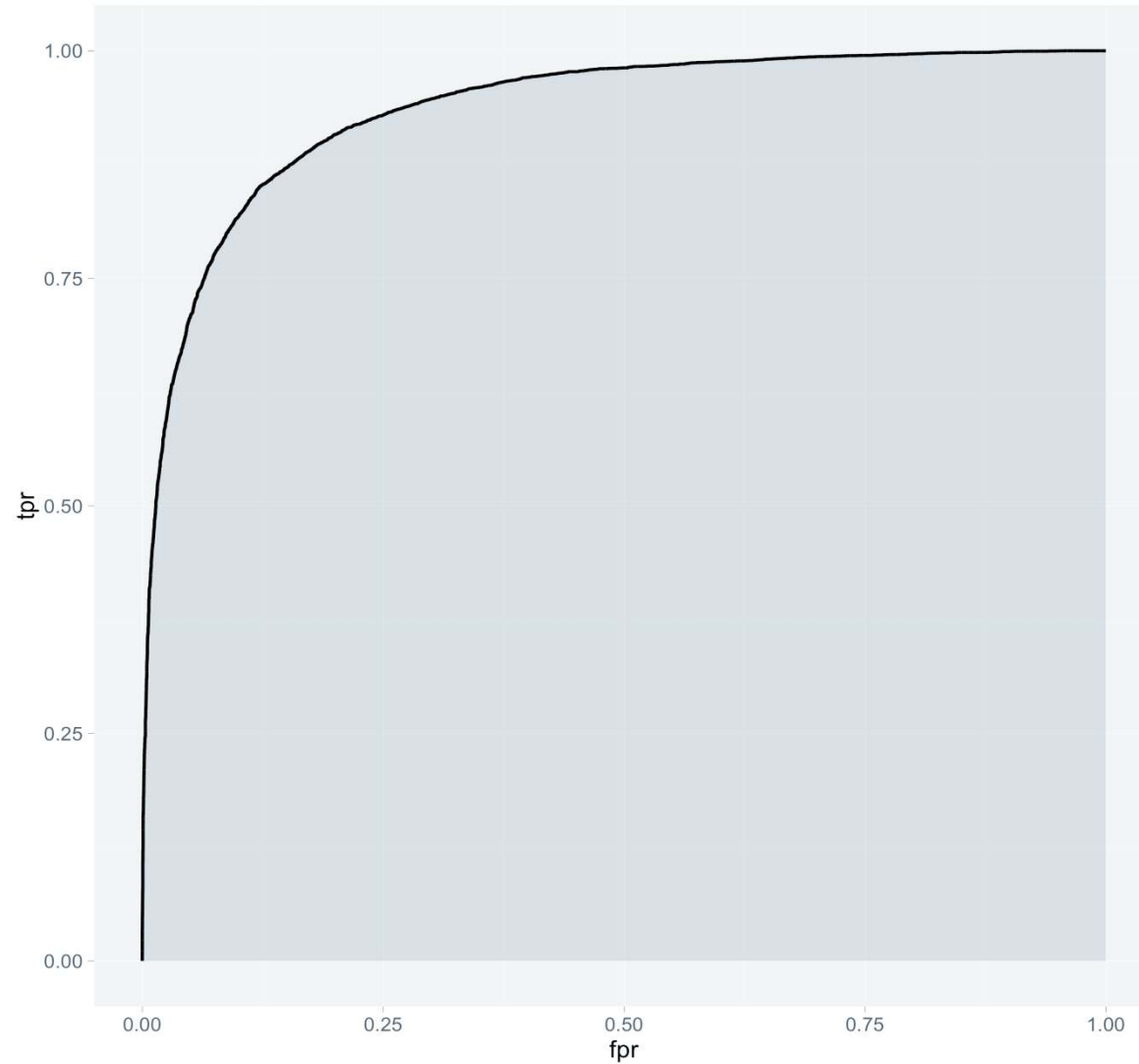






# Оценка эффективности

ROC Curve w/ AUC=0.936450612680921



Area under curve for  
gradient boosting  
machine = 0.93

# Результаты

- Были испытаны множество моделей (LR, SVM, NN, GBM)
- Наибольшую эффективность показала GBM
  - AUC на тесте – 0.93
  - AUC на валидации – 0.95
- Разработано Web – приложение для предсказания силы связывания МНС-пептид
- [https://github.com/VladimirGI/mhc\\_pred](https://github.com/VladimirGI/mhc_pred)