

**Critical Assessment of the Methods and the Features Used  
for Hot Spot Prediction at Protein-Protein Interfaces**

by

**Selin Karagülle**

**A Thesis Submitted to the  
Graduate School of Engineering  
in Partial Fulfillment of the Requirements for  
the Degree of**

**Master of Science**

**in**

**Computational Science and Engineering**

**Koc University**

**May 2014**

Koc University  
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Selin Karagülle

and have found that it is complete and satisfactory in all respects,  
and that any and all revisions required by the final  
examining committee have been made.

Committee Members:

---

Prof. Attila Gürsoy

---

Asst. Prof. Mehmet Sayar

---

Asst. Prof. Nathan Lack

Date:

---

## ABSTRACT

Hot spots are only a small subset of protein-protein interface residues but they account for the majority of the binding free energy. Measuring the binding free energy change upon mutating residues to alanine is an experimental way to determine hot spots. If mutation of a residue gives rise to a significantly large change in the binding free energy then this residue is defined as a hot spot. Experimental determination of hot spots is time-consuming, labor intensive and has high economic costs. Therefore, computational methods have been developed for hot spot prediction. These methods use training and testing data sets. However, there are no standard benchmark data sets for hot spot prediction. We present a new benchmark data set that is combination of 13 data sets and includes data of 1203 residues of 79 protein-protein complexes for computational hot spot prediction. The frequently used methods for hot spot prediction are machine-learning based and several features are combined in these methods. We reviewed literature, collected different features and critically assessed the effect of these features on results. As a result, seventy features that have strong effects are determined. The performances of different machine-learning methods, four servers and a plugin for a program using determined features on benchmark data set are compared. The results reveal that random forest classifier has the highest accuracy (80%) and although KFC2\_A has the highest F-measure (0.49) among existing methods, but it does not exceed the F-measure of naïve Bayes method (0.50). The benchmark data set, values of powerful features, and prediction results of four servers and a plugin can be viewed and downloaded via HotBase web interface located at <http://prism.cccb.ku.edu.tr/hotbase>

## ÖZET

Sıcak noktalar protein-protein ara yüzlerindeki aminoasitlerin sadece küçük bir alt kümesidir ama bağlanma serbest enerjisine büyük katkı sağlarlar. Sıcak noktalara deneysel olarak karar vermek için aminoasitlerin alanin aminoasidine mutasyonuna bağlı bağlanma serbest enerjilerinin değişimini ölçülür. Eğer aminoasidin mutasyonu bağlanma serbest enerjisinde çok büyük bir değişime yol açıyorsa bu aminoasit sıcak nokta olarak tanımlanır. Sıcak noktalara deneysel olarak karar vermek zaman alıcı, emek yoğunudur ve ekonomik maliyeti yüksektir. Bu sebeple, sıcak nokta tahmini için hesaplamalı yöntemler geliştirilmiştir. Bu yöntemler eğitim ve test setleri kullanır. Ancak, sıcak nokta tahmini için standart değerlendirme (benchmark) seti yoktur. Biz hesaplamalı sıcak nokta tahmini için 13 veri setinin birleşiminden oluşan ve 79 protein kompleksinin 1203 aminoasidi için verileri içeren yeni bir değerlendirme veri setini sunuyoruz. Makine öğrenme tabanlı metotlar sıcak nokta tahminleri için sıklıkla kullanılan yöntemlerdir ve bu yöntemlerde çeşitli özellikler birbirleriyle kombine edilirler. Biz literatürü taradık, değişik özellikler topladık ve bu özelliklerin sonuçlar üzerine etkisini eleştirel olarak değerlendirdik. Sonuç olarak güçlü etkisi olan yetmiş özellik tespit edildi. Belirlenen özellikler kullanılarak çeşitli makine öğrenme tabanlı metotların, sunucuların ve bir programın eklentisinin değerlendirme seti üzerindeki performansları kıyaslandı. Sonuçlara göre random forest sınıflayıcı en yüksek kesinliğe (%80) sahiptir ve KFC2\_A var olan diğer metotlar arasında en yüksek F-ölçü'süne (0.49) sahip olmasına rağmen naïve Bayes metodunun F-ölçü'sünü geçmez (0.50). Değerlendirme veri seti, güçlü özelliklerin değerleri ve dört sunucunun ve bir eklentinin tahmin sonuçları <http://prism.ccbb.ku.edu.tr/hotbase> adresinde yer alan HotBase internet ara yüzü aracılığıyla görülebilir ve indirilebilir.

## ACKNOWLEDGEMENTS

I would like to thank to my advisors **Prof. Attila Gürsoy** and **Prof. Özlem Keskin** for the chance of being a member of their research group and their support during my work. I am also thankful to my thesis committee members, **Asst. Prof. Mehmet Sayar** and **Asst. Prof. Nathan Lack** for their thoughtful critiques and final analysis.

I would like to acknowledge graduate fellowship provided by **Scientific and Technological Research Council of Turkey (TUBITAK)** as well.

Moreover, I would like to thank to all members of **COSBI (Computational Systems Biology)** Lab for good times during these 2.5 years.

I would like to give my deepest appreciation to my wonderful family, my father ***Mustafa Kemâl***, my mother ***Sevin***, my sisters ***Pelin*** and ***Rüya*** and my grandmother ***Hayriye***, for lifting me up when I was down, for settling me down when I was riled up, and holding steady beside me through all the rest.

## TABLE OF CONTENTS

<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>Nomenclature</b>	<b>x</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
<b>Chapter 2: Literature Review</b>	<b>3</b>
2.1 Protein-Protein Interfaces.....	3
2.2 Hot Spots .....	4
2.2.1 Characteristics of Hot Spots .....	5
2.2.2 Hot Spot Prediction .....	6
<b>Chapter 3: Methodology of Hotbase Creation</b>	<b>10</b>
3.1 Data Collection.....	11
3.2 Feature Determination.....	11
3.3 Model Evaluation .....	20
3.4 HotBase: Database for Hot Spot Prediction.....	21
<b>Chapter 4: Results and Discussion</b>	<b>24</b>
4.1 Benchmark Data Set.....	24
4.2 Features in HotBase .....	32
4.3 HotBase Website .....	34

4.4 Critical Assessment of the Results .....	38
<b>Chapter 5: Conclusion and Future Directions</b>	<b>45</b>
<b>Appendix A: Databases for Hot Spot Prediction</b>	<b>47</b>
A.1 Alanine Scanning Energetics Database (ASEdb).....	47
A.2 Binding Interface Database (BID).....	48
A.3 Structural Kinetic and Energetic Database of Mutant Protein Interactions (SKEMPI)..	48
A.4 Protherm .....	49
<b>Appendix B: ASEdb</b>	<b>51</b>
<b>Bibliography</b>	<b>63</b>
<b>Vita</b>	<b>69</b>

## LIST OF TABLES

<b>Table 3.1</b> Sources and number of residues that are taken from these sources. ....	11
<b>Table 3.2</b> Result of random forest model of Wang et al. ....	12
<b>Table 3.3</b> The impacts of features used in our study to result of Tuncbag et al.....	16
<b>Table 3.4</b> The effects of features to results of Chen et al.....	19
<b>Table 3.5</b> The results of sequence-based SVM on 10-fold cross validation using 2.0 kcal/mol training set.....	19
<b>Table 4.1</b> The 79 protein–protein complexes in HotBase. ....	26
<b>Table 4.2</b> List of features in HotBase.....	32
<b>Table 4.3</b> Results of Weka SVM classifier on 10-fold cross-validation. ....	39
<b>Table 4.4</b> Results of different classifiers on 10-fold cross-validation.....	40
<b>Table 4.5</b> The results of classifiers with four features. ....	40
<b>Table 4.6</b> Results of different methods. ....	41
<b>Table 4.7</b> Hot spot prediction performances on independent test set in two studies. ....	42
<b>Table 4.8</b> Cross validation results of different methods.....	43
<b>Table 4.9</b> Results of different methods on training and testing sets .....	44
<b>Table B.1</b> Data of ASEdb Database .....	51



## LIST OF FIGURES

<b>Figure 2.1</b>	Representation of the interface region and hot spot residues of Barnase-Barstar complex. In the left of the figure, dark orange area is the interface region of two chains. In the right part, the ball shaped atoms are atoms of hot spot residues. Purple ones are belong to barnase and blue ones are belong to barstar. ....	4
<b>Figure 3.1</b>	Flowchart of the methodology.....	10
<b>Figure 3.2</b>	Header row of “Complexes” table with explanations of the columns in bubbles. .	22
<b>Figure 3.3</b>	Header row of “Data Set” table with explanations of the columns in bubbles.	22
<b>Figure 3.4</b>	Header row of “Server Results” table with explanations in bubbles.....	23
<b>Figure 4.1</b>	The number and percentage of residues in the data set. ....	25
<b>Figure 4.2</b>	Distribution of hot spot residues.....	32
<b>Figure 4.3</b>	Homepage of Hotbase .....	35
<b>Figure 4.4</b>	"Complexes" webpage.....	35
<b>Figure 4.5</b>	"Data Set" webpage .....	36
<b>Figure 4.6</b>	"Server Results" webpage .....	36
<b>Figure 4.7</b>	"Feature Values" webpage.....	37
<b>Figure 4.8</b>	"ASEdb" webpage .....	37
<b>Figure 4.9</b>	"About Us" webpage .....	38
<b>Figure A.1</b>	The main page of ASEdb database.....	47
<b>Figure A.2</b>	The main page of WikiBID. ....	48
<b>Figure A.3</b>	The main page of SKEMPI database.....	49
<b>Figure A.4</b>	The main page of ProTherm .....	50

## NOMENCLATURE

<i>ASEdb</i>	Alanine Scanning Energetics Database
<i>ASA</i>	Accessible Surface Area
<i>BID</i>	Binding Interface Database
<i>CSU</i>	Contacts of Structural Units
<i>HSSP</i>	Homology driven Structures of Proteins (Database)
<i>KFC2</i>	Knowledge-based FADE and Contacts 2
<i>MCC</i>	Matthew's Correlation Coefficient
<i>PDB</i>	Protein Data Bank
<i>PSAIA</i>	Protein Structure and Interaction Analyzer
<i>PSSM</i>	Position-Specific Scoring Matrix
<i>SVM</i>	Support Vector Machine
<i>YASARA</i>	Yet Another Scientific Artificial Reality Application

## Chapter 1

### INTRODUCTION

Proteins play a central role in all essential molecular and cellular processes. In order to fulfill a function in the cell, they interact with each other through their interaction sites and trigger a series of reactions. Certain residues residing on the protein interaction or binding sites make dominant contributions to the free energy of binding.

This thesis is mainly motivated by several well-known facts. Experimental information about these hot spot residues is limited and fragmented. Machine-learning models have been processed by using different training sets and tested by making predictions against different test sets. Thus, comparing results of these models is not completely fair.

Two aims of this thesis are to provide a standard benchmark data set which combines multiple data sets and to present values of features that have strong effect on predictions of machine-learning models. Using the same features on standard benchmark data set, we will bring comparable results and then better insight to protein-protein interactions.

The third and most important purpose is to introduce unique source for hot spot prediction. To the best of our knowledge, HotBase is the first database which combines various resources (data sets, features, results of different methods) to predict hot spots together.

The remaining part of thesis is organized under three main chapters:

In chapter 2, an extended and most recent review focusing on the aspects of hot spot prediction is presented. This chapter includes background information on protein interfaces and machine-learning algorithms, as well as several studies focusing on the characterization of protein interfaces and identification of hot spots.

Chapter 3 introduces methods on new benchmark data set formation, determination of powerful features in recent studies and explanations of tables in Hotbase. First, benchmark data set will be formed by combining thirteen sources and removing redundant data (e.g. data of double or multiple mutated residues) from the data set. Then, the features are collected from three studies and after critical assessment, seventy features that have strong effects on results will be defined. At the end of this chapter, the tables of Hotbase are explained. The database includes tables of 79 protein complexes, data set with hyperlinks to articles that have related data in their data set, results of different methods and values of seventy features.

Chapter 4 presents benchmark data set, seventy features and web interface of HotBase and discusses obtained results from machine-learning based methods of Weka and different methods.

This thesis ends with a chapter explaining future directions, and finally with presenting major conclusions of the study.

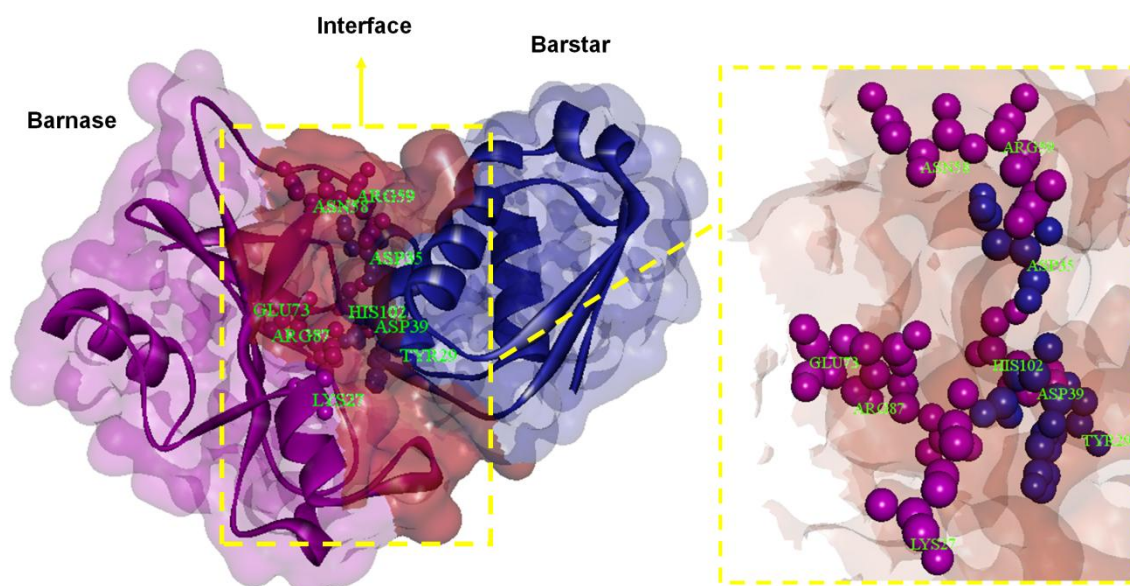
## Chapter 2

### LITERATURE REVIEW

#### 2.1 Protein-Protein Interfaces

Proteins cooperate with other proteins to perform biological function. They physically interact with their partners through their interfaces by weak non-covalent bonds. The particular region where two protein chains come into contact is termed a binding region or an interface. As an example, Figure 2.1 shows the interface region between barnase and barstar (Pdb ID: 1brs). To discriminate the interface from the rest of the surface, a myriad of studies have analyzed some characteristics of protein-protein interfaces like hydrophobicity, solvent accessibility, shape and electrostatic complementarity, evolutionarily conservation, flexibility, residue propensities and hydrogen bonding [1–5].

From these studies it has been found that interface residues are more conserved and less flexible than the rest of the surface [2, 6, 7]. The protein-protein interfaces have more hydrophobic and aromatic residues, less hydrophilic residues. Furthermore, hydrophobic and aromatic interface residues have high propensities [2]. On the other hand, Glaser et al. stated that polar residues are rich in small interfaces, while hydrophobic residues occurred more often in large interfaces [8].



**Figure 2.1** Representation of the interface region and hot spot residues of Barnase-Barstar complex. In the left of the figure, dark orange area is the interface region of two chains. In the right part, the ball shaped atoms are atoms of hot spot residues. Purple ones are belong to barnase and blue ones are belong to barstar.

## 2.2 Hot Spots

Studies on protein interfaces have revealed that the free energy contributions of interfacial residues to binding are not uniformly distributed. A small subset of interfacial residues named hot spots account for the majority of the binding free energy [9, 10]. In the right part of Figure 2.1, the hot spots of barnase-barstar complex are illustrated. Alanine Scanning Mutagenesis is an experimental method which is based on the fact that when a residue is mutated to alanine if it causes a significant drop in the binding free energy ( $\Delta\Delta G$ ) then it is a hot spot. Unfortunately, experimental determination of hot spots is time-

consuming, labor-intensive and has high economic costs. Therefore, there is a need for developing computational methods to identify hot spots [11].

### 2.2.1 Characteristics of Hot Spots

Bogan and Thorn reported that hot spots are abundant in tryptophan (trp), arginine (arg) and tyrosine (tyr). On the other hand, some residues like leucine (leu), methionine (met), serine (ser), threonine (thr) and valine (val) are disfavored [9]. The hot spots are surrounded by a ring of residues that are energetically less important and occlude bulk solvent from the hot spots (O-ring theory) [12]. “Double water exclusion” theory refines the “O-ring theory” and reveals that hot spots at protein-protein interfaces are water-free [13].

Keskin et al. showed that computational hot spots are not homogeneously distributed along the protein interfaces; rather they are clustered within locally tightly packed regions, called “*hot regions*” [14] and Cho et al. reached similar conclusion by showing significantly higher atomic packing density values of hot spots [15]. Del Sol and O’Meara and Keskin et al. illustrated that there is a correlation between hot spot residues and structurally conserved residues so the conserved and buried residues, are either experimental hotspot or in direct contact with an experimental hotspot [14, 16].

Kozakov et al. have recently demonstrated that hot spots are clustered in specific regions that are distinguishable due to their concave topology and those regions are patterned with hydrophobic and polar residues [17].

Understanding microenvironment of hot spots is also significant. Ye et al. demonstrated that alanine (ala), aspartic acid (asp), glycine (gly), histidine (his), isoleucine (ile), asparagine (asn), serine (ser) and tyrosine (tyr) are more likely to occur close by hot spots than non-hot spots [18].

To be a hot spot, residue must cause significant change in the binding free energy but there is no consensus in the literature for what value of change is significant. In recent studies on hot spot prediction, 1.0 kcal/mol ([19, 20]) and mostly 2.0 kcal/mol ([18, 21–28]) are used as the threshold to differentiate hot spots and non-hot spots, whereas Ofran and Rost chose 2.5 kcal/mol threshold to define hot spots and 0 kcal/mol to define non-hot spots [29]. Tuncbag et al. defined hot spots to be interface residues with higher than 2.0 kcal/mol and non-hot spots to be the ones with lower than 0.4 kcal/mol. Chen et al. and Xia et al. also used same definition of Tuncbag et al. [30, 31] while Cho et al. used two different cutoff values for the definition in their study, that is, 1.0 and 2.0 kcal/mol [15]. In most of the studies that have independent test set derived from BID, only ‘strong’ mutations are considered as hot spots [15, 18, 21, 22, 26, 27, 30], except Chen et al who consider both strong and intermediate mutations as hot spots [20].

### **2.2.2 Hot Spot Prediction**

Several computational methods have been developed to predict protein–protein interface hot spots. Machine-learning based methods have been preferred for hot spot prediction in recent years. Tom M. Mitchell provided a formal definition of machine learning: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” [32].

There are several types of machine-learning algorithms including supervised learning algorithm. Supervised learning algorithm takes a known set of data and known responses to the data (training set) and creates a model. The trained model can then generate predictions for the new data (testing set). Support vector machines (SVMs), J48 decision tree, random forest and naïve Bayes are commonly used supervised learning methods for hot spot prediction. SVMs are employed to find hyperplane that separates data instances of one kind



from those of another with maximum separation margin. The J48 decision tree classifier follows the following simple algorithm to classify a new item: In order to classify the data instances most clearly, it chooses the attribute that gives the highest information gain. It creates a tree node whose value is the chosen attribute and among the possible values of this attribute, if there is any value that unambiguously subdivides the instances into subclasses then it terminates that branch and assigns to it the target value. For the other cases, it searches another attribute that gives the highest information gain and continues in this manner until getting a clear decision of what combination of attributes gives a particular target value or running out of attributes. Random forests are ensemble learning methods that use many randomly generated decision trees for classification. To classify new data, each tree gives a classification and the classification having the most votes is chosen. Naïve Bayes is a probabilistic classifier based on applying Bayes theorem that every feature, regardless of the presence or absence of the other features, is independent given class.

Lise et al. combined the strengths of machine-learning and energy-based methods by considering the basic energy terms as input features of machine-learning models such as SVMs and Gaussian processes [37]. In their recent study, they integrated their approach with two additional SVM classifiers to overcome limitations on predictions involving arginine (arg) or glutamic acid (glu) residues [24]. Cho et al. proposed a hot spot prediction method based on protein structure, sequence and molecular interaction that used decision trees for feature selection and SVMs for classification [15]. Xia et al. also employed SVM classifiers with features such as protrusion index and solvent accessibility to predict hot spots [30]. Assi et al. applied Bayesian networks with energetic, structure-based and sequence-based features to predict hot spots [25]. Chen et al. generated sequence-based SVM model that utilized physicochemical features, Position Specific Scoring Matrix (PSSM), evolutionary conservation score, and sequence entropy [23]. Zhu and Mitchell

built two knowledge-based hot spot prediction methods (KFC2a and KFC2b) based on different feature combinations using SVMs [26]. Wang et al. proposed a random forest model that took hybrid features of residues that were defined with residues itself and its interacting residues of the opposite chain [21]. Chen et al. predict hot spots using IBk algorithm, a variant of the K-nearest neighbour classifier, as a classifier with only physicochemical characteristics of residues [31]. Furthermore, Ye et al. constructed the SVM model using network features and microenvironment features [18].

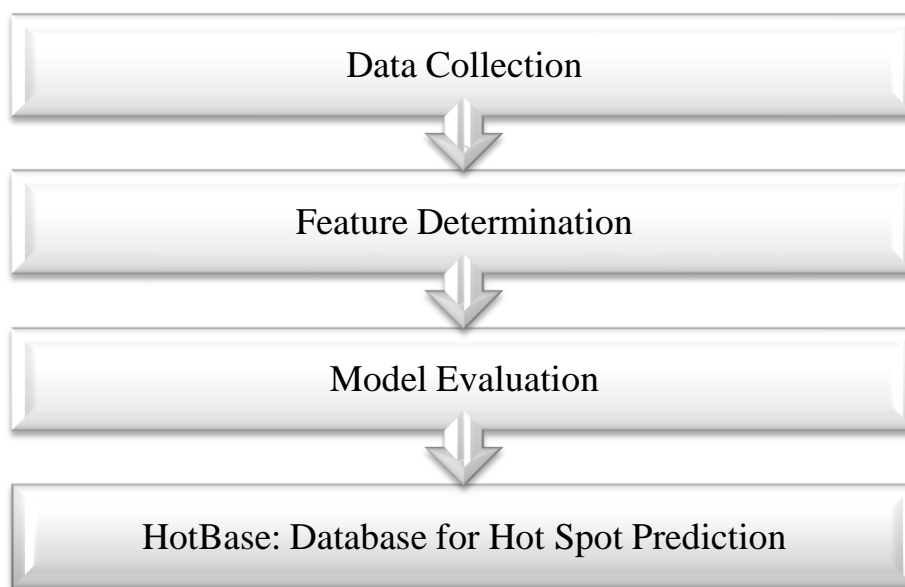
Besides machine-learning based methods, there are some other methods that are developed for hot spot prediction. Energy-based methods such as Robetta [34] and FOLDEF function [35], methods that are based on molecular dynamics simulations [36, 37] are developed and used for prediction, likewise the energy fluctuation model proposed by Erman predicted the locations of hot spots [38]. Ofran and Rost developed a knowledge-based method (ISIS) that was based on neural networks with features extracted from sequence environment and evolutionary profile and used ISIS to predict hot spots [29, 39]. Darnell et al. combined decision tree approach based on atomic contacts, physicochemical properties, and shape specificity with computational alanine scanning [22]. Grosdidier and Fernandez-Recio developed a method that predicted interface hot spots using protein-docking tools without protein complex knowledge [40]. Feature-based method of Guney et al. identified hot spots using solvent accessible surface areas, residue conservation and residue propensity [41] and Tuncbag et al. presented an empirical feature-based method by combining solvent accessibility and statistical pairwise residue potentials [42, 43]. BeatMusic predictor based on statistical potentials predicts the change in binding free energy by point mutations [44]. Haliloglu et al. applied Gaussian Network Model (GNM) to predict anchoring residues in the unbound molecules [45]. Furthermore, Ozbek et al. applied GNM both on unbound and complex structures to predict hot spots [46]. Wang et al. presented structure-based computational approach that determines hot spots through the

docking of a compound known as the inhibitor of the specific Protein-Protein Interaction (PPI) [47].

### Chapter 3

#### METHODOLOGY OF HOTBASE CREATION

In this chapter, a new database for computationally prediction of hot spots with machine-learning methods is presented. Figure 3.1 illustrates the flowchart of our methodology. Firstly, we collected data from literature and removed the redundancy in the data set. We determined the features that have strong effect on hot spot prediction in machine-learning studies. We evaluate power of our features in terms of precision, recall, specificity, accuracy and F-measure. Finally, we built our HotBase website with benchmark data set, powerful features and various server results.



**Figure 3.1** Flowchart of the methodology.

### 3.1 Data Collection

Residues at protein-protein interfaces that have experimentally determined binding free energy changes and interaction strengths are obtained from the literature. Thirteen different sources for data are listed in Table 3.1. The interface residues that do not have three-dimensional structure and the residues from single chain proteins are eliminated from the data set. In addition, there have been double or multiple mutated residues, non-interface residues and residues which were not mutated to alanine were removed as well. The final set of residues is called “benchmark data set”.

**Table 3.1** Sources and number of residues that are taken from these sources.

Source	Number of residues
Chen et al. [20]	522
Thorn and Bogan (ASEdb) [12]	476
Wang et al. [21]	436
Cho et al. [15]	390
Kortemme and Baker [48]	362
Lise et al. [33]	349
Darnell et al. [22]	349
Zhu and Mitchell [26]	252
Tuncbag et al. [42]	261
Fischer et al. (BID) [49]	233
Xia et al. [30]	153
Moal and Fernández-Recio [50]	41
Kumar et al. [51]	90

### 3.2 Feature Determination

In many studies on predictions of hot spots, different features have been combined to improve the performance of their models. In our study, we determined features that have strong effect on hot spot prediction using machine-learning models in previous studies that belong to Wang et al. [21], Tuncbag et al. [42] and Lise et al. [20]. Before calculating values of features in the study of Wang et al. [21], we firstly defined target residue, the

intra-contact residue (the nearest contact residue of the target residue in the same face of this interface) and the mirror-contact residue (the nearest contact residue of the target residue in the other face).

### A. The study of Wang et al.

In their model, they achieved the highest accuracy in comparison with previous studies. Table 3.2 shows their prediction results of 10-fold cross validation and in the independent test set. Considering their success, we decided to combine their descriptors with other features in our study.

**Table 3.2** Result of random forest model of Wang et al.

Testing	PR (%)	SE (%)	SP (%)	ACC (%)	MCC
10-fold	68.4	50.6	92.5	82.4	0.48
Test	70.8	44.7	92.0	77.6	0.43

MCC denotes Matthew's correlation coefficient.

#### A.1. Relative ASA (rel\_ASA) of i-th residue

Relative ASA (rel\_ASA) of i-th residue was calculated by PSAIA program [52]. The program calculated ASA using the 'rolling ball' algorithm which uses a sphere of 1.4 Å radius to probe the surface of the molecule. The formula of relative ASA (RASA) in this program is the ratio between the calculated ASA of i-th residue and standard ASA for a particular residue.

#### A.2. Relative Side-Chain ASA (rel\_s\_ch\_ASA)

Relative Side-Chain ASA was also calculated by PSAIA program [52] as rel\_ASA.

#### A.3. Atom contacts of a Residue (atom\_contacts)

The contact between two atoms were determined using Contacts of Structural Units (CSU) program and atom contacts of a residue were computed by summing these atom contacts between the residue and any other residue in the protein-protein interface:

$$\text{atom\_contacts}(i) = \sum_j \left\{ \sum_{\alpha \in i} \sum_{\beta \in j} \text{atom\_contact}(\alpha, \beta) \right\} \quad (3.1)$$

where  $\text{atom\_contact}(\alpha, \beta)$  is equal to 1 if atom  $\alpha$  is interacting with atom  $\beta$ , otherwise, it is equal to 0.

#### A.4. Atom Contact Areas of a Residue (**atom\_contact\_areas**)

Atom contact areas of a residue were calculated to represent the interaction area between the target residue and the other face of the interface. Atom contact areas of residue  $i$  are defined as sum of all atom contact areas between the residue  $i$  and any other residue  $j$  located at the other face:

$$\text{atom\_contact\_areas}(i) = \sum_{j \in \text{the other face}} \left\{ \sum_{\alpha \in i} \sum_{\beta \in j} \text{atom\_contact\_area}(\alpha, \beta) \right\} \quad (3.2)$$

where  $\text{atom\_contact\_area}(\alpha, \beta)$  is the contact area between two interacting atoms  $\alpha$  and  $\beta$ .

#### A.5. Residue Contacts of a Residue

If there is one pair of contact atoms from two residues individually then the residues have contact information [21]. Residue contacts of a residue were the sum of these residue contacts between the residue and other residues in the interface:

$$\text{residue\_contacts}(i) = \sum_j \text{residue\_contact}(i, j) \quad (3.3)$$

where  $\text{residue\_contact}(i, j)$  is equal to 1 if residue  $i$  interacts with residue  $j$ , otherwise, it is equal to 0.

## Depth Index

The depth of an atom refers to the distance from its closest solvent accessible atom. PSAIA program [52] provides several depth indices. The depth indices that are combined with other features in our study:

### A.6. Average Depth Index (ave\_dpx)

Average depth index value of all atom values.

### A.7. Standard Deviation of Depth Index (sd\_dpx)

### A.8. Side-Chain Average Depth Index (s\_ch\_ave\_dpx)

Average value of all side-chain atom values.

### A.9. Standard Deviation of Side-Chain Depth Index (sd\_s\_ch\_dpx)

### A.10. Relative Depth Index (rel\_dpx) of the ith Residue upon Complex Formation

The formula of relative depth index:

$$\text{rel\_dpx}(i) = \frac{\text{ave\_dpx}_{\text{comp}}(i) - \text{ave\_dpx}_{\text{mono}}(i)}{\text{ave\_dpx}_{\text{comp}}(i)} \quad (3.4)$$

### A.11. Relative Side-Chain Depth Index (rel\_s\_ch\_dpx) of the ith Residue upon Complex Formation

The formula of relative side-chain depth index:

$$\text{rel\_s\_ch\_dpx}(i) = \frac{\text{s\_ch\_ave\_dpx}_{\text{comp}}(i) - \text{s\_ch\_ave\_dpx}_{\text{mono}}(i)}{\text{s\_ch\_ave\_dpx}_{\text{comp}}(i)} \quad (3.5)$$

### A.12. The Class of Residues

We clustered 20 amino acids into six classes based on their dipoles and volumes of the side chains: Class 1: A, G, V; Class 2: I, L, F, P; Class 3: Y, M, T, S, C; Class 4: H, N, Q, W; Class 5: R, K and Class 6: D, E.



### **A.13. Secondary Structure**

To identify hot spots, we used three secondary structure types, i.e. helix, strand and loop. Secondary structures obtained by using dictionary of protein secondary structure algorithm [53] were used.

### **Physicochemical Features**

In our study, there are two physicochemical feature groups which are from the study of Wang et al. and from the study of Chen et al. respectively. All values of features are taken from AAindex database [54]. Hydrophobicity, hydrophilicity, polarity, and propensities are in common for two groups but the values of features are different and Wang et al. made extra calculations on the values while Chen et al. did not. In group one; there are six physicochemical features:

#### **A.14. Hydrophobicity**

Hydrophobicity is a property of an amino acid that is repelled by water molecule.

#### **A.15. Hydrophilicity**

Hydrophilicity means to have affinity for water

#### **A.16. Isoelectric Point**

Isoelectric point is the pH at which an amino acid carries the least electrical charge.

#### **A.17. Mass**

#### **A.18. Polarity**

Polarity is the difference in charge between atoms in an amino acid.

#### **A.19. Polarizability**

Polarizability is the ease of distortion of the electron cloud of an amino acid by electric field.

We defined these features of a residue by summing its and its interacting residues' values according to provided method by Wang *et al.* The second group of physicochemical features will be given in Chen *et al.* section.

### Evolutionary conservation score

Conservation score is commonly used feature in the study of hot spot prediction using machine-learning methods. We obtained evolutionary conservation scores from ConSurf Server Database [55]. ConSurf Server Database calculates scores using Rate4Site algorithm. This algorithm considers phylogenetic relations between homologues. The color scale, output of this algorithm, represents the conservation score (e.g. 9: Conserved, 1: Variable) [56].

According to recent studies, single conservation score by itself is not a good feature discriminating hot spots from non-hot spots [6, 57].

#### B. The Study of Tuncbag et al.

Prediction results of the study of Tuncbag et al. [42] showed that combining conservation score with other features was reasonable. The impact of single conservation score to the results did not exceed the impact of combinations with one or two different features in terms of accuracy and F-measure (F1) (Table 3.3).

The contribution of pairwise potential to the results of prediction was generally appreciated when both two-feature (relCompasa+Score) and three-feature (relCompasa+(Score+PP)) cases were observed so in our study, we combined that feature with other features.

**Table 3.3** The impacts of features used in our study to result of Tuncbag et al.

	Model	Data Set	P	R	S	A	F1
Single feature performances on ASEdb and BID data sets (Empirical formulas)	Score $\geq 7.0$	Training set	0.50	0.33	0.79	0.61	0.40
		Test set	0.52	0.46	0.60	0.54	0.49
	relCompASA $\leq 20.0$	Training set	0.55	0.81	0.58	0.67	0.65
		Test set	0.60	0.67	0.59	0.63	0.63
	PP $\geq 18.0$	Training set	0.56	0.55	0.73	0.66	0.56
		Test set	0.69	0.70	0.71	0.71	0.70
Two features performances on ASEdb and BID data sets	relCompASA $\leq 20.0$ and Score $\geq 7.0$	Training set	0.61	0.29	0.88	0.65	0.40
		Test set	0.71	0.32	0.88	0.61	0.44

		PP $\geq$ 18.0 and Score $\geq$ 7.0	Training set	0.57	0.14	0.94	0.63	0.22
			Test set	0.75	0.33	0.90	0.63	0.46
		relCompASA $\leq$ 20.0 and PP $\geq$ 18.0	Training set	0.64	0.52	0.82	0.70	0.57
			Test set	0.73	0.59	0.79	0.70	0.65
Multiple features performances on ASEdb and BID data sets		relCompASA $\leq$ 20.0 and (Score $\geq$ 7.0 or PP $\geq$ 18.0)	Training set	0.63	0.67	0.75	0.72	0.65
			Test set	0.67	0.63	0.71	0.67	0.65

Score denotes conservation score, relcompasa denotes relative ASA in complex state and PP denotes pairwise potential

### B.1. Knowledge-Based Solvent Mediated Inter-Residue Potentials

Knowledge-based solvent mediated inter-residue potentials are equations related to known three-dimensional structures. In a study which set out to determine hot spot, Tuncbag *et al.* [42] utilized those contact potentials which were taken from Keskin *et al.* [58]. Contact potential between residues  $i$  and  $j$  formulated as;

$$\text{Pair}(i, j) = \begin{cases} \text{contact potential of type } (i, j) & \text{if } d(i, j) \leq 7.0 \text{ and } |i - j| \geq 4 \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

where  $\text{Pair}(i, j)$  is the contact potential of residues  $i$  and  $j$  and  $d(i, j)$  is the distance between two residue centers.

Overall contact potential of residue  $i$  was defined as the absolute of sum of its pair potentials:

$$PP_i = \text{abs} \left( \sum_{j=1}^n \text{Pair}(i, j) \right) \quad \text{for } |i - j| \geq 4 \quad (3.7)$$

### **B.2. The Relative ASA of i-th Residue in Complex State**

The ASA of each residue in monomer state and in complex state was calculated using NACCESS and then  $\text{relCompASA}_i$  is formulated as follows:

$$\text{relCompASA}_i = \frac{\text{ASA in Complex}_i}{\text{maxASA}_i} \times 100 \quad (3.8)$$

where ‘ $\text{relCompASA}_i$ ’ is the relative ASA of i-th residue in complex state.

### **B.3. The Relative Difference of ASA Between Monomer and Complex State of i-th Residue**

The ASA of each residue in monomer state and in complex state were calculated using NACCESS. The formula of  $\text{rel}\Delta\text{ASA}_i$ :

$$\text{rel}\Delta\text{ASA}_i = \frac{[\text{ASA in Monomer}_i] - [\text{ASA in Complex}_i]}{\text{maxASA}_i} \times 100 \quad (3.9)$$

where ‘ $\text{rel}\Delta\text{ASA}_i$ ’ is the relative difference of ASA between monomer and complex state of i-th residue.

### **C. The Study of Chen et al.**

Chen et al. also used conservation score by combining it with other sequence-based features. Their sequence-based feature set comprised physicochemical features, evolutionary conservation score, sequence entropy, and position-specific scoring matrix. In Table 3.4, prediction using sequence-based features has slightly good results to the prediction using the second feature set whereas combining two feature sets did not have significant effect to the results.

**Table 3.4** The effects of features to results of Chen et al.

Features	Testing	P	R	F1	AUC
Phy	10-fold	0.63	0.51	0.55	0.67
	Test	0.59	0.53	0.56	0.58
ECS	10-fold	0.58	0.27	0.34	0.51
	Test	0.74	0.18	0.28	0.68
SE	10-fold	0.57	0.53	0.54	0.60
	Test	0.59	0.61	0.60	0.52
PSSM	10-fold	0.65	0.54	0.58	0.65
	Test	0.64	0.48	0.55	0.66
Phy+PSSM+ECS+SE	10-fold	0.65	0.65	0.65	0.68
	Test	0.69	0.68	0.68	0.68
Phy+ASA+BC	10-fold	0.65	0.60	0.61	0.70
	Test	0.62	0.70	0.66	0.62
Phy+ASA+BC+PSSM+ECS+SE	10-fold	0.66	0.68	0.66	0.72
	Test	0.65	0.63	0.64	0.66

ASA denotes accessible surface area; BC denotes biochemical contacts; Phy means physicochemical features; ECS denotes evolutionary conservation score; SE means sequence entropy; and PSSM is the abbreviation of Position-Specific Scoring Matrix.

On the other hand, the results of sequence-based SVM on 10-fold cross validation using 2.0 kcal/mol training set was remarkably high so we added only sequence-based features into our feature set (Table 3.5).

**Table 3.5** The results of sequence-based SVM on 10-fold cross validation using 2.0 kcal/mol training set.

Method	Testing	P	R	F1
Sequence-based SVM	10-fold cross validation	0.78	0.80	0.79

### Physicochemical Features

The physicochemical features in the study of Chen *et al.* [20] are formed the second group of physicochemical features. There are six physicochemical features:

#### C.1. Hydrophobicity

**C.2. Hydrophilicity****C.3. Polarity****C.4. Polarizability****C.5. Propensities**

Different amino acids have different propensities for forming secondary structures.

**C.6. Average Accessible Surface Area****C.7. Sequence Entropy**

Sequence entropy in HSSP (homology-derived structures of proteins) database [59] is a measure of sequence variability at a given sequence position. Sequence entropy values are normalized over the range 0-100 and the lower values indicate the more conserved positions.

**C.8. Sequence Profile**

Sequence profile is a Position-Specific Scoring Matrix (PSSM) which is taken from multiple sequence alignment and is obtained by PSI-BLAST [60] searching against NCBI non-redundant database. In this derivation process, the parameters are the BLOSUM62 substitution matrix and E-value threshold of 0.001. The PSSM scores are normalized in the range 0.0-1.0:

$$f(x) = \begin{cases} 0.0 & \text{if } x < -5 \\ 0.5 + 0.1x & \text{if } -5 \leq x < 5 \\ 1.0 & \text{if } x \geq 5 \end{cases} \quad (3.1)$$

where  $x$  is an element of PSSM.

**3.3 Model Evaluation**

Accuracy is the ratio of number of correctly predicted residues to number of all predicted residues and formulated as follows,

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}} \quad (3.2)$$

where true positive, false positive, true negative, false negative stands for correctly predicted hot spot residues, non-hot spot residues incorrectly predicted as hot spots, correctly predicted non-hot spot residues and hot spot residues incorrectly predicted as non-hot spots respectively.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3.3)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3.4)$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (3.5)$$

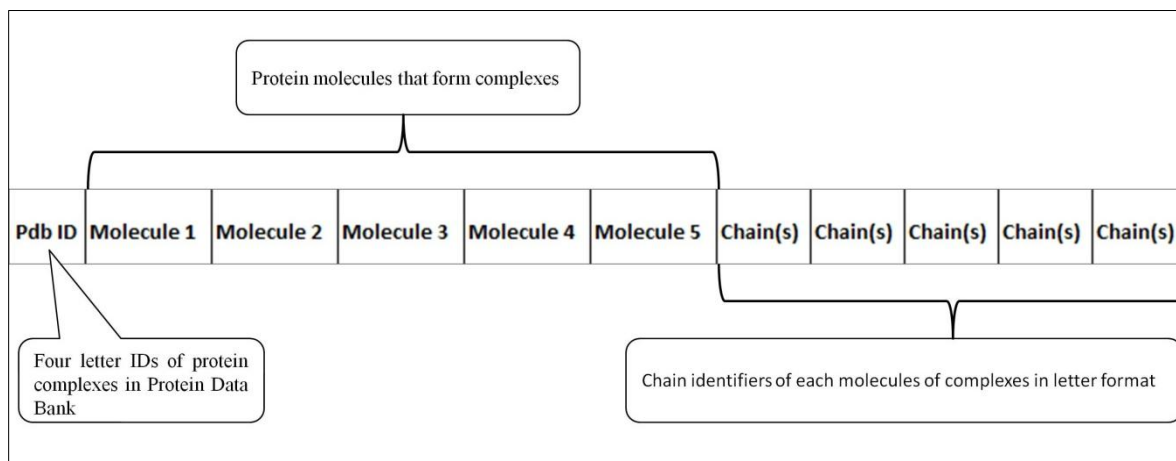
where recall is the proportion of number of correctly classified hot spot residues to the number of all hot spot residues; precision is the ratio of number of correctly classified hot spot residues to the number of all residues classified as hot spots; specificity is the proportion of number of correctly predicted non-hot spot residues to the number of all non-hot spot residues. F-measure ( $F_1$  score) is the weighted harmonic mean of precision and recall, formulated as;

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3.6)$$

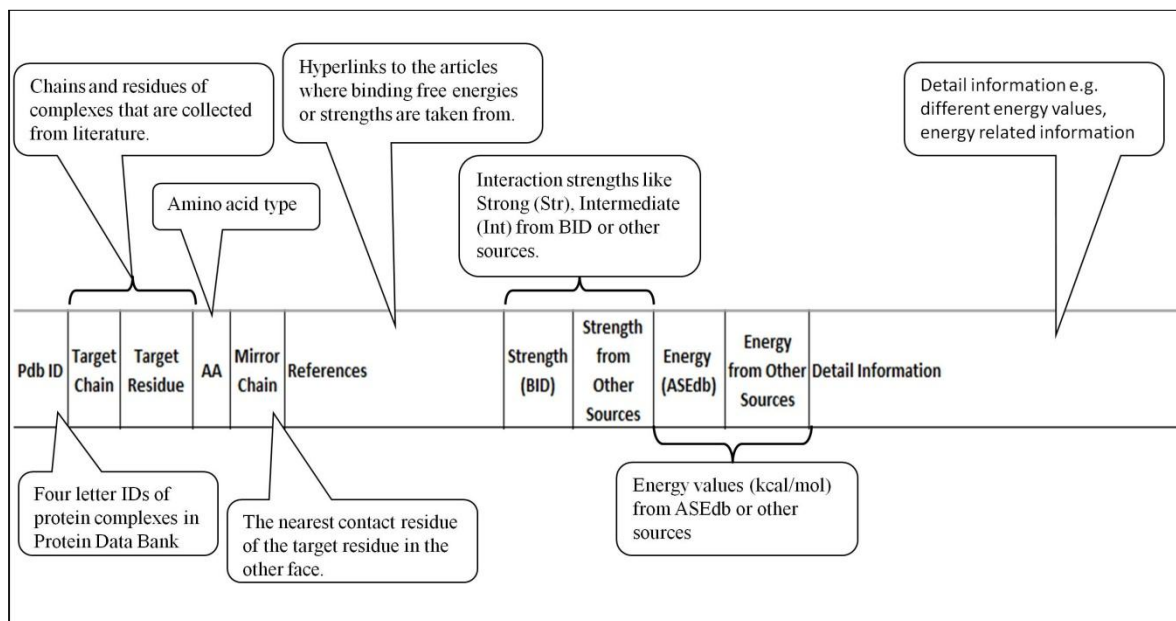
### 3.4 HotBase: Database for Hot Spot Prediction

We gather all data related to hot spot prediction into a “Hotbase” database. There are mainly four tables in Hotbase: complexes, data set, feature values, server results. Figure 3.2

and Figure 3.3 show explanations of columns of “Complexes” table and “Data Set” table by showing header rows with bubbles.



**Figure 3.2** Header row of “Complexes” table with explanations of the columns in bubbles.



**Figure 3.3** Header row of “Data Set” table with explanations of the columns in bubbles.



The results of BeAtMusic [44] and Robetta [54] servers, FoldX plugin of YASARA[61, 62], KFC2[26] and HotPoint [43] servers are given in “Server Results” table and Figure 3.4 illustrates header row of the table with explanations in bubbles. BeAtMuSiC server that is based on statistical potentials has been developed to predict the effects of mutations on protein–protein binding affinities. Robetta and FoldX are energy-based methods and the outputs are binding free energies. In our study, we used FoldX plugin of YASARA. KFC2 server has two methods KFC2\_A and KFC2\_B that have different feature compositions and provides classifications and confidence of predictions. HotPoint server makes predictions based on relative ASA in complex state and pairwise potentials.

Pdb ID	Target Chain	Target Residue	AA	Mirror Chain	Strength (BID)	Strength from Other Sources	Energy (ASEdb)	Energy from Other Sources	BeAtMuSiC	Robetta	Yasara-Foldx	KFC2_A	KFC2A_Conf	KFC2_B	KFC2B_Conf	HotPoint
--------	--------------	----------------	----	--------------	----------------	-----------------------------	----------------	---------------------------	-----------	---------	--------------	--------	------------	--------	------------	----------

**Figure 3.4** Header row of “Server Results” table with explanations in bubbles.

## Chapter 4

### RESULTS AND DISCUSSION

#### 4.1 Benchmark Data Set

In this study, we used data of thirteen different data sets that were not separate data sets. Figure 4.1 shows number of co-owned residues and percentage of intersections of two data sets. For instance, Xia et al. [30] and Cho et al. [15] are co-own the data of 153 residues and this is 39.23% of Cho et al. (row 5, column 13) and 100% of the data set of Xia et al. (row 13, column 5). In the last column, how much of our data set intersect with other data sets in terms of residue number and percentage are presented. Our data mostly come from the data set of Chen et al. [20] and ASEdb [12]. They form 43.39% and 39.57% of the whole data set respectively.

Figure 4.1 The number and percentage of residues in the data

Sources	ASEdb	BID	Chen et al.	Cho et al.	Darnell et al.	Kortenne and Baker	Kumar et al.	Lise et al.	Moal and Fernández-Reco	Tunchag et al.	Wang et al.	Xia et al.	Zhu and Mitchell	Our Data Set
ASEdb		35 (7.35%)	223 (46.85%)	150 (31.51%)	200 (42.02%)	236 (49.58%)	1 (0.21%)	211 (44.33%)	11 (2.31%)	127 (26.68%)	195 (40.97%)	84 (17.65%)	72 (15.13%)	476 (39.57%)
BID	35 (15.02%)		118 (50.64%)	108 (46.35%)	90 (38.63%)	23 (9.87%)	0 (0.00%)	53 (22.75%)	10 (4.29%)	67 (28.76%)	111 (47.64%)	20 (8.58%)	96 (41.20%)	233 (19.37%)
Chen et al.	223 (42.72%)	118 (22.61%)		255 (48.85%)	251 (48.08%)	153 (29.31%)	0 (0.00%)	217 (41.57%)	11 (2.11%)	178 (34.10%)	269 (51.53%)	97 (18.58%)	165 (31.61%)	522 (43.39%)
Cho et al.	150 (38.46%)	108 (27.69%)	255 (65.38%)		177 (45.38%)	225 (57.69%)	0 (0.00%)	240 (61.54%)	18 (4.62%)	130 (33.33%)	368 (94.36%)	153 (39.23%)	252 (64.62%)	390 (32.42%)
Darnell et al.	200 (57.31%)	90 (25.79%)	251 (71.92%)	177 (50.72%)		134 (38.40%)	0 (0.00%)	189 (54.15%)	10 (2.87%)	250 (71.63%)	218 (62.46%)	78 (22.35%)	117 (33.52%)	349 (29.01%)
Kortenne and Baker	236 (65.19%)	23 (6.35%)	153 (42.27%)	225 (62.15%)	134 (37.02%)		0 (0.00%)	255 (70.44%)	12 (3.31%)	85 (23.48%)	251 (69.34%)	135 (37.29%)	107 (29.56%)	362 (30.09%)
Kumar et al.	1 (1.11%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	90 (7.48%)
Lise et al.	211 (60.46%)	53 (15.19%)	217 (62.18%)	240 (68.77%)	189 (54.15%)	255 (73.07%)	0 (0.00%)		12 (3.44%)	116 (33.24%)	289 (82.81%)	143 (40.97%)	116 (33.24%)	349 (29.01%)
Moal and Fernández-Reco	11 (26.83%)	10 (24.39%)	11 (26.83%)	18 (43.90%)	10 (24.39%)	12 (29.27%)	0 (0.00%)	12 (29.27%)		8 (19.51%)	12 (29.27%)	15 (36.59%)	13 (31.71%)	41 (3.41%)
Tunchag et al.	127 (48.66%)	67 (25.67%)	178 (68.20%)	130 (49.81%)	250 (95.79%)	85 (32.57%)	0 (0.00%)	116 (44.44%)	8 (3.07%)		153 (58.62%)	78 (29.89%)	101 (38.70%)	261 (21.70%)
Wang et al.	195 (44.72%)	111 (25.46%)	269 (61.70%)	368 (84.40%)	218 (50.00%)	251 (57.57%)	0 (0.00%)	289 (66.28%)	12 (2.75%)	153 (35.09%)		145 (33.26%)	240 (55.05%)	436 (36.24%)
Xia et al.	84 (54.90%)	20 (13.07%)	97 (63.40%)	153 (100%)	78 (50.98%)	135 (88.24%)	0 (0.00%)	143 (93.46%)	15 (9.80%)	78 (50.98%)	145 (94.77%)		91 (59.48%)	153 (12.72%)
Zhu and Mitchell	72 (28.57%)	96 (38.10%)	165 (65.48%)	252 (100%)	117 (46.43%)	107 (42.46%)	0 (0.00%)	116 (46.03%)	13 (5.16%)	101 (40.08%)	240 (95.24%)	91 (36.11%)		252 (20.95%)

Table 4.1 lists the protein complexes from which the data are derived. In our study, the interface residues are hot spots if their observed binding free energy changes are greater than or equal to 2.0 kcal/mol or their interaction strengths are labeled as “Strong”. The others are considered as non-hot spots. The benchmark data set consists of 1203 residues (240 hot spots, 963 non-hot spots).

**Table 4.1** The 79 protein–protein complexes in HotBase.

Pdb ID	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5	Chain(s)	Chain(s)	Chain(s)	Chain(s)	Chain(s)
1A22	Human Growth Hormone	Human Growth Hormone Receptor				A	B			
1A4Y	Ribonuclease Inhibitor	Angiogenin				A, D	B, E			
1AHW	Immunoglobulin Fab5g9 (Light Chain)	Immunoglobulin Fab5g9 (Heavy Chain)	Tissue Factor			A, D	B, E	C, F		
1AON	Groel	Groel-Groes Complex				A, B, C, D, E, F, G, H, I, J, K, L, M, N	O, P, Q, R, S, T, U			
1AWC	Mouse Gabp Alpha	Mouse Gabp Beta	DNA	DNA		A	B	D	E	
1AYF	Bovine Adrenodoxin (Oxidized)					A, B				
1B26	Glutamate Dehydrogenase					A, B, C, D, E, F				
1BNI	Barnase Wildtype Structure at Ph 6.0					A, B, C				
1BP3	Growth Hormone	Prolactin Receptor				A	B			
1BRS	Barnase	Barstar				A, B, C	D, E, F			
1BSR	Bovine Seminal Ribonuclease					A, B				
1BXI	Colicin E9 Immunity Im9	Colicin E9				A	B			
1C2R	Cytochrome C2					A, B				
1C4Z	Ubiquitin-Protein Ligase E3A	Ubiquitin Conjugating Enzyme E2				A, B, C	D			
1CBW	Bovine Chymotrypsin	Bovine Chymotrypsin	Bovine Chymotrypsin	Basic Pancreatic		A, F	B, G	C, H	D, I	

		in	in	Trypsin Inhibitor						
1CDC	CD2, N-Terminal Domain (1-99), Truncated Form					A, B				
1CDL	Calmodulin	Calcium/Calmodulin-Dependent Protein Kinase Type II Alpha Chain				A, B, C, D	E, F, G, H			
1DAN	Blood Coagulation Factor VIIA (Light Chain)	Blood Coagulation Factor VIIA (Heavy Chain)	Soluble Tissue Factor	Soluble Tissue Factor		L	H	T	U	
1DDM	Numb Protein	Numb Associate Kinase				A	B			
1DFJ	Ribonuclease A	Ribonuclease Inhibitor				E	I			
1DKG	Nucleotide Exchange Factor Grpe	Molecular Chaperone Dnak				A, B	D			
1DN2	Immunoglobulin Lambda Heavy Chain	Engineered Peptide				A, B	E, F			
1DQJ	Anti-Lysozyme Antibody Hyhel-63 (Light Chain)	Anti-Lysozyme Antibody Hyhel-63 (Heavy Chain)	Lysozyme			A	B	C		
1DVA	Des-Gla Factor Viia (Heavy Chain)	Des-Gla Factor Viia (Light Chain)	Peptide E-76			H, I	L, M	X, Y		
1DVF	Idiotopic Antibody Fv D1.3	Idiotopic Antibody Fv D1.3	Anti-Idiotopic Antibody Fv E5.2	Anti-Idiotopic Antibody Fv E5.2		A	B	C	D	
1DX5	Thrombin (Light Chain)	Thrombin Inhibitor Glu-Gly-Arg-0qe	Thrombomodulin	Thrombin (Heavy Chain)		A, B, C, D	E, F, G, H	I, J, K, L	M, N, O, P	
1DZI	Integrin	Collagen				A	B, C, D			
1EBP	Erythropoietin Receptor	Erythropoietin Mimetics Peptide 1				A, B	C, D			
1ES7	Bone Morphogenetic Protein-2	Bone Morphogenetic Protein				A, C	B, D			

		Receptor IA								
1F47	Cell Division Protein FtsZ	Cell Division Protein ZipA				A	B			
1FAK	Blood Coagulation Factor Viia (Light Chain)	Blood Coagulation Factor Viia (Heavy Chain)	Soluble Human Tissue Factor	Protein (5L15)		L	H	T	I	
1FC1	Human FC Fragment					A, B				
1FC2	Fragment B of Protein A Complex	Immunoglobulin FC				C	D			
1FCC	Igg1 MO61 FC	Streptococcal Protein G (C2 Fragment)				A, B	C, D			
1FE8	The Von Willebrand Factor	Immunoglobulin Igg Ru5 (Heavy Chain)	Immunoglobulin Igg Ru5 (Light Chain)			A, B, C	H, I, J	L, M, N		
1FOE	T-Lymphoma Invasion and Metastasis Inducing Protein 1	RAS-Related C3 Botulinum Toxin Substrate				A, C, E, G	B, D, F, H			
1G3I	ATP-Dependent HSLU Protease ATP-Binding Subunit HSLU	ATP-Dependent Protease HSLV				A, B, C, D, E, F, S, T, U, V, W, X	G, H, I, J, K, L, M, N, O, P, Q, R			
1G6N	Catabolite Gene Activator Protein					A, B				
1GC1	CD4	Envelope Protein Gp120	Antibody 17B	Antibody 17B		C	G	L	H	
1GL4	Nidogen-1	Basement Membrane-Specific Heparan Sulfate Proteoglycan Core Protein				A	B			
1HFY	Alpha-Lactalbumin					A, B				
1HMV	HIV-1 Reverse Transcriptase (Subunit P66)	HIV-1 Reverse Transcriptase (Subunit P51)				A, C, E, G	B, D, F, H			
1IHB	Cyclin-					A, B				

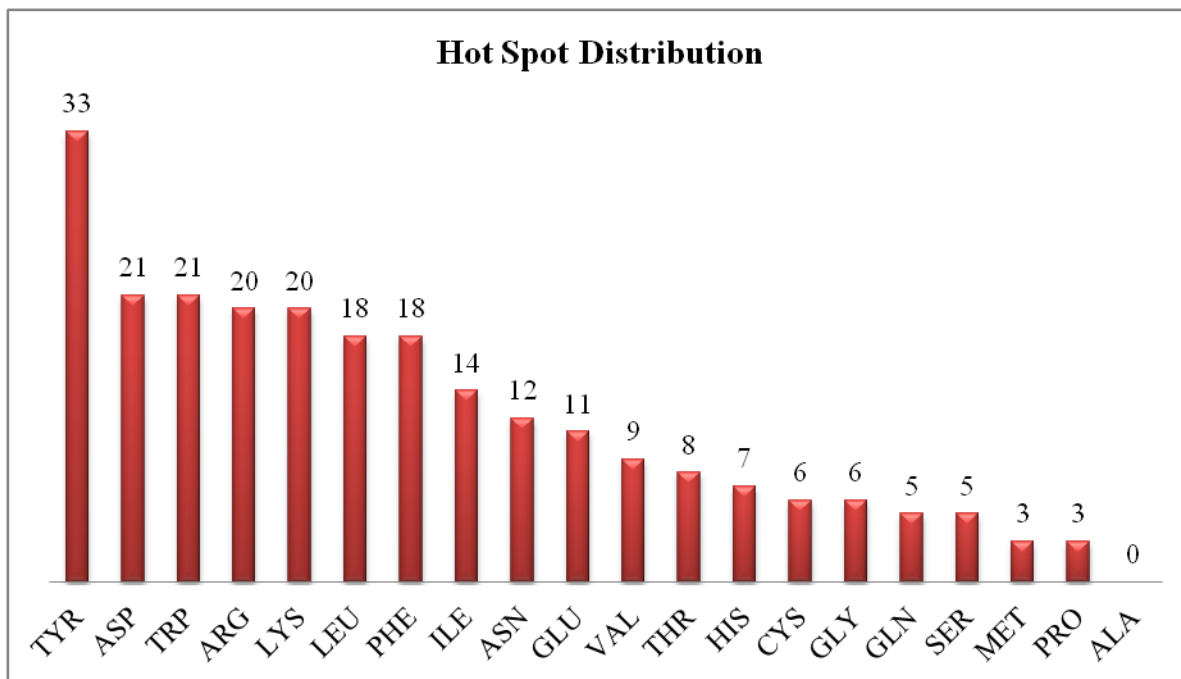
	Dependent Kinase 6 Inhibitor									
IIO4	Caat/Enhancer Binding Protein Beta	Runt-Related Transcription Factor 1	Core-Binding Factor, Beta Subunit	CSF-1R Promoter	CSF-1R Promoter	A, B	C	D	E (DNA)	F (DNA)
IJAT	Ubiquitin-Conjugating Enzyme E2-17.5 KDA	Ubiquitin-Conjugating Enzyme Variant Mms2				A	B			
IJCK	14.3.D T Cell Antigen Receptor	Staphylococcal Enterotoxin C3				A, C	B, D			
IJDH	Beta-Catenin	hTcF-4				A	B			
IJMA	Glycoprotein D	Herpesvirus Entry Mediator				A	B			
IJPP	Beta-Catenin	Adenomatous Polyposis Coli				A, B	C, D			
IJRH	Antibody A6	Antibody A6	Interferon-Gamma Receptor Alpha Chain			L	H	I		
IJTG	Beta-Lactamase Tem	Beta-Lactamase Inhibitory Protein				A, C	B, D			
IK4U	Phagocyte Nadph Oxidase Subunit P67 Phox	Phagocyte Nadph Oxidase Subunit P47 Phox				S	P			
IKTZ	Transforming Growth Factor Beta 3	TGF-Beta Type II Receptor				A	B			
ILFD	RALGDS	RAS				A, C	B, D			
ILQB	Elongin B	Elongin C	Von Hippel-Lindau Disease Tumor Suppressor	Hypoxia-Inducible Transcription Factor 1 Alpha		A	B	C	D	
IMQ8	Intercellular Adhesion Molecule-1	Integrin Alpha-L				A, C	B, D			
INFI	Nf-Kappa-B P65	Nf-Kappa-B P50	I-Kappa-B-Alpha			A, C	B, D	E, F		
INMB	Fab Nc10	Fab Nc10	N9 Neuramidase			H	L	N		
INUN	Fibroblast Growth	Fibroblast Growth				A	B			

	Factor-10	Factor Receptor 2 Isoform 2								
1OMW	G-Protein Coupled Receptor Kinase 2	Guanine Nucleotide-Binding Protein G(I)/G(S)/G(T) Beta	Guanine Nucleotide-Binding Protein G(I)/G(S)/G(O) Gamma-2 Subunit			A	B	G		
1RGG	Ribonuclease					A, B				
1RN1	Ribonuclease T1 Isozyme					A, B, C				
1T0J	Voltage-Gated Calcium Channel Subunit Beta2a	Voltage-Gated Calcium Channel Subunit Beta2a	Voltage-Dependent L-Type Calcium Channel Alpha-1c Subunit			A	B	C		
1TPK	Tissue Plasminogen Activator					A, B, C				
1UB4	MazF Protein	MazE Protein				A, B	C			
1VFB	Igg1-Kappa D1.3 Fv (Light Chain)	Igg1-Kappa D1.3 Fv (Heavy Chain)	Hen Egg White Lysozyme			A	B	C		
1WQ5	Tryptophan Synthase Alpha Chain					A, B				
2GOX	Complement C3	Fibrinogen-Binding Protein				A, C	B, D			
2HHB	Hemoglobin (Deoxy) (Alpha Chain)	Hemoglobin (Deoxy) (Beta Chain)				A, C	B, D			
2NMB	Numb Protein	Gppy Peptide				A	B			
2PCC	Cytochrome C Peroxidase	Iso-1-Cytochrome C				A, C	B, D			
2PTC	Beta-Trypsin	Trypsin Inhibitor				E	I			
2VLJ	HLA Class I Histocompatibility Antigen, A-2 Alpha Chain	Beta-2-Microglobulin	Flu Matrix Peptide	JM22 TCR Alpha Chain	JM22 TCR Beta Chain	A	B	C	D	E
3BPL	Interleukin-4	Interleukin-4 Receptor Alpha Chain	Cytokine Receptor Common Gamma Chain			A	B	C		
3BT2	Urokinase-	Anti-Upar	Anti-uPAR	Anti-uPAR	Urokinase	A	B	L	H	U



	Type Plasminogen Activator	Antibody, Light Chain	Antibody, Light Chain	Antibody, Heavy Chain	Plasminogen Activator Surface Receptor					
3BUK	Neurotrophin-3	Tumor Necrosis Factor Receptor Superfamily Member 16				A, B	C, D			
3HFM	Hyhel-10 Igg1 Fab (Light Chain)	Hyhel-10 Igg1 Fab (Heavy Chain)	Hen Egg White Lysozyme			L	H	Y		
3HHR	Human Growth Hormone	Human Growth Hormone Receptor				A	B, C			
3SAK	Tumor Suppressor P53					A, B, C, D				

As previously mentioned, Bogan and Thorn reported that hot spots are abundant in tryptophan (trp), arginine (arg) and tyrosine (tyr), whereas some residues like leucine (leu), methionine (met), serine (ser), threonine (thr) and valine (val) are disfavored [9]. In Figure 4.2, hot spot distribution in our data set is presented and amino acid preferences in hot spots are similar to preferences in the report. Tyrosine (tyr), aspartic acid (asp), tryptophan (trp) and arginine (arg) are much more likely to be in hot spots than others. Furthermore, there are totally 240 hot spot residues in our data set and the percentages of methionine (met), serine (ser), threonine (thr) and valine (val) residues in hot spots are less than 4%.



**Figure 4.2** Distribution of hot spot residues.

#### 4.2 Features in HotBase

There are values of seventy features in our database. In Table 4.2, complete list of features is given.

**Table 4.2** List of features in HotBase.

Column	Feature
1	Secondary structure of target residues (ss_target)
2	Secondary structure of intra-contact residues (ss_intra)
3	Secondary structure of mirror-contact residues (ss_mirror)
4	Class of target residues (class_target)
5	Class of intra-contact residues (class_intra)
6	Class of mirror-contact residues (class_mirror)
7	Atom contact areas of target residues (atom_contact_areas_target)
8	Atom contacts of target residues (atom_contacts_target)
9	Residue contacts of target residues (residue_contacts_target)
10	Hydrophobicity of target residues (hydrophobicity_target)
11	Hydrophilicity of target residues (hydrophilicity_target)
12	Isoelectric point of target residues (isopoint_target)

13	Mass of target residues (mass_target)
14	Polarity of target residues (polarity_target)
15	Polarizability of target residues (polarizability_target)
16	Average depth index of target residues (ave_dpx_target)
17	Standard deviation of depth index of target residues (sd_dpx_target)
18	Side-chain average depth index of target residues (s_ch_ave_dpx_target)
19	Standard deviation of side-chain depth index of target residues (sd_s_ch_dpx_target)
20	Relative accessible surface area of target residues (rel_ASA_target)
21	Relative side-chain accessible surface area of target residues (rel_s_ch_ASA_target)
22	Relative depth index of target residues (rel_dpx_target)
23	Relative side-chain depth index of target residues (rel_s_ch_dpx_target)
24	Atom contact areas of intra-contact residues (atom_contact_areas_intra)
25	Atom contacts of intra-contact residues (atom_contacts_intra)
26	Residue contacts of intra-contact residues (residue_contacts_intra)
27	Hydrophobicity of intra-contact residues (hydrophobicity_intra)
28	Hydrophilicity of intra-contact residues (hydrophilicity_intra)
29	Isoelectric point of intra-contact residues (isopoint_intra)
30	Mass of intra-contact residues (mass_intra)
31	Polarity of intra-contact residues (polarity_intra)
32	Polarizability of intra-contact residues (polarizability_intra)
33	Average depth index of intra-contact residues (ave_dpx_intra)
34	Standard deviation of depth index of intra-contact residues (sd_dpx_intra)
35	Side-chain average depth index of intra-contact residues (s_ch_ave_dpx_intra)
36	Standard deviation of side-chain depth index of intra-contact residues (sd_s_ch_dpx_intra)
37	Relative accessible surface area of intra-contact residues (rel_ASA_intra)
38	Relative side-chain accessible surface area of intra-contact residues (rel_s_ch_ASA_intra)
39	Relative depth index of intra-contact residues (rel_dpx_intra)
40	Relative side-chain accessible surface area of intra-contact residues (rel_s_ch_dpx_intra)
41	Atom contact areas of mirror-contact residues (atom_contact_areas_mirror)
42	Atom contacts of mirror-contact residues (atom_contacts_mirror)
43	Residue contacts of mirror-contact residues (residue_contacts_mirror)
44	Hydrophobicity of mirror-contact residues (hydrophobicity_mirror)
45	Hydrophilicity of mirror-contact residues (hydrophilicity_mirror)
46	Isoelectric point of mirror-contact residues (isopoint_mirror)
47	Mass of mirror-contact residues (mass_mirror)
48	Polarity of mirror-contact residues (polarity_mirror)
49	Polarizability of mirror-contact residues (polarizability_mirror)
50	Average depth index of mirror-contact residues (ave_dpx_mirror)
51	Standard deviation of depth index of mirror-contact residues (sd_dpx_mirror)
52	Side-chain average depth index of mirror-contact residues (s_ch_ave_dpx_mirror)
53	Standard deviation of side-chain depth index of mirror-contact residues (sd_s_ch_dpx_mirror)
54	Relative accessible surface area of mirror-contact residues (rel_ASA_mirror)
55	Relative side-chain accessible surface area of mirror-contact residues (rel_s_ch_ASA_mirror)

56	Relative depth index of mirror-contact residues (rel_dpx_mirror)
57	Relative side-chain accessible surface area of mirror-contact residues (rel_s_ch_dpx_mirror)
58	Hydrophobicity from the work of Chen et al. (hydrophobicity_rig)
59	Hydrophilicity from the work of Chen et al. (hydrophilicity_rig)
60	Polarity from the work of Chen et al. (polarity_rig)
61	Polarizability from the work of Chen et al. (polarizability_rig)
62	Propensities from the work of Chen et al. (propensities_rig)
63	ASA from the work of Chen et al. (Average_accessible_surface_area_rig)
64	Sequence_Profile
65	Sequence_Entropy
66	Conservation_Score
67	Relative ASA in complex state (RelCompASA)
68	Relative ASA in monomer state (RelMonomerASA)
69	Pairwise Potential (Potential)
70	ASA difference between monomer state and complex state (DiffASA)

### 4.3 HotBase Website

In Figure 4.3, introduction webpage of HotBase is illustrated. In “Complexes” webpage (Figure 4.4), 79 complexes and their chains are given. In “Data Set” webpage (Figure 4.5), binding strength, binding free energies and detail information about 1203 residues of 79 complexes from 13 different studies are presented. We introduce prediction results of Robetta, KFC2, Hotpoint and FoldX plugin of YASARA program in “Server Results” webpage (Figure 4.6). Dash (“-”), represents no information case. The “Feature Values” webpage (Figure 4.7) contains values of seventy features. ASEdb is no longer being actively maintained so we provide all its data in “ASEdb” webpage (Figure 4.8). “About Us” includes a photo of our lab members and link to our lab webpage (Figure 4.9).

**HotBase**  
Database for Hot Spot Prediction

**Welcome to HotBase!!**

Hot spots are small subset of protein-protein interface residues but they are accounting for the majority of the binding free energy. Measuring the binding free energy change upon mutating residues to alanine is an experimental way to determine hot spots. If mutation of a residue gives rise to a significantly large change in the binding free energy then this residue is a hot spot. Experimental determination of hot spots is time-consuming, labor intensive and has high economic costs. Therefore, computational methods are developed for hot spot prediction.

We offer a new standard benchmark data set, seventy feature values for computationally hot spot prediction and BeAtMuSIC, Robetta, Yasara-Foldx, KFC2 and HotPoint results.

HotBase: Database for Hot Spot Prediction (v 1.00)  
For any problems or comments please send e-mail to Selin Karagulle

Figure 4.3 Homepage of Hotbase

**HotBase**  
Database for Hot Spot Prediction

**List of Complexes**

Pdb ID	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5	Chain(s)	Chain(s)	Chain(s)	Chain(s)	Chain(s)
<a href="#">1A2Z</a>	Human Growth Hormone	Human Growth Hormone Receptor				A	B			
<a href="#">1A4Y</a>	Ribonuclease Inhibitor	Angiogenin				A, D	B, E			
<a href="#">1AHV</a>	Immunoglobulin Fab5g9 (Light Chain)	Immunoglobulin Fab5g9 (Heavy Chain)	Tissue Factor			A, D	B, E	C, F		
<a href="#">1AON</a>	Groel	Groel-Groes Complex				A, B, C, D, E, F, G, H, I, J, K, L, M, N	O, P, Q, R, S, T, U			
<a href="#">1AWC</a>	Mouse Gabp Alpha	Mouse Gabp Beta	DNA	DNA		A	B	D	E	
<a href="#">1AYF</a>	Bovine Adrenodoxin (Oxidized)					A, B				
<a href="#">1B26</a>	Glutamate Dehydrogenase					A, B, C, D, E, F				
<a href="#">1BNI</a>	Barnase Wildtype Structure at Ph 6.0					A, B, C				
<a href="#">1BP3</a>	Growth Hormone	Prolactin Receptor				A	B			
<a href="#">1BR5</a>	Barnase	Barstar				A, B, C	D, E, F			
<a href="#">1BSR</a>	Bovine Seminal Ribonuclease					A, B				
<a href="#">1BXI</a>	Colicin E9 Immunity Im9	Colicin E9				A	B			

HotBase: Database for Hot Spot Prediction (v 1.00)  
For any problems or comments please send e-mail to Selin Karagulle

Figure 4.4 "Complexes" webpage.

**HotBase**  
Database for Hot Spot Prediction

KOC UNIVERSITY

Home  
Complexes  
Data Set  
Server Results  
Feature Values  
ASEdb  
About us

Download Benchmark Dataset

Pdb ID	Target Chain	Target Residue	AA	Mirror Chain	References	Strength (BID)	Strength from Other Sources	Energy (ASEdb)	Energy from Other Sources	Detail Information
1A22	A	14	MET	B	13				0.10	
1A22	A	18	HIS	B	3, 5, 6, 7, 8, 13				-0.50	
1A22	A	21	HIS	B	3, 5, 6, 7, 13				0.20	
1A22	A	22	GLN	B	3, 5, 6, 7, 8, 13				-0.20	
1A22	A	25	PHE	B	3, 5, 6, 7, 11, 13				-0.40	In 11, the energy value is -0.42 kcal/mol
1A22	A	26	ASP	B	13				-0.05	
1A22	A	29	GLN	B	13				0.01	
1A22	A	42	TYR	B	3, 5, 6, 7, 8, 13				0.20	
1A22	A	45	LEU	B	3, 5, 7, 13				1.20	
1A22	A	46	GLN	B	3, 5, 6, 7				0.10	
1A22	A	51	SER	B	3, 5, 6, 7, 8, 13				0.30	
1A22	A	56	GLU	B	3, 5, 7, 13				0.40	
1A22	A	62	SER	B	3, 5, 6, 7, 13				0.10	
1A22	A	63	ASN	B	3, 5, 6, 7, 13				0.30	
1A22	A	64	ARG	B	3, 5, 7, 13				1.60	
1A22	A	65	GLU	B	3, 5, 6, 7, 8, 13				-0.50	

HotBase: Database for Hot Spot Prediction (v 1.00)  
For any problems or comments please send e-mail to Selin Karagulle

Figure 4.5 "Data Set" webpage

**HotBase**  
Database for Hot Spot Prediction

KOC UNIVERSITY

Home  
Complexes  
Data Set  
Server Results  
Feature Values  
ASEdb  
About us

Download Server Results

Pdb ID	Target Chain	Target Residue	AA	Mirror Chain	Strength (BID)	Strength from Other Sources	Energy (ASEdb)	Energy from Other Sources	BeAtMuSic	Robetta	Yasara-Foldx	KFC2_A	KFC2A_Conf	KFC2_B	KFC2B_Conf	HotPoint
1A22	A	14	MET	B				0.10	0.61	0.00	2.21	-	-	-	-	H
1A22	A	18	HIS	B				-0.50	2.00	2.29	0.32	NH	-0.39	NH	-0.41	NH
1A22	A	21	HIS	B				0.20	0.82	0.83	1.00	H	0.5	H	0.1	NH
1A22	A	22	GLN	B				-0.20	1.05	0.05	-1.54	NH	-2.35	NH	-0.99	NH
1A22	A	25	PHE	B				-0.40	2.44	1.29	0.51	NH	-0.87	NH	-0.41	H
1A22	A	26	ASP	B				-0.05	0.05	-0.05	-1.18	-	-	-	-	NH
1A22	A	29	GLN	B				0.01	0.11	0.01	-0.16	-	-	-	-	NH
1A22	A	42	TYR	B				0.20	1.81	2.02	1.31	NH	-0.3	NH	-0.15	NH
1A22	A	45	LEU	B				1.20	2.05	1.15	2.33	H	1.31	H	0.28	H
1A22	A	46	GLN	B				0.10	0.56	1.04	-0.07	NH	-1.04	NH	-0.79	NH
1A22	A	51	SER	B				0.30	0.83	-0.02	-0.06	NH	-0.13	NH	-0.45	H
1A22	A	56	GLU	B				0.40	1.15	0.97	-1.27	NH	-1.86	NH	-0.95	NH
1A22	A	62	SER	B				0.10	0.78	-0.17	0.54	H	0.56	NH	-0.59	NH
1A22	A	63	ASN	B				0.30	1.18	0.39	0.75	NH	-0.89	NH	-0.72	NH
1A22	A	64	ARG	B				1.60	1.81	2.00	0.10	H	0.54	NH	-0.04	NH
1A22	A	65	GLU	B				-0.50	0.40	-0.11	-0.53	NH	-1.93	NH	-0.9	NH
1A22	A	68	GLN	B				0.60	1.02	1.81	0.88	NH	-1.23	NH	-0.71	NH
1A22	A	164	TYR	B				0.30	1.58	0.99	2.42	H	0.26	H	0.17	H

HotBase: Database for Hot Spot Prediction (v 1.00)  
For any problems or comments please send e-mail to Selin Karagulle

Figure 4.6 "Server Results" webpage

**HotBase**  
Database for Hot Spot Prediction

KOC UNIVERSITY

Home  
Complexes  
Data Set  
Server Results  
Feature Values  
ASEdb  
About us

[Download Feature Values](#)

Pdb ID	Target Chain	Target Residue	AA	Mirror Chain	Strength (BID)	Strength from Other Sources	Energy (ASEdb)	Energy from Other Sources	ss_target	ss_intra	ss_mirror	class_target	class_intra	class_mirror	atom_contact_areas_target	atom_con
1A22	A	14	MET	B				0.10	helix	helix	loop	cla3	cla1	cla1	0.00	
1A22	A	18	HIS	B				-0.50	helix	helix	loop	cla4	cla5	cla4	89.70	
1A22	A	21	HIS	B				0.20	helix	helix	loop	cla4	cla4	cla4	45.30	
1A22	A	22	GLN	B				-0.20	helix	helix	loop	cla4	cla4	cla4	0.40	
1A22	A	25	PHE	B				-0.40	helix	helix	loop	cla2	cla1	cla4	59.00	
1A22	A	26	ASP	B				-0.05	helix	helix	loop	cla6	cla3	cla4	0.00	
1A22	A	29	GLN	B				0.01	helix	helix	loop	cla4	cla3	cla3	0.00	
1A22	A	42	TYR	B				0.20	helix	helix	strand	cla3	cla3	cla3	87.00	
1A22	A	45	LEU	B				1.20	helix	helix	strand	cla2	cla2	cla2	91.20	
1A22	A	46	GLN	B				0.10	loop	loop	loop	cla4	cla4	cla4	81.60	
1A22	A	51	SER	B				0.30	helix	loop	strand	cla3	cla2	cla2	11.00	
1A22	A	56	GLU	B				0.40	helix	helix	loop	cla6	cla3	cla4	26.50	
1A22	A	62	SER	B				0.10	loop	loop	loop	cla3	cla4	cla2	75.80	
1A22	A	63	ASN	B				0.30	loop	loop	loop	cla4	cla3	cla3	27.60	
1A22	A	64	ARG	B				1.60	helix	loop	loop	cla5	cla4	cla6	105.50	

HotBase: Database for Hot Spot Prediction (v 1.00)  
For any problems or comments please send e-mail to Selin Karagulle

Figure 4.7 "Feature Values" webpage

**HotBase**  
Database for Hot Spot Prediction

KOC UNIVERSITY

Home  
Complexes  
Data Set  
Server Results  
Feature Values  
ASEdb  
About us

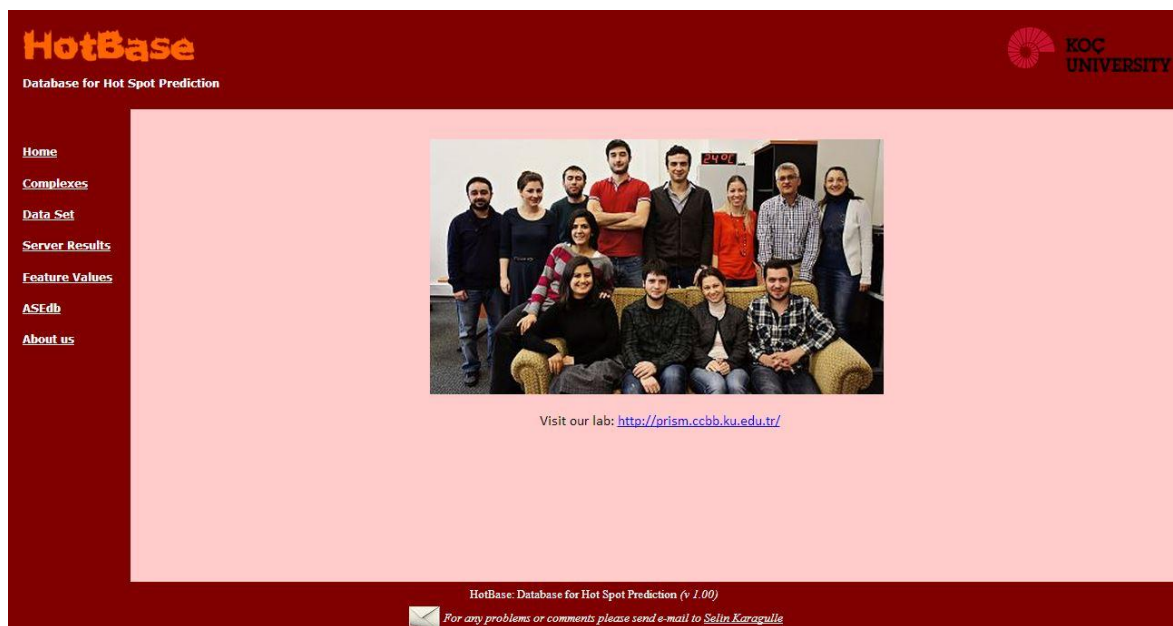
[Download ASEdb](#)

Mutated Protein	Partner (Type)	PDB	AA	Position	ddG	Monomer	Complex	Delta	RefID
Angiogenin	RNase inhibitor (P)	1a4y(D)	R	5	2.30	141.48	41.35	100.13	50
Angiogenin	RNase inhibitor (P)	1a4y(D)	H	8	0.90	109.67	85.62	24.05	6
Angiogenin	RNase inhibitor (P)	1a4y(D)	Q	12	0.30	37.86	15.09	22.77	6
Angiogenin	RNase inhibitor (P)	1a4y(D)	H	13	-0.30	13.26	8.56	4.70	49
Angiogenin	RNase inhibitor (P)	1a4y(D)	R	31	0.20	157.22	37.30	119.92	50
Angiogenin	RNase inhibitor (P)	1a4y(D)	R	32	0.90	164.82	102.66	62.16	50
Angiogenin	RNase inhibitor (P)	1a4y(D)	R	33	0.30	46.54	46.54	0.00	50
Angiogenin	RNase inhibitor (P)	1a4y(D)	R	66	0.20	130.80	130.80	0.00	50
Angiogenin	RNase inhibitor (P)	1a4y(D)	N	68	0.20	95.50	84.38	11.12	6
Angiogenin	RNase inhibitor (P)	1a4y(D)	R	70	-0.20	45.34	45.34	0.00	50
Angiogenin	RNase inhibitor (P)	1a4y(D)	H	84	0.20	98.57	34.60	63.97	67
Angiogenin	RNase inhibitor (P)	1a4y(D)	W	89	0.20	171.99	87.22	84.77	67
Angiogenin	RNase inhibitor (P)	1a4y(D)	E	108	-0.30	71.77	27.87	43.90	6
Angiogenin	RNase inhibitor (P)	1a4y(D)	H	114	0.65	87.25	24.63	62.62	49
RNase inhibitor	Angiogenin (P)	1a4y(D)	W	261	0.10	89.01	57.01	32.00	67
RNase inhibitor	Angiogenin (P)	1a4y(D)	W	263	1.20	68.99	3.03	65.96	67
RNase inhibitor	Angiogenin (P)	1a4y(D)	E	287	0.10	40.97	40.97	0.00	67
RNase inhibitor	Angiogenin (P)	1a4y(D)	S	289	0.00	7.54	-	-	67

HotBase: Database for Hot Spot Prediction (v 1.00)  
For any problems or comments please send e-mail to Selin Karagulle

Figure 4.8 "ASEdb" webpage





**Figure 4.9** "About Us" webpage

#### 4.4 Critical Assessment of the Results

In the first step of assessment, the 127 residues of protein complexes whose calculated sequence identities using PISCES sequence culling server [63] more than 35% are eliminated. After removing 127 residues, the resulting data set have 230 hot spots and 846 non-hot spots.

We have assessed the success of the machine-learning based methods, four servers and a plugin for a program by comparing accuracy (A), recall (R), precision (P), specificity (S) and F-measure (F1). A majority (78.62%) of our data set are non-hot spots. Therefore, precision and recall are relatively low while specificity is very high.

Weka [64] statistical results of support vector machines (SVM) classifier on 10-fold cross-validation with different random seed values are represented in Table 4.3. We change



random seed values to randomize our data. Standard deviations of results are very low in other words; the results are not variable while 10 folds are differed.

**Table 4.3** Results of Weka SVM classifier on 10-fold cross-validation.

Random seed for XVal / % Split	Precision	Recall	Specificity	Accuracy	F-measure
1	0.51	0.23	0.94	0.79	0.32
2	0.58	0.26	0.95	0.80	0.36
3	0.51	0.23	0.94	0.79	0.32
4	0.51	0.23	0.94	0.79	0.32
5	0.51	0.24	0.94	0.79	0.33
6	0.54	0.25	0.94	0.79	0.34
7	0.52	0.21	0.95	0.79	0.30
8	0.59	0.26	0.95	0.80	0.36
9	0.53	0.24	0.94	0.79	0.33
10	0.55	0.24	0.95	0.80	0.34
<b>Mean</b>	0.54	0.24	0.94	0.79	0.33
<b>Standard Deviation</b>	0.03	0.01	0.00	0.01	0.02

In Table 4.4, accuracy, precision, recall, specificity and F-measure are not changed remarkably, when we remove six physicochemical features from the study of Chen et al [20]. Naïve Bayes classifier gives the lowest accuracy (0.71) but at the same time, recall (0.69) and F-measure (0.50) are the highest among classifiers. Furthermore, all values of naïve Bayes classifier are greater than the values of random selection. Random forest classifier has the highest accuracy (0.80) and precision (0.55). This value of precision means nearly half of the hot spots are classified correctly but the precision value of random selection is 0.20 and all precision values in Table 4.4 are greater than 0.20. We have high non-hot spot rate as we explained before and for this reason, recalls are very low but interestingly they are lower than the recall of random selection. On the other hand, for the

same reason, specificities are very high and they are higher than the specificity of random selection. This high rate of non-hot spots indirectly causes low F-measures excepting naïve Bayes classifier that are presented in the last column of Table 4.4. Nevertheless, all f-measures are greater than the f-measure of random selection.

**Table 4.4** Results of different classifiers on 10-fold cross-validation.

Classifier	Precision	Recall	Specificity	Accuracy	F-measure
<b>SVM (Mean)</b>	0.54	0.24	0.94	0.79	0.33
<b>SVM (64 features)</b>	0.53	0.23	0.94	0.79	0.32
<b>Random Forest</b>	0.55	0.36	0.92	0.80	0.44
<b>J48</b>	0.45	0.38	0.87	0.77	0.41
<b>Decision Table</b>	0.45	0.21	0.93	0.78	0.29
<b>Naïve Bayes</b>	0.40	0.69	0.71	0.71	0.50
<b>Random Selection</b>	0.20	0.50	0.50	0.50	0.29

We performed feature selection using Weka to determine relative importance of features. The four most discriminating features are side-chain average depth index of target residues (s\_ch\_ave\_dpx\_target), relative side-chain accessible surface area of target residues (rel\_s\_ch\_ASA\_target), polarity of intra-contact residues (polarity\_intra) and hydrophilicity from the work of Chen et al. (hydrophilicity\_rig). The results are in Table 4.5 and they are close to the results with whole feature set but SVM classifier did not give meaningful result with those four features.

**Table 4.5** The results of classifiers with four features.

Classifier	Precision	Recall	Specificity	Accuracy	F-measure
<b>Random Forest</b>	0.44	0.39	0.87	0.76	0.41
<b>J48</b>	0.56	0.27	0.94	0.80	0.36
<b>Decision Table</b>	0.49	0.18	0.95	0.79	0.26
<b>Naïve Bayes</b>	0.47	0.32	0.90	0.78	0.38

In the last step of our analysis, we run a plugin for a program and servers to compare the results of the machine-learning classifiers with theirs. The threshold for experimentally

observed change in binding free energy is 1.00 kcal/mol in the paper of Robetta server so we computed statistical rates for 1.00 and 2.00 kcal/mol thresholds for that server. The number of residues that have prediction results is represented in the second column of Table 4.6. In Table 4.6, all values except recall are greater than the values of random selection and this is the same as in Table 4.4. Moreover, BeAtMuSiC server has the highest accuracy (0.79) and it has the lowest recall (0.33) and F-measure (0.39) at the same time. Although the accuracy of BeAtMuSiC is the highest in Table 4.6; it is lower than the accuracy of random forest classifier. Furthermore, precision of KFC\_B (0.50) does not exceed the precision of random forest (0.55) and recall of Robetta ( $\geq 1.00$  kcal) (0.64) is lower than recall of naïve Bayes classifier (0.69). Specificities of three machine-learning based methods of Weka are greater than the specificities of methods in Table 4.6. The highest F-measure among the methods in Table 4.6 (KFC2\_A, 0.49) is also less than the highest F-measure of machine-learning based methods (naïve Bayes, 0.50) but none of them passes 0.50 because of high non-hot spot rate.

**Table 4.6** Results of different methods.

Name	Total	Precision	Recall	Specificity	Accuracy	F-measure
BeAtMuSiC	1203	0.48	0.33	0.91	0.79	0.39
Robetta ( $\geq 2.00$ kcal/mol)	1082	0.43	0.39	0.87	0.77	0.41
KFC2_B	895	0.50	0.40	0.86	0.75	0.45
HotPoint	1203	0.37	0.55	0.76	0.72	0.44
Yasara-FoldX	1203	0.36	0.53	0.77	0.72	0.43
Robetta ( $\geq 1.00$ kcal/mol)	1082	0.36	0.64	0.71	0.70	0.46
KFC2_A	897	0.41	0.59	0.71	0.68	0.49
Random Selection	1203	0.20	0.50	0.50	0.50	0.29

The prediction performances of some methods in the studies of Tuncbag et al. [42] and Zhu and Mitchell [26] are presented in Table 4.7. The prediction performances in table 4.7 are almost the same as in Table 4.6. In Table 4.6 and in the study of Zhu and Mitchell in Table

4.7, KFC2\_B has the highest precision value and HotPoint has the lowest value, KFC2\_A has higher recall value than KFC2\_B and HotPoint in Table 4.6. Furthermore, Robetta (Zhu and Mitchell) has the highest specificity value in Table 4.7 and Robetta ( $\geq 2.00$  kcal/mol) in Table 4.6 has also the highest specificity value. For the value of accuracy, the highest and the lowest values are different but KFC2\_B has greater value than KFC2\_A in two tables. KFC2\_A has the highest and Robetta (in Table 4.6, Robetta ( $\geq 2.00$  kcal/mol)) has the lowest F-measure values in two tables. In the study of Tuncbag et al., the recall and F-measure values of HotPoint are greater than Robetta (Table 4.7) and the recall and F-measure values of HotPoint is also greater than the values of Robetta ( $\geq 2.00$  kcal/mol) in Table 4.6. On the other hand, in Table 4.7, the precision value of HotPoint is greater than Robetta whereas Robetta ( $\geq 2.00$  kcal/mol) has higher precision value in Table 4.6.

**Table 4.7** Hot spot prediction performances on independent test set in two studies.

Source	Method	Precision	Recall	Specificity	Accuracy	F-measure
<b>Zhu and Mitchell</b>	KFC2_B	0.65	0.62	0.85	0.78	0.63
	KFC2_A	0.57	0.85	0.71	0.75	0.68
	Robetta	0.52	0.33	0.86	0.70	0.41
	HotPoint	0.49	0.62	0.71	0.68	0.55
<b>Tuncbag et al.</b>	Robetta	0.63	0.57	-	-	0.60
	HotPoint	0.73	0.59	-	-	0.65

Comparing our study with previous studies on hot spots prediction is difficult because data sets, hot spot definitions, features, models and evaluation measures are different. Besides, to compare power of prediction based only on the obtained results is not entirely fair. Nonetheless, for completeness we gathered the cross validation results of recent studies into Table 4.8 and results of the methods on training and testing sets into Table 4.9.

**Table 4.8** Cross validation results of different methods

<b>Cross validation</b>	<b>Sets</b>	<b>Testing</b>	<b>P</b>	<b>R</b>	<b>S</b>	<b>A</b>	<b>F1</b>	<b>MCC</b>
<b>MINERVA<sup>a</sup></b>	Training 2	10-fold cross validation	0.73	0.58	0.89		0.65	
<b>RF<sup>b</sup></b>		10-fold cross validation	0.68	0.51	0.93	0.82		0.48
<b>Chen et al.<sup>c</sup></b>	Training 1	10-fold cross validation	0.65	0.65			0.65	
<b>Chen et al.<sup>c</sup></b>	Training 2	10-fold cross validation	0.78	0.80			0.79	
<b>KFC<sup>d</sup></b>		16-fold ‘‘leave one protein complex out’’ cross-validation	0.49	0.58			0.53	
<b>KFCA<sup>d</sup></b>		16-fold ‘‘leave one protein complex out’’ cross-validation	0.44	0.72			0.55	
<b>HSPred<sup>e</sup></b>		16-fold cross validation	0.61	0.69			0.65	0.54
<b>KFC2a</b>		standard leave one out cross validation	0.77	0.78	0.78	0.78	0.78	0.56
<b>KFC2b</b>		standard leave one out cross validation	0.80	0.55	0.87	0.71	0.65	0.44
<b>sbSVM<sup>f</sup></b>	Training 1	10-fold cross validation	0.50	0.82	0.74	0.76	0.62	
<b>sbSVM<sup>f</sup></b>	Training 2	10-fold cross validation	0.73	0.89	0.68	0.79	0.80	
<b>DBAC<sup>g</sup></b>		leave-one-out cross-validation	0.67	0.58	0.93	0.86	0.62	

<sup>a</sup>MINERVA, an acronym of MINE Residue Value, by Cho et al. [15]

<sup>b</sup>RF, Random Forest Model of Wang et al.[21]

<sup>c</sup>Results of the method of Chen et al.[20]

<sup>d</sup>Results of the method of Darnell et al.[22]

<sup>e</sup>Results of the method of Lise et al.[24]

<sup>f</sup>sbSVM, semi-supervised boosting SVM by Xu et al.[27]

<sup>g</sup>DBAC, an acronym of deeply buried atomic contacts, by Li et al.[28]

**Table 4.9** Results of different methods on training and testing sets

Evaluation on sets	Sets	P	R	S	A	F1	MCC
<b>MINERVA<sup>a</sup></b>	Training 1-Test	0.53	0.62	0.76		0.57	
<b>MINERVA<sup>a</sup></b>	Training 2-Test	0.65	0.44	0.90		0.52	
<b>Ye et al.</b>			0.45	0.94	0.79		0.47
<b>RF<sup>b</sup></b>		0.71	0.45	0.92	0.78		0.43
<b>Tuncbag et al.</b>	Training	0.64	0.52	0.82	0.70	0.57	
<b>Tuncbag et al.</b>	Test	0.73	0.59	0.79	0.70	0.65	
<b>Chen et al.<sup>c</sup></b>	Training 1-Test	0.69	0.68			0.68	
<b>Chen et al.<sup>c</sup></b>	Training 2-Test	0.65	0.64			0.64	
<b>KFC<sup>d</sup></b>		0.51	0.36			0.42	
<b>KFCA<sup>d</sup></b>		0.53	0.48			0.51	
<b>RS-MCLP<sup>h</sup></b>		0.18	0.76			0.28	0.27
<b>PCRPI<sup>i</sup></b>	Ab+ Training set-Test	0.79	0.64			0.71	
<b>KFC2a</b>		0.57	0.85	0.71	0.75	0.68	0.52
<b>KFC2b</b>		0.65	0.62	0.85	0.78	0.63	0.47
<b>sbSVM<sup>f</sup></b>	Training 1-Test	0.46	0.77	0.60	0.66	0.58	
<b>sbSVM<sup>f</sup></b>	Training 2-Test	0.51	0.82	0.64	0.70	0.63	
<b>APIS<sup>j</sup></b>		0.57	0.72	0.76		0.64	
<b>Ibk10<sup>k</sup></b>		0.65	0.92			0.76	

a, b, c, d, f are same with **Table 2.1**

<sup>h</sup>RS-MCLP, an acronym of rough set-based multiple criteria linear programming, by Chen et al. [23]

<sup>i</sup>PCRPI, an acronym of Presaging Critical Residues in Protein interfaces, by Assi et al. [25]

<sup>j</sup>APIS, a combined model based on Protrusion Index and Solvent accessibility, by Xia et al. [30]

<sup>k</sup>Ibk, an acronym of instance-based learning with parameter k and Ibk10 is a model which is ensemble of the ten most frequent classifiers by Chen et al. [31]

## Chapter 5

### CONCLUSION AND FUTURE DIRECTIONS

To investigate molecular mechanisms behind protein-protein interactions, the studies mainly focus on a specific area of protein-protein complexes, their interfaces. Hot spots are accounting only a small subset of all interface residues but they are essential for protein-protein recognition or binding. This important role makes them potential targets in drug design.

In this thesis, we introduced new benchmark data set for hot spot prediction. This benchmark data set is combination of 13 data sets and includes data of 1203 residues of 79 protein-protein complexes.

For hot spot prediction, machine-learning based methods are the commonly preferred computational methods. We critically assessed machine-learning method features and given results. Consequently, we defined seventy features that have strong impact on results. Several machine-learning methods (SVM, random forest, J48, decision table, naïve bayes) are applied using these features then the results of these methods are compared with results of other ML-based methods and servers. The results reveal that random forest classifier has the highest accuracy (80%) and although KFC2\_A has the highest F-measure (0.49) among existing methods, but it does not exceed the F-measure of naïve Bayes method (0.50).

The benchmark data set, features, prediction results of four servers and a plugin for a program are deposited in HotBase database. A web interface for the HotBase database is available at <http://prism.cccb.ku.edu.tr/hotbase> and allows users to browse the data of all residues and to download all data tables. For proceeding studies, we plan to improve database by adding values of new features to predict hot spots more accurately.

We believe that a better insight into protein-protein interactions will be developed by evaluating novel hot spot prediction models that uses values of same features on our new standard benchmark data set.

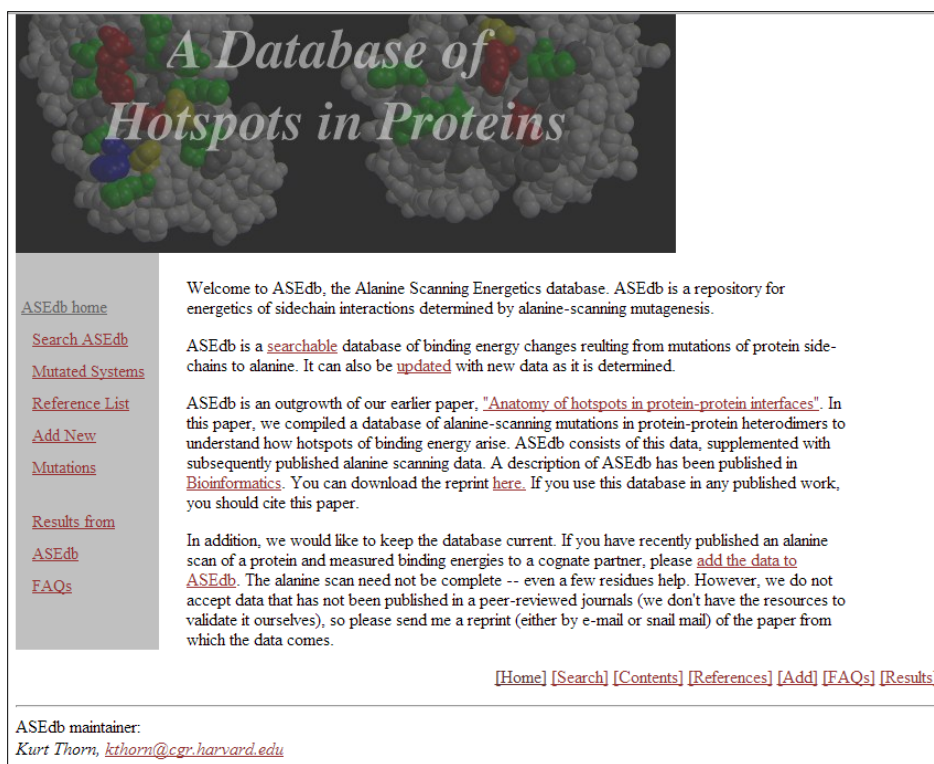


## Appendix A

### DATABASES FOR HOT SPOT PREDICTION

#### A.1 Alanine Scanning Energetics Database (ASEdb)

ASEdb [12] is searchable a database of binding free energy changes resulting from mutations in protein–protein, protein–nucleic acid, and protein–small molecule interactions to alanine. There are 1925  $\Delta\Delta G$  values of 46 monomers and 548  $\Delta\Delta G$  values of 23 dimers. ASEdb is currently not available, only MySQL dump of the data can be downloaded from <http://nic.ucsf.edu/asedb/>



*A Database of Hotspots in Proteins*

[ASEdb home](#)

[Search ASEdb](#)

[Mutated Systems](#)

[Reference List](#)

[Add New](#)

[Mutations](#)

[Results from ASEdb](#)

[FAQs](#)

Welcome to ASEdb, the Alanine Scanning Energetics database. ASEdb is a repository for energetics of sidechain interactions determined by alanine-scanning mutagenesis.

ASEdb is a [searchable](#) database of binding energy changes resulting from mutations of protein side-chains to alanine. It can also be [updated](#) with new data as it is determined.

ASEdb is an outgrowth of our earlier paper, "[Anatomy of hotspots in protein-protein interfaces](#)". In this paper, we compiled a database of alanine-scanning mutations in protein-protein heterodimers to understand how hotspots of binding energy arise. ASEdb consists of this data, supplemented with subsequently published alanine scanning data. A description of ASEdb has been published in [Bioinformatics](#). You can download the reprint [here](#). If you use this database in any published work, you should cite this paper.

In addition, we would like to keep the database current. If you have recently published an alanine scan of a protein and measured binding energies to a cognate partner, please [add the data to ASEdb](#). The alanine scan need not be complete -- even a few residues help. However, we do not accept data that has not been published in a peer-reviewed journals (we don't have the resources to validate it ourselves), so please send me a reprint (either by e-mail or snail mail) of the paper from which the data comes.

[\[Home\]](#) [\[Search\]](#) [\[Contents\]](#) [\[References\]](#) [\[Add\]](#) [\[FAQs\]](#) [\[Results\]](#)

ASEdb maintainer:  
Kurt Thorn, [kthorn@cgr.harvard.edu](mailto:kthorn@cgr.harvard.edu)

**Figure A.1** The main page of ASEdb database.

## A.2 Binding Interface Database (BID)

BID [49] is a protein-protein interaction database in which protein descriptions, interaction descriptions, bond formation, and the strength of each amino acid's contribution to binding are mined from literature and provided to users. Currently data on more than 467 interacting protein pairs are available with over 7000 hot spots documented. User can access BID database at <http://tsailab.chem.pacific.edu/wikiBID/>

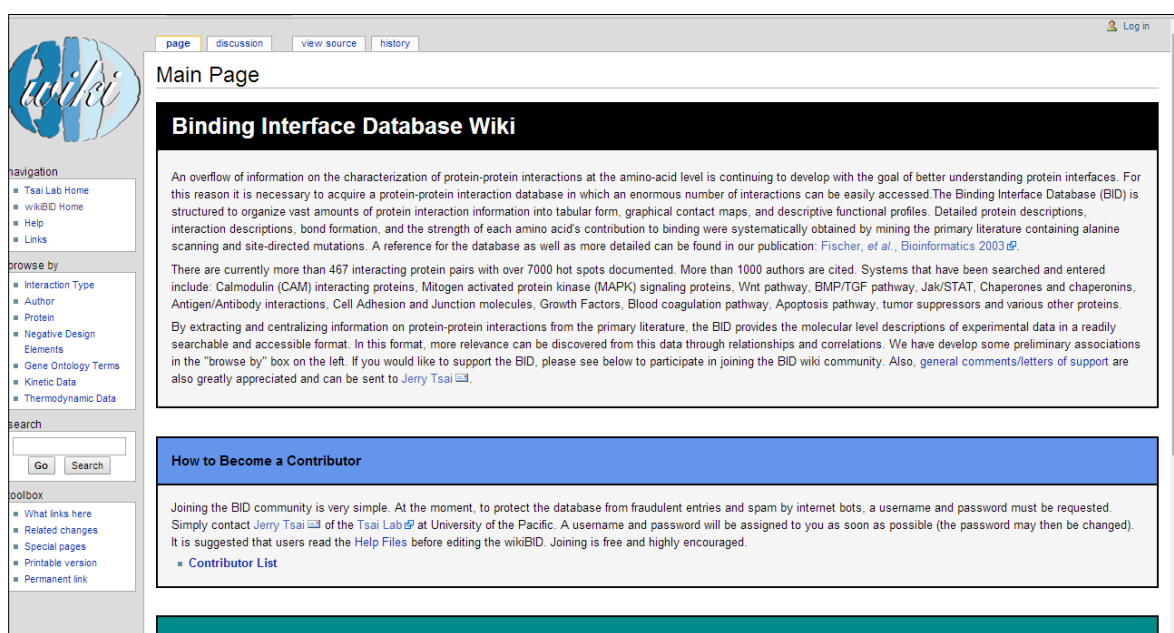
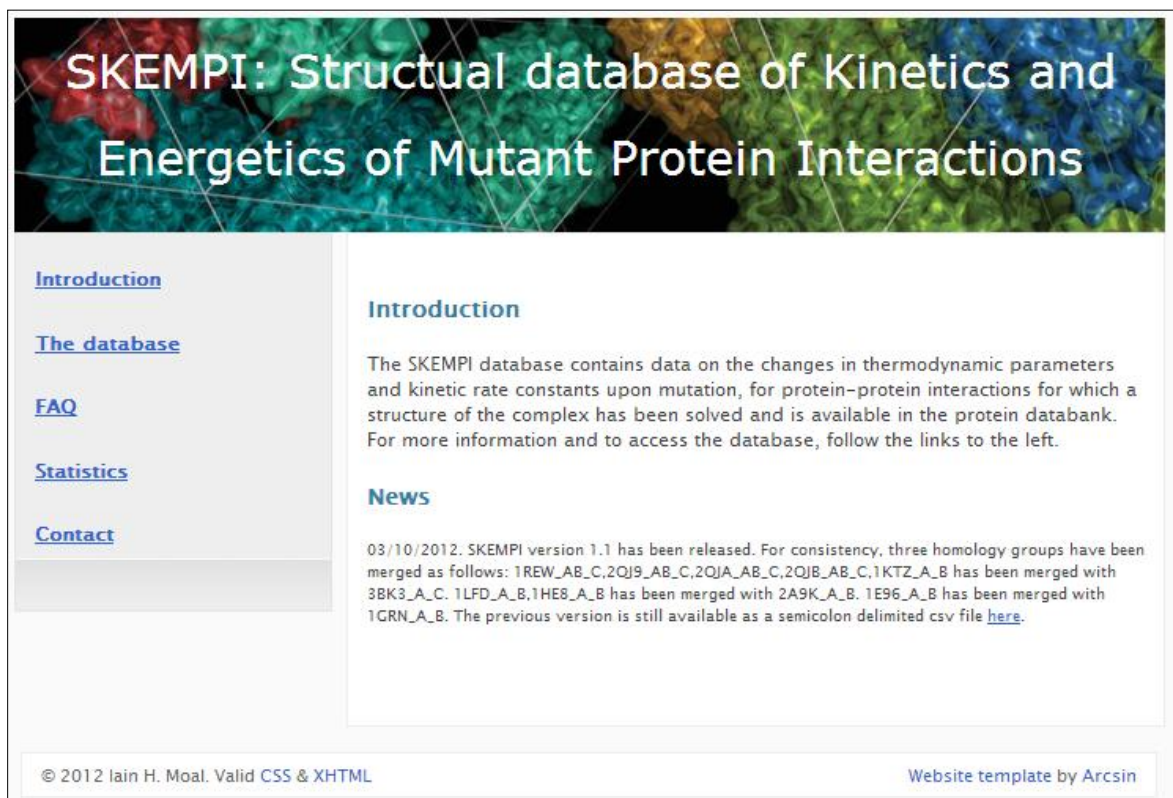


Figure A.2 The main page of WikiBID.

## A.3 Structural Kinetic and Energetic Database of Mutant Protein Interactions (SKEMPI)

SKEMPI database [58] contains 3,047 experimentally measured changes in binding free energy, entropy, enthalpy and rate constants upon mutation for 158 structures of 85 protein-protein complexes from the literature. SKEMPI database is available at [http://life.bsc.es/pid/mutation\\_database/](http://life.bsc.es/pid/mutation_database/)



**SKEMPI: Structural database of Kinetics and Energetics of Mutant Protein Interactions**

[Introduction](#)

[The database](#)

[FAQ](#)

[Statistics](#)

[Contact](#)

**Introduction**

The SKEMPI database contains data on the changes in thermodynamic parameters and kinetic rate constants upon mutation, for protein-protein interactions for which a structure of the complex has been solved and is available in the protein databank. For more information and to access the database, follow the links to the left.

**News**

03/10/2012. SKEMPI version 1.1 has been released. For consistency, three homology groups have been merged as follows: 1REW\_AB\_C,2QJ9\_AB\_C,2QJA\_AB\_C,2QJB\_AB\_C,1KTZ\_A\_B has been merged with 3BK3\_A\_C. 1LFD\_A\_B,1HE8\_A\_B has been merged with 2A9K\_A\_B. 1E96\_A\_B has been merged with 1GRN\_A\_B. The previous version is still available as a semicolon delimited csv file [here](#).

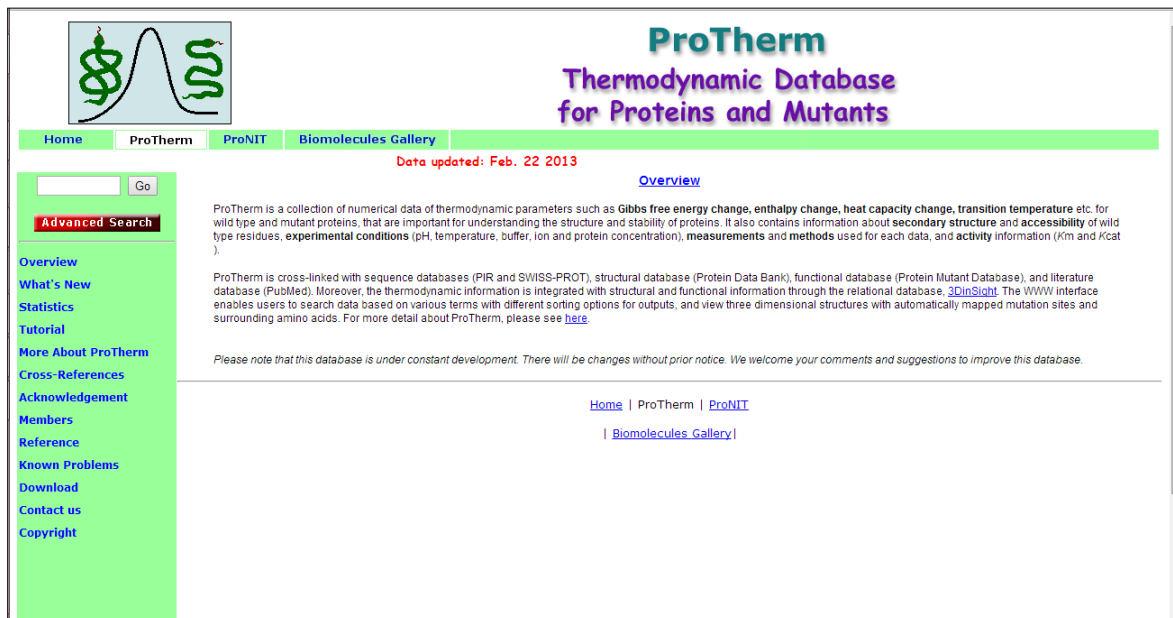
© 2012 Iain H. Moal. Valid CSS & XHTML

Website template by Arcsin

**Figure A.3** The main page of SKEMPI database.

#### A.4 Protherm

ProTherm [51] is a database that document experimentally determined thermodynamic parameters such as  $\Delta\Delta G$ , enthalpy change ( $\Delta H_{cal}$ ,  $\Delta H_{vH}$ ), heat capacity change ( $\Delta C_p$ ), transition temperature ( $T_m$ ) etc. for wild type and mutant proteins. Currently it contains 25,820 entries from 740 proteins. Database is available at <http://www.abren.net/protherm/>



The screenshot shows the main page of the ProTherm website. At the top left is a logo featuring a green caduceus and a black protein structure. To the right, the title "ProTherm" is displayed in large green letters, followed by "Thermodynamic Database for Proteins and Mutants" in purple. Below the title is a green navigation bar with links for "Home", "ProTherm", "ProNIT", and "Biomolecules Gallery". A red text notification states "Data updated: Feb. 22 2013". A search bar with a "Go" button and a red "Advanced Search" button is located on the left. A vertical green sidebar on the left contains a list of navigation links: Overview, What's New, Statistics, Tutorial, More About ProTherm, Cross-References, Acknowledgement, Members, Reference, Known Problems, Download, Contact us, and Copyright. The main content area contains an "Overview" link, a paragraph describing the database's scope (Gibbs free energy change, enthalpy change, heat capacity change, transition temperature, etc.), and a note about its integration with other databases (PIR, SWISS-PROT, Protein Data Bank, Protein Mutant Database, PubMed, 3DinSight). A footer section includes a disclaimer and navigation links for Home, ProTherm, ProNIT, and Biomolecules Gallery.

**Figure A.4** The main page of ProTherm

## Appendix B

## ASEdb

ASEdb is currently not available and the data in ASEdb are presented in **Table B.1**

**Table B.1** Data of ASEdb Database

Mutated Protein	Partner (Type)	Pdb	AA	Pos.	DDG	Monomer	Complex	Delta	RefID
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	R	5	2.3	141.48	41.35	100.13	50
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	H	8	0.9	109.67	85.62	24.05	6
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	Q	12	0.3	37.86	15.09	22.77	6
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	H	13	-0.3	13.26	8.56	4.7	49
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	R	31	0.2	157.22	37.3	119.92	50
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	R	32	0.9	164.82	102.66	62.16	50
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	R	33	0.3	46.54	46.54	0	50
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	R	66	0.2	130.8	130.8	0	50
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	N	68	0.2	95.5	84.38	11.12	6
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	R	70	-0.2	45.34	45.34	0	50
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	H	84	0.2	98.57	34.6	63.97	67
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	W	89	0.2	171.99	87.22	84.77	67
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	E	108	-0.3	71.77	27.87	43.9	6
Angiogenin	Rnase Inhibitor (P)	1a4y(D)	H	114	0.65	87.25	24.63	62.62	49
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	W	261	0.1	89.01	57.01	32	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	W	263	1.2	68.99	3.03	65.96	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	E	287	0.1	40.97	40.97	0	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	S	289	0	7.54	-	-	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	W	318	1.5	48.83	6.3	42.53	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	K	320	-0.3	76.64	37.17	39.47	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	E	344	0.2	39.8	23.79	16.01	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	W	375	1	61.26	4.67	56.59	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	E	401	0.9	33.16	13.57	19.59	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	Y	434	3.3	107.79	9.32	98.47	6
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	D	435	3.5	86.34	6.06	80.28	6
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	Y	437	0.8	173.71	30.33	143.38	6
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	R	457	-0.2	141.97	139.19	2.78	67
Rnase Inhibitor	Angiogenin (P)	1a4y(D)	I	459	0.7	49.06	25.2	23.86	67
Tissue Factor	Fab 5g9 (P)	1ahw(D)	Y	156	>4.000	70.67	4.79	65.88	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	T	167	0	47.71	6.21	41.5	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	T	170	1	18.22	6.93	11.29	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	L	176	1	67.39	67.39	0	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	D	178	-0.5	65.12	50.8	14.32	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	T	197	1.3	99.93	91.06	8.87	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	V	198	-0.3	88.51	81.36	7.15	24
Tissue Factor	Fab 5g9 (P)	1ahw(D)	N	199	1.1	52.07	51.39	0.68	24
Barstar	Barstar (P)	1brs(D)	K	27	5.4	75.69	21.06	54.63	48
Barstar	Barstar (P)	1brs(D)	Y	29	3.4	148.57	62.42	86.15	48
Barstar	Barstar (P)	1brs(D)	D	35	4.5	112.8	1.8	111	48
Barstar	Barstar (P)	1brs(D)	D	39	7.7	85.37	4.47	80.9	48
Barstar	Barstar (P)	1brs(D)	T	42	1.8	37.8	21.81	15.99	48
Barstar	Barstar (P)	1brs(D)	D	54	-0.8	30.15	30.15	0	47
Barstar	Barstar (P)	1brs(D)	N	58	3.1	18.1	15.48	2.62	47
Barstar	Barstar (P)	1brs(D)	R	59	5.2	177.78	33.55	144.23	48
Barstar	Barstar (P)	1brs(D)	E	60	-0.2	107.71	42.28	65.43	47
Barstar	Barstar (P)	1brs(D)	E	73	2.8	18.82	9.35	9.47	47
Barstar	Barstar (P)	1brs(D)	E	76	1.3	75.01	53.31	21.7	48
Barstar	Barstar (P)	1brs(D)	E	80	0.5	97.91	97.91	0	48
Barstar	Barstar (P)	1brs(D)	R	87	5.5	5.75	0.59	5.16	48
Barstar	Barstar (P)	1brs(D)	H	102	6	96.17	0.9	95.27	47
Im9	E9 Dnase (P)	1bxi(D)	C	23	0.92	19.77	9.16	10.61	56
Im9	E9 Dnase (P)	1bxi(D)	N	24	0.14	79.19	72.78	6.41	56
Im9	E9 Dnase (P)	1bxi(D)	D	26	0.34	92.17	92.17	0	56
Im9	E9 Dnase (P)	1bxi(D)	T	27	0.73	12.82	7.49	5.33	56
Im9	E9 Dnase (P)	1bxi(D)	S	28	0.17	71.24	71.23	0.01	56
Im9	E9 Dnase (P)	1bxi(D)	S	29	0.96	42.49	35.93	6.56	56
Im9	E9 Dnase (P)	1bxi(D)	E	30	1.41	124.5	23.99	100.51	56
Im9	E9 Dnase (P)	1bxi(D)	E	31	0.31	123.14	123.14	0	56
Im9	E9 Dnase (P)	1bxi(D)	E	32	0.22	85.05	85.04	0.01	56
Im9	E9 Dnase (P)	1bxi(D)	L	33	3.42	33.91	0.51	33.4	56
Im9	E9 Dnase (P)	1bxi(D)	V	34	2.58	86.31	24.81	61.5	56
Im9	E9 Dnase (P)	1bxi(D)	K	35	0.19	106.6	106.6	0	56
Im9	E9 Dnase (P)	1bxi(D)	L	36	0.91	45.85	45.84	0.01	56
Im9	E9 Dnase (P)	1bxi(D)	V	37	1.66	16.2	0	16.2	56
Im9	E9 Dnase (P)	1bxi(D)	T	38	0.9	54.97	35.73	19.24	56
Im9	E9 Dnase (P)	1bxi(D)	E	41	2.08	52.49	19.65	32.84	56
Im9	E9 Dnase (P)	1bxi(D)	E	42	0.66	93.05	93.05	0	56
Im9	E9 Dnase (P)	1bxi(D)	T	44	0.3	0.51	0.51	0	56
Im9	E9 Dnase (P)	1bxi(D)	E	45	0.21	78.58	78.58	0	56
Im9	E9 Dnase (P)	1bxi(D)	S	48	0.01	34.8	32.66	2.14	56
Im9	E9 Dnase (P)	1bxi(D)	G	49	1.49	10.4	1.79	8.61	56
Im9	E9 Dnase (P)	1bxi(D)	S	50	2.19	33.41	0.38	33.03	56
Im9	E9 Dnase (P)	1bxi(D)	D	51	5.92	63.64	27.01	36.63	56
Im9	E9 Dnase (P)	1bxi(D)	Y	55	4.63	151.8	61.17	90.63	56
Im9	E9 Dnase (P)	1bxi(D)	P	56	1.24	26.74	21.33	5.41	56
Im9	E9 Dnase (P)	1bxi(D)	D	60	0.51	46.17	46.17	0	56
Im9	E9 Dnase (P)	1bxi(D)	S	63	0.87	41.97	41.97	0	56
Im9	E9 Dnase (P)	1bxi(D)	N	69	0.28	83.78	83.78	0	56

Bpti	Chymotrypsin (P)	1cbw(D)	T	11	0.2	61.29	23.43	37.86	4
Bpti	Chymotrypsin (P)	1cbw(D)	P	13	-0.1	65.8	19.53	46.27	4
Bpti	Chymotrypsin (P)	1cbw(D)	K	15	2	147.63	0	147.63	4
Bpti	Chymotrypsin (P)	1cbw(D)	R	17	0.5	175.76	24.65	151.11	4
Bpti	Chymotrypsin (P)	1cbw(D)	I	19	0.1	101.73	63.63	38.1	4
Bpti	Chymotrypsin (P)	1cbw(D)	R	20	0.3	38.26	38.26	0	4
Bpti	Chymotrypsin (P)	1cbw(D)	V	34	0	60.11	46.06	14.05	4
Bpti	Chymotrypsin (P)	1cbw(D)	R	39	0.2	159.06	94.28	64.78	4
Bpti	Chymotrypsin (P)	1cbw(D)	K	46	0.1	133.87	133.87	0	4
Cd2	Cd48 (P)	1cdm(D)	D	28	>1.740	17.45	-	-	13
Cd2	Cd48 (P)	1cdm(D)	E	29	>1.740	8.42	-	-	13
Cd2	Cd48 (P)	1cdm(D)	R	31	>1.740	8.57	-	-	13
Cd2	Cd48 (P)	1cdm(D)	E	33	0.24	26.99	-	-	13
Cd2	Cd48 (P)	1cdm(D)	R	34	-0.06	59.5	-	-	13
Cd2	Cd48 (P)	1cdm(D)	L	38	>1.740	3.05	-	-	13
Cd2	Cd48 (P)	1cdm(D)	E	41	0.06	0.18	-	-	13
Cd2	Cd48 (P)	1cdm(D)	K	43	0.4	19.55	-	-	13
Cd2	Cd48 (P)	1cdm(D)	F	49	>1.740	32.36	-	-	13
Cd2	Cd48 (P)	1cdm(D)	K	51	0.15	51.33	-	-	13
Cd2	Cd48 (P)	1cdm(D)	E	56	0.11	73.73	-	-	13
Cd2	Cd48 (P)	1cdm(D)	Y	81	>1.740	8.1	-	-	13
Cd2	Cd48 (P)	1cdm(D)	R	86	-0.21	56.6	-	-	13
Cd2	Cd48 (P)	1cdm(D)	R	87	-0.94	62.18	-	-	13
Tissue Factor	Factor Viaa (P)	1dan(D)	K	15	-0.4	103.32	100	3.32	30
Tissue Factor	Factor Viaa (P)	1dan(D)	S	16	-0.13	2.11	2.11	0	44
Tissue Factor	Factor Viaa (P)	1dan(D)	T	17	0.1	58.71	33.02	25.69	30
Tissue Factor	Factor Viaa (P)	1dan(D)	N	18	0.2	12.6	1.33	11.27	30
Tissue Factor	Factor Viaa (P)	1dan(D)	K	20	2.6	88.35	12.46	75.89	30
Tissue Factor	Factor Viaa (P)	1dan(D)	T	21	-0.2	0	0	0	30
Tissue Factor	Factor Viaa (P)	1dan(D)	I	22	0.7	32.28	4.56	27.72	30
Tissue Factor	Factor Viaa (P)	1dan(D)	E	24	0.7	63.94	49.2	14.74	30
Tissue Factor	Factor Viaa (P)	1dan(D)	E	26	0.1	75.9	75.9	0	30
Tissue Factor	Factor Viaa (P)	1dan(D)	K	28	0.1	158.79	158.79	0	30
Tissue Factor	Factor Viaa (P)	1dan(D)	V	33	-0.2	23.52	23.52	0	45
Tissue Factor	Factor Viaa (P)	1dan(D)	V	36	-0.1	0	0	0	45
Tissue Factor	Factor Viaa (P)	1dan(D)	Q	37	0.55	32.19	4.89	27.3	30
Tissue Factor	Factor Viaa (P)	1dan(D)	I	38	-0.1	4.04	4.04	0	45
Factor VII	Tissue Factor (P)	1dan(D)	L	39	0	120.35	17.2	103.15	15
Tissue Factor	Factor Viaa (P)	1dan(D)	K	41	0.35	100.16	96.95	3.21	30
Tissue Factor	Factor Viaa (P)	1dan(D)	I	42	0	85.56	82.03	3.53	15
Tissue Factor	Factor Viaa (P)	1dan(D)	G	42	-0.1	81.27	63.17	18.1	30
Tissue Factor	Factor Viaa (P)	1dan(D)	S	43	0.2	21.75	21.75	0	45
Tissue Factor	Factor Viaa (P)	1dan(D)	D	44	0.7	118.68	67.52	51.16	30
Tissue Factor	Factor Viaa (P)	1dan(D)	W	45	1.6	72.32	23.22	49.1	45
Tissue Factor	Factor Viaa (P)	1dan(D)	K	46	0.25	91.11	52	39.11	30
Tissue Factor	Factor Viaa (P)	1dan(D)	S	47	0.05	31.57	8.63	22.94	30
Tissue Factor	Factor Viaa (P)	1dan(D)	K	48	0.4	37.75	14.92	22.83	30
Tissue Factor	Factor Viaa (P)	1dan(D)	F	50	0.4	109.15	7.97	101.18	30
Tissue Factor	Factor Viaa (P)	1dan(D)	Y	51	-0.1	134.82	100.71	34.11	45
Tissue Factor	Factor Viaa (P)	1dan(D)	T	52	0.4	21.66	21.66	0	30
Tissue Factor	Factor Viaa (P)	1dan(D)	D	58	2.18	48.42	5.23	43.19	30
Tissue Factor	Factor Viaa (P)	1dan(D)	L	59	0	0.05	0.05	0	44
Tissue Factor	Factor Viaa (P)	1dan(D)	T	60	2.3	4.59	4.59	0	45
Tissue Factor	Factor Viaa (P)	1dan(D)	D	61	0.24	72.22	21.96	50.26	44
Tissue Factor	Factor Viaa (P)	1dan(D)	K	62	0	119.99	48	71.99	15
Tissue Factor	Factor Viaa (P)	1dan(D)	E	62	0	11.83	10.75	1.08	44
Tissue Factor	Factor Viaa (P)	1dan(D)	I	63	0	0	0	0	44
Factor VII	Tissue Factor (P)	1dan(D)	Q	64	0.8	70.22	9.7	60.52	15
Tissue Factor	Factor Viaa (P)	1dan(D)	V	64	0	10.84	10.84	0	44
Factor VII	Tissue Factor (P)	1dan(D)	S	67	0	48.68	48		

Tissue Factor	Factor Viiia (P)	1dan(D)	E	128	0.1	78.92	64.43	14.49	30
Tissue Factor	Factor Viiia (P)	1dan(D)	D	129	0	36.6	36.6	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	R	131	0	133.85	89.44	44.41	44
Tissue Factor	Factor Viiia (P)	1dan(D)	T	132	0	0	0	0	44
Tissue Factor	Factor Viiia (P)	1dan(D)	L	133	0	49.06	7.5	41.56	30
Tissue Factor	Factor Viiia (P)	1dan(D)	R	135	0.55	113.13	44.28	68.85	30
Tissue Factor	Factor Viiia (P)	1dan(D)	R	136	-0.06	142.66	142.66	0	44
Tissue Factor	Factor Viiia (P)	1dan(D)	N	138	0	104.27	88.43	15.84	44
Tissue Factor	Factor Viiia (P)	1dan(D)	T	139	0	66.56	66.56	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	F	140	1.5	35.2	21.89	13.31	30
Tissue Factor	Factor Viiia (P)	1dan(D)	R	144	0	44.24	44.24	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	D	145	0	65.61	65.61	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	V	146	0.2	8.19	8.19	0	44
Tissue Factor	Factor Viiia (P)	1dan(D)	F	147	-0.06	0.18	0.18	0	44
Tissue Factor	Factor Viiia (P)	1dan(D)	I	152	0.2	31.44	31.44	0	30
Factor Vii	Tissue Factor (P)	1dan(D)	K	157	0	156.82	156.82	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	Y	157	0	24.94	24.94	0	44
Factor Vii	Tissue Factor (P)	1dan(D)	V	158	0	48.44	48.44	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	K	159	0.2	61.29	0	61.29	30
Factor Vii	Tissue Factor (P)	1dan(D)	P	160	0	71.67	71.67	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	K	161	0	57.45	57.45	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	163	0	44.44	44.44	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	S	163	0	92.76	68.16	24.6	30
Tissue Factor	Factor Viiia (P)	1dan(D)	T	167	0.2	52.83	52.83	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	K	169	0.1	113.85	113.85	0	30
Factor Vii	Tissue Factor (P)	1dan(D)	L	171	0	36.4	36.4	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	V	172	0	26.91	26.91	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	T	172	0	72	72	0	30
Factor Vii	Tissue Factor (P)	1dan(D)	N	173	0	107.94	107.94	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	176	0	35.62	35.62	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	L	176	0.1	75.02	75.02	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	V	179	0.11	3.12	3.12	0	44
Tissue Factor	Factor Viiia (P)	1dan(D)	K	181	0	163.65	163.65	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	Y	185	-0.35	37.26	37.26	0	30
Tissue Factor	Factor Viiia (P)	1dan(D)	S	195	0	61.98	61.98	0	30
Factor Vii	Tissue Factor (P)	1dan(D)	D	196	0	26.34	26.34	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	R	196	0.46	5.61	5.61	0	44
Factor Vii	Tissue Factor (P)	1dan(D)	K	197	0	165.34	165.34	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	T	197	0.11	101.71	101.71	0	44
Factor Vii	Tissue Factor (P)	1dan(D)	I	198	0	17.59	17.59	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	V	198	0.11	87.4	87.4	0	44
Factor Vii	Tissue Factor (P)	1dan(D)	K	199	0	151.26	151.26	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	N	199	0	63.61	63.61	0	44
Factor Vii	Tissue Factor (P)	1dan(D)	N	200	0	94.19	94.19	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	202	0	206.19	206.19	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	N	203	0	54.98	54.98	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	T	203	0.1	74.77	49.16	25.61	30
Factor Vii	Tissue Factor (P)	1dan(D)	I	205	0	48.59	48.59	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	V	207	0	10.16	10.16	0	15
Tissue Factor	Factor Viiia (P)	1dan(D)	V	207	-0.2	54.98	0	54.98	30
Tissue Factor	Factor Viiia (P)	1dan(D)	E	208	0	74.1	48.28	25.82	30
Factor Vii	Tissue Factor (P)	1dan(D)	E	210	0	10.75	10.75	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	212	0	25.25	25.25	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	L	213	0	21.9	21.9	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	S	214	0	47.65	47.65	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	215	0	99.12	99.12	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	H	216	0	140.24	140.24	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	217	0	42.49	42.49	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	219	0	38.45	38.45	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	220	0	14.81	14.81	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	221	0	28.39	28.39	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	223	0	48.03	48.03	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	224	0	126.6	126.6	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	227	0	8.46	8.46	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	S	232	0	25.91	25.91	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	V	235	0	85.3	85.3	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	T	238	0	63.09	63.09	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	T	239	0	9.25	9.25	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	N	240	0	38.7	38.7	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	247	0	37.52	37.52	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	H	249	0	105.49	105.49	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	250	0	102	102	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	256	0	61.72	61.72	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	266	0	81.98	81.98	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	270	0	61.5	61.5	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	271	0	70.37	69.04	1.33	15
Factor Vii	Tissue Factor (P)	1dan(D)	T	272	0	11.9	11.9	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	F	275	0	108.51	37.36	71.15	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	277	0.51	158.84	38.21	120.63	15
Factor Vii	Tissue Factor (P)	1dan(D)	F	278	0	81.89	56.05	25.84	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	286	0	27.55	27.55	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	L	288	0	113.21	113.21	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	289	0	57.87	57.21	0.66	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	290	0	204.81	204.81	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	T	293	0	49.14	49.14	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	L	295	0	106.22	106.22	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	296	0	66.65	66.65	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	M	298	0	25.56	25.56	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	N	301	0	59.62	59.62	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	P	303	0.6	17.97	17.97	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	304	0.65	18.9	16.05	2.85	15
Factor Vii	Tissue Factor (P)	1dan(D)	L	305	0.7	23.75	23.75	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	M	306	0.5	122.47	11.67	110.8	15
Factor Vii	Tissue Factor (P)	1dan(D)	T	307	0	30.35	24.32	6.03	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	308	0	128.6	55.79	72.81	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	309	0.41	57.67	15.21	42.46	15
Factor Vii	Tissue Factor (P)	1dan(D)	L	311	0.3	91.67	84.04	7.63	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	312	0	107.76	102.8	4.96	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	313	0	88.19	88.19	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	S	314	0	18.49	4.01	14.48	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	315	0	164.72	164.72	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	K	316	-0.54	170.95	170.95	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	V	317	0	39.65	39.65	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	319	0	111.46	111.46	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	325	0	109.33	105.43	3.9	15
Factor Vii	Tissue Factor (P)	1dan(D)	Y	332	0	57.07	57.07	0	15

Factor Vii	Tissue Factor (P)	1dan(D)	S	333	0	26.3	26.3	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	D	334	0	73.92	73.92	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	S	336	0	42.36	42.36	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	K	337	0	75.58	75.58	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	K	341	0	91.09	86.08	5.01	15
Factor Vii	Tissue Factor (P)	1dan(D)	H	351	0	72.35	72.35	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	353	0	75.15	75.15	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	W	356	0	36.08	36.08	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	366	0	36.38	21.53	14.85	15
Factor Vii	Tissue Factor (P)	1dan(D)	T	370	0	42.09	42.09	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	V	371	0	97.76	97.76	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	379	0.51	52.48	21.61	30.87	15
Factor Vii	Tissue Factor (P)	1dan(D)	Q	382	0	23.6	23.6	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	385	0	116.3	116.3	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	392	0	161.04	161.04	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	E	394	0	100.47	100.47	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	R	396	0	137.47	137.47	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	V	399	0	31.82	31.82	0	15
Factor Vii	Tissue Factor (P)	1dan(D)	L	400	0	56.88	56.88	0	15
Rnase Inhibitor	Rnase A (P)	1dfj(D)	E	206	1	47.04	22.96	24.08	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	W	261	1.3	46.57	7.39	39.18	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	W	263	2.2	53.13	9.68	43.45	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	E	287	1.3	50.71	26.81	23.9	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	S	289	0.8	2.22	0.04	2.18	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	W	318	1	76.46	21.74	54.72	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	K	320	1.3	89.92	47.75	42.17	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	E	344	1.6	43.06	37.97	5.09	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	E	401	1.3	44.08	17.74	26.34	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	Y	434	5.9	113.76	6.04	107.72	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	D	435	3.6	85.89	23.16	62.73	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	Y	437	2.6	165.03	6.76	158.27	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	R	457	0.8	163.65	135.03	28.62	67
Rnase Inhibitor	Rnase A (P)	1dfj(D)	I	459	0.3	49.16	13.87	35.29	67
D1.3	E5.2 (P)	1dvr(D)	D	H100	2.8	79.41	32.46	46.95	11
E5.2	D1.3 (P)	1dvr(D)	Q	H100	1.6	97.98	7.4	90.58	19
E5.2	D1.3 (P)	1dvr(D)	R	H100a	4.1	102.18	7.79	94.39	19
D1.3	E5.2 (P)	1dvr(D)	Y	H101	>4.000	90.62	22.6	68.02	11
D1.3	E5.2 (P)	1dvr(D)	T	H30	0.9	63.98	45.95	18.03	11
E5.2	D1.3 (P)	1							



Appendix B: ASEdb

Hgh	Hghbp_Site? (P)	3hhr(D)	N	12	0.1	59.14	0	59.14	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	M	14	0.1	29.43	29.43	0	10
Hgh	Hghbp_Site? (P)	3hhr(D)	L	15	0.15	106.35	35.9	70.45	8
Hgh	Hghbp_Site? (P)	3hhr(D)	R	16	0.24	93.51	12.74	80.77	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	H	18	-0.5	103.16	37.92	65.24	10
Hgh	Hghbp_Site? (P)	3hhr(D)	R	19	0.05	64.6	20.23	44.37	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	H	21	0.2	28.35	2.65	25.7	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	Q	22	-0.2	58.68	44.46	14.22	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	F	25	-0.4	95.28	32.28	63	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	D	26	-0.2	47.64	47.53	0.11	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	Q	29	-0.6	79.3	76.97	2.33	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	39	0.27	46.28	46.28	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	P	41	0.26	22.73	22.73	0	3
Hgh	Hghbp_Site1 (P)	3hhr(D)	Y	42	0.2	135.59	42.94	92.65	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	42	1.22	18.04	18.04	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	43	2.2	35.8	7.04	28.76	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	44	1.8	65.48	37	28.48	7
Hgh	Hghbp_Site1 (P)	3hhr(D)	L	45	1.2	58.3	0.57	57.73	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	Q	46	0.1	127.18	36.56	90.62	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	S	47	-0.01	14.42	14.42	0	3
Hgh	Hghbp_Site1 (P)	3hhr(D)	P	48	0.4	93.65	71.86	21.79	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	S	51	0.3	18.03	8.68	9.35	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	T	51	0.07	37.49	37.49	0	3
Hgh	Hghbp_Site? (P)	3hhr(D)	F	54	0.87	28.64	28.64	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	S	55	0.11	0.31	0.31	0	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	E	56	0.4	59.95	58.71	1.24	10
Hgh	Hghbp_Site? (P)	3hhr(D)	S	57	0.2	50.8	50.8	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	I	58	1.64	0	0	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	P	59	0.38	61.96	61.96	0	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	P	61	1.2	3.18	3.18	0	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	S	62	0.2	64.67	12.66	52.01	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	N	63	0.3	71.09	32.54	38.55	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	R	64	1.6	178.63	63.47	115.16	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	E	65	-0.5	122.67	121.48	1.19	10
Hgh	Hghbp_Site? (P)	3hhr(D)	E	66	0.43	62.11	62.11	0	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	Q	68	0.6	52.81	27.6	25.21	10
Hgh	Hghbp_Site? (P)	3hhr(D)	Q	69	-0.05	105.78	105.78	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	K	70	0.52	57.11	57.11	0	8
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	70	0.41	49.73	49.73	0	3
Hgh	Hghbp_Site? (P)	3hhr(D)	S	71	0.41	31.28	31.28	0	8
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	71	0.59	21.84	21.84	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	N	72	0.2	68.1	58.78	9.32	7
Hgh	Hghbp_Site? (P)	3hhr(D)	L	73	-0.2	1.02	1.02	0	8
Hghbp_Site 1	Hgh (P)	3hhr(D)	T	73	0.1	0	0	0	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	Q	74	0	-	-	-	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	75	-0.1	73.7	73.7	0	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	W	76	0.6	87.63	31.66	55.97	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	T	77	-0.25	58.34	33.89	24.45	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	Q	78	-0.41	61.9	61.9	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	79	-0.08	39.21	39.21	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	W	80	0	60.25	59.98	0.27	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	81	0.14	122.17	122.17	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	82	0.77	61.4	61.4	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	91	0.18	136.78	136.78	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	Y	95	0.2	53.62	53.62	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	N	97	-0.25	41.88	41.88	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	S	98	-0.1	60.79	59.76	1.03	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	S	99	-0.5	71.72	71.72	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	F	100	-0.01	69.52	69.52	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	S	102	-0.2	27.91	2.73	25.18	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	I	103	1.8	7.84	1.53	6.31	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	W	104	>4.500	140.93	1.71	139.22	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	I	105	2	37.76	25.14	12.62	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	C	108	0	10	0.03	9.97	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	110	0.04	51.59	51.59	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	120	-0.2	78.39	55.84	22.55	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	121	0.1	81.63	56.57	25.06	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	C	122	0	38.03	0.11	37.92	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	S	124	0.2	11.28	0.08	11.2	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	V	125	1.13	0	0	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	D	126	1	29.22	2.37	26.85	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	127	1	100.58	30.53	70.05	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	I	128	0.58	42.85	42.85	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	V	129	-0.05	1.21	1.21	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	Q	130	-0.03	54.3	54.3	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	D	132	1.07	35.83	35.83	0	3
Hgh	Hghbp_Site1 (P)	3hhr(D)	Y	164	0.3	26.67	1.69	24.98	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	D	164	1.6	39.74	22.82	16.92	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	I	165	2.2	64.73	54.57	10.16	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	Q	166	0	133.31	91.28	42.03	7
Hgh	Hghbp_Site1 (P)	3hhr(D)	R	167	0.3	46.4	14.47	31.93	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	167	0	135.32	97.7	37.62	7
Hgh	Hghbp_Site1 (P)	3hhr(D)	K	168	-0.2	64.69	7.44	57.25	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	W	169	>4.500	88.24	0.39	87.85	7
Hgh	Hghbp_Site1 (P)	3hhr(D)	D	171	0.8	62.72	11.69	51.03	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	K	172	2	44.95	13.16	31.79	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	173	0.08	25.68	25.68	0	3
Hgh	Hghbp_Site1 (P)	3hhr(D)	E	174	-0.9	34.42	10.48	23.94	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	V	174	-0.7	6.03	6.03	0	7
Hgh	Hghbp_Site1 (P)	3hhr(D)	T	175	2	57.8	9.02	48.78	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	175	0.35	20.28	20.28	0	3
Hgh	Hghbp_Site1 (P)	3hhr(D)	F	176	1.9	6.41	0	6.41	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	R	178	2.4	104.91	37.99	66.92	10
Hgh	Hghbp_Site1 (P)	3hhr(D)	I	179	0.8	32.18	0.48	31.7	10
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	179	0	19.78	19.78	0	3
Hgh	Hghbp_Site? (P)	3hhr(D)	V	180	0	0	0	0	8
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	180	0.2	23.18	23.18	0	3
Hgh	Hghbp_Site? (P)	3hhr(D)	Q	181	0.27	12.06	12.06	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	C	182	1.01	28.18	1.46	26.72	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	R	183	0.5	62.68	62.34	0.34	10
Hgh	Hghbp_Site? (P)	3hhr(D)	S	184	-0.05	7.27	7.27	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	V	185	0.87	22.66	22.66	0	8
Hgh	Hghbp_Site1 (P)	3hhr(D)	E	186	0	107.65	102.43	5.22	10
Hgh	Hghbp_Site? (P)	3hhr(D)	G	187	0.34	26.43	26.43	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	S	188	-0.2	14.36	14.36	0	8
Hgh	Hghbp_Site? (P)	3hhr(D)	F	191	0.19	-	-	-	8

Hghbp_Site 1	Hgh (P)	3hhr(D)	T	194	0.2	78.42	78.42	0	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	T	195	-0.1	48.33	48.33	0	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	203	0.24	74.89	74.89	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	D	205	0.37	87.36	87.36	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	206	0.2	52.02	52.02	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	207	-0.08	43.18	43.18	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	209	-0.06	21.64	21.64	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	211	0.06	65.39	65.39	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	213	-0.19	55.61	55.61	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	K	215	0.49	39.19	39.19	0	3
Hghbp_Site 1	Hgh (P)	3hhr(D)	Q	216	0.9	14.83	8.96	5.87	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	R	217	0.2	90	54.07	35.93	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	N	218	0.3	120	12.54	107.46	7
Hghbp_Site 1	Hgh (P)	3hhr(D)	E	224	0.11	124.06	124.06	0	3
Hgh	Mab 1 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 2 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 3 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 10 (P)	3hhr(M)	F	1	0	177.51	-	-	27
Hgh	Mab 1 (P)	3hhr(M)	P	2	0	106.95	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	P	2	0	106.95	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	P	2	0	106.95	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	P	2	0	106.95	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	P	2	0	106.95	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	P	2	0	106.95	-	-	27







Hgh	Mab 20 (P)	3hhr(M)	Q	40	0	59.67	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	Q	40	0	59.67	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	Q	40	0.64	59.67	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	Q	40	0.41	59.67	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	Q	40	0	59.67	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	Y	42	0	135.59	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	Y	42	1.76	135.59	-	-	27
Hghbp	Mab1288 (P)	3hhr(M)	E	42	-0.52	18.04	-	-	3
Hghbp	Mab13e1 (P)	3hhr(M)	E	42	-0.81	18.04	-	-	3
Hghbp	Mab263 (P)	3hhr(M)	E	42	-0.11	18.04	-	-	3
Hghbp	Mab367 (P)	3hhr(M)	E	42	-0.54	18.04	-	-	3
Hghbp	Mab349 (P)	3hhr(M)	E	42	-0.15	18.04	-	-	3
Hghbp	Mab5 (P)	3hhr(M)	E	42	-0.2	18.04	-	-	3
Hgh	Mab 11 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	S	43	0	47.23	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	S	43	0.41	47.23	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	F	44	0	1.78	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	F	44	0	1.78	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	F	44	0	1.78	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	F	44	0	1.78	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	F	44	0	1.78	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	F	44	0	1.78	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	F	44	>3.640	1.78	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	F	44	>3.640	1.78	-	-	27
Hghbp	Mab1288 (P)	3hhr(M)	E	44	-0.36	65.48	-	-	3
Hghbp	Mab13e1 (P)	3hhr(M)	E	44	-0.1	65.48	-	-	3
Hghbp	Mab263 (P)	3hhr(M)	E	44	0.09	65.48	-	-	3
Hghbp	Mab367 (P)	3hhr(M)	E	44	0.28	65.48	-	-	3
Hghbp	Mab349 (P)	3hhr(M)	E	44	0.24	65.48	-	-	3
Hghbp	Mab5 (P)	3hhr(M)	E	44	0.52	65.48	-	-	3
Hgh	Mab 11 (P)	3hhr(M)	K	45	0	58.3	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	K	45	0	58.3	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	K	45	0	58.3	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	K	45	0	58.3	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	K	45	1.76	58.3	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	K	45	0.94	58.3	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	K	45	>3.640	58.3	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	Q	46	0	127.18	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	Q	46	0	127.18	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	Q	46	0	127.18	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	Q	46	0	127.18	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	Q	46	0	127.18	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	Q	46	1.35	127.18	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	Q	46	0	127.18	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	Q	46	>3.640	127.18	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	N	47	0	63.07	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	N	47	>3.640	63.07	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	P	48	0	93.65	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	P	48	0	93.65	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	P	48	0	93.65	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	P	48	0	93.65	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	P	48	0	93.65	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	P	48	0	93.65	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	P	48	>3.640	93.65	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	Q	49	0.64	128.24	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	Q	49	0	128.24	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	Q	49	0	128.24	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	Q	49	0	128.24	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	Q	49	0	128.24	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	Q	49	0	128.24	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	Q	49	0	128.24	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	Q	49	0.41	128.24	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	T	50	0.64	13.77	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab10 (P)	3hhr(M)	T	50	0	13.77	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	T	50	0.64	13.77	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	P	61	0.41	3.18	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	P	61	0.41	3.18	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab10 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab10 (P)	3hhr(M)	T	60	0	9.74	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	P	61	0.41	3.18	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	P	61	0.41	3.18	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab10 (P)	3hhr(M)	P	61	0	3.18	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab10 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Mab10 (P)	3hhr(M)	S	62	0	64.67	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	N	63	0.85	71.09	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	N	63	0	71.09	-	-	27

Hgh	Mab 6 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab10 (P)	3hhr(M)	S	51	0	18.03	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	L	52	0	108.75	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	L	52</					

Hgh	Mab 12 (P)	3hhr(M)	N	63	0	71.09	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	N	63	0.41	71.09	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	N	63	0.41	71.09	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	N	63	0	71.09	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	N	63	0	71.09	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	N	63	>3.640	71.09	-	-	27
Hgh	Mab10 (P)	3hhr(M)	N	63	0	71.09	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	R	64	0.35	178.63	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	R	64	0	178.63	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	R	64	2.16	178.63	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	R	64	0.41	178.63	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	R	64	0.41	178.63	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	R	64	0.41	178.63	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	R	64	1.35	178.63	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	R	64	3.64	178.63	-	-	27
Hgh	Mab10 (P)	3hhr(M)	R	64	0	178.63	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	E	65	0.54	122.67	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	E	65	0	122.67	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	E	65	0.64	122.67	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	E	65	0	122.67	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	E	65	0	122.67	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	E	65	0	122.67	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	E	65	0	122.67	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	E	65	0	122.67	-	-	27
Hgh	Mab10 (P)	3hhr(M)	E	65	0.81	122.67	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	E	66	0.07	62.11	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	E	66	0	62.11	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	E	66	0	62.11	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	E	66	0.41	62.11	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	E	66	0.41	62.11	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	E	66	0	62.11	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	E	66	0	62.11	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	E	66	0.41	62.11	-	-	27
Hgh	Mab10 (P)	3hhr(M)	E	66	0	62.11	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	Q	68	0.11	52.81	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	Q	68	1.14	52.81	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Mab10 (P)	3hhr(M)	Q	68	0	52.81	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	Q	69	-0.21	105.78	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	Q	69	0	105.78	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	Q	69	0	105.78	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	Q	69	0	105.78	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	Q	69	0	105.78	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	Q	69	0.41	105.78	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	Q	69	0.41	105.78	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	Q	69	0	105.78	-	-	27
Hgh	Mab10 (P)	3hhr(M)	Q	69	0.41	105.78	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	K	70	0.24	57.11	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	K	70	1.35	57.11	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Mab10 (P)	3hhr(M)	K	70	0	57.11	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	S	71	0.61	31.28	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Mab10 (P)	3hhr(M)	S	71	0	31.28	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	L	73	0.35	1.02	-	-	9
Hgh	Mab 11 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	L	73	0.41	1.02	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab10 (P)	3hhr(M)	L	73	0	1.02	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab10 (P)	3hhr(M)	E	74	0	56.26	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	L	82	0	0	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	L	82	0	0	-	-	27
Hghbp	Mab1288 (P)	3hhr(M)	E	82	-0.82	61.4	-	-	3
Hghbp	Mab13e1 (P)	3hhr(M)	E	82	-0.74	61.4	-	-	3
Hghbp	Mab263 (P)	3hhr(M)	E	82	1.28	61.4	-	-	3
Hghbp	Mab3b7 (P)	3hhr(M)	E	82	0.2	61.4	-	-	3
Hghbp	Mab3d9 (P)	3hhr(M)	E	82	0.16	61.4	-	-	3
Hghbp	Mab5 (P)	3hhr(M)	E	82	0.21	61.4	-	-	3
Hgh	Mab 8 (P)	3hhr(M)	I	83	0	0	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	I	83	0	0	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	Q	84	0	30.6	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	Q	84	0	30.6	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	S	85	0	0.57	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	S	85	0	0.57	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	L	87	0	30.45	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	L	87	0	30.45	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	L	87	0	30.45	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	L	87	0	30.45	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	E	88	0	119.79	-	-	27

Hgh	Mab 21 (P)	3hhr(M)	E	88	0	119.79	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	E	88	1.35	119.79	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	E	88	1.22	119.79	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	P	89	0	3.74	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	P	89	0	3.74	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	P	89	0	3.74	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	P	89	0	3.74	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	P	89	2.16	3.74	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	P	89	0	3.74	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	V	90	0	2.6	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	V	90	0	2.6	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	V	90	0	2.6	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	V	90	0	2.6	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	V	90	0	2.6	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	V	90	0	2.6	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	Q	91	0	93.53	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	Q	91	0	93.53	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	Q	91	0	93.53	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	Q	91	0	93.53	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	Q	91	0	93.53	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	Q	91	0	93.53	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	F	92	0	80.99	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	F	92	0	80.99	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	F	92	0	80.99	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	F	92	0	80.99	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	F	92	0	80.99	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	F	92	0	80.99	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	L	93	0	0	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	L	93	0	0	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	L	93	0	0	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	L	93	0	0	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	L	93	0	0	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	L	93	0	0	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	R	94	0	133.09	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	R	94	0	133.09	-	-	27
Hgh	Mab 2 (P)	3hhr(M)	R	94	0	133.09	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	R	94	0	133.09	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	R	94	0	133.09	-	-	27
Hgh	Mab 3 (P)	3hhr(M)	R	94	0	133.09	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	R	94	2.29	133.09	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	R	94	2.29	133.09	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 2 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 3 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	S	95	0	56.06	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	S	95	>3.640	56.06	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	V	96	0	13.33	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	V	96	0	13.33	-	-	27
Hgh	Mab 2 (P)	3hhr(M)	V	96	0	13.33	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	V	96	0	13.33	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	V	96	0	13.33	-	-	27

Hghbp	Mab13e1 (P)	3hr(M)	W	104	-0.02	140.93	-	-	3
Hghbp	Mab263 (P)	3hr(M)	W	104	0	140.93	-	-	3
Hghbp	Mab3b7 (P)	3hr(M)	W	104	0.2	140.93	-	-	3
Hghbp	Mab3d9 (P)	3hr(M)	W	104	0	140.93	-	-	3
Hghbp	Mab5 (P)	3hr(M)	W	104	0.36	140.93	-	-	3
Hgh	Mab 1 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 18 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 19 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 2 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 20 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 21 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 3 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 8 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 9 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab10 (P)	3hr(M)	S	106	0	37.98	-	-	27
Hgh	Mab 1 (P)	3hr(M)	D	107	0	47.34	-	-	27
Hgh	Mab 15 (P)	3hr(M)	D	107	0	47.34	-	-	27
Hgh	Mab 2 (P)	3hr(M)	D	107	0	47.34	-	-	27
Hgh	Mab 3 (P)	3hr(M)	D	107	0	47.34	-	-	27
Hgh	Mab 8 (P)	3hr(M)	D	107	0.94	47.34	-	-	27
Hgh	Mab 9 (P)	3hr(M)	D	107	>3.640	47.34	-	-	27
Hgh	Mab10 (P)	3hr(M)	D	107	0	47.34	-	-	27
Hgh	Mab 1 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab 15 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab 2 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab 3 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab 8 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab 9 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab10 (P)	3hr(M)	S	108	0	73.29	-	-	27
Hgh	Mab 1 (P)	3hr(M)	N	109	0	43.86	-	-	27
Hgh	Mab 15 (P)	3hr(M)	N	109	0	43.86	-	-	27
Hgh	Mab 2 (P)	3hr(M)	N	109	0	43.86	-	-	27
Hgh	Mab 3 (P)	3hr(M)	N	109	0	43.86	-	-	27
Hgh	Mab 8 (P)	3hr(M)	N	109	0	43.86	-	-	27
Hgh	Mab 9 (P)	3hr(M)	N	109	>3.640	43.86	-	-	27
Hgh	Mab10 (P)	3hr(M)	N	109	0	43.86	-	-	27
Hgh	Mab 1 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab 15 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab 2 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab 3 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab 8 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab 9 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab10 (P)	3hr(M)	V	110	0	0	-	-	27
Hgh	Mab 1 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab 15 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab 2 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab 3 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab 8 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab 9 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab10 (P)	3hr(M)	Y	111	0	84.9	-	-	27
Hgh	Mab 1 (P)	3hr(M)	D	112	0	84.46	-	-	27
Hgh	Mab 15 (P)	3hr(M)	D	112	0	84.46	-	-	27
Hgh	Mab 2 (P)	3hr(M)	D	112	>3.640	84.46	-	-	27
Hgh	Mab 3 (P)	3hr(M)	D	112	1.76	84.46	-	-	27
Hgh	Mab 8 (P)	3hr(M)	D	112	0	84.46	-	-	27
Hgh	Mab 9 (P)	3hr(M)	D	112	>3.640	84.46	-	-	27
Hgh	Mab10 (P)	3hr(M)	D	112	0	84.46	-	-	27
Hgh	Mab 1 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab 15 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab 2 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab 3 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab 8 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab 9 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab10 (P)	3hr(M)	L	113	0	12.99	-	-	27
Hgh	Mab 1 (P)	3hr(M)	K	115	0	80.6	-	-	27
Hgh	Mab 15 (P)	3hr(M)	K	115	0	80.6	-	-	27
Hgh	Mab 2 (P)	3hr(M)	K	115	0	80.6	-	-	27
Hgh	Mab 3 (P)	3hr(M)	K	115	0	80.6	-	-	27
Hgh	Mab 8 (P)	3hr(M)	K	115	0.41	80.6	-	-	27
Hgh	Mab 9 (P)	3hr(M)	K	115	0	80.6	-	-	27
Hgh	Mab10 (P)	3hr(M)	K	115	0	80.6	-	-	27
Hgh	Mab 1 (P)	3hr(M)	D	116	0	59.73	-	-	27
Hgh	Mab 15 (P)	3hr(M)	D	116	0	59.73	-	-	27
Hgh	Mab 2 (P)	3hr(M)	D	116	>3.640	59.73	-	-	27
Hgh	Mab 3 (P)	3hr(M)	D	116	2.49	59.73	-	-	27
Hgh	Mab 8 (P)	3hr(M)	D	116	0	59.73	-	-	27
Hgh	Mab 9 (P)	3hr(M)	D	116	>3.640	59.73	-	-	27
Hgh	Mab10 (P)	3hr(M)	D	116	0	59.73	-	-	27
Hgh	Mab 1 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab 15 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab 2 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab 3 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab 8 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab 9 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab10 (P)	3hr(M)	L	117	0	1.57	-	-	27
Hgh	Mab 1 (P)	3hr(M)	E	118	0	18.14	-	-	27
Hgh	Mab 15 (P)	3hr(M)	E	118	0	18.14	-	-	27
Hgh	Mab 2 (P)	3hr(M)	E	118	0	18.14	-	-	27
Hgh	Mab 3 (P)	3hr(M)	E	118	0	18.14	-	-	27
Hgh	Mab 8 (P)	3hr(M)	E	118	0	18.14	-	-	27
Hgh	Mab 9 (P)	3hr(M)	E	118	1.35	18.14	-	-	27
Hgh	Mab10 (P)	3hr(M)	E	118	0	18.14	-	-	27
Hgh	Mab 1 (P)	3hr(M)	E	119	0	117.78	-	-	27
Hgh	Mab 15 (P)	3hr(M)	E	119	0	117.78	-	-	27
Hgh	Mab 2 (P)	3hr(M)	E	119	0	117.78	-	-	27
Hgh	Mab 3 (P)	3hr(M)	E	119	0	117.78	-	-	27
Hgh	Mab 8 (P)	3hr(M)	E	119	0.41	117.78	-	-	27
Hgh	Mab 9 (P)	3hr(M)	E	119	2.64	117.78	-	-	27
Hgh	Mab10 (P)	3hr(M)	E	119	0	117.78	-	-	27
Hgh	Mab 1 (P)	3hr(M)	I	121	0	0	-	-	27
Hgh	Mab 15 (P)	3hr(M)	I	121	0	0	-	-	27
Hgh	Mab 2 (P)	3hr(M)	I	121	0	0	-	-	27
Hgh	Mab 3 (P)	3hr(M)	I	121	0	0	-	-	27
Hgh	Mab 8 (P)	3hr(M)	I	121	0	0	-	-	27
Hgh	Mab 9 (P)	3hr(M)	I	121	0	0	-	-	27
Hgh	Mab10 (P)	3hr(M)	I	121	0	0	-	-	27

Hgh	Mab 1 (P)	3hr(M)	Q	122	0	80.34	-	-	27
Hgh	Mab 15 (P)	3hr(M)	Q	122	0	80.34	-	-	27
Hgh	Mab 2 (P)	3hr(M)	Q	122	0	80.34	-	-	27
Hgh	Mab 3 (P)	3hr(M)	Q	122	0	80.34	-	-	27
Hgh	Mab 8 (P)	3hr(M)	Q	122	0.41	80.34	-	-	27
Hgh	Mab 9 (P)	3hr(M)	Q	122	0	80.34	-	-	27
Hgh	Mab10 (P)	3hr(M)	Q	122	0	80.34	-	-	27
Hgh	Mab 1 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab 15 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab 2 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab 3 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab 8 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab 9 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab10 (P)	3hr(M)	T	123	0	49.56	-	-	27
Hgh	Mab 1 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab 15 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab 2 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab 3 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab 8 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab 9 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab10 (P)	3hr(M)	L	124	0	1.05	-	-	27
Hgh	Mab 1 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab 15 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab 2 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab 3 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab 8 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab 9 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab10 (P)	3hr(M)	M	125	0	17.37	-	-	27
Hgh	Mab12b8 (P)	3hr(M)	V	125	-0.55	0	-	-	3
Hghbp	Mab13e1 (P)	3hr(M)	V	125	-0.74	0	-	-	3
Hghbp	Mab263 (P)	3hr(M)	V	125	-0.24	0	-	-	3
Hghbp	Mab3b7 (P)	3hr(M)	V	125	0	0	-	-	3
Hghbp	Mab3d9 (P)	3hr(M)	V	125	0.22	0	-	-	3
Hghbp	Mab5 (P)	3hr(M)	V	125	0.08	0	-	-	3
Hgh	Mab 1 (P)	3hr(M)	G	126	0	35.72	-	-	27
Hgh	Mab 15 (P)	3hr(M)	G	126	0	35.72	-	-	27
Hgh	Mab 2 (P)	3hr(M)	G	126	0	35.72	-	-	27
Hgh	Mab 3 (P)	3hr(M)	G	126	0	35.72	-	-	27
Hgh	Mab 8 (P)	3hr(M)	G	126	0	35.72	-	-	27
Hgh	Mab 9 (P)	3hr(M)	G	126	0	35.72	-	-	27
Hgh	Mab10 (P)	3hr(M)	G	126	1.29	35.72	-	-	27
Hgh	Mab12b8 (P)	3hr(M)	D	126	-0.26	29.22	-	-	3
Hghbp	Mab13e1 (P)	3hr(M)	D	126	-0.37	29.22	-	-	3
Hghbp	Mab263 (P)	3hr(M)	D	126	0	29.22	-	-	3
Hghbp	Mab3b7 (P)	3hr(M)	D	126	0.2	29.22	-	-	3
Hghbp	Mab3d9 (P)	3hr(M)	D	126	0.34	29.22	-	-	3
Hghbp	Mab5 (P)	3hr(M)	D	126	0.21	29.22	-	-	3
Hgh	Mab 1 (P)	3hr(M)	R	127	3.11	69.22	-	-	27
Hgh	Mab 15 (P)	3hr(M)	R	127	>4.050	69.22	-	-	27
Hgh	Mab 2 (P)	3hr(M)	R	127	0	69.22	-	-	27
Hgh	Mab 3 (P)	3hr(M)	R	127	0	69.22	-	-	27
Hgh	Mab 8 (P)	3hr(M)	R	127	0	69.22	-	-	27
Hgh	Mab 9 (P)	3hr(M)	R	127	0	69.22	-	-	27
Hgh	Mab10 (P)	3hr(M)	R	127	>3.640	69.22	-	-	27
Hgh	Mab 1 (P)	3hr(M)	L	128	0	8.51	-	-	27
Hgh	Mab 15 (P)	3hr(M)	L	128	0	8.51	-	-	





Hgh	Mab 1 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	I	179	2.16	32.18	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	I	179	2.29	32.18	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Mab10 (P)	3hhr(M)	I	179	0	32.18	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	V	180	-0.29	0	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	V	180	0.64	0	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Mab10 (P)	3hhr(M)	V	180	0	0	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	Q	181	0	12.06	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Mab10 (P)	3hhr(M)	Q	181	0	12.06	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	R	183	0.56	62.68	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	R	183	>3.640	62.68	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	R	183	>3.640	62.68	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	R	183	>3.640	62.68	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	R	183	>3.640	62.68	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	R	183	2.7	62.68	-	-	27
Hgh	Mab 18 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 19 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 20 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 21 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	R	183	1.76	62.68	-	-	27
Hgh	Mab10 (P)	3hhr(M)	R	183	0	62.68	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	S	184	-0.12	7.27	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Mab10 (P)	3hhr(M)	S	184	0	7.27	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	V	185	0.28	22.66	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	V	185	0.94	22.66	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	V	185	0.64	22.66	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Mab10 (P)	3hhr(M)	V	185	0	22.66	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	E	186	0	107.65	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	E	186	0.41	107.65	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Mab10 (P)	3hhr(M)	E	186	0	107.65	-	-	27
Hgh	Hprlbp +Zn (P)	3hhr(M)	G	187	0	26.43	-	-	9
Hgh	Mab 1 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	G	187	0.94	26.43	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	G	187	0.64	26.43	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	G	187	0.64	26.43	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab10 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab 17 (P)	3hhr(M)	G	187	0	26.43	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 10 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 3 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 2 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 1 (P)	3hhr(M)	S	188	0	14.36	-	-	27
Hgh	Mab 16 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 15 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 14 (P)	3hhr(M)	F	191	0.41	-	-	-	27
Hgh	Mab 13 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 12 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 11 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 10 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 9 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 8 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 7 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 6 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 5 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 4 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 3 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 2 (P)	3hhr(M)	F	191	0	-	-	-	27
Hgh	Mab 1 (P)	3hhr(M)	F	191	0	-	-	-	27
Il-8	Il-8r (P)	3i8(M)	E	4	2.7	-	-	-	22
Il-8	Il-8r (P)	3i8(M)	L	5	2.7	120.35	-	-	22
Il-8	Il-8r (P)	3i8(M)	R	6	4.05	199.51	-	-	22
Il-8	Il-8r (P)	3i8(M)	I	10	1.99	142.2	-	-	22
Il-2 (Human)	Il-2 Receptor (P)	3ink(M)	R	38	1.58	168.27	-	-	46
Il-2 (Human)	Il-2 Receptor (P)	3ink(M)	F	42	1.96	83.98	-	-	46
Bfgf	Fgfr1b (P)	4fgf(M)	R	22	0.18	52.53	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	Y	24	2.86	59.67	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	K	26	0.01	61.74	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	R	44	0.71	125.88	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	K	46	0.07	130.79	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	E	96	4.33	39.96	-	-	61
Bfgf	Fgfr1b (P)	4fgf(M)	R	97	0.41	142.17	-	-	61
Bfgf	Fgfr1b (P)	4fgf(M)	N	101	1.16	65.47	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	Y	103	3.06	38.3	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	R	107	0.6	75.39	-	-	61
Bfgf	Fgfr1b (P)	4fgf(M)	R	109	0.22	98.57	-	-	61
Bfgf	Fgfr1b (P)	4fgf(M)	K	110	-0.33	72.57	-	-	61
Bfgf	Fgfr1b (P)	4fgf(M)	Y	111	0.17	36.66	-	-	53
Bfgf	Fgfr1b (P)	4fgf(M)	T	112	0.51	63.			



---

**BIBLIOGRAPHY**

1. Keskin O, Tuncbag N, Gursoy A: **Characterization and prediction of protein interfaces to infer protein-protein interaction networks.** *Curr Pharm Biotechnol* 2008, **9**:67–76.
2. Yan C, Wu F, Jernigan RL, Dobbs D, Honavar V: **Characterization of Protein-Protein Interfaces.** *Protein J* 2008, **27**:59–70.
3. Chen J, Ma X, Yuan Y, Pei J, Lai L: **Protein-Protein Interface Analysis and Hot Spots Identification for Chemical Ligand Design.** *Curr Pharm Des* 2013.
4. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R: **Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect.** *Protein Sci Publ Protein Soc* 1997, **6**:53–64.
5. Tsai CJ, Xu D, Nussinov R: **Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes.** *Protein Sci Publ Protein Soc* 1997, **6**:1793–1805.
6. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES: **Are protein-protein interfaces more conserved in sequence than the rest of the protein surface?** *Protein Sci Publ Protein Soc* 2004, **13**:190–202.
7. Cole C, Warwicker J: **Side-chain conformational entropy at protein-protein interfaces.** *Protein Sci Publ Protein Soc* 2002, **11**:2860–2870.
8. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N: **Residue frequencies and pairing preferences at protein-protein interfaces.** *Proteins Struct Funct Bioinforma* 2001, **43**:89–102.
9. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces.** *J Mol Biol* 1998, **280**:1–9.
10. Clackson T, Wells JA: **A hot spot of binding energy in a hormone-receptor interface.** *Science* 1995, **267**:383–386.

11. DeLano WL: **Unraveling hot spots in binding interfaces: progress and challenges.** *Curr Opin Struct Biol* 2002, **12**:14–20.
12. Thorn KS, Bogan AA: **ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions.** *Bioinformatics* 2001, **17**:284–285.
13. Li J, Liu Q: **“Double water exclusion”: a hypothesis refining the O-ring theory for the hot spots at protein interfaces.** *Bioinformatics* 2009, **25**:743–750.
14. Keskin O, Ma B, Nussinov R: **Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues.** *J Mol Biol* 2005, **345**:1281–1294.
15. Cho K, Kim D, Lee D: **A feature-based approach to modeling protein–protein interaction hot spots.** *Nucleic Acids Res* 2009, **37**:2672–2687.
16. Del Sol A, O’Meara P: **Small-world network approach to identify key residues in protein-protein interaction.** *Proteins* 2005, **58**:672–682.
17. Kozakov D, Hall DR, Chuang G-Y, Cencic R, Brenke R, Grove LE, Beglov D, Pelletier J, Whitty A, Vajda S: **Structural conservation of druggable hot spots in protein–protein interfaces.** *Proc Natl Acad Sci* 2011.
18. Ye L, Kuang Q, Jiang L, Luo J, Jiang Y, Ding Z, Li Y, Li M: **Prediction of hot spots residues in protein–protein interface using network feature and microenvironment feature.** *Chemom Intell Lab Syst* 2014, **131**:16–21.
19. Kortemme T, Kim DE, Baker D: **Computational Alanine Scanning of Protein-Protein Interfaces.** *Sci Signal* 2004, **2004**:pl2.
20. Chen R, Chen W, Yang S, Wu D, Wang Y, Tian Y, Shi Y: **Rigorous assessment and integration of the sequence and structure based features to predict hot spots.** *BMC Bioinformatics* 2011, **12**:311.
21. Wang L, Liu Z-P, Zhang X-S, Chen L: **Prediction of hot spots in protein interfaces using a random forest model with hybrid features.** *Protein Eng Des Sel PEDS* 2012, **25**:119–126.
22. Darnell SJ, Page D, Mitchell JC: **An automated decision-tree approach to predicting protein interaction hot spots.** *Proteins Struct Funct Bioinforma* 2007, **68**:813–823.

23. Chen R, Zhang Z, Wu D, Zhang P, Zhang X, Wang Y, Shi Y: **Prediction of protein interaction hot spots using rough set-based multiple criteria linear programming.** *J Theor Biol* 2011, **269**:174–180.
24. Lise S, Buchan D, Pontil M, Jones DT: **Predictions of hot spot residues at protein-protein interfaces using support vector machines.** *PLoS One* 2011, **6**:e16774.
25. Assi SA, Tanaka T, Rabbitts TH, Fernandez-Fuentes N: **PCRPI: Presaging Critical Residues in Protein interfaces, a new computational tool to chart hot spots in protein interfaces.** *Nucleic Acids Res* 2010, **38**:e86.
26. Zhu X, Mitchell JC: **KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features.** *Proteins* 2011, **79**:2671–2683.
27. Xu B, Wei X, Deng L, Guan J, Zhou S: **A semi-supervised boosting SVM for predicting hot spots at protein-protein Interfaces.** *BMC Syst Biol* 2012, **6**(Suppl 2):S6.
28. Li Z, Wong L, Li J: **DBAC: A simple prediction method for protein binding hot spots based on burial levels and deeply buried atomic contacts.** *BMC Syst Biol* 2011, **5**(Suppl 1):S5.
29. Ofran Y, Rost B: **Protein-protein interaction hotspots carved into sequences.** *PLoS Comput Biol* 2007, **3**:1169–1176.
30. Xia J-F, Zhao X-M, Song J, Huang D-S: **APIS: accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility.** *BMC Bioinformatics* 2010, **11**:174.
31. Chen P, Li J, Wong L, Kuwahara H, Huang JZ, Gao X: **Accurate prediction of hot spot residues through physicochemical characteristics of amino acid sequences.** *Proteins Struct Funct Bioinforma* 2013, **81**:1351–1362.
32. Mitchell T: **Machine Learning**. McGraw-Hill, New York. 1997.
33. Lise S, Archambeau C, Pontil M, Jones DT: **Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods.** *BMC Bioinformatics* 2009, **10**:365.
34. Kim DE, Chivian D, Baker D: **Protein structure prediction and analysis using the Robetta server.** *Nucleic Acids Res* 2004, **32**(suppl 2):W526–W531.

35. Guerois R, Nielsen JE, Serrano L: **Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations.** *J Mol Biol* 2002, **320**:369–387.
36. González-Ruiz D, Gohlke H: **Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding.** *Curr Med Chem* 2006, **13**:2607–2625.
37. Rajamani D, Thiel S, Vajda S, Camacho CJ: **Anchor residues in protein–protein interactions.** *Proc Natl Acad Sci U S A* 2004, **101**:11287–11292.
38. Erman B: **Relationships between ligand binding sites, protein architecture and correlated paths of energy and conformational fluctuations.** *ArXiv11050956 Cond-Mat Physicsphysics Q-Bio* 2011.
39. Ofra Y, Rost B: **ISIS: interaction sites identified from sequence.** *Bioinformatics* 2007, **23**:e13–e16.
40. Grosdidier S, Fernández-Recio J: **Identification of hot-spot residues in protein-protein interactions by computational docking.** *BMC Bioinformatics* 2008, **9**:447.
41. Guney E, Tuncbag N, Keskin O, Gursoy A: **HotSprint: database of computational hot spots in protein interfaces.** *Nucleic Acids Res* 2007, **36**(Database):D662–D666.
42. Tuncbag N, Gursoy A, Keskin O: **Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy.** *Bioinforma Oxf Engl* 2009, **25**:1513–1520.
43. Tuncbag N, Keskin O, Gursoy A: **HotPoint: hot spot prediction server for protein interfaces.** *Nucleic Acids Res* 2010, **38**(Web Server):W402–W406.
44. Dehouck Y, Kwasigroch JM, Rooman M, Gilis D: **BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations.** *Nucleic Acids Res* 2013, **41**(Web Server issue):W333–339.
45. Haliloglu T, Seyrek E, Erman B: **Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model.** *Phys Rev Lett* 2008, **100**:228102.
46. Ozbek P, Soner S, Haliloglu T: **Hot Spots in a Network of Functional Sites.** *PLoS ONE* 2013, **8**:e74320.

47. Wang L, Hou Y, Quan H, Xu W, Bao Y, Li Y, Fu Y, Zou S: **A compound-based computational approach for the accurate determination of hot spots: *In Silico* Determination of Hot Spots.** *Protein Sci* 2013, **22**:1060–1070.
48. Kortemme T, Baker D: **A simple physical model for binding energy hot spots in protein–protein complexes.** *Proc Natl Acad Sci* 2002, **99**:14116–14121.
49. Fischer TB, Arunachalam KV, Bailey D, Mangual V, Bakhru S, Russo R, Huang D, Paczkowski M, Lalchandani V, Ramachandra C: **The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces.** *Bioinformatics* 2003, **19**:1453–1454.
50. Moal IH, Fernández-Recio J: **SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models.** *Bioinformatics* 2012, **28**:2600–2607.
51. Kumar MDS, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A: **ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions.** *Nucleic Acids Res* 2006, **34**(Database issue):D204–D206.
52. Mihel J, Sikić M, Tomić S, Jeren B, Vlahovićek K: **PSAIA–protein structure and interaction analyzer.** *BMC Struct Biol* 2008, **8**:21.
53. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.** *Biopolymers* 1983, **22**:2577–2637.
54. Kawashima S, Kanehisa M: **AAindex: amino acid index database.** *Nucleic Acids Res* 2000, **28**:374.
55. Gershon Celniker, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N: **ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function.** *Isr J Chem* 2013, **53**:199–206.
56. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N: **ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures.** *Nucleic Acids Res* 2005, **33**(suppl 2):W299–W302.
57. Valdar WS, Thornton JM: **Conservation helps to identify biologically relevant crystal contacts.** *J Mol Biol* 2001, **313**:399–416.

- 
58. Keskin O, Bahar I, Jernigan RL, Badretdinov AY, Ptitsyn OB: **Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions.** *Protein Sci* 2008, **7**:2578–2586.
59. Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, Sander C, Vriend G: **A series of PDB related databases for everyday needs.** *Nucleic Acids Res* 2011, **39**(Database issue):D411–D419.
60. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389–3402.
61. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L: **The FoldX web server: an online force field.** *Nucleic Acids Res* 2005, **33**(Web Server issue):W382–W388.
62. Durme JV, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F: **A graphical interface for the FoldX forcefield.** *Bioinformatics* 2011, **27**:1711–1712.
63. Wang G, Dunbrack RL Jr: **PISCES: a protein sequence culling server.** *Bioinforma Oxf Engl* 2003, **19**:1589–1591.
64. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH: **The WEKA Data Mining Software: An Update.** *SIGKDD Explor Newsl* 2009, **11**:10–18.

## VITA

Selin Karagulle was born in Izmir, Turkey, on March 19, 1988. She graduated as valedictorian in Mathematics & Computer Science with a double major in Software Engineering from Bahcesehir University, Istanbul in 2011.

From September 2011 to September 2013, she worked as teaching and research assistant at Koç University, Istanbul, Turkey. She had a scholarship from TUBITAK (The Scientific and Technological Research Council of Turkey) for her graduate study. She has worked on “assessment of the methods and features used for hot spot residue prediction at the protein-protein interfaces”, “preparing benchmark data set for hot spot prediction” and “designing a database that includes features, methods and various server results” during her MSc study.

She has attended ISMB’13 (Berlin, Germany) where she has presented a poster about “Critical Assessment of the Methods and the Features Used for Hot Spot Residue Prediction at Protein-Protein Interfaces”.

She will start to work in IT industry but she may return to academia later.